

TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KİTLESEL AÇIK ÇEVİRİMİÇİ KURSLARDA
KATILIMCILARIN İNGİLİZCE DİL GRUPLARININ
TESPİTİNE DAYALI DAVRANIŞ VE PERFORMANS
ANALİZİ

İsmail DURU

DOKTORA TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Banu DİRİ

Eş-Danışman
Doç. Dr. Gülüstan DOĞAN

Ocak, 2020

TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KİTLESEL AÇIK ÇEVİRİMİÇİ KURSLARDA KATILIMCILARIN
İNGİLİZCE DİL GRUPLARININ TESPİTİNE DAYALI DAVRANIŞ VE
PERFORMANS ANALİZİ

İsmail DURU tarafından hazırlanan tez çalışması 03.01.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Banu DİRİ
Yıldız Teknik Üniversitesi
Danışman

Doç. Dr. Gülüstan DOĞAN
Yıldız Teknik Üniversitesi
Eş-Danışman

Jüri Üyeleri

Prof. Dr. Banu DİRİ, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Şule GÜNDÜZ ÖĞÜDÜCÜ, Üye
İstanbul Teknik Üniversitesi

Doç. Dr. Songül VARLI, Üye
Yıldız Teknik Üniversitesi

Doç. Dr. Arzucan ÖZGÜR, Üye
Boğaziçi Üniversitesi

Dr. Öğr. Üyesi Yunus Emre SELÇUK, Üye
Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Banu DİRİ sorumluluğunda tarafımca hazırlanan KİTLESEL AÇIK ÇEVİRİMİÇİ KURSLARDA KATILIMCILARIN İNGİLİZCE DİL GRUPLARININ TESPİTİNE DAYALI DAVRANIŞ VE PERFORMANS ANALİZİ başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

İsmail DURU

İmza

Bu tez çalışması, '1059B141601346' nolu Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) 2214-A doktora tezi sırası araştırma bursu ve '2016-04-01-DOP05' nolu Yıldız Teknik Üniversitesi BAP (DOP) projesi tarafından desteklenmiştir.

Çok kıymetli aileme

TEŞEKKÜR

Tezimi hazırladığım sürede, sürekli olarak hiç sıkılmadan, yorulmadan bilgi ve deneyimini paylaşan, motive eden ve önerileriyle bana her zaman destek olan değerli tez danışmanım Sayın Prof. Dr. Banu DİRİ'ye ve eş danışmanım Sayın Doç. Dr. Gülüstan DOĞAN'a teşekkürlerimi sunarım. Tez izleme dönemleri boyunca komitemde yer alan, önerileriyle tezime katkıda bulunan Prof. Dr. Şule GÜNDÜZ ÖĞÜDÜCÜ, Doç. Dr. Arzucan ÖZGÜR, Doç. Dr. Songül VARLI'ya teşekkür ederim. Tezim boyunca işbirliği fırsatı bulduğumuz sayın Dr. Öğretim Üyesi Ayşe Saliha SUNAR'a ve Doç. Dr. Su WHITE'a ve tez çalışmalarımın bir bölümünde destek olan Muhammed Emir ÖZÇEVİK'e de teşekkürlerimi sunarım.

İsmail DURU

İÇİNDEKİLER

SİMGE LİSTESİ	x
KISALTMA LİSTESİ	xi
ŞEKİL LİSTESİ	xiii
TABLO LİSTESİ	xix
ÖZET	xxiv
ABSTRACT	xxvi
1 Giriş	1
1.1 Literatür Özeti	6
1.1.1 Eğitimsel Veri Madenciliği	6
1.1.2 Öğrenme Analitiği	7
1.1.3 Derin Öğrenme	8
1.1.4 Kitlesel Çevrim İçi Açık Kurs (MOOC)	9
1.1.5 Kitlesel Çevrim İçi Açık Kurslarda Öğrenme Analitiği	11
1.1.6 Kitlesel Açık Çevrim İçi Kurslarda Kullanım Analizi	12
1.1.7 Kitlesel Açık Çevrim İçi Kurslarda Performans Tahmini	12
1.1.8 Kitlesel Açık Çevrim İçi Kurslarda Derin Öğrenme Kullanımı	15
1.1.9 Dil Perspektifinden Kitlesel Açık Çevrimiçi Kursların Analizi	17
1.2 Tezin Amacı ve Yöntem	20
1.3 Tezin Katkıları	25
1.3.1 Tezin Akademik Yayın Çalışmaları	26
1.4 Hipotezler	29
2 Veri	31
2.1 Kullanılan Veriler ve Özellikleri	33
2.1.1 Kayıtlar (Enrolments) Verisi ve Özellikleri	34
2.1.2 Kurs Adımları Aktivitesi (Step Activity) Verisi ve Özellikleri	35
2.1.3 Yorumlar (Comments) Verisi ve Özellikleri	35
2.1.4 Takipler (Followings) Verisi ve Özellikleri	36

3	Kullanıcıların İngilizce Dil Gruplarının Tespiti	38
3.1	Kurs Öncesi Anketlerdeki Ülke Bilgilerinden Kullanıcıların İngilizce Dil Gruplarının Tespiti	39
3.2	Kullanıcı Yorumlarının 'Manuel' Olarak Ön İncelemesi	44
3.3	Oluşturulan Düzenli İfadelere Dayalı Yöntemle İngilizce Dil Gruplarının Güncellenmesi	46
3.3.1	"Understanding Language 4: Learning and Teaching" Kursu (MOOC) İçin Sonuçlar	55
3.3.2	"Understanding Language: Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonu İçin Sonuçlar	58
3.3.3	Farklı Türden Kurslarda (MOOC'lar) Sonuçlar	60
3.4	Derin Öğrenme İle İngilizce Dil Grubu Tespiti	68
3.4.1	Standart Özellikleri Oluşturma	69
3.4.2	Kullanıcı Yorum Metinlerinin Ön işleme	69
3.4.3	Kullanılan Derin Öğrenme Modelleri	71
3.5	Derin Öğrenme Modellerinde Eğitim ve Test Senaryoları	77
3.5.1	Birinci Senaryo: Aynı Kurs (MOOC) Serisi İle Eğitim ve Test	79
3.5.2	İkinci Senaryo: Aynı Kurs (MOOC) Serisinin Farklı İterasyonlarının Eğitim ve Test Verileri Olarak Kullanımı	79
3.5.3	Üçüncü Senaryo: Farklı Türden Kursların (MOOC'lar) Test Verisi Olarak Kullanımı	80
3.5.4	Dördüncü Senaryo: Sadece Oluşturulan Standart Özelliklerin Kullanımı	81
3.6	Sonuçlar	82
4	İngilizce Dil Gruplarının Kullanım Analizleri	84
4.1	Veri Önışleme İle Standart Özelliklerin Çıkarımı	84
4.2	"Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri	91
4.2.1	Kurs Adımlarındaki Davranış Analizleri	91
4.2.2	Kurs Tartışma Forumlarındaki Davranış Analizleri	93
4.2.3	Takip Etkileşimlerindeki Davranış Analizleri	99
4.3	"Understanding Language : Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonunda İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri	101
4.3.1	Kurs Adımlarındaki Davranış Analizleri	101
4.3.2	Kurs Tartışma Forumlarındaki Davranış Analizleri	103
4.4	Farklı Türden Kurslarda (MOOC'lar) İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri	109

4.4.1	Kurs Adımlarındaki Davranış Analizleri	110
4.4.2	Kurs Tartışma Forumlarındaki Davranış Analizleri	112
4.5	Sonuç ve Gelecek Çalışmalar	121
5	İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri	124
5.1	Sık Kullanılan Geleneksel Makine Öğrenmesi Algoritmaları İle İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri	125
5.1.1	"Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri	126
5.1.2	"Understanding Language : Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonunda İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri	130
5.1.3	Farklı Türden Kurslarda (MOOC'lar) İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri	134
5.2	Derin Öğrenme İle İngilizce Dil Gruplarının Kurs Sonu Performansı Tahminlerinin Gerçekleştirilmesi	140
5.3	Derin Öğrenme Modellerinde Eğitim ve Test Senaryoları	143
5.3.1	Birinci Senaryo: Aynı Kurs (MOOC) Serisi İle Eğitim ve Test . .	145
5.3.2	İkinci Senaryo: Aynı Kurs (MOOC) Serisinin Farklı İterasyonlarının Eğitim ve Test Verileri Olarak Kullanımı	146
5.3.3	Üçüncü Senaryo: Farklı Türden Kursların (MOOC'lar) Test Verisi Olarak Kullanımı	146
5.3.4	Dördüncü Senaryo: Sadece Oluşturulan Standart Özelliklerin Kullanımı	148
5.4	Sonuçlar	148
6	İngilizce Dil Gruplarının Haftalık Kurs Performans Analizi	151
6.1	Veri Önışleme	151
6.2	"Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Haftalık Kurs Performans Analizleri .	152
6.3	Farklı Türden Kurslarda (MOOC'lar) İngilizce Dil Gruplarının Haftalık Kurs Performans Analizleri	158
6.4	Sonuçlar	169
7	Sonuç ve Öneriler	172
7.1	Gelecek Çalışmalar	181
A	Ekler	183
A.1	İngilizce Dil Gruplarındaki Ülkeler ve ISO-2 Formatındaki Kodları . . .	183

A.2 Dil, Ülke ve Millet Bilgisi Tespit İçin Oluşturulan Düzenli İfadeler ve Sembolleri	184
Kaynakça	187
Tezden Üretilmiş Yayınlar	196

SİMGE LİSTESİ

[]	Yukarı Yuvarlama
[]	Aşağı Yuvarlama

KISALTMA LİSTESİ

ABD	Amerika Birleşik Devletleri
ACT	American College Testing
Bidirectional LSTM	İki Yönlü Uzun-Kısa Süreli Bellek
CLIL	İçerik ve Dille Bütünleşik Öğrenme
cMOOC	Bağlantıcı Kitlesele Çevrim İçi Açık Kurs
CNN	Konvolüsyonel Sinir Ağları
CSV	Virgülle Ayrılmış Değerler
DT	Karar Ağacı
E-Öğrenme	Elektronik Öğrenme
EVM	Eğitimsel Veri Madenciliği
GPU	Grafik İşleme Ünitesi
GRU	Kapılı Tekrarlamalı Ünite
JMOOC	Japon MOOC Platformu
IBD	İngilizce Birincil Dil
IID	İngilizce İkincil Dil
IRD	İngilizce Resmi Dil
KNN	K-En Yakın Komşu
k-12	Kindergarten 12th grade
LMOOC	Language MOOC
LSTM	Uzun-Kısa Süreli Bellek
MLP	Çok Katmanlı Algılayıcı
MLR	Çok Kategorili Lojistik Regesyon

MOOC	Kitlesel Açık Çevrimiçi Kurs
ÖA	Öğrenme Analitiği
NA	Not Available
NB	Naive Bayes
RF	Rasgele Orman
RNN	Tekrarlı Sinir Ağları
SAT	Scholastic Aptitude Testing
SQL	Structured Query Language
SVM	Destek Vektör Makinesi
UL	Understanding Language : Learning and Teaching
URL	Uniform Resource Loader

ŞEKİL LİSTESİ

Şekil 1.1	Araştırmanın dört adımı	23
Şekil 1.2	R programlama dili ve Shiny kütüphanesi ile geliştirilen uygulamanın işlem ve veri akışı	26
Şekil 3.1	"Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketteki ülke bilgilerine göre İngilizce dil gruplarındaki katılımcıların yüzdeleri (birinci aşama)	40
Şekil 3.2	"Understanding Language 4: Learning and Teaching" kursu katılımcılarının kurs performansı özelliğinin aldığı değerlerin dağılımlarına göre yüzdeleri (birinci aşama)	41
Şekil 3.3	"Understanding Language 4: Learning and Teaching" kursunda "resmi dil olarak İngilizce" ve "ikincil dil olarak İngilizce" konuşan katılımcıların kurs performansı özelliğinin aldığı değerlerin dağılımlarına göre yüzdeleri (birinci aşama)	41
Şekil 3.4	"Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketteki ülke bilgilerine göre İngilizce dil grupları yüzdeleri (birinci ve ikinci aşama)	43
Şekil 3.5	"Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketlerdeki ülke bilgileri ile İngilizce dil gruplarının kurs performanslarına göre yüzdeleri (ikinci aşama)	44
Şekil 3.6	İngilizce birincil ve resmi dili olan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti	45
Şekil 3.7	İngilizce ikincil dili olan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti	45
Şekil 3.8	İngilizce resmi dili olup birincil dili olmayan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti	45
Şekil 3.9	Düzenli ifadeleri kullanarak kullanıcıların İngilizce dil gruplarına ayrılımları için önerdiğimiz model	48
Şekil 3.10	Düzenli ifadeler ile İngilizce dil grubu tespiti modelinin akış şeması	49

Şekil 3.11 İngilizce dil bilgisini çıkarmak için yazılan düzenli ifade kalıplarının bir örneği (l3)	51
Şekil 3.12 Millet bilgisini çıkarmak için yazılan düzenli ifade kalıplarının bir örneği (n1)	54
Şekil 3.13 Ülke bilgisini çıkarmak için yazılan düzenli ifadelerin bir örneği (c3)	55
Şekil 3.14 Link (URL) temizleme düzenli ifadesi	70
Şekil 3.15 Oluşturulan standart özellikler ve yorumlarla İngilizce dil grubu tahmini için Derin Öğrenme modeli	74
Şekil 3.16 Sadece yorum metinleri ile İngilizce dil grubu tahmini için Derin Öğrenme modeli	75
Şekil 3.17 Sadece oluşturulan standart özelliklerle İngilizce dil grubu tahmini için Derin Öğrenme modeli	76
Şekil 3.18 Kullanılan test verisi ve denenen mimarilere göre farklı senaryolar .	77
Şekil 4.1 İngilizce dil gruplarındaki kullanıcıların kursa katılımları - "Understanding Language 4: Learning and Teaching" kursu	92
Şekil 4.2 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu	92
Şekil 4.3 İngilizce dil gruplarındaki katılımcıların tartışma forumlarına katılım yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu	93
Şekil 4.4 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Understanding Language 4: Learning and Teaching" kursu	94
Şekil 4.5 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Understanding Language 4: Learning and Teaching" kursu	95
Şekil 4.6 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yaptıkları yorum ve aldıkları beğeni sayıları - "Understanding Language 4: Learning and Teaching" kursu	96
Şekil 4.7 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki ana yorumlarının aldıkları cevap yorum sayıları - "Understanding Language 4: Learning and Teaching" kursu	96
Şekil 4.8 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Understanding Language 4: Learning and Teaching" kursu	97
Şekil 4.9 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki ana yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language 4: Learning and Teaching" kursu	98

Şekil 4.10 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language 4: Learning and Teaching" kursu	98
Şekil 4.11 İngilizce dil gruplarındaki katılımcıların aynı kurstan takip ettikleri katılımcı sayıları - "Understanding Language 4: Learning and Teaching" kursu	99
Şekil 4.12 İngilizce dil gruplarındaki katılımcıları aynı kurstan takip eden katılımcı sayıları - "Understanding Language 4: Learning and Teaching" kursu	100
Şekil 4.13 İngilizce dil grupları arasında kullanıcıların takip etme etkileşimlerinin yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu	101
Şekil 4.14 İngilizce dil gruplarındaki kullanıcıların kursa katılımlarının azalımı - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	102
Şekil 4.15 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Understanding Language: Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	103
Şekil 4.16 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	104
Şekil 4.17 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	104
Şekil 4.18 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu .	105
Şekil 4.19 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	106
Şekil 4.20 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları cevap yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	107
Şekil 4.21 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu .	107

Şekil 4.22 İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	108
Şekil 4.23 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu	109
Şekil 4.24 İngilizce dil gruplarındaki kullanıcıların kursa katılımlarının azalımı - "Exploring Our Oceans 4" kursu	110
Şekil 4.25 İngilizce dil gruplarındaki kullanıcıların kursa katılımları - "Web Science 1" kursu	110
Şekil 4.26 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Exploring Our Oceans 4" kursu	111
Şekil 4.27 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Web Science 1" kursu	111
Şekil 4.28 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Exploring Our Oceans 4" kursu	112
Şekil 4.29 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Web Science 1" kursu	113
Şekil 4.30 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Exploring Our Oceans 4" kursu	113
Şekil 4.31 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Web Science 1" kursu	114
Şekil 4.32 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Exploring Our Oceans 4" kursu	114
Şekil 4.33 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Web Science 1" kursu	115
Şekil 4.34 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Exploring Our Oceans 4" kursu	116
Şekil 4.35 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Web Science 1" kursu	116
Şekil 4.36 İngilizce dil gruplarının kurs forumlarındaki yorumlarının aldıkları cevap sayıları - "Exploring Our Oceans 4" kursu	117
Şekil 4.37 İngilizce dil gruplarının kurs forumlarındaki yorumlarının aldıkları cevap sayıları - "Web Science 1" kursu	117
Şekil 4.38 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Exploring Our Oceans 4" kursu	118

Şekil 4.39	İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Web Science 1" kursu	118
Şekil 4.40	İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Exploring Our Oceans 4" kursu	119
Şekil 4.41	İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Web Science 1" kursu	119
Şekil 4.42	İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Exploring Our Oceans 4" kursu	120
Şekil 4.43	İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Web Science 1" kursu .	120
Şekil 5.1	Oluşturulan standart özellikler ve yorumlar ile kurs sonu performansı tahmini için Derin Öğrenme modeli	141
Şekil 5.2	Sadece yorumlarla kurs sonu performansı tahmini İçin Derin Öğrenme modeli	142
Şekil 5.3	Sadece oluşturulan standart özelliklerle kurs sonu performansı tahmini İçin Derin Öğrenme modeli	143
Şekil 5.4	Kullanılan test verisi ve denenen mimarilere göre farklı kurs sonu performansı tahmini senaryoları	144
Şekil 6.1	İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk (accuracy) değerleri - "Understanding Language 4: Learning and Teaching" kursu	152
Şekil 6.2	İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu	153
Şekil 6.3	İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu	154
Şekil 6.4	İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu	155
Şekil 6.5	İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu	158
Şekil 6.6	İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu	158

Şekil 6.7	İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu	160
Şekil 6.8	İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu	160
Şekil 6.9	İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu	161
Şekil 6.10	İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu	162
Şekil 6.11	İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu	163
Şekil 6.12	İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu	163

TABLO LİSTESİ

Tablo 1.1	edX verisi	27
Tablo 1.2	Detaylı olarak incelenen çalışmalarda kullanılan verilerin bilgileri .	28
Tablo 1.3	Detaylı olarak incelenen çalışmalardaki yöntem ve tahmin hedefi bilgileri	28
Tablo 2.1	Kullanılan FutureLearn platformu verileri ve özellikleri	34
Tablo 3.1	İngilizce'nin birincil dil veya İngilizce'nin resmi dil olduğu ülkelerin ISO-2 formatında kodları	42
Tablo 3.2	Yorumlar (Comments) verisinden kullanılan özellikler ve anlamları	44
Tablo 3.3	Kayıtlar (Enrolments) verisinden kullanılan özellikler ve anlamları .	44
Tablo 3.4	Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce'nin birincil ve resmi dil olduğu İngilizce dışındaki diller . .	52
Tablo 3.5	Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce'nin resmi dil olup birincil dil olmadığı İngilizce dışındaki diller	53
Tablo 3.6	Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce'nin ikincil dil olduğu İngilizce dışındaki diller	54
Tablo 3.7	Katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları - "Understanding Language 4: Learning and Teaching" kursu	56
Tablo 3.8	Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı (kurs öncesi anketinden ülke bilgisi bilinenler) - "Understanding Language 4: Learning and Teaching" kursu	57
Tablo 3.9	Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı (kurs öncesi anketinde ülke bilgisi yer almayanlar) - "Understanding Language 4: Learning and Teaching" kurs serisi	58
Tablo 3.10	"Understanding Language : Learning and Teaching" (UL) kurs serisinin ilk yedi iterasyonunda katılımcı ve yorum sayıları	59

Tablo 3.11 "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonundaki katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları . .	59
Tablo 3.12 Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı - "Understanding Language: Learning and Teaching" kurs serisinin ilk yedi iterasyonu	60
Tablo 3.13 "Exploring Our Oceans 4" kursunda katılımcı ve yorum sayıları . . .	61
Tablo 3.14 "Web Science 1" kursunda katılımcı ve yorum sayıları	61
Tablo 3.15 "AginCourt 3" kursunda katılımcı ve yorum sayıları	61
Tablo 3.16 "Research Project 5" kursunda katılımcı ve yorum sayıları	62
Tablo 3.17 "Exploring Our Oceans 4" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları	62
Tablo 3.18 "Web Science 1" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları . .	63
Tablo 3.19 "AginCourt 3" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları . .	63
Tablo 3.20 "Research Project 5" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları	64
Tablo 3.21 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Exploring Our Oceans 4" kursu	65
Tablo 3.22 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Web Science 1" kursu . .	66
Tablo 3.23 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "AginCourt 3" kursu . . .	67
Tablo 3.24 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Research Project 5" kursu	68
Tablo 3.25 Derin Öğrenme ile İngilizce dil grubu tahmini için kurs aktivite kayıtlarından faydalanarak oluşturulan standart özellikler (veri dosyalarına göre)	69
Tablo 3.26 Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri (birinci senaryo)	79

Tablo 3.27	Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri (ikinci senaryo)	80
Tablo 3.28	Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri - "Exploring Our Oceans 4" kursu (üçüncü senaryo)	80
Tablo 3.29	Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri - "Web Science 1" kursu (üçüncü senaryo)	81
Tablo 3.30	Oluşturulan standart özellikler kullanılarak Derin Öğrenme ile İngilizce dil grubu tahmini doğruluk yüzdeleri (dördüncü senaryo)	82
Tablo 4.1	Veri dosyalarına göre oluşturulan özelliklerin başlıcaları	85
Tablo 4.2	Veri dosyalarına göre oluşturularak analizlerde kullanılan özellikler	91
Tablo 5.1	Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu	127
Tablo 5.2	Kurs performansı tahmininde sık kullanılan yedi geleneksel algoritma için tahmin doğruluk (accuracy) değerleri - "Understanding Language 4: Learning and Teaching" kursu	128
Tablo 5.3	Rastgele Orman algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma ve F-ölçüm metriklerinin aldıkları değerler - "Understanding Language 4: Learning and Teaching" kursu	128
Tablo 5.4	Rastgele Orman algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik (denenen diğer algoritmalarından farklı olarak birinci hafta tamamlanan adım sayısı ikinci hafta tamamlanan adım sayısından daha önemlidir) - "Understanding Language 4: Learning and Teaching" kursu	129
Tablo 5.5	İngilizce dil gruplarında Rastgele Orman algoritmasının karmaşıklık matrisleri - "Understanding Language 4: Learning and Teaching" kursu	130
Tablo 5.6	Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Understanding Language: Learning and Teaching" kurs serisinin ilk yedi iterasyonu	131
Tablo 5.7	Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu	132
Tablo 5.8	Karar Ağacı algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu	132

Tablo 5.9	Karar Ağacı algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu	133
Tablo 5.10	İngilizce dil gruplarında Karar Ağacı algoritmasının karmaşıklık matrisleri - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu	133
Tablo 5.11	Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Exploring Our Oceans 4" kursu	135
Tablo 5.12	Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Web Science 1" kursu	135
Tablo 5.13	Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Exploring Our Oceans 4" kursu	136
Tablo 5.14	Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Web Science 1" kursu	137
Tablo 5.15	Rastgele Orman algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Exploring Our Oceans 4" kursu	137
Tablo 5.16	Destek Vektör Makinesi algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Web Science 1" kursu	138
Tablo 5.17	Rastgele Orman algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Exploring Our Oceans 4" kursu	138
Tablo 5.18	Destek Vektör Makinesi algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Web Science 1" kursu . . .	139
Tablo 5.19	İngilizce dil gruplarında Rastgele Orman algoritmasının karmaşıklık matrisleri - "Exploring Our Oceans 4" kursu	139
Tablo 5.20	İngilizce dil gruplarında Destek Vektör Makinesi algoritmasının karmaşıklık matrisleri - "Web Science 1" kursu	140
Tablo 5.21	Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri (birinci senaryo)	146
Tablo 5.22	Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri (ikinci senaryo)	146
Tablo 5.23	Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri - "Exploring Our Oceans 4" (üçüncü senaryo) . .	147
Tablo 5.24	Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri - "Web Science 1" (üçüncü senaryo)	147

Tablo 5.25	Oluşturulan standart özellikler kullanılarak Derin Öğrenme ile kurs sonu performansı tahmini doğruluk yüzdeleri (dördüncü senaryo) .	148
Tablo 6.1	"Understanding Language 4: Learning and Teaching" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Rastgele Orman ve Karar Ağacı algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik . .	156
Tablo 6.2	"Exploring Our Oceans 4" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Rastgele Orman ve Karar Ağacı algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik	165
Tablo 6.3	"Web Science 1" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Çok Kategorili Lojistik Regresyon ve Destek Vektör Makinesi algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik	167
Tablo A.1	Birincil dili ve resmi dili İngilizce olan ülkelerin İngilizce isimleri ve ISO-2 formatındaki kodları	183
Tablo A.2	İngilizce yalnızca resmi dili olan ülkelerin İngilizce isimleri ve ISO-2 formatındaki kodları	184
Tablo A.3	Katılımcı yorumlarından dil bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları	185
Tablo A.4	Katılımcı yorumlarından ülke bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları	186
Tablo A.5	Katılımcı yorumlarından millet bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları	186

KİTLESEL AÇIK ÇEVİRİMİÇİ KURSLARDA KATILIMCILARIN İNGİLİZCE DİL GRUPLARININ TESPİTİNE DAYALI DAVRANIŞ VE PERFORMANS ANALİZİ

İsmail DURU

Bilgisayar Mühendisliği Anabilim Dalı
Doktora Tezi

Danışman: Prof. Dr. Banu DİRİ
Eş-Danışman: Doç. Dr. Gülüstan DOĞAN

Günümüzde, isteyen her katılımcıya ücretsiz olarak 'Kitlesele Açık Çevrimiçi Kurs' (MOOC) olarak adlandırılan kursları sunan platformlar mevcuttur. Bu platformlardaki kurslar, açık iletişim ve video etkileşimlerine öncelik verecek şekilde ayrıık alt adımlar halinde yapılandırılmıştır. Dünya üzerinden birçok kurum MOOC oluşturmaya çalışmakta ve farklı coğrafi konumlardan milyonlarca insanı kendine çekmektedir. MOOC'ların kazandıkları önemli popülerliğe rağmen, dil, zorluk seviyesi ve öğrenme yaklaşımı gibi katılımcıların bireysel ihtiyaçlarına göre esnek olmamak gibi önemli problemleri vardır. Bu durumun bir sonucu olarak MOOC'lardaki önemli problemlerden biri, katılımcıların çok düşük tamamlama oranlarıdır (% 10'dan az).

Ana MOOC sağlayıcıları, çoğunlukla Amerika Birleşik Devletleri ve İngiltere gibi İngilizce konuşulan ülkeler tabanlıdır ve sağladıkları MOOC'ların dili İngilizce'dir. Bununla birlikte katılımcıların çoğunluğu İngilizce'yi yabancı dil olarak konuşmaktadır. Bu durumun katılımcılarda, özgüven, kullanım davranışları, iletişim tarzları ve kursu tamamlama performanslarında farklılıklara neden olduğu yapılan çalışmalar tarafından ortaya konulmuştur. Katılımcıların İngilizce dil becerilerindeki değişkenlik kaçınılmaz olarak, en fazla birincil dili İngilizce olmayan katılımcılar için bir engel oluşturmaktadır. Bu engele karşı bir çözüm üretebilmek için MOOC'larda İngilizce dilbilgisi eksikliği sebebi ile zorluk yaşayan katılımcıların tespit

edilmeleri gerekmektedir. Kurs devam ederken öğrencilerin etkileşimlerini anlamak, öğrenme yollarını karakterize etmek, kursu bırakma oranlarını en aza indirmek ve öğretmen müdahalelerini başlatmaya yardımcı olabilmek kritik öneme sahiptir.

MOOC'larda, öğrencinin içeriklere hangi sıklıkta, ne kadar süreyle, hangi zaman dilimlerinde ulaştığı, ne kadar ilerlediği, ne zaman, ne kadar ve hangi konularda kimlerle paylaşımda bulunduğu bilgileri kayıt altına alınabilmektedir. Bu bilgiler, MOOC platform sağlayıcıları ve kursları hazırlayanların, katılımcıların ihtiyaçlarını tespit edip bu ihtiyaçlara göre kursları yapılandırmalarına yardımcı olabilir. Bununla birlikte, İngilizce'yi ikincil olarak konuşan katılımcıların tespiti ve onların MOOC'lardaki performans etkileşimleri ile ilgili çok fazla çalışma yoktur. Aynı zamanda, ikincil dil olarak İngilizce konuşanların MOOC'lardaki davranışlarını araştıran sınırlı sayıda araştırma vardır.

Tezde, MOOC katılımcıları ana dillerine göre gruplandırılmış, ardında da bu grupların kullanım ve performans analizleri gerçekleştirilmiştir. Bunun için öncelikle katılımcıların, oluşturulan yöntemlerle İngilizce dil gruplarına ayrılmasına odaklanılmıştır. Katılımcıları kendi İngilizce dil geçmişlerine göre otomatik veya yarı otomatik olarak tespit etme yöntemleri belirlendikten sonra kullanım davranışları ve kurs performansları analiz edilmiştir.

Tez kapsamında, içerdiği sosyalleşme özellikleri, dil odaklı analiz yapma ve işbirliği imkânlarının yüksek olması gibi nedenlerle, çoğunlukla İngiltere tabanlı FutureLearn MOOC platformunda Southampton Üniversitesi tarafından sunulan farklı türden MOOC'ların verileri kullanılmıştır.

Tezde uygulanan metodoloji, oluşturulan farklı yöntemlerle İngilizce dil gruplarının tespit edilmesi, güncellenmesi, kullanıcı davranışlarının analiz edilmesi, performanslarının tahmininde kullanılabilecek en uygun ve performanslarına en fazla etki eden özelliklerle, performans tahminlerinde en başarılı algoritmaları belirlemede kullanılabilecek birkaç alt aşamayı içermektedir. Bu sonuçlar ayrıca MOOC platformu geliştiricileri tarafından, kurs içinde katılımcıların performanslarını iyileştirebilmek ve erken müdahale sağlayabilmek için strateji geliştirmede kullanılabilir.

Anahtar Kelimeler: Performans tahmini, dil tespiti, kullanım analizi, Kitlesel Açık Çevrimiçi Kurs, ikincil dil olarak İngilizce

ABSTRACT

BEHAVIOUR AND PERFORMANCE ANALYSIS OF LEARNERS IDENTIFIED IN ENGLISH LANGUAGE BASED GROUPS ON MASSIVE OPEN ONLINE COURSES

İsmail DURU

Department of Computer Engineering

Doctor of Philosophy Thesis

Advisor: Prof. Dr. Banu DIRI

Co-advisor: Assoc. Prof. Dr. Gulustan DOGAN

Educational platforms offer Massive Open Online Courses (MOOCs), where learners can study online courses for free of charge. The design of MOOCs enables learners to interact with video lectures as the main module and communicate with others through supportive tools such as discussion forums. Institutions and universities create content for MOOCs and attract millions of people worldwide. Despite the hype around MOOCs, there are a number of deficiencies of MOOCs, namely, language, difficulty of courses, one-size-fits-all approach which denies learners' personal differences such as learning approaches. One of the consequences of these deficiencies is the low completion rate of courses (less than 10%).

The biggest MOOC providers are the UK and the US-based and the majority of MOOCs are provided in English language. However, the majority of participants speaks English as a second language. Studies show that some learners feel low self-confidence due to language skills and it causes different behaviour patterns in course activities and course completion. The difference in English-language abilities inevitably is an obstacle especially for those who speak English as a second language. In order to overcome this obstacle, it is important to detect the participants whose having difficulties due to their lack of English language skills. Understanding learners' interactions during the course, characterizing learning patterns, lowering course drop-outs, and enabling learning interventions are also crucially important.

MOOCs log learners' activities such as how frequently, when, and which page they visited, how far they completed the course, when and with whom they interacted in a discussion. This information might be helpful for MOOC providers and course authors to detect the needs of learners and organise their course to meet these detected needs. Having said that, the identification of learners who speak English as a second language and analysing course performance have not been widely investigated yet. There very little number of studies published on behaviour analysis of MOOC learners who speak English as a second language.

In this thesis, MOOC participants were categorised by their first language and then their behaviour and course performances were analysed. In order to complete the analysis, firstly, this thesis focused on categorising participants based on their first language whether or not it is English. This thesis proposes semi and fully automatic methods for categorising participants based on the English language. The behaviour patterns and course performances of learners were then analysed based on the language groups.

This thesis mainly uses data collected through a selection of MOOCs delivered by the University of Southampton on the UK-based MOOC platform, FutureLearn. FutureLearn has been chosen due to its rich social tools integrated into the platform and their welcoming approach towards international research collaborations. The methodology applied in the thesis consists of several sub-steps, which are: identifying English language groups by different methods, statistical analysis of learners' behaviours, identifying the best possible and most important features that can be used in performance prediction, identifying the most successful machine learning algorithms to predict learners' course performance. The finding of the thesis can also be used by MOOC providers for identifying strategies to provide early course interventions and help MOOC participants to improve their course performance.

Keywords: Prediction of performance, Identification of language, Usage analysis, Massive Open Online Courses, English as a second language

1

Giriş

Günümüzde isteyen her katılımcıya ücretsiz kayıtla çevrim içi kurslar sunan platformlar, teknolojiyle güçlendirilmiş öğrenmedeki trend uygulamalarından biri haline gelmiştir. Bu kurslar 'Kitlesel Açık Çevrim İçi Kurs' (MOOC) olarak isimlendirilmekte ve Dünya üzerinden birçok kurum MOOC oluşturmaya çalışmaktadır.

MOOC'lar tüm uzaktan eğitim alan öğrenciler için İnternet'ten ücretsiz olarak erişime açık olan derslerdir [1]. MOOC'larda bir tek derse kayıt olan öğrenci sayısı 10 bin ile 200 bin arasında değişebilmektedir. Böylelikle MOOC'lar büyük ölçekli sayısal veri (öğrencinin kurs yorumları, zamansal ve coğrafi konum verileri gibi) toplamak için potansiyel bir kaynak niteliği taşımaktadır [2].

MOOC'larda öğrenme, açık iletişime öncelik verilererek ayrı adımlara bölünmekte ve video bileşenleri alt yazılar içermektedir.

MOOC'lar sürdürülebilir eğitim için büyük bir potansiyele sahiptir. Her yıl milyonlarca katılımcı, giderek artan çeşitlilik gösteren ve uluslararası katılımcı nüfusunun ihtiyaçlarını karşılamak amacı ile tasarlanan MOOC'lara kaydolmaktadır. ClassCentral raporuna göre, 2018 yılında ¹, 100 milyon insan, 900'ün üzerinde enstitü tarafından sağlanan 11,4 bin adet MOOC'a katılım sağlamıştır. Katılımcılardan 20 milyonu ilk defa bir MOOC'a kayıt olanlardan oluşmaktadır ve bu rakam bir önceki yılki ilk defa katılan katılımcı sayısı olan 23 milyondan biraz daha azdır.

Çin Eğitim Bakanlığı, 2016 yılında 10 milyondan fazla Çin'li öğrencinin yıl sonuna kadar MOOC'lara katılacağını öngörmüştür ve bu sayı 2014 yılındaki sayıdan yaklaşık yedi kat daha fazladır ². Buna rağmen aynı habere göre Çin'deki üniversitelerin % 90'ı MOOC sağlamayı planlamamıştır.

Giderek daha yüksek seviyede öğretim kaynağı sağlayan MOOC'lar, hem resmi

¹<https://www.class-central.com/report/mooc-stats-2018>

²<http://thepienews.com/news/chinese-mooc-learners-top-10-million-year-end>

(örgün), hem de yaygın öğretimin genişlemesini sürdürmek için uygulanabilir bir yol sağlamaktadırlar. Resmi olarak, harmanlanmış (karma) MOOC'lar, akademisyenlerin dışarıdan üretilen kaynakları yüz yüze öğretimlerine dahil edebilecekleri bir araç sağlamaktadırlar. Bu durum öğretim kapasitesini hızla arttırmak ve büyütme için bir araç olarak kullanılmaktadır (örneğin Suudi Arabistan'daki kadın üniversitelerinde) [3].

MOOC'larda öğretme ve öğrenmeyi geliştirmek için, araştırmacılar MOOC'ları, paydaşlar [4], tasarım [5], MOOC pedagojisi [6], öğrenci etkileşimi [7], yüksek öğrenim üzerindeki etkileri [8], MOOC forumları [9], performans tahmini [10–13], muhtemel iyileştirmeler [14] gibi farklı açılardan incelemektedirler.

MOOC'lardaki ana kaygılardan biri az kişiselleştirilmiş platform tasarımı ve çok farklı ülkelerden katılımcılara sunulan statik içeriktir ³. Bu kişiselleştirme problemi yaş, etnik köken, cinsiyet, eğitim durumu gibi birçok farklı açıdan ele alınabilir. Örneğin, sosyal bilimlerde tecrübesi olan bir öğrenci teknik bir konuda zorlanabilir ve konuyla ilgili temel bilgilere ihtiyaç duyabilir [15].

MOOC'lar, üniversite eğitimini ölçeklendirme potansiyeline sahiptirler ve binlerce öğrencinin tek bir çevrim içi kursa katılmasını sağlamaktadırlar. Bununla birlikte, Coursera, edX veya Iversity gibi başarılı MOOC platformları bile çok düşük tamamlama oranları sorunlarıyla karşı karşıya kalmaktadırlar. MOOC'ların güçlü bir yanı, çok sayıda katılan katılımcılara ücretsiz olarak uzaktan erişim sağlayabilmeleridir. MOOC'lar giderek daha çeşitli ve uluslararası topluluklardan gelen öğrencilerin ihtiyaçlarını karşılamaya çalıştıkça, bu güç aynı zamanda potansiyel problemleri de beraberinde getirmektedir. Katılımcıların geçmişleri, önceden aldıkları eğitimsel kazanımlar, yaş, katıldıkları yer ve ilk dil dahil olmak üzere birçok faktöre göre değişmektedir. MOOC yazarları, ortaya çıkan ihtiyaçlara karşı kursların iyi organize edilmesini ve yapılandırılmasını sağlamalıdır.

Gayri resmi olarak, bireysel katılımcılar genellikle eğitime bağımsız olarak erişmek ve mesleki becerilerini güncellemek, eğitime çok az bir ücret karşılığında veya hiçbir ücret ödemediği erişim sağlamak için MOOC'ları kullanır. Bununla birlikte, MOOC'ların gelişmekte olan ülkelerde kullanımı, önemli sorun alanlarından biri olduğu belirlenen dil kaynaklı engeller sebebi ile, düşünüldüğü kadar basit değildir [16]. Katılımcıların dil becerilerindeki değişkenlik kaçınılmaz olarak, en fazla birincil dili İngilizce olmayan kullanıcılar için bir engel oluşturmaktadır.

Birçok araştırmacı tarafından, MOOC'lardaki düşük tamamlanma oranlarının

³<https://computinged.wordpress.com/2013/01/04/moocs-are-a-fundamental-misperception-of-how-learning-works>

nedenlerinden biri olarak MOOC'lardaki herkese uyan tek bir model yapısı anlayışı gösterilmektedir [17, 18]. MOOC'ların içeriği bireysel katılımcıların dil, zorluk seviyesi ve öğrenme yaklaşımları gibi farklılık gösteren kişiye özel ihtiyaçlara göre değişmemektedir.

MOOC'ların kazandıkları önemli popülerliğe rağmen, dil, zorluk seviyesi, ve öğrenme yaklaşımı gibi katılımcıların bireysel ihtiyaçlarına göre esnek olmamak gibi önemli problemleri vardır. MOOC'lar birçok farklı ülkeden milyonlarca öğrenciyi çekse de, ana MOOC sağlayıcıları çoğunlukla Amerika Birleşik Devletleri (ABD) ve İngiltere gibi İngilizce konuşulan ülkeler tabanlıdır ⁴. Shah [19] tarafından yayınlanan 2015 yılı Class-Central istatistiklerine göre MOOC'ların % 75'i İngilizce olarak yayınlanmıştır ⁵. Bu oran 2014 yılı için % 80'dir. Aynı İnternet sitesi ⁶ 2016 yılı için farklı türde bir inceleme analizi yayınlamıştır. Analizde yeni MOOC katılımcılarının % 25'inin İngilizce dışındaki ülkeler tarafından sağlanan MOOC'lara katılım sağladıklarını tespit etmişlerdir ⁷. Bu durum bazı katılımcıların kurs videolarında konuşulan dili veya öğretmenin aksanını anlamalarında zorlanmalarına neden olabilir [20].

MOOC'ların çoğunun dili İngilizce ve sağlayıcıları ABD'ye dayalı olsa da, katılımcıların demografik özellikleri farklılık göstermektedir. 2017 yılında yayınlanan MIT üniversitesi raporuna göre, katılımcılarının % 71'i uluslararası katılımcılardır ⁸. En iyi MOOC sağlayıcılarının çoğunluğu İngilizce dili tabanlıdır. 2018 yılında en iyi MOOC sağlayıcısı değişmemiştir. Katılımcı sayılarına göre ABD tabanlı MOOC sağlayıcılar Coursera (37 milyon); edX (18 milyon); Udacity (10 milyon) ve İngiltere tabanlı MOOC sağlayıcı FutureLearn (8,7 milyon) İngilizce tabanlı iken, XuetangX (14 milyon) Çin tabanlıdır⁹.

Çin'de birçok MOOC Mandarin dilinde sunulsa da, Çin'deki öğrencilerin birçoğu İngilizce teknik kelime hazinelerini ve uzmanlıklarını geliştirmek için İngilizce MOOC'lara kaydolmayı seçmektedir ¹⁰.

Yerel olarak başlatılan bazı MOOC platformları, hedef kitleye yönelik yerel dillerde kurslar sunmaktadır. Örneğin, Japon MOOC platformu (JMOOC), web sitesinde, birçok Japon'un İngilizce öğrenmekte zorluk yaşaması sebebi ile, Japonca konuşan topluluklara hizmet vermek için JMOOC'u başlattıklarını belirtmiştir. ¹¹.

⁴<https://www.class-central.com/report/moocs-2015-stats>

⁵<https://www.class-central.com/report/moocs-2015-stats>

⁶<https://www.class-central.com/report/moocs-2015-stats>

⁷<https://www.class-central.com/report/moocs-stats-and-trends-2016>

⁸<https://news.mit.edu/2017/mooc-study-offers-insights-into-online-learner-engagement-behavior-0112>

⁹<https://www.class-central.com/report/mooc-stats-2018>

¹⁰<http://thepienews.com/news/chinese-mooc-learners-top-10-million-year-end>

¹¹<http://www.jmooc.jp/en/about>

Tüm bu gelişmelerin yanında, İngilizce olarak sunulan MOOC'lar, ikincil dil olarak İngilizce konuşan katılımcıların dikkatini çekmeye devam etmektedir. Bu katılımcıların platforma katılmalarına yardımcı olmak için bazı kişiselleştirilmiş hizmetler de sağlanmaktadır. Örneğin, Coursera katılımcılarından kurs içeriği çevirisine gönüllü olarak katkıda bulunmalarını talep etmektedir ¹². Böylece, katılımcıların kurstan daha yüksek bir düzeyde fayda sağlamaları mümkün olabilecektir.

MOOC'lar, İnternet bağlantısına ve uygun bir cihaza sahip olan tüm kullanıcılara açık ve serbest bir şekilde içeriklerini sundukları için eğitimde demokrasi sağlayıcısı olarak düşünülmektedirler. Bununla birlikte, Dillahunta ve diğerleri gerçeğin çok farklı olduğunu kanıtlayan bir takım yönlerini de ortaya koymuşlardır [21]. Yazarlar, MOOC'larda İngilizce'nin birincil eğitim dili olarak kullanmasının tartışılması gerektiğini ifade etmişlerdir. Uluslararası katılımcıların büyük bir kısmı, ikinci dil olarak İngilizce konuşan katılımcılar için İngilizce ile iletişim kurmanın olası zorlukları sorununu gündeme getirmektedirler. Bu tür engellerin üstesinden gelmek için bir çözüm, EMMA ¹³ gibi bazı kurumlar tarafından MOOC'ların birden fazla Avrupa dilinde sunulması için geliştirilmiştir. Mamgain ve diğerleri, bazı MOOC sağlayıcılarının (Coursera gibi) öğrencilere İngilizce transkript seçeneğini ve farklı dillerde alt yazıları sağlamalarına dikkat çekmişlerdir [22]. Eriksson ve diğerleri çalışmalarında, MOOC öğrencileri için İngilizce'deki alt yazıların yardımcı olduğunu belirtmişlerdir [20].

Yapılan araştırmalar, kullanıcıların hangi İngilizce dil gruplarında bulunduklarının kursu başarı ile tamamlamalarında etkili olduklarını göstermektedir [23]. Birincil dilinde öğrenim gören kullanıcıların kursu bitirmeye yönelik özgüveni ve motivasyonu diğer kullanıcılardan daha yüksektir [24]. Barak ve diğerleri motivasyonu yüksek olan öğrencilerin kursu tamamlamasının daha olası olduğunu iddia etmektedirler [24]. Ek olarak, yazarlara göre ilk dillerinde okuyan öğrenciler kursu tamamlama konusunda daha fazla güven duymakta ve bu da motivasyon seviyelerini arttırmaktadır. Barak ve diğerleri farklı kültürel geçmişlerden gelen öğrencilerin öğrenme yöntemlerinde, iletişim tarzlarında ve davranış kalıplarında farklılık gösterebileceklerini belirtmektedirler [24].

Katılımcıların öğrenme davranış kalıplarını belirleme ve sınıflandırma için MOOC'lar ile nasıl etkileşim kurduklarını inceleyen çok sayıda çalışma vardır [25, 26]. Bazı araştırmacılar, MOOC araştırmalarında ortak hedeflerden biri olan katılım düzeyi ve kurs tamamlama seviyelerini belirlemek ve tahmin etmek için farklı faktörlere

¹²<http://www.coursera.community/#gtc>

¹³<https://platform.europeanmoocs.eu/>

odaklanmaktadır. Örneğin, Kızılcec ve diğerleri temel olarak zamanlı ödev yüklemelerini göz önünde bulundurmuşlar ve katılımcıların davranış modellerini (denetleme, geride, takipte (yolunda) ve ayrılmış gibi) belirlemişlerdir [25]. Ardından da katılımcıları etkileşim seviyelerine göre (deneyleyici, tamamlayıcı, bağlantısını kesen ve örnekleyici olarak) gruplandırmışlardır.

Kumar ve Shastry, kurs devam ederken öğrencilerin etkileşimlerini anlamının, öğrencilerin öğrenme yollarını karakterize etmeye, kursu bırakma oranlarını en aza indirmeye ve öğretmen müdahalelerini başlatmaya yardımcı olabileceğini belirtmişlerdir [27].

MOOC'larda yer alan kurs forumları, öğrencilerin kursa katılımları ve kursa yönelik tutumları hakkında iyi bir gösterge olmaktadır. Bunun sonucu olarak, kurs forumlarının analizi bize MOOC forumları hakkında daha fazla bilgi sahibi olma imkanı sağlamaktadır [9]. Bir öğrencinin kursta başarılı olmasının muhtemel olduğunun önemli bir göstergesi, sohbetler yolu ile tartışma forumlarına katılmasıdır. Çünkü bu katılım biraz İngilizce dil akıcılığı gerektirmektedir. Cho ve Byun, ikincil dil olarak İngilizce konuşan katılımcıların MOOC'larda nasıl çalıştıklarına dair çok az kanıt bulunduğunu belirtmişlerdir [28]. Kızılcec ve diğerlerinin açıkladığı gibi, İngilizce'nin tipik olarak birincil dil olmadığı en az gelişmiş ülkelerden gelen insanlar zayıf İngilizce dil becerilerinden ötürü daha az yeterli olma korkusu ile cesaretleri kırılmış hissedebilirler [29]. Bu faktör, onların sohbetlere daha az katkıda bulunmalarına sebep olabilir, bu da kursa katılımlarının azalmasına ve öğrencilerin kursu bırakma ihtimalinin artmasına neden olabilir.

Araştırmacılar, katılımcıları kurslardaki aktivite sıralı kayıt (log) ve tıklama akışı verilerine göre sınıflandırmak için istatistiksel analizleri ve Öğrenme Analitikleri yöntemlerini ve makine öğrenmesi algoritmalarını kullanmaktadırlar. Örneğin, Milligan ve diğerleri katılımcılarla görüşmeler yapmışlar ve katılımcıları ifadelerine ve kurs aktivitelerine göre sınıflandırmışlardır [30]. Gillani ve diğerleri katılımcıların tartışma formlarındaki yorumlarında yer alan ifadelerini sınıflandırma için kullanmışlardır [26].

Uchidiuno ve diğerleri MOOC katılımcılarının dilini tespit etmek için İnternet tarayıcı dili tercihlerini kullanmış ve kullanıcıların video etkileşimlerini analiz etmişlerdir [31]. Yazarlar İnternet tarayıcı dil tercihlerinin İngilizce ikincil dil olan kullanıcıları daha doğru tespit etmek için faydalı olduğunu göstermişlerdir.

Bu çalışmalara rağmen, bildiğimiz kadarıyla, İngilizce'yi ikincil olarak konuşan konuşmacıların tespiti ve onların MOOC'lardaki performans etkileşimleri ile ilgili çok fazla çalışma yoktur. Aynı zamanda, ikincil dil olarak İngilizce konuşanların MOOC'lar

içindeki davranışlarını araştıran sınırlı sayıda araştırma vardır.

1.1 Literatür Özeti

Bu bölümde literatür özeti aşağıda listelenen dokuz alt başlık altında incelenmiştir.

- Eğitimsel Veri Madenciliği
- Öğrenme Analitiği
- Derin Öğrenme
- Kitlel Çevrim İçi Açık Kurs (MOOC)
- Kitlel Çevrim İçi Açık Kurslarda Öğrenme Analitiği
- Kitlel Açık Çevrim İçi Kurslarda Kullanım Analizi
- Kitlel Açık Çevrim İçi Kurslarda Performans Tahmini
- Kitlel Açık Çevrim İçi Kurslarda Derin Öğrenme Kullanımı
- Dil Perspektifinden Kitlel Açık Çevrim İçi Kursların Kullanımı

1.1.1 Eğitimsel Veri Madenciliği

Eğitimsel Veri Madenciliği'nin literatürde farklı tanımları yapılmıştır. Bu tanımlardan ilkinke göre, Eğitimsel Veri Madenciliği, eğitimsel sistemlere ait ham verilerden tasarım modellerine bilgi sağlama ve araştırma sonuçlarını cevaplama için kullanılabilecek kullanışlı bilgiler oluşturma sürecidir [32]. İkinci tanıma göre, Eğitimsel Veri Madenciliği bağımsız bir araştırma alanı olarak ortaya çıkmış ve 2008 yılında yıllık Eğitimsel Veri Madenciliği Uluslararası Konferansı ve Eğitimsel Veri Madenciliği Dergisi'nin kurulmasıyla sonuçlanmıştır [33].

Eğitimsel Veri Madenciliği metotları genelde eğitimsel verideki anlamlı hiyerarşinin çoklu seviyede açıkça kullanması ile daha geniş olan veri madenciliği literatüründen farklılık göstermektedir [33]. Eğitimsel Veri Madenciliği'nin amacını gerçekleştirmek için Psikometri literatüründen metotlar genellikle makine öğrenmesi ve veri madenciliği literatüründen metotlar ile entegre edilmektedir [33].

Son yıllarda, Eğitim Veri Madenciliği'nin araştırma alanları aşağıdaki gibidir [34];

- Çevrimdışı eğitim (Yüzyüze eğitime dayalı bilgi ve beceriyi aktarmayı deneme ve psikolojik olarak insanın nasıl öğrendiği ile ilgili)

- E-öğrenme ve öğrenme yönetim sistemi
- Akıllı eğitim sistemi ve uyarlanabilir eğitimsel hiper medya sistemi

1.1.2 Öğrenme Analitiği

Literatürde Öğrenme Analitiği terimi için birkaç tanım alternatifi bulunmaktadır. Birinci tanıma göre Öğrenme Analitiği, öğretmenin ve öğrenmenin gerçekleştiği ortamın anlaşılması ve optimize edilmesi için öğrencilerin ve onların içeriklerine ait verilerin ölçülmesi, toplanması, analiz edilmesi ve raporlanmasıdır [35]. İkinci tanım, Öğrenme Analitiği'ni öğrenmeyi iyileştirmek için katılımcılarla ilgili verilerin analizi ve gösterimi, öğretmenlerin eğitimi anlayabileceği yeni bir mercek olarak tanımlanmaktadır [36]. Üçüncü tanıma göre Öğrenme Analitiği, Büyük Veri tekniklerinin öğrenmeyi iyileştirmek için uygulanmasıdır [33]. Dördüncü ve bizim yer vereceğimiz son tanıma göre Öğrenme Analitiği, öğrenme ve eğitimi geliştirmek için karmaşık analitik araçların kullanıldığı yeni bir alandır [37]. Öğrenme teknolojisi satıcıları analitik paketler sağlamaktadırlar. Örneğin, Blackboard, Desire2Learn, Instructure ve Tribal tüm analitik araçlarını yayınlamıştır ve bunlara ek olarak Moodle topluluğunda da bir aktivite görülmektedir. Yüksek profilli MOOC sağlayıcıların hepsi (Coursera, Udacity ve edX) uygulamaları hakkında bilgilendirme sağlamak için analitik araçlarını kullanmaktadırlar [33].

Öğrenme Analitiği, diğer modelleme çalışmalarında sıklıkla kullanılan çok çeşitli türde öğrenen karakteristik verilerinden de yararlanmaktadır. Aslında, Öğrenme Analitiği, öğrenme çıktıları bilgilendirmesi ve karakteristik verilerin öğrenilmesiyle kayıt, kalıcılık, devamlılık ve mezuniyet modellerinin iyileştirilmesi olarak görülebilir. Diğer veri türleri arasında [38]:

- Test puanları (örneğin, ACT, SAT)
- Kurs notları
- Demografik, psikografik veriler
- Öğrenme stilleri, özellikleri, karakteristikleri veya tercihlerinin verileri
- Öğrenme/içerik yönetim sistemi aktivite verisi
- Anket verisi

yer almaktadır.

Öğrenme Analitiği'nde yer alan kişiler aşadaki dört paydaş grubuna ayrılabilirler [39]:

- Öğrenen: Öğrenme ile etkileşim kuran herkes.
- Öğretmen: Doğrudan öğrenmeyi kolaylaştırma ile ilgili herkes. Öğretim asistanlarını, yardımcı öğretim görevlilerini, fakülte yardımcılarını, fakülteyi, akademik personeli ve bazı MOOC'lardaki akranları içermektedir.
- Yönetici: Öğretmenlerin organizasyonundan veya yönetiminden sorumlu olan kişilerdir. Bölüm düzeyinde ve kurumsal düzeyde yönetimi içermektedir (örneğin, yöneticiler, bölüm başkanları, dekanlar, başkanlar, başkan yardımcıları, rektörler ve onların vekilleri gibi).
- Politika Yapıcı: Yerel, bölgesel, eyalet, ulusal veya uluslararası/hükümetler arası düzeyde ve fon sağlayıcılar dahil olmak üzere politikaların belirlenmesinden sorumlu olan herkes.

Öğrenme Analitiği bir dizi soruyu ele almak ve çeşitli öğrenme durumları hakkında fikir vermek için uygulanabilir. Öğrenme Analitiği'nin uygulanabileceği alanları açıklayıcı örnekler [38]:

- Sonuç başarısını tahmin etme
- Kurs ve program göstergeleri
- Müfredat değerlendirme
- Öğrenme çıktılarına öncelik verme
- Kursu ve öğretim politikalarını belirleme
- Akademik kalitenin tanımlanması

1.1.3 Derin Öğrenme

Derin Öğrenme, yapay sinir ağlarının bir alt alanı olan bir makine öğrenmesi tekniğidir. Sinir ağları beynin sinir sisteminden esinlenerek yapılandırılmış matematiksel modellerdir. Beyindeki sinir ağına benzer şekilde, her sinir hücresi bir girdi alan, aldığı bilgiyi beyinde yönetilebilir bilgilere dönüştüren ve bir çıktı üreten matematiksel bir fonksiyondur. Yapay sinir ağlarında, algoritmaya ilk verinin girdiği bir giriş katmanı, algoritmanın verilerden öğrendiği gizli katmanlar ve son olarak sonuç tahminin üretildiği bir çıkış katmanı vardır. Gizli katmanların sayısı çok olduğunda, sinir ağı Derin Öğrenme olarak adlandırılır [40].

Derin Öğrenme modellerinde, gizli katmanların yapılandırılmasındaki çeşitlilik, farklı Derin Öğrenme modellerini oluşturur.

Derin Öğrenme'nin kullanım potansiyeli, Goodfellow ve diğerleri tarafından tanımlanmıştır [40]. Derin Öğrenme, kavramlar hiyerarşisi açısından Dünya'yı anlamak ve ondan öğrenmek için öğrenci davranışını analiz etmemizi sağlar [40]. Derin Öğrenme'nin hiyerarşik konsepti, bilgisayarların karmaşık kavramları, daha basit katmanla başlayıp inşa ederek öğrendiği birçok katman içerir.

Derin Öğrenme yeni bir yöntem olmasa da, çok sayıda katmanı işlemek için güçlü bir bilgisayara ihtiyaç vardır. 2000'li yılların başlarında bilgisayar mimarisindeki gelişmeler, büyük verileri işleyebilmemizi sağlamıştır [41]. Sonuç olarak, Derin Öğrenme yaygın olarak kullanılan yöntemlerden biri haline gelmiştir. Son Derin Öğrenme uygulamalarında yüzlerce katman oluşturulmuştur. Derin Öğrenme'nin bugün kullanıma daha uygun olmasının bir başka nedeni de, üretilen büyük miktarda verinin mevcudiyetidir. Derin Öğrenme modeli; Rastgele Orman modeli, Destek Vektör Makinesi gibi diğer makine öğrenmesi teknikleriyle karşılaştırıldığında algoritmayı her katmanda beslemek için daha büyük veri gerektirmektedir.

Derin Öğrenme modeli, bilgisayarların daha basit katmandan yapılandırmaya başlayarak karmaşık kavramları öğrendiği birçok katmanı içeren yapılandırılmış hiyerarşik katmanlardır [40].

Derin Öğrenme modellerinden biri olan Konvolüsyonel Sinir Ağları, konvolüsyonel katmanları da içeren gizli katmanlara sahiptir. Bu katmanlar verinin iç yapısını öğrenebilir ve filtreler uygulayarak kalıpları (yolları) tespit edebilirler. Konvolüsyonel Sinir Ağları, öğrenebilen ve filtreler uygulayarak kalıpları tespit edebilen evrişimli katmanlar içeren gizli katmanlara sahiptir. Bu filtreler metni kelime/cümle dizileri ile filtrelemektedirler [42].

1.1.4 Kitlel Çevrim İçi Açık Kurs (MOOC)

Kitlel Açık Çevrim İçi Kurslar (MOOC'lar), tüm uzaktan eğitim alan öğrenciler için ücretsiz olarak İnternet'ten erişime açık olan kurslardır. MOOC'lar, kitleler için yaşam boyu öğrenme fırsatlarına daha geniş bir erişim imkanı sağlayarak, eğitimi küresel düzeyde ilerletme potansiyeline sahiptirler [43]. MOOC'ların muazzam başarısı kısmen, küresel olarak ulaşılabilirlik özelliği, kurs yayında olduğu süre içinde Dünya üzerinden herkesin istediği zaman kursa kayıt olup/bırakabilmesine dayandırılabilir [44].

Coursera, edX ve Udacity gibi MOOC sağlayıcıların artan popülerlikleri ile birlikte,

MOOC'lar eğitimcilerin, bilgisayar bilimcilerinin ve genel kamuoyunun dikkatini çekmiştir. MOOC'ların amacı, eğitimi ücretsiz yaparak tüm Dünya'dan erişilebilir hale getirmektir. MOOC'lar farklı yaş gruplarından, eğitimsel geçmişlerden ve milletlerden çeşitli öğrencileri kendilerine çekmişlerdir [45].

Bağlantıcı Kitlesele Çevrim İçi Açık Kurslar (cMOOC'lar) ile, öğrencilerin kendi ilgi alanlarına ve önceden edindikleri bilgi ve becerilere dayanarak, kursun hangi yönleriyle ilgilenmek istedikleri, ilgilenmeleri gerekenler, kendi öğrenmelerini yönetmeleri ve planlamaları konusunda daha büyük bir tartışma vardır. Ayrıca, cMOOC'larda öğrencilerin görevleri ve ödevleri herhangi bir sırayla tamamlayabilecekleri açık müfredat yapılarıyla, genellikle desteklenmeyen ve daha yapılandırılmış doğrusal kurslarda dahi bulunmayan yollarla öğrencilerin tekrar ziyaret etme ve gözden geçirme imkanları vardır [10].

Son yıllarda, MOOC'ların popülerliğindeki artış, hem bu kurslara kaydolana öğrenci sayısına hem de MOOC formatında ders veren üniversite sayısına yansımıştır. Bu kadar büyük bir ölçekte yüksek kaliteli eğitim deneyimlerini öğrencilere sağlama konusunda birçok zorluk olsa da, MOOC'lar, eğitim araştırmacılarına, öğrencilerin çevrim içi öğrenme yoluyla bilgilerini ve anlayışlarını nasıl geliştirdiklerini daha iyi anlamaları için önemli fırsatlar oluşturmuşlardır [10].

Popüler MOOC platformlarının bazı örnekleri aşağıdaki gibi listelenebilir;

- edX Platformu: EdX, efsanevi profesörler tarafından desteklenen, Harvard, MIT ve Berkeley (California) gibi saygın üniversiteler tarafından kullanılan ve kaliteli eğitim materyalleri ile kurslara genişletilmiş erişim imkânları sunan bir platformdur [46]. EdX platformu farklı alanlardan çok sayıda kursu tek bir yerde sunmaktadır. EdX kursları sertifika alma imkânı sunacak şekilde prestijli üniversiteler tarafından organize edilmektedir. Gerekli olan İnternet bağlantısı ile katılımcılar, tek bir yerden öğrenme materyallerine erişim, öğretmene danışma ve edinilen bilgilerin değerlendirilebilmesi imkanlarına sahip olmaktadır [47].
- Coursera Platformu: Coursera, Andrew Ng tarafından Stanford Üniversitesi'nden ayrılarak kurulmuş ve daha ilk yılında çok hızlı büyümeye ulaşan kar amacı gütmeyen MOOC sağlayıcı platform olarak ortaya çıkmıştır [48].
- Udacity Platformu: David Stavens, Thrun ve Michael Sokolsky ile birlikte Udacity platformunu kurmuştur. Platform, Dr. Thrun'un 'Introduction to

Artificial Intelligence' kursunun 150.000'in üzerinde kayıt alması ile devrim niteliğinde büyümeye başlamıştır [48].

- FutureLearn Platformu: FutureLearn, The Open University'nin sahibi olduğu ve Coursera ve edX gibi Amerikan platformlarıyla yarışan özel bir şirkettir. Platform, ilk kursunu Eylül 2013'te başlatmıştır ve ilk yılında 650.000'den fazla öğrenci platforma kayıt olmuş ve bu katılımcılar 1,4 milyonun üzerinde kursa kaydolmuşlardır [49]. FutureLearn, sohbetlerle öğrenmeyi teşvik eden ve sosyal yapılandırmacı yaklaşımı benimseyen MOOC platformlarından biridir [50]. İngiltere merkezli FutureLearn MOOC platformunun işbirliği yaptığı kuruluşlarından biri de Southampton Üniversitesi'dir [50].

1.1.5 Kitlese Çevrim İçi Açık Kurslarda Öğrenme Analitiği

Öğrenme Analitiği, birçok farklı uygulama alanına sahiptir ve Kitlese Açık Çevrim İçi Kurslarda (MOOC'larda) Öğrenme Analitiği, öğrenci davranışını anlama, performansını iyileştirme ve öğrenme sürecini görselleştirme için kullanılmaktadır.

Öğrenme Analitiği bir MOOC'da öğrenme aktivitelerini analiz ederek, gerekli müdahaleleri belirlemeye yardımcı olabilir [11].

Araştırmacılar, öğrencilerin aktiviteleri hakkında fikir sahibi olmak, ihtiyaçları tespit etmek ve olası tasarım değişikliklerini önermek için öğrenme analitiklerinin kullanımını araştırmaktadırlar. Öğrenme analitiklerini kullanmak, öğrencilerin ihtiyaçlarını tahmin etmede ve bu tür ihtiyaçlara verilebilecek cevapları uygun olarak belirleyebilmeye de yardımcı olabilir.

Öğrenme Analitiği, MOOC'larda öğrenci davranışını ve performansını anlamak için sistematik bir yaklaşım sunmaktadır. Öğrenme Analitiği'ni kullanmak, öğrenci davranışı ve performansının ölçülmesini, analiz edilmesini ve tahmininin hızlı bir şekilde yapılmasını sağlar [51].

Standardize testler ve sınavlar gibi öğrenci değerlendirme teknikleri ile bir zaman noktasında öğrencilerin belli özellikleri hakkında bilgi toplanabilmektedir. Öğrencilerin ilerlemelerini açıklamayı sağlayacak şekilde öğrenciler hakkında bilgi toplamak daha komplike tekniklerle onların aktivitelerinin sürekli olarak kayıt altına alınmasını gerektirir [52].

Literatürde, Öğrenme Analitiği'ni MOOC'larda kullanımı ile ilgili birçok çalışma mevcuttur. Örneğin, Ramesh ve diğerleri, kurs ilerlerken öğrencinin katılımının anlaşılmasının, öğrenme yollarını karakterize etmeye ve böylece odaklanmış eğitmen

müdahalesi ile kursu bırakma oranlarını en aza indirmeye yardımcı olacağını belirtmişlerdir [53]. Papadakis ve diğerleri tarafından yürütülen bir çalışma, MOOC platformunun içeriğinin ve satın alımlarının işlemlerinin sunum şeklinin, katılımın azalmasına neden olabilecek önemli bir faktör olduğunu göstermektedir [54]. Örneğin, eğitmen müdahalesini kolaylaştırmak için kursun ve oyunlaştırma unsurlarının mobil uygulamasının iyileştirilmesi, İngilizce'yi ikinci dil konuşmacıların kullanımları üzerinde etkili olabilir.

Bir kurs ilerledikçe öğrencinin katılımını anlamak, öğrencinin öğrenme modellerini karakterize etmeye ve kursu bırakma oranlarını en aza indirmeye yardımcı olur, eğitmen müdahalesi başlamasını sağlar [53].

1.1.6 Kitlese Açık Çevrim İçi Kurslarda Kullanım Analizi

Öğrencilerin sanal öğrenme ortamları üzerinden öğrenme materyalleri ile etkileşim kurdukları uzaktan eğitimde, makine öğrenmesi yöntemleri bütünleşik öğrenci kaynaklarına uygulanarak hangi öğrencilerin müdahalelerden ve rehberlikten daha fazla faydalanacakları ortaya konulduktan sonra onlara sağlanacak destek daha iyi yönetilebilir [55].

Klusener ve diğerleri öğrenme davranışını analiz etmek ve gerekli müdahaleleri belirlemeye yardımcı olacak yöntemleri ortaya koymak için forum faaliyetinden türetilen öğrenci profillerini kullanmışlardır [11]. Brinton ve diğerleri forum katılımının azalması ile ilişkili faktörleri incelemiş ve kurs konularını buna göre sıralamıştır [56]. Tanımlanan faktörleri ve iş bölümlerini sıralama mekanizmasının, MOOC'larda kişiselleştirilmiş önerilerde bulunmada kullanılabileceğini belirtmişlerdir. DeBoer ve Breslow, katılımcıların MOOC'lardaki her bir tıklamalarının, öğrenme süreçlerini ve öğrencinin kursa karşı tutumunu tahmin etmeye yardımcı olan çevrim içi davranışların bir parçası olduğunu ortaya koymuşlardır [57].

Kurs forumu, MOOC'ların önemli bir parçası ve öğrencilerin katılımı ve kurslara karşı tutumları gibi birçok şeyin bir göstergesidir. Sonuç olarak, kurs forumunun analizi MOOC'lar hakkında daha fazla bilgi sahibi olmamızı sağlar [9].

1.1.7 Kitlese Açık Çevrim İçi Kurslarda Performans Tahmini

Kitlese Açık Çevrim İçi Kurslarda (MOOC'lar) performans tahmini, kurs bırakma tahmini gibi en popüler alanlardan biridir. Öğrenci performansını önceden tahmin etmek, zamanında öğrenci performansını iyileştirmek ve kursu bırakmanın önüne geçmek veya başarısızlığı azaltmak için müdahalelerde bulunmak, hem öğrenciler

hem de ğretim kurumları iin faydalar saėlamaktadır. ğrencinin performansını arttırmak ve ayrılma/başarısızlık durumlarına zamanında müdahale etmek iin ğrenci performansı tahmini, hem ğrenciler hem de eėitim kurumları iin faydalıdır.

Araştırmacılar doėru tahminlerde bulunmak iin eşitli makine ğrenmesi tekniklerini kullanmaktadırlar. Makine ğrenme algoritmaları, katılımcıların MOOC'lardaki nceki katılım yollarından ğrenmekte ve gelecekteki katılımları (aktiviteleri) hakkında tahminde bulunmaktadırlar. Katılımcı kursu tamamlayacak mı?, Katılımcı sertifika alacak mı?, gibi [58].

MOOC'larda her tıklama ile araştırmacılar, ğrencilerin ders boyunca takip ettikleri yolları izleyebilmekte ve kurs materyallerinin kullanımının ğrencilerin başarıları üzerindeki etkisinin nedensel tahminlerine yaklaşabilmektedirler [59].

MOOC'larda, her etkileşim kayıt altına alındığı iin, kurs boyunca performansı modelleyebiliriz [57].

Literatürde ğrencilerin MOOC'larda olan performans analizine odaklanan birçok alışma ([10], [11], [13], [12], [60]) mevcuttur. Bunlardan biri, ğrencilerin ön bilgi ve becerilerinin derecesini ve MOOC ile olan ilişkilerini ğrenme Analitiėi ile ölçmüş ve MOOC sonu performansını tahmin etmiştir [10].

Bazı alışmalar, MOOC katılımcılarının sosyal katılımlarını anlamaya odaklanmışlardır. Örneėin, Tucker ve diėerleri forum aktivite kayıtlarından ğrenci performansını tahmin etmeyi göz önünde bulundurdular [44].

ğrenci performansı kurs ev ödevlerinde, testlerde ve sınavlarda elde edilen notlara göre belirlenir. Sınıf ii ğrenme ortamlarına benzer şekilde, MOOC'lara kayıtlı ğrenciler, genellikle kurs konularının ve ödevlerinin tartışılabilceėi ğrenme gruplarını kendileri organize eder ve oluşturur [44].

Bir alışma, MOOC'larda kurstan ayrılma ve performans tahminlerini daha doėru yapmak iin geleneksel özellikleri daha zengin ve daha fazla ayrıntılara sahip bilgilerle genişletmiştir [12].

Başka bir alışma, MOOC'lardaki ğrenme aktivite verilerini analiz ederek farklı ğrencilerin aktivite özellikleri arasındaki farkları bulmayı amaçlamıştır. alışmada, ğrencilerin MOOC'lardaki aktivitelerine dayanarak bunu gerçekleştirmek iin ğrenciler motivasyonlarına göre farklı gruplara ayrılmıştır. Ardından sertifika kazanan ğrencilere not tahmini uygulanmıştır. Tahminlemenin doėruluėu, sınıflandırma modelinin parametrelerinin daha fazla ğrenciye uyacak şekilde daha ince bir ölekte ayarlanabilmesi sayesinde iyileştirilmiştir [13].

Farklı bir çalışma, öğrencilerin Melbourne Üniversitesinde geliştirilen iki farklı MOOC olan 'Makroekonomi İlkeleri' ve 'Ayrık Optimizasyon' kurslarının açılış oturumlarındaki etkileşimlerini incelemiştir. Bu iki kurs, üretim ekibi tarafından geliştirilip aynı MOOC platformunda yayınlanmasına rağmen kurs yapıları ve uygulamaları tamamen farklı olduğu için seçilmişlerdir. Kurslar isim, konu alanı, ön koşul bilgileri, müfredat tasarımı ve ödev tasarımı gibi sadece birkaç yönden çeşitlilik göstermiştir. Çalışma, öğrencilerin ilk iki haftadaki notlarının kurstaki final notunun iyi bir göstergesi hipotezini doğrulamayı amaçlamıştır [14].

Ayrıca, öğrenci kurs performansının erkenden tahmin edilebilmesini sağlayan göstergeleri tespit etmeye odaklanarak Moodle aktivite verilerini analiz eden farklı bir çalışma da vardır [61]. Çalışmadaki analizler, ilgili üç hipotezin büyük ölçüde veriler tarafından desteklendiğini göstermiştir. Çalışmanın sonuçlarına göre katılımcının erken yükleme yapması iyi bir işarettir, yüksek düzeyde aktivitede iyi sonuçlar alınacağı öngörülmektedir ve akşam aktivitesi gündüz aktivitesinden daha iyidir. Çalışmanın ana katkısı, iyi veya kötü öğrenci performansının erken göstergelerini sağlayıp sağlamayacağını belirlemek için büyük miktarda Moodle aktivite verisinin analiz edilmiş olmasıdır [61].

Alternatif bir çalışma, motivasyon ve katılımın öğrencilerin bir MOOC'daki performansına etkisini, özellikle de MOOC'un sonuna kadar devam eden öğrencileri nasıl etkilediğini incelemiştir. Çalışma sonuçlarına göre, performansın en güçlü belirleyicisinin katılım olduğu ve ardından bunu motivasyonun takip ettiği bulunmuştur. Motivasyon, kurs boyunca hem öğrencilerin katılımlarından etkilenmiş hem de öğrencilerin katılımlarını etkilemiştir. Dahası, durumsal ilgi, genel içsel motivasyon ve katılımın performans üzerindeki etkisine aracılık etmede çok önemli bir rol oynamıştır [62].

Başka bir çalışma, en üst düzeyde katılımda bulunanları karakterize etmek ve erken tespitlerini kolaylaştırmayı amaçlayan bazı öngörülerini sağlamak için gerçek bir MOOC ile ilgili beş farklı sosyal araçtan gelen ampirik verileri analiz etmiş ve raporlamıştır. Bu analizin sonuçları, en çok katılımda bulunanların diğerlerinden daha iyi final notlarına sahip olduklarını göstermiştir. Ayrıca, katılımcıların genel performansları (final puanlarına göre ölçülmüştür) ile beş sosyal araca gönderdikleri mesaj sayıları arasında orta düzeyde pozitif bir korelasyon gözlemlenmiştir [63].

Ayrıca, öğrencilerin forum aktivite kayıtlarını kullanarak öğrenci performansını analiz eden çalışmalar da vardır. Bunlardan biri, öğrencilerin performans ve öğrenme çıktıları üzerindeki etkilerini ölçmek amacıyla MOOC'larda bulunan öğrenciler tarafından oluşturulan metinsel verileri (çevrim içi tartışma forumları gibi) işlemeyi

hedeflemiştir [44].

Bir çalışmada, başarılı öğrencilerin özellikleri forum aktivitelerinden türetilmiş ve bir öğrenme profiliyle birleştirilmiştir. Bu öğrenme profilinden, 'riskli öğrenci' olarak sınıflandırılan öğrenciler için geri bildirim üretilebileceği ifade edilmiştir. Iversity MOOC'larındaki öğrencileri, bir forumdaki cevap sayıları veya oy sayıları gibi özellikleri kullanarak sınıflandıran, makine öğrenmesine dayalı bir analitik aracı geliştirmişlerdir. Bu araç sayesinde başarılı öğrencilerin özellikleri sezgisel bir şekilde belirlenebilir ve görselleştirilebilir olduğunu belirtmişlerdir [11].

MOOC forum kayıtlarını kullanarak öğrenci performansını analiz eden bir başka çalışma, öğrencilerin akademik performansları ile ders forumuna katılımları arasındaki ilişki üzerine odaklanmıştır. Çalışma, ayrıca forumdaki davranışları anlamak için hem öğrencilerin hem de öğretim asistanlarının ders forumundaki gönderilerinin anlamlarını da incelemiştir. Çalışmada bulunan sonuç daha yüksek puan alanların forumda daha aktif olma eğiliminde olduklarıdır. Bununla birlikte yüksek puan alanların aynı zamanda, puanları düşük olanlara göre daha yüksek bir oranda kursla ilgisi olmayan yorumlar da yazmaktadırlar. Öğretim asistanları forumda çok aktif değildirler ve sınırlı etkileri vardır [9].

Bir öğrencinin MOOC'u tamamlama olasılığını kesin olarak anlamak için performansını tahmin etmek ve öncelikle hangi bilgilerin gerekli olduğunun belirlenmesi gerekir. Bu amaçla, bilgi ve bunun elde edildiği verileri biçimlendirmek için bir dizi adım gerçekleştirilmesi gerekmektedir [64].

1.1.8 Kitlesel Açık Çevrim İçi Kurslarda Derin Öğrenme Kullanımı

MOOC'lar binlerce katılımcıdan elde edilen büyük veriler ürettiği için, literatürde Derin Öğrenme'nin MOOC araştırmalarına farklı amaçlarla adapte edildiğine dair örnekler vardır.

Örneğin, Xi ve Duhave, risk altındaki öğrencilerin bireysel kurs bırakma olasılıkları ile kişiselleştirme ve önceliklendirmenin gerçekleştirilmesi için Derin Öğrenme yöntemlerini kullanmışlardır [65]. Seçilen temel algoritmalarla karşılaştırıldığında Derin Öğrenme her hafta için her bir bireyin kursu bırakma ihtimalini tahmin etmede en başarılı algoritma olarak bulunmuştur.

Wang ve diğerleri Konvolüsyonel Sinir Ağları'nı ve Tekrarlamalı Sinir Ağları'nı birleştiren bir Derin Öğrenme modeli (ConRec) önermişlerdir [66]. Çalışmalarında önerdikleri Derin Öğrenme modeli en iyi sonuçları vermemiştir. Seçilen bazı sınıflandırma algoritmaları, Derin Öğrenme modelinden biraz daha iyi performans

göstermiştir. Bununla birlikte çalışmanın sonuçları, Derin Öğrenme'nin kullanımının, ham verinin kullanılması durumunda özellik mühendisliği yapılmadan oldukça verimli bir model olduğunu da göstermektedir.

Derin Öğrenme, yalnızca kursu tamamlama tahmini için kullanılan bir yöntem değildir. Tang ve Pardos, bir katılımcının hangi kurs sayfasına geçiş yapacağını tahmin etmek için Tekrarlı Sinir Ağları'nı (RNN) kullanmışlardır [67]. 13 edX MOOC'una uygulama sonuçlarına göre, önerilen model çoğunlukla temel modellerden daha iyi performans göstermiştir.

Yang ve diğerleri kurs notu tahmini için bir Zaman Serisi Yapay Sinir Ağı'nda video izleme tıklama akışları ve ödev notlarının bir kombinasyonunu kullanmışlardır [60]. Yazarlar, davranışların hiçbirinin özellikle performansla korelasyonu olmadığını, ancak bunların kombinasyonunun tahminsel analizlerde kullanım için efektif olduğunu göstermişlerdir. Bu örneklerde, kullanıcıların bir materyalde harcadıkları zaman, belli bir sayfanın görüntülenme sayısı veya belli araçları hiç kullanıp kullanmadıkları gibi sayısal veya ikili sayı (Boolean) veriler kullanılırken, tez kapsamında daha fazla, katılımcılar tarafından gönderilen yorumların Derin Öğrenme yöntemi ile kurs tamamlama başarılarının sınıflandırılmasındaki potansiyel kullanımıyla ilgilenilmiştir.

Nümerik (sayısal) verilerin ötesinde, tartışma forumlarındaki gönderiler katılımcıların kurs ve içeriği ile ilgili düşüncelerinin, sorularının ve duygularının yansıması ile ilgili daha zengin veriler sağlamaktadır. Bu nedenle, katılımcıların tutumlarını anlamak, ayrılma durumlarını tahmin etmek ve yardıma ihtiyacı olanları belirlemek için tartışma forumlarının içeriğinin kullanımı ile ilgilenen çalışmalar da yapılmıştır [68].

Geleneksel ve modern makine öğrenme tekniklerinin yanı sıra, forum yorumlarından elde edilen metin verilerine Derin Öğrenme yöntemi uygulanmıştır. Örneğin, Chaplot ve diğerleri, öğrencinin kurstan ayrılacağı haftayı tam olarak belirlemek için forum gönderilerine anlam analizi uygulamıştır [69]. Yazarlar, önerilen sinir ağını beslemek için nümerik veri ile birlikte yorumların anlam sonucunu da kullanmışlar ve algoritmayı esnek yayılma sezgiselliği ile eğitmişlerdir.

Wei ve diğerleri, farklı türde MOOC verilerindeki forum gönderilerinin kargaşa ve aciliyet içeren gibi sınıflandırılması için Konvolüsyonel Yapay Sinir Ağı'na ve Uzun-Kısa Süreli Yapay Sinir Ağı'na dayanan Derin Öğrenme modeli oluşturmuş (ConvL) ve gelecek iyileştirmeler için umut verici sonuçlar elde etmişlerdir [42].

Özetle, Derin Öğrenme, şuan için MOOC'larda risk altındaki öğrencileri tahmin etmek için diğer tüm sınıflandırma algoritmalarından daha iyi sonuç veren mucizevi

çözüm değildir. Bununla birlikte, Derin Öğrenme, etiketli olmayan ham verilerin ulaşılabilir olduğu (tartışma forumlarındaki öğrencilerin mesajları gibi) bazı MOOC içerikleri üzerinde daha iyi çalışan yöntemlerden biridir. İncelenen çalışmalar, özellik çıkarımı yapmadan kursun tamamlanmasının tahmin edilmesinde kullanılacak forum yazılarının anlamsal analizinde muhtemel çeşitli Derin Öğrenme uygulamalarının yolunu açmaktadırlar.

Mevcut literatür, MOOC'ların Derin Öğrenme'den nasıl yararlandığı hakkında örnekler sunmaktadır. Biz araştırmamızda, farklı MOOC'larda ikincil dil olarak İngilizce konuşanların kurs performans tahmini için Derin Öğrenme'den faydalanmayı amaçladık. Bu tez çalışmasında, sıkça kullanılan birkaç Derin Öğrenme modeli uygulanmıştır. Bu modeller Konvolüsyonel Sinir Ağları (CNN), Uzun-kısa süreli bellek (LSTM), İki yönlü Uzun-kısa süreli bellek (Bidirectional LSTM), Kapılı Tekrarlamalı Ünite (GRU), Konvolüsyonel Sinir Ağları (CNN) ve Uzun-kısa süreli bellek (LSTM) birlikte kullanımı (Konvolüsyon ve LSTM) modelleridir.

1.1.9 Dil Perspektiften Kitlese Açık Çevrimiçi Kursların Analizi

MOOC adı verilen, yaygın olarak kullanılan uzaktan eğitim platformlarının çoğunda, kursların dili İngilizce'dir, ancak yine de birçok ülkeden katılımcıları vardır. Bu durum katılımcıların kullanım davranışlarında ve performanslarında farklılıklara neden olmaktadır.

Dil Perspektifinden MOOC'lardaki mevcut çalışmalar dört ana temaya ayrılabilir:

1. İngilizce dışında bir dilde oluşturulan bir MOOC üzerinde katılımcıların etkileşimleri [70], [71], [72], [73]
2. İngilizce dilinde oluşturulan bir MOOC üzerinde İngilizce dili öğrenenlerin etkileşimleri [74], [75], [76]
3. ikincil dil olarak İngilizce konuşan katılımcıların İngilizce öğrenmeleri için oluşturulan bir dil MOOC'u üzerindeki etkileşimleri [77], [78], [79], [80]
4. İngilizce'nin ikincil dili olduğu kullanıcılar için İngilizce dilini kullanarak oluşturulmuş bir MOOC üzerindeki etkileşimleri [28], [81], [82], [83], [84].

Uchidiuno ve diğerleri, çalışmalarında MOOC'ların erişilebilirliği hakkında uluslararası öğrencilerin motivasyonlarını anlamak için onlarla röportaj yapmışlardır [85]. Yazarlar içeriğin tercüme edilmesinin bir çözüm olabileceğini önerirken, bunun herkes için uygun olmayabileceğini de belirtmişlerdir. Bununla birlikte, ikinci dil

olarak İngilizce konuşan katılımcıların ihtiyaçlarına ve motivasyonuna dayanarak uyarlanmış araçların daha etkili olabileceğini ifade etmektedirler.

Uchidiuno ve diğerleri, bir başka çalışmalarında İngilizce olarak verilen MOOC'lara katılan İngilizce dili öğrenenlerin katılımını (etkileşimini) araştırmışlardır [75]. Çalışmada, İngilizce, bu dil öğrenenlerin ilk dili olmasa da, katılımcılar profesyonel olarak ve motivasyonla dil ile ilgilendiklerini tespit etmişlerdir. Bu nedenle, İngilizce öğrenenlerin özel olarak araştırılması gerekliliğine ihtiyaç duyulduğunu savunmuşlardır.

Başka bir örnekte, Colas ve diğerleri, MOOC'larda, ek olarak başka dillerdeki tartışma forumlarının yer almasının, katılımı bir iyileşme ile sonuçlandığını gözlemlemişlerdir [86]. Ancak, kullandıkları forum tartışmalarını izlemek için yedi rehber ekibini işe alma yöntemleri genişletilebilir değildir. Ek olarak, vaka çalışmaları MOOC'ların dar bir perspektifine odaklanmıştır [86].

Eriksson ve diğerleri öğrencilerin kurstan ayrılma sebeplerini inceleyen çalışmalarında, bazı öğrencilerin videodaki konuşma dilini ve bazen de öğretmenlerin aksanını anlamada zorluk çektiklerini ortaya koymuşlardır [20]. Bu sonucun aksine, Aboshady ve diğerleri, Mısırlı tıp öğrencilerinin İngilizce olarak sunulan bir MOOC'a bakış açılarını incelemelerinde İngilizce'yi önemli bir engel olarak bulmamışlardır (katılan öğrencilerin % 20'sinden azı zorlanmıştır) [87]. Bu durum tıp öğrencilerinin aldıkları ileri düzey eğitimden kaynaklanmış olabilir [87].

Uchidiuno ve diğerleri tarafından yapılan farklı bir çalışma İngilizce dili öğrenenlerin, İngilizce ilk dili olanlarla karşılaştırıldığında, metin içeriği ile daha fazla etkileşim gösterirken, video ve görsel destek olmadığında içerikle daha az etkileşimde bulunduklarını göstermişlerdir [76].

Cho ve Byun, ikincil dil olarak İngilizce konuşan katılımcıların İngilizce olarak sunulan bir MOOC'daki katılımları üzerine çalışma yapmışlardır [28]. Çalışmalarında, 24 Koreli üniversite öğrencisinin tecrübelerini incelemişlerdir. Aynı zamanda, dilin, ikincil dil olarak İngilizce konuşan katılımcılar arasında aktif katılım için potansiyel bir engel olduğunu tespit etmişlerdir. Ek olarak, kültürel olarak tanınmayan öğretme ve öğrenme uygulamaları ikincil dil olarak İngilizce konuşan katılımcılar için bazı zorluklar ortaya koymaktadırlar [28]. Bu tür engeller, özellikle tartışma forumları gibi nispeten yapılandırılmamış faaliyetlerde yüksek olabilir. Bu nedenle, ikincil dil olarak İngilizce konuşan katılımcıların başarılı olmak için ek desteğe ihtiyaçları olabilir.

Rimbaud ve diğerleri, MOOC'larda ikincil dil olarak İngilizce konuşanları destekleyecek adaptif MOOC eksikliğine dikkat çekmişlerdir [88]. Çalışmalarında,

“İçerik ve Dille Bütünleşik Öğrenme (CLIL)” nin ikincil dil olarak İngilizce konuşan öğrencilerin ihtiyaçları için bir çözüm olabileceğini önermişlerdir. De Waard ve Demeulenaere [89], 5. sınıf K-12 öğrencileri için dil, sosyal ve çevrim içi öğrenme becerilerini artırmak için MOOC’ları ve CLIL yöntemini karma bir öğrenme ortamında birleştirmişlerdir. Aynı zamanda, ikincil dil olarak İngilizce konuşan öğrenciler için adaptif bir desteğin faydalı olabileceğine dair bazı kanıtlar sunmuşlardır.

Reilly ve diğerleri, otomatik kompozisyon notlandırma sistemlerini test etmiş ve ikincil dil olarak İngilizce konuşan öğrencilerin dezavantajlı olduklarını gözlemlemişlerdir. Bu durumun nedeni otomatik notlandırma sistemi tarafından verilen notların, insan eğitmenler tarafından verilen notlardan önemli ölçüde düşük olmasından kaynaklandığını belirtmişlerdir [90]. Yazarlar, MOOC’ların bu dezavantajı ele almaları ve çok kültürlü ve dilsel açıdan farklı katılımcılar için önlemler almaları gerektiğini ifade etmişlerdir.

Yakın tarihli bir çalışmada Calvo ve diğerleri, uluslararası katılımcılardan elde edilen verilere dayanan sonuçlara göre, sosyal girişimcilik eğitime odaklanan bir MOOC’daki, dil ve kültür tabanlı engelleri, katılımcıların MOOC’lara erişimini engelleyici olarak tespit etmişlerdir [91].

Daha sonraki MOOC’ların erişilebilirliklerinin önündeki engelleri ortaya koymakla ilgili davranışsal ve uygulama çalışmaları, kültürel ve dilsel konulara odaklanmışlardır. MOOC katılımcılarının büyük bir bölümünün gelişmekte olan ülkeler [16] kökenli oldukları açıktır. Bu sebeple, MOOC sağlayıcıları tarafından, büyümekte olan bu zengin eğitim kaynaklarının etkin bir şekilde kullanılmasını artırmak için çalışmalar yapılması gerekmektedir [16]. Kullanıcılarla ilişkili temel özellikleri otomatik olarak tanımlamanın yollarını bulma onların yaklaşık dilsel geçmişleri gibi, MOOC platform sağlayıcılarına ve kurs yazan ekiplerinin kişiselleştirmeye yönelik geniş yünden yaklaşımları gerçekçi bir şekilde ele almalarını sağlar. Ayrıca, bu yaklaşım potansiyel olarak etkili yerelleştirmeyi sağlamak için veri üretilmesi için kullanılabilir [92]]. Sosyal olarak aktif katılımcıların (çevrim içi forumları olan MOOC’larda) kursu tamamlaması muhtemeldir [7].

Literatürdeki çalışmalar göz önüne alındığında, araştırmacıların MOOC’lardaki ikincil dil olarak İngilizce konuşan katılımcıların içerikle nasıl etkileşim kurdukları, katılımcıların ihtiyaçları ve motivasyonları ile ilgilendikleri görülmektedir. Ancak, ikincil dil olarak İngilizce konuşan katılımcıların ihtiyaçlarını daha iyi anlayabilmek için bu alanın hala derinlemesine incelenmesi gerekmektedir. MOOC’lar, ikincil dil olarak İngilizce konuşan katılımcıların ihtiyaçlarına göre doğru yardım sağlayacak şekilde iyileştirilebilirler.

Tez çalışmalarından birinde, birincil dili İngilizce'den farklı olan yani ikinci dil olarak İngilizce konuşan katılımcıları tespiti ile ilgili zorluklar ortaya konulmuştur [23]. Oluşturulan düzenli ifadelerle İngilizce dil gruplarını tespit yöntemi, çalışmada ortaya bulguları destekler nitelikte olup, kullanıcıların İngilizce dil gruplarının tespitini otomatikleştirmek için tüm kullanıcı yorumlarında yer alan ülke, şehir, millet ve dil bilgilerini tespit edebilecek düzenli ifadelerin oluşturulması ve yorumlara uygulanması ve gerekli olan durumlarda İngilizce dil gruplarının güncellenmesi üzerinedir. Daha sonra çeşitli konulardaki MOOC'ların elde edilen büyük verilere Derin Öğrenme yöntemi uygulanarak İngilizce dil grubu tahmini de denenmiştir.

1.2 Tezin Amacı ve Yöntem

Kitlesel Açık Çevrim İçi Kurslar (MOOC'lar) farklı coğrafi konumlardan milyonlarca insanı kendine çekmektedir. MOOC'lar çoğunlukla İngilizce dilinde hazırlanırken, katılımcıların bazıları İngilizce'yi yabancı (ikincil) dil olarak konuşmaktadırlar.

MOOC'lardaki bazı katılımcıların video derslerinde dili anlama konusunda zorlandıklarını ve diğer öğrencilerle iletişim kurmakta isteksiz olduklarını bildiren çalışmalar mevcuttur. Aynı zamanda, MOOC'larda çeşitli zorluklar yaşayan katılımcılara yardımcı olmak için bazı kişiselleştirme servisleri de önerilmiştir.

MOOC'larda, İngilizce dil bilgisi eksikliği sebebi ile zorluk yaşayan katılımcılara destek sağlanmalıdır. Bunu gerçekleştirebilmek içinse, ilk olarak İngilizce'yi ikinci dil olarak konuşan katılımcıların tespit edilmeleri gerekmektedir.

Tezde, çeşitli kültürel geçmişlerden gelen ve çoğunlukla birincil dilleri İngilizce olmayan uluslararası MOOC katılımcıların İngilizce dil bilgilerinin farklılıkları göz önünde bulundurulmuştur. Bu sebeple, ilk olarak farklı türdeki katılımcıları kendi İngilizce dil geçmişlerine göre otomatik olarak tespit etme yöntemlerini belirlemeye daha sonrada olası öğrenme yollarını ve kurs performanslarını tahmin etmeye odaklanılmıştır.

Tezde ortaya konulan İngilizce dil gruplarındaki kullanıcıların davranışsal analiz sonuçları, farklı İngilizce dil gruplarından kullanıcıların kursu tamamlama performanslarını tahmininde güvenilir bir yöntem belirlemek için kullanılmıştır. Davranış analiz sonuçları, aynı zamanda katılımcıların MOOC'lara katılımlarını sürdürmeleri için gerekli çalışma stratejilerini belirlemek için de değerli olabilir.

Literatürdeki çalışmalarda, MOOC katılımcılarının sosyal etkileşimlerinin önemi ile ilgili bazı bulgular ortaya konulmuştur. Bu bulgulardan, İngilizce dili odaklı kullanım

analizlerinde sosyal etkileşimleri de incelenin katılımcıların davranışlarını daha iyi anlamayı sağlayabileceği kanaatine varılmıştır. Tezde, literatürdeki çalışmalardan farklı olarak, MOOC'lardaki ikincil dil olarak İngilizce konuşan katılımcıların davranışları farklı açılardan detaylı olarak incelenmiştir. Tezin amaçlarından biri, farklı İngilizce dil gruplarındaki katılımcıların, özellikle de, MOOC içeriğinin dilinden farklı birincil dilleri olanların sosyal etkileşimlerini araştırmaktır.

Tezde aynı zamanda, İngilizce'yi ikinci dil olarak konuşan katılımcıların MOOC'lardaki ihtiyaçları ve davranışları tespit edilerek, literatürdeki araştırmalara katkıda bulunmak da amaçlamaktadır. Tezdeki bulgular, İngilizce'yi ikinci dil olarak konuşan katılımcılara ve MOOC'ları harmanlanmış kampüs eğitimlerinde yabancı dilde bir materyal olarak kullanan kurs eğitmenlerine yardımcı olabilir.

Tezin amaçları;

- FutureLearn verisinden katılımcıların İngilizce dil gruplarının otomatik olarak tespiti
- Farklı İngilizce dil gruplarındaki katılımcıların MOOC'lardaki davranışlarını incelemek ve kullanım yollarını belirlemek
- Farklı İngilizce dil gruplarındaki katılımcıların, kursu bırakma veya sertifika kazanma durumlarını tahmin etmek
- Katılımcıların İngilizce dil gruplarına göre MOOC kurslarındaki geçmiş aktivitelerinden performanslarını tahmin ederek kişiselleştirilmiş bir müdahaleyi mümkün hale getirmek

Tezde aşağıdaki araştırmalara katkıda bulunmak istenilmektedir:

- MOOC'lardaki forum mesajlarının kullanımı
- Farklı Derin Öğrenme modellerini kullanarak kurs performansı ve İngilizce dil grubu tahmin başarılarını karşılaştırmak
- Farklı disiplinlerden olan MOOC'larda çeşitli Derin Öğrenme modellerini deneyerek en uygun mimarının belirlenmesi ve sonuçların yorumlanması

Tezde aşağıdaki araştırma soruları incelenmiştir;

- İngilizce'yi ikinci dil olarak konuşan katılımcılar ile diğerleri arasında kurs tamamlama, kurs forumlarına katılım ve diğer katılımcılarla etkileşim davranışları yönünden bir fark var mı?
- İngilizce dil gruplarının kullanımları arasında farklılıklar varsa, bu farklılıklar bir kurs sonu performansı tahmin modeli oluşturmak için kullanılabilir mi?
- Düzenli ifade kalıpları; dilleri, uyrukları veya ülkeleri veya şehirleriyle ilgili bilgilerin mevcut olması durumunda, bir MOOC'da bulunan katılımcıların kendi yorumlarından İngilizce dil gruplarını tespit sürecini otomatikleştirmek için kullanılabilir mi?
- Derin Öğrenme, bir MOOC'da bulunan katılımcıların dilleri, uyrukları veya ülkeleriyle ilgili herhangi bir bilgi olmaması durumunda İngilizce dil gruplarının tespitinde kullanılabilir mi?

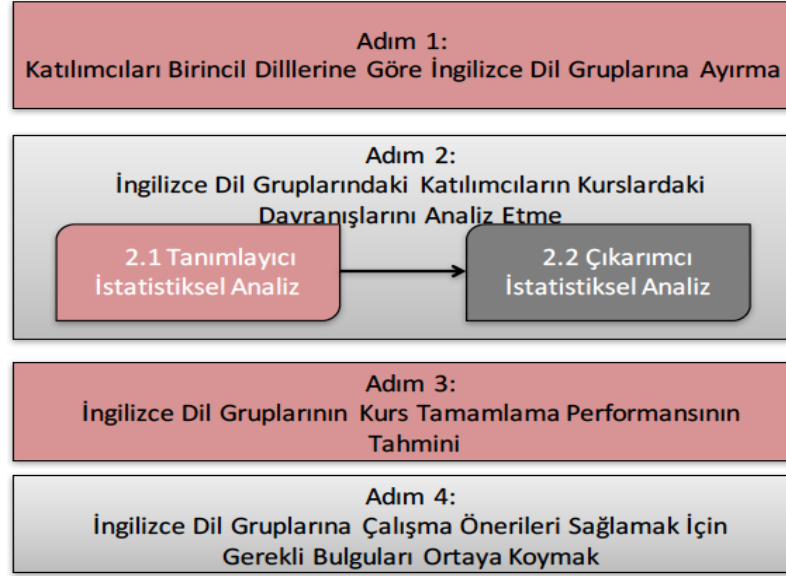
Metodolojiyi açıklamaya veri hakkında kısa bir bilgilendirme ile başlanacaktır.

Bu araştırma, Yıldız Teknik Üniversitesi ve Southampton Üniversitesi'nin etik olarak onaylanmış bir işbirliği ile gerçekleştirilmiştir. İngiltere merkezli FutureLearn MOOC platformu, Southampton Üniversitesi tarafından sunulan her bir MOOC için bir veri kümesi sunmaktadır.

Tezde uygulanan metodoloji, İngilizce dil gruplarının tespit edilmesi ve güncellenmesi, katılımcı davranışlarının analiz edilmesi, performanslarının tahmin edilmesi için en uygun modelin belirlenmesi, tahmin modelini iyileştirme, katılımcılar ve MOOC platformu geliştiricileri tarafından kullanım stratejisi geliştirmede kullanılabilecek faydalı bulgular üretme gibi birkaç alt aşamayı içermektedir.

Şekil 1.1'de, tezde uygulanan metodolojinin dört ana adımı gösterilmektedir. Birinci adımda, katılımcıların İngilizce dil gruplarını tespit yöntemleri önerildikten sonra, ikinci adımda her gruptaki katılımcıların ayrıntılı olarak kursları kullanım analizleri gerçekleştirilmiştir. Üçüncü adımda, kullanım analizlerinden elde edilen bulgulara dayanarak sık kullanılan geleneksel makine öğrenme algoritmalarına ve oluşturulan Derin Öğrenme modellerine dayalı performans tahminleri gerçekleştirilmiştir. Performans tahminleri, ilk olarak kurs sonuna kadar olan aktivite kullanım verileri ile ardından da ilk haftadan başlayarak kümülatif haftalık kullanım verileri ile gerçekleştirilmiştir. Dördüncü adımda ise, MOOC katılımcılarının İngilizce dil gruplarına göre çalışma stratejileri önerilebilmesi için gerçekleştirilen analizlerden elde edilen bulgular paylaşılmıştır.

Tezde, çoğunlukla FutureLearn platformuna ait "Understanding Language: Learning and Teaching" MOOC serisinin ilk sekiz iterasyonuna ve aynı platforma ait farklı türlerden MOOC'lara ait veriler kullanarak analizler gerçekleştirilmiştir. Tez kapsamında aynı zamanda, katılımcıların FutureLearn platformundaki sosyal etkileşimleri de detaylı olarak analiz edilmiştir.



Şekil 1.1 Araştırmanın dört adımı

Araştırmanın ilk adımında, katılımcıların İngilizce dil gruplarına ayrılması gerçekleştirilmiştir. FutureLearn platformunda (diğer uzaktan eğitim platformlarında da olduğu gibi), katılımcılar kayıt olurken ülke bilgilerini vermek zorunda değillerdir. İngilizce dil gruplarına karar vermek için ilk aşamada, katılımcıların kurs öncesi anket cevaplarındaki ülke bilgileri kullanılmıştır. İkinci aşamada, MOOC katılımcıların kurs öncesi anket cevaplarında yer alan lokasyon (ülke) bilgileri ve kurs forumlarında yaptıkları yorumlar birlikte değerlendirilerek İngilizce dil grupları güncellenmiştir. Katılımcıların yorumlarının incelenmesi sonuçlarına göre, bazı kullanıcıların, kurs öncesi anket cevaplarında verdikleri ülke bilgilerinin kursu kullandıkları sırada yaşadıkları yerlere ait olup, kendi ülkelerinden farklı olduğu anlaşılmıştır. Bu nedenle, İngilizce dil gruplarına ayırmanın ikinci aşamasında, daha doğru gruplandırma yapma ve tezin sonraki adımlardaki analizlerin güvenilirliği için, tartışma forumlarındaki katılımcı yorumlarından düzenli ifadeleri kullanarak İngilizce dil grupları tespit edilmiştir. Katılımcının tespit edilen İngilizce dil grubu, daha önceden kurs öncesi anketinde verdiği ülke bilgisinden elde edilen İngilizce dil grubundan farklı ise katılımcının İngilizce dil grubu güncellenmiştir. Oluşturulan düzenli ifade kalıpları ile İngilizce dil gruplarını güncellemek için çoğunlukla katılımcıların ülke, millet, dil ve şehir gibi bilgileri kullanılmıştır. Düzenli ifadelerle İngilizce dil grubu

tahmini algoritması, yalnızca birincil dil, millet, ülke, şehir gibi bilgilerini paylaşan katılımcıları tespit edebilmektedir. Bununla birlikte, katılımcılar farklı türden kurslarda kendilerini ifade etmek için kullanabilecekleri başka cümle yapılarını tespit etmek için farklı düzenli ifade kalıplarına da ihtiyaç duyulabilir. Derin Öğrenme yöntemiyle ise, katılımcıların yorumlarında belirli bilgileri paylaşıp paylaşmamalarına bağlı olmadan katılımcıların İngilizce dil gruplarının tespit edilebilmesi mümkündür. Bu nedenle, İngilizce dil gruplarına ayırmanın üçüncü aşaması olarak, oluşturulan Derin Öğrenme modelleri (farklı mimariler ile birlikte) denenmiştir.

Araştırmanın ikinci adımında, İngilizce dil gruplarındaki katılımcıların davranışlarını incelemek için Öğrenme Analitiği teknikleri kullanılmıştır. Bunun için, R programlama dili (başlangıçta MySQL programlama dili de denenmiştir.) ve Veri Madenciliği teknikleri ile FutureLearn platformu verilerinde yer alan kullanıcıların etkileşimleri ile ilgili özelliklerden faydalananarak, kullanım ve performans analizlerinde ihtiyaç duyulan özellikler oluşturulmuştur. Araştırmanın ikinci adımında, gerçekleştirilen İngilizce dil gruplarının davranış analizlerinde farklı İngilizce dil gruplarındaki katılımcıların bazen benzer davranışlar gösterirken diğer durumlarda katılımcıların davranışlarında farklılıklar görülmüştür.

Araştırmamızın üçüncü adımı, İngilizce dil gruplarındaki katılımcıların davranışlarındaki farklılıkları kullanarak kurs sonu performans tahmini için en uygun modeli belirleyip, katılımcıların kursu tamamlama durumlarını tahmin etmek için kullanılmasıdır. Bu sebeple, ilk olarak katılımcıların davranışlarını incelenmiş ve kurs performans tahmini için gerekli özellikler çıkarılmıştır.

Tezde, kurs performans tahminleri ilk olarak aşağıda listelenen yaygın olarak kullanılan sınıflandırma algoritmalarıyla gerçekleştirilmiştir. Bu algoritmalar;

- Çok Katmanlı Algılayıcı (MLP);
- Karar Ağacı (DT);
- K-En Yakın Komşu (KNN);
- Çok Kategorili Lojistik Regresyon (MLR);
- Naif (Naive) Bayes (NB);
- Rastgele Orman (RF);
- Destek Vektör Makinesi (SVM).

Modellerin eğitimleri için, 10-kat çapraz doğrulama yöntemi kullanılmıştır.

MOOC'lar, katılımcıların aktivitelerinden büyük miktarda veri ürettiği için, Derin Öğrenme modelinin, MOOC katılımcılarının gelecekteki performanslarını tahmin etmek için yararlı bir araç olabileceği düşünülmüştür. Bu sebeple, katılımcıların kurs sonu performans tahminleri için, Derin öğrenme modelleri (farklı mimariler de denenerek) önerilmiş ve önerilen modellerle elde edilen sonuçlar, makine öğrenme algoritmaları ile karşılaştırılmıştır.

Aynı zamanda, kurs devam ederken, başarısız olması muhtemel katılımcıları erkenden tespit edebilmek adına, ilk hafta verisinden başlayarak, kümülatif haftalık kullanım verileri ile (ilk hafta verisi ile, ilk iki hafta verisi ile gibi) kurs performans tahminleri de gerçekleştirilmiştir.

Araştırmanın son adımında, MOOC katılımcıları için (özellikle de İngilizce'yi ikinci dil olarak konuşan katılımcılar için) çalışma stratejilerini oluşturmada kullanılabilecek bulguların ortaya konulmasına odaklanılmıştır. Bu bulgular, tezde gerçekleştirilen İngilizce dil gruplarının tespiti, kurs kullanım ve performans analizlerinin sonuçlarından elde edilebilecek bulgulardır. Ortaya konulan bulguların, katılımcılara doğru çalışma stratejisi önermede değerli olabileceğine inanılmaktadır.

1.3 Tezin Katkıları

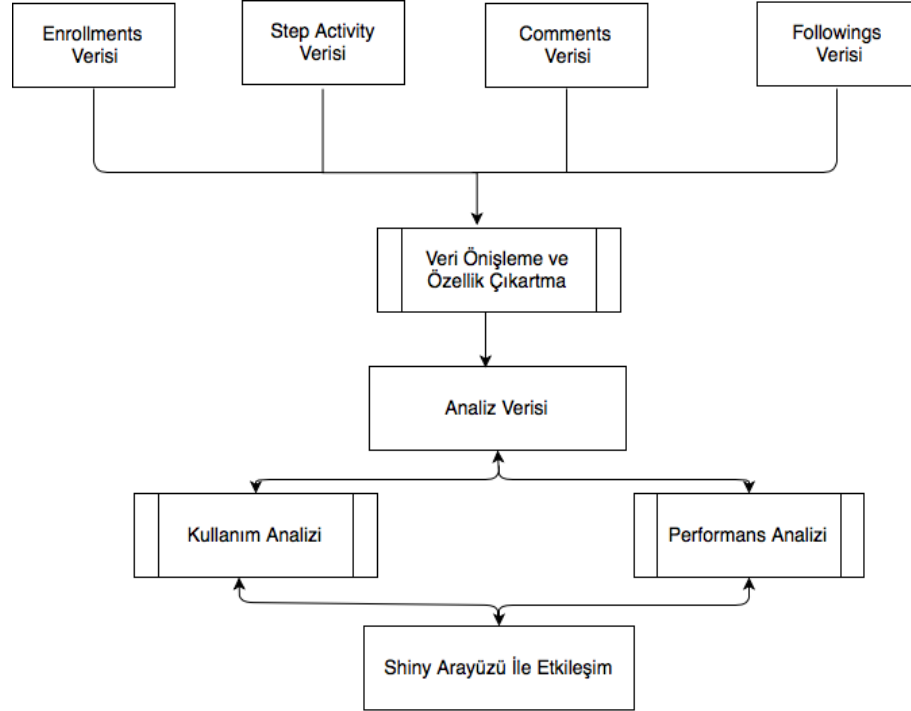
Literatürde, bildiğimiz kadarıyla, MOOC'larda İngilizce'yi ikincil dil olarak konuşanlar ile birincil veya resmi dil olarak konuşanlar arasındaki farkları, bu farklılıkların oluşturduğu problemleri ve olası çözüm yollarını araştıran yeterli çalışma yoktur. Bu nedenle, tezde bu sorun ele alınmıştır. İlk olarak, İngilizce'yi ikincil dil olarak konuşanlar ve onların ortak davranış yolları tespiti üzerine yoğunlaşmıştır. Daha sonraki iki araştırma hedefi;

- 1) Tespit edilen etkileşim yolları ile İngilizce dil gruplarındaki kullanıcıların gelecek katılımlarını ve sertifika kazanma durumlarının tahmini.
- 2) İngilizce'yi ikincil dil olarak konuşan kullanıcıların MOOC kullanımları için tanımlanabilecek stratejiler için gerekli bulguları ortaya koyma.

Tezdeki çalışmaların sonuçları, İngilizce'yi ikinci dil olarak konuşan katılımcılara ve MOOC'ları yabancı dilde karma kampüs eğitiminde bir materyal olarak kullanabilecek olan eğitmenlere yardımcı olabilecektir.

Şekil 1.2'de, R programlama dili ve Shiny kütüphanesi kullanılarak İngilizce dil

gruplarının davranış ve performans analizlerini gerçekleştirmek için oluşturulan altyapının temel işlem akışı yüzeysel olarak gösterilmiştir.



Şekil 1.2 R programlama dili ve Shiny kütüphanesi ile geliştirilen uygulamanın işlem ve veri akışı

1.3.1 Tezin Akademik Yayın Çalışmaları

Tezdeki ilk çalışmada, MOOC'lardaki katılımcıların başarılarında demografik ve kullanım özelliklerinin önemini anlamak için edX ¹⁴ platformu verileri kullanılmıştır [93]. Çalışmada cinsiyet bilgilerinin ve aktivite kullanımının edX platformu MOOC'undaki katılımcıların performansına olan etkisinin analizine odaklanılmıştır. Tablo 1.1'de yer verilen 14 Mayıs 2014'te oluşturulan HarvardX ve MITx kişi-kurs (2013 akademik yılı) verisi kullanılmıştır ¹⁵. Verilerin büyük boyutu nedeniyle daha hızlı analiz için veri formatı, *virgülle ayrılmış değerler* (CSV) formatından MySQL formatına dönüştürülmüştür. Dönüştürülen veriler, MySQL Workbench programında yazılan SQL (Structured Query Language) sorguları ile analiz edilerek kullanıcıların profillerinin ve platform kullanımının performanslarına olan etkileri incelenmiştir. Ardından SQL sorgularını görselleştirmek için grafikler oluşturulmuş ve oluşturulan grafikler yorumlanarak cinsiyet ve aktivite kullanımının kullanıcı performansı üzerindeki etkisi analiz edilmiştir [93].

¹⁴<https://www.edx.org>

¹⁵<https://doi.org/10.7910/DVN/26147>

Tablo 1.1 edX verisi

Kurum	Kurs Kodu	Kurs Başlığı	Dönem
HarvardX	CB22x	HereosX	İlkbahar (İlkb)-Yaz 2013
HarvardX	CS50x	CS50x	Sonbahar-İlkb 2012-2013
HarvardX	ER22x	JucticeX	İlkb-Yaz 2013
HarvardX	PH207x	HealthStat	Sonbahar (Sonb) 2012
HarvardX	PH278x	HealthEnv	Yaz 2013
MITx	14.73x	Poverty	İlkb 2013
MITx	2.01x	Structures	İlkb, Yaz 2013
MITx	3.091x	SSChem	Sonb 2012, İlkb 2013
MITx	6.002x	Circuits	Sonb 2012, İlkb 2013
MITx	6.00x	CS	Sonb 2012, İlkb 2013
MITx	7.00x	Biology	İlkb 2013
MITx	8.02x	E & M	İlkb 2013
MITx	8.MReV	MechRev	Yaz 2013

Tez çalışmasının ikinci adımı olarak, MOOC'larda Öğrenme Analitiği'nin kullanımı ve performans tahmini ile ilgili mevcut çalışmaların kısa bir incelemesi gerçekleştirilmiştir [94]. Çalışmada, MOOC'larda kullanıcıların performans analizi ve Öğrenme Analitiği ile ilgili son gelişmeler ve çalışmalar hakkında konuyla ilgili araştırmacılara rehberlik etmek amaçlanmıştır. Bu amaçla, MOOC'larda performans tahmininin ve Öğrenme Analitiği'nin kullanımına ilişkin kısa bir inceleme yapılmıştır. Çalışmada, okuyucuların konuyu tanımlarına yardımcı olmak için öncelikle temel kavramlarla ilgili literatür bilgileri açıklanmıştır. Ardından özelliklerin önem düzeyi ve ilişkilerinin daha ayrıntılı anlaşılması için bazı çalışmalar hakkında detaylı bilgiler verilmiştir. Daha sonra, MOOC'larda öğrenci performans tahmininin ve Öğrenme Analitikleri'nin kullanımı hakkındaki bulgular özetlenmiştir. Mevcut çalışmaların incelemesi kapsamında, sekiz çalışma detaylı olarak incelenmiş ve inceleme sonuçları sunulmuştur. Tablo 1.2, incelenen çalışmalara ait veri bilgilerini gösterirken, Tablo 1.3 ve Tablo 1.2'de veri bilgileri verilen çalışmalarda kullanılan yöntemleri ve tahmin hedefi parametrelerini aynı sıra ile ortaya koymaktadır [94].

Tablo 1.2 Detaylı olarak incelenen çalışmalarda kullanılan verilerin bilgileri

Veri Tanımı	Veri Miktarı
2013 Discrete Optimization (Coursera)	37.777 öğrencinin verisi
1) Digital Story Telling , 2) Math. Teach. (Iversity)	106.327 öğrencinin verisi
Vanderbilt Univ, Soft. Archit.	6.953 öğrencinin verisi
Harvard-MITx Person Course 2013	641.138 kayıt
UCD Moodle Verisi	360 kurs ve 2.194 ödev
Princ. Of Macro Econ., Disc. Opt. (Coursera)	22.731 öğrencinin verisi
Melbourne Univ. (Princ. Of Macro Econ.)	862 öğrencinin verisi
Dig. Tech. Of F. (MiriadaX 2013)	5.595 katılımcının verisi

Tablo 1.3 Detaylı olarak incelenen çalışmalardaki yöntem ve tahmin hedefi bilgileri

Yöntemler	Tahmin Hedefi
Stepwise Regresyon	MOOC Performans (Perf) Tahmini
10-kat çapraz doğrulama	MOOC Perf Tahmini
Pearson's R Korelasyon	Kurs Bırakma ve Perf Tahmini
Farklı çekirdeklerle SVM	MOOC Final Notu Tahmini
Korelasyon Analizi	Üç MOOC Perf Analiz Hipotezi Testi
Tekrarlama yaklaşımı, Durum Geçiş Diy.	Perf Analizi Tekniği Geliştirme
Cronbach's alpha katsayısı, Korelasyon	MOOC Final Notu Tahmini
Bağımsız t test, Korelasyon, Norm., Lin. Reg.	MOOC Perf ve Paylaşım İlişkisi

Tablo 1.2'de, incelenen çalışmalardaki veri odaklı farklılıklar olan platform farkı, kurs farkı ve farklı veri miktarı yer almaktadır. Tablo 1.3'den de katılımcıların kurs performans analizi ile ilgili olarak kurs sonu notu tahmini, performans (sertifika alma durumu) tahmini ve kursu bırakma tahmini gibi birkaç tahmin türünün gerçekleştirildiği anlaşılmaktadır. Tablo 1.3, MOOC'larda kurs performans analizlerinde, incelenen çalışmalarda korelasyon ve regresyon yöntemlerinin, geniş kullanımına sahip olduğunu da ortaya koymaktadır.

Tezdeki sonraki çalışmalarda, FutureLearn platformu verileri kullanılmıştır. FutureLearn platformundan ilk olarak "Understanding Language 4: Learning and Teaching" MOOC'undan faydalanılmıştır [95], [23], [96]. "Understanding Language 4: Learning and Teaching" kursunda, kayıtlı olan 25.597 katılımcı içinden sadece % 12,9 (3.305) katılımcı, kayıt sürecinde nerede yaşadıklarına dair bilgiler vermişlerdir. Daha doğru grup ayrımı için, katılımcıların ülke ve dil bilgileri ile kendilerini tanıttıkları kursun birinci haftasının beşinci adımına ait tartışma forumundaki yorumları incelenmiştir. Bununla birlikte, ikincil dil olarak İngilizce konuşan katılımcıları, yorumlardan elle tespit etmek zaman alıcı ve birden fazla

MOOC'daki çok sayıda katılımcıya uygulamak için güvenilir olmayan bir yöntemdir. Katılımcıların tüm yorumlarının inceleme sonuçlarına göre ülke bilgisini giren 3.305 katılımcıdan 710'u tartışma forumlarında birincil dilleri hakkında yorum yapmıştır. Katılımcıların yorumları, İngilizce dil grubu tespit etme amaçlı düzenli ifade kalıplarını oluşturmak için kullanılmıştır [97]. Ardından, düzenli ifadelerle tespit edilen İngilizce dil grup bilgileri kullanılarak katılımcıların önceki aşamadaki İngilizce dil grupları güncellenmiştir. Tez kapsamında gerçekleştirilen sonraki çalışmalarda, İngilizce dil grubu perspektifinden katılımcıların davranışları da analiz edilmiştir [95], [23], [96], [98].

Tez kapsamındaki bir diğer çalışma, kursta İngilizce'yi ikincil dil olarak konuşan katılımcıları belirlemeye yönelik denenen yaklaşımlar ve karşılaşılan zorlukların belirlenmesidir [23]. Çalışmada belirtilen yöntemle göre kurs katılımcıları, ülkelerindeki İngilizce dilinin yasal statüsüne göre üç gruba ayrılmıştır (1- İngilizce'yi ana ve resmi dil olarak konuşanlar, 2- İngilizce'yi yalnızca resmi dil olarak konuşanlar, 3- İngilizce'yi ikincil dil olarak konuşanlar) [23]. Ardından gerekli olan durumlarda, katılımcıların kurs tartışma forumlarında yaptıkları yorumlara göre İngilizce dil grupları güncellenmiştir.

1.4 Hipotezler

Tezin hipotezleri aşağıdaki gibi listelenebilir:

- MOOC'lardaki tartışma forumlarındaki yorumlarda katılımcıların kendi ülke, şehir, dil, millet bilgileri mevcutsa katılımcıların ait oldukları İngilizce dil grupları düzenli ifade kalıplarına dayalı bir algoritma oluşturularak otomatik olarak tespit edilebilir.
- MOOC kullanıcılarının yorumlarından İngilizce dil gruplarının tespiti için Derin öğrenme yöntemi ile tüm kurslara uygulanabilecek bir İngilizce dil grubu tespit algoritması oluşturulabilir.
- İngilizce dil gruplarının kurs kullanımları arasında İngilizce dil yeteneklerindeki değişime bağlı olarak belirgin farklılıklar olmalıdır.
- Derin Öğrenme modellerinde, veri boyutu arttıkça kurs performans tahmininde diğer algoritmalara göre daha yüksek tahmin başarısına ulaşılabilecektir.
- İngilizce'yi ikincil dil olarak konuşan katılımcıların MOOC forumlarında daha az aktif olmaları beklenir.

- İngilizce'yi ikincil dil olarak konuşan katılımcılar, diğer katılımcılara göre sosyal etkileşim sağlanan alanlarda daha az popüler olmaktadır.
- İngilizce'yi ikincil dil olarak konuşan katılımcıların kurstaki aktiviteleri tamamlama ve devamlılık oranları daha düşük olmaktadır.
- Kurs aktiviteleri ile performans arasındaki ilişkilerin sonuçları yorumlanarak, İngilizce dil grupları kullanıcılarına ve platform sağlayıcılarına kursu başarı ile tamamlayabilmeleri için kullanım stratejileri önerilerilebilir.
- İngilizce dil gruplarının kurs başarılarını belirleyen faktörler, hem kurs içinde hem de İngilizce dil grupları arasında değişkenlik gösterebilir.

Tez kapsamında Edx, Coursera ve FutureLearn gibi popüler MOOC platformlarının verileri incelenmiş olup, araştırma konumuz için içerdiği sosyalleşme özellikleri, İngilizce dili odaklı analiz yapma ve işbirliği imkanlarının yüksek olması sebebi ile en fazla katkı sağlayabileceğini düşündüğümüz FutureLearn verileri ile çalışmalara devam edilmiştir.

FutureLearn verisinin sağladığı avantajlardan bazıları aşağıdaki gibi listelenebilir:

- Sertifika, not bilgisi özellikleri ile performans analizi yapmamıza elverişli olması
- Yeterli miktarda veri içermesi ve verinin işbirliği kapsamında kullanıma açık olması
- Geniş özellik kümesi içermesi (site içi görevler, forum aktiviteleri, kullanıcıların demografik özellikleri, kullanım özellikleri ve zamanları) ile farklı üniversite ve kurslar arasında karşılaştırmalı analizler yapma imkanı sunması

FutureLearn platformundaki MOOC'lar genellikle Birleşik Krallık'tan olan katılımcıların katılımı ile gerçekleşirken, "Understanding Language: Learning and Teaching" kurs serisi MOOC'ları birincil dili İngilizce olmayan ülkelerdeki İngilizce öğrenenlerin ve öğretmenlerin yoğun katılımı ile gerçekleşmiştir. Bu nedenle analizlere bu kurs verileri ile başlanmıştır. Aynı zamanda kurs serisinin İngilizce dili öğrenme ile ilgili olması, buna bağlı olarak da forum mesajlarının diller hakkında olması kurslara birçok uluslararası İngilizce öğretmeni ve öğrencisinin kayıt olması, İngilizce'yi ikincil dil olarak konuşan konuşmacıların tespiti için zengin bir veri kaynağı oluşturmuştur. "Understanding Language: Learning and Teaching" kurs serisi MOOC'ları, Southampton Üniversitesi ile British Council arasındaki işbirlikle 2014-2018 yılları arasında yayınlanmıştır. Kurslara katılımlar analiz edilirken genellikle kursların yayında kaldıkları son tarihlere kadar yapılan aktiviteler dikkate alınmıştır.

Kurs serisinden veri kaynağı olarak ilk aşamada, serinin dördüncü versiyonu olan "Understanding Language 4: Learning and Teaching" kursu verileri kullanılmıştır. "Understanding Language 4: Learning and Teaching" kursu '2016-04-04' ile '2016-05-02' tarih aralığında dört haftalık MOOC olarak planlanmış olup, '2016-05-14' tarihine kadar yayında kalmıştır. Kurs serisinin dördüncü sürümü ile yapılan analizlerde kursun yayında kaldığı son gün olan 2016-05-14'e kadar olan aktivite kayıtları kullanılmıştır. "Understanding Language 4: Learning and Teaching" kursu için 25.000'den fazla katılımcı kayıt yaptırmıştır. Kayıt olan öğrencilerden 22.000'den fazlası için ülke bilgisi bulunmamaktadır. Benzer durum cinsiyet, eğitim durumu, çalışma durumu gibi bilgiler için de geçerlidir.

Veri kaynağı olarak "Understanding Language: Learning and Teaching" kurs serisinin ilk yedi versiyonundan faydalanılmaya başlanması ile birlikte elde edilen büyük boyutlu veri ile Derin Öğrenme yöntemi denenmeye uygun hale gelmiştir.

Veri üzerinde, analizlerden önce veri ön işleme gerçekleştirilmiştir. Veri ön işleme ile gereksiz veri özelliklerinden kurtulmak, özellik seçimi ve özellik çıkarımı gibi gerekli görülen işlemler gerçekleştirilmiştir. Gerçekleştirilen veri ön işlemlere tezde ilgili bölümlerin içeriklerinde yer verilmiştir.

Kurs verilerinden kullanım analizi ve performans tahmini için ihtiyaç duyduğumuz özellikleri oluşturmak için veri madenciliği teknikleri (R programlama diliyle) kullanılmıştır. Büyük Verileri analiz için de Google Cloud platformundan faydalanılmıştır.

R dili, çoğu bilgisayar platformunda serbestçe kullanılabilen, çok kaliteli grafikleri olan gelişmiş bir istatistiksel hesaplama sistemidir. R dilini yüksek lisans ve daha ileri seviyeye kadar kullanılan istatistiksel yöntemler sunan gelişmiş bir sistemdir. Birçok araştırma öğrencisi R dilini araştırma aracı olarak kullanmaktadır [99]. R programlama dili, farklı sistemlerde hesaplama motoru (computing engine) olarak kullanımı diğer dillere göre nispeten kolay olan ve açık kaynaklı sistemler için lisans sorunu olmayan bir dildir [48]. R dilinin bir avantajı, özelleştirilmiş paketler yazarak özel bir uygulama alanına kolayca uyarlanabilmesidir. R dilinin bir başka avantajı, Post-Script cihazı ile kaliteli grafik çıktı yayını elde etmenin basit olmasıdır [59].

Büyük Veri, yapılandırılmış ve yapılandırılmamış olan büyük hacimli veri demektir [100]. Büyük Veri terimi, büyük veri kümelerinin yanında, yüksek miktarda veri ile çok büyük veriler arasındaki farkı tanımlayan başka özelliklere de sahiptir [101]. Büyük Veri'de, veriler sadece bir organizasyon içindeki etkileşimlerin ve faaliyetlerin yan ürünü değil; hükümetler, şirketler ve yüksek öğretim kurumları için kritik bir değer katmanıdır [102]. Büyük Veri Analizi, işe alım ve kabul işlemleri,

finansal planlama, donör izleme ve öğrenci performans izlemesi dahil olmak üzere yüksek öğretim, idari ve öğretim uygulamaları gibi çeşitli alanlara uygulanabilir [100].

2.1 Kullanılan Veriler ve Özellikleri

Analizler için, FutureLearn platformundaki Kayıt Bilgileri (Enrolments), Aktiviteler (Step Activity) ve Yorumlar (Comments) olmak üzere üç temel veriden faydalanılmıştır. Veriler, katılımcıların platformdaki kurs aktivitelerinden otomatik olarak üretilmiştir. Bu verilere ek olarak FutureLearn platformunun bir yıllık kullanıcı takip etkileşimlerini içeren *Followings* verisinden elde edilen kursa ait kayıtlardan da faydalanılmıştır.

Kullanılan dört veri ve tanımları aşağıdaki gibidir;

- Enrolments: *Enrolments* verisi, FutureLearn platformundaki kurslara katılan kullanıcıların demografik özellikleri ile kursa kayıt, tamamlama, ayrılma ve sertifika alma bilgilerini içermektedir.
- Step Activity: *Step Activity* verisi, katılımcının bir kurs eğitim adımı ile ilgili adım sayfasını (öğrenme birimi) açma zamanını, adımı tamamlama durumunu ve tamamlama zamanı (tamamlaması durumunda) bilgilerini içermektedir. Kursta hiçbir adıma başlamayan kullanıcıların *Step Activity* verisinde kaydı yoktur. Kursu tamamlamak için birden fazla adımı (en az % 50) tamamlamak gerekmektedir. Bu sebeple, veride bazı kullanıcılara ait birden fazla (başladıkları adım sayısı kadar) kayıtları yer almaktadır.
- Comments: *Comments* verisi, tartışma forumlarındaki yorumların içeriği, yorumların hangi yoruma cevap olduğu, yorum yapan kişi, yorumun yapıldığı adım, yorum zamanı ve yorumun aldığı beğeni sayısı bilgilerini içermektedir. Kursa kayıtlı kullanıcılardan hiçbir yorum yapmayanlar olabildiği gibi birden fazla yorum yapan kullanıcılar da olabildiği için veride bazı kullanıcılara ait id'ler (author_id) yer almazken bazı kullanıcıların id'leri tekrarlı olarak yer alabilmektedir.
- Followings: *Followings* verisi, FutureLearn platformundaki kullanıcıların takip etme ilişkileri ile takip işlemlerinin kim tarafından ve ne zaman gerçekleştirildiği bilgilerini içermektedir.

Tablo 2.1'de kullanılan verilere ve bu verilerde yer alan özelliklerin isimlerine yer verilmiştir.

Tablo 2.1 Kullanılan FutureLearn platformu verileri ve özellikleri

Enrolments	Step Activity	Comments	Followings
learner_id enrolled_at unenrolled_at role fully_participated_at purchased_statement_at gender Country age_range highest_education_level employment_status employment_area	learner_id Step week_number step_number first_visited_at last_completed_at	id author_id parent_id step week_number step_number text timestamp moderated likes	followed_user_id followed_user_role follower_user_id follower_user_role created_at

2.1.1 Kayıtlar (Enrolments) Verisi ve Özellikleri

Enrolments verisinde yer alan özellikler ve bu özelliklerin açıklamaları aşağıdaki gibidir:

- learner id: Her kullanıcı için oluşturulmuş benzersiz bir id değerini ifade etmektedir.
- role: Kullanıcının sistemdeki rolünü ifade etmektedir. Kullanıcının sahip olabileceği roller; öğrenen (learner), yorumları gözden geçirip düzenleyebilen (moderator), organizasyonel değişiklikleri yapabilen (organization admin), sistem yöneticisi (admin) olarak sıralanabilir.
- gender: Kullanıcının cinsiyetini göstermektedir.
- country: Kurs öncesi anket bilgilerine göre kullanıcının hangi ülkeden olduğu bilgisini göstermektedir. Veride kurs öncesi anketinde ülke bilgisini veren kullanıcıların ISO2 formatında olan ülke kodları yer almaktadır. Ülke bilgisini girmeyen kullanıcılara ait olan "NA" ya da "Unknown" değerleri görmezden gelinerek, bunların dışındaki değerler anlamlı ülke bilgisi olarak kabul edilmiştir.
- age range: Kullanıcının hangi yaş aralığında olduğunu göstermektedir. "<18", "18-25", "26-35", "36-45", "46-55", "55-65" ve ">65" değerlerinden birisini alabilir.
- highest education level: Kullanıcının mezun olduğu en yüksek okul seviyesini göstermektedir.
- employment status: Kullanıcının iş/çalışma durumunu göstermektedir.
- employment area: Kullanıcının mesleğini göstermektedir.

- `enrolled_at`: Kullanıcının kursa kayıt olduğu zamanı göstermektedir.
- `unenrolled_at`: Kullanıcının kursu bırakma zamanını (eğer bıraktı ise) göstermektedir.
- `fully_participated_at`: Kullanıcının kursu tamamladığı zamanı (eğer tamamladı ise) göstermektedir.
- `purchased_statement_at`: Kullanıcının sertifika satın alma zamanını (eğer sertifika satın aldı ise) göstermektedir.

"Understanding Language 4 : Learning and Teaching" kursuna katılan kullanıcılara ait *Enrolments* verisi 25.597 satırdan oluşmaktadır ve satır sayısı kadar kullanıcı verisi içermektedir.

2.1.2 Kurs Adımları Aktivitesi (Step Activity) Verisi ve Özellikleri

Step Activity verisinde yer alan özellikler ve tanımları aşağıdaki gibi listelenebilir:

- `learner_id`: Her kullanıcı için benzersiz bir değerdir. *Enrolments* verisinde karşılığı olan learner id'ye sahip olan kullanıcıyı ifade etmektedir. Her bir kullanıcının başladığı adım sayısı kadar id'si (`learner_id`) tekrar eder.
- `step`: Hafta ve adım değeri ile birlikte aktivitenin ait olduğu adımı ifade eder.
- `first_visited_at`: Aktivite adımı (eğitim birimi) sayfasının ilk görüntülendiği zamanı göstermektedir.
- `last_completed_at`: Kurs adımının (eğitimin) tamamlandığı tarihi göstermektedir.
- `week_number`: Kurs adımının ait olduğu haftayı göstermektedir.
- `step_number`: Aktivitenin ait olduğu ilgili haftadaki, alt kurs adımı numarasıdır.

"Understanding Language 4: Learning and Teaching" kursuna ait *Step Activity* verisi 197.978 satırdan oluşmaktadır.

2.1.3 Yorumlar (Comments) Verisi ve Özellikleri

Comments verisinde yer alan özellikler ve tanımları aşağıdaki gibi listelenebilir:

- id: Yorumun veya cevabın id'si olup, her yorum için benzersiz bir değer almaktadır.
- author_id: Yorumu yazan kişinin id'sidir. *Enrolments* verisindeki learner_id özelliğine karşılık gelmektedir.
- Parent_id: Yorumun cevap olarak yazıldığı yoruma ait id'dir. Yorum, başka bir yoruma cevap olarak yazılmışsa parent id özelliği cevap olarak yazıldığı yorumun id değeridir. Parent id değeri boş ise yorumun ana yorum olduğunu gösterir.
- week_number: Yorumun kursun kaçınıcı haftasında yapıldığını göstermektedir.
- step_number: Ana yorumun veya cevap yorumun ait olduğu haftadaki alt adımın numarasıdır.
- step: Yorumun kursun hangi adımı için yazıldığını göstermektedir.
- text: Ana yorum veya cevap yorum metnine karşılık gelmektedir.
- timestamp: Yorumun yazıldığı zamanı göstermektedir.
- likes: Yorumun diğer kullanıcılar tarafından toplamda kaç beğeni aldığını göstermektedir.

"Understanding Language 4: Learning and Teaching" kursuna ait *Comments* verisi, 50.332 satırdan oluşmaktadır.

2.1.4 Takipler (Followings) Verisi ve Özellikleri

Followings verisi, FutureLearn platformundaki kurslara başlayan kullanıcıların takip etme ve edilmeleri ile ilgili bilgileri içermektedir. *Followings* verisinde yer alan özelliklerin tanımları aşağıdaki gibidir:

- followed_user_id: Takip edilen kullanıcı id'sidir. *Enrolments* verisindeki learner_id özelliğine karşılık gelmektedir.
- followed_user_role: Takip edilen kullanıcının rolüdür (learner, moderator, organisation admin, admin).
- follower_user_id: Takip eden kullanıcının id'sidir. *Enrolments* verisindeki learner_id özelliğine karşılık gelmektedir.

- follower_user_role: Takip eden kullanıcının rolüdür (learner, moderator, organisation admin, admin).
- created_at: Takip etme/edilme zamanını göstermektedir.

Kullanılan 2016 yılı *Followings* verisi 1.298.285 satırdan oluşmaktadır. *Followings* verisinden "Understanding Language 4: Learning and Teaching" kursuna kayıtlı kullanıcıların kursun yayınlandığı süre içerisindeki takip etme/edilme davranışları analiz edilmiştir.

Kullanıcıların İngilizce Dil Gruplarının Tespiti

Bu bölümde, kullanıcıların İngilizce dil grubu tespitinde kullanılan yöntemlere ve elde edilen sonuçlara yer verilmiştir. Kullanıcıların İngilizce dil gruplarının doğru tespit edilmesi, tezin sonraki adımlarında gerçekleştirilen analizlerin, “ikincil dil olarak İngilizce” konuşan katılımcılar için faydalı olabilecek çıkarımların güvenilirliği, MOOC kullanıcıları ve sağlayıcıları için faydalı olabilecek tasarım stratejilerinin doğru belirlenmesine yardımcı olmak için çok önemlidir. Bu sebeple ilk olarak MOOC'larda, “ikincil dil olarak İngilizce” konuşan katılımcıları otomatik ve doğru bir şekilde tespit etmeye odaklanılmıştır.

MOOC'larla ilgili mevcut araştırma çalışmalar ([103], [60]) genellikle, kullanıcının ilk diline veya kursun diline bakmaksızın kursa kayıtlı olan tüm katılımcılara yöneliktir. Ancak, araştırmalar ([20], [24]) MOOC'larda ikinci dil olarak İngilizce konuşan katılımcıların, dil becerilerinden yoksun oldukları için kursları takip ederken ve başkalarıyla etkileşime geçerken bazı zorluklarla karşılaşabileceklerini göstermektedir.

Kullanıcıların İngilizce dil gruplarının tespitinde, veri kaynağı olarak öncelikle '2016-04-04' ve '2016-05-14' tarihleri arasında yayında kalan "Understanding Language 4: Learning and Teaching" MOOC'u kullanılmıştır. Kurs, Dünya üzerinden birçok uluslararası İngilizce öğretmeni ve öğrencisinin katılımı ile gerçekleştiği için analizlerde öncelikli olarak tercih edilmiştir. Kurs 25.598 kullanıcı kayıt olurken, 3.305 tanesinin kurs öncesi anketinde lokasyon (ülke) bilgileri mevcuttur.

FutureLearn MOOC'larındaki katılımcıların İngilizce dil gruplarının tespitinde kurs öncesi anketlerinde yer alan ülke bilgileri kullanılabilir. Bununla birlikte, MOOC'lardaki katılımcılar genellikle kurs öncesi anket sorularını cevaplamamaktadırlar. Bu problemin üstesinden gelmek için çeşitli yaklaşımlar önerilmiştir. Örneğin Uchidiuno ve diğerleri [31], katılımcıların birincil dillerini, İnternet tarayıcılarında sahip oldukları kullanım dili tercihleriyle tespit etmişlerdir. Bununla birlikte, katılımcıların pratik yapmak gibi bir amaç için İnternet tarayıcı dili tercihi olarak İngilizce'yi seçmiş olma ihtimalleri de vardır.

Tez çalışmasının ilk adımı olarak, katılımcıları İngilizce bilgilerine göre gruplandırmak için en fazla kullanıcıya uygulanabilir ve en fazla doğruluğa sahip yöntemler bulunmaya çalışılmıştır. Ardından katılımcıların davranış ve kurs sonu performansları analiz edilmiştir.

Birkaç adımda gerçekleştirilen katılımcıların İngilizce dil gruplarını tespitinden sonra, gruplardaki katılımcıların genel kurs katılımları arasındaki farklılıklar incelenmiş ve ardından iyileştirmeler gerçekleştirilerek İngilizce dil gruplarında gerekli görülen güncellemeler yapılmıştır.

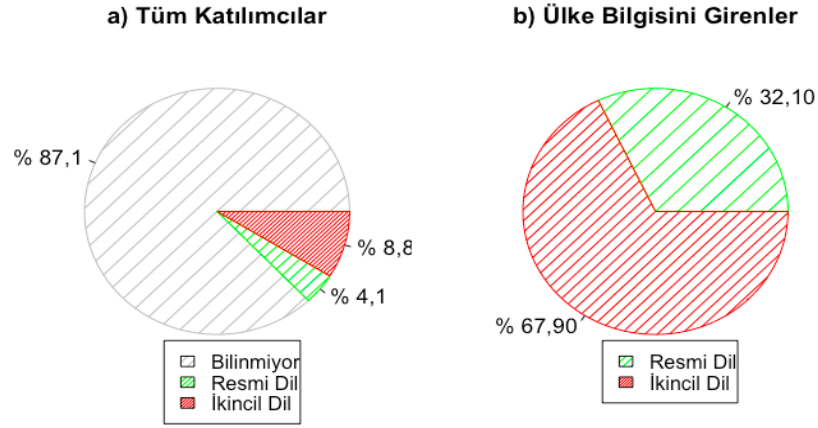
3.1 Kurs Öncesi Anketlerdeki Ülke Bilgilerinden Kullanıcıların İngilizce Dil Gruplarının Tespiti

Katılımcıların İngilizce dil gruplarını tespit etmek için ilk olarak, yalnızca FutureLearn platformundaki kurs öncesi anketlerinden elde edilen katılımcıların konum (ülke) bilgileri kullanılmıştır. Bu bilgiler katılımcıların Kayıt (Enrolments) verisinde yer alan 'country' özelliğinde tutulmaktadır.

Kurs öncesi anketlerden İngilizce dil gruplarının belirlenmesinin ilk aşamasında, katılımcılar kurs öncesi anketlerde verdikleri ülkelerin resmi dillerinin İngilizce olup olmamasına göre gruplandırılmışlardır. Ülkeleri, resmi dil bilgilerine göre gruplandırmak için bir Wikipedia makalesinden¹ faydalanılmıştır. İlk aşama sonucunda katılımcılar, i) Resmi dil olarak İngilizce konuşanlar ii) İkincil dil olarak İngilizce konuşanlar iii) Veride ülke bilgisi yer almayanlar olarak İngilizce dil gruplarına ayrılmışlardır.

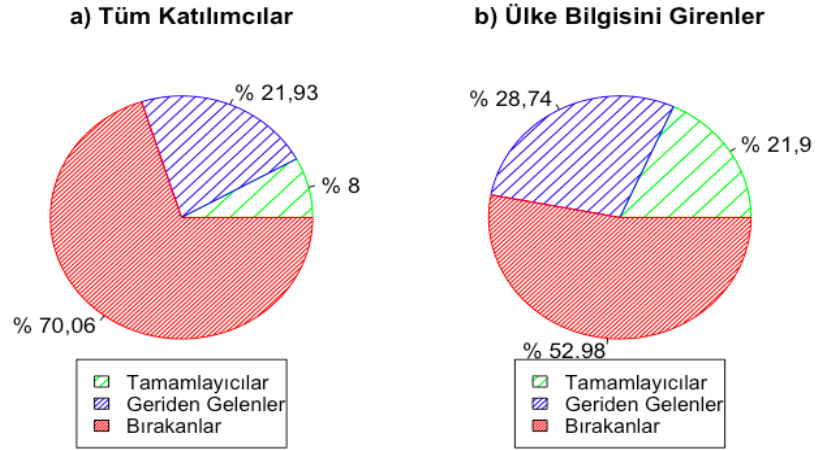
"Understanding Language 4: Learning and Teaching" MOOC'u katılımcılarının ilk aşamadaki gruplandırmalarından sonra oluşturulan İngilizce dil grupları ve katılımcı yüzdelere Şekil 3.1'de yer verilmiştir. Şekil 3.1-a'da yer alan pasta grafiği, kursa katılan tüm katılımcılar dikkate alındığında İngilizce dil gruplarındaki katılımcıların dağılımlarını göstermektedir. Bu dağılımlara göre, katılımcıların sadece % 12,9'u kurs öncesi anketinde ülke bilgisini vermiştir. Şekil 3.1-b'de yer alan pasta grafiği ise, ilk aşamada sadece kurs öncesi ankette ülke bilgisini veren katılımcılar dikkate alındığında İngilizce dil gruplarındaki katılımcıların dağılımını vermektedir. Bu pasta grafiğine göre, kurs öncesi anketinde ülke bilgisini veren katılımcıların, yaklaşık olarak üçte ikisinin İngilizce'nin ikincil dil olan ülkelerdendir.

¹ https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language



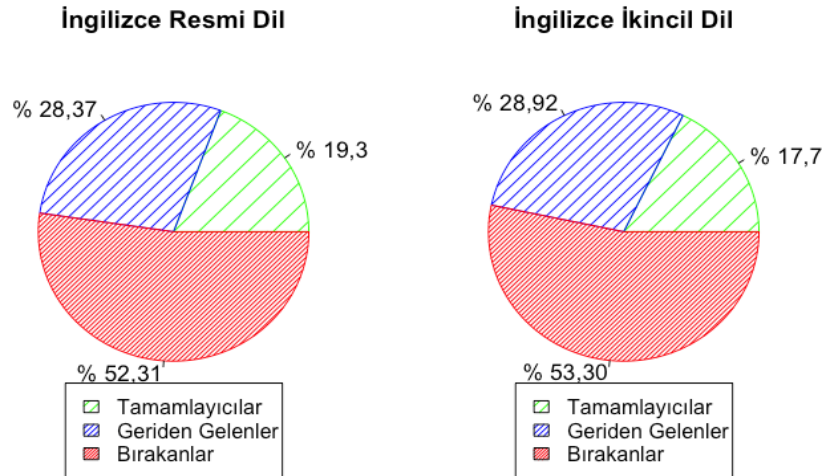
Şekil 3.1 "Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketteki ülke bilgilerine göre İngilizce dil gruplarındaki katılımcıların yüzdeleri (birinci aşama)

Şekil 3.2'de, "Understanding Language 4: Learning and Teaching" MOOC'unda kayıtlı katılımcıların kurs performanslarının aldıkları değerlerin yüzdelerine göre dağılımları gösterilmiştir. Şekil 3.2-a'da yer alan pasta grafiği, tüm katılımcılar dikkate alındığında kurs performansı özelliğinin aldığı değerlerin dağılımını göstermektedir. Buna göre katılımcıların sadece % 8'i 'tamamlayıcı' değerini alırken (% 70,06)'sı 'bırakanlar' değerini almıştır. Şekil 3.2-b'de yer alan pasta grafiği ise, sadece kurs öncesi ankette ülke bilgisini veren katılımcılar dikkate alındığında, ilk aşamada kurs performansı özelliğinin değerlerine göre dağılımı göstermektedir. Bu pasta grafiğinden kurs öncesi anketinde ülke bilgisini veren katılımcıların kurs tamamlama yüzdelerinin ('tamamlayıcı' değerini alma) daha yüksek (% 21,9) ve kurstan ayrılan ('bırakanlar' değerini alan) katılımcı yüzdesinin daha düşük olduğu (% 52,98) anlaşılmaktadır.



Şekil 3.2 "Understanding Language 4: Learning and Teaching" kursu katılımcılarının kurs performansı özelliğinin aldığı değerlerin dağılımlarına göre yüzdeleri (birinci aşama)

Şekil 3.3’de, kurs önceki anket bilgilerine göre İngilizce dil gruplarına ayırmanın birinci aşamasının yeterliliğini ölçmek adına, İngilizce resmi dil ve İngilizce ikincil dil olan katılımcıların kurs performans özelliklerinin aldıkları değerlerin dağılımlarına yer verilmiştir. Şekil 3.3’e göre, ilk aşamada oluşturulan İngilizce dil gruplarının performanslarının aldıkları değerlerin dağılımında gruplar arasında belirgin bir fark görülmemiştir. Bu sebeple İngilizce dili gruplandırma yönteminin ikinci aşamasının denenmesine karar verilmiştir.



Şekil 3.3 "Understanding Language 4: Learning and Teaching" kursunda "resmi dil olarak İngilizce" ve "ikincil dil olarak İngilizce" konuşan katılımcıların kurs performansı özelliğinin aldığı değerlerin dağılımlarına göre yüzdeleri (birinci aşama)

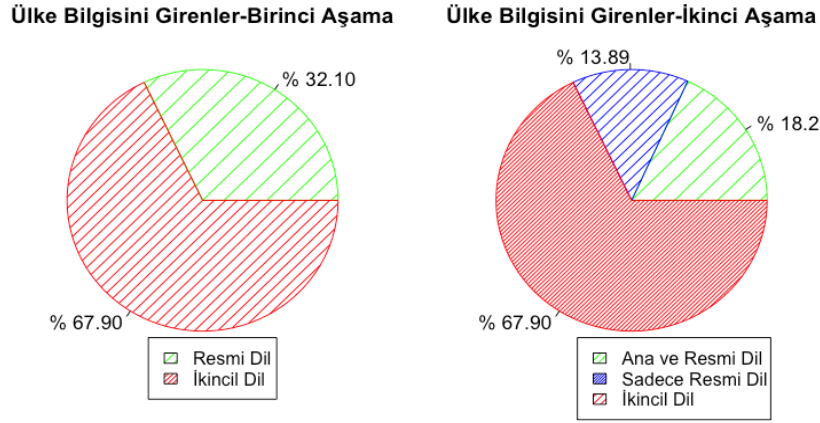
Kurs öncesi anketlerdeki ülke bilgilerine göre İngilizce dil gruplarına ayırmanın ikinci

aşamasında, İngilizce'nin resmi dil olduğu ülkeler aynı Wikipedia makalesinden ¹ faydalanılarak iki alt gruba ayrılmıştır. İkinci aşamadaki gruplandırma işlemini gerçekleştirmek için Tablo 3.1 oluşturulmuştur. Tablo 3.1'de birincil ve resmi dili İngilizce olan ülkeler ile birincil dili İngilizce olmayıp resmi dili İngilizce olan ülkelerin ISO-2 formatında kodlarına yer verilmiştir. Tablo 3.1'de yer verilen ISO-2 formatında kodların karşılık geldiği ülkelerin isimlerine (kurs öncesi anketlerinde yer alan İngilizce isimlerine), tezin ekinde yer verilmiştir. Kurs öncesi anketlerdeki ülke bilgilerine göre İngilizce dil gruplarına ayırmanın ikinci aşamasında, gruplandırma için kurs öncesi anketdeki ülke bilgilerinin *Enrolments* verisinde yer alan ISO-2 formatındaki kodları ile Tablo 3.1'deki ülkelerin ISO-2 formatındaki kodları karşılaştırılmaktadır. Ülke, bu iki İngilizce dil grubundan birine ait ise bu kullanıcı gruplarından birine, değil ise üçüncü İngilizce dil grubu olan "ikincil dil olarak İngilizce" konuşanlar dil grubuna eklenmektedir.

Tablo 3.1 İngilizce'nin birincil dil veya İngilizce'nin resmi dil olduğu ülkelerin ISO-2 formatında kodları

Birincil dili ve resmi dili İngilizce olanlar									
US	GB	AU	SL	CA	GH	NZ	NG	IE	JE
TT	TC	GG	BB	JM	AG	BS	BZ	CK	DM
GD	GY	LR	KN	LC	VC	AI	BM	VG	KY
FK	IM	NF	PN	VI	IO	GG	MS	SH	
Yalnızca resmi dili İngilizce olanlar									
MY	KE	PK	PH	CM	BW	LK	BD	TZ	SG
IN	NA	ZA	RW	FJ	BN	MV	SD	IL	ET
ZW	SS	UG	ZM	PG	MT	LS	GM	CW	HK
PR	MU	BI	FM	KI	MH	NR	NP	NU	PW
WS	SC	SB	SZ	TO	TV	VU	BH	CY	ER
JO	KW	MV	MM	QA	AE	AS	GI	GU	MP
SX	CX	CC	TK						

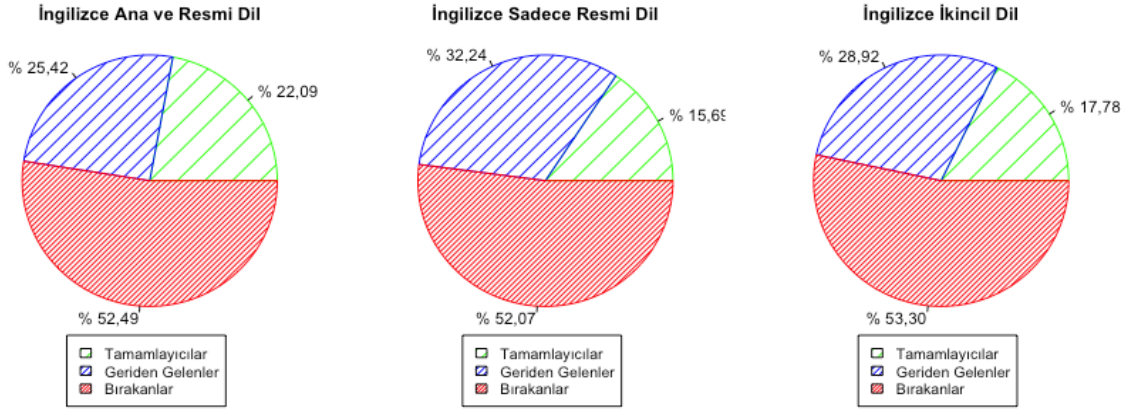
Şekil 3.4'de kurs öncesi anketlerdeki ülke bilgilerine göre İngilizce dil gruplarına ayırmanın birinci ve ikinci aşamalarında İngilizce dil gruplarındaki katılımcıların dağılımlarına yer verilmiştir.



Şekil 3.4 "Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketteki ülke bilgilerine göre İngilizce dil grupları yüzdeleri (birinci ve ikinci aşama)

Şekil 3.4'e göre İngilizce dil gruplarına ayırmanın ikinci aşamasında, "Understanding Language 4: Learning and Teaching" MOOC'undaki katılımcıların % 18,2'si İngilizce'nin birincil ve resmi dil olduğu bir ülkeden, % 13,89'u İngilizce'nin yalnızca birincil dil olduğu bir ülkeden ve % 67,90'ı da "ikincil dil olarak İngilizce" konuşulan bir ülkedendir.

Şekil 3.5, "Understanding Language 4: Learning and Teaching" MOOC'unda, kurs öncesi anketindeki ülke bilgilerine göre İngilizce dil gruplarına ayırmanın ikinci aşaması sonucunda elde edilen İngilizce dil gruplarında kurs performansı özelliklerinin dağılımlarını göstermektedir. Şekil 3.5'e göre, birinci aşama sonuçlarından farklı olarak, İngilizce dil gruplarındaki kurs performansı özelliklerinin değerleri arasındaki fark daha belirgin hale gelmiştir (özellikle 'tamamlayıcı' değerini alan kullanıcı yüzdelerinde).



Şekil 3.5 "Understanding Language 4: Learning and Teaching" kursunda kurs öncesi anketlerdeki ülke bilgileri ile İngilizce dil gruplarının kurs performanslarına göre yüzdeleri (ikinci aşama)

3.2 Kullanıcı Yorumlarının 'Manuel' Olarak Ön İncelemesi

FutureLearn platformu, diğer MOOC platformları gibi haftalar ve bu haftaların alt adımlarından oluşmaktadır. Örneğin, birinci haftanın ilk adımları olarak 1.1, 1.2, 1.3 gibi. "Understanding Language: Learning and Teaching" MOOC serisinde, birinci haftanın beşinci adımına (adım 1.5) ait tartışma forumunda, katılımcıların nerede yaşadıkları, ilk dillerinin ne olduğu ve İngilizce'de ne derece akıcı oldukları hakkında bilgi vermeleri istenmiştir. Tablo 3.2 ve Tablo 3.3'de, İngilizce dil gruplarına ayırmada kullanıcı yorumlarının elle (manuel) olarak ön incelenmesi aşamasında *Enrolments* ve *Comments* verilerinden kullanılan özelliklere ve açıklamalarına yer verilmiştir.

Tablo 3.2 Yorumlar (Comments) verisinden kullanılan özellikler ve anlamları

Özellik Adı	Anlamı
author_id	kullanıcının kimliği
id	yorumun kimliği
text	yorumun kendisi
step	yorumun gerçekleştirildiği adım

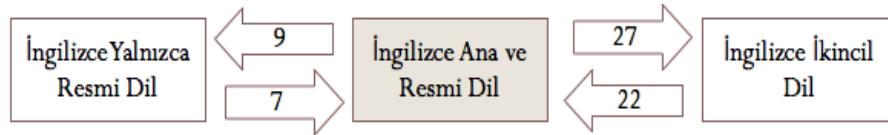
Tablo 3.3 Kayıtlar (Enrolments) verisinden kullanılan özellikler ve anlamları

Özellik Adı	Anlamı
learner_id	kullanıcının kimliği
country	kullanıcının iki haneli ülke kodu

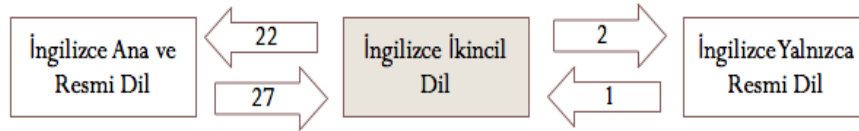
Tez kapsamında gerçekleştirilen ön çalışmalardan birinin sonuçları [23], bazı katılımcıların ilk dillerinin yaşadıkları ülkenin resmi dillerinden farklı olabildiğini

ortaya koymuştur. Çalışmada [23], elle (manuel) yapılan inceleme sonuçlarına göre hem "Understanding Language 4: Learning and Teaching" kursunun birinci haftasının beşinci adımına ait tartışma forumunda ilk dilleri hakkında bilgi veren hem de kurs öncesi anketinde ülke bilgilerini veren 694 katılımcıdan 68'inin yanlış İngilizce dil grubunda gruplandırıldığı tespit edilmiştir (694 katılımcının yaklaşık % 10'u). Bazı katılımcılar da, kursun birinci haftasının beşinci adımında yorum yapmış olsalar da, kurs öncesi anketinde ülkeleri hakkında hiçbir bilgi vermemişlerdir. Şekil 3.6, Şekil 3.7 ve Şekil 3.8 İngilizce dil gruplarının birinci haftasının beşinci kurs adımının tartışma forumu yorumlarının gözlemlenmesinden edinilen bilgilere göre güncellenirken gruplar arasında yer değiştiren katılımcı sayılarını göstermektedir.

Şekil 3.6 ve Şekil 3.7'den, "Understanding Language 4: Learning and Teaching" kursunda birincil ve "ikincil dil olarak İngilizce" konuşan katılımcılardan İngilizce dil grubu değişen sayısının kayda değer olduğu anlaşılmaktadır.

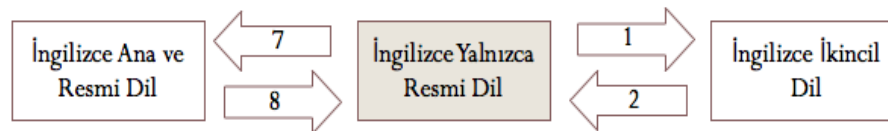


Şekil 3.6 İngilizce birincil ve resmi dili olan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti



Şekil 3.7 İngilizce ikincil dili olan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti

Şekil 3.8, "Understanding Language 4: Learning and Teaching" kursunda birinci haftanın beşinci adımındaki kurs forumunun yorum incelemelerine göre, en az grup değişikliğinin İngilizce'nin resmi olup birincil dili olmadığı ülkelerden gelen katılımcıların grubunda olduğunu göstermektedir.



Şekil 3.8 İngilizce resmi dili olup birincil dili olmayan kullanıcıların grubunda kursun birinci haftasının beşinci adımındaki yorumlara göre grup güncelleme özeti

"Understanding Language 4: Learning and Teaching" kursunda, birinci haftanın beşinci adımındaki yorumlara göre, yanlış gruplandırılan katılımcıların İngilizce dil gruplarının güncellenmesinden sonra, "birincil ve resmi dil olarak İngilizce" konuşan

katılımcılar grubunda 603, “sadece resmi dil olarak İngilizce” konuşan katılımcılar grubunda 350, son olarak da “ikincil dil olarak İngilizce” konuşan katılımcılar grubunda da 2.349 katılımcı yer almıştır.

Kursun birinci haftasının beşinci adımındaki forum yorumlarının elle yapılan inceleme sonuçlarına göre İngilizce dil grupları güncellenen katılımcıların örnekleri;

- Birleşik Krallık'ta yaşayan ve veride ülke isminin ISO-2 formatındaki kodu olan "GB" olan bir katılımcı, ilk dilinin İngilizce'den farklı olduğunu belirtmiştir. Katılımcı, ayrıca çeşitli Avrupa dillerine çalıştığını ve 'İngiliz Dili ve Edebiyatı' bölümüne devam ettiğini de söylemiştir.
- Güney Avustralya'dan katılan bir katılımcı, ilk dilinin İngilizce olmadığını belirtmiştir. Katılımcı, her gün İngilizce kullandığını ve ikinci dil olarak İngilizce konuşanlarla iletişim kurmasının kolay olduğunu, ancak birincil dil olarak İngilizce konuşanlarla iletişim kurmanın onun için daha karmaşık olabileceğini söylemiştir.
- Birleşik Krallık'ta bulunan bir başka katılımcı, iki dilli (Ukraynaca ve Rusça) olarak yetiştirildiğini ve İngilizce öğretmeni olduğunu söylemiştir.

Katılımcı yorumlarının gözlemlenmesi, İngilizce dil gruplandırma yönteminin iyileştirilebileceğini göstermesi bakımından önemli olmakla birlikte, zaman alıcı ve birden fazla MOOC'da çok sayıda katılımcıya uygulamak için güvenilir olmayan bir yöntemdir [23]. Bu sebeple bir sonraki adımda, MOOC'daki katılımcıların İngilizce dil gruplarını daha doğru bir şekilde tespit etmek için düzenli ifadelerle dayalı bir yöntem kullanılmasına karar verilmiştir.

3.3 Oluşturulan Düzenli İfadelerle Dayalı Yöntemle İngilizce Dil Gruplarının Güncellenmesi

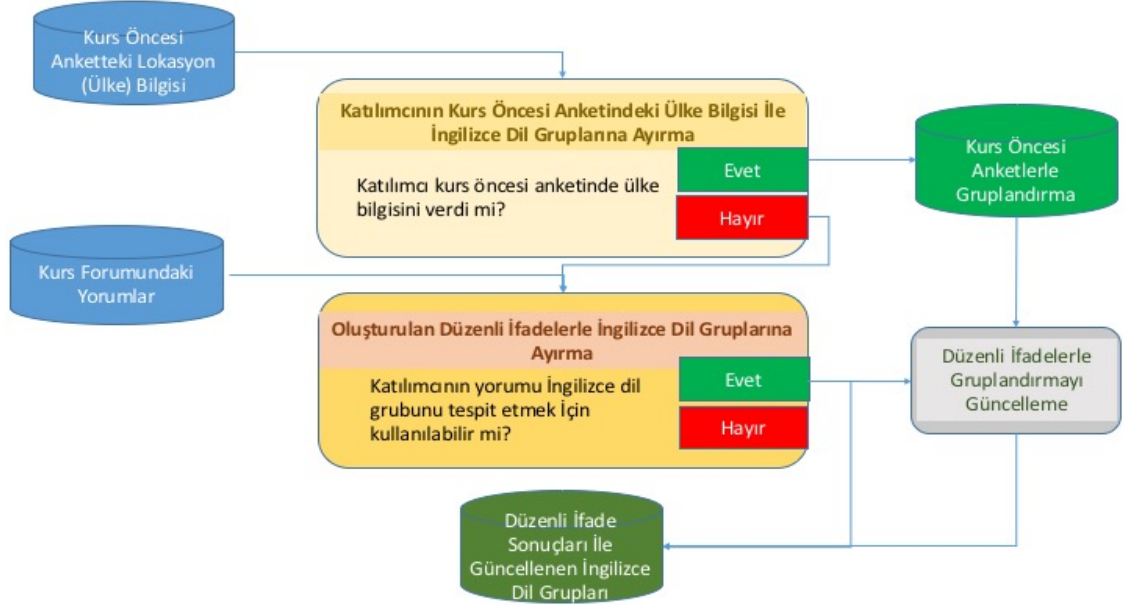
Düzenli ifadeler, güçlü, esnek ve verimli metin işleme yolu sunmakta ve genel desen notasyonu gibi ifadelerle metinleri tanımlamamızı ve ayrıştırmamızı (dil bilgisi bakımından) sağlamaktadırlar [104]. Düzenli ifadeler, bir katarın kısa yoldan belirlenmesini sağlayarak, metin içerisinde belli bir bilginin aranmasına imkan vermektedirler. Düzenli ifadeler, hemen hemen bütün platformlarda aynı söz dizimi kurallarına sahiptirler. Bir karakterin/kelimenin, bir cümlenin veya yazının herhangi bir bölümünde var olup olmadığını kısa yollardan belirlenmesini sağlamaktadırlar.

Literatürde, MOOC'ların içeriğini düzenli ifadeler kullanarak analiz eden çalışmalar vardır. Örneğin, Acosta ve Otero [105], MOOC'larda açık bir soruya katılımcıların

verdikleri cevapları otomatik olarak değerlendirmek için düzenli ifadelerden yararlanmışlardır. Shukla ve Kakkar [106] video içeriğinin değerlendirilmesinde yardımcı olabilecek anahtar kelimeleri çıkarmak amacı ile bir MOOC'un video transkriptindeki isim parçalarını tespit etmede düzenli ifadeleri kullanmışlardır. An ve diğerleri [107], MOOC'larda dış ve iç kaynakları birbirine bağlamak için hashtag içeren düzenli ifade kullanmışlardır. Bu çalışmalar, MOOC'larda düzenli ifadelerin yorum metinleri ve transkript metinlerinden hashtag'leri, anahtar kelimeleri veya belirli kelimeleri tespit etmek için nasıl kullanıldığını göstermektedir.

Tezde katılımcıların kurs forumlarındaki birincil dil, ülke/şehir veya milletleri ile bağlantılı kelime kümeleri ile İngilizce dil gruplarını tespit etmek için düzenli ifadeler kullanılmıştır.

İngilizce dil gruplarının tespiti için düzenli ifade kalıplarını oluşturmanın ilk adımı olarak, katılımcıların ilk dilleri hakkında konuşmak için hangi kelimeleri ve cümleleri kullandıkları tespit edilmiştir. Düzenli ifadeler, bu bilgilerin sık geçtiği cümle yapıları tespit edilerek bu cümleleri eşleyebilecek şekilde oluşturulmuşlardır. Oluşturulan düzenli ifade kalıpları ile eşleşen cümleler üzerinde dil, ülke, şehir ya da millet bilgilerine karşılık gelen yerlerdeki bilgilerin bu tür bilgilerden biri olup olmadığı ülke, şehir, dil ve milliyet bilgilerini içeren listeler üzerinden kontrol edilmekte ve istenen bilgileri içermeyen eşleşmeler elenmektedir. Kullanıcıların tüm yorumları dikkate alınarak düzenli ifadelerle yorumları eşleşen kullanıcı sayısı arttırılmıştır. Bu adımda, benzerlik metriklerinden de yararlanılmıştır. Şekil 3.9, anket bilgilerini ve tartışma forumlarındaki yorumları kullanarak bir katılımcının hangi İngilizce dil grubuna ait olduğunu tespit etme sürecimizi özetleyen sistem modelini göstermektedir.



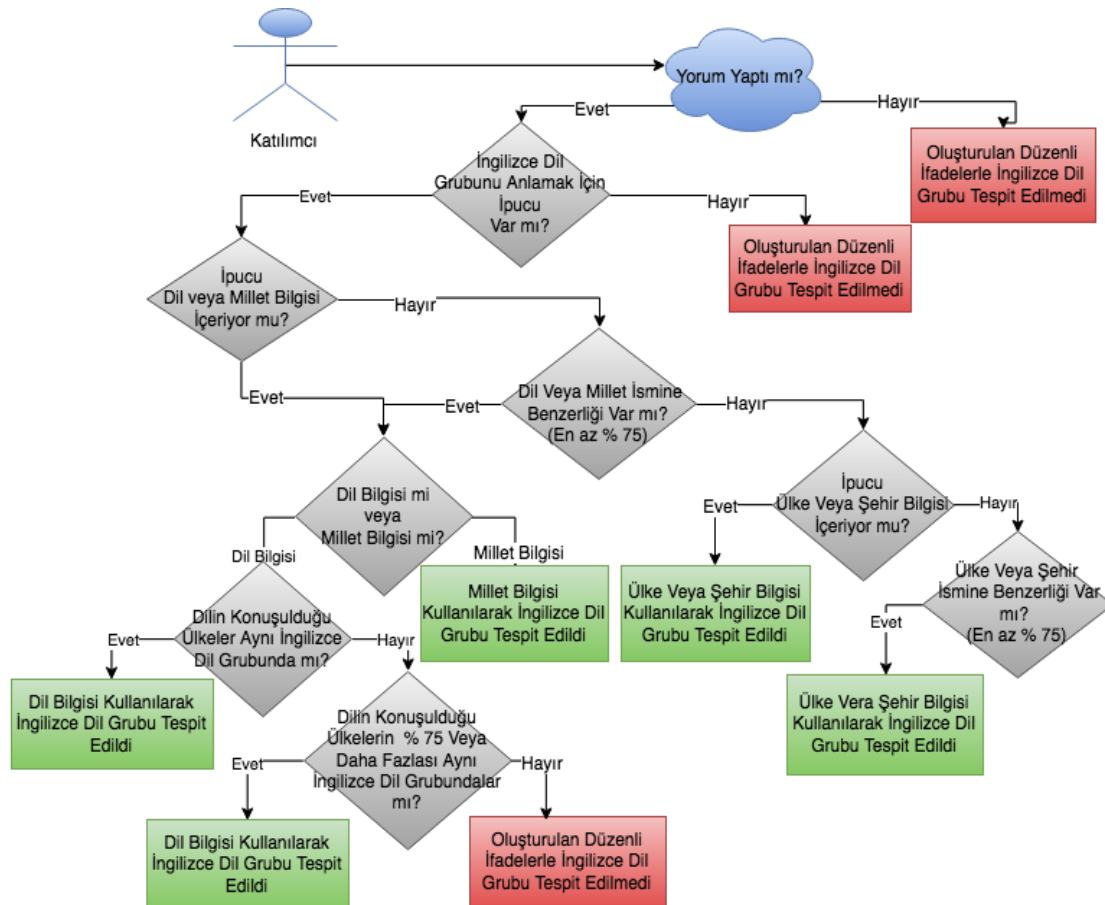
Şekil 3.9 Düzenli ifadeleri kullanarak kullanıcıların İngilizce dil gruplarına ayrılımları için önerdiğimiz model

Düzenli ifadelerin uygulanması sonrası İngilizce dil grupları ve içinde yer alan katılımcıların özellikleri;

1. **Resmi ve birincil dil olarak İngilizce:** Kurs öncesi anketinde veya kurs forumunda İngilizce'nin resmi ve birincil dili olduğunu belirten katılımcılardır. Örneğin Birleşik Krallık'tan olduğunu belirten katılımcılar. Aynı zamanda; kurs forumlarındaki yorumlarda verdikleri şehir, millet veya dil bilgilerinden bu bilginin çıkarılabildiği katılımcılar.
2. **Sadece resmi dil olarak İngilizce:** Kurs öncesi ankette veya forum tartışmalarında İngilizce'nin resmi dillerinden biri olup, birincil dil olmadığı bir ülkeden olduğunu belirten katılımcılar. Örneğin Hindistan gibi. Buna ek olarak şehir, millet veya dil bilgilerinden bu bilginin çıkarılabildiği katılımcılar.
3. **İkincil dil olarak İngilizce:** Kurs öncesi ankette veya tartışma forumunda İngilizce'nin ikincil dil olduğu bir ülkeden olduğunu belirten katılımcılar. Örneğin Türkiye gibi. Aynı zamanda İngilizce'nin resmi bir dili olmadığını belirten veya verdikleri şehir, millet veya dil bilgilerinden bu bilginin çıkarılabildiği katılımcılar.

Şekil 3.10, katılımcıların yorumlarından İngilizce dil gruplarını düzenli ifade ve benzerlik metriğini kullanarak tespit etme algoritmamızın akış şemasını göstermektedir.

Şekil 3.10'da gösterildiği gibi, düzenli ifadelerle İngilizce dil grubu tespiti algoritmamızda, tüm özel koşullar göz önünde bulundurulmuştur. Örneğin, algoritmamız İngilizce'den başka bir dil tespit ettiğinde, o dili konuşanın hangi İngilizce dil grubuna ait olduğunu dilin konuşulduğu ülkeler üzerinden tespit etmeye çalışmaktadır.



Şekil 3.10 Düzenli ifadeler ile İngilizce dil grubu tespiti modelinin akış şeması

Lokasyon (ülke/şehir), dil veya millet bilgileri içeren cümle türlerini tespit etmek için 22 farklı düzenli ifade kalıbı oluşturulmuştur. Oluşturulan düzenli ifade kalıplarına tezin ekinde yer verilmiştir. Düzenli ifade kalıplarını şekillendirmek için tartışma forumlarından seçilen bazı cümle örnekleri:

- I was **born** in **Turkey**.
- I am a **native speaker** of **Spanish**.
- My **mother's tongue** is **Albanian**.
- I live in the UK but I **am** actually **French**.

- I **am** a typical **Japanese**.
- I only **speak English**.
- My **first tongue** is **English**.
- I **grew up** in **Australia**.

Oluşturulan düzenli ifade kalıplarından her biri, birden fazla örnek cümle türü ile eşleşebilmektedir. Örneğin, aşağıdaki düzenli ifade;

*"I was | were brought up | born | grow up | grew"
" | grown | grown up country | city"*

"I was born in Turkey" ve "I grew up in Paris" cümleleri ile eşleşmektedir.

Algoritma, tartışma forumlarındaki cümleleri işlerken düzenli ifade kalıpları ile eşleştirmekte ve ülke, şehir, millet veya dil ile ilgili bilgileri tespit etmektedir. Ardından, bu bilgileri katılımcıların kurs öncesi anketindeki ülke bilgilerinden elde edilen İngilizce dil grubu bilgisi ile karşılaştırmaktadır. Yorumlardan elde edilen İngilizce dili grup bilgileri, kurs öncesi anketindeki ülke bilgilerinden elde edilen ile aynı değilse, katılımcıların İngilizce dil grup bilgileri tartışma forumlarındaki yorumlardan elde edilen İngilizce dil grubu bilgilerine göre güncellenmektedir.

İngilizce dil gruplarına ayırmanın hassasiyetini arttırmak için, algoritma en az % 75 kelime benzerliği kabulü ölçütüyle oluşturulmuştur.

Yorumlardan Dil Bilgisinin Tespiti: İngilizce dil grubu bilgisinin yorumlardan tespitinde hedef bazı cümle taslakları aşağıdaki gibidir:

- my first tongue is **language** (dil benim birincil dilimdir)
- my mother language is **language** (benim birincil dilim dil dir)
- my native speaker is **language** (dil benim yerli konuşma dilimdir)
- I speak **language** (Ben dil konuşurum)
- which is my mother tongue **language** (ayrıca benim birincil dilim)

Kullanıcının dil bilgisini tespit ederken Dünya’da konuşulan dillerin listesi ² referans alınmıştır. Ayrıca, kullanıcının bir veya iki harfi yanlış yazması ihtimali dikkate alınarak R programlama dilindeki "Record Linkage" kütüphanesinde yer alan *Levenshtein Sim* benzerlik fonksiyonu yardımı ile benzerlik eşik oranı olan % 75 ve üzeri olan benzerlik değerlerine sahip eşleşmeler doğru olarak kabul edilmiştir.

Yorumlara uygulanan düzenli ifade kalıplarının eşleştikleri olumlu cümlelerde, “language” yazılan yere “English” yazan kullanıcılar, “birincil ve resmi dil olarak İngilizce” grubuna eklenmişlerdir.

Kullanıcıların dilini tespit etmek için yazılan düzenli ifade kalıplarından birisi (I3), Şekil 3.11’de yer almaktadır. Örneğin, “English language is my first tongue” cümlesini içeren bir yorum, Şekil 3.11’de gösterilen kalıpla eşleşmektedir.

```
\b(?:our|ma?[yi])\s+(?:(:first|1st|mother(?:\'s)?|native|official|main|national|primary|best)\s*)*[\s-]*
(?:t[aou]{1,2}n[gq](?:[ue]{1,2}s?)?|l[ae]nguage|speaker)
(?:(:\s+(?:l[ae]nguage\s+)?(?:(:it|which|was|what)\s+)?
(?:is|was|are)\s+(?:mainly\s+)?(?:the\s+)?|,\s+)(\w{4,}))
```

Şekil 3.11 İngilizce dil bilgisini çıkarmak için yazılan düzenli ifade kalıplarının bir örneği (I3)

Kullanıcının yorumda yazdığı dilin İngilizce dışında bir dil olması durumunda kullanıcının ait olduğu İngilizce dil grubunu belirlemek için Tablo 3.4, Tablo 3.5 ve Tablo 3.6 hazırlanmıştır². Tablo 3.4, Tablo 3.5 ve Tablo 3.6’dan dilin konuşulduğu ülke veya ülkelerin ait oldukları İngilizce dil gruplarının bilgisi (en az % 75’i aynı İngilizce dil grubunda ise) elde edilerek İngilizce dil grupları belirlenmiştir. Tabloları oluşturmak için aşağıdaki referans listeleri kullanılmıştır:

1. İngilizce’nin resmi dil olduğu ülkelerin listesi³.
2. Ülkelerin ve onların resmi ve azınlık dillerinin listesi⁴.
3. Ülkelerin en büyük şehirlerinin listesi⁵ (Bazı katılımcılar ülke isimlerini yazmayıp bulundukları şehri yazdıkları için bu listeye ihtiyaç duyulmuştur. Bu liste ile şehrin ait olduğu ülke bulunmakta, ülke bilgisinden de İngilizce dil grubu tespit edilmektedir).

²https://en.wikipedia.org/wiki/List_of_official_languages_by_country_and_territory

³https://en.wikipedia.org/wiki/List_of_territorial_entities_where_English_is_an_official_language

⁴https://en.wikipedia.org/wiki/List_of_official_languages_by_country_and_territory

⁵https://en.wikipedia.org/wiki/List_of_largest_cities

4. Bazı katılımcılar yorumlarında ülkeleri yerine milletlerini ifade etmişlerdir. Milletler, milletlerin ait oldukları ülkeler ve ülkelerde konuşulan dillerinin listesi ⁶. Bu liste, milletlerin ait oldukları ülkeleri tespit etmek için kullanılmıştır. Ülke bilgisinden de İngilizce dil grubu tespit edilmektedir.
5. Ülkeler ve ISO-2 formatındaki ülke kodlarının listesi ⁷.

Kullanıcı yorumlarından elde edilen dil, eğer sadece bir ülkede konuşuluyorsa; o dili konuşan katılımcılar, dilin konuşulduğu ülkenin ait olduğu İngilizce dil grubuna eklenmektedirler.

Kullanıcının yorumlarında belirttiği dil, eğer birden fazla ülkede konuşuluyorsa ve dilin konuşulduğu ülkeler de aynı İngilizce dil grubunda ise, dilin konuşulduğu ülkelerin hangi İngilizce dil grubuna ait olduklarına bakılarak, kullanıcının İngilizce dil grubu belirlenir.

Kullanıcı yorumlardan elde edilen dilin konuşulduğu ülkelerin hepsi aynı İngilizce dil grubunda değil ise dilin konuşulduğu ülkelerin ait oldukları İngilizce dil gruplarının yüzdelik dilimlerine bakılır. Örneğin, Greek (Yunanca) dilinin konuşulduğu dört ülkeden üçü, “ikincil dil olarak İngilizce” konuşan katılımcıların grubuna ait iken, bir tanesi de “resmi dili ve birincil dili İngilizce” konuşan katılımcıların grubuna ait ise katılımcı % 75 eşik değeri ile Greek dilinin konuşulduğu ülkelerin çoğunun İngilizce dil grubu olan “ikincil dil olarak İngilizce” grubuna eklenir.

Tablo 3.4, konuşuldukları ülkelerde % 75 veya daha fazla yüzde ile İngilizce’nin birincil ve resmi dil olduğu İngilizce dışındaki dilleri içermektedir. Örneğin, bir katılımcı ilk dilinin Yoruba olduğunu belirtmişse, “birincil ve resmi dil olarak İngilizce” konuşan katılımcı olarak gruplandırılmaktadır.

Tablo 3.4 Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce’nin birincil ve resmi dil olduğu İngilizce dışındaki diller

Konuşulan dil	İngilizce dil grubu yüzdesi
Irish	% 100
Yoruba	% 100
Igbo	% 100
Manx	% 100

Tablo 3.5, İngilizce’nin en az % 75 ile, konuşuldukları ülkelerde resmi bir dil olduğu ancak birincil dil olmadığı İngilizce dışındaki dillerin listesini göstermektedir.

⁶<http://www.vocabulary.cl/Basic/Nationalities.htm>

⁷<https://github.com/luke/ISO-3166-Countries-with-Regional-Codes/blob/master/all/all.csv>

İngilizce'nin resmi bir dil olduğu ancak baskın olarak konuşulan yerel dil olmadığı bazı ülkeler vardır. Örneğin, insanların MOOC'lara büyük önem verdiği Hindistan'da, İngilizce de dahil olmak üzere birkaç resmi dil vardır. Hindistan'daki bazı bölgelerde insanlar çok iyi derecede İngilizce bilseler de, bazı yerel hükümetler İngilizce dilinde eğitimi desteklemediği için, İngilizce daha az akıcı olarak konuşulabilmektedir [108]. Bu nedenle, Hindu gibi diller, İngilizce'nin resmi bir dil olduğu ancak birincil dil olmadığı İngilizce dil grubunda yer almaktadırlar.

Tablo 3.5 Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce'nin resmi dil olup birincil dil olmadığı İngilizce dışındaki diller

Konuşulan dil	İngilizce dil grubu yüzdesi
Hindu	% 100
Urdu	% 100
Tamil	% 100
Bangla	% 100
Sindhi	% 100
Afrikaans	% 100
Tagalog	% 100
Shona	% 100
Telugu	% 100
Marathi	% 100
Zulu	% 100
Burmese	% 100

İspanyolca 20'den fazla ülkede resmi bir dildir⁸. İngilizce, İspanyolca'nın konuşulduğu ülkelerin % 87'sinde (Tablo 3.6) ikincil dildir. Bu nedenle, eğer bir katılımcı İspanyolca'nın ilk dili olduğunu belirtmişse, "ikincil dil olarak İngilizce" grubunda yer almaktadır. Başka bir örnek ise Tamil dilidir. Tamil dilinin konuşulduğu her ülkede (Tablo 3.6) İngilizce sadece resmi dildir. Bu sebeple, katılımcının ilk dili Tamil ise İngilizce'nin ikincil dil olduğu dil grubunda yer almaktadır.

⁸https://en.wikipedia.org/wiki/List_of_countries_where_Spanish_is_an_official_language

Tablo 3.6 Konuşuldukları ülkelerde % 75 veya daha fazla yüzdeyle İngilizce'nin ikincil dil olduğu İngilizce dışındaki diller

Konuşulan dil	İngilizce dil grubu yüzdesi
Portuguese	% 100
Spanish	% 87
Chinese	% 80
Indonesian	% 100
Valencian	% 100
Italian	% 86
Dutch	% 100
German	% 100
Greek	% 75
Korean	% 100
Persian	% 100
Galician	% 100
Kurdish	% 100
Bahasa	% 100
Tatar	% 100
Quechua	% 100

Yorumlardan Millet Bilgisinin Tespiti: Kullanıcıların millet bilgilerini tespit edebilmek için yorumlar içerisinde millet bilgisinin çıkarılabileceği cümle yapıları belirlenmiş ve bunları eşleyebilecek düzenli ifade kalıpları oluşturulmuştur. Millet bilgilerinin tespitinde kullanılan bazı cümle taslakları aşağıdaki gibidir;

- as a native **nation** (Bir yerli milletinden olarak)
- I was a native **nation** (Ben yerli milletindendim)

Şekil 3.12, katılımcıların milletlerini tespit etmek için kullanılan düzenli ifade kalıplarından birini (n1) göstermektedir. Örneğin, “I am a typical Japanese” cümlesini içeren bir yorum Şekil 3.12’de gösterilen kalıpla eşleşmektedir.

```
\b(?:i\s*\s*m|i\s*am|im|we|am)\s+(?:a\s+|an\s+|are\s+)?  
(?:typical\s+)?(\w{4,})\b[\.]*(!\s+\w+[eo]r)
```

Şekil 3.12 Millet bilgisini çıkarmak için yazılan düzenli ifade kalıplarının bir örneği (n1)

Yorumlardan millet bilgisinin tespitinde *Levensthain Sim* benzerlik fonksiyonu kullanılarak, % 75 ve üzeri kelime benzerliğine sahip eşleşmeler doğru kabul edilmişlerdir.

Yorumlardan Ülke/Şehir Bilgisinin Tespiti: Kullanıcıların ülke bilgilerini tespit edebilmek için yorumlar içerisinde ülke bilgisinin çıkarılabileceği cümle yapıları

belirlenmiş ve bu cümle kalıplarını tespit için düzenli ifadeler yazılmıştır. Ülke bilgilerinin tespitinde kullanılan bazı cümle taslakları aşağıdaki gibidir;

- I was born **country** (Ben ülkede doğdum)
- I grew up **country** (Ben ülkede büyüdüm)
- I come from **country** (Ben ülkeden geldim)
- having grown up in **country** (ülkede büyümüş olmak)

Yorumlardaki cümlelerden elde edilen ülke bilgilerinde, kullanıcı bir veya iki harfi yanlış yazmasından kaynaklanan hataları engellemek için *Levensthain Sim* benzerlik fonksiyonu ile eşik değeri olarak belirlenen % 75 ve üzeri kelime benzerliği kabulüyle ülke bilgileri tespit edilmiştir. R programlama dilindeki "Country Code" kütüphanesinin *getCountryCode* fonksiyonu ile ülke isimleri ülke koduna dönüştürülerek (Tablo 3.1) İngilizce dil grupları belirlenmiştir.

Şekil 3.13, kullanıcıların ülkelerini tespit etmek için kullanılan düzenli ifade kalıplarından birini (c3) göstermektedir. Örneğin, "I came from northern Turkey" cümlesini içeren bir yorum, Şekil 3.13'te gösterilen kalıpla eşleşmektedir.

```
\\b(?:i|we|i\\s*\\s*a?m?|i\\s*am|im|(?::my\\s+name\\s+is\\s+\\w+))\\s+[^\\.]*?(?:c[oa]me\\s+)?from(?:\\s+here)?(?:\\s+the)?\\s+(?:east(?:ern)?|north(?:ern)?|west(?:ern)?|south(?:ern)?\\s+)?(?:of\\s+)?(\\w{2,}|?!\\bmy)(?:\\s+republic)?
```

Şekil 3.13 Ülke bilgisini çıkarmak için yazılan düzenli ifadelerin bir örneği (c3)

3.3.1 "Understanding Language 4: Learning and Teaching" Kursu (MOOC) İçin Sonuçlar

"Understanding Language 4: Learning and Teaching" kursuna ait yorumlar (Comments) verisi, 50.332 satırdan oluşmaktadır (ön işleme sonrasında toplamda 4.164 farklı katılımcının 50.309 farklı yorumu vardır). Tablo 3.7, "Understanding Language 4: Learning and Teaching" kursunda, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.7 Katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları - "Understanding Language 4: Learning and Teaching" kursu

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
3305	602 (% 18,21)	459 (% 13,89)	2244 (% 67,90)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
674	148 (% 21,96)	68 (% 10,09)	458 (% 67,95)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
3305	643 (% 19,46)	434 (% 13,13)	2.228 (% 67,41)

Tablo 3.8 ve Tablo 3.9’da, "Understanding Language 4: Learning and Teaching" kursu yorumlarından oluşturulan ülke, şehir, millet ve dil odaklı düzenli ifade kalıpları ile İngilizce dil grubu tespit edilen katılımcı sayılarına yer verilmiştir. Tablo 3.8 ve Tablo 3.9’da ülke odaklı düzenli ifadeler ‘c’ harfi ile, millet odaklı düzenli ifadeler ‘n’ harfi ile ve dil odaklı düzenli ifadeler de ‘l’ harfi ile temsil edilmişlerdir. Bu harflerden sonra gelen rakamlar ise odak noktasında göre düzenli ifade kalıbının sırasını göstermektedir. Örneğin; c5 ülke odaklı beşinci düzenli ifade kalıbını göstermektedir.

Tablo 3.8 Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı (kurs öncesi anketinden ülke bilgisi bilinenler) - "Understanding Language 4: Learning and Teaching" kursu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	3	0	8
c2	3	0	8
c3	16	29	111
c4	16	29	111
c5	17	31	115
c6	17	31	120
l1	46	44	339
l2	73	45	351
l3	109	48	448
l4	109	48	448
l5	109	48	448
l6	125	50	477
l7	125	50	477
l8	125	50	477
l9	127	52	478
l10	128	52	478
l11	128	53	478
l12	128	53	478
l13	128	53	478
n1	139	55	513
n2	142	55	513

"Understanding Language 4: Learning and Teaching" kursunda veri ön işleme sonrasında toplamda 4.164 farklı katılımcının yorumları mevcuttur. Bu katılımcılardan 2.626 tanesinin (% 63,06) oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilmiştir. İngilizce dil grubu tespit edilen katılımcılardan 674 tanesi, kurs öncesi anketlerinde ülke bilgilerini vermişlerdir. Bu katılımcılardan % 78,19'unun tespit edilen İngilizce dil grupları ile katılımcıların kurs öncesi anketinde verdikleri ülke bilgilerinin İngilizce dil grupları aynı iken, 147 (% 21,81) tanesinin yorumlarından oluşturulan düzenli ifade kalıpları ile elde edilen İngilizce dil grup bilgilerinin, kurs öncesi anketlerinde verdikleri ülke bilgilerinden elde edilenden farklı oldukları görülerek, İngilizce dil grupları yorumlardan elde edilen İngilizce dil grupları ile güncellenmiştir.

Sonuç olarak, oluşturduğumuz düzenli ifade kalıpları ile katılımcılar hakkında kurs forumlarındaki konuşmaları sırasında birincil dilleri, şehirleri/ülkeleri veya milletleri hakkında bilgi verdikleri sürece, İngilizce dil grupları tespit edilebilmektedir.

Tablo 3.9 Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı (kurs öncesi anketinde ülke bilgisi yer almayanlar) - "Understanding Language 4: Learning and Teaching" kurs serisi

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	7	1	16
c2	7	1	16
c3	39	58	350
c4	39	58	350
c5	40	59	352
c6	42	60	360
l1	141	119	994
l2	208	120	1.034
l3	283	139	1.243
l4	283	139	1.243
l5	283	139	1.243
l6	333	147	1.329
l7	333	147	1.329
l8	333	147	1.329
l9	340	148	1.333
l10	340	148	1.333
l11	340	148	1.334
l12	343	149	1.348
l13	343	149	1.348
n1	378	162	1.426
n2	384	163	1.434

3.3.2 "Understanding Language: Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonu İçin Sonuçlar

"Understanding Language: Learning and Teaching" kurs serisi, FutureLearn platformu üzerinde, British Council ve Southampton Üniversitesi işbirliği ile 2014 - 2018 yılları arasında yayınlanmıştır.

"Understanding Language: Learning and Teaching" kurs serisinin ilk yedi iterasyonunda her bir iterasyona ait toplam katılımcı, toplam yorum, yorum yapan katılımcı sayıları ile kurs öncesi ankette ülke bilgilerini veren katılımcı sayısı bilgileri, Tablo 3.10'da yer almaktadır. Tablo 3.10'a göre, katılımcıların sadece % 5,87'si kurs öncesi anket cevaplarında ülke bilgilerini vermişlerdir. Katılımcıların % 15,99'u tarafından 421.375 tane yorum yapılmıştır. Bu sebeple, İngilizce dil grupları tespit edilen katılımcı yüzdesini ve gruplamanın doğruluğunu arttırmak açısından yorumlar daha büyük potansiyele sahiptir.

Tablo 3.10 "Understanding Language : Learning and Teaching" (UL) kurs serisinin ilk yedi iterasyonunda katılımcı ve yorum sayıları

	Katılımcı Sayısı	Yorum Sayısı	Yorum Yapan Katılımcı Sayı ve Yüzdesi	Kurs öncesi anketinde ülke bilgisini veren katılımcı Sayı ve Yüzdesi
UL 1	58.781	145.425	11.315 (% 19,25)	1.793 (% 3,05)
UL 2	41.,912	86.139	7.585 (% 18,10)	1.516 (% 3,6)
UL 3	44.293	58.285	5.199 (%11,74)	1.207 (% 2,7)
UL 4	25.597	50.332	4.164 (% 16,27)	3.305 (% 12,9)
UL 5	19.866	37.637	2.431 (% 12,24)	2.292 (% 11,5)
UL 6	10.275	18.616	1.319 (% 12,84)	1.019 (% 9,9)
UL 7	12.893	24.941	2.153 (% 16,70)	1.413 (% 10,96)
Toplam	213.617	421.375	34.166 (% 15,99)	12.545 (% 5,87)

Tablo 3.11, "Understanding Language : Learning and Teaching" kurs serisinde, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.11 "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonundaki katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
12.545	2.154 (% 17,17)	1.569 (% 12,51)	8.822 (% 70,32)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
2616	406 (% 15,52)	296 (% 11,31)	1914 (% 73,17)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
12.545	2.297 (% 18,31)	1.509 (% 12,03)	8.739 (% 69,66)

"Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonunda, toplamda 31.166 katılımcı tarafından 421.375 tane yorum yazılırken, yorumlardan 344.260 tanesi düzenli ifade kalıplarımızla eşleşmiş olup, bu

eşleşmelerle 21.621 katılımcının İngilizce dil grubu tespit edilmiştir. Tablo 3.12, "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonunda, düzenli ifade kalıpları kullanılarak İngilizce dil grupları tespit edilen katılımcı sayılarını göstermektedir. Tabloda (Tablo 3.12) ülke, millet ve dil odaklı düzenli ifadeler sırası ile "c", "n" ve "l" harfi ile temsil edilmektedir. Bu harflerden sonra gelen rakamlar ise odak noktasında göre düzenli ifade kalıplarının sıralarını göstermektedir. Örneğin; "c5" ülke odaklı beşinci düzenli ifade kalıbına karşılık gelmektedir.

Tablo 3.12 Oluşturulan düzenli ifade kalıpları ile İngilizce dil grupları tespit edilen katılımcı sayısının kümülatif olarak artışı - "Understanding Language: Learning and Teaching" kurs serisinin ilk yedi iterasyonu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	75	49	202
c2	75	49	202
c3	433	1.902	4.424
c4	433	1.902	4.424
c5	438	1.143	4.485
c6	459	1.161	4.654
l1	1.313	1.762	11.427
l2	1.725	1.768	11.672
l3	2.634	1.940	13.699
l4	2.634	1.940	13.700
l5	2.634	1.940	13.700
l6	3.205	2.016	14.427
l7	3.205	2.016	14.429
l8	3.205	2.016	14.429
l9	3.253	2.028	14.478
l10	3.254	2.028	14.479
l11	3.255	2.031	14.378
l12	3.284	2.044	14.611
l13	3.287	2.044	14.636
n1	3.659	2.173	15.663
n2	3.708	2.176	15.737

3.3.3 Farklı Türden Kurslarda (MOOC'lar) Sonuçlar

Tablo 3.13'te, "Exploring Our Oceans 4" MOOC'unun toplam katılımcı, yorum, katılımcılar tarafından gönderilen yorum sayıları ile kurs öncesi ankette ülke bilgilerini veren katılımcı sayısı bilgileri yer almaktadır. Tablo 3.13'e göre katılımcıların sadece % 12,44'ü kurs öncesi anket cevaplarında ülke bilgilerini vermişlerdir. Katılımcıların % 10,32'si ise, 3.200 tane yorum yapmışlardır. İngilizce dil gruplarının tespit edilen katılımcı sayısını ve gruplandırmanın doğruluğunu arttırmak açısından bu yorumlardan faydalanılmıştır.

Tablo 3.13 "Exploring Our Oceans 4" kursunda katılımcı ve yorum sayıları

Katılımcı Sayısı	Yorum Sayısı	Yorum Yapan Katılımcı Sayı ve Yüzdesi	Kurs Öncesi Anketinde Ülke Bilgisini Veren Katılımcı Sayı ve Yüzdesi
3.867	3.200	399 (% 10,32)	481 (% 12,44)

Tablo 3.14’de, "Web Science 1" MOOC’unda; toplam katılımcı, yorum, katılımcılar tarafından gönderilen yorum sayıları ile kurs öncesi ankette ülke bilgilerini veren katılımcı sayısı bilgileri yer almaktadır. Tablo 3.14’e göre katılımcıların sadece % 2,87’si kurs öncesi anket cevaplarında ülke bilgilerini vermişlerdir. Katılımcıların % 16,08’i ise, 18.414 adet yorum yapmışlardır. İngilizce dil gruplarının tespit edilen katılımcı sayısını ve gruplandırmanın doğruluğunu arttırmak açısından bu yorumlardan da faydalanılmıştır.

Tablo 3.14 "Web Science 1" kursunda katılımcı ve yorum sayıları

Katılımcı Sayısı	Yorum Sayısı	Yorum Yapan Katılımcı Sayı ve Yüzdesi	Kurs Öncesi Anketinde Ülke Bilgisini Veren Katılımcı Sayı ve Yüzdesi
13.673	18.414	2.198 (% 16,08)	392 (% 2,87)

Tablo 3.15’de, "AginCourt 3" MOOC’unda; toplam katılımcı, yorum, katılımcılar tarafından gönderilen yorum sayıları ile kurs öncesi ankette ülke bilgilerini veren katılımcı sayısı bilgileri yer almaktadır. Tablo 3.15’e göre katılımcıların % 14,89’u kurs öncesi anket cevaplarında ülke bilgilerini vermişlerdir. Katılımcıların % 14,22’si ise, 14.096 adet yorum yapmışlardır. İngilizce dil gruplarının tespit edilen katılımcı sayısını ve gruplandırmanın doğruluğu arttırmak açısından bu yorumlardan da faydalanılmıştır.

Tablo 3.15 "AginCourt 3" kursunda katılımcı ve yorum sayıları

Katılımcı Sayısı	Yorum Sayısı	Yorum Yapan Katılımcı Sayı ve Yüzdesi	Kurs Öncesi Anketinde Ülke Bilgisini Veren Katılımcı Sayı ve Yüzdesi
7.314	14.096	1.040 (% 14,22)	1.089 (% 14,89)

Tablo 3.16’da, "Research Project 5" MOOC’unda; toplam katılımcı, yorum, katılımcılar tarafından gönderilen yorum sayıları ile kurs öncesi ankette ülke bilgilerini veren katılımcı sayısı bilgileri yer almaktadır. Tablo 3.16’ya göre katılımcıların % 10,66’sı kurs öncesi anket cevaplarında ülke bilgilerini vermişlerdir. Katılımcıların % 15,96’sı ise, 8.111 adet yorum yapmışlardır. İngilizce dil gruplarının tespit edilen katılımcı

sayısını ve gruplandırmanın doğruluğu arttırmak açısından bu yorumlardan da faydalanılmıştır.

Tablo 3.16 "Research Project 5" kursunda katılımcı ve yorum sayıları

Katılımcı Sayısı	Yorum Sayısı	Yorum Yapan Katılımcı Sayı ve Yüzdesi	Kurs Öncesi Anketinde Ülke Bilgisini Veren Katılımcı Sayı ve Yüzdesi
6.568	8.111	1.048 (% 15,96)	700 (% 10,66)

Tablo 3.17, "Exploring Our Oceans 4" kursunda, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.17 "Exploring Our Oceans 4" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
481	274 (% 56,96)	46 (% 9,56)	161 (% 33,47)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
17	1 (% 5,88)	3 (% 17,65)	13 (% 76,47)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
481	267 (% 55,51)	48 (% 9,98)	166 (% 34,51)

Tablo 3.18, "Web Science 1" kursunda, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.18 "Web Science 1" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
392	309 (% 78,83)	17 (% 4,34)	66 (% 16,84)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
12	9 (% 75)	0 (% 0)	3 (% 25)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
392	304 (% 77,56)	20 (% 5,10)	68 (% 17,35)

Tablo 3.19, "Agincourt 3" kursunda, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.19 "Agincourt 3" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
1.089	512 (% 47,02)	115 (% 10,56)	462 (% 42,42)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
27	20 (% 74,07)	2 (% 7,41)	5 (% 18,52)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
1.089	500 (% 45,91)	117 (% 10,74)	472 (% 43,34)

Tablo 3.20, "Research Project 5" kursunda, kurs öncesi anket bilgileri ve oluşturulan düzenli ifadelerle katılımcıların İngilizce dil gruplarının tespiti ve güncellenmesi aşamalarında elde edilen sonuçları göstermektedir.

Tablo 3.20 "Research Project 5" kursunda katılımcıların kurs öncesi anketlerindeki ülke bilgilerinden oluşturulan İngilizce dil gruplarının oluşturulan düzenli ifadelerle tespiti ve güncellenmesi adımlarının sonuçları

I. Kursa nerelerden katıldıkları sorusuna cevap veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
700	204 (% 29,14)	182 (% 26)	314 (% 44,86)
II. Hem Kurs öncesi anketi yanıtlayıp hem de tartışma forumunda konuşulan dil / ülke hakkında ek bilgi veren katılımcılar			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
52	11 (% 21,15)	15 (% 28,85)	26 (% 50)
III. Varsayılan dil grubunu (kurs öncesi anketten elde edilen) tartışmalardan elde edilen bilgilerle güncelleme			
Toplam	Birincil ve Resmi Dil	Yalnızca Resmi Dil	İkincil Dil
700	202 (% 28,86)	179 (% 25,57)	319 (% 45,57)

"Exploring Our Oceans 4" kursunda, 399 katılımcı tarafından 3.200 tane yorum yazılırken, 17 katılımcının İngilizce dil grubu yorumlarından tespit edilmiştir. Tablo 3.21, "Exploring Our Oceans 4" kursunda düzenli ifade kalıpları kullanılarak İngilizce dil grupları tespit edilen katılımcı sayılarını göstermektedir.

Tablo 3.21 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Exploring Our Oceans 4" kursu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	0	0	0
c2	0	0	0
c3	0	2	11
c4	0	2	11
c5	0	3	11
c6	1	3	11
l1	1	3	11
l2	1	3	11
l3	1	3	11
l4	1	3	11
l5	1	3	11
l6	1	3	11
l7	1	3	11
l8	1	3	11
l9	1	3	11
l10	1	3	11
l11	1	3	11
l12	1	3	11
l13	1	3	11
n1	1	3	12
n2	1	3	13

"Web Science 1" kursunda, 2.198 katılımcı tarafından 18.414 tane yorum yazılırken, 128 katılımcının İngilizce dil grubu yorumlarından tespit edilmiştir. Tablo 3.22, "Web Science 1" kursunda düzenli ifade kalıpları kullanılarak İngilizce dil grupları tespit edilen katılımcı sayılarını göstermektedir.

Tablo 3.22 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Web Science 1" kursu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	1	1	1
c2	1	1	1
c3	22	12	67
c4	22	12	67
c5	22	13	71
c6	22	15	74
l1	22	15	74
l2	22	15	74
l3	22	15	74
l4	22	15	74
l5	22	15	74
l6	22	15	75
l7	22	15	75
l8	22	15	75
l9	22	15	75
l10	22	15	75
l11	22	15	75
l12	22	15	75
l13	22	15	75
n1	23	15	81
n2	25	15	88

"Agincourt 3" kursunda, 1.040 katılımcı tarafından 14.096 tane yorum yazılırken, 74 katılımcının İngilizce dil grubu yorumlarından tespit edilmiştir. Tablo 3.23, "Agincourt 3" kursunda düzenli ifade kalıpları kullanılarak İngilizce dil grupları tespit edilen katılımcı sayılarını göstermektedir.

Tablo 3.23 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Aginccourt 3" kursu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	3	0	0
c2	3	0	0
c3	16	9	25
c4	16	9	25
c5	16	9	26
c6	16	9	27
l1	16	9	27
l2	16	9	27
l3	16	9	27
l4	16	9	27
l5	16	9	27
l6	16	9	27
l7	16	9	27
l8	16	9	27
l9	16	9	27
l10	16	9	27
l11	16	9	27
l12	16	9	27
l13	16	9	27
n1	27	10	33
n2	28	10	36

"Research Project 5" kursunda, 1.048 katılımcı tarafından 8.111 tane yorum yazılırken, 192 katılımcının İngilizce dil grubu yorumlarından tespit edilmiştir. Tablo 3.24, "Research Project 5" kursunda düzenli ifade kalıpları kullanılarak İngilizce dil grupları tespit edilen katılımcı sayılarını göstermektedir.

Tablo 3.24 Oluşturulan düzenli ifadelerle İngilizce dil grubu tespit edilen katılımcı sayısının kümülatif olarak artışı - "Research Project 5" kursu

Düzenli İfade Kalıpları	Birincil ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
c1	0	0	0
c2	0	0	0
c3	22	47	85
c4	22	47	85
c5	24	48	88
c6	26	51	92
l1	26	51	92
l2	26	51	92
l3	26	51	92
l4	26	51	92
l5	26	51	92
l6	26	51	92
l7	26	51	92
l8	26	51	92
l9	26	51	92
l10	26	51	92
l11	26	51	92
l12	26	51	92
l13	26	51	92
n1	38	52	101
n2	38	52	102

3.4 Derin Öğrenme İle İngilizce Dil Grubu Tespiti

Derin Öğrenme, verileri etiketlemeye gerek kalmadan çok çeşitli ve yapılandırılmamış veri kümelerinden öğrenmeyi sağlamaktadır. MOOC'larda, Derin Öğrenme çeşitli amaçlar için kullanılmıştır. Örneğin, Liao ve diğerleri [109] MOOC'ların değerlendirme süreci kalitesini arttırmak amacı ile MOOC öğrencilerinin çoklu soru performansını tahmin etmek için Derin Öğrenme'yi kullanmıştır. Liu ve diğerleri [110] ile Wang ve diğerleri [66] kursu bırakacak olan katılımcıları tahmin etmek için Derin Öğrenme'yi kullanmışlardır.

Tezde, katılımcıların İngilizce dil gruplarını tespit etmek için, kurs öncesi anket bilgileri ile, oluşturulan düzenli ifade kalıpları ve kurs öncesi anket bilgileri ile ve Derin öğrenme ile olmak üzere üç yöntem önerilmiştir.

"Understanding Language : Learning and Teaching" kurs serisinden farklı türdeki kurslarda, katılımcıların birincil dil, şehir/ülke veya uyruk bilgilerden birini verme ihtimalleri daha düşüktür. Böyle durumlarda, yorumlardan çıkarılan özellikler veya özel olarak tanımlanmış desenler vermeden öğrenebilecek bir algoritmaya ihtiyaç

duyulmaktadır. Bu nedenle, tartışma grubuna ne yazdıklarına bakılmaksızın (ilk dilleri ile ilgili bilgi vermeseler de), katılımcıların İngilizce dil gruplarını tespit etmek için Derin Öğrenme yönteminin de denenmesine karar verilmiştir.

Derin öğrenme ile kullanıcıların dil grubu tahmininde hem "oluşturulan standart özellikler" hem de "yorum metinleri" (ön işlemeden sonra) kullanılmıştır.

3.4.1 Standart Özellikleri Oluşturma

Derin Öğrenme ile İngilizce dil grubu tahmininde kullanılan ilk veri, FutureLearn platformundaki katılımcıların aktivite özelliklerinden faydalananarak oluşturulan özellikleri içermektedir. Bölüm 4'te standart özelliklerin oluşturulma kuralları ve ham verilerden faydalanan özellikler hakkında detaylı bilgilendirme yapılmıştır. Oluşturulan özellikler, tez kapsamında gerçekleştirdiğimiz çalışmalardan birinde [111] kurs sonu performans tahminleri için de kullanılmıştır.

Tablo 3.25, "oluşturulan standart özellikler"den İngilizce dil grubu tahmininde kullanılanları göstermektedir. Bu özelliklere ek olarak, Derin Öğrenme modelini eğitmek için kurs tamamlama durumu da tahmin özelliği olarak kullanılmıştır. Düzenli ifadelere dayalı modelimizle tespit edilen İngilizce dil grupları da oluşturulan modellerin tahmin sonuçlarının doğruluk testleri için kullanılmıştır (İngilizce dil gruplarının kategorilerinin değerleri *One Hot Encoding tekniği* ile üç boyutlu bir vektöre dönüştürülmüştür).

Tablo 3.25 Derin Öğrenme ile İngilizce dil grubu tahmini için kurs aktivite kayıtlarından faydalananarak oluşturulan standart özellikler (veri dosyalarına göre)

Step Activity	Comments
başlanan adım sayısı	yorum sayısı
tamamlanmadan bırakılan adımı sayısı	ana yorum sayısı
ilk hafta tamamlanan adım sayısı	cevap yorum sayısı
ikinci hafta tamamlanan adım sayısı	yorum uzunluk ortalaması
üçüncü hafta tamamlanan adım sayısı	ana yorum uzunluk ortalaması
dördüncü hafta tamamlanan adım sayısı	cevap yorum uzunluk ortalaması
	yorum beğeni sayısı ortalaması

3.4.2 Kullanıcı Yorum Metinlerinin Ön işleme

Derin Öğrenme modellerinde kullanılan ikinci veri kümesi kullanıcı yorum metinlerini içermektedir. Derin Öğrenme modellerine girecek yorumların metinlerinde gerçekleştirilen ön işleme adımları sırası ile aşağıda listelenmiştir.

1. Yorumlarda geçen İnternet sitesi linkleri, Şekil 3.14'te yer alan link temizleme

düzenli ifadesi kullanılarak silinmiştir.

```
str_replace_all(text, "[a-zA-Z0-9@:%._\\+~#={2,256}\\.[a-z]{2,6}\\b  
([-a-zA-Z0-9@:%._\\+~#?&\\/=]*)", "")
```

Şekil 3.14 Link (URL) temizleme düzenli ifadesi

2. Kurs performansı veya İngilizce dil grubu bilgileri bilinmeyen kullanıcılara ait yorumlar silinmiştir.
 3. Uzunluğu 100 karakterden kısa olan yorumlar yetersiz bilgi içerdiğinden silinmiştir.
 4. İngilizce dil gruplarının bilgileri (“birincil ve resmi dil olarak İngilizce” : 0, yalnızca resmi dil olarak İngilizce : 1, “ikincil dil olarak İngilizce” : 2 olarak) sayısal değere dönüştürülmüştür.
 5. Yorumlarda yer alan ‘1-9’ aralığındaki rakamlar sıfıra dönüştürülmüştür.
 6. Yorumlardaki karakterlerin tamamı küçük harfe çevrilmiştir.
 7. “Glove Global Vectors For Word Representation”dan [112] faydalanmak için yorum metinlerinde en sık geçen 20.000 kelime bulunmuştur.
 8. Yorumlar kelimelerine ayrılarak, her kelime yerine “Glove Global Vectors For Word Representation”da [112] kelimelere karşılık gelen indisler bulunarak atanmıştır. Böylece, tüm yorumlar sayı dizileri haline dönüştürülmüştür.
 9. Derin Öğrenme modelleri için sayı dizilerine dönüştürülen yorumların uzunluklarını eşitlemek gerektiği için, 200 kelimeden uzun olan yorumlarda ilk 200 kelimeye karşılık gelen ve ön işlemenin sekizinci adımda oluşturulan sayı dizileri kullanılmıştır. 200 kelimeden az kelime içeren yorumlarda ise sayı dizilerinin başlarına sıfırlar eklenerek uzunlukları 200’e eşitlenmiştir.
 10. Katılımcı yorumları, yorum yapan *author_id*’lerin % 80’i eğitim ve % 20’si test kümesinde yer alacak şekilde rastgele seçilerek eğitim ve test kümelerine ayrılmıştır.
 11. Derin Öğrenme ile İngilizce dil grubu tahmini için her İngilizce dil grubunda eşit sayıda yorum olacak şekilde, yorumlar üç farklı örnekleme yöntemi ile (aşırı örnekleme, ortalama örnekleme (her dil grubunda toplam yorum sayısının ortalaması kadar) ve aşağı örnekleme) ile dengelenmiştir.
- Not: Bölüm 5’te yer verilen Derin Öğrenme ile performans tahmini için her kurs

performansı deęerinden eřit sayıda yorum olacak řekilde, yorumlar aynı üç farklı örnekleme yöntemi ile dengelenmiřtir.

12. Birden fazla kursa kayıtlı yani tekrarlı kullanıcıların rastgele olarak bir kurs verisi seçilerek benzersiz kullanıcı verilerini içerecek řekilde yorumlar verisi güncellenmiřtir.
13. On ikinci adımda elde edilen yorum verisinden yorumların en uzun % 50'si seçilmiřtir (Seçilen benzersiz kullanıcıların yorumlarının sayısı ikiye eřit veya büyük ise 'n' uzunluęu göstermek üzere ' $n/2 + 1$ ' tane yorum, bir tane ise tamamı alınmaktadır)

3.4.3 Kullanılan Derin Öğrenme Modelleri

Derin Öğrenme kullanırken, yorumlar harf seviyesi, hece seviyesi, kelime seviyesi, cümle seviyesi veya paragraf seviyesi gibi farklı seviyelerde işlenebilmektedir. Harf seviyesi ve hece seviyesinde anlamlı bir ilişki kurmak için çok daha güçlü bir işlemciye ihtiyaç duyulabilmektedir. Cümle veya paragraf seviyeleri de İngilizce dil grubu tahmini problemini çözecek kadar ayrıntılı değildir. Bu sebeple, modelde yer verilen *Embedding* (Kodlama) katmanı, kelime düzeyinde oluşturulmuřtur.

Word Embeddings, Doğal Dil İşleme'de, özellikle de sınıflandırma problemlerinde en sık kullanılan tekniklerden biridir [113]. Yaygın olarak kullanıldığı için, birçok farklı kütüphane ve bunun için önceden eğitilmiş vektör verileri vardır. Tezde, "Glove Global Vectors For Word Representation" kullanılmıştır [112]. "Glove Global Vectors For Word Representation" 300 boyutlu bir vektör formatında 400.000 farklı kelime sunmaktadır. Bu gösterimde, benzer anlamı olan veya anlam olarak yakın ilişkileri olan kelimelerin birbirlerine yakın vektör deęerleri vardır. Bunun karşısında da zıt veya uzak anlamı olan kelimelerin de uzak deęerleri vardır [112].

Derin Öğrenme İngilizce dil grubu tahmini için kullanılan veri türüne göre üç farklı model ile gerçekleştirilmiştir. İlk modelde "yorum metinleri ve oluşturulan standart özellikler" kullanılırken, ikinci modelde "sadece yorumlar", üçüncü modelde ise "sadece oluşturulan standart özellikler" modelin eğitiminde kullanılmıştır.

Katılımcıların "yorumlar ve oluşturulan standart özelliklerle" İngilizce dil grubu tahmini Derin Öğrenme modelinde kullanılan katmanlar ve açıklamaları:

1. **Girdi (Input) Katmanı:** Girdi katmanı, modelin ilk katmanı olarak girdi verilerini almak için kullanılmıştır.

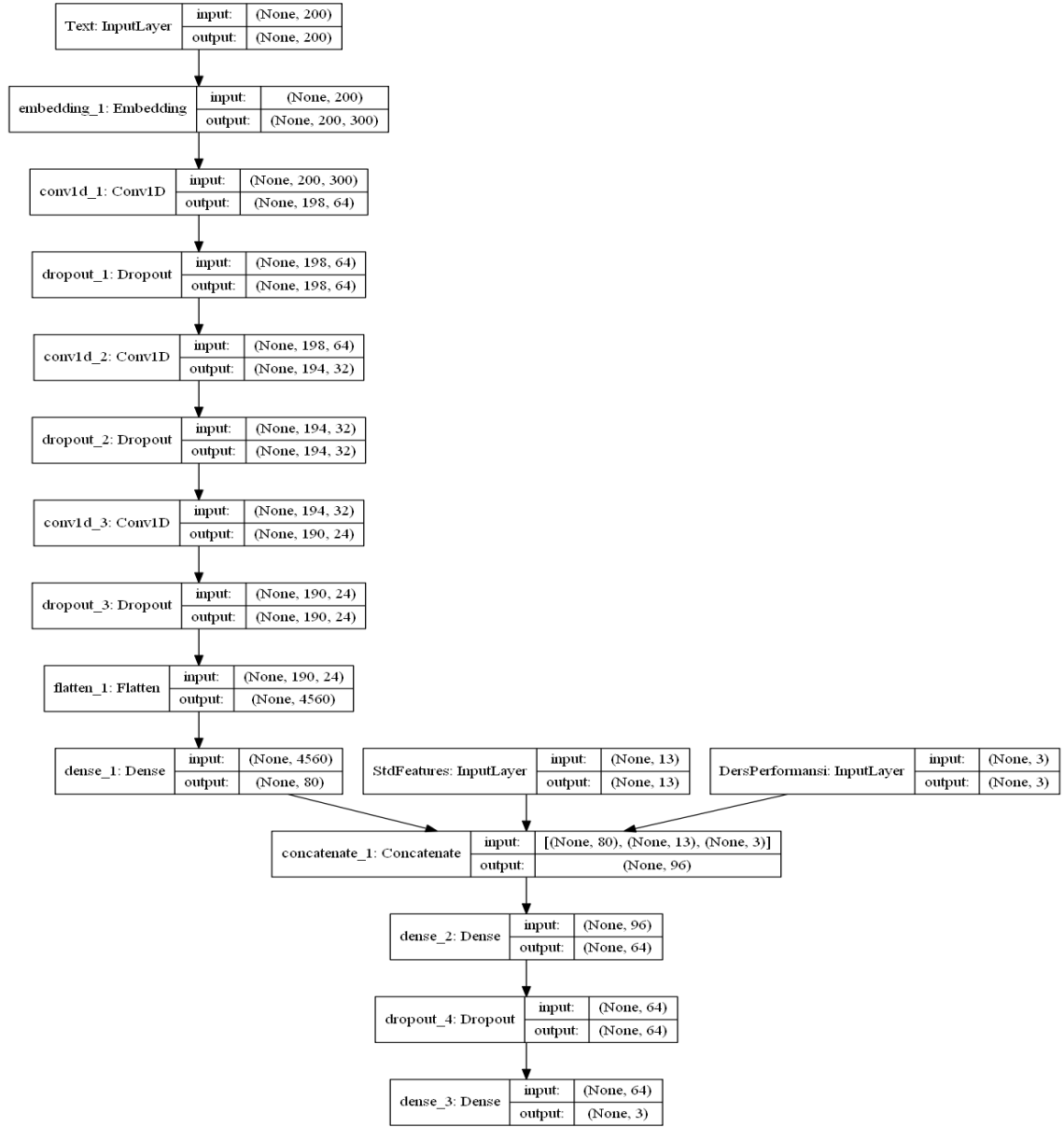
2. **Kodlama (Embedding) Katmanı:** *Embedding* katmanında, her kelime indeksi bir girdi olarak alınmakta ve kelimeler 300 boyutlu vektörlere dönüştürülmektedir. Bu işlemin *Embedding* katmanında yapılması, eğitim süresini kısaltmaktadır (özellikle Grafik İşleme Ünitesi (GPU) ile kullanıldığında veri ön işleme adımı kelime vektöre dönüştürmeye göre daha hızlıdır). Katılımcıların yorumlarından en sık 20.000 kelime *Embedding* katmanında “Glove Global Vectors For Word Representation” ile vektör formatlarına dönüştürülmüşlerdir. Ardından bu vektörler, *Evrişim* katmanına aktarılmıştır.

3. Yorum metinlerini İşlemede Denenen Mimariler:

- **Evrişim (Konvolüsyon) katmanı (1 Boyutlu):** *Evrişim* katmanı, standart bir katmandan yorumları bir bütün olarak değil, kısmen ve ardışık olarak önceden tanımlanmış çekirdek büyüklüğü ile işlemesi yönü ile farklıdır. Katman, İngilizce dil grubu bilgisini elde etmek için katmana parametre olarak verilen sayıda komşu kelimeleri (çekirdek boyutu) kontrol etmektedir. Problemimizde bilgi, yorumun herhangi bir yerinde bulunabileceği için, konum bilgisi problemimiz için anahtar bir gösterge değildir. Modelimizde *Evrişim* katmanları, sırasıyla üç komşu kelimeyi, beş komşu kelimeyi ve yedi komşu kelimeyi aramak için kullanılmışlardır. İngilizce dil grubunu tespit etmek için kelimeleri belirli bir komşulukta işlemek daha doğru bir yöntem sunabileceği için kullanılabilecek farklı mimarilerden ilk olarak, *Evrişim* mimarisi uygulanmıştır.
- **LSTM:** LSTM yöntemi uzun vadel bağımlılıkları modelleyen bellek hücreleri içeren ve özellikle metin analizi gibi dizi öğrenme problemleri için faydalıdır. Her bir hücrede üç kapı vardır (giriş (input), çıkış (output) ve unutma (forget)). Bağımlılıklara atanan ağırlıklara göre, hücre hangi bilgilerin ne zaman unutulacağına karar verir. Bu nedenle, geçmişten geleceğe bilgileri daha uzun süre taşır. Bu yönü ile *Evrişim* katmanından farklıdır. Çünkü *Evrişim* katmanı her çalışma sonunda hangi bilgilerin kesinlikle bırakılacağına karar verir.
- **İki Yönlü (Bidirectional) LSTM:** Geleneksel LSTM’e ek olarak iki yönlü LSTM geçmiş ve gelecek arasındaki bilgileri ileri ve geri taşır.
- **GRU :** Üç kapı yerine sadece iki kapıya sahip olan LSTM’nin daha basit versiyonudur.
- **Konvolüsyon+LSTM:** Eğitim sürecini ve tahmin doğruluk oranlarını iyileştirmek amacı ile Konvolüsyon ve LSTM katmanlarının hibrit katmanlar oluşturacak şekilde birlikte kullanıldığı mimaridir.

4. **Bırakma (Drop-out) Katmanı:** Eğitim sürecinde oluşabilecek aşırı uyum probleminden kaçınmak için Bırakma katmanı kullanılmıştır. Bırakma katmanı, belirlenen oranda (% 50) rastgele seçilen sinir ağı çıktılarının bir sonraki katmana aktarılmasını engellemektedir. Böylece model daha fazla bilgi öğrenmeye zorlanmaktadır.
5. **Düzleştirme (Flatten) Katmanı:** Düzleştirme katmanı, matris formatındaki veriyi tek boyutlu vektör formatına dönüştürmektedir.
6. **Birleştirme (Concatenate) Katmanı:** Birleştirme katmanı iki veya daha fazla katmandan gelen bilgileri birleştirmekte ve bir sonraki katmana göndermektedir. Birleştirme katmanı, üç katmanın çıkış vektörlerini birleştirmek için kullanılmıştır. Birleştirme katmanının çıktısı kendisinden sonra gelen Yoğun (Dense) katmanına gönderilmiştir.
7. **Yoğun (Dense) Katmanı:** Yoğun katmanı, hem kendinden önceki hem de kendinden sonraki katmandaki tüm sinir ağlarına tamamen bağlantılı bir katmandır. Yoğun katmanı, sinir ağlarının mantıklı kararlar vermelerine yardımcı olmaktadır. Modelimizdeki, son Yoğun katmanı, üç İngilizce dil grubundan birini gösterecek şekilde tahmin çıktıları üretmektedir.

Şekil 3.15, İngilizce dil grubu tahmini için oluşturulan özellikler ve ön işlenmiş yorum verileri ile birlikte çalışan Derin Öğrenme modelimizi (Konvolüsyon mimarisi kullanıldığı durumda) göstermektedir.



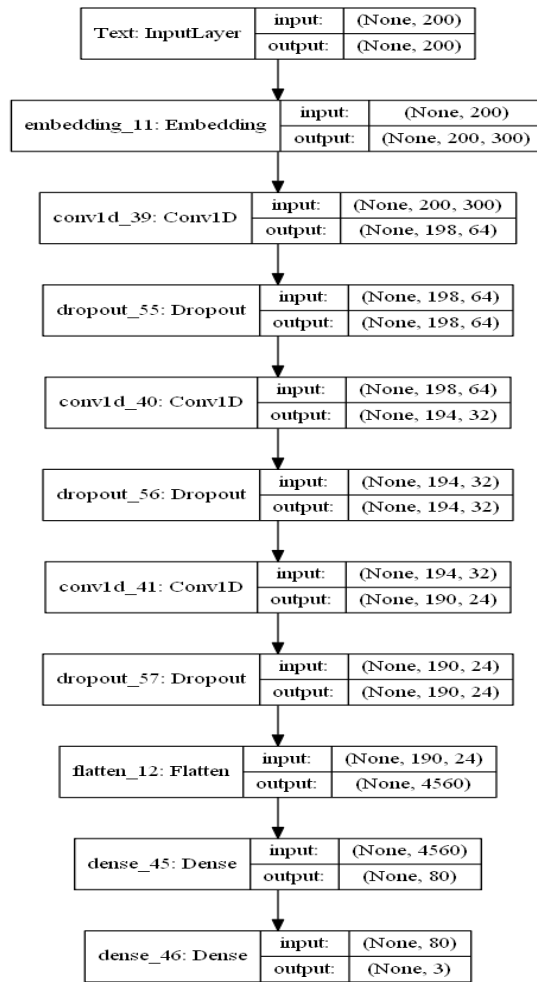
Şekil 3.15 Oluşturulan standart özellikler ve yorumlarla İngilizce dil grubu tahmini için Derin Öğrenme modeli

Şekil 3.15'te yer verilen model, öncelikle Girdi katmanı ile yorumları girdi olarak almaktadır. Girdi katmanından sonra *Embedding* katmanında, 200 kelime uzunluklu yorumlardaki her bir kelime “Glove global vectors for word representation”da karşılık gelen 300 boyutlu bir vektör formatına dönüştürülmektedir. Ardından beş farklı mimari (her biri ayrı modelde) denenmiştir. Örneğin Konvolüsyon mimarisi kullanıldığında, Evrişim katmanları kullanılarak “sadece yorumlar” ile eğitilmektedir (Birinci Evrişim katmanı art arda üç komşu kelimeyi aramak, ikincisi beş komşu kelimeyi aramak, üçüncüsü sırasıyla yedi komşu kelimeyi aramak için kullanılmıştır). Her Evrişim katmanından sonra, daha iyi öğrenme için birer tane Bırakma katmanı yer almaktadır. Daha sonra, Düzleştirme katmanı ile sonuçlar tek boyutlu hale

getirilmekte ve Yoğun katmanı ile modelin ikinci kısmında (oluşturduğumuz standart özelliklerle birlikte) tekrar girdi olarak kullanılacak bir eğitim çıktısı üretmektedir.

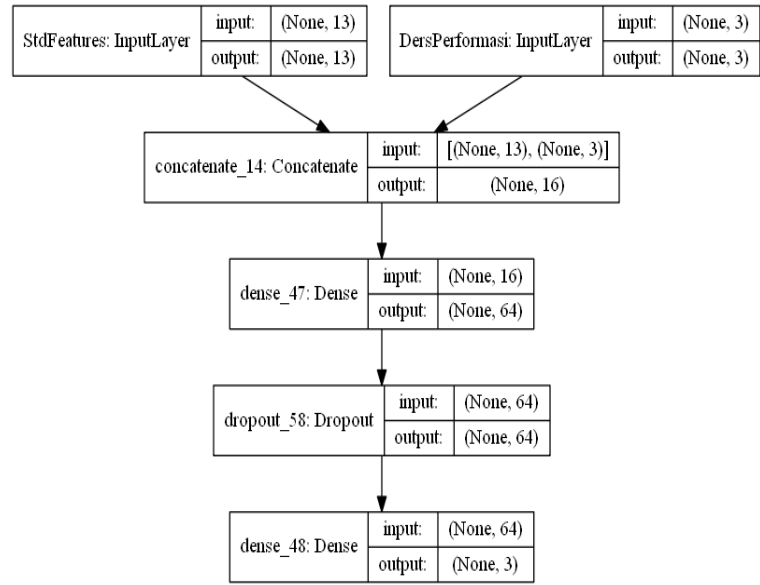
Şekil 3.15'teki modelin ikinci kısmında üç girdi alınmıştır. Bu girdiler; modelin ilk kısmının eğitim çıktısı, Tablo 3.25'te yer verilen oluşturulan özellikler ve her katılımcı için kurs bitirme durumu özelliğidir. Model, bu üç girdi Birleştirme katmanında birleştirmektedir. Modelin ikinci kısmı, Yoğun ve Bırakma katmanları ile eğitimden sonra, tespit edilen İngilizce dil kategorileri bilgilerini çıktı olarak vermektedir.

Şekil 3.16, “sadece yorumlar” ile İngilizce dil grubu tahmini Derin Öğrenme modelimizi (Konvolüsyon mimarisi kullanıldığı durumda) göstermektedir.



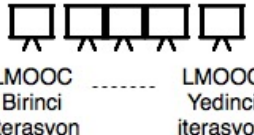
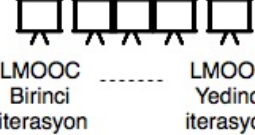
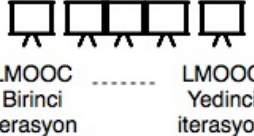
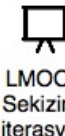
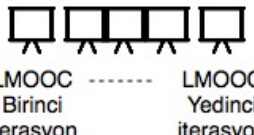
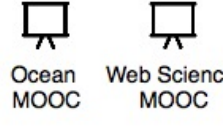
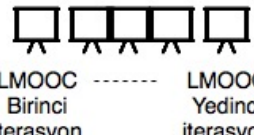
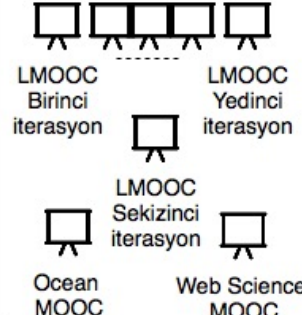
Şekil 3.16 Sadece yorum metinleri ile İngilizce dil grubu tahmini için Derin Öğrenme modeli

Şekil 3.17, ”sadece oluşturulan standart özellikler” ile İngilizce dil grubu tahmini Derin Öğrenme modelimizi göstermektedir.



Şekil 3.17 Sadece oluşturulan standart özelliklerle İngilizce dil grubu tahmini için Derin Öğrenme modeli

3.5 Derin Öğrenme Modellerinde Eğitim ve Test Senaryoları

	Eğitim Verisi	Test Verisi	Özellikler	
Senaryo 1			 Yorumlar  Yorumlar ve Oluşturulan Özellikler	  
Senaryo 2			 Yorumlar  Yorumlar ve Oluşturulan Özellikler	  
Senaryo 3			 Yorumlar  Yorumlar ve Oluşturulan Özellikler	  
Senaryo 4			 Oluşturulan Özellikler	  

Şekil 3.18 Kullanılan test verisi ve denenen mimarilere göre farklı senaryolar

Şekil 3.18’de Derin Öğrenme ile İngilizce dil grubu tahmininde uygulanan ve eğitimde kullanılan veriler ve testte kullanılan kurslar, ve denenen mimarilere göre değişiklik gösteren dört farklı senaryoya yer verilmiştir. Senaryolarda kullanılan modellerin yapısı eğitimde kullanılan verilere göre değişmektedir. İlk üç senaryoda veri olarak “sadece yorumlar” veya “yorum metinleri ve oluşturulan standart özelliklere” sahip veriler kullanılmıştır. Dördüncü senaryoda ise veri olarak “sadece oluşturulan standart özellikler” kullanılmıştır. Tüm senaryolarda eğitim için "Understanding Language: Learning and Teaching" MOOC serisinin ilk yedi iterasyonuna ait veriler (Derin Öğrenme daha büyük veri kümeleriyle daha iyi performans gösterdiği için)

kullanılmıştır. Aynı zamanda, ilk üç senaryo, iki ayrı model (eğitim için kullanılan verilere göre) ve beş farklı mimari (Konvolüsyon, LSTM, İki Yönlü LSTM, GRU ve Konvolüsyon+LSTM) denenerek oluşturulmuştur. Bununla birlikte, dördüncü senaryoda eğitim için "sadece oluşturulan standart özellikler" kullanıldığı için tek bir model ve mimari yapısı kullanılmıştır.

Derin öğrenme modellerinden yorum odaklı olarak elde edilen tahmin sonuçları kullanıcı odaklı hale getirilerek (İngilizce dil grubu özelliğine göre) hem diğer makine öğrenme yöntemleri ile karşılaştırılabilir hem de daha yüksek başarılı elde edilir hale getirilmiştir. Sonuçların kullanıcı odaklı hale getirilmesi adımları:

1. Model uygulandığında elde edilen üç değer içeren vektör sonucu, "0", "1", "2" değerlerinden birini alacak şekilde (0='Birincil ve resmi dil İngilizce', 1='Sadece resmi dil İngilizce', 2= 'İkincil dil olarak İngilizce') dönüştürülmüştür.
2. Yorum yapan kullanıcı yorumların sayısı n olmak üzere, en az $\lceil n/2 \rceil$ kadar yorum aynı İngilizce dil grubunda ise o katılımcı doğru gruplandırılmış, değil ise yanlış gruplandırılmış olarak kabul edilmiştir.

Şekil 3.18'de yer alan birinci senaryoda, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda yorum yapan katılımcılardan rastgele seçim ile % 80'i eğitim verisi ve % 20'si de test verisi olarak kullanılmıştır.

Şekil 3.18'de yer verilen ikinci senaryoda, test verisi olarak eğitimde kullanılan MOOC serisinin sekizinci iterasyonu olan "Understanding Language 8 : Learning and Teaching" kursunun yorum yapan katılımcılarına ait veriler kullanılmıştır.

Şekil 3.18'de yer verilen üçüncü senaryoda, Oluşturulan Derin Öğrenme modellerinin, katılımcıların İngilizce yazma yeterliliğinde daha belirgin farklılıklar ve değişkenliklere sahip olduğu kurslarda daha iyi sonuç vereceği düşünüldüğü için, farklı kurslardaki İngilizce dil grubu tahmini başarılarını anlamak için, test verisi olarak eğitimde kullanılan MOOC serisinden farklı türde iki kurs kullanılmıştır. Test için kullanılan kurslardan ilki, altı haftalık bir fen bilimleri MOOC'u olan "Web Science 1", ikincisi ise dört haftalık bir doğa bilimleri MOOC'u olan "Exploring Our Oceans 4"dur.

Şekil 3.18'de yer verilen dördüncü senaryoda ilk üç senaryodan farklı olarak hem eğitimde hem de testte "sadece oluşturulan standart özellikler" kullanılmıştır. Dördüncü senaryoda, "sadece oluşturulan standart özellikler" üzerinde ilk üç senaryoda denenen yöntemlerin tamamı (aynı MOOC serisini % 80'e % 20

ayırarak eğitim ve test, aynı MOOC serisinin farklı iterasyonu üzerinde test, farklı türden MOOC'lar üzerinde test) denenmiştir. "Sadece oluşturulan standart özellikler" kullanıldığında, kurs performans tahmininde en yüksek başarı elde edildiği gözlemlendiği için, kullanılan tüm İngilizce dil grupları bilinen katılımcılara ait veriden, her bir İngilizce dil grubuna ait veriler (İngilizce ana ve resmi dil, İngilizce sadece resmi dil ve İngilizce ikincil dil) elde edilerek, katılımcıların İngilizce dil gruplarının tahmin başarılarında değişim olup olmadığı gözlemlenmek istenmiştir.

3.5.1 Birinci Senaryo: Aynı Kurs (MOOC) Serisi İle Eğitim ve Test

Tablo 3.26'da yer alan sonuçlara göre, aynı kurs serisinin % 20 lik bölümü test verisi olarak kullanıldığında "sadece yorumlar" ile İngilizce dil grubu tahmininde Konvolüsyon mimarisi kullanıldığında en yüksek İngilizce dil grubu tahmini doğruluğu gerçekleşirken, "yorum metinleri ve oluşturulan standart özelliklerle" İngilizce dil grubu tahmininde Konvolüsyon ve LSTM hibrit mimarisi ile en yüksek İngilizce dil grubu tahmini doğruluğu elde edilmiştir.

Tablo 3.26 Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri (birinci senaryo)

	Sadece Yorumlarla (%)	Oluşturulan Standart Özellikler ve Yorumlarla (%)
Konvolüsyon	82,01	76,92
LSTM	73,35	69,68
İki Yönlü LSTM	72,27	70,64
GRU	70,14	68,12
Konvolüsyon ve LSTM	77,88	79,44

3.5.2 İkinci Senaryo: Aynı Kurs (MOOC) Serisinin Farklı İterasyonlarının Eğitim ve Test Verileri Olarak Kullanımı

Tablo 3.27'de yer alan sonuçlara göre, aynı kurs serisinin sekizinci sürümü test verisi olarak kullanıldığında, hem "sadece yorumlar" ile İngilizce dil grubu tahmininde, hem de "yorum metinleri ve oluşturulan standart özelliklerle" İngilizce dil grubu tahmininde GRU mimarisi kullanıldığında en yüksek doğrulukla İngilizce dil grubu tahmini gerçekleştirilmiştir.

Tablo 3.27 Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri (ikinci senaryo)

	Sadece Yorumlarla (%)	Oluşturulan Standart Özellikler ve Yorumlarla (%)
Konvolüsyon	54,52	53,62
LSTM	56,56	56,79
İki Yönlü LSTM	57,01	57,92
GRU	59,73	59,95
Konvolüsyon ve LSTM	54,98	50,23

3.5.3 Üçüncü Senaryo: Farklı Türden Kursların (MOOC'lar) Test Verisi Olarak Kullanımı

Üçüncü senaryoda, test için kullanılan MOOC'lardan doğa bilimleri MOOC'u olan "Exploring Our Oceans 4" MOOC'u, "Understanding Language : Learning and Teaching" kurs serisi MOOC'ları gibi dört haftalık bir MOOC iken "Web Science 1" MOOC'u altı haftalık bir MOOC'dur. Bu farklılıktan ötürü "Web Science 1" MOOC'unun ilk dört haftasına ait aktivite kullanım verileri İngilizce dil grubu tahmininde kullanılmıştır.

Tablo 3.28 Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri - "Exploring Our Oceans 4" kursu (üçüncü senaryo)

	Sadece Yorumlarla (%)	Oluşturulan Standart Özellikler ve Yorumlarla (%)
Konvolüsyon	47,12	49,04
LSTM	41,35	45,19
İki Yönlü LSTM	42,31	48,08
GRU	40,38	37,5
Konvolüsyon ve LSTM	47,12	44,23

Tablo 3.28'de yer alan sonuçlara göre, "Exploring Our Oceans 4" kursu test verisi olarak kullanıldığında, “sadece yorumlar” ile İngilizce dil grubu tahmininde Konvolüsyon mimarisi ile Konvolüsyon ve LSTM hibrit mimarileri kullanıldığında en yüksek doğrulukla İngilizce dil grubu tahmini gerçekleştirilirken, “yorum metinleri ve oluşturulan standart özelliklerle” İngilizce dil grubu tahmininde Konvolüsyon

mimarisi ile en yüksek İngilizce dil grubu tahmini doğruluğu elde edilmiştir.

Tablo 3.29 Farklı Derin Öğrenme modelleri ile İngilizce dil grubu tahmini doğruluk yüzdeleri - "Web Science 1" kursu (üçüncü senaryo)

	Sadece Yorumlarla (%)	Oluşturulan Standart Özellikler ve Yorumlarla (%)
Konvolüsyon	53,98	54,67
LSTM	42,91	43,94
İki Yönlü LSTM	42,56	46,71
GRU	44,64	40,14
Konvolüsyon ve LSTM	39,79	40,14

Tablo 3.29’da yer alan sonuçlara göre, "Web Science 1" kursu test verisi olarak kullanıldığında, hem “sadece yorumlar” ile hem de “yorum metinleri ve oluşturulan standart özelliklerle” İngilizce dil grubu tahmininde Konvolüsyon mimarisi ile en yüksek İngilizce dil grubu tahmini doğruluğu elde edilmiştir. “Sadece yorumlar” ile İngilizce dil grubu tahmininde üçüncü senaryoda, denenen doğa bilimler kursu olan "Exploring our Oceans 4" kursunda GRU mimarisi, fen bilimleri kurslarından olan "Web Science 1" kursunda Konvolüsyon ve LSTM hibrit mimarisi denenen mimariler içinde en kötü İngilizce dil grubu tahmini sonuçlarını vermişlerdir. “Yorum ve oluşturulan standart özelliklerle” İngilizce dil grubu tahmininde üçüncü senaryoda denenen doğa bilimleri kursu olan "Exploring our Oceans 4" kursunda GRU mimarisi, fen bilimleri kurslarından olan "Web Science 1" kursunda GRU mimarisi ile Konvolüsyon ve LSTM hibrit mimarisi denenen mimariler içinde en kötü İngilizce dil grubu tahmini sonucunu vermiştir.

3.5.4 Dördüncü Senaryo: Sadece Oluşturulan Standart Özelliklerin Kullanımı

Tablo 3.30, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi sürümünde katılımcıların "sadece oluşturulan standart özellikleri" ile Derin Öğrenme modelleri eğitilerek, farklı test verileri üzerinde testlerin gerçekleştirildiği üç ayrı duruma yer verilmiştir. İlk durumda aynı kurs verilerinin % 80’i eğitim, % 20’si test olarak ayrılarak, ikinci durumda aynı kurs serisinin tamamı eğitim verisi ve sekizinci sürümü test verisi olarak ardından da üçüncü durumda aynı kurs serisinin tamamı eğitim biri fen bilimleri kursu (Web Science 1) ve diğeri doğa bilimler kursu (Exploring Our Oceans 4) olan iki farklı kurs da (ayrı ayrı) test verileri olarak denemesi ile elde edilen İngilizce dil grubu tahmin başarı sonuçlarını göstermektedir.

Tablo 3.30 Oluşturulan standart özellikler kullanılarak Derin Öğrenme ile İngilizce dil grubu tahmini doğruluk yüzdeleri (dördüncü senaryo)

	Oluşturulan Standart Özelliklerle (%)
Aynı MOOC Serisi Üzerinde Eğitim ve Test	32,60
Eğitimde Kullanılan MOOC Serisinin Farklı İterasyonu Üzerinde Test	27,99
Eğitimde Kullanılan MOOC Serisinden Farklı Türde bir Üzerinde Test - "Exploring Our Oceans 4"	30,61
Eğitimde Kullanılan MOOC Serisinden Farklı Türde bir Üzerinde Test - "Web Science 1"	30,12

Tablo 3.30'daki sonuçlara göre veri olarak sadece standart özelliklerin kullanılması İngilizce dil grubu tahmini için yetersizdir.

3.6 Sonuçlar

Bu bölümde, FutureLearn platformundan "Understanding Language : Learning and Teaching" MOOC serisinin ilk sekiz iterasyonu, "Exploring Our Oceans 4" kursu ve "Web Science 1" MOOC'larında kurs öncesi anketlerdeki ülke bilgileri ve kendi oluşturduğumuz düzenli ifade kalıpları ile İngilizce dil grubu tespiti ve aynı zamanda farklı Derin Öğrenme modelleri (denenen beş farklı mimari ile) ile İngilizce dil grubu tahminleri gerçekleştirilmiştir. Elde edilen sonuçların özeti aşağıdaki gibi listelenebilir.

- Sadece kurs öncesi anket bilgileri ile katılımcıların İngilizce dil grubu tespitinde, resmi dil olarak İngilizce konuşanları “birincil ve resmi dil olarak İngilizce” ve “sadece resmi dil olarak İngilizce” olarak iki alt gruba ayırmak, gruplar arasındaki farklı daha belirgin hale getirmiştir.
- Düzenli ifadelerle katılımcıların İngilizce dil gruplarını, yorumlarındaki dil, ülke, şehir ve millet gibi dile yönelik bilgileri kullanarak otomatik olarak tespit edebilen bir model üretilmiştir. Düzenli ifadelerin kurs öncesi anketlerle birlikte İngilizce dil grubu tahmininde kullanılmasının, İngilizce dil grubu tespit edilen katılımcıların sayısını ve tespitin doğruluğunu artırabileceği görülmüştür.
- Oluşturulan düzenli ifadelerle İngilizce dil grubu tespiti yöntemi "Understanding Language: Learning and Teaching" MOOC serisinde, farklı türden MOOC'lara göre daha iyi sonuç vermiştir.

- Farklı İngilizce dil gruplarındaki katılımcıların doğru tespiti, davranışlarını ve performanslarını analiz etme sonuçlarının güvenilirliği açısından önemlidir. Bu sebeple ilerleyen bölümlerdeki analizlerde oluşturulan düzenli ifadeler ile tespit edilen İngilizce dil gruplarının verileri kullanılmıştır.
- Farklı Derin Öğrenme mimarileri ile İngilizce dil grubu tahmini gerçekleyebilecek modeller önerilerek, oluşturulan düzenli ifadelerle dil grubu tahmini modeline daha genel bir alternatif sunulmuştur. Modelde kullanılan farklı Derin Öğrenme mimarilerinin üç farklı örnekleme yöntemi başarıları karşılaştırılmış (aşırı örnekleme, ortalama örnekleme, aşağı örnekleme) ve genellikle aşırı örnekleme (oversampling) ile en yüksek doğruluk değerleri elde edilmiştir.
- Derin Öğrenme ile İngilizce dil grubu tahmini başarısının seçilen verinin özelliklerine, kullanılan Derin Öğrenme modeli ve katmanlara, katmanların optimizasyon parametre değerlerine ve veri kümesi büyüklükleri gibi bir dizi faktöre bağlı olarak değiştiği gözlemlenmiştir.
- Derin Öğrenme modelleri (beş farklı mimari denenerek) ile İngilizce dil grubu tahmininde, “sadece yorumlar” veri kaynağı olarak kullanıldığı durumda, tüm senaryolarda genellikle Konvolüsyon mimarisi ile en iyi sonuç alınmıştır. Veri kaynağı olarak “yorumlar ve oluşturulan standart özellikler” kullanıldığında, "Understanding Language: Learning and Teaching" MOOC serisinin hem eğitim hem de testte yer aldığı senaryolardan ilkinde Konvolüsyon ve LSTM mimarilerinin birlikte kullanıldığı hibrit yapı ile başarılı tahmin sonucu, ikinci senaryoda GRU mimarisi ile en başarılı tahmin sonucu elde edilirken, farklı kursların test verisi olarak kullanıldığı üçüncü senaryoda Konvolüsyon mimarisi ile en başarılı tahmin sonucu elde edilmiştir.
- Derin öğrenme ile İngilizce dil grubu tahmini senaryolarında, "Understanding Language: Learning and Teaching" MOOC serisinin ilk yedi iterasyonunun % 80 eğitim ve % 20 test olarak ayrıldığı ilk senaryoda en yüksek tahmin doğruluğu elde edilirken, aynı kurs serisinin ilk yedi iterasyonu eğitim verisi ve sekizinci iterasyonu test verisi olarak kullanıldığında en iyi ikinci tahmin doğruluk başarısı elde edilmiştir. Farklı türden kursların test verisi olarak kullanıldığı üçüncü senaryoda tahmin başarıları düşmüştür.
- Veri kaynağına bağlı olarak en iyi İngilizce dil grubu tahmini başarısı "sadece yorumlar" veya "yorumlar ve oluşturulan standart özellikler" kullanıldığı iki durum arasında kullanılan mimariye ve senaryoya bağlı olarak değişmektedir. Bununla birlikte, bu iki durumda da sadece standart özelliklerle gerçekleştirilen tahmin başarısına göre daha yüksek tahmin başarısı elde edilmiştir.

İngilizce Dil Gruplarının Kullanım Analizleri

Bu bölümde, İngilizce dil gruplarının davranışları analiz edilerek katılımcılara MOOC'larda çalışma stratejileri önermek için gerekli olan bilgilerin sağlanması, İngilizce dil bilgisinin önemli olduğu ve eksikliği sebebi ile zorlanan aktiviteleri belirlemek hedeflenmiştir.

Bu sebeple, farklı İngilizce dil gruplarındaki katılımcıların tartışma forumlarını nasıl kullandıkları veya kurs adımlarını ne sıklıkta tamamladıkları, diğer kullanıcılarla takip etkileşimlerine girme durumları gibi kurs etkileşimleri ile ilgili istatistiksel analiz sonuçlarına yer verilmiştir. Bölümde, ilk olarak analizlerde kullanılabilecek özelliklerin verilerden nasıl oluşturulduğuna, ardından da istatistiksel kullanım analizi sonuçlarına yer verilmiştir.

4.1 Veri Ön işleme İle Standart Özelliklerin Çıkarımı

FutureLearn platformu tarafından sağlanan verilerden faydalanarak analizlerde kullanılabilecek özellikler çıkarılmış ve Tablo 4.1'de bu özellikler ve özellikleri oluşturmak için kaynak olarak kullanılan verilere yer verilmiştir.

Tablo 4.1 Veri dosyalarına göre oluşturulan özelliklerin başlıcaları

Enrolments	Step Activity	Comments	Followings
kullanıcı id	başlanan adım sayısı	yorum sayısı	kurs verilirken takip etme sayısı
İngilizce dil grubu	bitirilen adım sayısı	ana yorum sayısı	kurs verilirken takip edilme sayısı
çalışılan sektör	tamamlanmadan bırakılan adım sayısı	cevap yorum sayısı	kurs içi takip etme sayısı
cinsiyet	adım tamamlama yüzdesi	beğeni alan yorum sayısı	kurs içi takip edilme sayısı
çalışma durumu	tamamlanmadan bırakılan adım yüzdesi	yorum uzunluk ortalaması	
kullanıcı rolü	tamamlanan ilk hafta adımı sayısı	ana yorum uzunluk ortalaması	
yaş aralığı	tamamlanan ikinci hafta adımı sayısı	cevap yorum uzunluk ortalaması	
en yüksek eğitim seviyesi	tamamlanan üçüncü hafta adımı sayısı	yorum beğeni sayısı ortalaması	
kurs performansı	tamamlanan dördüncü hafta adımı sayısı	ana yorum beğeni sayısı ortalaması	
		cevap yorum beğeni sayısı ortalaması	
		yorum beğeni sayısı toplamı	
		ana yorum beğeni sayısı toplamı	
		cevap yorum beğeni sayısı toplamı	

Enrolments verisinden oluşturulan özellikler ve açıklamaları

Kullanıcı id: Kullanıcı id özelliği, *Enrolments* verisinde her bir kullanıcı için yer alan 'learner_id' özelliğinin karşılığı olarak aynı değerleri içerecek şekilde kullanılmıştır.

İngilizce dil grubu: İngilizce dil grubu özelliği değerleri, tezde bölüm 3'de yer verilen kurs öncesi anket ve düzenli ifade yöntemleri ile İngilizce dil grubu tespit yöntemi uygulanarak elde edilen İngilizce dil gruplarına göre oluşturulmuştur. Bu yöntemle göre, İngilizce dil grubu bilinmeyen kullanıcılar için de 'bilinmiyor' değerini almaktadır.

Çalışılan sektör: *Enrolments* verisindeki 'employment_area' özelliğinden faydalanan oluşturulmuştur. 'Employment_area' özelliğinin "Unknown" değerini alması durumunda 'bilinmiyor'; "teaching_and_education" değerini alması

durumunda “eğitim”; bunun dışında aldığı tüm değerleri için de ‘eğitim dışı’ değerini almaktadır

Çalışma durumu: *Enrolments* verisinde yer alan ‘employment_status’ özelliğinden faydalanarak oluşturulmaktadır.

‘Employment_status’ özelliğinin "full_time_student" değerini alması durumunda ‘öğrenci’; "not_working" değerini alması durumunda ‘çalışmayan’; "self_employed" veya "working_full_time" değerlerinden birini alma durumunda ‘çalışan’; "working_part_time" değerini alması durumunda da ‘yarı zamanlı çalışan’ değerlerini almaktadır.

Kullanıcı rolü: Kullanıcı rolü özelliği, kursa kayıtlı kullanıcıların kurstaki rolünü göstermektedir. *Enrolments* verisindeki ‘role’ özelliğinin karşılığı olarak aynı değerleri alacak şekilde kullanılmaktadır ("Unknown" değerinin karşılığında ‘bilinmiyor’ değerini almaktadır).

Cinsiyet: Cinsiyet özelliği, *Enrolments* verisindeki ‘gender’ özelliğinin karşılığı olarak aynı değerleri alacak şekilde kullanılmıştır ("Unknown" değerinin karşılığında ‘bilinmiyor’ değerini almaktadır).

Yaş Aralığı: Yaş Aralığı özelliği *Enrolments* verisindeki ‘age_range’ özelliğinden kursa kayıtlı kullanıcıların yaş aralıklarını göstermek üzere aynı değerleri alacak şekilde oluşturulmuştur ("Unknown" değerinin karşılığında ‘bilinmiyor’ değerini almaktadır).

En yüksek eğitim seviyesi: En yüksek eğitim seviyesi özelliği, *Enrolments* verisindeki ‘highest_education_level’ özelliğinden faydalanarak oluşturulmaktadır.

‘Highest_education_level’ özelliğinin "less_than_secondary", "secondary" veya "tertiary" değerleri için ‘eğitim seviyesi düşük’; "apprenticeship", "professional", "university_degree", "university_doctorate" veya "university_masters" değerleri için ‘eğitim seviyesi yüksek’ değerlerini almaktadır.

Kurs Performansı: Katılımcının kurs sonu performansını göstermek üzere *Enrolments* verisindeki ‘unenrolled_at’, ‘fully_participated_at’ ve ‘purchased_statement_at’ özelliklerinden faydalanarak oluşturulmaktadır.

- Kurstaki tüm adımların en az % 50’sini tamamlayanlar ile sertifika satın alanlar (kurs adımlarının en az % 90’ını tamamlama şartı vardır) ‘tamamlayıcılar’ değerini almaktadırlar.
- Hiçbir kurs adımına başlamayan veya başladığı kurs adımlarından hiçbirini

tamamlamayan katılımcılar 'bırakanlar' değerini almaktadırlar.

- Başladığı kurs adımlarından en az birini tamamlamakla birlikte, tamamladığı kurs adımı sayısı, kurstaki toplam adım sayısının yarısından az olan katılımcılar 'yavaş tempolu katılımcılar' değerini almaktadırlar.

Step Activity verisinden oluşturulan özellikler ve açıklamaları

Başlanan adım sayısı: Başlanan adım sayısı özelliği *Step Activity* verisinde yer alan kullanıcıların kurs aktivitelerinin başladıkları tarih özellikleri kullanılarak ve her bir kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id'lere göre hesaplanmıştır.

Bitirilen adım sayısı: *Step Activity* verisindeki kullanıcıların kurs aktivitelerini bitirdikleri tarihleri gösteren 'last_completed_at' özelliği kullanılarak, özelliğin değerinin boş olmadığı her bir kayıt için kullanıcının *Enrolments* verisinden karşılık geldiği 'learner_id'ye göre bitirilen adım sayısı hesaplanmıştır.

Tamamlanmadan bırakılan adım sayısı: *Step Activity* verisindeki kullanıcı aktivitelerinin başladıkları ve bitirdikleri zaman özellikleri kullanılarak, bitirdikleri zaman özelliği boş olan aktiviteleri gerçekleştiren kullanıcının *Enrolments* verisinde karşılık geldiği 'learner_id' üzerinden tamamlanmadan bırakılan adım sayısı hesaplanmıştır.

Kurs adımı tamamlama yüzdesi: *Step Activity* verisindeki kullanıcıların aktivitelere başladıkları ve bitirdikleri zaman özelliklerinden her bir kullanıcının *Enrolments* verisinde karşılık geldiği 'learner_id'lere göre başladıkları aktivitelere bitirdikleri zaman boş olmayan aktivite sayılarının, toplam başlanan aktivite sayıları içindeki yüzdeleri göstermektedir.

Tamamlanmadan bırakılan kurs adımı yüzdesi: *Step Activity* verisinde kullanıcıların kurs aktivitelerine başladıkları ve bitirdikleri tarihlere göre, başladıkları aktivitelere bitirdikleri zaman değerleri boş olanların toplam başladıkları aktivite sayıları içindeki yüzdeleri (*Enrolments* verisinde kullanıcıya karşılık gelen 'learner_id' üzerinden) göstermektedir.

Tamamlanan ilk hafta kurs adımı sayısı: Kullanıcıların kursun ilk haftasına ait kaç aktiviteye başladıklarını göstermektedir. *Step Activity* verisindeki 'week_number' özelliğinin değeri "1" olan ve bitirme zamanı özelliğinin değeri boş olmayan aktiviteler için katılımcıların *Enrolments* verisindeki 'learner_id'leri üzerinden hesaplanmaktadır.

Tamamlanan ikinci hafta kurs adımı sayısı: Kullanıcıların kursun ikinci haftasına ait olan kaç aktiviteye başladıklarını göstermektedir. *Step Activity* verisinde

'week_number' özelliğinin değeri "2" olan ve bitirme zamanı özelliğinin değeri boş olmayan aktiviteler için katılımcıların *Enrolments* verisindeki 'learner_id'leri üzerinden hesaplanmaktadır.

Tamamlanan üçüncü hafta kurs adımı sayısı: Kullanıcıların kursun üçüncü haftasına ait olan kaç aktiviteye başladıklarını göstermektedir. *Step Activity* verisinde 'week_number' özelliğinin değeri "3" olan ve bitirme zamanı özelliğinin değeri boş olmayan aktiviteler için katılımcıların *Enrolments* verisindeki 'learner_id'leri üzerinden hesaplanmaktadır.

Tamamlanan dördüncü hafta kurs adımı sayısı: Kullanıcıların kursun dördüncü haftasına ait olan kaç aktiviteye başladıklarını göstermektedir. *Step Activity* verisinde 'week_number' özelliğinin değeri "4" olan ve bitirme zamanı özelliğinin değeri boş olmayan aktiviteler için katılımcıların *Enrolments* verisindeki 'learner_id'leri üzerinden hesaplanmaktadır.

Comments Verisinden Oluşturulan Özellikler ve Açıklamaları

Yorum sayısı: *Comments* verisinde her bir kullanıcı için yaptığı yorumlardaki 'author_id' özelliğinden faydalanarak 'author_id'lerin *Enrolments* verisinde karşılık geldiği 'learner_id'lere göre kullanıcıların yaptıkları yorum sayıları hesaplanmaktadır.

Ana yorum sayısı: *Comments* verisinde kullanıcının yaptığı yorumlardan başka bir yoruma cevap olmayanlar ('parent_id' özelliği boş değer alanlar) için kullanıcının *Enrolments* verisinde karşılık geldiği 'learner_id'lere göre kullanıcıların yaptıkları ana yorum sayıları hesaplanmaktadır.

Cevap yorum sayısı: *Comments* verisinde kullanıcının yaptığı yorumlardan başka bir yoruma cevap olanlar ('parent_id' özelliği boş değer almayanlar) için kullanıcının *Enrolments* verisinde karşılık geldiği 'learner_id'lere göre cevap yorum sayısı hesaplanmaktadır.

Begeni alan yorum sayısı: *Comments* verisinde her bir kullanıcının yaptığı yorumların aldığı beğeni sayılarını gösteren 'likes' özelliği bulunmaktadır. Bu özelliğin aldığı değer "0"dan farklı olduğu yorumlar için kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id' özelliği üzerinden beğeni alan yorum sayısı özelliğinin değeri hesaplanmaktadır.

Yorum uzunluk ortalaması: *Comments* verisinde kullanıcının yaptığı yorumların metinlerini içeren 'text' özelliğinden karakter sayıları, *length* fonksiyonu ile hesaplanıp *mean* fonksiyonu ile de bunların ortalaması alınarak, kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id' özelliği üzerinden yorum uzunluk ortalaması özelliğinin

değeri hesaplanmaktadır.

Ana yorum uzunluk ortalaması: *Comments* verisinde kullanıcının yaptığı ana yorumların ('parent_id' özelliği değeri boş olan) metinlerini içeren 'text' özelliğinden karakter sayıları *length* fonksiyonu ile hesaplanıp *mean* fonksiyonu ile bunların ortalaması alınarak, kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id' özelliği üzerinden ana yorum uzunluk ortalaması özelliğinin değeri hesaplanmaktadır.

Cevap yorum uzunluk ortalaması: *Comments* verisinde kullanıcının yaptığı cevap yorum ('parent_id' özelliği değeri boş olmayan) metnini içeren 'text' özelliğinin karakter sayısı *length* fonksiyonu ile hesaplanıp *mean* fonksiyonu ile de bunların ortalaması alınarak, kullanıcıya *Enrolments* verisinde karşılık geldiği 'learner_id' özelliği üzerinden cevap yorum uzunluk ortalaması özelliğinin değeri hesaplanmaktadır.

Yorum beğeni sayısı ortalaması: *Comments* verisinde kullanıcının yaptığı yorumların aldığı beğeni sayısını gösteren 'likes' özelliğinin aldığı değerlerin *mean* fonksiyonu ile ortalaması alınarak, kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id' özelliği üzerinden yorum beğeni sayısı ortalaması özelliğinin değeri hesaplanmaktadır.

Ana yorum beğeni sayısı ortalaması: *Comments* verisinde kullanıcının yaptığı ana yorumların ('parent_id' özelliği değeri boş olan) aldığı beğeni sayılarını gösteren 'likes' özelliğinin aldığı değerlerin *mean* fonksiyonu ile ortalaması alınarak *Enrolments* verisinde kullanıcıya karşılık gelen 'learner_id' özelliği üzerinden ana yorum beğeni sayısı ortalaması özelliğinin değeri hesaplanmaktadır.

Cevap yorum beğeni sayısı ortalaması: *Comments* verisinde kullanıcının yaptığı ana yorumların ('parent_id' özelliği değeri boş olmayan) aldığı beğeni sayılarını gösteren 'likes' özelliğinin aldığı değerlerin *mean* fonksiyonu ile ortalaması alınarak kullanıcıya *Enrolments* verisinde karşılık geldiği 'learner_id' özelliği üzerinden cevap yorum beğeni sayısı ortalaması özelliğinin değeri hesaplanmaktadır.

Yorum beğeni sayısı toplamı: *Comments* verisinde kullanıcının yaptığı yorumların aldıkları beğeni sayılarını gösteren 'likes' özelliğinin aldığı değerler *sum* fonksiyonu ile toplanarak kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id' özelliği üzerinden yorum beğeni sayısı toplamı özelliğinin değeri hesaplanmaktadır.

Ana yorum beğeni sayısı toplamı: *Comments* verisinde her bir kullanıcının yaptıkları yorumların aldıkları beğeni sayılarını gösteren 'likes' özelliği vardır. Yorumun ana yorum olduğunu gösteren 'parent_id' özelliğinin boş değer alması durumları için *sum* fonksiyonu ile kullanıcıya *Enrolments* verisinde karşılık gelen 'learner_id'ler üzerinden

kullanıcıların ana yorum beğeni sayısı toplamı özelliğinin değeri hesaplanmaktadır. **Cevap yorum beğeni sayısı toplamı:** *Comments* verisinde her bir kullanıcının yaptıkları yorumların aldıkları beğeni sayılarını gösteren 'likes' özelliği vardır. Yorumun cevap yorum olduğunu gösteren 'parent_id' özelliğinin boş değer almaması durumları için *sum* fonksiyonu ile kullanıcılara *Enrolments* verisinde karşılık gelen 'learner_id'lere göre kullanıcıların cevap yorumları için toplam beğeni sayıları hesaplanarak cevap yorum beğeni sayısı toplamı özelliğine değer olarak atanmaktadır.

Followings verisinden oluşturulan özellikler ve açıklamaları:

Kurs verilirken takip etme sayısı: Kursun verildiği zamana ait katılımcıların takip etme sayılarını göstermektedir. Özellik, FutureLearn platformunda, Southampton Üniversitesi tarafından verilen tüm kurslardaki kullanıcılar arasında gerçekleştirilen takip etme ve edilme durumlarına ait bilgileri içeren verinin 'created_at' (takip etme işlemi zamanı) ve 'follower_user_id' (takip eden kullanıcı id) özellikleri ile kurs zamanı ve kurs kullanıcılarına ait olma kontrolleri ile filtrelenmesi ile elde edilmiştir.

Kurs verilirken takip edilme sayısı: Kursun verildiği zamanda katılımcıların takip edilme sayılarını gösteren özelliktir. FutureLearn platformundaki tüm kurslarda, kullanıcılar arasında gerçekleştirilen takip etme ve edilme durumlarına ait bilgileri içeren verinin 'created_at' (takip etme işlemi zamanı) ve 'followed_user_id' (takip edilen kullanıcı id) özellikleri ile kurs zamanı ve kurs kullanıcılarına ait olma kontrolleri ile filtrelenmesi ile elde edilmiştir.

Kurs içi takip etme sayısı: Kursun verildiği zamanda gerçekleştirilen ve kursun katılımcılarının yine aynı kursun katılımcılarını kurs zamanı içinde takip etme sayılarını gösteren özelliktir.

Kurs verilirken takip edilme sayısı: Kursun verildiği zamanda gerçekleştirilen ve kursun katılımcılarının yine aynı kursun katılımcıları tarafından kurs zamanı içinde takip edilme sayılarını gösteren özelliktir.

Oluşturduğumuz özelliklerden kullanım analizlerinde kullandıklarımıza Tablo 4.2'de yer verilmiştir.

Tablo 4.2 Veri dosyalarına göre oluşturularak analizlerde kullanılan özellikler

1-Step Activity	2-Comments	3-Followings
1. başlanan adım sayısı 2. tamamlanmadan bırakılan adım sayısı 3. tamamlanan ilk hafta adımı sayısı 4. tamamlanan ikinci hafta adımı sayısı 5. tamamlanan üçüncü hafta adımı sayısı 6. tamamlanan dördüncü hafta adımı sayısı	1. yorum sayısı 2. ana yorum sayısı 3. cevap yorum sayısı 4. yorum uzunluk ortalaması 5. ana yorum uzunluk ortalaması 6. cevap yorum uzunluk ortalaması 7. yorum beğeni sayısı ortalaması	1. kurs verilirken takip etme sayısı 2. kurs verilirken takip edilme sayısı

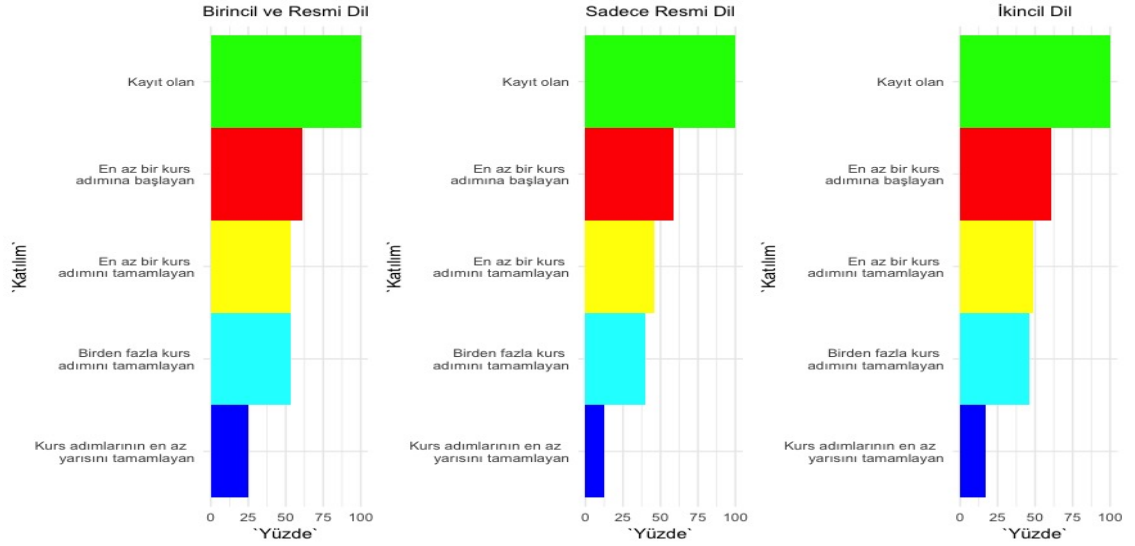
4.2 "Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri

"Understanding Language 4: Learning and Teaching" MOOC'u üzerinde gerçekleştirilen İngilizce dil gruplarının kullanım anaizleri, kullanılan veri kaynaklarına göre kurs adımlarındaki davranış analizleri, kurs tartışma forumlarındaki davranış analizleri ve takip etkileşimleri davranış analizleri olarak üç alt başlık altında incelenmiştir.

4.2.1 Kurs Adımlarındaki Davranış Analizleri

Farklı MOOC katılımcılarının kurslara katılım düzeyleri değişiklik göstermektedir. Bununla birlikte, kurs katılımcılarının art arda gelen haftalar boyunca toplam katılımlarını azaltma yüzdeleri, neredeyse her MOOC'da yaygın olarak gözlenmiştir. Clow [114], bu davranış biçimini Katılım Hunisi (*Funnel of Participation*) olarak tanımlamıştır. Tez çalışması kapsamında yapılan incelemelerde bu düşüş eğilimi doğrulanmıştır.

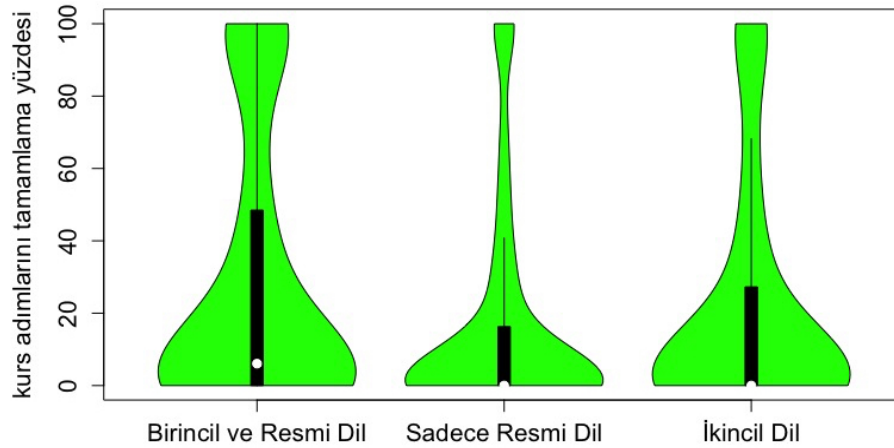
Şekil 4.1, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce dil gruplarındaki katılımcıların azalan devam etme yüzdelerini göstermektedir.



Şekil 4.1 İngilizce dil gruplarındaki kullanıcıların kursa katılımları - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.1'den, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce'yi birincil ve resmi dil olarak konuşan katılımcıların grubunda kurs adımlarına katılımın (birden fazla kurs adımına) ve kurs bitirme yüzdelerinin en yüksek olduğu anlaşılmaktadır.

Violin grafikleri olasılık yoğunluğu (density), ortanca değer (median) ve çeyrekler arası (interquartile) aralığı bilgilerini ortaya koyan ve farklı kategorilerden örneklerin dağılımını göstermede faydalı olan grafiklerdir. Şekil 4.2'deki Violin grafiği, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce dil gruplarındaki katılımcıların kurs sonundaki adımları tamamlama yüzdelerini göstermektedir (katılımcıların başladıkları kurs adımlarının tüm kurs adımları içindeki yüzdelerini).



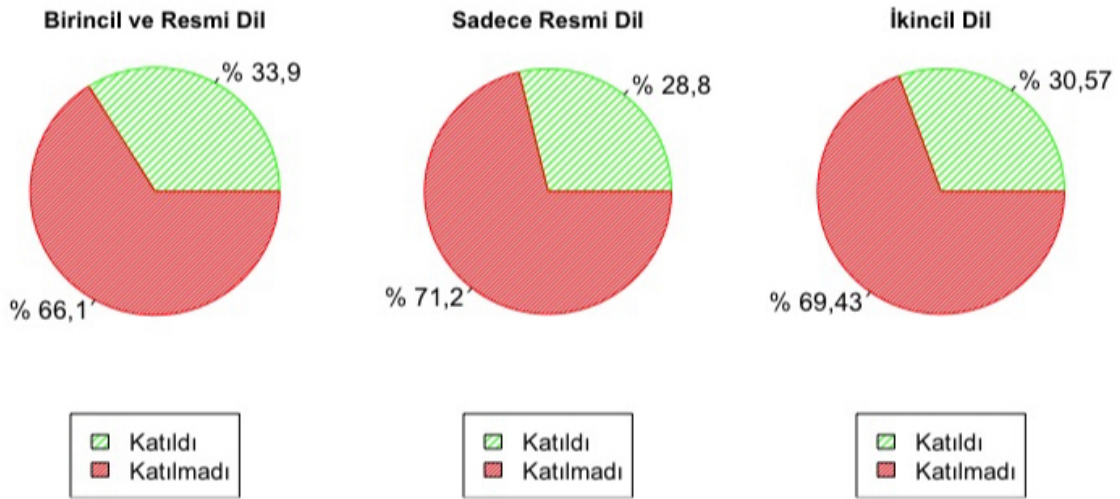
Şekil 4.2 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.2'ye göre, “birincil ve resmi dil olarak İngilizce” konuşan katılımcıların kurs adımlarının genelini tamamlama yüzdesi, diğer İngilizce dil gruplarındaki katılımcılara göre daha yüksektir.

4.2.2 Kurs Tartışma Forumlarındaki Davranış Analizleri

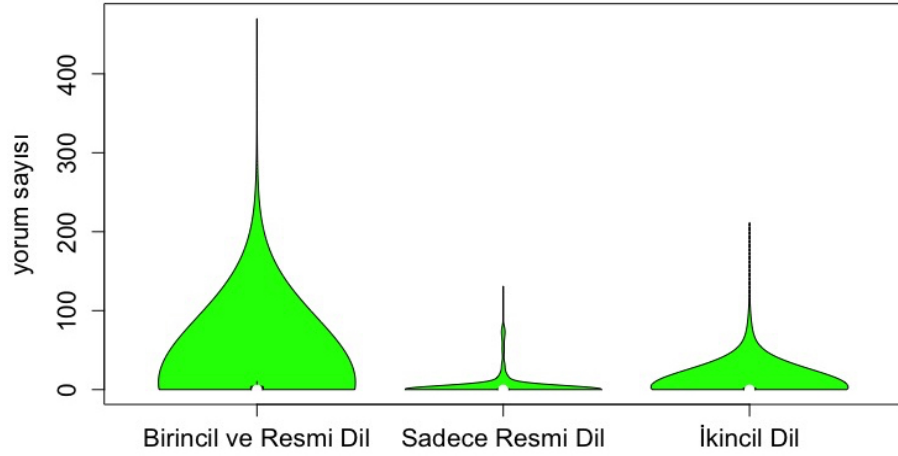
FutureLearn platformu MOOC’larında, katılımcılar kurs forumlarına ana yorum göndererek, yorumu yanıtlayarak veya ana yorumu/cevap yorumu beğenerek katkıda bulunabilirler. Bu eylemlerle gerçekleştirilen sosyal katılımlar, genel kurs katılımlarına göre daha azdır.

Şekil 4.3, "Understanding Language 4: Learning and Teaching" MOOC’unda İngilizce dil gruplarındaki katılımcıların tartışma forumlarında aktif (en az bir ana yorum veya cevap yorum yazan) veya pasif (ana veya cevap yorumu olmayan) olarak katılımlarının dağılımlarını göstermektedir. Şekil 4.3’e göre İngilizce dil gruplarının tartışma forumlarına katılım yüzdeleri arasında çok az fark vardır.



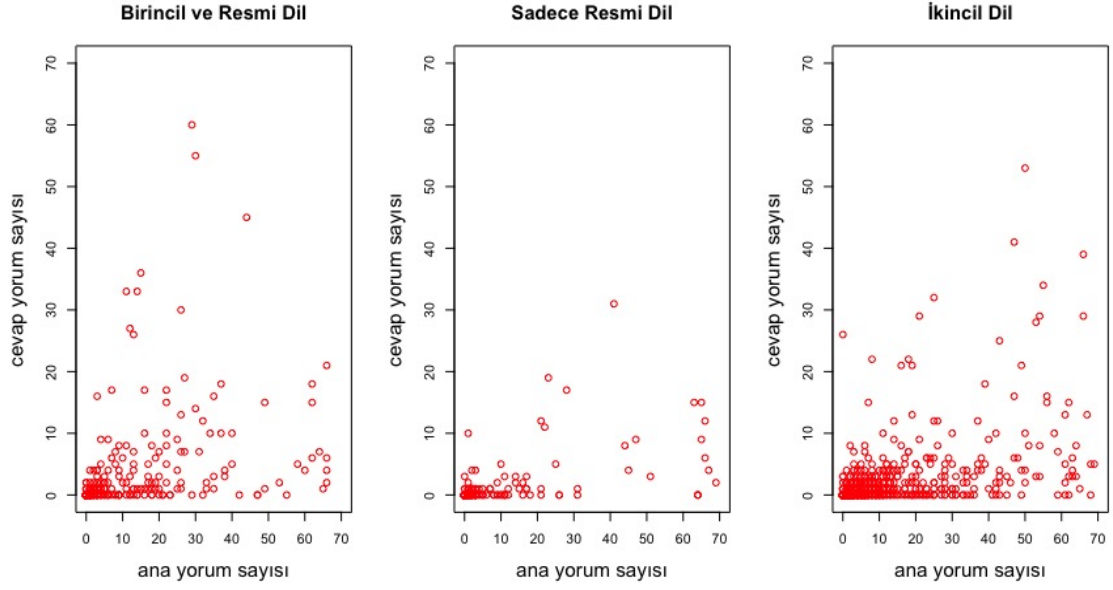
Şekil 4.3 İngilizce dil gruplarındaki katılımcıların tartışma forumlarına katılım yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.4, "Understanding Language 4: Learning and Teaching" MOOC’unda İngilizce dil gruplarında katılımcı başına düşen toplam yorum sayılarını (ana yorum veya cevap yorum) göstermektedir. Şekil 4.4’e göre, hem ana hem de resmi dili İngilizce olan katılımcılar tartışma forumlarında daha fazla sayıda yorum yapmışlardır. Sadece resmi dili İngilizce olanlar grubundaki katılımcılar ise genel olarak en az sayıda yorum yapmışlardır. İngilizce dil gruplarında, 400’e yakın sayıda yorum yapan (aykırı değerlere sahip) bazı katılımcılar olmakla birlikte, katılımcıların çoğunluğu herhangi bir yorum yapmamışlardır.



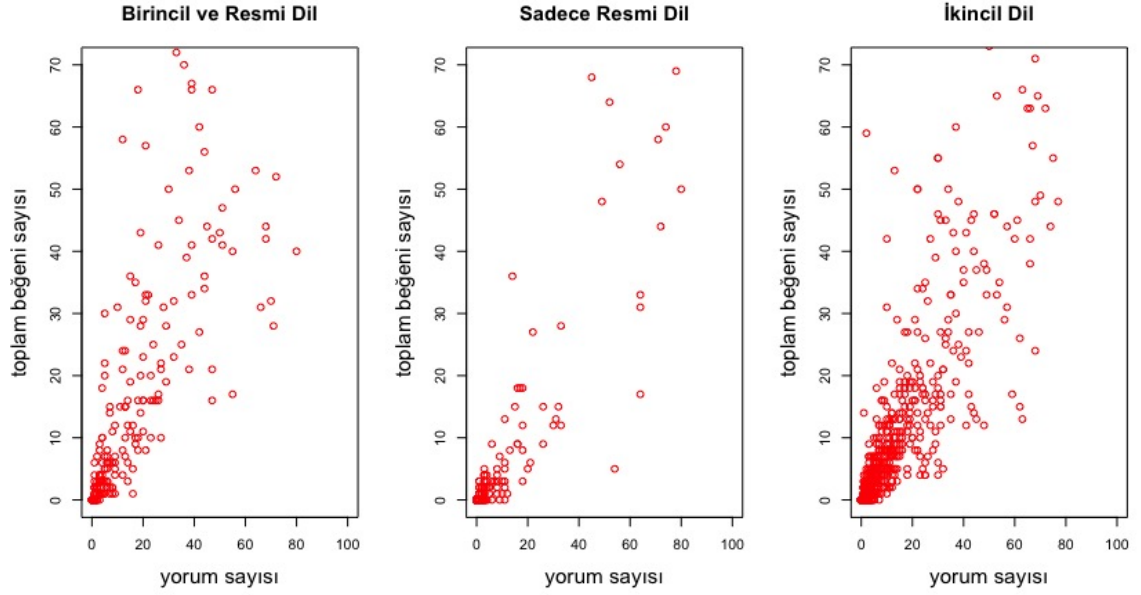
Şekil 4.4 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları
- "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.5, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların toplam sayıları ile ana yorumlarının aldıkları cevap yorumların toplam sayılarını göstermektedir. Şekil 4.5'te daha net bir gösterim için ana yorumların ve cevap yorumların sayıları 70'le sınırlandırılmıştır. (çok büyük değerlere sahip aykırı değerler içerildiğinde grafik çok küçük ve belirsiz olduğu için net bir gösterim için 70'ten fazla olan yorum sayılarına yer verilmemiştir). Şekil 4.5'e göre, İngilizce dil gruplarının toplam ana yorumların sayılarına göre aldıkları toplam cevap yorum sayıları çok benzer eğilimdedir. Katılımcılar tarafından kaç tane ana yorum yapıldığına bağlı olmaksızın, katılımcıların aldıkları toplam cevap yorum sayısı genellikle ondan az sayıdadır.



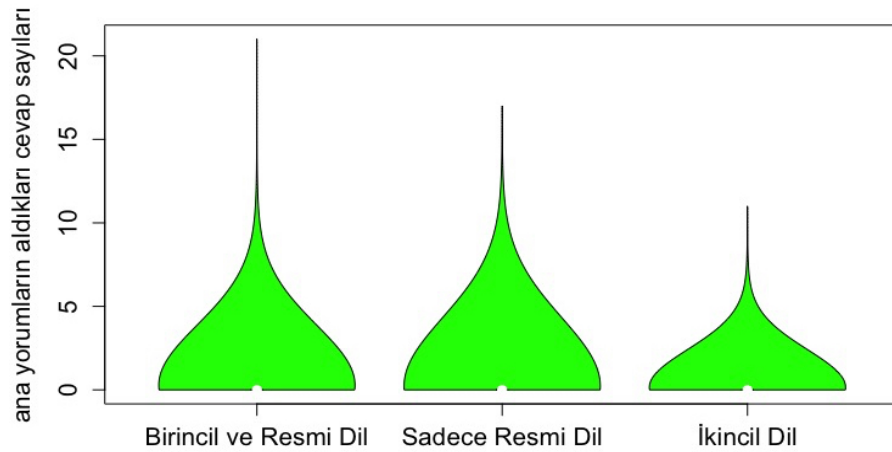
Şekil 4.5 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.6, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce dil gruplarındaki katılımcıların yaptıkları yorumların aldıkları toplam beğeni sayılarını göstermektedir. Şekil 4.6'da daha net bir gösterim için azami yorum ve beğeni sayıları 70'le sınırlandırılmıştır. Şekil 4.6'ya göre İngilizce dil gruplarındaki yorum başına düşen beğeni sayılarının eğilimi benzer olmakla birlikte ana ve resmi dil olarak İngilizce konuşan dil grubunda diğer İngilizce dil gruplarına göre biraz daha fazladır.



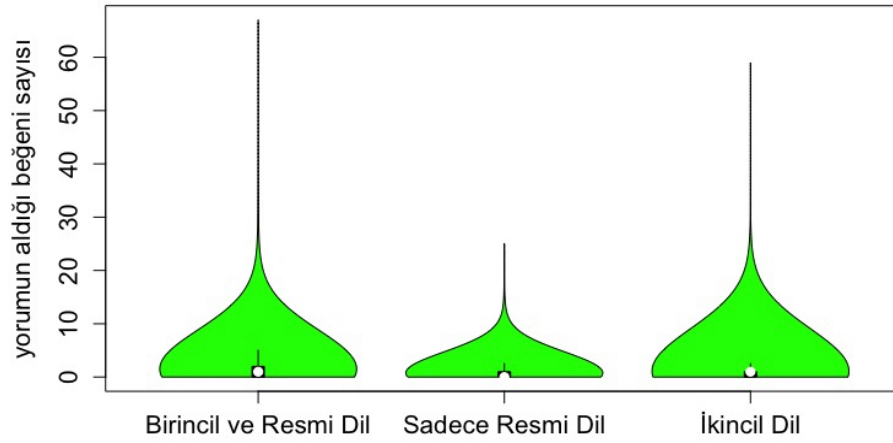
Şekil 4.6 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yaptıkları yorum ve aldıkları beğeni sayıları - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.7, "Understanding Language 4: Learning and Teaching" MOOC’unda İngilizce dil gruplarındaki katılımcılar tarafından yapılan ana yorumların aldıkları cevap yorum sayılarını göstermektedir. Şekil 4.7’ye göre “birincil ve resmi dil olarak İngilizce” konuşan katılımcıların yorumları, diğer İngilizce dil gruplarındaki katılımcıların yorumlarından daha fazla cevap almaya eğilimlidirler. Bununla birlikte İngilizce gruplarının hepsinde, katılımcılar tarafından yapılan yorumlar çoğunlukla herhangi bir cevap almamışlardır.



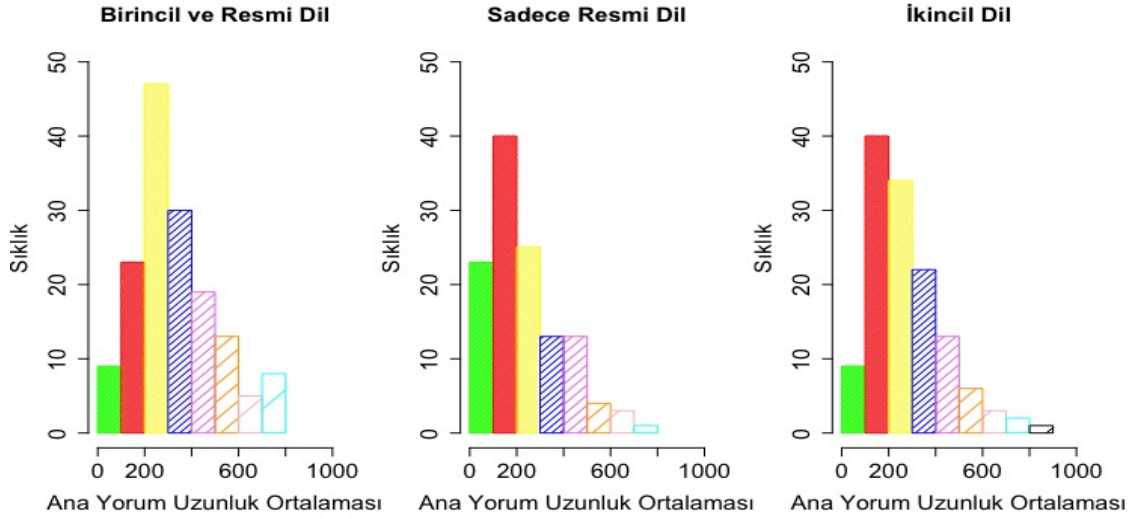
Şekil 4.7 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki ana yorumlarının aldıkları cevap yorum sayıları - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.8, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce dil gruplarındaki katılımcılar tarafından kurs forumlarında yapılan yorumların aldıkları beğeni sayılarını göstermektedir. İngilizce dil gruplarındaki katılımcıların çoğunun yaptıkları yorumlar, beğeni almamış veya 25'in altında bir beğeni almışsa da, birincil ve resmi dili İngilizce olan katılımcıların çok sayıda beğeni alan yorumları da vardır. İkincil dil olarak İngilizce konuşan katılımcıların bazılarının yorumları da çok sayıda beğeni almakla birlikte, birincil ve resmi dili İngilizce olan katılımcıların yorumlarının aldıkları beğeni miktarları biraz daha fazla olmuştur.



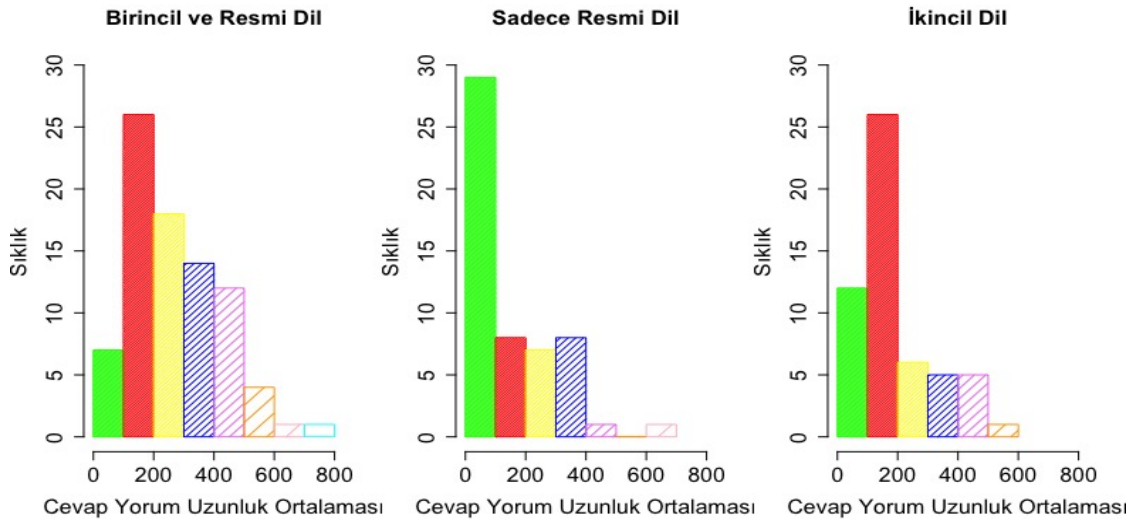
Şekil 4.8 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Understanding Language 4: Learning and Teaching" kursu

"Understanding Language 4: Learning and Teaching" MOOC'unda, İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların ve cevap yorumların ortalama uzunluklarına (ortalama karakter sayıları) Şekil 4.9 ve Şekil 4.10'da yer verilmiştir. İngilizce dil gruplarının ana ve cevap yorumlarının ortalama uzunlukları arasındaki farkları açıkça ortaya koymak için (gruplardaki katılımcı sayıları çok değişkenlik göstermektedir) Şekil 4.9 ve Şekil 4.10'da İngilizce dil gruplarından eşit sayıda katılımcı kullanılmıştır.



Şekil 4.9 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki ana yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.9'a göre, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce'yi birincil ve resmi dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları daha fazla oranda 200 karakterden fazla iken, "ikincil dil olarak İngilizce" konuşan katılımcılarla İngilizce'yi yalnızca resmi dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları daha az oranda 200 karakterden fazladır.



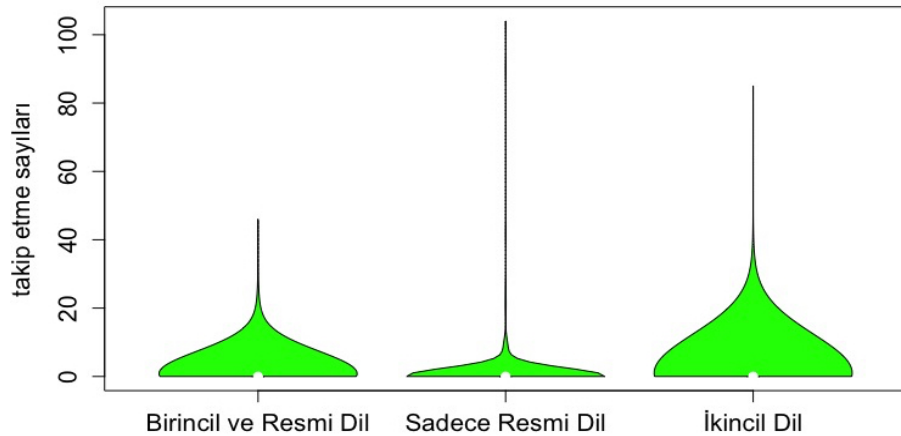
Şekil 4.10 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.10'da İngilizce dil gruplarındaki katılımcıların cevap yorumlarının uzunluklarının ortalamalarına yer verilmiştir. Şekil 4.10'a göre, yalnızca resmi dil olarak İngilizce konuşan katılımcıların yaptıkları cevap yorumları, düşük oranda 100 karakterden, "ikincil dil olarak İngilizce" konuşan katılımcıların cevap yorumlarının ortalama uzunlukları da düşük oranda 200 karakterden daha uzun olmuştur. Bununla birlikte, "birincil ve resmi dil olarak İngilizce" konuşan katılımcıların cevap yorumlarının ortalama uzunlukları genel olarak 100 karakterden daha uzundur.

4.2.3 Takip Etkileşimlerindeki Davranış Analizleri

FutureLearn platformu MOOC'larında katılımcıların takip etme davranışları ile ilgili Sunar ve diğerleri tarafından yapılan bir araştırma [7], takip etme davranışlarının (özellikle de katılımcıların yorum yaparak tartışma forumlarına katıldıkları zaman) kursun tamamlanmasıyla pozitif olarak ilişkili (kolerasyon sahibi) olduklarını ortaya koymuştur. Tezde, İngilizce dil gruplarındaki katılımcıların takip etme etkileşimleri detaylı olarak incelenmiştir.

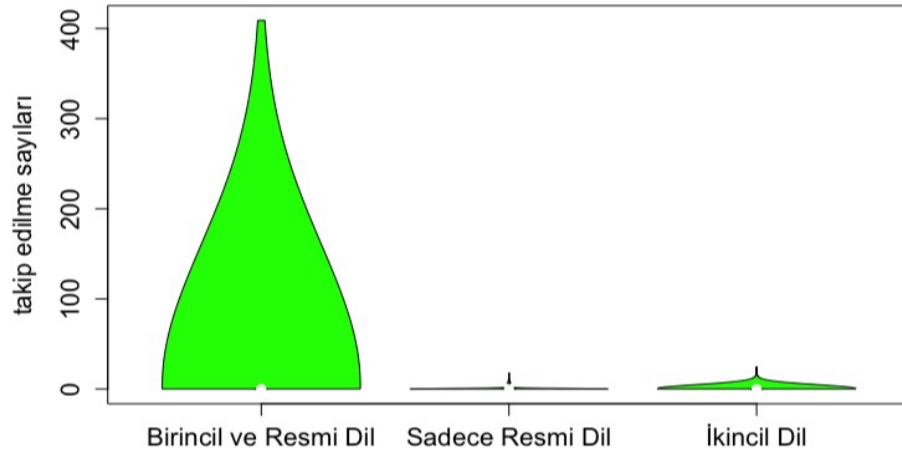
Şekil 4.11, İngilizce dil gruplarındaki katılımcıların, aynı kursa kayıtlı kaç farklı katılımcıyı kurs zamanı içinde takip ettiklerini göstermektedir. Şekil 4.11'e göre, "birincil ve resmi dil olarak İngilizce" konuşan katılımcılar ve "ikincil dil olarak İngilizce" konuşan katılımcılar genellikle daha fazla sayıda katılımcıyı takip etmektedir. "Sadece resmi dil olarak İngilizce" konuşan katılımcılar, genel olarak çok az sayıda takip etme davranışına sahiplerken, yüksek sayıda takip etme davranışına sahip bazı katılımcıların da olduğu (aykırı değerler) gözlemlenmiştir.



Şekil 4.11 İngilizce dil gruplarındaki katılımcıların aynı kurstan takip ettikleri katılımcı sayıları - "Understanding Language 4: Learning and Teaching" kursu

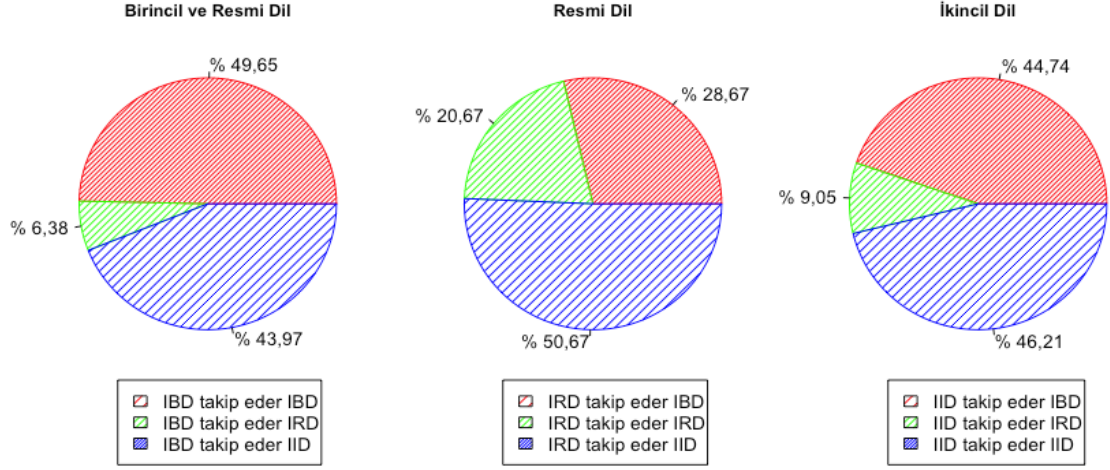
Şekil 4.12, İngilizce dil gruplarındaki katılımcıların aynı kursa kayıtlı kaç katılımcı tarafından kurs zamanı içinde takip edildiklerini göstermektedir. Şekil 4.12'ye göre

“birincil ve resmi dil olarak İngilizce” konuşan katılımcılar, en fazla sayıda katılımcı tarafından takip edilmişlerdir. “Sadece resmi dil olarak İngilizce” konuşan katılımcılar ile “ikincil dil olarak İngilizce” konuşan katılımcılar çok nadir olarak takip edilmiş olup, neredeyse hiç bir zaman 20’den fazla aynı kursa kayıtlı katılımcı tarafından takip edilmemişlerdir.



Şekil 4.12 İngilizce dil gruplarındaki katılımcıları aynı kurstan takip eden katılımcı sayıları - "Understanding Language 4: Learning and Teaching" kursu

Şekil 4.13, İngilizce dil grupları arasındaki takip etme etkileşimlerinin yüzdelerini göstermektedir. Şekil 4.13’e göre takip etme işlemlerinin büyük bir çoğunluğu hem ana hem de resmi dil olarak İngilizce (IBD) konuşan katılımcılarla, “ikincil dil olarak İngilizce” (IID) konuşan katılımcılar tarafından gerçekleştirilmiştir. “Sadece resmi dil olarak İngilizce” (IRD) konuşan katılımcılar az sayıda kişiyi takip etmiş ve az sayıda kişi tarafından takip edilmişlerdir. İkincil dil olarak İngilizce konuşan katılımcıların toplam takip etme sayılarının çok olmasında, bu gruptaki katılımcı sayısının diğer gruplardan çok daha fazla olmasından kaynaklanmış olabileceği düşünülmüştür.



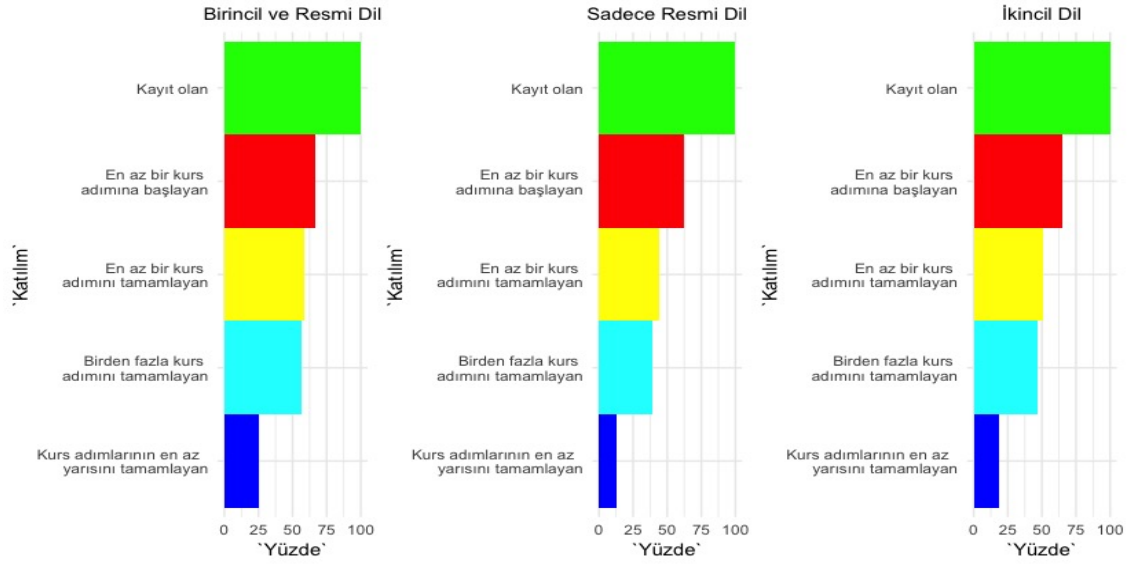
Şekil 4.13 İngilizce dil grupları arasında kullanıcıların takip etme etkileşimlerinin yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu

4.3 "Understanding Language : Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonunda İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri

"Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunun verileri üzerinde gerçekleştirilen İngilizce dil gruplarının istatistiksel kullanım analizleri, kullanılan veri kaynaklarına göre kurs adımlarındaki davranış analizleri ve kurs tartışma forumlarındaki davranış analizleri olarak iki alt başlık altında incelenmiştir. **MOOC serisinde yer verilen iterasyonların tamamının gerçekleştirildiği zaman aralığına ait takip verisi FutureLearn platformu tarafından sağlanmadığı için MOOC serisine ait takip davranış analizlerine yer verilmemiştir.**

4.3.1 Kurs Adımlarındaki Davranış Analizleri

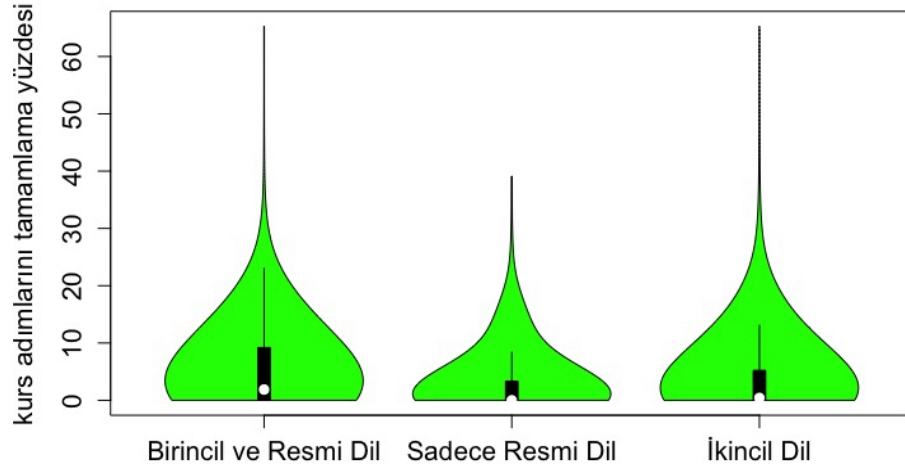
Şekil 4.14, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarındaki katılımcıların azalan devam etme yüzdelerini göstermektedir.



Şekil 4.14 İngilizce dil gruplarındaki kullanıcıların kursa katılımlarının azalımı - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.14'ten İngilizce'yi birincil ve resmi dil olarak konuşan katılımcıların grubunda kurs adımlarına katılımının ve kurs bitirme yüzdelerinin en yüksek olduğu ve MOOC serisindeki azalan devam etme yüzdeleri, serisinin dördüncü MOOC'u olan "Understanding Language 4 : Learning and Teaching" MOOC'undaki ile çok benzer olduğu görülmektedir.

Şekil 4.15'deki Violin grafiği, İngilizce dil gruplarındaki katılımcıların kurs sonunda adımları tamamlama yüzdelerini göstermektedir (katılımcıların başladıkları kurs adımlarının tüm kurs adımları içindeki yüzdelerini). Şekil 4.15'deki yüzdeler hesaplanırken, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonundaki toplam kurs adımı sayısı dikkate alınmıştır. Katılımcıların katılmadıkları kurslar da, kurs adımı tamamlama yüzdelerini düşürdüğü için, genel olarak "Understanding Language 4: Learning and Teaching" MOOC'una göre kurs adımlarını tamamlama yüzdeleri daha düşüktür.



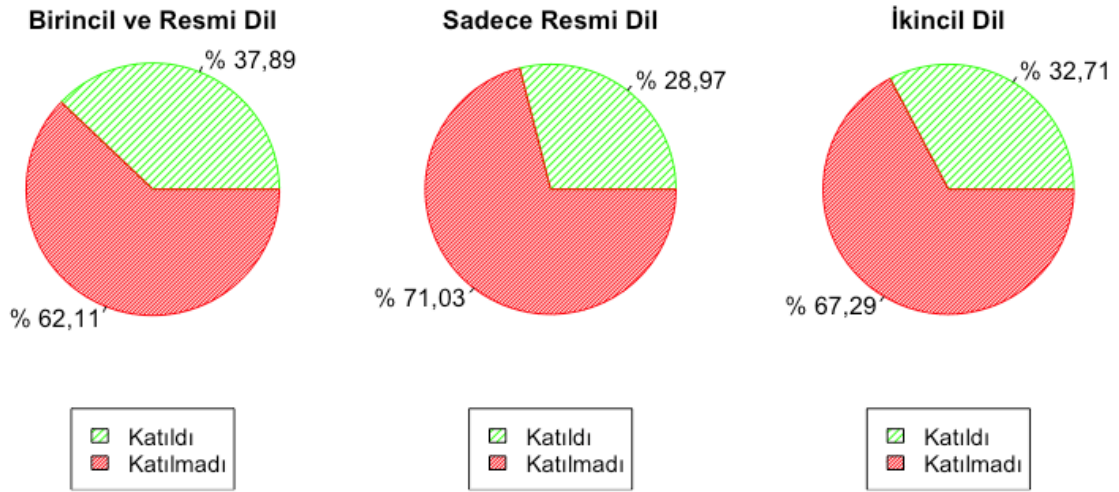
Şekil 4.15 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Understanding Language: Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.15'e göre, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda "birincil ve resmi dil olarak İngilizce" konuşan katılımcıların kurs adımlarının genelini tamamlama yüzdesi, diğer İngilizce dil gruplarındaki katılımcılara göre daha yüksektir.

4.3.2 Kurs Tartışma Forumlarındaki Davranış Analizleri

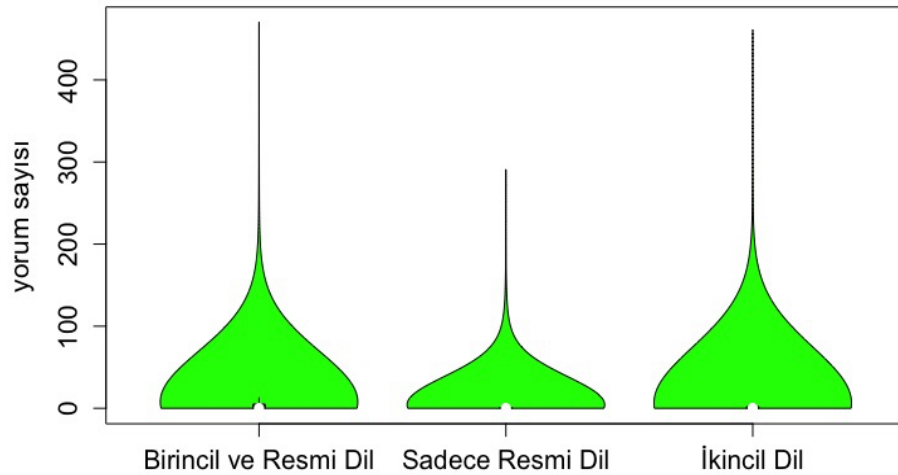
Şekil 4.16, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil grupları arasında tartışma forumlarında aktif (en az bir ana yorum veya cevap yorum yazan) veya pasif katılımcıların (ana veya cevap yorumu olmayan) tartışma forumlarına katılımlarını göstermektedir.

Şekil 4.16'ya göre, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarının tartışma forumlarına katılım yüzdeleri arasındaki fark aynı MOOC serisinin dördüncü iterasyonu olan "Understanding Language 4 : Learning and Teaching" MOOC'una göre biraz daha belirgindir ve İngilizceyi ana ve resmi dil olarak konuşan katılımcılar diğer İngilizce dil gruplarındaki katılımcılara göre biraz daha fazla yüzdeyle tartışma forumlarına katılmışlardır.



Şekil 4.16 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

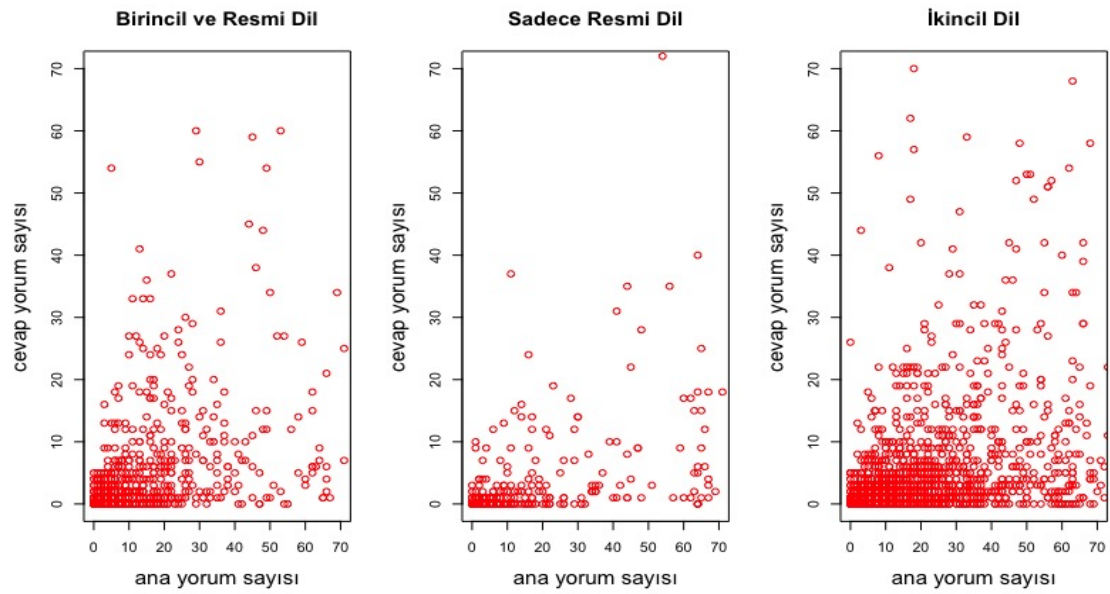
Şekil 4.17, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarında katılımcı başına düşen toplam yorum sayılarını (ana yorum veya cevap yorum) göstermektedir. Şekil 4.17'e göre, aynı MOOC serisinin dördüncü iterasyonu olan "Understanding Language 4 : Learning and Teaching" MOOC'undan farklı olarak İngilizce dil gruplarındaki katılımcıların yorum sayıları arasında büyük fark yoktur. Bu durumda iterasyon sayısının artması sebebi ile İngilizce dil gruplarındaki katılımcı sayısında meydana gelen artışların etkili olmuş olabileceği düşünülmüştür.



Şekil 4.17 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

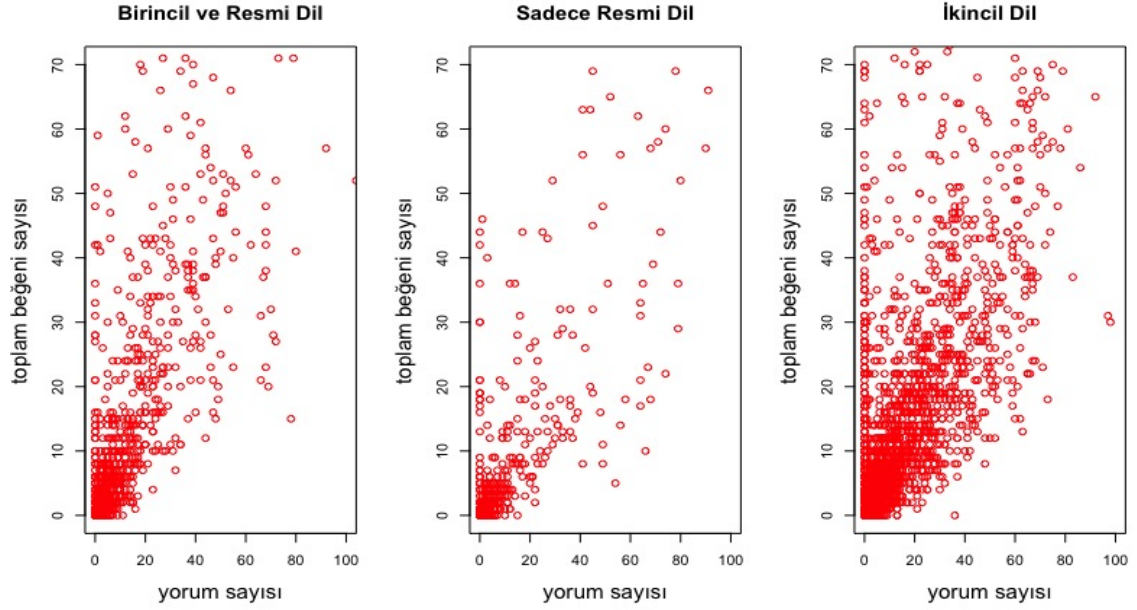
Şekil 4.18, "Understanding Language : Learning and Teaching" MOOC serisinin ilk

yedi iterasyonunda İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların toplam sayıları ile ana yorumlarının aldıkları cevap yorumların toplam sayılarını göstermektedir. Şekil 4.18'de daha net bir gösterim için ana yorumların ve cevap yorumların sayıları 70'le sınırlandırılmıştır. Şekil 4.18'e göre, İngilizce dil gruplarının toplam ana yorumların sayılarına göre aldıkları toplam cevap yorum sayıları benzer eğilimdedir. Katılımcılar tarafından kaç tane ana yorum yapıldığına bağlı olmaksızın, İngilizce'yi sadece resmi dil olarak konuşan katılımcıların aldıkları toplam cevap yorum sayısı genellikle beşten, diğer İngilizce dil gruplarındaki katılımcıların aldıkları toplam cevap yorum sayısı ise ondan az sayıdadır.



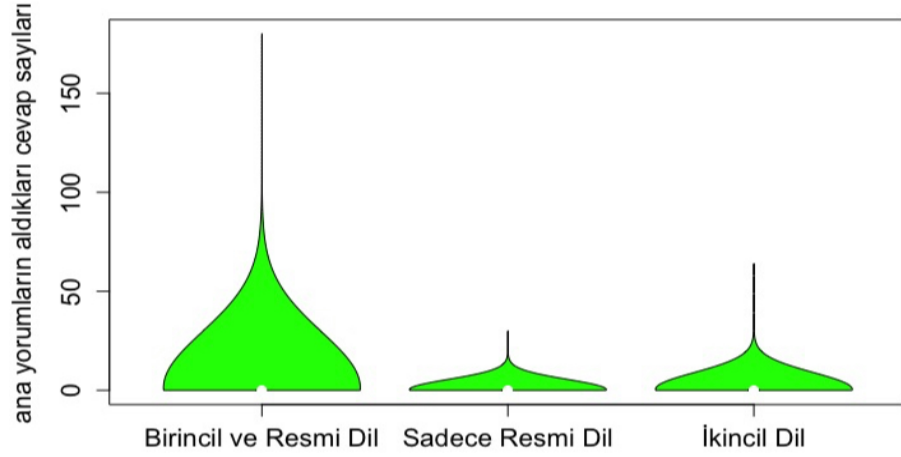
Şekil 4.18 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.19, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarındaki katılımcıların yaptıkları yorumların aldıkları toplam beğeni sayılarını göstermektedir. Şekil 4.19'da daha net bir gösterim için azami yorum ve beğeni sayıları 70'le sınırlandırılmıştır. Şekil 4.19'a göre İngilizce dil gruplarındaki yorum başına düşen beğeni sayılarının eğilimi benzerdir.



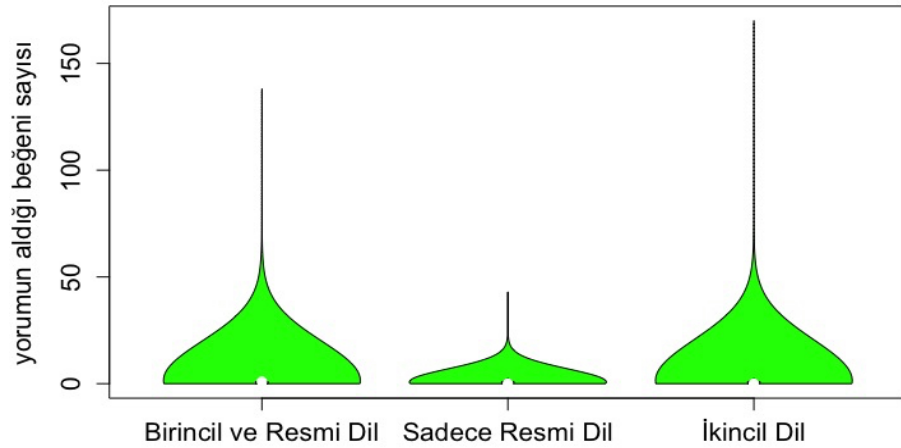
Şekil 4.19 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.20, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarındaki katılımcılar tarafından yapılan ana yorumların aldıkları cevap sayılarını göstermektedir. Şekil 4.20'ye göre ana ve resmi dil olarak İngilizce konuşan katılımcıların yorumları, diğer İngilizce dil gruplarındaki katılımcıların yorumlarından daha fazla cevap almaya eğilimlidirler. Aynı MOOC serisinin dördüncü iterasyonu olan "Understanding Language 4 : Learning and Teaching" MOOC'undan farklı olarak ana ve resmi dil olarak İngilizce konuşan katılımcıların yorumları ile diğer İngilizce dil gruplarındaki katılımcıların yorumlarının aldıkları beğeni sayıları arasındaki fark belirgindir.



Şekil 4.20 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları cevap yorum sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

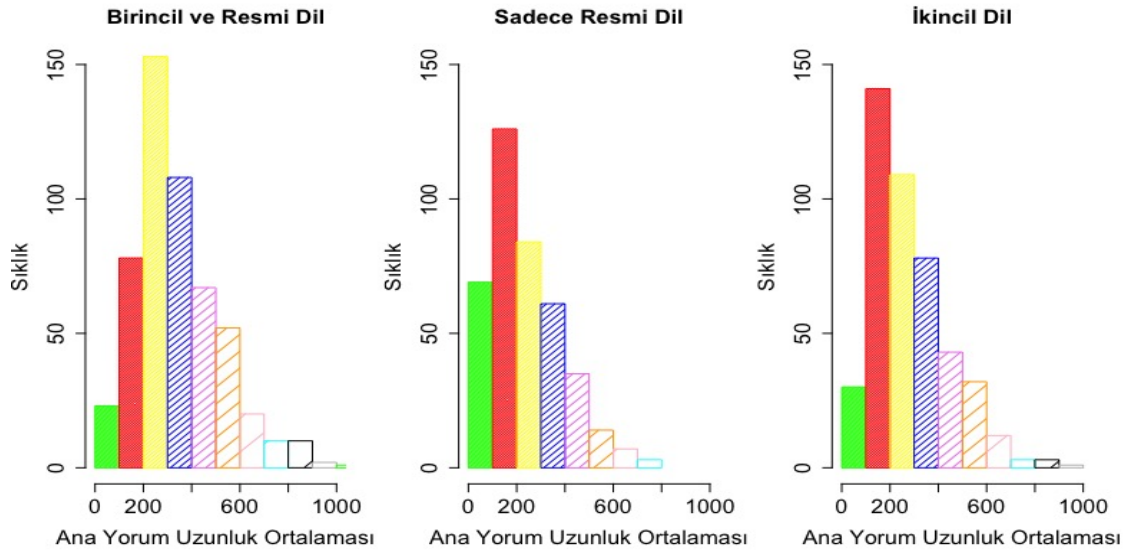
Şekil 4.21, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarındaki katılımcılar tarafından kurs forumlarında yapılan yorumların aldıkları beğeni sayılarını göstermektedir. Şekil 4.21'e göre, İngilizce dil gruplarındaki katılımcıların çoğunun yaptıkları yorumlar, beğeni almamış veya 50'nin altında bir beğeni almışsa da, birincil ve resmi dili İngilizce olan katılımcılarla "ikincil dil olarak İngilizce" konuşan çok sayıda beğeni alan yorumları da vardır. İkincil dil olarak İngilizce olan katılımcıların yorumlarının aldıkları beğeni miktarlarından fazla olmasında katılımcı sayısının fazlalığının etkisinin olabileceği düşünülmüştür.



Şekil 4.21 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

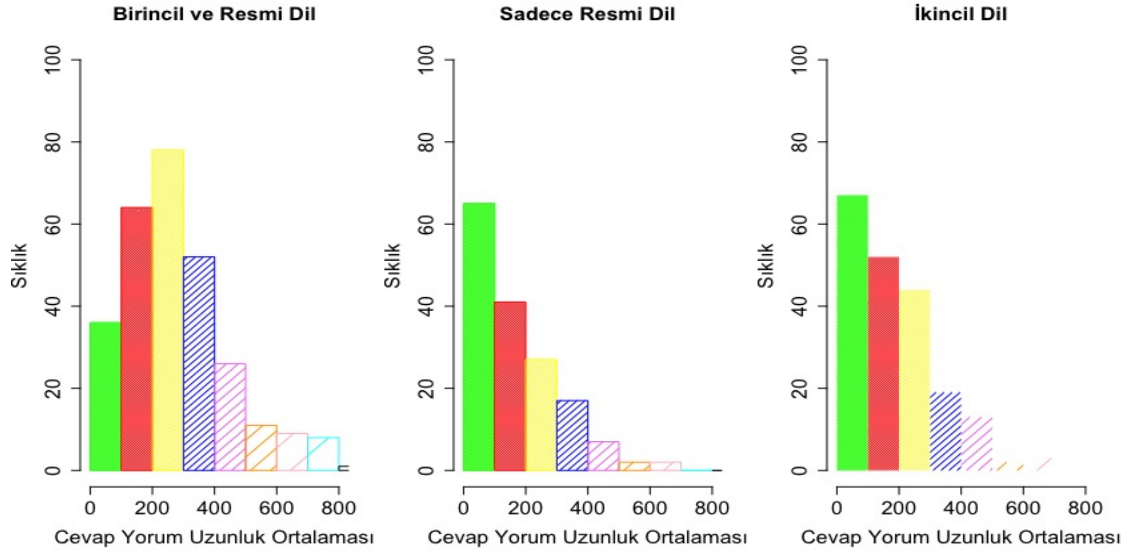
"Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi

iterasyonunda İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların ve cevap yorumların ortalama uzunluklarına (ortalama karakter sayıları) Şekil 4.22 ve Şekil 4.23’de yer verilmiştir. İngilizce dil gruplarının ana ve cevap yorumlarının ortalama uzunlukları arasındaki farkları açıkça ortaya koymak için (gruplardaki katılımcı sayıları çok değişkenlik göstermektedir) Şekil 4.22 ve Şekil 4.23’de İngilizce dil gruplarından eşit sayıda katılımcı kullanılmıştır.



Şekil 4.22 İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.22’ye göre, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce’yi birincil ve resmi dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları en fazla 200-300 arası karakter uzunluğunda iken, “ikincil dil olarak İngilizce” konuşan katılımcılarla İngilizce’yi yalnızca resmi dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları en fazla 100-200 arası karakter uzunluğundadır.



Şekil 4.23 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Understanding Language : Learning and Teaching" Kurs (MOOC) serisinin ilk yedi iterasyonu

Şekil 4.23’de yer alan İngilizce dil gruplarındaki katılımcıların cevap yorumlarının uzunluklarının ortalamalarına göre, cevap yorum uzunluklarının ortalamaları arasındaki fark, ana yorumların uzunluklarının ortalamaları arasındaki farka göre (Şekil 4.22) daha belirgindir. Şekil 4.23’e göre, “birincil ve resmi dil olarak İngilizce” konuşan katılımcıların yaptıkları cevap yorumların ortalama uzunlukları en fazla 200-300 karakter arasında, “sadece resmi dil olarak İngilizce” konuşan katılımcılarla “ikincil dil olarak İngilizce” konuşan katılımcıların cevap yorumlarının ortalama uzunlukları da en fazla 0-100 arası uzunluğunda olmuştur.

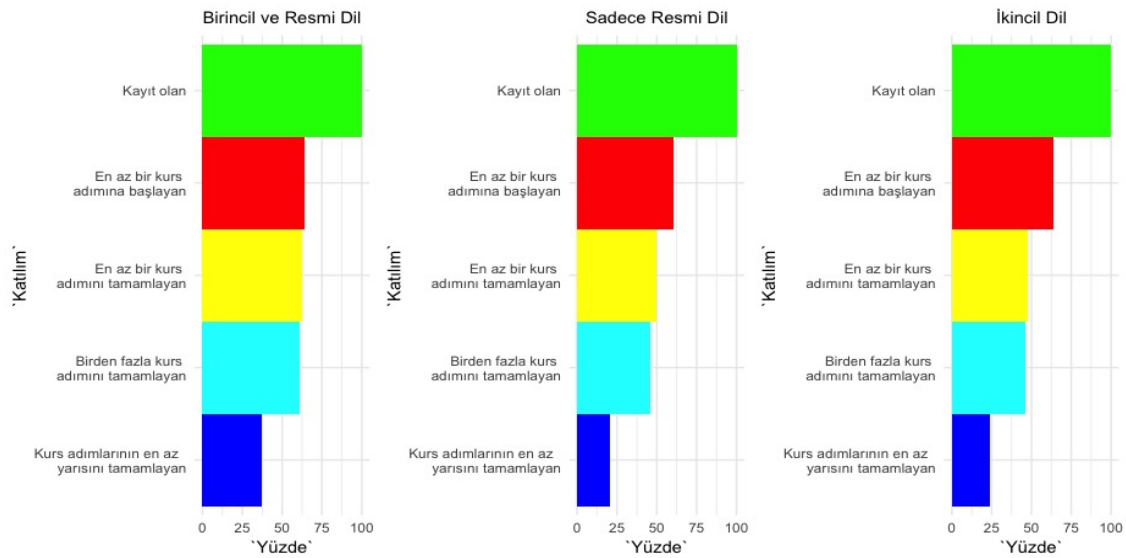
4.4 Farklı Türden Kurslarda (MOOC’lar) İngilizce Dil Gruplarının Karşılaştırmalı Kullanım Analizleri

"Understanding Language : Learning and Teaching" MOOC serisinden farklı türlerde MOOC’larda İngilizce dil gruplarının kullanım analiz sonuçlarında farklılık olup olmayacağını gözlemlemek amacı ile, FutureLearn platformuna ait "Exploring Our Oceans 4" ve "Web Science 1" kursları üzerinde de kullanım analizleri gerçekleştirilmiştir. Kullanılan veri kaynaklarına göre kurs adımlarındaki davranış analizleri, kurs tartışma forumlarındaki davranış analizleri olarak iki alt başlık altında incelenmiştir. **MOOC’ların gerçekleştirildiği zamanlara ait takip verisi FutureLearn platformu tarafından sağlanmadığı için MOOC’lara ait takip davranış analizlerine yer verilmemiştir.**

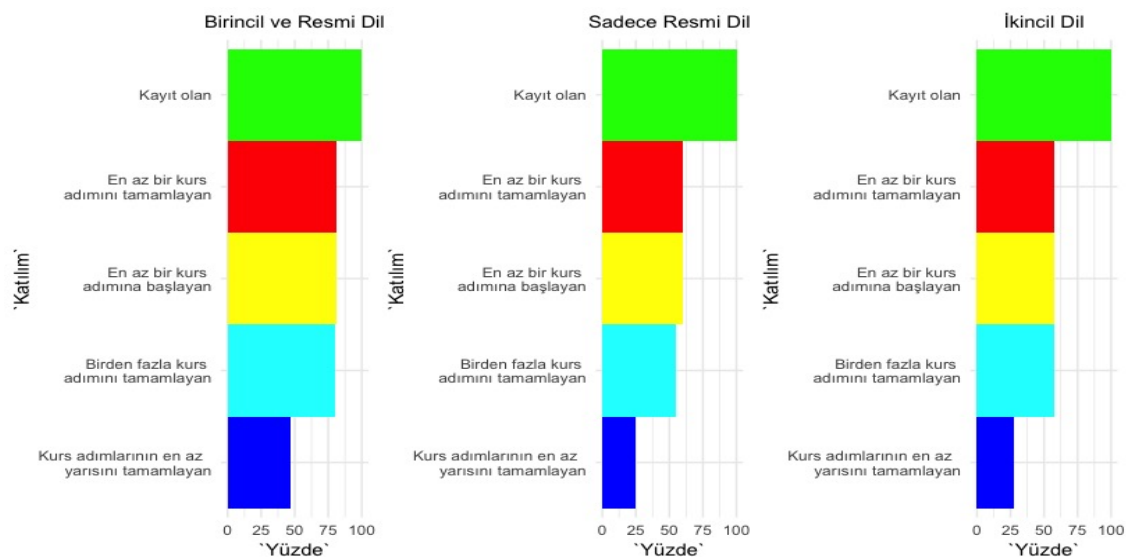
"Exploring Our Oceans 4" ve "Web Science 1" kurslarında "Understanding Language : Learning and Teaching" MOOC serisi kurslarından farklı olarak ana ve resmi dil olarak İngilizce konuştuğu tespit edilen katılımcıların sayıları, "ikincil dil olarak İngilizce" konuştuğu tespit edilen katılımcı sayılarından fazla olmuştur.

4.4.1 Kurs Adımlarındaki Davranış Analizleri

Şekil 4.24 ve Şekil 4.25 "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında İngilizce dil gruplarındaki azalan devam etme eğilimlerini göstermektedir.



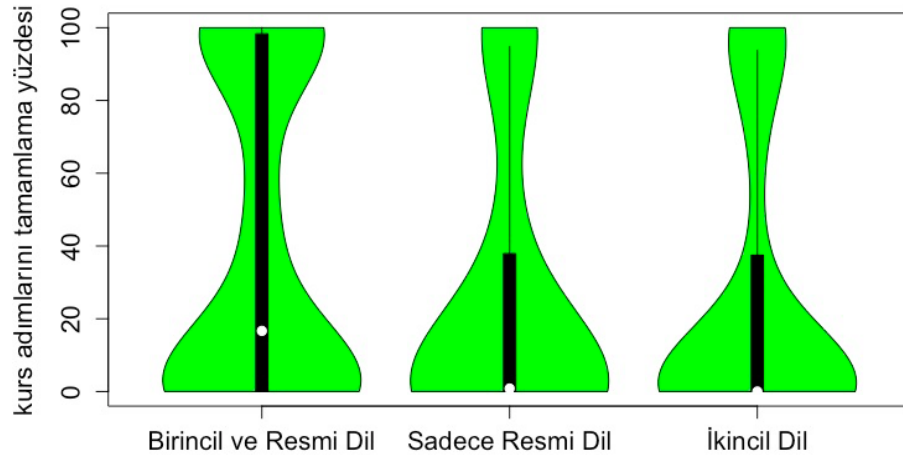
Şekil 4.24 İngilizce dil gruplarındaki kullanıcıların kursa katılımlarının azalımı - "Exploring Our Oceans 4" kursu



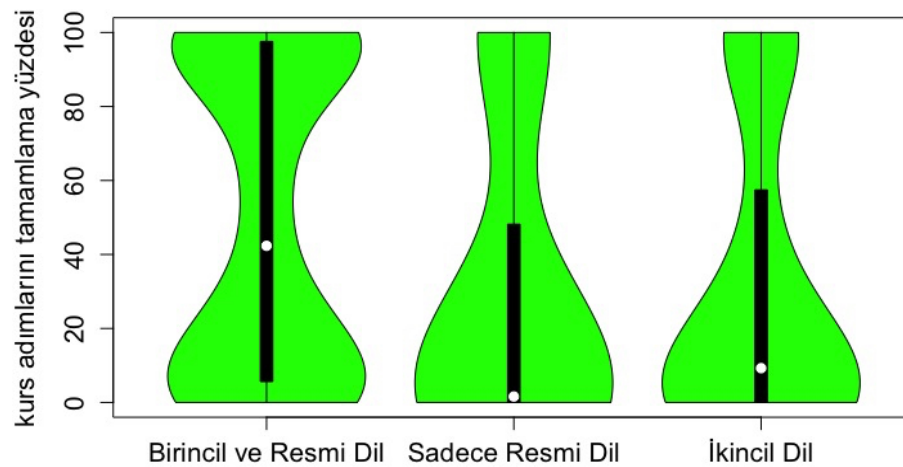
Şekil 4.25 İngilizce dil gruplarındaki kullanıcıların kursa katılımları - "Web Science 1" kursu

Şekil 4.26 ve Şekil 4.27'den İngilizce'yi birincil ve resmi dil olarak konuşan katılımcıların grubunda kurs adımlarına katılımının ve kurs bitirme yüzdelerinin en yüksek olduğu ve MOOC serisindeki azalan devam etme yüzdelerinde kurs ilerledikçe "Web Science 1" MOOC'unda, "Exploring Our Oceans 4" MOOC'una göre daha az düşüş olduğu görülmektedir. Her iki MOOC'daki azalan devam etme yüzdelerinde kurs ilerledikçe görülen düşüşler "Understanding Language : Learning and Teaching" MOOC serisi kurslarına göre daha azdır.

Şekil 4.26 ve Şekil 4.27'deki Violin grafikleri, İngilizce dil gruplarındaki katılımcıların "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında kurs sonunda adımlarını tamamlama yüzdelerini göstermektedir (katılımcıların başladıkları kurs adımlarının tüm kurs adımları içindeki yüzdelerini).



Şekil 4.26 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Exploring Our Oceans 4" kursu

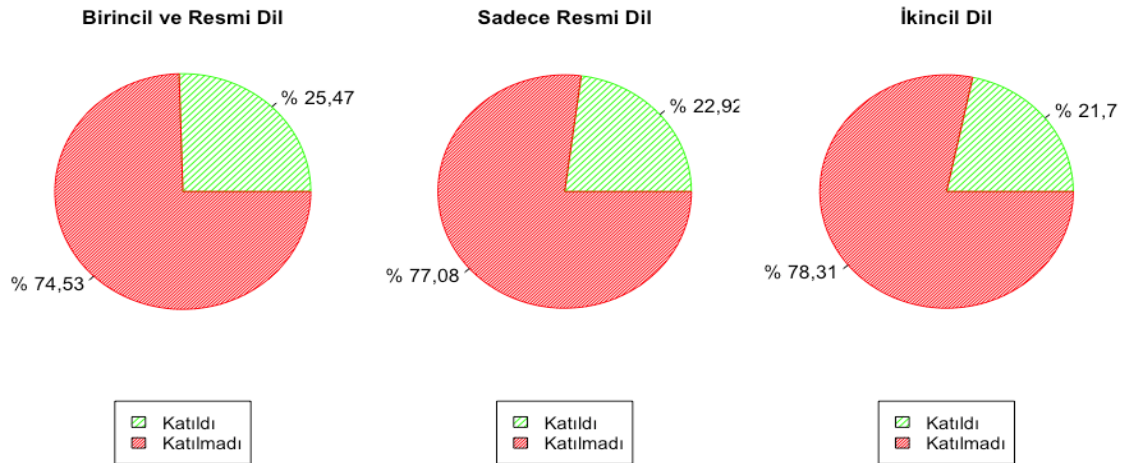


Şekil 4.27 İngilizce dil gruplarındaki katılımcıların kurs adımlarının genelini tamamlama yüzdeleri - "Web Science 1" kursu

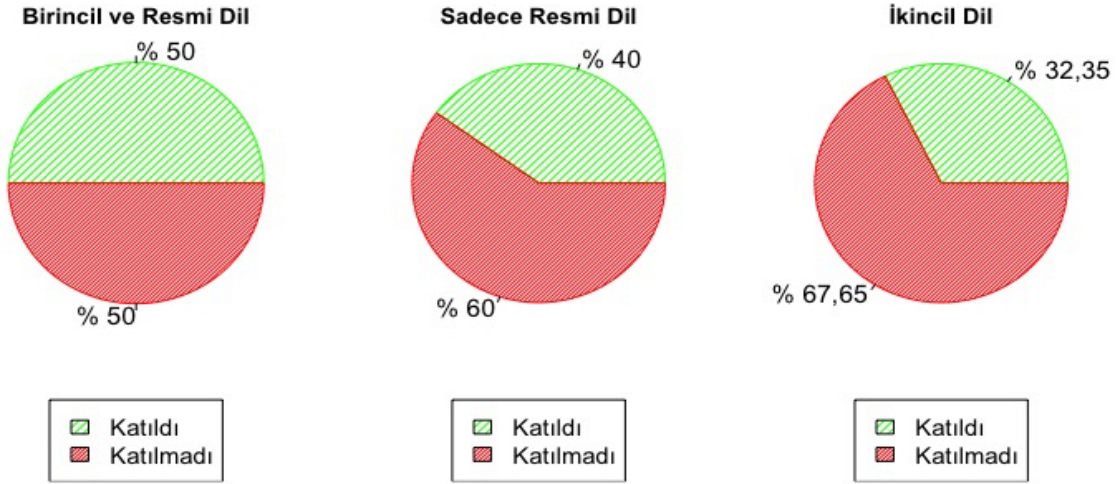
Şekil 4.26 ve Şekil 4.27'ye göre ana ve resmi dil olarak İngilizce konuşan katılımcıların "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında genel kurs adımlarını tamamlama yüzdesi, diğer İngilizce dil gruplarına göre daha yüksektir. Kurslar arasında ise İngilizce dil gruplarının genel kurs adımlarını tamamlama yüzdeleri "Web Science 1" MOOC'unda daha yüksektir.

4.4.2 Kurs Tartışma Forumlarındaki Davranış Analizleri

Şekil 4.28 ve Şekil 4.29, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında İngilizce dil grupları arasında tartışma forumlarında aktif (en az bir ana yorum veya cevap yorum yazan) veya pasif katılımcıların (ana veya cevap yorumu olmayan) tartışma forumlarına katılımlarını göstermektedir. Şekil 4.28 ve Şekil 4.29'a göre, "Exploring Our Oceans 4" MOOC'unda İngilizce gruplarının tartışma forumlarına katılım miktarları "Understanding Language : Learning and Teaching" MOOC serisi kurslarına göre daha az iken, "Web Science 1" MOOC'unda daha fazladır. Şekil 4.28 ve Şekil 4.29'dan aynı zamanda her iki MOOC'da da ana ve resmi dil olarak İngilizce konuşan katılımcıların tartışma forumlarına katılımları diğer İngilizce dil gruplarına göre daha fazla olduğu ve İngilizce dil gruplarının tartışma forumlarına katılım yüzdeleri arasındaki farkın "Web Science 1" MOOC'unda daha fazla olduğu da anlaşılmaktadır (bu farkta, kurslarda ana ve resmi dil olarak İngilizce konuşan katılımcı sayısının fazlalığının da etkili olmu olabileceği düşünülmüştür).

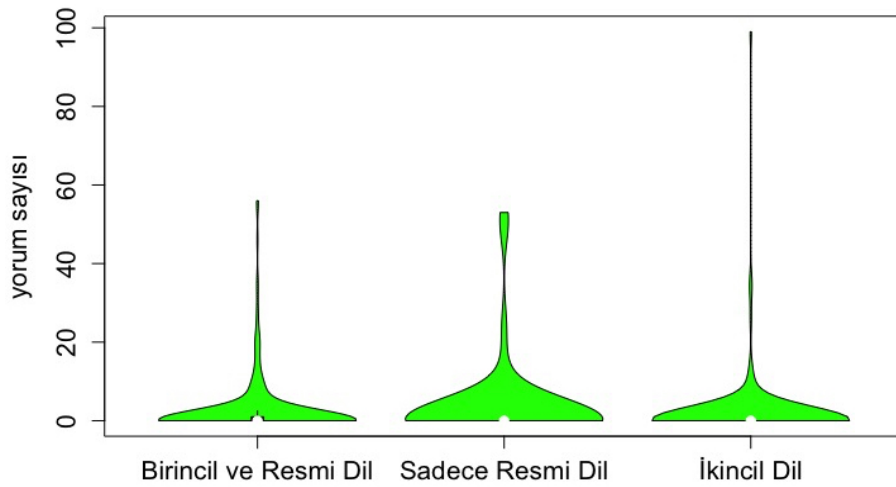


Şekil 4.28 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Exploring Our Oceans 4" kursu

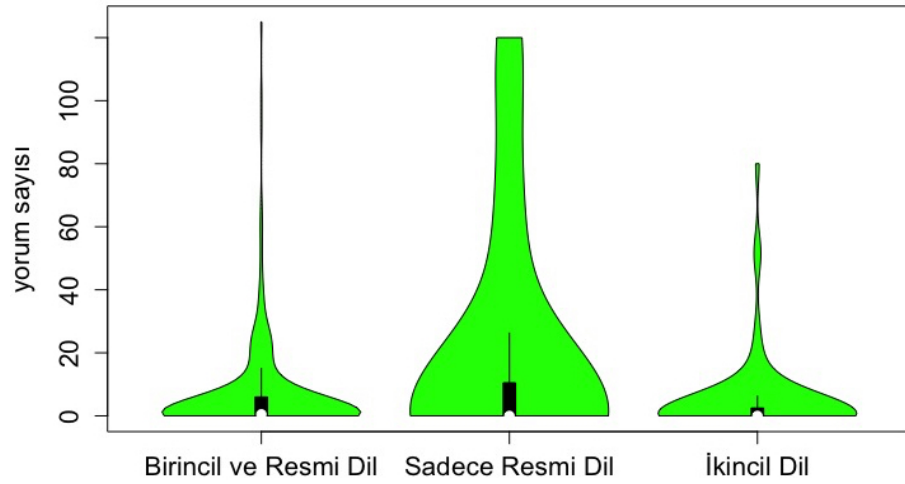


Şekil 4.29 İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri - "Web Science 1" kursu

Şekil 4.30 ve Şekil 4.31, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarında katılımcı başına düşen toplam yorum sayılarını (ana yorum veya cevap yorum) göstermektedir. Şekil 4.30 ve Şekil 4.31'e göre, "Understanding Language : Learning and Teaching" MOOC serisi kurslarından farklı olarak "sadece resmi dil olarak İngilizce" konuşan katılımcılar tartışma forumlarında daha fazla yorum yapmışlardır. İngilizce dil gruplarındaki katılımcıların tartışma forumlarındaki yorum sayıları arasındaki fark "Exploring Our Oceans 4" MOOC'unda az iken, "Web Science 1" MOOC'unda "sadece resmi dil olarak İngilizce" konuşan katılımcıların tartışma forumlarındaki yorum sayıları ile diğer İngilizce dil gruplarındaki yorum sayıları arasındaki fark fazla olmuştur.

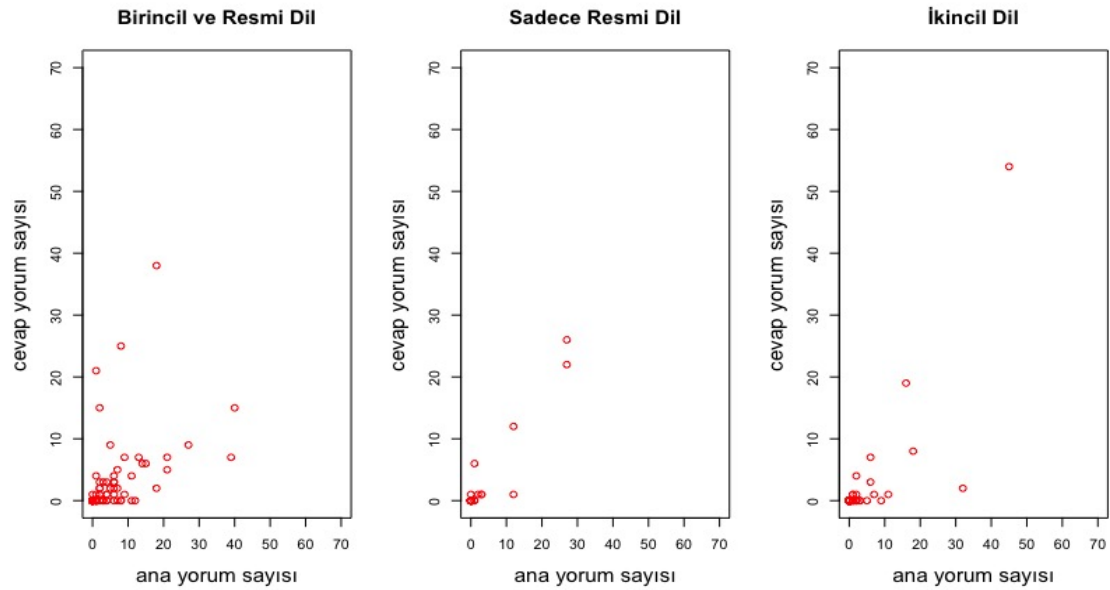


Şekil 4.30 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Exploring Our Oceans 4" kursu

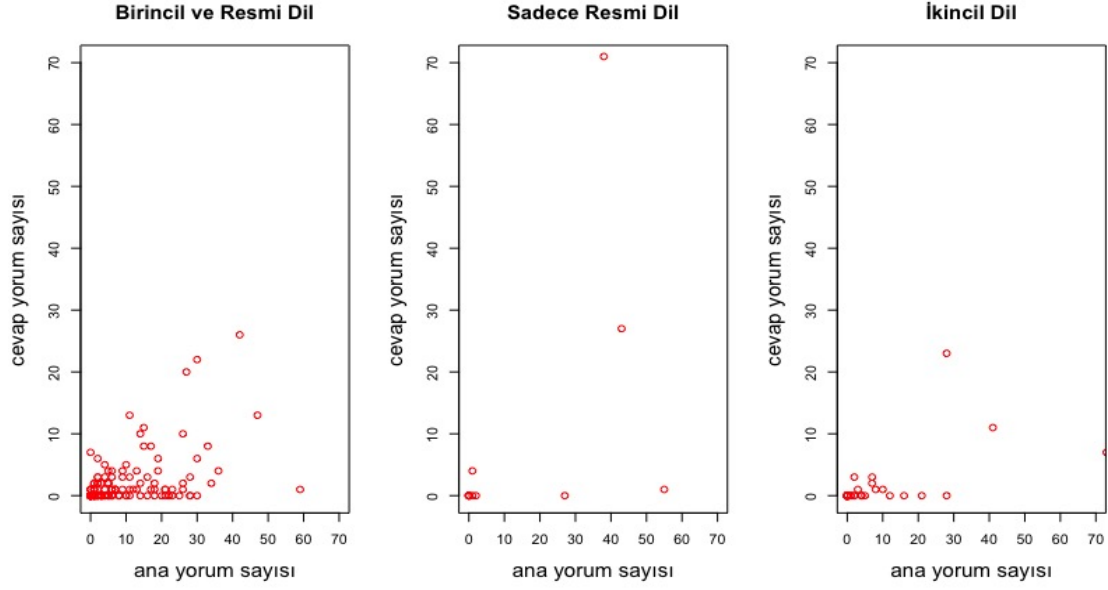


Şekil 4.31 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum sayıları - "Web Science 1" kursu

Şekil 4.32 ve Şekil 4.33, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların toplam sayıları ile ana yorumlarının aldıkları cevap yorumlarının toplam sayılarını göstermektedir. Şekil 4.32 ve Şekil 4.33'te daha net bir gösterim için ana ve cevap yorumlarının sayıları 70'le sınırlandırılmıştır. Şekil 4.32 ve Şekil 4.33'e göre, İngilizce dil gruplarının toplam ana yorumların sayısına karşı aldıkları toplam cevap sayısı eğilimleri benzerdir. Katılımcılar tarafından kaç tane ana yorum yapıldığına bağlı olmaksızın, katılımcılar genellikle 10'dan az sayıda cevap yorum almaktadırlar.

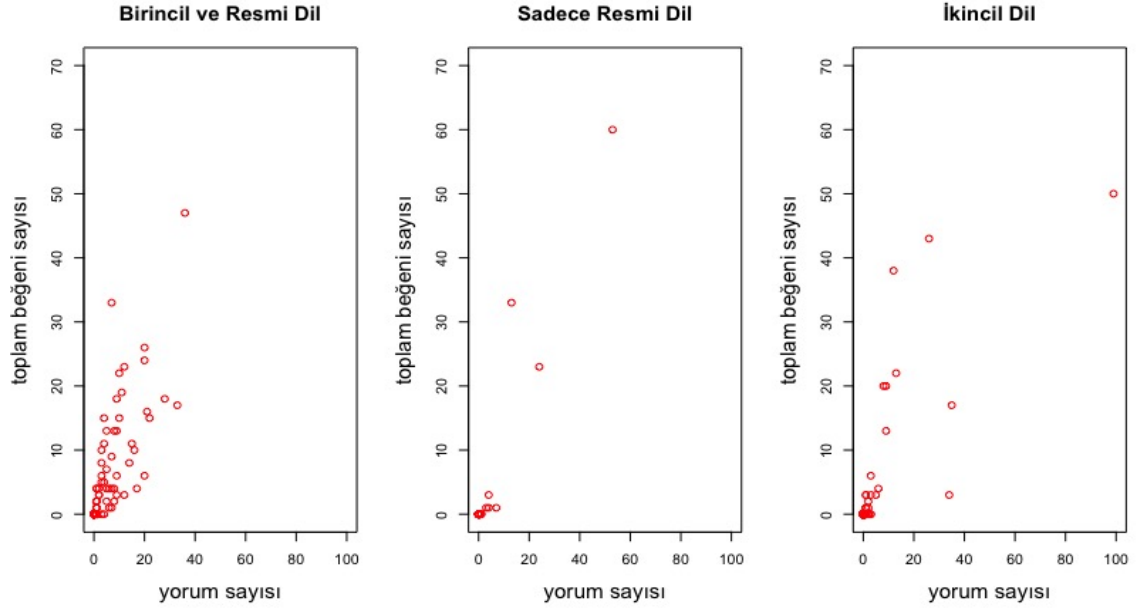


Şekil 4.32 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Exploring Our Oceans 4" kursu

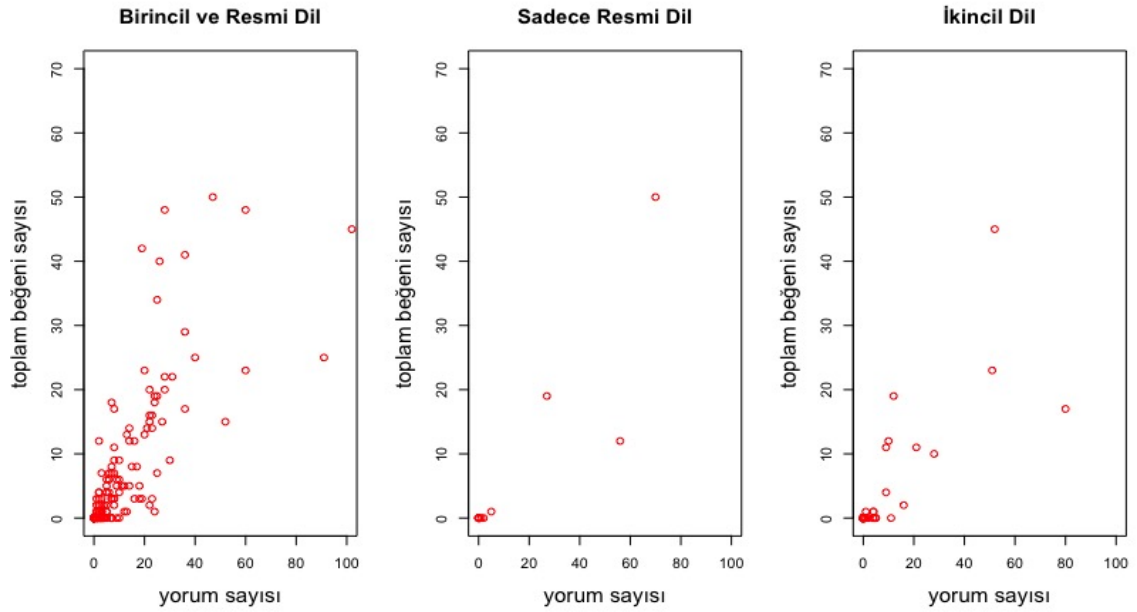


Şekil 4.33 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki toplam ana yorum ve cevap yorum sayıları - "Web Science 1" kursu

Şekil 4.34 ve Şekil 4.35, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarındaki katılımcıların yaptıkları yorumların aldıkları toplam beğeni sayılarını göstermektedir. Şekil 4.34 ve Şekil 4.35'te daha net bir gösterim için azami yorum ve beğeni sayıları 70'le sınırlandırılmıştır. Şekil 4.34 ve Şekil 4.35'e göre ana ve resmi dil olarak İngilizce konuşan katılımcıların dil gruplarındaki yorumların aldıkları beğeni sayıları diğer İngilizce dil gruplarındaki katılımcıların yorumların aldıkları beğeni sayılarından fazladır (bu durumda ana ve resmi dil olarak İngilizce konuşan katılımcıların sayılarının fazla olmasının da etkisinin olmuş olabileceği düşünülmüştür)



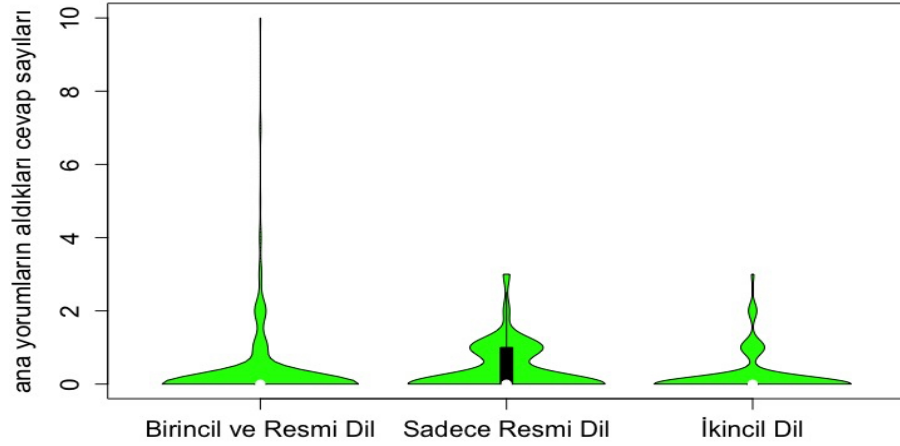
Şekil 4.34 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Exploring Our Oceans 4" kursu



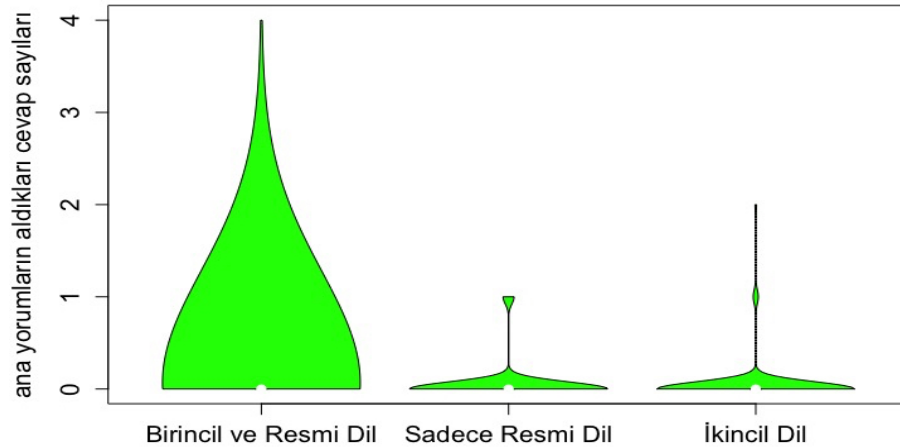
Şekil 4.35 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorum ve beğeni sayıları - "Web Science 1" kursu

Şekil 4.36 ve Şekil 4.37, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarındaki katılımcılar tarafından yapılan ana yorumların aldıkları cevap sayılarını göstermektedir. Şekil 4.36 ve Şekil 4.37'ye göre "Exploring Our Oceans 4" MOOC'unda İngilizce dil gruplarındaki katılımcıların yorumlarının aldıkları cevap yorum sayıları benzer iken, "Web Science 1" ana ve resmi dil olarak İngilizce konuşan

katılımcıların yorumları, diğer İngilizce dil gruplarındaki katılımcıların yorumlarından daha fazla cevap almaya eğilimlidirler ve gruplar arasında beğeni sayıları arasındaki fark belirgindir.

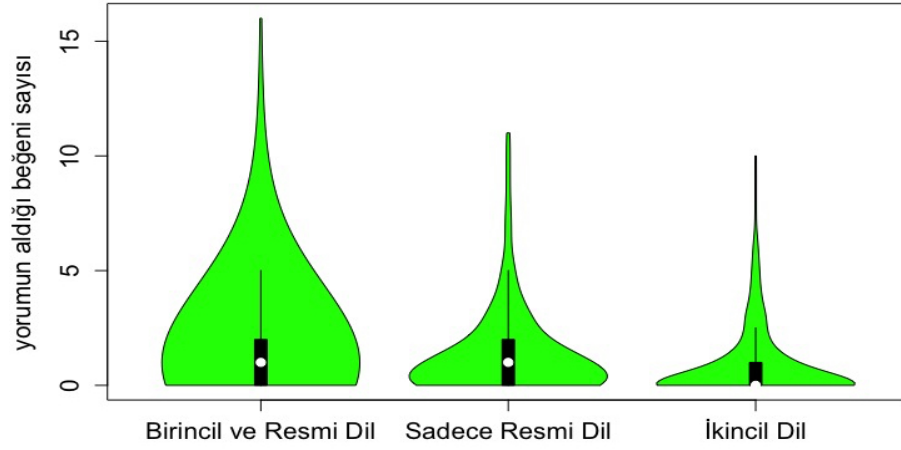


Şekil 4.36 İngilizce dil gruplarının kurs forumlarındaki yorumlarının aldıkları cevap sayıları - "Exploring Our Oceans 4" kursu

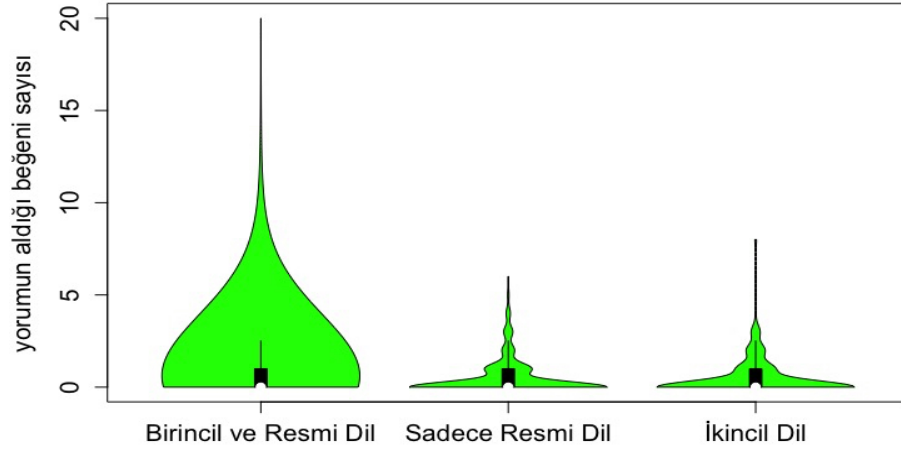


Şekil 4.37 İngilizce dil gruplarının kurs forumlarındaki yorumlarının aldıkları cevap sayıları - "Web Science 1" kursu

Şekil 4.38 ve Şekil 4.39, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarındaki katılımcılar tarafından kurs forumlarında yapılan yorumların aldıkları beğeni sayılarını göstermektedir. Şekil 4.38 ve Şekil 4.39'a göre, İngilizce dil gruplarındaki katılımcıların çoğunun yaptıkları yorumlar, beğeni almamış veya 20'nin altında bir beğeni almışsa da, birincil ve resmi dili İngilizce olan katılımcıların çok sayıda beğeni alan yorumları da vardır.

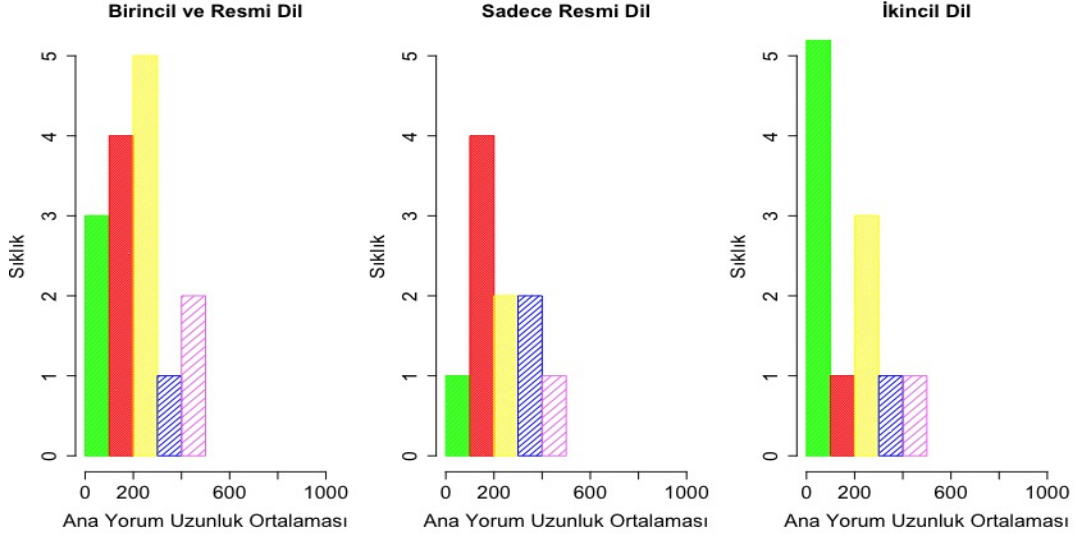


Şekil 4.38 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Exploring Our Oceans 4" kursu

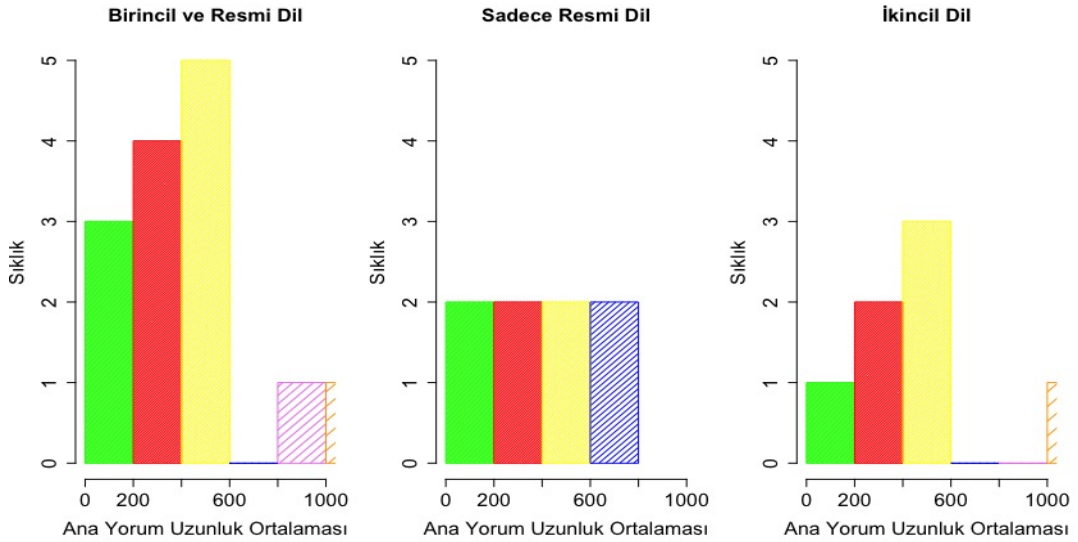


Şekil 4.39 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki yorumlarının aldıkları beğeni sayıları - "Web Science 1" kursu

"Exploring Our Oceans 4" ve "Web Science 1" MOOC'larındaki İngilizce dil gruplarındaki katılımcıların yaptıkları ana yorumların ve cevap yorumların ortalama uzunluklarına (ortalama karakter sayıları) Şekil 4.40, Şekil 4.41, Şekil 4.42 ve Şekil 4.43'de yer verilmiştir. İngilizce dil gruplarının ana ve cevap yorumlarının ortalama uzunlukları arasındaki farkları açıkça ortaya koymak için (gruplardaki katılımcı sayıları çok değişkenlik göstermektedir) Şekil 4.40, Şekil 4.41, Şekil 4.42 ve Şekil 4.43'de İngilizce dil gruplarından eşit sayıda katılımcı kullanılmıştır.

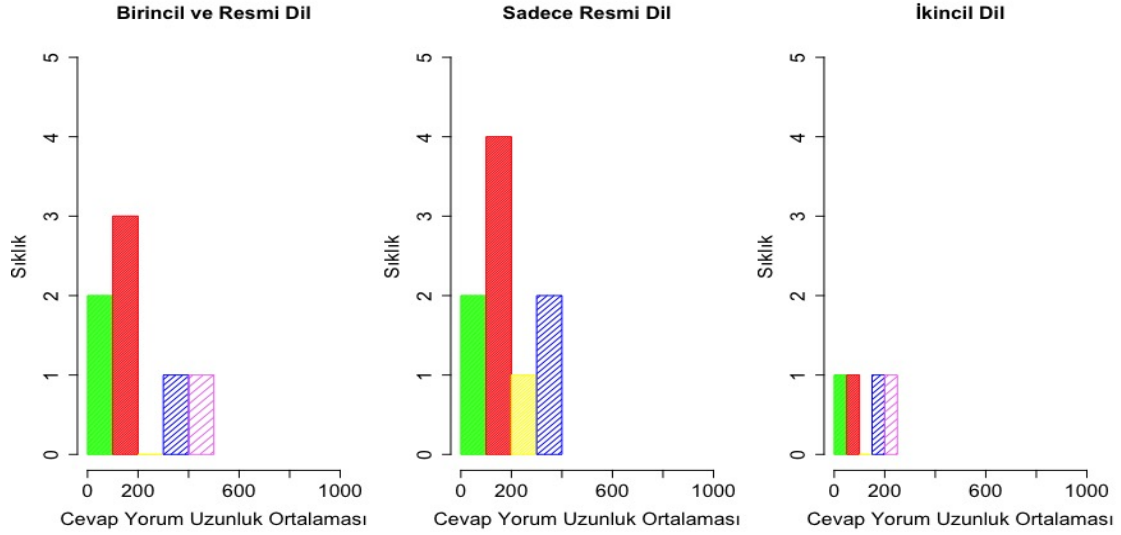


Şekil 4.40 İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Exploring Our Oceans 4" kursu

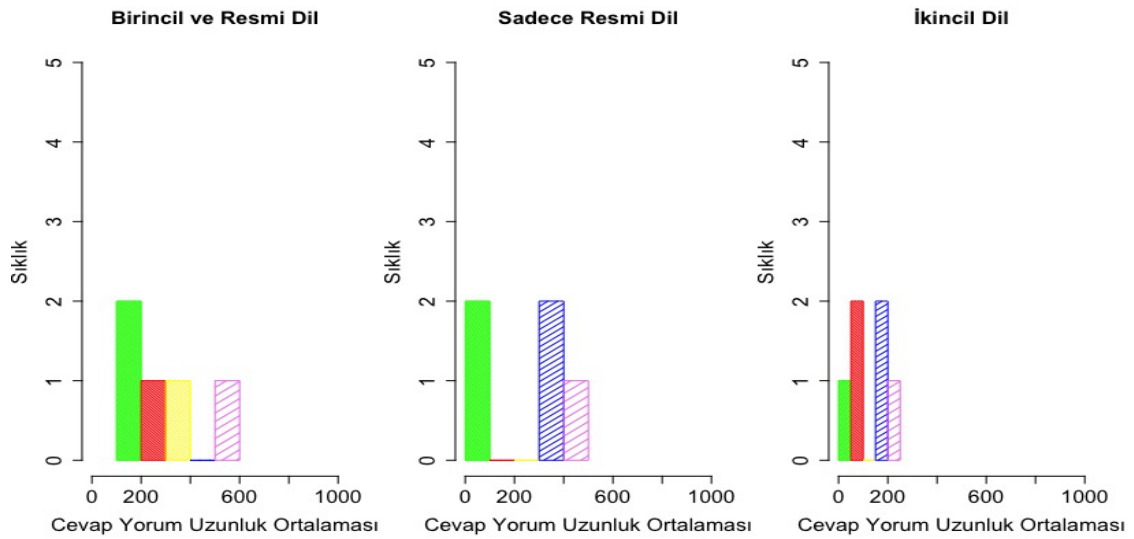


Şekil 4.41 İngilizce dil gruplarının kurs forumlarındaki yorumlarının uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Web Science 1" kursu

Şekil 4.40 ve Şekil 4.41'e göre, "Exploring Our Oceans 4" MOOC'unda İngilizce'yi birincil ve resmi dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları en fazla 200-300 arası karakter uzunluğunda iken, "sadece resmi dil olarak İngilizce" konuşan katılımcıların ana yorum uzunlukları en fazla 100-200 arası karakter uzunluğunda ve İngilizce'yi ikincil dil olarak konuşan katılımcıların yaptıkları ana yorumların ortalama uzunlukları en fazla 0-100 arası karakter uzunluğundadır. "Web Science 1" MOOC'unda İngilizce dil gruplarının ana yorum uzunluklarının dağılımı arasındaki fark ise daha azdır.



Şekil 4.42 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Exploring Our Oceans 4" kursu



Şekil 4.43 İngilizce dil gruplarındaki katılımcıların kurs forumlarındaki cevap yorumlarının ortalama uzunlukları (her bir İngilizce dil grubundan eşit sayıda katılımcının rastgele seçimi ile) - "Web Science 1" kursu

Şekil 4.42 ve Şekil 4.43’de yer alan İngilizce dil gruplarındaki katılımcıların cevap yorumlarının uzunluklarının ortalamalarına göre, "Exploring Our Oceans 4" ve "Web Science 1" MOOC’larında İngilizce dil gruplarındaki katılımcıların cevap yorum uzunlukları arasında belirgin bir fark yoktur.

4.5 Sonuç ve Gelecek Çalışmalar

Bu bölümde gerçekleştirilen analizlerin sonuçları ile FutureLearn platformu üzerinde hem İngilizce dil gruplarının kullanımları arasındaki farklar ortaya konulmuş hem de bir sonraki bölümde (Bölüm 5), İngilizce dil gruplarının kurs tamamlama tahmini modellerinde kullanılacak özellikler belirlenmiştir.

Kurs adımlarının kullanım analizleri ile ilgili sonuçlar aşağıdaki gibi özetlenebilir:

- Kullanılan tüm MOOC'larda kurs zamanı boyunca katılımcıların kursa katılımları azalmaktadır. Bununla birlikte "Web Science 1" MOOC'unda azalma oranı en az olup, bu MOOC'u "Exploring Our Oceans 4" MOOC'u takip etmektedir. Buradan kurs türünün katılım miktarlarının kurs zamanı içinde azalmaya etki ettiği sonucuna varılmıştır.
- Tüm kurslarda ana ve resmi dil olarak İngilizce konuşan katılımcıların kurs adımlarına katılım ve adımları tamamlama oranları diğer İngilizce dil gruplarına göre daha yüksektir.

Kurs tartışma forumlarının kullanım analizleri ile ilgili sonuçlar aşağıdaki gibi özetlenebilir:

- İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdelerine göre "Exploring Our Oceans 4" kursunda katılım en azdır. "Understanding Language : Learning and Teaching" MOOC serisinde de tartışma forumlarına katılım az olup, "Web Science 1" MOOC'unda ise fazladır.
- İngilizce dil gruplarında katılımcıların tartışma forumlarına katılım yüzdeleri arasındaki fark "Exploring Our Oceans 4" MOOC'unda ve "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda az iken, "Web Science 1" MOOC'unda ise fazladır.
- "Understanding Language : Learning and Teaching" MOOC serisi ve "Exploring Our Oceans 4" MOOC'unda yorum sayılarına göre İngilizce dil grupları arasında önemli fark yoktur. "Understanding Language 4: Learning and Teaching" MOOC'unda ana ve resmi dil olarak İngilizce konuşan katılımcıların yorum sayıları fazla iken, "Web Science 1" MOOC'unda ise "sadece resmi dil olarak İngilizce" konuşan katılımcıların yorum sayıları daha fazladır. Buradan İngilizce dil gruplarındaki katılımcıların yaptıkları yorum sayısının kurs türüne göre farklılık gösterebildiği anlaşılmaktadır.

- Kullanılan tüm MOOC'larda kurs tartışma forumlarına gönderilen ana yorumların aldıkları toplam cevap yorum sayıları, ana yorum sayısına bağlı olmaksızın genellikle 10'u geçmemiştir.
- Kullanılan tüm MOOC'larda kurs tartışma forumlarına gönderilen yorumların aldıkları beğeni sayıları yorum sayısı ile doğru orantılı olarak artmaktadır. Bununla birlikte, kurslarda ana ve resmi dil olarak İngilizce konuşan katılımcıların yorumları daha fazla beğeni almaya eğilimlidir.
- "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında ana ve resmi dil olarak İngilizce konuşan katılımcıların dil grubundaki yorumların aldıkları beğeni sayıları diğer İngilizce dil gruplarına göre belirgin olarak fazla iken, "Understanding Language : Learning and Teaching" MOOC serisinde İngilizce dil grupları arasındaki fark daha azdır.
- "Web Science 1" MOOC'u dışında kullanılan tüm MOOC'larda ana ve resmi dil olarak İngilizce konuşan katılımcıların kurs tartışma forumlarına gönderdikleri yorumlar daha uzun olma eğilimindedir ve uzunlukları en fazla 200-300 karakter arasında yoğunlaşmaktadır. "Web Science 1" MOOC'unda ise İngilizce dil gruplarının tartışma forumlarına gönderdikleri ana yorumların uzunlukları arasındaki fark daha azdır. Buradan hem katılımcıların ait oldukları İngilizce dil grubunun hem de kurs türünün katılımcıların ana yorum uzunluklarını etkilediği anlaşılmaktadır.
- "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda ana ve resmi dil olarak İngilizce konuşan katılımcıların kurs tartışma forumlarına gönderdikleri cevap yorumlar daha uzun olma eğilimindedir. Bununla birlikte "Exploring Our Oceans 4" ve "Web Science 1" İngilizce dil gruplarının tartışma forumlarına gönderdikleri cevap yorumların uzunlukları arasındaki fark daha azdır. Buradan hem katılımcıların ait oldukları İngilizce dil grubunun hem de kurs türünün katılımcıların cevap yorum uzunluklarını etkilediği anlaşılmaktadır.

Kurs takip etkileşimlerinin kullanım analizleri ile ilgili sonuçlar aşağıdaki gibi özetlenebilir:

- “Sadece resmi dil olarak İngilizce” konuşan katılımcılar diğer İngilizce dil gruplarına göre az sayıda takip etme gerçekleştirmektedir.
- Hem ana hem de resmi dil olarak İngilizce konuşan katılımcıların takip edilme sayıları diğer İngilizce dil gruplarına göre çok fazla sayıdadır.

- “Sadece resmi dil olarak İngilizce” konuşan katılımcılar az sayıda kişi tarafından takip edilmişlerdir.
- İkincil dil olarak İngilizce konuşan katılımcılarda takip etme sayısının fazla olmasında bu İngilizce dil grubundaki katılımcı sayısının fazla olmasının etkili olduğu düşünülmüştür.

İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri

Katılımcıların MOOC platformlarında birbirleriyle ve kurs içerikleri ile etkileşimlerinin ürettiği büyük miktarda veriler, katılımcıların gelecekteki kurs performanslarını tahmin etmek ve gerektiğinde çalışmalarına müdahale etmek için kullanılabilir.

Bölüm 4 (İngilizce Dil Gruplarının Kullanım Analizleri)'teki analizlerin sonuçlarına göre, İngilizce dil gruplarının kurs performanslarında ve kurs performanslarını belirleyen en önemli özelliklerde bazı farklılıkların olması beklenmektedir. Bu bölümde, katılımcıların kurs sonu performans tahminleri, sık kullanılan yedi geleneksel makine öğrenmesi algoritması ve oluşturulan Derin Öğrenme modelleri (farklı mimariler denenerek) ile gerçekleştirilerek elde edilen sonuçlara yer verilmiştir.

Bölümde, katılımcıların kurs sonu performansları analiz edilmiş ve kurs sonu performans tahmininde hangi özelliklerin önemli olduğu incelenmiştir. Aynı zamanda farklı algoritmalar ve mimariler denenerek, İngilizce dil gruplarının kurs sonu performansı tahmini için uygun algoritma ve Derin Öğrenme modeli belirlenmeye çalışılmıştır.

Kurs sonu performans analizlerinde veri kaynağı olarak, sırası ile FutureLearn platformundan "Understanding Language 4: Learning and Teaching" kursu, "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu, "Exploring Our Oceans 4" kursu ve son olarak "Web Science 1" kursu verileri kullanılmıştır.

Kurs performans tahminlerinden önce belirlenen kurallarla kurs tamamlama özelliği oluşturulmuş ve tahmin sonuçlarının güvenilirliği için özelliğin aldığı her farklı değerden eşit sayıda katılımcıya yer verilerek veriler dengelenmiştir.

Oluşturulan kurs sonu performansı özelliğinin alabileceği farklı değerler ve açıklamaları:

- **Bırakanlar:** Resmi olarak kurstan ayrılan veya kurs adımlarının hiçbirini tamamlamayan katılımcılardır.
- **Yavaş tempolu katılımcılar:** Kursun en az bir adımını tamamlayan ancak tamamladıkları adım sayısı kurstaki tüm adımların yarısından az olan katılımcılardır.
- **Tamamlayıcılar:** Kurs adımlarının en az yarısını tamamlayan veya katılım sertifikası alan katılımcılardır. FutureLearn'e ¹ göre kurs adımlarının % 50'sini veya fazlasını tamamlayan katılımcılar aktif katılımcılardır ve 'tamamlamış katılımcılar' (completers) olarak isimlendirilirler. FutureLearn'de, tüm kurs adımlarının % 90'ı veya üzerinde adım tamamlayan katılımcılar, kurs sertifika sınavına girmeye hak kazanmaktadırlar. Sertifika satın alan katılımcılar ile % 50 veya üzerinde kurs adımı tamamlayan katılımcılar daha güvenilir performansı tahmini için tek bir katılımcı grubunda 'tamamlamış katılımcılar' birleştirilmişlerdir.

5.1 Sık Kullanılan Geleneksel Makine Öğrenmesi Algoritmaları İle İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri

Kurs sonu performansı tahmini özellikleri oluşturulurken, katılımcıların kursa katılım verilerinden faydalanılmıştır.

Kurs tamamlama özelliğinin alabileceği farklı değerlere sahip katılımcı sayılarının dağılımına göre algoritmalarından gerçekçi tahmin sonuçları alabilmek için veriyi dengelemek gerektiği görülmüştür. Verileri dengelemek için rastgele aşağı örnekleme yöntemi kullanılmıştır. Bu yöntem, diğer örnekleme yöntemlerinde gerekli olan büyük miktarlarda örnek kopyalamaya gerek kalmadan verinin dengelenmesini sağlamaktadır. Tüm İngilizce dil gruplarında katılımcıların kurs sonu performansı özelliğinin alabileceği farklı değerlerden ('bırakanlar', 'yavaş tempolu katılımcılar', 'tamamlayıcılar'), en az sayıda katılımcısı olan referans olarak alınarak rastgele aşağı örnekleme gerçekleştirilmiştir. Rastgele aşağı örneklemede kurs sonu performansı tahmini için kullanılacak katılımcılar rastgele seçilmektedir.

Kurs sonu performans tahminleri, aşağıdaki yaygın olarak kullanılan geleneksel makine öğrenmesi algoritmalarıyla gerçekleştirilmiştir:

- Çok Katmanlı Algılayıcı (MLP);

¹<https://about.futurelearn.com/research-insights/learners-learning-know>

- Karar Ağacı (DT);
- K-En Yakın Komşu (KNN);
- Çok Kategorili Lojistik Regresyon (MLR);
- Naive Bayes (NB);
- Rastgele Orman (RF);
- Destek Vektör Makinesi (SVM).

Katılımcıların kurs sonu performansı tahmini için, Bölüm 4'te kurs katılımları kullanılarak oluşturulan özellikler kullanılmıştır. Algoritmaların eğitimi için de 10 kat çapraz doğrulama yöntemi uygulanmıştır.

5.1.1 "Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri

Kurs sonu performansı analizleri ilk olarak, "Understanding Language 4: Learning and Teaching" MOOC'u üzerinde gerçekleştirilmiştir.

"Understanding Language 4: Learning and Teaching" MOOC'unda her bir İngilizce dil grubu için kurs sonu performansını gösteren özelliğin aldığı değerlerin sayıları ve yüzdeleri Tablo 5.1'de gösterilmektedir. Tablo 5.1'e göre İngilizce dil grubu bilinen katılımcıların kurs sonu performansı özelliğinin aldığı değerlerin sayılarının dağılımı dengesizdir. "Birincil ve resmi dil olarak İngilizce" konuşan katılımcılarda, kurs tamamlama özelliğinin aldığı değerlerin dağılımı yönünden diğer İngilizce dil gruplarına göre belirgin farklılıklar vardır. Farklılıklardan biri, İngilizce dil grupları içinde, "birincil ve resmi dil olarak İngilizce" konuşanlarda 'tamamlayıcılar' değerini alanların yüzdesi en yüksek iken 'bırakanlar' değerini alanların yüzdesi en düşüktür ("birincil ve resmi dil olarak İngilizce" konuşan katılımcıların kursu tamamlama ihtimali daha yüksek iken diğer İngilizce dil gruplarındaki katılımcılarda kursu bırakma ihtimali daha yüksektir). Bir diğer fark da, "birincil ve resmi dil olarak İngilizce" konuşan katılımcılardan kurs sonu performans özelliğinin aldığı değerlere göre 'yavaş tempolu katılımcılar' sayısı ile 'tamamlayıcılar' değerini alan katılımcıların sayıları neredeyse eşitken diğer İngilizce dil gruplarında 'yavaş tempolu katılımcılar' değerinin yüzdesi 'tamamlayıcılar' değerinin yüzdesinden çok daha büyüktür.

Tablo 5.1 Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Understanding Language 4: Learning and Teaching" kursu

	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil	Tüm İngilizce Dil Grubu Bilinenler
Bırakanlar	316 (% 49,14)	237 (% 54,61)	1.198 (% 53,77)	1.751 (% 52,98)
Yavaş tempolu katılımcılar	165 (% 25,66)	140 (% 32,26)	645 (% 28,95)	950 (% 28,74)
Tamamlayıcılar	162 (% 25,19)	57 (% 13,13)	385 (% 17,28)	604 (% 18,28)
Toplam	643	434	2.228	3.305

Tablo 5.1'e göre, İngilizce dil gruplarında 'tamamlayıcılar' değerini alan katılımcıların sayısı en düşük iken, 'bırakanlar' değerini alan katılımcıların sayısı ise en yüksektir. Aynı zamanda, İngilizce dil gruplarında yer alan katılımcı sayıları önemli miktarda değişiklik göstermektedir.

Tablo 5.1'de, 'tamamlayıcılar' değerine sahip olan katılımcı sayısı en düşük (604) olduğu için tüm İngilizce dil grubu bilinen katılımcıların verisinden her kurs performansı özelliği değeri için 604'er katılımcı rastgele olarak seçilmiştir. Aynı yöntemle, her bir kurs performansı kategorisi için (kursu bırakma, yavaş tempolu katılımcılar, tamamlayıcılar), birincil ve resmi dili İngilizce olan katılımcıların verisinden 162, sadece resmi dili İngilizce olan katılımcıların verisinden 57, "ikincil dil olarak İngilizce" konuşan katılımcıların verisinden 385 katılımcı aşağı örnekleme ile seçilmiştir.

Tablo 5.2, denenen her algoritma için kurs sonu performansı tahmini doğruluk sonuçlarını göstermektedir. Sonuçlara göre, İngilizce dil gruplarının kurs performansı tahmin başarılarının ortalamaları alındığında, Rastgele Orman algoritması, kurs performansı tahmininde en iyi doğruluk yüzdesine sahip olurken bu algoritmadan sonra en başarılı algoritma Karar Ağacı olmuştur. K-En Yakın Komşu, Çok Kategorili Lojistik Regresyon, Destek Vektör Makinesi algoritmaları da katılımcıların kurs sonu performansı tahmininde iyi tahmin doğruluğuna sahip olmuşlardır. Katılımcıların kurs sonu performansını tahmin ederken en kötü performansı Çok Katmanlı Algılayıcı algoritması göstermiştir.

Tablo 5.2 Kurs performansı tahmininde sık kullanılan yedi geleneksel algoritma için tahmin doğruluk (accuracy) değerleri - "Understanding Language 4: Learning and Teaching" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Çok Katmanlı Algılayıcı	0,730	0,823	0,778	0,794
Karar Ağacı	0,9658	0,967	0,947	0,967
K-En Yakın Komşu	0,940	0,901	0,889	0,949
Çok Kategorili Lojistik Regresyon	0,921	0,975	0,977	0,930
Naive Bayes	0,829	0,870	0,807	0,843
Rastgele Orman	0,970	0,973	0,965	0,977
Destek Vektör Makinesi	0,928	0,955	0,924	0,936

Tablo 5.3, Rastgele Orman modeli ile kurs sonu performansı tahmininde Tutarlılık (Precision), Anma (Recall) ve F-ölçüm (F-measure) metriklerinin aldıkları değerleri göstermektedir. Bu değerlere göre de, uygulanan modeller içinde Rastgele Orman modeli en iyi kurs sonu performansı tahmini doğruluk yüzdesine sahip olduğu anlaşılmaktadır.

Tablo 5.3 Rastgele Orman algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma ve F-ölçüm metriklerinin aldıkları değerler - "Understanding Language 4: Learning and Teaching" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Tutarlılık	0,970	0,974	0,965	0,974
Anma	0,968	0,977	0,965	0,976
F-ölçüm	0,969	0,976	0,965	0,975

"Understanding Language 4: Learning and Teaching" MOOC'unda, kurs sonu performansı tahmininde en önemli üç özellik Rastgele Orman algoritması dışındaki tüm algoritmalar için 1) başlanan adım sayısı, 2) ikinci haftaya ait tamamlanan adım sayısı ve 3) birinci haftaya ait tamamlanan adım sayısı olarak bulunmuştur. İngilizce

dil gruplarında kurs sonu performansı tahmininde en iyi sonuca sahip algoritma olan Rastgele Orman algoritması, Tablo 5.4’de gösterildiği gibi diğer algoritmalarından farklılık gösteren ilk üç önemli özelliğe ve farklı önem derecesi sıralamasına sahip olmuştur

Tablo 5.4 Rastgele Orman algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik (denenen diğer algoritmalarından farklı olarak birinci hafta tamamlanan adım sayısı ikinci hafta tamamlanan adım sayısından daha önemlidir) - "Understanding Language 4: Learning and Teaching" kursu

Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan birinci hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan birinci hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan birinci hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan birinci hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı

Rastgele Orman algoritması ile kurs sonu performansı tahmininde, “sadece resmi dili İngilizce “ olan katılımcılar ve “ikincil dil olarak İngilizce” konuşan katılımcılar için ‘başlanan adım sayısı’ en önemli tahmin edici özellik olarak bulunmuştur. “Birincil ve resmi dil olarak İngilizce” konuşan katılımcılarda ise en önemli kurs sonu performansı tahmin özelliği ‘tamamlanan birinci hafta adım sayısı’ olarak bulunmuştur.

Rastgele Orman modelinin karmaşıklık matrisini gösteren Tablo 5.5’e göre, algoritma tüm İngilizce dil gruplarının kurs sonu performansı tahmininde ‘yavaş tempolu katılımcılar’ değerini alan katılımcılarda en büyük tahmin hata yüzdesine sahiptir. Tüm İngilizce dil grubu bilinenlerin ve “ikincil dil olarak İngilizce” konuşanların verilerinde, Rastgele Orman algoritması en iyi tahmin doğruluk sonucunu, ‘tamamlayıcılar’ değerini alan katılımcıların kurs sonu performansı tahmininde vermiştir. “Birincil ve resmi dili İngilizce“ olan katılımcılarla, “sadece resmi dili İngilizce“ olan katılımcıların verilerinde, Rastgele Orman algoritması en iyi tahmin doğruluk sonucunu, ‘kursu bırakan’ katılımcıların kurs sonu performansı tahmininde vermiştir.

FutureLearn platformuna ait "Understanding Language 4: Learning and Teaching" kursunda, kurs sonu performansı tahmininde Rastgele Orman algoritması en iyi tahmin doğruluğuna sahip olmuştur. Tahmin algoritmalarının doğruluğu çeşitli faktörlere göre değişebilmektedir. Rastgele Orman algoritmasının, kurs sonu performansı tahmininde ve önemli özellikleri belirlemede en iyi algoritma olmasının

Tablo 5.5 İngilizce dil gruplarında Rastgele Orman algoritmasının karmaşıklık matrisleri - "Understanding Language 4: Learning and Teaching" kursu

Tüm İngilizce Dil Grubu Bilinenler	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	591	11	34
Tamamlayıcılar	8	592	0
Bırakanlar	5	1	570
İngilizce Birincil ve Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	158	5	4
Tamamlayıcılar	2	157	0
Bırakanlar	2	0	158
İngilizce Yalnızca Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	53	2	2
Tamamlayıcılar	3	55	0
Bırakanlar	1	0	55
İngilizce İkincil Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	376	7	12
Tamamlayıcılar	4	377	1
Bırakanlar	5	1	372

sebebinin algoritmanın sınıflandırma için çoklu karar süreçlerinin bir koleksiyonunu kullanması olabileceği düşünülmüştür.

Daha büyük veri kullanımı ve farklı Derin Öğrenme model ve mimarileri ile kurs sonu performans analizlerinde iyileştirme sağlamak amacı ile ve kullanıcı yorum metinlerinden de faydalanmak için bir sonraki adımda, Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunun verilerinin kullanılmasına karar verilmiştir.

5.1.2 "Understanding Language : Learning and Teaching" Kurs (MOOC) Serisinin İlk Yedi İterasyonunda İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri

MOOC'larda kurs performansı tahmininde sık kullanılan yedi geleneksel algoritma ile kurs sonu performansı analizleri ikinci olarak "Understanding Language 4: Learning and Teaching" MOOC'undan sonra aynı MOOC verilerinin de yer aldığı "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonu üzerinde gerçekleştirilmiştir.

"Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda İngilizce dil gruplarının kurs sonu performansını gösteren özelliğin aldığı değerlerin sayıları ve yüzdeleri Tablo 5.6'da gösterilmektedir. Tablo 5.6'ya

göre İngilizce dil grubu bilinen katılımcıların kurs sonu performansı özelliğinin aldığı değerlerin dağılımları dengesizdir. “Birincil ve resmi dil olarak İngilizce” konuşan katılımcılarda, kurs tamamlama özelliğinin aldığı değerlerin dağılımı yönünden diğer İngilizce dil gruplarına göre belirgin farklılıklar vardır. İngilizce dil grupları içinde, “birincil ve resmi dil olarak İngilizce” konuşanlarda ‘tamamlayıcılar’ değerini alanların yüzdesi en yüksek iken ‘bırakanlar’ değerini alanların yüzdesi en düşüktür (“birincil ve resmi dil olarak İngilizce” konuşan katılımcıların kursu tamamlama ihtimali daha yüksek iken diğer İngilizce dil gruplarındaki katılımcılarda kursu bırakma ihtimali daha yüksektir). Tablo 5.6’ya göre, İngilizce dil gruplarında yer alan katılımcı sayıları da önemli miktarda değişiklik göstermektedir.

Tablo 5.6 Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Understanding Language: Learning and Teaching" kurs serisinin ilk yedi iterasyonu

	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil	Tüm İngilizce Dil Grubu Bilinenler
Bırakanlar	1.043 (% 49,04)	818 (% 58,51)	4.350 (% 53,94)	6.221 (% 53,63)
Yavaş tempolu katılımcılar	496 (% 23,32)	378 (% 27,04)	2.106 (% 26,11)	2.980 (% 25,69)
Tamamlayıcılar	588 (% 27,64)	202 (% 14,45)	1.609 (% 19,95)	2.399 (% 20,68)
Toplam	2.127	1.398	8.065	11.600

Tablo 5.6’da, ‘tamamlayıcılar’ değerine sahip olan katılımcı sayısı en düşük (2.399) olduğu için tüm İngilizce dil grubu bilinen katılımcıların verisinden her kurs performansı özelliği değeri için 2.399’ar katılımcı rastgele olarak seçilmiştir. Aynı yöntemle, her bir kurs performansı kategorisi için (kursu bırakma, yavaş tempolu katılımcılar, tamamlayıcılar), “birincil ve resmi dili İngilizce” olan katılımcıların verisinden 496, sadece resmi dili İngilizce olan katılımcıların verisinden 202, “ikincil dil olarak İngilizce” konuşan katılımcıların verisinden 1.609 katılımcı aşağı örnekleme ile seçilmiştir.

Tablo 5.7, denenen her algoritma için kurs sonu performansı tahmini doğruluk sonuçlarını göstermektedir. Sonuçlara göre, Karar Ağacı algoritması, kurs performansı tahmininde en iyi doğruluk yüzdesine sahip olmuştur. Çok Katmanlı Algılayıcı algoritması en kötü performansı göstermiştir.

Tablo 5.7 Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Çok Katmanlı Algılayıcı	0,711	0,694	0,713	0,764
Karar Ağacı	0,934	0,907	0,924	0,940
Çok Kategorili Lojistik Regresyon	0,850	0,832	0,871	0,857
Naive Bayes	0,780	0,754	0,817	0,778
Rastgele Orman	0,932	0,903	0,924	0,932
Destek Vektör Makinesi	0,895	0,865	0,880	0,886

Tablo 5.8, Karar Ağacı modeli ile kurs sonu performansı tahmininde Tutarlılık, Anma ve F-ölçüm metriklerinin aldıkları değerleri göstermektedir. Bu değerlere göre de, uygulanan modeller içinde Karar Ağacı modeli en iyi kurs sonu performansı tahmini doğruluk yüzdesine sahip olduğu anlaşılmaktadır.

Tablo 5.8 Karar Ağacı algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Tutarlılık	0,938	0,912	0,926	0,943
Anma	0,9342	0,906	0,933	0,940
F-ölçüm	0,936	0,909	0,929	0,941

"Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda, kurs sonu performansı tahmininde en önemli üç özellik, en başarılı algoritma olan Karar Ağacı algoritması için Tablo 5.9'da gösterildiği gibi 1) başlanan adım sayısı, 2) tamamlanan ikinci hafta adımı sayısı ve 3) tamamlanan birinci hafta adımı sayısı olarak bulunmuştur.

Karar Ağacı algoritması ile kurs sonu performansı tahmininde, tüm İngilizce dil

Tablo 5.9 Karar Ağacı algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu

Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. tamamlanan birinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. tamamlanan birinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan birinci hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. tamamlanan birinci hafta adımı sayısı

gruplarındaki katılımcılar için en önemli tahmin edici özellik 'başlanan adım sayısı' iken 'tamamlanan ikinci hafta adımı sayısı' ve 'tamamlanan birinci hafta adımı sayısı' üçüncü en önemli tahmin özellikler olmuşlardır.

Karar Ağacı modelinin karmaşıklık matrisini gösteren Tablo 5.10'a göre, algoritma tüm İngilizce dil gruplarının kurs sonu performansı tahmininde 'tamamlayıcılar' değerini alan katılımcılarda en iyi tahmin hata yüzdesine sahipken, 'yavaş tempolu katılımcılar' değerini alan katılımcılarda en büyük tahmin hata yüzdesine sahiptir.

Tablo 5.10 İngilizce dil gruplarında Karar Ağacı algoritmasının karmaşıklık matrisleri - "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu

Tüm İngilizce Dil Grubu Bilinenler	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	2.325	117	228
Tamamlayıcılar	58	2.238	11
Bırakanlar	16	44	2.160
İngilizce Birincil ve Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	468	15	79
Tamamlayıcılar	23	472	9
Bırakanlar	5	9	408
İngilizce Yalnızca Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	187	15	10
Tamamlayıcılar	5	183	2
Bırakanlar	10	4	190
İngilizce İkincil Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	1.563	70	137
Tamamlayıcılar	40	1.507	7
Bırakanlar	6	32	1.465

FutureLearn platformuna ait "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda, kurs sonu performansı tahmininde Karar Ağacı algoritması en iyi tahmin doğruluğuna sahip olmuştur.

5.1.3 Farklı Türden Kurslarda (MOOC'lar) İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri

"Understanding Language : Learning and Teaching" MOOC serisinden farklı türde (Sosyal Bilimler, Biyoloji, Doğa Bilimi gibi) MOOC'larda "ikincil dil olarak İngilizce" konuşan katılımcıların karşılaşacağı zorluk seviyesi farklı olabilir. Örneğin, Matematik kursunda daha az İngilizce dil zorluğu ile karşı karşıya kalabilirler. Bu sebeple, hem İngilizce dil bilgisinin farklı kurslarda katılımcı başarısındaki önemini hem de tahmin modellerinin farklı kurslardaki kurs sonu performansı tahmini başarılarını gözlemlemek için FutureLearn platformuna ait "Exploring Our Oceans 4" ve "Web Science 1" kursları da veri kaynağı olarak kullanılmıştır.

Kurs sonu performans analizleri ile katılımcıların hangi kurslardaki İngilizce performanslarının daha yüksek başarı ile tahmin edilebildiği ve İngilizce dil gruplarında kurs sonu performanslarının kurs içeriğine bağlı olarak değişimi ortaya konulmak istenilmiştir.

"Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında İngilizce dil gruplarının kurs sonu performansını gösteren özelliğin aldığı değerlerin sayıları ve yüzdeleri Tablo 5.11 ve Tablo 5.12'de gösterilmektedir. Tablo 5.11 ve Tablo 5.12'ye göre "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında İngilizce dil grubu bilinen katılımcıların kurs sonu performansı özelliğinin aldığı değerlerin dağılımları dengesizdir. "Understanding Language : Learning and Teaching" MOOC serisi kursları ile benzer şekilde "birincil ve resmi dil olarak İngilizce" konuşan katılımcılarda, kurs tamamlama özelliğinin aldığı değerlerin dağılımı yönünden diğer İngilizce dil gruplarına göre belirgin farklılıklar vardır. İngilizce dil grupları içinde, "birincil ve resmi dil olarak İngilizce" konuşanlarda 'tamamlayıcılar' değerini alanların yüzdesi en yüksek iken 'bırakanlar' değerini alanların yüzdesi en düşüktür ("birincil ve resmi dil olarak İngilizce" konuşan katılımcıların kursu tamamlama ihtimali daha yüksek iken diğer İngilizce dil gruplarındaki katılımcılarda kursu bırakma ihtimali daha yüksektir). Tablo 5.11 ve Tablo 5.12'ye göre İngilizce dil gruplarında yer alan katılımcı sayıları da önemli miktarda değişiklik göstermektedir

Tablo 5.11 Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Exploring Our Oceans 4" kursu

	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil	Tüm İngilizce Dil Grubu Bilinenler
Bırakanlar	114 (% 42,70)	24 (% 50)	93 (% 56,02)	231 (% 48,02)
Yavaş tempolu katılımcılar	53 (% 19,85)	14 (% 29,17)	33 (% 19,88)	100 (% 20,79)
Tamamlayıcılar	100 (% 37,47)	10 (% 20,83)	40 (% 24,10)	150 (% 31,19)
Toplam	267	48	166	481

Tablo 5.11 ve Tablo 5.12'ye göre "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında, 'yavaş tempolu katılımcılar' değerlerine sahip olan katılımcı sayıları en düşük (100 ve 89) olduğu için tüm İngilizce dil grubu bilinen katılımcıların verisinden her kurs performansı özelliği değeri için "Exploring Our Oceans 4" kursunda 100'er ve "Web Science 1" 89'ar katılımcı rastgele olarak seçilmiştir. Aynı yöntemle, her bir kurs performansı kategorisi için (kursu bırakma, yavaş tempolu katılımcılar, tamamlayıcılar), birincil ve resmi dili İngilizce olan katılımcıların verisinden "Exploring Our Oceans 4" kursunda 53'er ve "Web Science 1" kursunda 71'er, sadece resmi dili İngilizce olan katılımcıların verisinden "Exploring Our Oceans 4" kursunda 10'ar ve "Web Science 1" kursunda 5'er, "ikincil dil olarak İngilizce" konuşan katılımcıların verisinden "Exploring Our Oceans 4" kursunda 33'er ve "Web Science 1" kursunda 12'şer katılımcı aşağı örnekleme ile seçilmiştir.

Tablo 5.12 Veriyi dengelemeden önce İngilizce dil gruplarında kurs tamamlama özelliğinin değerlerine göre katılımcı sayıları ve yüzdeleri - "Web Science 1" kursu

	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil	Tüm İngilizce Dil Grubu Bilinenler
Bırakanlar	106 (% 34,87)	9 (% 45)	41 (% 60,29)	156 (% 39,80)
Yavaş tempolu katılımcılar	71 (% 23,36)	6 (% 30)	12 (% 17,65)	89 (% 22,70)
Tamamlayıcılar	127 (% 41,78)	5 (% 25)	15 (% 22,06)	147 (% 37,50)
Toplam	304	20	68	392

Tablo 5.13 ve Tablo 5.14'e göre "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında, denenen her algoritma için kurs sonu performansı tahmini

doğruluk sonuçlarını göstermektedir. Sonuçlara göre, İngilizce dil gruplarının kurs performansı tahmin başarılarının ortalamaları alındığında, "Exploring Our Oceans 4" kursunda Rastgele Orman ve "Web Science 1" kursunda Destek Vektör Makinesi algoritmaları, kurs performansı tahmininde en iyi doğruluk yüzdelere sahip olmuşlardır. "Exploring Our Oceans 4" kursunda Naive Bayes ve "Web Science 1" kursunda Çok Katmanlı Algılayıcı algoritmaları, çoğunlukla en kötü performansı tahmini sonuçlarına sahip olmuşlardır. "Exploring Our Oceans 4" kursunda Naive Bayes algoritması dışındaki algoritmaların tahmin başarıları genel olarak iyi olurken "Web Science 1" kursunda en başarılı algoritma olan Destek Vektör Makinesi algoritması ile diğer algoritmaların tahmin başarı sonuçları arasında belirgin farklar oluşmuştur.

Tablo 5.13 Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Exploring Our Oceans 4" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Çok Katmanlı Algılayıcı	0,873	0,868	0,933	0,818
Karar Ağacı	0,957	0,962	0,667	0,970
K-En Yakın Komşu	0,907	0,881	0,900	0,939
Çok Kategorili Lojistik Regresyon	0,920	0,943	1	0,939
Naive Bayes	0,823	0,774	0,733	0,919
Rastgele Orman	0,953	0,943	0,867	0,960
Destek Vektör Makinesi	0,943	0,950	0,800	0,929

Tablo 5.14 Kurs performansı tahmininde sık kullanılan geleneksel algoritma için tahmin doğruluk değerleri - "Web Science 1" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Çok Katmanlı Al- gılayıcı	0,640	0,615	0,800	0,667
Karar Ağacı	0,8202	0,831	0,333	0,861
K-En Yakın Komşu	0,802	0,817	0,600	0,722
Çok Katego- rili Lojistik Regresyon	0,779	0,803	1	0,833
Naive Bayes	0,734	0,751	0,667	0,833
Rastgele Orman	0,817	0,803	0,933	0,694
Destek Vektör Makinesi	0,948	0,958	1	0,972

Tablo 5.15 ve Tablo 5.16, "Exploring Our Oceans 4" MOOC'unda Rastgele Orman ve "Web Science 1" MOOC'unda Destek Vektör Makinesi modelleri ile kurs sonu performansı tahmininde Tutarlılık, Anma ve F-ölçüm metriklerinin aldıkları değerleri göstermektedir. Bu değerlere göre de, uygulanan modeller içinde "Exploring Our Oceans 4" MOOC'unda Rastgele Orman ve "Web Science 1" MOOC'unda Destek Vektör Makinesi modelinin en iyi kurs sonu performansı tahmini doğruluk yüzdelere sahip oldukları anlaşılmaktadır.

Tablo 5.15 Rastgele Orman algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Exploring Our Oceans 4" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Tutarlılık	0,955	0,945	0,867	0,970
Anma	0,953	0,943	0,867	0,970
F-ölçüm	0,9539	0,944	0,867	0,970

Tablo 5.16 Destek Vektör Makinesi algoritması ile kurs sonu performansı tahmininde Tutarlılık, Anma, F-ölçüm metriklerinin aldıkları değerler - "Web Science 1" kursu

	Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
Tutarlılık	0.948	0.958	1	0.974
Anma	0.948	0.958	1	0.972
F-ölçüm	0.948	0.958	1	0.973

Tablo 5.17 ve Tablo 5.18, "Exploring Our Oceans 4" ve "Web Science- 1" MOOC'larında kurs sonu performansı tahmininde en iyi tahmin sonuçlarına sahip algoritmalar için en önemli üç tahmin özelliğini göstermektedir. Kurs sonu performansı tahmininde en önemli üç özellik, "Exploring Our Oceans 4" MOOC'unda, en iyi tahmin sonucuna sahip Rastgele Orman algoritması için yalnızca resmi dil olarak İngilizce konuşan katılımcı grubu dışındaki tüm İngilizce dil gruplarında 1) baslanan adım sayısı, 2) tamamlanan ilk hafta adım sayısı ve 3) tamamlanan üçüncü hafta adım sayısı olurken, yalnızca resmi dil olarak İngilizce konuşan katılımcı grubunda ise 'tamamlanan üçüncü hafta adım sayısı' en önemli üç tahmin özelliği içinde yer almamış ve 'tamamlanan ikinci hafta adım sayısı' en önemli üçüncü tahmin özelliği olmuştur. "Web Science- 1" MOOC'unda ise, en iyi tahmin sonucuna sahip Destek Vektör Makinesi algoritması için en önemli üç tahmin özelliği ve sıralaması yalnızca resmi dil olarak İngilizce konuşan katılımcı grubu dışındaki tüm İngilizce dil gruplarında 1) ana yorum uzunluk ortalaması, 2) yorum uzunluk ortalaması ve 3) başlanan adım sayısı olurken, yalnızca resmi dil olarak İngilizce konuşan katılımcı grubunda ise 'yorum uzunluk ortalaması' en önemli tahmin özelliği, 'ana yorum uzunluk ortalaması' ise en önemli ikinci tahmin özelliği olmuştur.

Tablo 5.17 Rastgele Orman algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Exploring Our Oceans 4" kursu

Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. baslanan adım sayısı 2. tamamlanan ilk hafta adım sayısı 3. tamamlanan üçüncü hafta adım sayısı	1. baslanan adım sayısı 2. tamamlanan ilk hafta adım sayısı 3. tamamlanan üçüncü hafta adım sayısı	1. baslanan adım sayısı 2. tamamlanan ilk hafta adım sayısı 3. tamamlanan ikinci hafta adım sayısı	1. baslanan adım sayısı 2. tamamlanan ilk hafta adım sayısı 3. tamamlanan üçüncü hafta adım sayısı

Tablo 5.18 Destek Vektör Makinesi algoritması ile kurs sonu performansı tahmininde en önemli olan üç özellik - "Web Science 1" kursu

Tüm İngilizce Dil Grupları	İngilizce Birincil ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. yorum uzunluk ortalaması 2. ana yorum uzunluk ortalaması 3. başlanan adım sayısı	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı

"Exploring Our Oceans 4" MOOC'unda, Rastgele Orman modelinin ve "Web Science 1" MOOC'unda ise, Destek Vektör Makinesi modelinin karmaşıklık matrisini gösteren Tablo 5.19 ve Tablo 5.20'ye göre, Rastgele Orman algoritması İngilizce dil gruplarının kurs sonu performansı tahmininde genellikle 'yavaş tempolu katılımcılar' kurs performansı değerini alan katılımcılarda ve Destek Vektör Makinesi algoritması ise 'bırakanlar' kurs performansı değerini alan katılımcılarda en büyük tahmin hata yüzdesine sahiptir. İki algoritmada en iyi tahmin doğruluk sonucunu tüm İngilizce dil gruplarının verilerinde 'tamamlayıcılar' kurs performansı değerini alan katılımcıların tahmininde sahip olmuşlardır.

Tablo 5.19 İngilizce dil gruplarında Rastgele Orman algoritmasının karmaşıklık matrisleri - "Exploring Our Oceans 4" kursu

Tüm İngilizce Dil Grubu Bilinenler	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	96	1	9
Tamamlayıcılar	1	99	0
Bırakanlar	3	0	91
İngilizce Birincil ve Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	50	1	5
Tamamlayıcılar	0	52	0
Bırakanlar	3	0	48
İngilizce Yalnızca Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	7	0	1
Tamamlayıcılar	1	10	0
Bırakanlar	2	0	9
İngilizce İkincil Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	32	0	2
Tamamlayıcılar	0	33	0
Bırakanlar	1	0	31

Tablo 5.20 İngilizce dil gruplarında Destek Vektör Makinesi algoritmasının karmaşıklık matrisleri - "Web Science 1" kursu

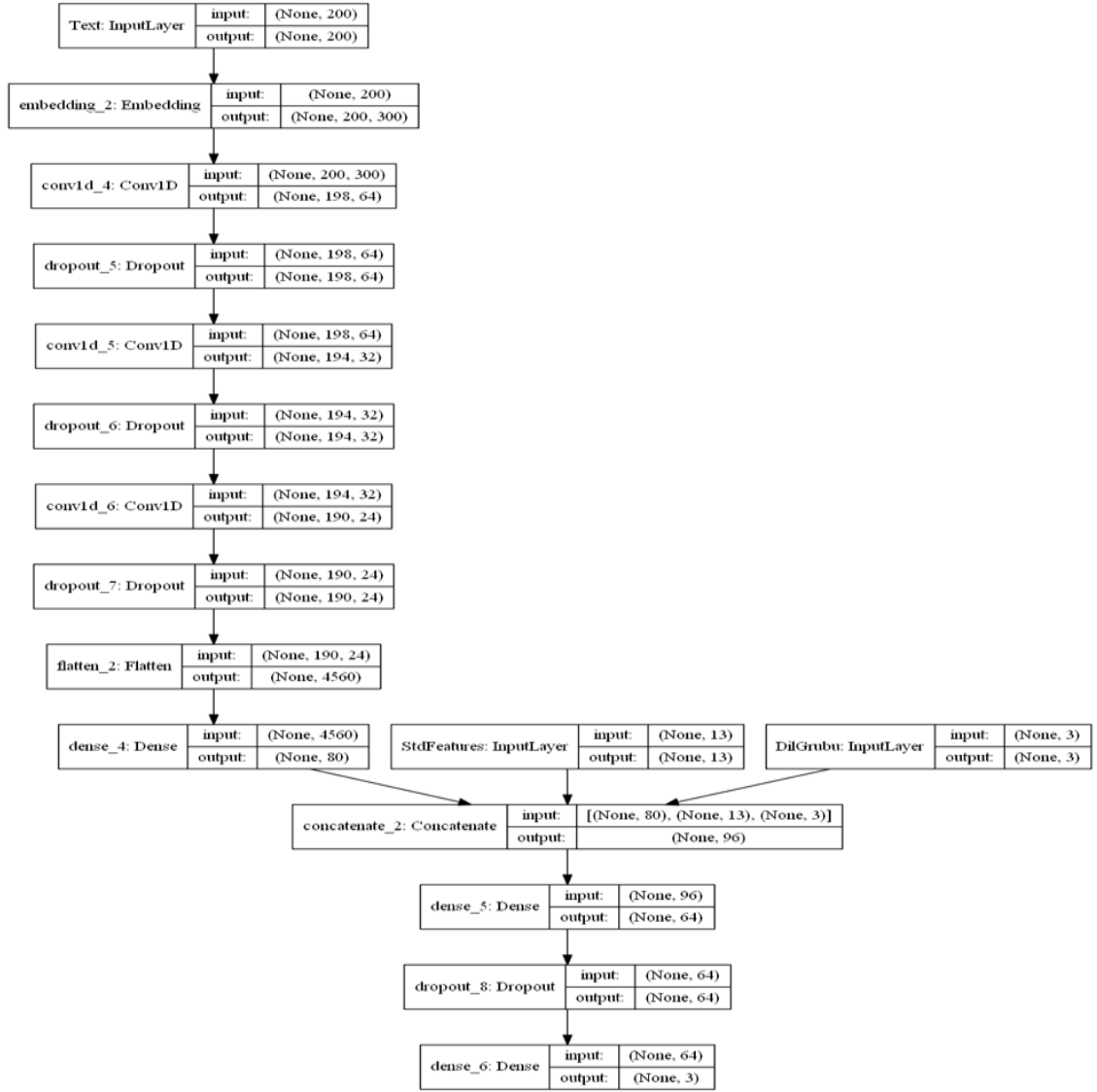
Tüm İngilizce Dil Grubu Bilinenler	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	85	0	9
Tamamlayıcılar	0	89	1
Bırakanlar	4	0	79
İngilizce Birincil ve Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	66	0	4
Tamamlayıcılar	0	71	0
Bırakanlar	5	0	67
İngilizce Yalnızca Resmi Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	5	0	0
Tamamlayıcılar	0	5	0
Bırakanlar	0	0	5
İngilizce İkincil Dil	Yavaş tempolu katılımcılar	Tamamlayıcılar	Bırakanlar
Yavaş tempolu katılımcılar	12	0	1
Tamamlayıcılar	0	12	0
Bırakanlar	0	0	11

5.2 Derin Öğrenme İle İngilizce Dil Gruplarının Kurs Sonu Performansı Tahminlerinin Gerçekleştirilmesi

Derin Öğrenme ile İngilizce dil gruplarının kurs sonu performansı tahmininde Bölüm 3'te yer verilip, Bölüm 4'te oluşturulma adımları açıklanan “oluşturulan standart özellikler” ile Bölüm 3'te açıklanan ön işlemler sonucu elde edilen katılımcı yorumları kullanılmıştır. Derin Öğrenme ile kurs sonu performansı tahmininde, İngilizce dil grubu tahmininden farklı olarak, İngilizce dil grubu bilgisi tahmin edici özellik olarak, kurs sonu performansı özelliği ise tahmin edilen (hedef) özellik olarak belirlenmiştir.

Derin Öğrenme ile kurs sonu performansı tahmini için kullanılan veri türüne göre üç farklı model ile gerçekleştirilmiştir. İlk modelde “yorumlar ve oluşturulan standart özellikler” kullanılırken, ikinci modelde “sadece yorumlar”, üçüncü modelde ise “sadece oluşturulan standart özellikler” modelin eğitiminde kullanılmıştır.

Şekil 5.1, kurs sonu performansı tahmini için oluşturulan özellikler, İngilizce dil grubu ve ön işlenmiş yorum verileri ile birlikte çalışan Derin Öğrenme modelimizi (Konvolüsyon mimarisi kullanıldığı durumda) göstermektedir.

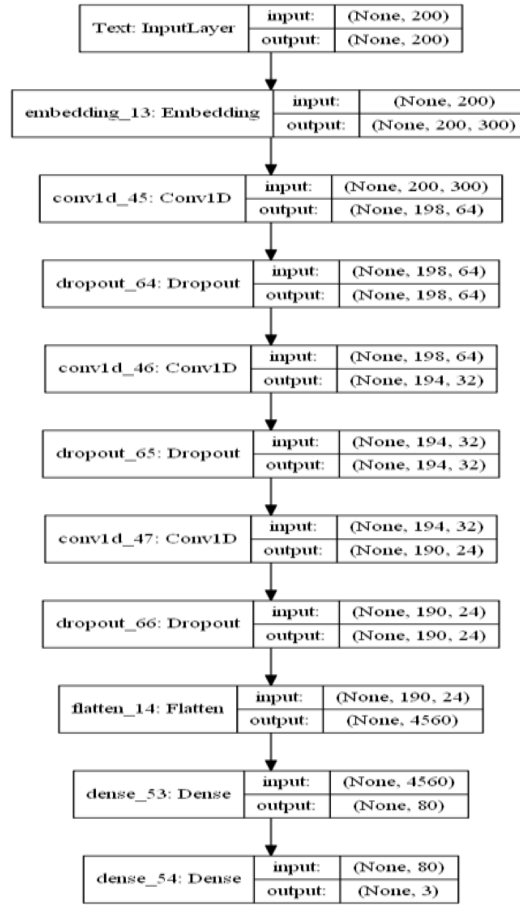


Şekil 5.1 Oluşturulan standart özellikler ve yorumlar ile kurs sonu performansı tahmini için Derin Öğrenme modeli

Şekil 5.1’de yer verilen “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmini Derin Öğrenme modeli, Bölüm 3’te yer verilen İngilizce dil grubu tahmini modelinde de olduğu gibi öncelikle Giriş (Input) katmanı ile yorumları girdi olarak almaktadır. Ardından Evrişim (Convolution) katmanlarını kullanarak “sadece yorumlarla” eğitilmektedir (Birinci Evrişim katmanı art arda üç komşu kelimeyi aramak, ikincisi beş komşu kelimeyi aramak, üçüncüsü sırasıyla yedi komşu kelimeyi aramak için kullanılmıştır). Her Evrişim katmanından sonra, daha iyi öğrenme için birer tane Bırakma (Drop-out) katmanı yer almaktadır. Daha sonra, Düzleştirme (Flatten) katmanı ile sonuçlar tek boyutlu hale getirilmekte ve Yoğun (Dense) katmanı ile modelin ikinci kısmında (oluşturduğumuz standart özelliklerle birlikte) tekrar girdi olarak kullanılacak bir çıktı üretmektedir.

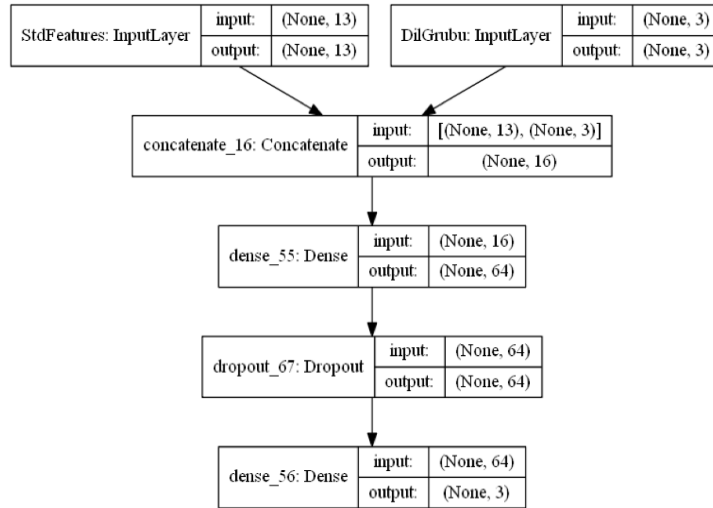
Şekil 5.1’deki modelin ikinci kısmında üç girdi alınmıştır: Modelin ilk kısmının çıktısı, Bölüm 3’te yer verilen “oluşturulan standart özellikler” ve her katılımcının İngilizce dil grubu özelliği (kurs öncesi anket ve düzenli ifadelerle tespit edilen). Model, bu üç girişi Birleştirme (Concatenate) katmanında birleştirmektedir. Modelin ikinci kısmı, Yoğun (Dense) ve Bırakma (Drop-out) katmanları ile eğitimden sonra, tespit edilen kurs sonu performansı değerlerini çıktı olarak vermektedir.

Şekil 5.2, sadece yorumlar ile kurs sonu performansı tahmini Derin Öğrenme modelimizi (Konvolüsyon mimarisi kullanıldığı durumda) göstermektedir.



Şekil 5.2 Sadece yorumlarla kurs sonu performansı tahmini İçin Derin Öğrenme modeli

Şekil 5.3, “sadece oluşturulan standart özellikler” ile kurs sonu performansı tahmini Derin Öğrenme modelimizi göstermektedir.



Şekil 5.3 Sadece oluşturulan standart özelliklerle kurs sonu performansı tahmini için Derin Öğrenme modeli

5.3 Derin Öğrenme Modellerinde Eğitim ve Test Senaryoları

Şekil 5.4’de Derin Öğrenme ile kurs sonu performansı tahmininde eğitimde kullanılan veriler, testte kullanılan kurslar ve denenen model ve mimariye göre değişiklik gösteren dört farklı senaryoya yer verilmiştir. Senaryolarda kullanılan modellerin yapısı eğitimde kullanılan verilere göre değişmektedir. İlk üç senaryoda veri olarak “sadece yorumlar” veya “yorumlar ve oluşturulan standart özellikler” kullanılmıştır. Dördüncü senaryoda ise veri olarak “sadece oluşturulan standart özellikler” kullanılmıştır. Tüm senaryolarda eğitim için "Understanding Language: Learning and Teaching" MOOC serisinin ilk yedi iterasyonuna ait veriler (Derin Öğrenme daha büyük veri kümeleriyle daha iyi performans gösterdiği için) kullanılmıştır. Aynı zamanda, ilk üç senaryo, iki ayrı model (eğitim için kullanılan verilere göre) ve beş farklı mimari (Konvolüsyon, LSTM, Bidirectional LSTM, GRU ve Konvolüsyon+LSTM) denenerek oluşturulmuştur. Bununla birlikte, dördüncü senaryoda eğitim için oluşturulan sadece standart özellikler kullanıldığı için tek bir model ve mimari yapısı kullanılmıştır.

	Eğitim Verisi	Test Verisi	Özellikler	
Senaryo 1	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon	Yorumlar Yorumlar ve Oluşturulan Özellikler	Tamamlandı Yavaş İlerleyen Bıraktı
Senaryo 2	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon	LMOOC Sekizinci iterasyon	Yorumlar Yorumlar ve Oluşturulan Özellikler	Tamamlandı Yavaş İlerleyen Bıraktı
Senaryo 3	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon	Ocean MOOC Web Science MOOC	Yorumlar Yorumlar ve Oluşturulan Özellikler	Tamamlandı Yavaş İlerleyen Bıraktı
Senaryo 4	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon Aşağıdaki İngilizce Dil Gruplarının Verileri Oluşturuldu Tüm Dil Grubu Bilinenler Ana ve Resmi Dil Sadece Resmi Dil İkincil Dil	LMOOC Birinci iterasyon ----- LMOOC Yedinci iterasyon LMOOC Sekizinci iterasyon Ocean MOOC Web Science MOOC Aşağıdaki İngilizce Dil Gruplarının Verileri Oluşturuldu Tüm Dil Grubu Bilinenler Ana ve Resmi Dil Sadece Resmi Dil İkincil Dil	Oluşturulan Özellikler	Tamamlandı Yavaş İlerleyen Bıraktı

Şekil 5.4 Kullanılan test verisi ve denenen mimarilere göre farklı kurs sonu performansı tahmini senaryoları

Derin Öğrenme modellerinden yorum odaklı olarak elde edilen tahmin sonuçları kullanıcı odaklı hale getirilerek (kurs sonu performansı göre) hem diğer makine öğrenme yöntemleri ile karşılaştırılabilir hem de daha yüksek başarı alınacak hale getirilmiştir. Sonuçların kullanıcı odaklı hale getirilmesi adımları:

1. Model uygulandığında elde edilen üç değer içeren vektör sonucu, “0”, “1”, “2” değerlerinden birini alacak şekilde dönüştürülmüştür.
2. Yorum yapan kullanıcıların yorumlarının sayısı 'n' olmak üzere, en az $\lceil n/2 \rceil$ kadar yorum aynı kurs sonu performans değerine sahip ise o katılımcı doğru

gruplandırılmış, değil ise yanlış gruplandırılmış olarak kabul edilmiştir.

Şekil 5.4'de yer alan birinci senaryoda, "Understanding Language : Learning and Teaching" MOOC serisinin ilk yedi iterasyonunda yorum yapan katılımcılardan rastgele seçim ile % 80'i eğitim verisi ve % 20'si de test verisi olarak kullanılmıştır.

Şekil 5.4'de yer verilen ikinci senaryoda, test verisi olarak eğitimde kullanılan MOOC serisinin sekizinci iterasyonu olan "Understanding Language 8 : Learning and Teaching" kursunun yorum yapan katılımcılara ait veriler kullanılmıştır.

Şekil 5.4'de yer verilen üçüncü senaryoda, oluşturulan Derin Öğrenme modellerinin, katılımcıların İngilizce yazma yeterliliğinde daha belirgin farklılıklar ve değişkenliklere sahip olduğu kurslarda daha iyi sonuç vereceği düşünüldüğü için, farklı kurslardaki kurs sonu performans tahmin başarılarındaki etkilerini anlamak için, test verisi olarak eğitimde kullanılan MOOC serisinden farklı türde iki kurs kullanılmıştır. Test için kullanılan kurslardan ilki, altı haftalık fen bilimleri MOOC'u olan "Web Science 1", ikincisi ise dört haftalık doğa bilimleri MOOC'u olan "Exploring Our Oceans 4"dur.

Şekil 5.4'de yer verilen dördüncü senaryoda ilk üç senaryodan farklı olarak hem eğitimde hem de testte “sadece oluşturulan standart özellikler” kullanılmıştır. Dördüncü senaryoda, “sadece oluşturulan standart özellikler” üzerinden ilk üç senaryoda denenilen yöntemlerin tamamı (aynı MOOC serisini % 80'e % 20 ayırarak eğitim ve test, aynı MOOC serisinin farklı iterasyonu üzerinde test, farklı türden MOOC'lar üzerinde test) denenmiştir.

5.3.1 Birinci Senaryo: Aynı Kurs (MOOC) Serisi İle Eğitim ve Test

Tablo 5.21'de yer alan sonuçlara göre, aynı kurs serisinin % 20 lik bölümü test verisi olarak kullanıldığında “sadece yorumlarla” kurs sonu performansı tahmin edildiğinde Evrişim (Konvolüsyon) mimarisi ile en yüksek kurs sonu performansı tahmini doğruluğu gerçekleşirken, “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmin edildiğinde GRU mimarisi ile en yüksek kurs sonu performansı tahmini doğruluğu gerçekleşmiştir.

Tablo 5.21 Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri (birinci senaryo)

	Sadece Yorumlarla (%)	Yorumlar ve Oluşturulan Standart Özelliklerle (%)
Konvolüsyon	68,05	82,95
LSTM	61,62	85,92
Bidirectional LSTM	60,76	84,62
GRU	60,49	87,05
Konvolüsyon ve LSTM	66,99	81,70

5.3.2 İkinci Senaryo: Aynı Kurs (MOOC) Serisinin Farklı İterasyonlarının Eğitim ve Test Verileri Olarak Kullanımı

Tablo 5.22’de yer alan sonuçlara göre, aynı kurs serisinin sekizinci sürümü test verisi olarak kullanıldığında, “sadece yorumlarla” kurs sonu performansı tahmininde Evrişim (Konvolüsyon) mimarisi ile en yüksek doğrulukla kurs sonu performansı tahmini gerçekleştirilirken, “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmininde GRU mimarisi ile en yüksek kurs sonu performansı tahmini doğruluğu elde edilmiştir.

Tablo 5.22 Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri (ikinci senaryo)

	Sadece Yorumlarla (%)	Yorumlar ve Oluşturulan Standart Özelliklerle (%)
Konvolüsyon	55,88	69,91
LSTM	51,36	76,92
Bidirectional LSTM	48,19	75,79
GRU	51,58	84,62
Konvolüsyon ve LSTM	52,71	72,62

5.3.3 Üçüncü Senaryo: Farklı Türden Kursların (MOOC’lar) Test Verisi Olarak Kullanımı

Üçüncü senaryoda, test için kullanılan MOOC’lardan doğa bilimleri MOOC’u olan "Exploring Our Oceans 4" MOOC’u, "Understanding Language : Learning and Teaching" kurs serisi MOOC’ları gibi dört haftalık bir MOOC iken "Web Science 1" MOOC’u altı haftalık bir MOOC’dur. Bu farklılıktan ötürü "Web Science 1" MOOC’unun

ilk dört haftasına ait aktivite kullanım verileri kurs sonu performansı tahmininde kullanılmıştır.

Tablo 5.23 Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri - "Exploring Our Oceans 4" (üçüncü senaryo)

	Sadece Yorumlarla (%)	Yorumlar ve Oluşturulan Standart Özelliklerle (%)
Konvolüsyon	57,69	74,04
LSTM	66,35	74,04
Bidirectional LSTM	73,08	75,96
GRU	68,27	77,89
Konvolüsyon ve LSTM	63,46	71,15

Tablo 5.23’de yer alan sonuçlara göre, 'Exploring Our Oceans 4' kursu test verisi olarak kullanıldığında, “sadece yorumlarla” kurs sonu performansı tahmininde iki yönlü (Bidirectional) LSTM mimarisi ile en yüksek doğrulukla kurs sonu performansı tahmini gerçekleştirilirken, “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmininde GRU mimarisi ile en yüksek kurs sonu performansı tahmini doğruluğu elde edilmiştir.

Tablo 5.24 Farklı Derin Öğrenme modelleri ile kurs sonu performansı tahmini doğruluk yüzdeleri - "Web Science 1" (üçüncü senaryo)

	Sadece Yorumlarla (%)	Yorumlar ve Oluşturulan Standart Özelliklerle (%)
Konvolüsyon	57,79	70,93
LSTM	58,82	74,39
Bidirectional LSTM	57,79	73,70
GRU	59,52	79,24
Konvolüsyon ve LSTM	55,02	71,28

Tablo 5.24’de yer alan sonuçlara göre, "Web Science 1" kursu test verisi olarak kullanıldığında, hem “sadece yorumlarla”, hem de “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmininde GRU mimarisi ile en yüksek kurs sonu performansı tahmini doğruluğu elde edilmiştir. Hem “sadece yorumlar” ile hem de “yorumlar ve oluşturulan standart özelliklerle” kurs sonu performansı tahmininde üçüncü senaryoda denenen "Exploring our Oceans 4" (doğa bilimleri kursu) ve "Web Science 1" (fen bilimleri kursu) kursunda, Konvolüsyon ve LSTM hibrit mimarisi

denenen mimariler içinde en kötü kurs sonu performansı tahmini sonucunu vermiştir.

5.3.4 Dördüncü Senaryo: Sadece Oluşturulan Standart Özelliklerin Kullanımı

Tablo 5.25, "Understanding Language : Learning and Teaching" MOOC'unun, ilk yedi iterasyonunda katılımcıların oluşturulan standart özellikleri ile Derin Öğrenme modellerinin eğitilerek aynı kurs verileri (% 80 eğitim, % 20 test olarak ayrılarak), aynı kurs serisinin sekizinci sürümü ardından da "Exploring our Oceans 4" (doğa bilimleri kursu) ve "Web Science 1" (fen bilimleri kursu) MOOC kurslarının veri ile test edilerek alınan başarıyı göstermektedir.

Tablo 5.25 Oluşturulan standart özellikler kullanılarak Derin Öğrenme ile kurs sonu performansı tahmini doğruluk yüzdeleri (dördüncü senaryo)

	Tüm İngilizce Dil Grupları	Ana ve Resmi Dil	Sadece Resmi Dil	İkincil Dil
Aynı MOOC Serisi Üzerinde Eğitim ve Test (%)	83,47	84,31	84,48	84,44
Eğitimde Kullanılan MOOC Serisinin Farklı İterasyonu Üzerinde Test (%)	83,78	84,72	84,84	84,55
Eğitimde Kullanılan MOOC Serisinden Farklı Türde MOOC Üzerinde Test - "Exploring Our Oceans 4" (%)	83,78	84,02	84,84	84,55
Eğitimde Kullanılan MOOC Serisinden Farklı Türde MOOC Üzerinde Test - "Web Science 1" (%)	83,51	84,72	84,04	83,60

5.4 Sonuçlar

- Geleneksel makine öğrenme algoritmaları ile kurs sonu performans tahmininde genellikle, Rastgele Orman ve Karar Ağacı algoritmaları en yüksek doğruluğa sahip olurken, Web Science-1 kursunda, Destek Vektör Makinesi algoritması en başarılı algoritma olmuştur. Naive Bayes ve Çok Katmanlı algılayıcı algoritmaları ise en başarısız iki algoritma olmuştur.
- Geleneksel makine öğrenme algoritmaları ile en başarılı kurs sonu performansı tahmin sonuçları; "Understanding Language : Learning and Teaching" kurs

serisinde ikincil dil olarak İngilizce konuşanlar için, diğer MOOC'larda ise sadece resmi dil olarak İngilizce konuşanlar için elde edilmiştir.

- 'Başlanan adım sayısı', 'ilk hafta tamamlanan adım sayısı' ve 'ikinci hafta tamamlanan adım sayısı' özellikleri, Web Science-1 kursu dışında denenen tüm MOOC'larda en önemli kurs sonu performans tahmini özelliklerinden olurken, Web Science-1 kursunda, İngilizce dil gruplarının tamamında en önemli ilk iki özellik, yorumlarla ilgili özellikler olmuştur.
- Geleneksel makine öğrenme algoritmaları ve oluşturulan Derin öğrenme modelleri (farklı mimarilerle) denenerek kurs sonu performans tahmini başarısı iyileştirilmeye çalışılmıştır. Rastgele Orman ve Karar Ağacı gibi yüksek başarı gösteren geleneksel algoritmalar, önerilen Derin Öğrenme modellerine göre daha iyi sonuç verirken, Naive Bayes ve Çok Katmanlı algılayıcı algoritmaları ise daha kötü sonuç vermişlerdir.
- Geleneksel makine öğrenme algoritmaları ile kurs sonu performansı tahmininde, rastgele aşağı örnekleme yöntemi denenmiştir. Derin Öğrenme ile kurs sonu performansı tahmininde ise üç farklı örnekleme yöntemi (rastgele aşağı örnekleme, rastgele ortalama örnekleme, rastgele aşırı örnekleme) denenmiş ve en başarılı sonuçlara yer verilmiştir.
- Derin Öğrenme modelleri ile kurs sonu performansı tahmininde, denenen beş farklı mimariden GRU mimarisi genel olarak "yorumlar ve oluşturulan standart özellikler" kullanıldığında en başarılı tahmin sonuçlarına sahip olmaktadır. "Sadece yorumlar" ile kurs sonu performansı tahmininde "Understanding Language : Learning and Teaching" kurs serisinin hem eğitim hem de testte kullanıldığı ilk iki senaryoda Evrişim (Konvolüsyon) en başarılı tahmin doğruluğuna sahipken "Exploring Our Oceans 4" kursu test verisi olarak kullanıldığında İki Yönlü (Bidirectional) LSTM mimarisi ile "Web Science 1" kursu test verisi olarak kullanıldığında ise GRU mimarisi ile en başarılı tahmin sonuçları elde edilmiştir.
- Derin Öğrenme modelleri ile kurs sonu performansı tahmininde, İngilizce dil grubu tahmininden farklı olarak eğitim ve test için kullanılan veriler değiştirildiğinde tahmin başarı yüzdelerinde fazla düşüş olmamıştır. "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonunun %80 eğitim ve %20 test verisi olarak kullanıldığı ilk senaryoda tahmin başarı yüzdeleri biraz daha yüksek olmakla birlikte, ikinci ve üçüncü senaryolar arasında tahmin başarı sonuçları birbirine oldukça yakın olmuştur.

- Derin Öğrenme modelleri ile kurs sonu performansı tahmininde kullanılan veri türüne göre en başarılı tahmin sonuçlar “sadece oluşturulan standart özellikler” kullanıldığında elde edilirken, ikinci sırada “yorumlar ve oluşturulan standart özellikler” kullanılarak gerçekleştirilen tahminlerde elde edilmiştir. Sonuçlara göre ‘sadece yorumların’ kullanılması kurs sonu performansı tahmini için yetersiz kalmaktadır.

Bir sonraki bölümde (Bölüm 6), İngilizce dil gruplarının kurs boyunca performansları için önemli olan özellikleri ve bunların değişimlerini belirlemek, aynı zamanda da problemleri erkenden öngörebilmek adına performansı tahmininde sık kullanılan yedi geleneksel makine öğrenme algoritması, ilk haftadan başlayarak haftalık verilerle kümülatif olarak (ilk hafta verileri ile, ilk iki hafta verileri ile vb.) kurs sonu performansı tahmini gerçekleştirilmiştir.

İngilizce Dil Gruplarının Haftalık Kurs Performans Analizi

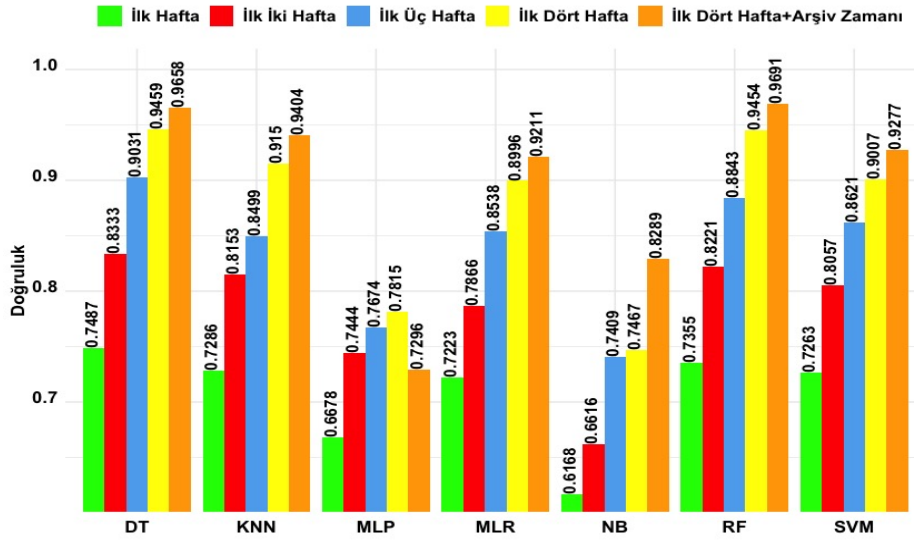
Haftalık kurs performans analizlerinde, Bölüm 5'te (İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri) yer verilen yedi geleneksel algoritma ve kümülatif haftalık veriler ile (ilk hafta, ilk iki hafta gibi) kurs performans tahminleri gerçekleştirilmiştir.

6.1 Veri Önışleme

Haftalık kurs performans tahmininde, Bölüm 5 (İngilizce Dil Gruplarının Kurs Sonu Performans Analizleri) ile aynı özellikler kullanılmış olmakla birlikte, kümülatif olarak ilgili haftaların kullanım verilerini elde etmek için zaman filtreleri uygulanmıştır. "İlk hafta" verileri için kursun başlangıç zamanından itibaren ilk yedi gün, "ilk iki hafta" verileri için ilk 14 gün, "ilk üç hafta" için ilk 21 gün, "ilk dört hafta" için ilk 28 gün, "ilk dört hafta +14 günlük arşiv zamanı" için kursun yayında kaldığı sürede gerçekleştirilen tüm aktiviteler dikkate alınmıştır. Örneğin, bir katılımcı, kursun tamamında 20 aktiviteye başlarken ve bunlardan dördüne "ilk hafta" başladı ise, kurs sonuna kadar olan aktiviteleri içeren veride başladığı aktivite sayısı özelliği 20 değerini alırken, aynı özellik "ilk hafta" sonuna kadar olan veride dört değerini alacaktır. Bir diğer örnek ise, bir katılımcı "ilk hafta"ya ait beş kurs adımı tamamlarken, bu adımlardan üçünü "ilk hafta" (kursun yayınlanma zamanından itibaren ilk yedi gün) içinde, kalan ikisini de kursun kalanında tamamlamış ise, "ilk hafta"ya ait veride 'tamamlanan ilk hafta adımı sayısı' özelliği üç değerini alırken, kursun tamamının aktivitelerini içeren veride aynı özellik beş değerini alacaktır.

6.2 "Understanding Language 4: Learning and Teaching" Kursunda (MOOC) İngilizce Dil Gruplarının Haftalık Kurs Performans Analizleri

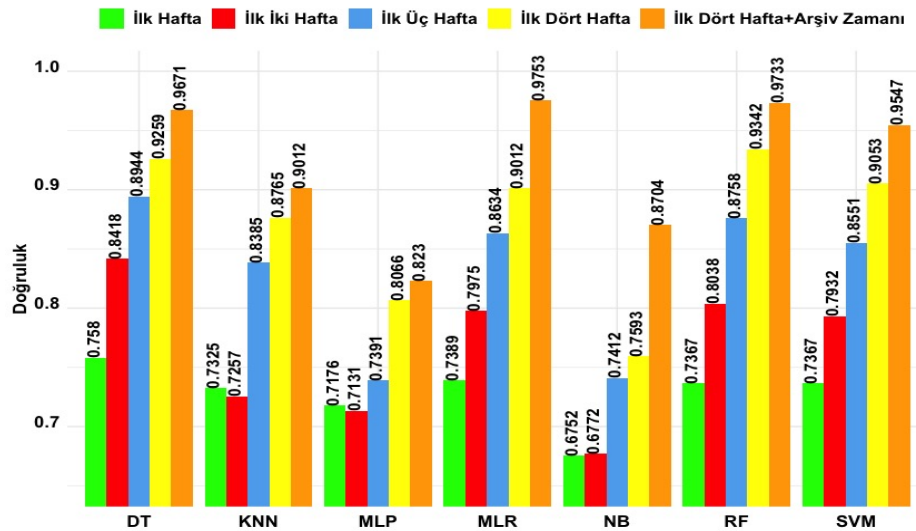
İngilizce dil grubu bilinen katılımcıların "Understanding Language 4: Learning and Teaching" MOOC'unda "ilk hafta", "ilk iki hafta", "ilk üç hafta", "ilk dört hafta"lık kurs dönemi ve "dört haftalık kurs dönemi + kursun arşiv zamanı" için, oluşturulan özellikler kullanılarak yedi geleneksel makine öğrenme algoritması ile gerçekleştirilen kurs performans tahmini sonuçları Şekil 6.1'de gösterilmiştir.



Şekil 6.1 İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk (accuracy) değerleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 6.1'e göre, "Understanding Language 4: Learning and Teaching" MOOC'unda, tüm İngilizce dil grubu bilinen katılımcıların "dört haftalık kurs dönemi + kursun arşiv zamanı"nda en iyi doğruluk yüzdesine sahip algoritma Rastgele Orman algoritması iken, bu algortimadan sonra Karar Ağacı (DT) algoritması en başarılı algoritma olmuştur. Tüm İngilizce dil grubu bilinen katılımcıların "ilk hafta", "ilk iki hafta", "ilk üç hafta" ve "ilk dört haftalık kurs dönemi" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Karar Ağacı olurken, ikinci sırada Rastgele Orman (RF) algoritması gelmektedir. Naive Bayes algoritması, tüm İngilizce dil grubu bilinen katılımcıların "ilk hafta", "ilk iki hafta", "ilk üç hafta" ve "ilk dört haftalık kurs dönemi" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en düşük doğruluk yüzdesine sahipken "dört haftalık kurs dönemi + kursun arşiv zamanı"nda en düşük doğruluk yüzdesine sahip algoritma Çok Katmanlı Algılayıcı (MLP) algoritması olmuştur.

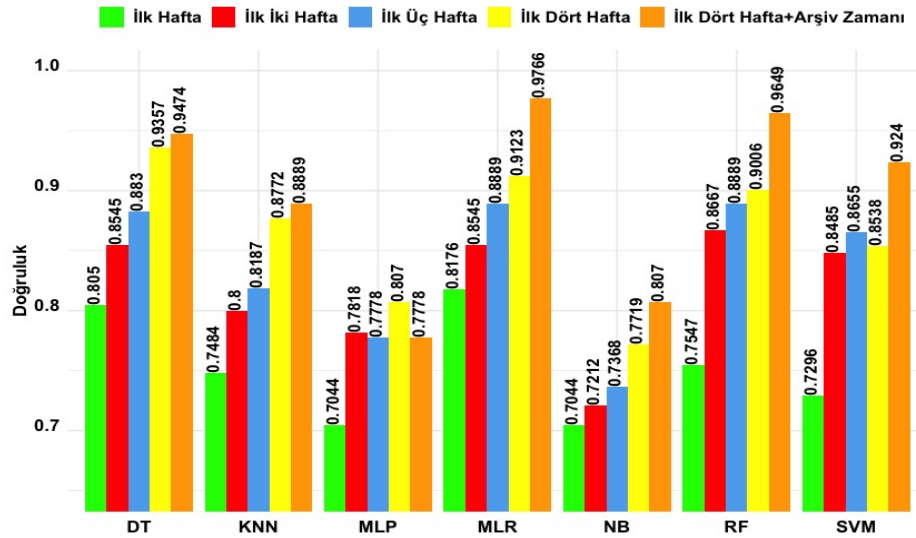
Şekil 6.2, "Understanding Language 4: Learning and Teaching" MOOC'unda birincil dili İngilizce olan katılımcıların kursun verildiği "ilk hafta", "ilk iki hafta", "ilk üç hafta", "ilk dört hafta"lık kurs dönemi ve "dört haftalık kurs dönemi + kursun arşiv zamanı"nda oluşturulan özellikler ve yaygın kullanılan yedi geleneksel algoritma ile gerçekleştirilen kurs performansı tahmini sonuçlarını göstermektedir.



Şekil 6.2 İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 6.2'ye göre, "Understanding Language 4: Learning and Teaching" MOOC'unda birincil dili İngilizce olan katılımcıların "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Çok Kategorili Lojistik Regresyon (MLR) olurken, ikinci sırada Rastgele Orman algoritması gelmektedir. "İlk dört haftalık kurs dönemi" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Rastgele Orman algoritması olurken, ikinci sırada Karar Ağacı algoritması gelmektedir. Birincil dili İngilizce olan katılımcıların "ilk iki hafta" ve "ilk üç hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı Karar Ağacı olurken, ikinci sırada Rastgele Orman algoritması gelmektedir. Birincil dili İngilizce olan katılımcıların "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde de en başarılı algoritma Karar Ağacı olurken, bu algorithmadan sonra Rastgele Orman, Destek Vektör Makinesi (SVM) algoritmaları en başarılı (iki algoritmanın başarı yüzdesi aynı) algoritmalar olmuşlardır. Naive Bayes ve Çok Katmanlı Algılayıcı (MLP) algoritmaları, birincil dili İngilizce olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinde en düşük doğruluk yüzdesine sahip olan iki algoritma olmuşlardır.

Şekil 6.3, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce sadece resmi dili olan katılımcıların kursun verildiği "ilk hafta", "ilk iki hafta", "ilk üç hafta", "ilk dört hafta"lık kurs dönemi ve "dört haftalık kurs dönemi + kursun arşiv zamanı"nda, oluşturulan özellikler ve yaygın kullanılan yedi geleneksel makine öğrenmesi algoritması ile gerçekleştirilen kurs performansı tahmini sonuçlarını göstermektedir.

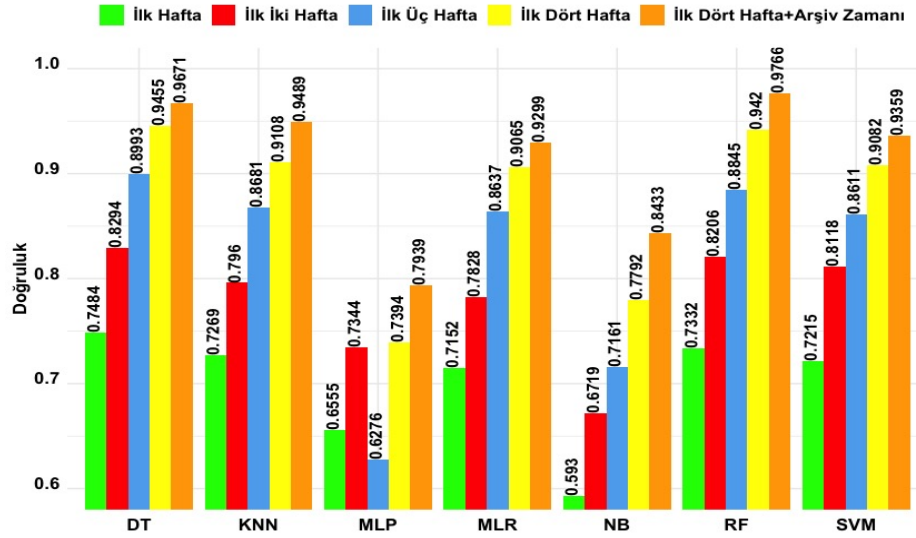


Şekil 6.3 İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 6.3'e göre "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce sadece resmi dili olan katılımcıların "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Çok Kategorili Lojistik Regresyon algoritması olurken, ikinci sırada Rastgele Orman algoritması gelmektedir. "İlk dört hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı Karar Ağacı olurken, ikinci sırada Çok Kategorili Lojistik Regresyon algoritması gelmektedir. "İlk üç hafta"lık kurs dönemi sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı iki algoritma Rastgele Orman ve Çok Kategorili Lojistik Regresyon algoritmaları olmuştur. "İlk iki hafta"lık kurs dönemi sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Rastgele Orman olurken, bu algoritmadan sonra Karar Ağacı ve Lojistik Regresyon algoritmaları gelmektedir. Son olarak, "ilk hafta"lık kurs dönemi sonuna kadar olan aktivite kullanım verileri ile en başarılı Çok Kategorili Lojistik Regresyon olurken ikinci sırada Karar Ağacı algoritması gelmektedir. Naive Bayes ve Çok Katmanlı Algılayıcı algoritmaları, İngilizce sadece resmi dili olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında

en düşük doğruluk yüzdesine sahip algoritmalar olmuşlardır.

Şekil 6.4, "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce ikincil dili olan katılımcıların kursun verildiği "ilk hafta", "ilk iki hafta", "ilk üç hafta", "ilk dört hafta"lık kurs dönemi ve "dört haftalık kursu dönemi + kursun arşiv zamanı"ndaki oluşturulan özellikler ve yaygın kullanılan yedi geleneksel algoritma ile gerçekleştirilen kurs performansı tahmini sonuçlarını göstermektedir.



Şekil 6.4 İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Understanding Language 4: Learning and Teaching" kursu

Şekil 6.4'e göre "Understanding Language 4: Learning and Teaching" MOOC'unda İngilizce ikincil dili olan katılımcıların "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Rastgele Orman algoritması olurken, ikinci sırada Karar Ağacı algoritması gelmektedir. İngilizce ikincil dili olan katılımcıların "ilk hafta", "ilk iki hafta", "ilk üç hafta" ve "ilk dört hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı Karar Ağacı olurken, ikinci sırada Rastgele Orman algoritması gelmektedir. Naive Bayes ve Çok Katmanlı Algılayıcı algoritmaları, İngilizce ikincil dili olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında en düşük doğruluk yüzdesine sahip algoritmalar olmuşlardır.

"Understanding Language 4: Learning and Teaching" MOOC'unda haftalık kurs performansı tahmininde en iyi doğruluğu veren Rasgele Orman ve Karar Ağacı algoritmalarının "ilk hafta" (ilk yedi gün) ve "ilk iki hafta" (ilk 14 gün) sonunda gerçekleştirilen İngilizce dil gruplarına göre kurs performans tahminlerinde en önemli üç özellik ve sıraları Tablo 6.1'de gösterilmektedir.

Tablo 6.1 "Understanding Language 4: Learning and Teaching" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Rastgele Orman ve Karar Ağacı algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik

a) Rastgele Orman: İlk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. yorum uzunluk ortalaması	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. yorum uzunluk ortalaması	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. ana yorum sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. yorum uzunluk ortalaması
b) Rastgele Orman: İlk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı
c) Karar Ağacı: İlk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. yorum sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. yorum sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. yorum uzunluk ortalaması	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. yorum sayısı
d) Karar Ağacı: İlk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı

Tablo 6.1-a'ya göre, Rastgele Orman algoritması için "Understanding Language 4: Learning and Teaching" MOOC'unda "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde "tüm İngilizce dil grubu bilinenler", "İngilizce ana ve resmi dil" ve "İngilizce ikinci dil" gruplarında ilk üç sırada 'tamamlanan ilk hafta adımı sayısı', 'başlanan adım sayısı' ve 'yorum uzunluk ortalaması' özellikleri yer alırken, "İngilizce yalnızca resmi dil" grubunda ise diğer üç gruptan farklı olarak üçüncü özellik olarak 'ana yorum sayısı' yer almıştır.

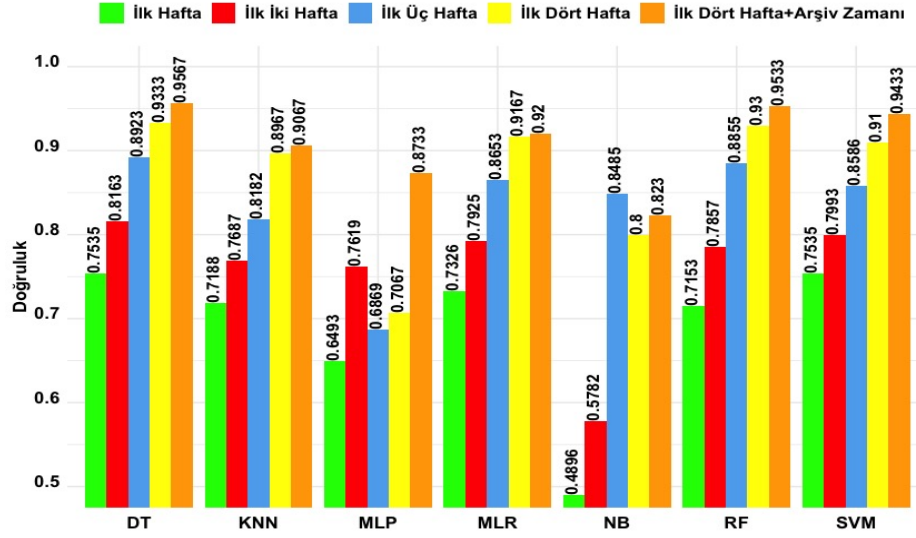
Tablo 6.1-b'ye göre, Rastgele Orman algoritması için "Understanding Language 4: Learning and Teaching" MOOC'unda "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde tüm İngilizce dil gruplarında en önemli ilk iki özellik "ilk hafta" sonundaki kurs performans tahmininde olan özellikler ile aynı iken, en önemli üçüncü özellik 'tamamlanan ikinci hafta adımı sayısı' olup, tüm İngilizce dil grupları için aynıdır. Kursun ilerleyen haftalarında Rasgele Orman algoritması ile kurs performans tahmininde önemli olan özellikler "ilk iki hafta" sonundaki ile uyumlu olmakta ve önemli değişim görülmemektedir.

Tablo 6.1-c'ye göre, Karar Ağacı algoritması için "Understanding Language 4: Learning and Teaching" MOOC'unda "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde Rastgele Orman algoritması ile benzer şekilde "İngilizce dil grubu bilinenler", "İngilizce ana ve resmi dil" ve "İngilizce ikinci dil" gruplarında ilk üç sırada 'başlanan adım sayısı', 'tamamlanan ilk hafta adımı sayısı' ve 'yorum sayısı' özellikleri yer almıştır. "İngilizce yalnızca resmi dil" grubunda ise ilk üç sırada 'tamamlanan ilk hafta adımı sayısı', 'başlanan adım sayısı' ve 'yorum uzunluk ortalaması' yer almıştır.

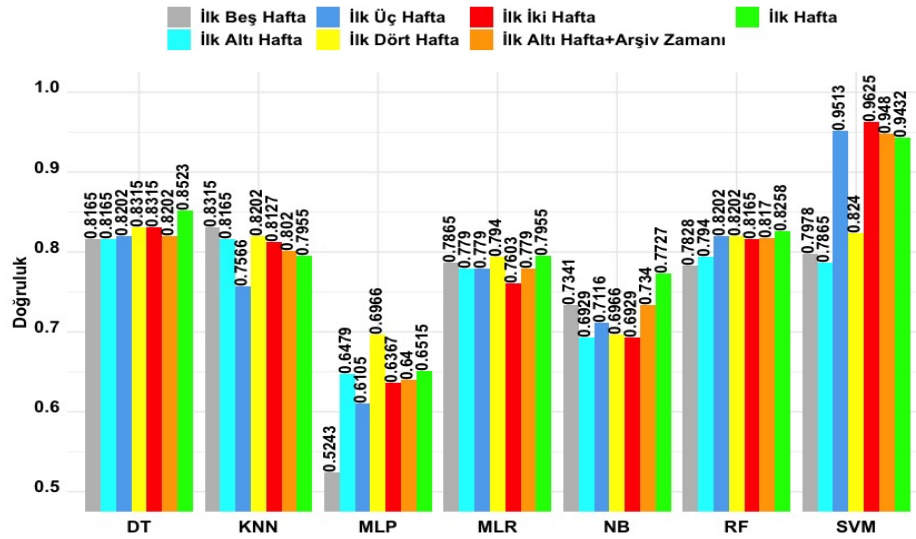
Tablo 6.1-d'ye göre, Karar Ağacı algoritması için "Understanding Language 4: Learning and Teaching" MOOC'unda "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde, "tüm İngilizce dil grubu bilinenler" ve "İngilizce ikinci dil" gruplarında en önemli özelliklerde ilk üç sırada 'başlanan adım sayısı', 'tamamlanan ilk hafta adımı sayısı' ve 'tamamlanan ikinci hafta adımı sayısı' özellikleri yer alırken, "İngilizce ana ve resmi dil" ve "İngilizce yalnızca resmi dil" grubunda ise ilk üç sırada 'tamamlanan ilk hafta adımı sayısı', 'başlanan adım sayısı' ve 'tamamlanan ikinci hafta adımı sayısı' özellikleri yer almıştır. Kursun ilerleyen haftalarında da Rasgele Orman algoritması ile kurs performans tahmininde önemli olan özellikler "ilk iki hafta" sonundaki ile uyumlu olmakta ve önemli değişim görülmemektedir.

6.3 Farklı Türden Kurslarda (MOOC'lar) İngilizce Dil Gruplarının Haftalık Kurs Performans Analizleri

Şekil 6.5 ve Şekil 6.6, "Exploring Our Oceans 4" ve "Web Science-1" MOOC'larında tüm İngilizce dil grubu bilinen katılımcıların kursun verildiği sürenin haftaları boyunca kümülatif olarak oluşturulan özellikler ve yaygın kullanılan yedi geleneksel makine öğrenmesi algoritması ile gerçekleştirilen kurs performans tahmini sonuçlarını göstermektedir.



Şekil 6.5 İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu

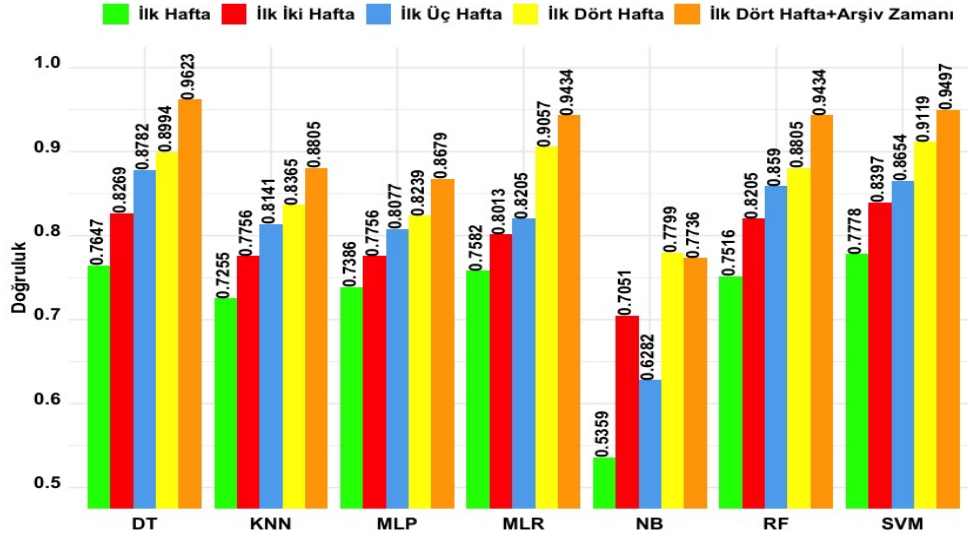


Şekil 6.6 İngilizce dil grubu bilinen tüm katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu

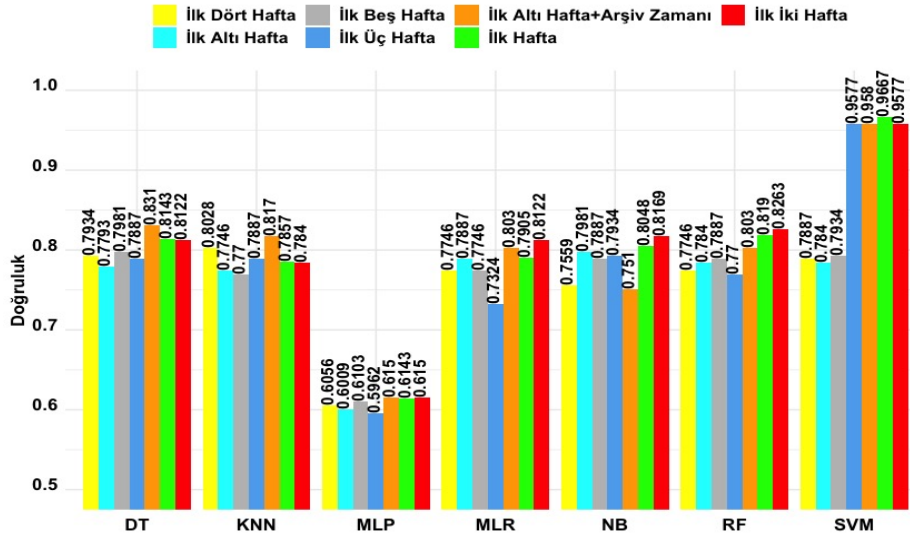
Şekil 6.5'e göre, "Exploring Our Oceans 4" MOOC'unda, tüm İngilizce dil grubu bilinen katılımcıların "ilk üç hafta" ve "ilk dört hafta"lık kurs dönemi sonuna kadar olan aktivite kullanım verileri ile ve "dört haftalık kurs dönemi + kursun arşiv zamanı"nda en iyi doğruluk yüzdesine sahip algoritma Karar Ağacı algoritması iken, bu algortimadan sonra Rastgele Orman algoritması en başarılı algoritma olmuştur. Tüm İngilizce dil grubu bilinen katılımcıların "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde de en başarılı algoritma Karar Ağacı algoritması olurken, bu algortimadan sonra en başarılı algoritma Destek Vektör Makinesi algoritması olmuştur. "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde de en başarılı algoritmalar Karar Ağacı ve Destek Vektör Makinesi algoritmaları olmuştur. Tüm İngilizce dil grubu bilinen katılımcıların "ilk hafta"dan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında en düşük doğruluk yüzdesine sahip algoritmalar Çok Katmanlı Algılayıcı ve Naive Bayes algoritmaları olmuştur.

Şekil 6.6'ya göre, "Web Science-1" MOOC'unda, tüm İngilizce dil grubu bilinen katılımcıların kümülatif olarak haftalık performans tahminlerinde en iyi doğruluk yüzdesine sahip algoritma Destek Vektör Makinesi algoritması iken, bu algortimadan sonra Rastgele orman ve Karar Ağacı algoritmaları en başarılı iki algoritma olmuşlardır. Tüm İngilizce dil grubu bilinen katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında en düşük doğruluk yüzdesine sahip olan algoritma Çok Katmanlı Algılayıcı algoritması olurken, Naive Bayes algoritması da en başarısız ikinci algoritma olmuştur.

Şekil 6.7 ve Şekil 6.8, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında birincil dili İngilizce olan katılımcıların kursun verildiği sürenin haftaları boyunca kümülatif olarak oluşturulan özellikler ve yaygın kullanılan yedi geleneksel makine öğrenmesi algoritması ile gerçekleştirilen kurs performans tahmini sonuçlarını göstermektedir.



Şekil 6.7 İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu



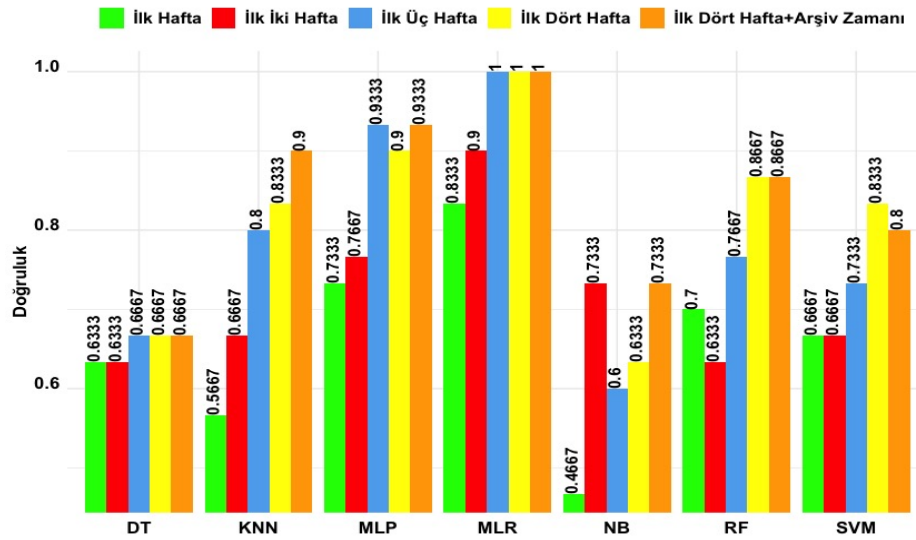
Şekil 6.8 İngilizce ana ve resmi dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu

Şekil 6.7'ye göre, "Exploring Our Oceans 4" MOOC'unda birincil dili İngilizce olan katılımcıların "ilk üç hafta" ve "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Karar Ağacı algoritması olurken ikinci sırada Destek Vektör Makinesi algoritması gelmektedir. "İlk hafta", "ilk iki hafta" ve "ilk dört hafta"lık kurs dönemi sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Destek Vektör Makinesi algoritması olurken ikinci sırada "ilk hafta" ve "ilk iki hafta"lık kurs dönemi sonuna kadar olan bölümde Karar Ağacı algoritması, "ilk dört hafta"lık kurs dönemi sonuna kadar olan bölümde ise Çok Kategorili Lojistik Regresyon

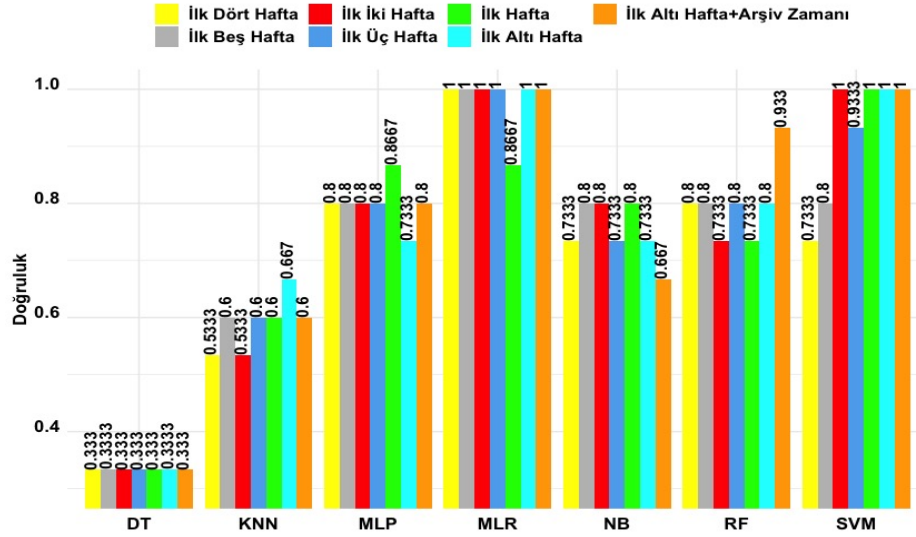
algoritması gelmektedir. Naive Bayes algoritması "ana ve resmi dili İngilizce" olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinde çoğunlukla en düşük doğruluk yüzdesine sahip algoritma olmuştur.

Şekil 6.8'e göre, "Web Science-1" MOOC'unda, birincil dili İngilizce olan katılımcıların kümülatif olarak haftalık performans tahminlerinde en iyi doğruluk yüzdesine sahip algoritma Destek Vektör Makinesi algoritması olmuştur. Bu algortimadan sonra Çok Katmanlı Algılayıcı Algoritması dışındaki algoritmaların performans tahmin başarı sonuçları birbirlerine yakın olup genel olarak en başarılı ikinci algoritma Rastgele Orman Algoritması olmuştur. Tüm İngilizce dil grubu bilinen katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında en düşük doğruluk yüzdesine sahip olan algoritma Çok Katmanlı Algılayıcı algoritması olmuştur.

Şekil 6.9 ve Şekil 6.10, "Exploring Our Oceans 4" ve "Web Science-1" MOOC'larında, "sadece resmi dili İngilizce" olan katılımcıların kursun verildiği sürenin haftaları boyunca kümülatif olarak oluşturulan özellikler ve yaygın kullanılan yedi geleneksel makine öğrenmesi algoritması ile gerçekleştirilen kurs performans tahmini sonuçlarını göstermektedir.



Şekil 6.9 İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu



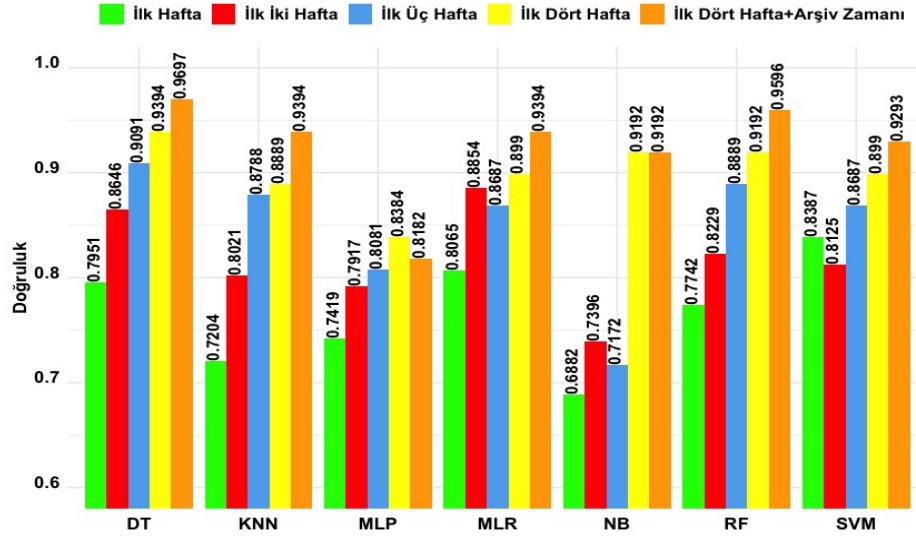
Şekil 6.10 İngilizce resmi dili olup birincil dili olmayan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu

Şekil 6.9'a göre, "Exploring Our Oceans 4" MOOC'unda, "sadece resmi dili İngilizce" olan katılımcıların "ilk hafta", "ilk iki hafta", "ilk üç hafta", "ilk dört hafta"lık kurs dönemi ve "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Çok Kategorili Lojistik Regresyon algoritması olurken ikinci sırada Çok Katmanlı Algılayıcı algoritması gelmektedir. Naive Bayes algoritması, "sadece resmi dili İngilizce" olan katılımcıların "ilk hafta"dan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinde çoğunlukla en düşük doğruluk yüzdesine sahip algoritma olurken K-En Yakın Komşu ve Karar ağacı algoritmaları diğer düşük doğruluk yüzdelere sahip algoritmalar olmuşlardır.

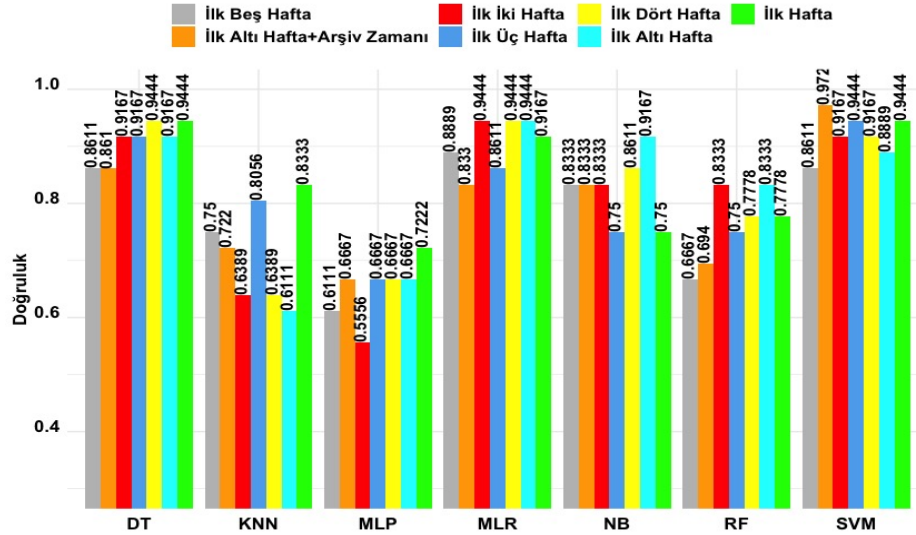
Şekil 6.10'a göre, "Web Science 1" MOOC'unda, İngilizce sadece resmi dili olan katılımcıların kümülatif olarak haftalık performans tahminlerinde en iyi doğruluk yüzdesine sahip iki algoritma Çok Kategorili Lojistik Regresyon ve Destek Vektör Makinesi algoritmaları olmuştur. "İlk hafta" sonuna kadar aktivite kullanım verilerinin kullanıldığı durumda en başarılı sonucu veren algoritma Destek Vektör Makinesi olurken, diğer tüm durumlarda Çok Kategorili Lojistik Regresyon algoritması en başarılı sonucu veren algoritma olmuştur. Tüm İngilizce dil grubu bilinen katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin tamamında en düşük doğruluk yüzdesine sahip olan algoritma Karar Ağacı algoritması olurken, en başarısız sonucu veren ikinci algoritma K En Yakın Komşu algoritması olmuştur.

Şekil 6.11, ve Şekil 6.12, "Exploring Our Oceans 4" ve "Web Science 1" MOOC'larında, İngilizce ikincil dili olan katılımcıların kursun verildiği sürenin haftaları boyunca

kümülatif olarak oluşturulan özellikler ve yaygın kullanılan yedi geleneksel makine öğrenmesi algoritması ile gerçekleştirilen kurs performans tahmini sonuçlarını göstermektedir.



Şekil 6.11 İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Exploring Our Oceans 4" kursu



Şekil 6.12 İngilizce ikincil dili olan katılımcıların haftalık kurs performans tahmini doğruluk değerleri - "Web Science 1" kursu

Şekil 6.11'e göre, "Exploring Our Oceans 4" MOOC'unda, "İngilizce ikincil dili" olan katılımcıların "ilk dört hafta" ve "dört haftalık kurs dönemi + kursun arşiv zamanı" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Karar Ağacı algoritması olurken ikinci sırada Rastgele algoritması

gelmektedir. "İlk iki hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde en başarılı algoritma Çok Kategorili Lojistik Regresyon algoritması olurken ikinci sırada Karar Ağacı algoritması gelmektedir. "İlk hafta" sonuna kadar olan aktivite kullanım verileri ile kurs performans tahmininde ise en başarılı algoritma Destek Vektör Makinesi olurken ikinci sırada Çok Kategorili Lojistik Regresyon algoritması gelmektedir. Naive Bayes algoritması, "İngilizce ikincil dili" olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin çoğundan düşük doğruluk yüzdesine sahip algoritma olurken bu algoritmayı Çok Katmanlı Algılayıcı algoritması takip etmektedir.

Şekil 6.12'ye göre, "Web Science 1" MOOC'unda, İngilizce ikincil dili olan katılımcıların kümülatif olarak haftalık performans tahminlerinde en iyi doğruluk yüzdesine sahip üç algoritma Çok Kategorili Lojistik Regresyon, Destek Vektör Makinesi ve Karar Ağacı algoritmasıdır. En iyi sonuç veren üç algoritmanın başarıları arasındaki fark düşük olmakla birlikte Destek Vektör Makinesi genel olarak en iyi sonucu veren algoritma olmuştur. İngilizce ikincil dili olan katılımcıların ilk haftadan itibaren gerçekleştirilen kümülatif haftalık kurs performans tahminlerinin çoğunda en düşük doğruluk yüzdesine sahip olan algoritma Çok Katmanlı Algılayıcı olurken, en başarısız sonucu veren ikinci algoritma K En Yakın Komşu algoritması olmuştur.

"Exploring Our Oceans 4" MOOC'unda, haftalık kurs performanslarının tahmininde en iyi doğruluğu veren Rasgele Orman ve Karar Ağacı algoritmalarının "ilk hafta" (ilk yedi gün) ve "ilk iki hafta" (ilk 14 gün) sonunda gerçekleştirilen İngilizce dil gruplarına göre kurs performans tahminlerinde en önemli üç özellik ve sıraları Tablo 6.2'de gösterilmektedir.

Tablo 6.2 "Exploring Our Oceans 4" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Rastgele Orman ve Karar Ağacı algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik

a) Rastgele Orman: İlk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanmadan bırakılan adım sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanmadan bırakılan adım sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanmadan bırakılan adım sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanmadan bırakılan adım sayısı
b) Rastgele Orman: İlk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı
c) Karar Ağacı: İlk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanmadan bırakılan adım sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanmadan bırakılan adım sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanmadan bırakılan adım sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanmadan bırakılan adım sayısı
d) Karar Ağacı: İlk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanan ikinci hafta adımı sayısı	1. başlanan adım sayısı 2. tamamlanan ilk hafta adımı sayısı 3. tamamlanan ikinci hafta adımı sayısı

Tablo 6.2-a'ya göre, Rastgele Orman algoritması için "Exploring Our Oceans

4" MOOC'unda, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde 'tamamlanan ilk hafta adımı sayısı', tüm İngilizce dil grupları bilinenlerin verileri ile "sadece resmi dil olarak İngilizce" konuşanların verilerinde en önemli özellikler olurken, 'başlanan adım sayısı', "ana ve resmi dili İngilizce" olanların verisi ve "ikincil dil olarak İngilizce" konuşanların verisinde en önemli özellik olmuştur. 'Tamamlanmadan bırakılan adım sayısı' tüm İngilizce dil gruplarının verilerinde en önemli üçüncü tahmin edici özellik olmuştur.

Tablo 6.2-b'ye göre, Rastgele Orman algoritması için "Exploring Our Oceans 4" MOOC'unda, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininden farklı olarak 'tamamlanmadan bırakılan adım sayısı' en önemli üç özellik içerisinde yer almamıştır. Tüm İngilizce dil gruplarının verilerinde en önemli tahmin edici özellik 'başlanan adım sayısı' olurken ikinci sırada 'tamamlanan ilk hafta adımı sayısı' ve üçüncü sırada 'tamamlanan ikinci hafta adımı sayısı' yer almıştır.

Tablo 6.2-c'ye göre, Karar Ağacı algoritması için "Exploring Our Oceans 4" MOOC'unda, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde 'tamamlanan ilk hafta adımı sayısı', 'ana ve resmi dil olarak İngilizce' dil grubu verileri ile en önemli özellik olurken 'başlanan adım sayısı, tüm diğer İngilizce dil grubu verilerinde en önemli özellik olmuştur. 'Tamamlanmadan bırakılan adım sayısı' tüm İngilizce dil gruplarının verilerinde en önemli üçüncü tahmin edici özellik olmuştur.

Tablo 6.2-d'ye göre, Karar Ağacı algoritması için "Exploring Our Oceans 4" MOOC'unda, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininden farklı olarak 'tamamlanmadan bırakılan adım sayısı' en önemli üç özellik içerisinde yer almamıştır. "Sadece resmi dil olarak İngilizce" konuşanların dışında tüm İngilizce dil grubu verilerinde en önemli tahmin edici özellik 'başlanan adım sayısı' olurken ikinci sırada 'tamamlanan ilk hafta adımı sayısı' ve üçüncü sırada 'tamamlanan ikinci hafta adımı sayısı' yer almıştır. "Sadece resmi dil olarak İngilizce" konuşanların verisinde ise 'tamamlanan ilk hafta adımı sayısı' en önemli tahmin edici özellik olurken ikinci sırada 'başlanan adım sayısı' ve üçüncü sırada 'tamamlanan ikinci hafta adımı sayısı' yer almıştır.

"Web Science 1" MOOC'unda, haftalık kurs performanslarının tahmininde en iyi doğruluğu veren iki algoritmadan Çok Kategorili Lojistik Regresyon algoritmasının

“ilk hafta” ve “ilk iki hafta” sonunda gerçekleştirilen İngilizce dil gruplarına göre kurs performans tahminlerinde en önemli üç özellik ve sıraları Tablo 6.3’de gösterilmektedir.

Tablo 6.3 "Web Science 1" kursunda haftalık performans tahmininde en başarılı iki algoritma olan Çok Kategorili Lojistik Regresyon ve Destek Vektör Makinesi algoritmaları ile ilk hafta ve ilk iki hafta sonunda kurs performans tahmininde en önemli üç özellik

a) Çok Kategorili Lojistik Regresyon: İlk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. cevap yorum sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. ana yorum sayısı	1. cevap yorum sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. ana yorum sayısı	1. tamamlanan ilk hafta adımı sayısı 2. başlanan adım sayısı 3. tamamlanmadan bırakılan adım sayısı	1. cevap yorum sayısı 2. cevap yorum uzunluk ortalaması 3. yorum uzunluk ortalaması
b) Çok Kategorili Lojistik Regresyon: İlk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. tamamlanan ikinci hafta adımı sayısı 2. yorum beğeni sayısı ortalaması 3. tamamlanan ilk hafta adımı sayısı	1. yorum beğeni sayısı ortalaması 2. cevap yorum sayısı 3. yorum sayısı	1. başlanan adım sayısı 2. tamamlanmadan bırakılan adım sayısı 3. ana yorum uzunluk ortalaması	1. ana yorum sayısı 2. tamamlanan ikinci hafta adımı sayısı 3. yorum sayısı
c) Destek Vektör Makinesi algoritması ile ilk hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. yorum uzunluk ortalaması 2. ana yorum uzunluk ortalaması 3. başlanan adım sayısı	1. yorum uzunluk ortalaması 2. ana yorum uzunluk ortalaması 3. cevap yorum uzunluk ortalaması	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı
d) Destek Vektör Makinesi algoritması ile ilk iki hafta sonunda en önemli üç özellik			
İngilizce Dil Grupları Bilinenler	İngilizce Ana ve Resmi Dil	İngilizce Yalnızca Resmi Dil	İngilizce İkincil Dil
1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı	1. ana yorum uzunluk ortalaması 2. yorum uzunluk ortalaması 3. başlanan adım sayısı

Tablo 6.3-a'ya göre, Çok Kategorili Lojistik Regresyon algoritması için "Web Science 1" MOOC'unda, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde, denenen diğer kurs verilerine göre en önemli olan özelliklerde farklılık olduğu ve "sadece resmi dil olarak İngilizce" konuşanların verileri dışında, yorumlarla ilgili özelliklerin daha önemli özellikler olduğu görülmüştür. 'Cevap yorum sayısı', "sadece resmi dil olarak İngilizce" konuşanların verileri dışında tüm İngilizce dil gruplarının verilerinde en önemli özellik olurken, "sadece resmi dil olarak İngilizce" konuşanların verilerinde 'tamamlanan ilk hafta adımı sayısı' en önemli özellik olmuştur.

Tablo 6.3-b'ye göre, Çok Kategorili Lojistik Regresyon algoritması için "Web Science 1" MOOC'unda, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininden farklı olarak katılımcıların başladıkları ve tamamladıkları kurs adımı sayıları ile ilgili özellikler daha önemli olmuştur. Bununla birlikte İngilizce dil gruplarının verileri arasında en önemli olan üç özellikte önemli farklılıkların olduğu görülmüştür. Örneğin; en önemli tahmin edici özellik, "İngilizce dil grubu bilinenler" in verilerinde 'tamamlanan ikinci hafta adımı sayısı', "ana ve resmi dil olarak İngilizce" konuşanların verilerinde 'yorum beğeni sayısı ortalaması', "sadece resmi dil olarak İngilizce" konuşanların verilerinde 'başlanan adım sayısı' olurken, "ikincil dil olarak İngilizce" konuşanların verilerinde ise 'ana yorum sayısı' olmuştur.

Tablo 6.3-c'ye göre, Destek Vektör Makinesi algoritması için "Web Science 1" MOOC'unda, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde, denenen diğer kurs verilerine göre en önemli olan özelliklerde farklılık olduğu ve yorumlarla ilgili özelliklerin daha önemli özellikler olduğu görülmüştür. İngilizce dil grubu verilerinin tamamında, 'ana yorum uzunluk ortalaması' ve 'yorum uzunluk ortalaması' en önemli iki özellik olmuştur. 'Başlanan adım sayısı', "sadece resmi dil olarak İngilizce" konuşanların verileri dışındaki İngilizce dil gruplarının verilerinin tamamında en önemli üçüncü özellik olurken, "sadece resmi dil olarak İngilizce" konuşanların verilerinde ise 'cevap yorum uzunluk ortalaması' en önemli üçüncü özellik olmuştur.

Tablo 6.3-d'ye göre, Destek Vektör Makinesi algoritması için "Web Science 1" MOOC'unda, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahmininde, "ilk hafta" sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen kurs performans tahminine benzer olarak yorumlarla ilgili özelliklerin daha önemli özellikler olduğu görülmüştür. İngilizce dil gruplarının verilerinin tamamında en önemli üç özellik sırası ile 'ana yorum uzunluk

ortalaması', 'yorum uzunluk ortalaması' ve 'başlanan adım sayısı' olmuştur.

6.4 Sonuçlar

"Understanding Language 4: Learning and Teaching" MOOC'unda, haftalık performans analizlerinde en iyi sonucu veren üç algoritma Rastgele Orman, Karar Ağacı ve Çok Kategorili Lojistik Regresyon algoritmaları olurken, Rastgele Orman ve Karar Ağacı algoritmaları genel olarak biraz daha iyi doğrulukla tahmin gerçekleştirmişlerdir. Üç algoritmanın tahminleri arasındaki başarı farkı ise düşüktür.

"Exploring Our Oceans 4" MOOC'unda, haftalık performans analizlerinde çoğunlukla en iyi sonucu veren iki algoritma Rastgele Orman ve Karar Ağacı algoritmaları olurken, Çok Kategorili Lojistik Regresyon ve Destek Vektör Makinesi algoritmalarının sonuçları da bu iki algoritmaya yakın olmuştur. Diğer İngilizce dil gruplarının verileri ile haftalık kurs performansı tahmin başarı sonuçlarından farklı olarak, "sadece resmi dil olarak İngilizce" konuşanların verilerinde en başarılı algoritma Çok Kategorili Lojistik Regresyon olurken bu algoritmayı Çok Katmanlı Algılayıcı algoritması takip etmiştir.

"Web Science 1" MOOC'unda, denenen diğer kurs verilerinden farklı olarak, haftalık kurs performans analizlerinde çoğunlukla en iyi sonucu veren iki algoritma Çok Kategorili Lojistik Regresyon ve Destek Vektör Makinesi algoritmaları olmuştur.

"Understanding Language 4: Learning and Teaching" MOOC'unda, hem "ilk hafta" hem de "ilk iki hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahmininde en başarılı iki algoritma için, 'tamamlanan ilk hafta adımı sayısı' ve 'başlanan adımı sayısı' en önemli iki özellik olmuştur. "İlk hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahmininde, yorumlarla ilgili özellikler en önemli üçüncü özellik olurken, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahmininde, 'tamamlanan ikinci hafta adımı sayısı' en önemli üçüncü özellik olmuştur.

"Exploring Our Oceans 4" MOOC'unda, "ilk iki hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahmininde en başarılı iki algoritma için "Understanding Language 4: Learning and Teaching" MOOC'una benzer şekilde, 'tamamlanan ilk hafta adımı sayısı' ve 'başlanan adımı sayısı' en önemli iki özellik olurken, farklı olarak 'başlanan adım sayısı' genellikle en önemli tahmin edici özellik olmuş ve yorumlarla ilgili özellikler en önemli özelliklerde yer almamıştır.

"Web Science 1" MOOC'unda, "ilk hafta" ve "ilk iki hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahmininde en başarılı

iki algoritma için en önemli özelliklerde, denenen diğer kurslardan farklı olarak yorumlara bağlı özelliklerin daha önemli olduğu görülmüştür.

Algoritmaların kümülatif olarak haftalık performans tahmin başarıları, "Web Science 1" kursu dışında genel olarak haftalar boyunca artış göstermiştir. "Understanding Language 4: Learning and Teaching" ve "Exploring Our Oceans 4" MOOC'larında, "ilk hafta" sonuna kadar olan aktivite kullanım verileriyle gerçekleştirilen kurs performans tahminleri genel olarak % 75 civarında doğruluk gösterirken, "Exploring Our Oceans 4" MOOC'unda "sadece resmi dil olarak İngilizce" konuşanların verilerinde tahmin doğrulukları daha düşük olmaktadır. "Ana ve resmi dil olarak İngilizce" konuşan katılımcıların dışında, kurs performans tahmini doğruluk yüzdeleri ilk haftadan sonra "Understanding Language 4: Learning and Teaching" kursunda % 80'in "Exploring Our Oceans 4" MOOC'unda % 70'in üzerine çıkmaktadır. "Understanding Language 4: Learning and Teaching" MOOC'unda "ikincil dil olarak İngilizce" konuşan katılımcıların, "ilk hafta" sonundaki kurs performans tahmini doğruluk sonuçları, diğer iki İngilizce dil grubundaki katılımcıların kurs performans tahmini doğruluk sonuçlarına göre daha düşüktür. Bu durum, "ikincil dil olarak İngilizce" konuşan katılımcılarda, kursu terk etme riski altında olanları doğru olarak tahmin etmenin daha zor olduğunu da göstermektedir. "Web Science 1" kursunda ise kümülatif olarak haftalar boyunca kurs performans tahminlerinin başarılarında hem artışlar hem de azalışlar görülebilmektedir.

Haftalık kurs performans tahminlerinde, "Web Science 1" MOOC'unda, en önemli ilk iki özellik katılımcıların tartışma forumlarındaki yorumları ile ilgili özellikler olurken, "Understanding Language 4: Learning and Teaching" MOOC'unda forumlarındaki yorumları ile ilgili özellikler en önemli üçüncü özellik olarak yer almışlardır. "Exploring Our Oceans 4" MOOC'unda ise yorumlarla ilgili özellikler, en önemli özelliklerin içinde yer almamıştır. Örneğin; "Understanding Language 4: Learning and Teaching" MOOC'unda, Rastgele Orman algoritmasında 'ortalama yorum uzunluğu', Karar Ağacı algoritmasında ise 'yorum sayısı' en önemli üçüncü kurs performansı tahmin edici özelliktir.

"Understanding Language 4: Learning and Teaching" kursunda, kümülatif olarak haftalık kurs performansı tahminlerinde, Naive Bayes ve Çok Katmanlı Algılayıcı algoritmaları en kötü doğruluk sonucuna sahip ve diğer algoritmaların tahmin sonuçları ile farkları belirgin iken, "Exploring Our Oceans 4" kursunda, Çok Katmanlı Algılayıcı algoritması diğer algoritmalara yakın tahmin başarı sonuçlarına sahip olmuştur. "Web Science 1" kursunda ise, "ikincil dil olarak İngilizce" konuşanların verileri dışında Çok Katmanlı Algılayıcı algoritması en kötü doğruluk sonucunu verirken, K En Yakın Komşu algoritması en kötü ikinci doğruluk sonucunu vermiştir.

Sonuç olarak, haftalık performans tahmininde en önemli olan özellikler; kurs türüne, kullanılan algoritmaya, İngilizce dil grubu verisine ve kursun başından itibaren kaç haftalık veri kullanıldığına bağlı olarak değişebilmektedir.

7 Sonuç ve Öneriler

Kitlesel Açık Çevrim İçi Kursların esnek öğrenme ortamlarından ötürü, katılımcılar kursları tamamlamak için farklı yollar izlemektedirler. Örneğin, bazı katılımcılar kursun yalnızca bir bölümü ile ilgilenirken, bazıları asla ödev teslim etmemekte ve bazıları da tartışma forumlarında daha aktif olmayı seçmektedirler. Bu sebeple, MOOC sağlayıcıları ve kurs yazarlarının, katılımcıların davranışlarını anlamaları, gelecekteki davranış ve performanslarını tahmin etmeleri ve gerekli destekleri belirlemeleri için önemlidir. Aynı zamanda, katılımcıların platformdaki öğrenme aktivitelerini sürdürmelerini sağlamak için nelere ihtiyaç duyduklarını da anlamaları gerekmektedir. Tez, MOOC literatüründeki bu önemli boşluğu doldurmaktadır.

Tezde, veri kaynağı olarak 2014 - 2018 yılları arasında yayında kalan FutureLearn platformuna ait "Understanding Language : Learning and Teaching" MOOC serisinin ilk sekiz iterasyonu ile FutureLearn platformunda yayınlanmış farklı türden MOOC'lardan elde edilen veriler kullanılmıştır. FutureLearn platformu MOOC'ları üzerindeki çalışmalar ilk olarak "Understanding Language 4: Learning and Teaching" MOOC'u üzerinde gerçekleştirilmiştir. Çalışmaların ikinci aşamasında aynı MOOC serisinin ilk yedi MOOC'u olan "Understanding Language : Learning and Teaching" kurs serisinden elde edilen daha büyük veriler üzerinde gerçekleştirilmiştir. Son olarak da, aynı MOOC serisinin sekizinci versiyonu ile, FutureLearn platformuna ait farklı MOOC'ların verileri kullanılmıştır.

Tezde, FutureLearn platformu MOOC'larındaki katılımcıların İngilizce dil grupları tespit edilerek, davranışları ve performansları üzerinde detaylı karşılaştırmalı analizler gerçekleştirilmiştir. Analizlerde ortaya konulan sonuçlar, MOOC sağlayıcıları ve katılımcılarına İngilizce dil gruplarının tespiti, İngilizce dil gruplarındaki katılımcıların kullanım ve performanslarına yönelik öneriler sunma, platformların tasarımlarını ve kurslardaki çalışma yollarını geliştirmeleri için faydalı olabilecektir. Sonuçlar aynı zamanda, MOOC sağlayıcıları ve yazarlarının öğrenme sürecine müdahale edebilmeleri ve katılımcılara daha uygun çalışma deneyimleri sunabilmeleri için kursların tasarım ve dinamiklerini iyileştirmelerine katkı sağlayacak bulguları ortaya

koymaktadır.

Tezde uygulanan yaklaşımlar;

1. İngilizce dil grubu tespitinde uygulanan yaklaşımlar

- (a) Katılımcıları İngilizce dili gruplarına ayırmak için kurs öncesi anketlerinde verdikleri ülke bilgilerine dayalı yöntem önermek.
- (b) Katılımcıları İngilizce dili gruplarına ayırmak için oluşturulan düzenli ifade kalıplarına dayalı ve kurs öncesi anketlerindeki ülke bilgilerine dayalı yöntem oluşturmak.
- (c) İngilizce dil gruplarının tespitini iyileştirmek için farklı veri türleri üzerinde kullanılabilecek Derin Öğrenme model ve mimarileri önermek ve elde edilen sonuçları kendi aralarında karşılaştırmak.

2. İngilizce dil gruplarının kullanım analizlerinde uygulanan yaklaşımlar

- (a) Kurs boyunca kursa katılımı ve kurs tamamlama durumunu anahtar göstergeler olarak kullanarak farklı İngilizce dil gruplarındaki katılımcıların kurs katılımlarının karşılaştırılması.
- (b) İngilizce dil gruplarının kursu kullanımları arasında değişiklik gösteren ve göstermeyen aktiviteleri belirlemek.
- (c) Farklı kurslar üzerinde İngilizce dil gruplarının davranışlarını karşılaştırmak.

3. Kurs sonu performans tahmininde uygulanan yaklaşımlar

- (a) Oluşturulan özellikler ve güncellenen İngilizce dil gruplarına dayalı en uygun kurs performansı tahmin model ve özelliklerini belirlemek.
- (b) Kurs performans tahmininde sıkça kullanılan geleneksel makine öğrenme algoritmaları ile kurs performans tahmini sonuçlarını, oluşturulan Derin Öğrenme modelleri (farklı mimarileri deneyerek) ile elde edilen kurs performans tahmini başarı sonuçları ile karşılaştırmak.
- (c) Katılımcıların kursun başından itibaren kümülatif olarak haftalık kullanım verilerini oluşturarak, bu verilerle kurs performans tahminlerini gerçekleştirmek ve elde edilen tahmin başarı sonuçlarını ve tahminde önemli olan özellikleri kurs sonu performans tahminleri ile karşılaştırmak.

Tezde, katılımcıların kursa katılımları ile ilgili elde edilen sonuçlar önceki akademik çalışmalardaki [114] çıkarımlara benzer şekilde, kursa katılımın kurs haftaları

boyunca sürekli olarak azaldığını ve kursu tamamlama oranlarının düşük olduğunu ortaya koymuştur. Tezde, kursa katılım ve kursu tamamlamada daha önce gerçekleştirilmemiş olan İngilizce dili gruplarının karşılaştırmalı analizleri yapılmış ve İngilizce dil gruplarının kursları kullanım ve performansları arasında önemli farklılıklar olduğu ortaya konulmuştur. Aynı zamanda, detaylı olarak İngilizce dil gruplarını tespit etme, İngilizce dil gruplarına göre katılımcıların kurs adımlarını tamamlama, forum tartışmalarına katılım ve birbirlerini takip etme konusundaki davranışların analizi ile kurs sonu ve haftalık performans analizleri gerçekleştirilmiş ve sonuçlar karşılaştırılmıştır. Örneğin, ilk dili İngilizce olmayan kişilerin çoğu (yalnızca resmi dil olarak İngilizce konuşanlar veya “ikincil dil olarak İngilizce” konuşanlar) kurs adımlarının hiçbirini tamamlamazlarken, ilk dili İngilizce olan katılımcılar daha fazla yüzdeyle kurs adımlarını tamamlamaktadırlar.

Tez çalışmasının sağladığı katkılar aşağıdaki gibi özetlenebilir;

1. İngilizce dil gruplarının tespiti için sağlanan katkılar

- (a) Kurs öncesi anketinde katılımcıların verdikleri ülke bilgilerinden İngilizce dil gruplarının tespiti aşamasında, katılımcıları sadece İngilizce’yi birincil dil olarak konuşup konuşmadıklarına göre gruplara ayırmak yerine “birincil ve resmi dil olarak İngilizce”, “sadece resmi dil olarak İngilizce” ve “ikincil dil olarak İngilizce” gruplarına ayırmanın daha doğru olacağı ortaya konulmuştur.
- (b) Hem oluşturulan düzenli ifade kalıpları hem de katılımcıların kurs öncesi anketinde verdikleri ülke bilgileri ile İngilizce dil grupları tespit edilmiştir. Önerilen düzenli ifade kalıplarını içeren yeni yöntem, katılımcılar milletleri, ülkeleri, şehirleri veya ilk dilleri hakkında bilgi verdiklerinde forum tartışmalarına gönderdikleri yorumları kullanarak ait oldukları İngilizce dil gruplarını tespit etmektedir. Yöntemi oluşturmak için ilk olarak katılımcıların nereden olduklarını, milletleri veya ilk dillerinin ne olduğunu söyledikleri cümle yapıları belirlenmiştir. Ardından bu cümle yapılarını tespit edebilecek düzenli ifade kalıpları oluşturulmuştur.
- (c) Oluşturulan düzenli ifade kalıplarına dayalı olarak önerilen yöntemin, kurs anketinden tespit edilenden daha fazla ve daha doğru olarak katılımcıların İngilizce dil grubunu tespit ettiği ortaya konulmuştur. Kurs öncesi anketteki ülke bilgilerinden elde edilen İngilizce dil grupları, düzenli ifade kalıpları ile elde edilenden farklı ise katılımcıların İngilizce dil grupları düzenli ifadelerle elde edilen grup olarak güncellenmiştir.
- (d) Oluşturulan Derin Öğrenme modelleri ile katılımcıların yorumlarında

milletleri, ülkeleri, şehirleri veya ilk dilleri hakkında bilgi vermedikleri durumlar için de, yorumlarından İngilizce dil gruplarının tahmininin gerçekleştirilmesi sağlanmıştır.

2. İngilizce dil gruplarının kullanım analizleri için sağlanan katkılar

- (a) İngilizce dil gruplarına göre katılımcıların kursla etkileşim davranışları analiz edilmiştir. Analiz sonuçlarına göre, “birincil ve resmi dil olarak İngilizce” konuşan katılımcıların daha aktif olarak kursla etkileşim kurdukları ve kursu tamamlama olasılıklarının daha yüksek olduğu gözlemlenmiştir.
- (b) İngilizce dil gruplarında, katılımcıların kullanım ve kurs performanslarının karşılaştırmalı analizlerini sağlamak için bir prototip araç geliştirilmiştir.
- (c) İngilizce dil gruplarındaki katılımcıların farkı kurslardaki aktivite kullanım sonuçları karşılaştırılarak, kurs türünün İngilizce dil gruplarının aktivite kullanımlarına olan etkileri ortaya konulmuştur.

3. İngilizce dil gruplarının kurs performans analizleri için sağlanan katkılar

- (a) Kurs sonuna kadar olan aktivite kullanım verileri ile gerçekleştirilen İngilizce dil gruplarının kurs sonu performans tahmininde denenen geleneksel makine öğrenmesi algoritmalarından en başarılı ve başarısız olanlar belirlenmiştir. Örneğin; "Understanding Language 4: Learning and Teaching" kursunda, "Exploring Our Oceans 4" kursunda ve "Web Science 1" kursunda kurs sonu performans tahmininde Rastgele Orman algoritması en başarılı algoritma olurken, "Understanding Language : Learning and Teaching" kurs serisinin ilk yedi iterasyonu ile gerçekleştirilen kurs sonu performans tahmininde en başarılı algoritma Karar Ağacı algoritması olmuştur.
- (b) Algoritmaların haftalık kurs performansı tahmini sonuçları ile kurs sonuna kadar olan aktivite kayıtları kullanılarak gerçekleştirilen kurs sonu performans tahmin sonuçları karşılaştırılmıştır. Örneğin; "Understanding Language 4: Learning and Teaching" kursunda, Rastgele Orman algoritması denenen diğer geleneksel makine öğrenme algoritmalarına göre kurs sonuna kadar olan aktivite kullanım kayıtları ile gerçekleştirilen kurs sonu performans tahmininde daha başarılı olurken, Karar Ağacı algoritması, kümülatif olarak ilk üç hafta sonunda (ilk hafta, ilk iki hafta ve ilk üç hafta) gerçekleştirilen kurs performans tahminlerinde en yüksek başarıya sahip olmuştur.

- (c) Geleneksel makine öğrenmesi algoritmaları ile İngilizce dil gruplarındaki katılımcıların kurs sonu performans tahminlerinde hangi davranışların daha belirleyici olduğu tespit edilmiştir. Örneğin; "Understanding Language 4: Learning and Teaching" kursunda, "birincil ve resmi dil olarak İngilizce" konuşan katılımcılar için en iyi kurs sonu performans tahmini doğruluğuna sahip olan Rastgele Orman algoritmasında, 'birinci haftada tamamlanan toplam adım sayısı' en belirleyici özellik iken, "sadece resmi dil olarak İngilizce" konuşan katılımcılar ve "ikincil dil olarak İngilizce" konuşan katılımcılar için 'başlanan adım sayısı' en belirleyici özelliktir.
- (d) MOOC'lar üzerinde gerçekleştirilen haftalık performans analizlerinde en önemli olan özelliklerin İngilizce dil grupları arasında değişiklik gösterdiği ve göstermediği durumlar ortaya konulmuştur. Örneğin; "Understanding Language 4: Learning and Teaching" kursunda, haftalık kurs performans tahmininde en başarılı doğruluk sonuçlarına sahip iki algoritmadan biri olan Rastgele Orman algoritması ile ilk hafta sonunda ve ilk iki hafta sonunda gerçekleştirilen kurs performans tahmininde, 'ilk hafta tamamlanan adım sayısı' tüm İngilizce dil gruplarının verisinde en belirleyici özelliktir. Karar Ağacı algoritması ile ilk hafta sonunda kurs performans tahmininde yalnızca resmi dili İngilizce olan katılımcılarda 'ilk hafta tamamlanan adım sayısı' en belirleyici özellik iken, diğer tüm İngilizce dil gruplarında 'başlanan adım sayısı' en belirleyici özellik olmuştur. Karar Ağacı algoritması ile ilk iki hafta sonunda kurs performans tahmininde, İngilizce'yi hem ana hem de resmi dil olarak konuşan katılımcılarla, İngilizce'yi yalnızca resmi dil olarak konuşan katılımcılarda 'ilk hafta tamamlanan adım sayısı' en belirleyici özellik iken, tüm İngilizce dil grubu bilinen katılımcılarla, İngilizce'yi ikincil dil olarak konuşan katılımcılarda 'başlanan adım sayısı' en belirleyici özellik olmuştur.
- (e) Haftalık performans analizlerinde en belirleyici olan özelliklerin bazılarının kurs türüne bağlı olmaksızın aynı kaldığı, bazılarının da kurs türüne bağlı olarak değişim gösterebildiği ortaya konulmuştur. Örneğin; kursun ilk haftası içinde gerçekleştirilen aktivite kullanımları ile gerçekleştirilen kurs performans tahminlerinde, 'başlanan adım sayısı', 'tamamlanan ilk hafta adım sayısı' kurs türüne bağlı olmaksızın en önemli üç özellik içinde yer almıştır. Bununla birlikte, forum aktivite kullanımları ile ilgili özelliklerden 'yorum sayısı' veya 'yorum uzunluk ortalaması' hem "Understanding Language 4: Learning and Teaching" kursunda hem de "Web Science 1" kursunda en önemli üç özellik içinde yer alırken; 'cevap yorum sayısı', 'ana yorum sayısı', 'ana yorum uzunluk ortalaması' ve 'yorum beğeni sayısı ortalaması' özellikleri "Web Science 1" kursunda en

önemli üç özellik içinde yer almıştır. 'Tamamlanmadan bırakılan adım sayısı', "Exploring Our Oceans 4" kursunda en önemli üç özellik içinde yer almaktadır.

- (f) Haftalık performans analizlerinde sık kullanılan yedi geleneksel makine öğrenmesi algoritmasından kurs türüne bağlı olarak en başarılı ve en başarısız olanlar ortaya konulmuştur. Örneğin; kümülatif haftalık kullanım verileri ile kurs performans tahmininde "Understanding Language 4: Learning and Teaching" kursunda Rastgele Orman algoritması ve Karar Ağacı algoritması genel olarak en başarılı, 'Naive Bayes' ve 'Çok katmanlı Algılayıcı' algoritmaları en başarısız algoritmalar olurken, "Exploring Our Oceans 4" kursunda, kullanılan İngilizce dil grubunun verisine bağlı olarak en başarılı ve en başarısız algoritmalar değişkenlik gösterebilmektedir.

İngilizce dil gruplarındaki katılımcıların kurs aktivite kullanımları arasındaki farklılıkları belirlemek, katılımcılara doğru çalışma stratejilerini belirleyebilmek için çok önemlidir. Bu sebeple tezde gerçekleştirdiğimiz ayrıntılı analizlerle ortaya konulan bulguların kullanılması, MOOC'larda çalışma deneyimlerinin geliştirilmesine ve farklı İngilizce dil gruplarındaki katılımcılara en uygun çalışma önerilerinin oluşturulmasında kullanılabilir.

Tezde, ilk olarak İngilizce dil gruplarına ayırma yöntemleri üzerinde çalışılmıştır. İngilizce gruplarına ayırmanın doğru yapılması, hem tez kapsamındaki MOOC analizlerin güvenilirliği hem de analiz sonuçlarında ortaya konulan bulguların MOOC platform sağlayıcıları tarafından katılımcıların başarı şansını arttırmak amacı kullanılarak ile platformları otomatik veya yarı otomatik olarak uyarlayabilmeleri adına önemlidir.

İngilizce dil gruplarındaki katılımcıların davranışlarının muntazam olup olmadıkları incelenip, aykırı değerler tespit edilmiştir. Ardından, oluşturulan özelliklerden kurs sonu performans özelliği ile yüksek derecede korelasyonlu olanlar tespit edilmiştir. İngilizce dil gruplarındaki katılımcıların sosyal etkileşimleri arasındaki farklılıklar da analiz edilerek, sonuçlara göre diğer kullanım özelliklerinden de en uygun olanlar seçilerek, İngilizce dil gruplarındaki katılımcıların kurs performanslarını (tamamlama, ayrılma veya yavaş ilerleme) tahmin için kullanılmışlardır.

Popüler MOOC platformlarının kaynak kodlarına sahip olmadığımız için, MOOC platformlarında doğrudan değişiklik yapamamız mümkün değildir. Bununla birlikte, önerdiğimiz İngilizce dil grubu tespit yöntemleri ile İngilizce dil gruplarındaki katılımcıların kullanım ve performans analizlerindeki bulgular, MOOC platform sağlayıcıları ve kurs içeriği oluşturanlar tarafından doğru çalışma stratejilerinin

önerilmesinde kullanılabilir. MOOC platformlarında katılımcıların ait oldukları İngilizce dil gruplarına göre kişiselleştirilmiş bir ortam sağlanabilirse, katılımcıların daha fazla etkileşimde bulunabilecekleri düşünülmektedir.

Katılımcıların kurs sonu performans tahminlerinde en önemli olan özellikler, kullanılan algoritmaya ve kullanılan İngilizce dil gruplarının verilerine bağlı olarak değişebilmektedir. Örneğin, "Understanding Language 4: Learning and Teaching" kursunda en iyi kurs sonu performans tahmini doğruluğuna sahip Rastgele Orman modelinde, İngilizce'yi ilk dil olarak konuşan katılımcılarda, 'tamamlanan birinci hafta adım sayısı' en belirleyici özellik olmasına rağmen, ikinci dil olarak İngilizce konuşan katılımcılarda 'başlanan adım sayısı' en belirleyici özellik olmuştur. Aynı kursta, denenen diğer altı sık kullanılan geleneksel kurs sonu performans tahmini algoritmasında, 'başlanan adım sayısı', tüm İngilizce dil grupların verisi ile kurs sonu performans tahmininde en belirleyici özellik olmuştur.

Tezde elde edilen sonuçlar aşağıdaki gibi özetlenebilir:

1. İngilizce dil gruplarının tespiti ile ilgili sonuçlar

- (a) Kurs öncesi anketlerinde katılımcıların verdikleri ülke bilgileri İngilizce dil gruplarının tespitinde kullanılabilmeyle birlikte, tek başına kullanıldığında elde edilen sonuçlarda %15'e kadar hatalı sonuçlar verebilmektedir.
- (b) Düzenli ifade kalıpları, katılımcılar MOOC forumlarında ilk dilleri, milletleri, şehirleri veya ülkeleri hakkında bilgi verdikleri süreçte, İngilizce'yi ikinci dil olarak konuşan katılımcıları tespit etme sürecini otomatikleştirmek için kullanışlıdır.
- (c) Kullanılan veri türüne, eğitim ve testte kullanılan kurs türüne ve Derin Öğrenme modeline ve denenen mimarilere (katmanlara) göre İngilizce dil grubu tahmini başarısı değişkenlik gösterebilmektedir. Derin öğrenme yöntemi, tüm yorum yapan katılımcıların İngilizce dil grubu tahminini yapabilmeyi sağlaması bakımından değerli olmakla birlikte, sonuçların güvenilirliği açısından, düzenli ifadelerle İngilizce dil grubu tespit yönteminin gerisinde kalmaktadır.

2. İngilizce dil gruplarının kullanım analizleri ile ilgili sonuçlar

- (a) İlk dili İngilizce olan katılımcılar, kursa katılımda genellikle diğer İngilizce dil gruplarındaki katılımcılardan daha etkinlerdir. Örneğin, forum tartışmalarına daha fazla sayıda yorum göndermekte, daha uzun yorumlar yazmaktadırlar ve başkaları tarafından takip edilmeleri daha muhtemeldir.

- (b) İlk dili İngilizce olan katılımcıların diğer İngilizce dil gruplarındaki katılımcılara göre daha fazla kurs adımı tamamlamaları beklenmektedir.
- (c) İngilizce dil gruplarındaki katılımcıların çoğu çok az sayıda yorum yaparken, daha fazla yorum yapan (aykırı değerlere sahip) katılımcıların genellikle İngilizce'yi ilk dili olarak konuşan katılımcılar grubundan olduğu görülmüştür.
- (d) İngilizce dil gruplarının davranışları arasındaki en belirgin fark, takip edilme davranışlarında görülmüştür. İlk dili İngilizce olan katılımcılar, diğer iki İngilizce dil grubundaki katılımcılardan çok daha fazla kişi tarafından takip edilmişlerdir. Bunun sebebi araştırılmaya ihtiyaç duyuyor olsa da sezgisel olarak, İngilizce'yi birincil ve resmi dil olarak konuşan katılımcılar tarafından yapılan yorumlardaki İngilizce dili hakimiyetinin başkalarının onları takip etmesine yol açmış olabileceği kanaatine varılmıştır.
- (e) Farklı İngilizce dil gruplarındaki katılımcılar, bazı aktivitelerde benzer davranışlar gösterebilir de davranışlarında önemli farklılıkların olduğu gözlemlenmiştir. Bu sebeple, katılımcıların İngilizce dil gruplarına göre platformların kişiselleştirilmesi gerektiği kanaatine varılmıştır.
- (f) İlk dili İngilizce olan katılımcıların diğer İngilizce dil gruplarındaki katılımcılara göre daha fazla sayıda kurs adımı tamamlama olasılıklarının yüksek olduğu ve forum tartışmalarına daha aktif olarak katıldıkları tespit edilmiştir. İkincil dil olarak İngilizce konuşan katılımcıları forum tartışmalarına daha aktif olarak katılmaları için cesaretlendiren MOOC platform tasarım ve altyapısı oluşturularak daha fazla sayıda kurs adımı tamamlamaları sağlanabilir.
- (g) İlk haftaya ait aktivitelerin kullanılma davranışları ile kurs tamamlama özelliği arasındaki korelasyonun bulunması daha önce Jiang ve diğerlerinin yaptığı çalışma sonuçlarını doğrulamıştır [115]. Bununla birlikte, Jiang ve diğerlerinin yaptığı çalışma, birinci haftadaki sosyal etkileşimlerin kurs tamamlama ile güçlü korelasyonunun olduğunu gösterirken, çalışmamızın bulguları, birinci haftadaki kurs adımlarıyla etkileşimin en belirleyici özellik olduğunu göstermektedir.

3. İngilizce dil gruplarının kurs sonu performans analizleri ile ilgili sonuçlar

- (a) Kurs sonuna kadar gerçekleştirilen aktivite kullanım verileri ile ve kümülatif olarak haftalık kurs verileri ile gerçekleştirilen kurs sonu performans tahmininde genel olarak en yüksek doğruluğa sahip olan iki algoritma Rastgele Orman ve Karar Ağacı algoritmaları olurken, Naive

Bayes ve Çok Katmanlı Algılayıcı algoritmaları en düşük doğruluğa sahip olmuşlardır. Bununla birlikte, kurs türüne ve kurs performans tahmininde kullanılan verinin kursun ne kadarlık kısmına ait olduğuna bağlı olarak, en başarılı ve başarısız olan algoritmaların değişebildiği gözlemlenmiştir.

- (b) Kurs performans tahmininde çoğunlukla 'başlanan adım sayısı', 'tamamlanan ilk hafta adım sayısı' ve 'tamamlanan ikinci hafta adım sayısı' en önemli özellikler olmakla birlikte; kurs türüne, kaç haftalık kurs verisi kullanıldığına ve kullanılan İngilizce dil grubu verisine bağlı olarak en önemli kurs performans tahmini özellikleri değişebilmektedir. Örneğin; "Understanding Language 4: Learning and Teaching" kursunda, 'ilk hafta tamamlanan adım sayısı (kurs adımı), en iyi kurs sonu performans tahmini doğruluğuna sahip Rastgele Orman algoritmasında, ilk dili İngilizce olan katılımcılar için en belirleyici özellik iken, 'başlanan adım sayısı' birincil dili İngilizce olmayan katılımcılar için (sadece resmi veya "ikincil dil olarak İngilizce" konuşanlar) en iyi kurs sonu performansı tahmin edici özellik olmuştur.
- (c) MOOC platformlarında "ikincil dil olarak İngilizce" konuşan katılımcılara ek öğrenme kolaylıkları sağlanmalıdır. Örneğin, platformdaki videolar otomatik olarak alt yazılı olarak verilebilir ve platform katılımcının "ikincil dil olarak İngilizce" konuşan katılımcıları tespit etmesi durumunda videolar daha yavaş iletilebilirler.

4. İngilizce dil gruplarının haftalık performans analizleri ile ilgili sonuçlar

- (a) Kurs tamamlama oranlarını iyileştirmek için, kurs devam ederken yardıma ihtiyaç duyan katılımcıları tespit edebilmek gerekmektedir. Bulgularımız, erken davranışların en önemli kurs performansı belirleyicileri olduklarını göstermektedir. Erken müdahale geri bildirimler, kursu başarılı tamamlama ihtimali düşük katılımcıları erkenden belirleyebilmeyi sağlamaktadır. Bu durum, MOOC'ların sürdürülebilirliği için de önemlidir.
- (b) Haftalık kurs performans tahmininde, kurs türüne bağlı olmaksızın çoğunlukla en başarılı tahmin sonucuna sahip algoritmalar Rastgele Orman ve Karar Ağacı algoritmaları olurken, Çok Kategorili Lojistik Regresyon ve Karar Destek Makineleri algoritmaları da bu algoritmalara yakın kurs performansı tahmin sonuçlarına sahip olmuşlardır. Naive Bayes ve Çok Katmanlı Algılayıcı algoritmaları çoğunlukla en başarısız performans tahminini gerçekleştiren algoritmalar olmuştur. Bununla birlikte, kurs tahminin kaç haftalık veri ile yapıldığına ve kullanılan İngilizce dil grubu verisine bağlı olarak başarılı ve başarısız olan algoritmalarda değişimler de

gözlemlenmiştir.

- (c) Algoritmaların haftalık performans tahminleri kümülatif olarak verisi kullanılan haftaların sayısı ile paralel olarak artış gösterirken, istisnai durumlar da oluşmuştur.
- (d) Haftalık kurs performans tahmininde en önemli olan özelliklerden 'başlanan adım sayısı' ve 'ilk hafta tamamlanan adım sayısı' gibi bazı özellikler kurs türüne bağlı olarak değişim göstermezken, kurs türüne bağlı olarak forum kullanımlarına bağlı özelliklerin veya kurs adımlarını tamamlama durumlarının da en önemli üç tahmin özelliği içinde yer aldıkları olmuştur.

Tezde, ilk olarak MOOC platformu katılımcılarının İngilizce dil gruplarının tespiti için yöntemler önerilmiştir. Önerilen İngilizce dil grubu tespit yöntemleri içinden en başarılı sonuç, oluşturulan düzenli ifadeler ve kurs öncesi ankelerindeki bilgilerle İngilizce dil grubu tespit yöntemi ile elde edilmiş ve sonraki aşamalarda bu yöntemin sonuçları kullanılmıştır. Ardından, İngilizce dil gruplarının sırası ile detaylı karşılaştırmalı kullanım analizleri, kurs sonu performans analizleri ve kümülatif olarak haftalık performans analizleri gerçekleştirilmiştir. Analizlerden elde edilen bulgular, MOOC platformu geliştiricileri ve kurs içerik sağlayıcılarının, katılımcıların İngilizce dil gruplarına göre kullanım ve performanslarını iyileştirilebilmelerini sağlayacak geliştirmeler yapabilmeleri için paylaşılmıştır.

7.1 Gelecek Çalışmalar

Gelecek çalışmalarda gerçekleştirilebilecek muhtemel iyileştirmelerden biri, katılımcıların İngilizce dilindeki akıcılıkları tespit edilerek İngilizce dili gruplandırma yöntemini iyileştirmeye çalışılmasıdır. Örneğin, tezde önerilen yöntemde aynı İngilizce dil grubunda yer alan Finlandiya ve Almanya'dan gelen katılımcıları aynı gruba yerleştirmek doğru olmayabilir. Çünkü Finlandiya ve Almanya'nın İngilizce'de farklı bir akıcılık seviyesine sahip olduğunu gösteren bazı istatistikler vardır. Aynı ülkeden gelen katılımcılarda bile İngilizce akıcılığı farklı seviyelerde olabilir. Örneğin İngiltere'de yaşayan iki Türk öğrencinin İngilizce bilgisi farklı olabilir. Bu durumda öğrencileri aynı kümeye yerleştirmek doğru olmayabilir. Bu sebeple, tez çalışma sonuçlarında gerçekleştirilebilecek potansiyel bir ilerleme, İngilizce ilk dili olmayan katılımcıları (sadece resmi dili İngilizce olan katılımcılar veya "ikincil dil olarak İngilizce" konuşan katılımcılar) dil akıcılıklarındaki farklılıklar ile otomatik olarak saptama yöntemleri belirleyerek İngilizce dil gruplarının oluşturulmasında bu yöntem sonuçlarının da dikkate alınmasıdır.

Tezde gerekleřtirilen İngilizce dil gruplarının davranıř ve performanslarının ayrıntılı analiz sonuçları, katılımcıların İngilizce dil gruplarına gre, en uygun alıřma kalıplarının onerilerini belirlemeye ve katılımcıların MOOC’lardaki alıřma deneyimlerini geliřtirmeye yardımcı olabilir. Bununla birlikte, FutureLearn platformundan kullanılan veriler, video etkileřimleri ile ilgili detaylı analizler iin yeterli ayrıntıyı iermemektedir. FutureLearn dıřındaki MOOC saėlayıcı platformların saėlayabileceėi detaylı video etkileřimleri gibi ek zellikler kullanılarak İngilizce dil grupları hakkında daha fazla bilgi edinilebileceėi dıřınılmektedir.

Tezde ortaya konulan bulgular, MOOC platformu saėlayıcıları (yazarlar ve platform oluřturucuları) tarafından, kursları ve platformları “ikincil dil olarak İngilizce” konuřan konuřmacıların en fazla yararlanabileceėi řekilde yeniden tasarlamak iin kullanılabilir.

İkincil dil olarak İngilizce konuřan katılımcıların davranıřları ile diėer İngilizce dil gruplarından katılımcıların davranıřları karřılařtırılarak elde edilen bulgular kullanılarak İngilizce dil gruplarındaki katılımcıların kurslarda daha bařarılı olabilmeleri iin zelleřtirilmiř destek ve mdahaleler saėlanabilirse, bu eřitsizlik potansiyel olarak ele alınabilir.

MOOC platformu saėlayıcılarına, MOOC platformunun sunduėu olanakların kullanımını ve aık kaynaklara eriřimi kolaylařtırmalarını saėlayacak platform tasarım stratejileri onerilebilir. Bylece, MOOC’larda azınlık dilleri konuřan katılımcılara daha etkin ve bařarılı kurs deneyimi sunulabilir.

A.1 İngilizce Dil Gruplarındaki Ülkeler ve ISO-2 Formatındaki Kodları

Tablo A.1 Birincil dili ve resmi dili İngilizce olan ülkelerin İngilizce isimleri ve ISO-2 formatındaki kodları

Ülke İsmi	Kod (ISO2)	Ülke İsmi	Kod (ISO2)
United States of America	US	United Kingdom	GB
Australia	AU	Sierra Leone	SL
Canada	CA	Ghana	GH
New Zealand	NZ	Nigeria	NG
Ireland	IE	Jersey	JE
Trinidad and Tobago	TT	Turks and Caicos Islands	TC
Guernsey	GG	Barbados	BB
Jamaica	JM	Antigua and Barbuda	AG
Bahamas	BS	Belize	BZ
Cook Islands	CK	Dominica	DM
Grenada	GD	Guyana	GY
Liberia	LR	Saint Kitts and Nevis	KN
Saint Lucia	LC	Saint Vincent and the Grenadines	VC
Anguilla	AI	Bermuda	BM
Virgin Islands (British)	VG	Cayman Islands	KY
Falkland Islands (Malvinas)	FK	Isle of Man	IM
Norfolk Island	NF	Papua New Guinea	PN
Virgin Islands (U.S.)	VI	British Indian Ocean Territory	IO
Guernsey	GG	Montserrat	MS
Saint Helena, Ascension and Tristan da Cunha	SH		

Tablo A.2 İngilizce yalnızca resmi dili olan ülkelerin İngilizce isimleri ve ISO-2 formatındaki kodları

Ülke İsmi	Kod (ISO2)	Ülke İsmi	Kod (ISO2)
Malaysia	MY	Kenya	KE
Pakistan	PK	Philippines	PH
Cameroon	CM	Botswana	BW
Sri Lanka	LK	Bangladesh	BD
Tanzania	TZ	Singapore	SG
India	IN	Namibia	NA
South Africa	ZA	Rwanda	RW
Fiji	FJ	Brunei Darussalam	BN
Malawi	MW	Sudan	SD
Israel	IL	Ethiopia	ET
Zimbabwe	ZW	South Sudan	SS
Uganda	UG	Zambia	ZM
Papua New Guinea	PG	Malta	MT
Lesotho	LS	Gambia	GM
Curaçao	CW	Hong Kong	HK
Puerto Rico	PR	Mauritius	MU
Burundi	BI	Micronesia (Federated States of)	FM
Kiribati	KI	Marshall Islands	MH
Nauru	NR	Nepal	NP
Niue	NU	Palau	PW
Samoa	WS	Seychelles	SC
Solomon Islands	SB	Eswatini	SZ
Tonga	TO	Tuvalu	TV
Vanuatu	VU	Bahrain	BH
Cyprus	CY	Eritrea	ER
Jordan	JO	Kuwait	KW
Maldives	MV	Myanmar	MM
Qatar	QA	United Arab Emirates	AE
American Samoa	AS	Gibraltar	GI
Guam	GU	Northern Mariana Islands	MP
Sint Maarten (Dutch part)	SX	Christmas Island	CX
Cocos (Keeling) Islands	CC	Tokelau	TK

A.2 Dil, Ülke ve Millet Bilgisi Tespit İçin Oluşturulan Düzenli İfadeler ve Sembolleri

Tablo A.3 Katılımcı yorumlarından dil bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları

I1	\\b(?:our ma?y \\s+(?:first 1st mother(?:\\s)? native official main national primary best)\\s*)*?\\s\\s\\s*(?:t[aou]{1,2}n[gq](?:[ue]{1,2}s)? l[ae]nguage speaker)(?:\\s+(?:l[ae]nguage\\s+)?(?:it which was what)\\s+)?(?:is was are)\\s+(?:mainly\\s+)?(?:the\\s+)? ,\\s+)(\\w{4,})
I2	\\b(?:i\\s*\\s*m i\\s*am we we\\s+are we\\s+re)\\s+(?:a\\s+)?native\\s+(\\w{4,})
I3	\\b(\\w{4,})\\s+(?:\\s*(\\s+)?(?:l[ae]nguage\\s+)?(?:is\\s+ as\\s+)?(?:also\\s+)?(?:ma?y\\s+)?(?:first 1st mother native of+icial main national primary best)\\s\\s\\s*(?:t[aou]{1,2}n[gq](?:[ue]{1,2})? l[ae]nguan?ge speaker)
I4	\\b(?:i\\s*\\s*m i\\s*am im)\\s+(?:a an)\\s+(\\w{4,})\\s+(?:l[ae]nguage\\s+)?native\\s+speaker
I5	\\b(?:i we my)\\s[\\s\\s]*?(\\w{4,})\\s+(?:is as)\\s+(?:ma?y\\s+ a\\s+)?(?:second)\\s+(?:foreign\\s+)?l[ae]nguage
I6	\\b(?:i we i\\s*\\s*m i\\s*am im)\\s+speak\\s+(?:only\\s+)?(\\w{4,})
I7	growing up with (\\w{4,})
I8	\\b(?:i we)\\s+use\\s+(\\w{4,})\\s+as\\s+(?:ma?y\\s+)?(?:first mother native official main national primary best)\\s\\s\\s*(?:t[aou]{1,2}n[gq](?:[ue]{1,2})? l[ae]nguage speaker)
I9	\\b(?:i we)\\s*(?:was were)?[\\s\\s]*?(?:born grown grown\\s+up grew grew\\s+up brought\\s+up)[\\s\\s]*?(?:speaking talking)(?:both\\s+)?\\s+(\\w{4,})
I10	\\b(?:i we)[\\s\\s]*?only\\s+speak\\s+(?:one\\s+)?l[ae]nguage\\s+(?:the\\s+) (?:[-]\\s+)?(\\w{4,})
I11	\\b(\\w{4,})\\s+which\\s+is\\s+(?:our ma?y)\\s+(?:first 1st mother native official main national primary best)\\s\\s\\s*(?:t[aou]{1,2}n[gq](?:[ue]{1,2})? l[ae]nguage speaker)
I12	\\b(?:i\\s*\\s*m i(?:\\s+too)?\\s*am im)\\s+a?\\s+(?:native\\s+)?speaker\\s+of\\s+(\\w{4,})

Tablo A.4 Katılımcı yorumlarından ülke bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları

c1	\\b(?:i we)\\s+(?:was were)?[^\.\s]*?(?:brought\\s+up born grow\\s+up grew\\s+up grew grown grown\\s+up)\\s[^\.\s]*?in\\s+(?:a?\\s+part\\s+of\\s+)?(?:the\\s+)?(?:east(?:ern)? north(?:ern)? west(?:ern)? south(?:ern)?\\s+)?(?:of\\s+)*?(\\w{2,})
c2	\\b(?:i we)\\s+(?:grew up grow up)\\s+in\\s+(?:the\\s+)?(?:east(?:ern)? north(?:ern)? west(?:ern)? south(?:ern)?\\s+)*?(?:of\\s+)?(\\w{2,})
c3	\\b(?:i we i\\s*' \\s*a?m i\\s*am im (?my\\s+name\\s+is\\s+\\w+))\\s+[^\.\s]*?(?:c[oa]me\\s+)?from(?:\\s+here)?(?:\\s+the)?\\s+(?:east(?:ern)? north(?:ern)? west(?:ern)? south(?:ern)?\\s+)?(?:of\\s+)?(\\w{2,})(?<!\\bmy)(?:\\s+republic)?
c4	\\b(?:having i\\s+was)\\s+grown\\s+up\\s+in\\s+(?:east(?:ern)? north(?:ern)? west(?:ern)? south(?:ern)?\\s+)*?(?:of\\s+)?(\\w{2,})
c5	\\b(?:my our)\\s+country\\s+(?:is\\s+)?(\\w{2,})
c6	\\b(?:i\\s*' \\s*m i\\s*am we we\\s+are we\\s+re)[^\.\s]{1,20}from(?:[^\.\s]{1,10})[,,-]\\s+(\\w+)(?<!\\bmy)

Tablo A.5 Katılımcı yorumlarından millet bilgilerini tespit etmek için oluşturulan düzenli ifade kalıpları ve kullanılan kısaltmaları

n1	\\b(?:i\\s*' \\s*m i\\s*am im we am)\\s+(?:a\\s+ an\\s+ are\\s+)?(?:typical\\s+)?(\\w{4,})\\b[^\.\s]*?(?!\\s+\\w+[eo]r)
n2	\\b[^\\w]as\\s+a\\s+(?:native\\s+)?(\\w{4,})

- [1] N. DEMIRCI, "What is Massive Open Online Courses (MOOCs) and what is promising us for learning?: A review-evaluative article about MOOCs.," *Necatibey Faculty of Education Electronic Journal of Science & Mathematics Education*, vol. 8, no. 1, 2014.
- [2] K. C. Green, "Massive Open Online Courses (MOOCs) and other digital initiatives," *Journal of Collective Bargaining in the Academy*, no. 8, p. 10, 2013.
- [3] F. Almutairi, "The Impact of Integrating MOOCs into Campus Courses on Student Engagement," PhD thesis, University of Southampton, 2018.
- [4] S. White, H. Davis, K. Dickens, M. Leon, and M. M. Sanchez-Vera, "MOOCs: What motivates the producers and participants?" In *International Conference on Computer Supported Education*, Springer, 2014, pp. 99–114.
- [5] F. Brouns and O. Firssova, "The role of learning design and Learning Analytics in MOOCs," 2016.
- [6] L. Guàrdia, M. Maina, and A. Sangrà, "MOOC design principles: A pedagogical approach from the learner's perspective," *eLearning Papers*, no. 33, 2013.
- [7] A. S. Sunar, S. White, N. A. Abdullah, and H. C. Davis, "How learners' interactions sustain engagement: A MOOC case study," *IEEE Transactions on Learning Technologies*, vol. 10, no. 4, pp. 475–487, Oct. 2017, ISSN: 2372-0050. DOI: 10.1109/TLT.2016.2633268.
- [8] S. White and S. White, "Learning designers in the 'third space': The socio-technical construction of MOOCs and their relationship to educator and learning designer roles in HE," *Journal of Interactive Media in Education*, no. 1, 2016.
- [9] Y. Feng, D. Chen, Z. Zhao, P. Xi, and H. Chen, "The impact of students and TAs' participation on students' academic performance in MOOC," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Aug. 2015, pp. 1149–1154. DOI: 10.1145/2808797.2809428.
- [10] G. Kennedy, C. Coffrin, P. De Barba, and L. Corrin, "Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance," in *Proceedings of the fifth international conference on learning analytics and knowledge*, ACM, 2015, pp. 136–140.
- [11] M. Klüsener and A. Fortenbacher, "Predicting students' success based on forum activities in MOOCs," in *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2, Sep. 2015, pp. 925–928. DOI: 10.1109/IDAACS.2015.7341439.

- [12] C. Ye and G. Biswas, "Early prediction of student dropout and performance in MOOCs using higher granularity temporal information," *Journal of Learning Analytics*, vol. 1, no. 3, pp. 169–172, 2014.
- [13] B. Xu and D. Yang, "Motivation classification and grade prediction for MOOCs learners," *Computational intelligence and neuroscience*, vol. 2016, p. 4, 2016.
- [14] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, "Visualizing patterns of student engagement and performance in MOOCs," in *Proceedings of the fourth international conference on learning analytics and knowledge*, ACM, 2014, pp. 83–92.
- [15] M. K. Khribi, M. Jemni, and O. Nasraoui, "Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval," in *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, Jul. 2008, pp. 241–245. DOI: 10.1109/ICALT.2008.198.
- [16] T. R. Liyanagunawardena, S. Williams, and A. A. Adams, "The impact and reach of MOOCs: A developing countries' perspective," *eLearning Papers*, pp. 38–46, 2014.
- [17] A. Bakki, L. Oubahssi, C. Cherkaoui, and S. George, "Motivation and engagement in MOOCs: How to increase learning motivation by adapting pedagogical scenarios?" In *Design for Teaching and Learning in a Networked World*, Springer, 2015, pp. 556–559.
- [18] H. A. Fasihuddin, G. D. Skinner, and R. I. Athauda, "Boosting the opportunities of open learning (MOOCs) through learning theories," *GSTF Journal on Computing (JoC)*, vol. 3, no. 3, p. 112, 2013.
- [19] D. Shah, *By the numbers: MOOCs in 2015-class central's MOOC report*, 2015.
- [20] T. Eriksson, T. Adawi, and C. Stöhr, "“time is the bottleneck”: A qualitative study exploring why learners drop out of MOOCs," *Journal of Computing in Higher Education*, vol. 29, no. 1, pp. 133–146, 2017.
- [21] T. R. Dillahunt, B. Z. Wang, and S. Teasley, "Democratizing higher education: Exploring MOOC use among those who cannot afford a formal education," *The International Review of Research in Open and Distributed Learning*, vol. 15, no. 5, 2014.
- [22] N. Mamgain, A. Sharma, and P. Goyal, "Learner's perspective on video-viewing features offered by MOOC providers: Coursera and edX," in *2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)*, Dec. 2014, pp. 331–336. DOI: 10.1109/MITE.2014.7020298.
- [23] I. Duru, A. S. Sunar, G. Dogan, and S. White, "Challenges of identifying second language English speakers in MOOCs," in *European Conference on Massive Open Online Courses*, Springer, 2017, pp. 188–196.
- [24] M. Barak, A. Wattad, and H. Haick, "Motivation to learn in Massive Open Online Courses: Examining aspects of language and social engagement," *Computers & Education*, vol. 94, pp. 49–60, 2016.

- [25] R. F. Kizilcec, C. Piech, and E. Schneider, “Deconstructing disengagement: Analyzing learner subpopulations in Massive Open Online Courses,” in *Proceedings of the third international conference on learning analytics and knowledge*, ACM, 2013, pp. 170–179.
- [26] N. Gillani, R. Eynon, M. Osborne, I. Hjorth, and S. Roberts, “Communication communities in MOOCs,” *arXiv preprint arXiv:1403.4640*, 2014.
- [27] S. S. Kumar and P. M. Shastry, “Analysis of student engagement and course completion in Massive Open Online Courses,” in *Integrated Intelligent Computing, Communication and Security*, Springer, 2019, pp. 447–458.
- [28] M.-H. Cho and M.-K. Byun, “Nonnative English-speaking students’ lived learning experiences with MOOCs in a regular college classroom,” *The International Review of Research in Open and Distributed Learning*, vol. 18, no. 5, 2017.
- [29] R. F. Kizilcec, A. J. Saltarelli, J. Reich, and G. L. Cohen, “Closing global achievement gaps in MOOCs,” *Science*, vol. 355, no. 6322, pp. 251–252, 2017.
- [30] C. Milligan, A. Littlejohn, and A. Margaryan, “Patterns of engagement in connectivist MOOCs,” *MERLOT Journal of Online Learning and Teaching*, vol. 9, no. 2, 2013.
- [31] J. Uchidiuno, A. Ogan, K. R. Koedinger, E. Yarzebinski, and J. Hammer, “Browser language preferences as a metric for identifying ESL speakers in MOOCs,” in *Proceedings of the Third ACM Conference on Learning@ Scale*, 2016, pp. 277–280.
- [32] C. Heiner, N. Heffernan, and T. Barnes, “Educational data mining,” in *Supplementary Proceedings of the 12th International Conference of Artificial Intelligence in Education*, 2007.
- [33] R. Baker *et al.*, “Data mining for education,” *International encyclopedia of education*, vol. 7, no. 3, pp. 112–118, 2010.
- [34] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010, ISSN: 1558-2442. DOI: 10.1109/TSMCC.2010.2053532.
- [35] S. B. Shum and R. Ferguson, “Social learning analytics,” *Journal of educational technology & society*, vol. 15, no. 3, pp. 3–26, 2012.
- [36] D. T. Tempelaar, B. Rienties, and B. Giesbers, “In search for the most informative data for feedback generation: Learning Analytics in a data-rich context,” *Computers in Human Behavior*, vol. 47, pp. 157–167, 2015.
- [37] T. Elias, *Learning analytics: Definitions, processes and potential*. retrieved july 29, 2011, 2011.
- [38] C. Bach, “Learning analytics: Targeting instruction, curricula and student support,” *Office of the Provost, Drexel University*, 2010.
- [39] D. Clow, “The learning analytics cycle: Closing the loop effectively,” 2012.
- [40] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT Press Cambridge, 2016, vol. 1.

- [41] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [42] X. Wei, H. Lin, L. Yang, and Y. Yu, "A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification," *Information*, vol. 8, no. 3, p. 92, 2017.
- [43] N. M. Dowell, O. Skrypnyk, S. Joksimovic, A. C. Graesser, S. Dawson, D. GaLevic, T. A. Hennis, P. de Vries, and V. Kovanovic, "Modeling learners' social centrality and performance through language and discourse," *International Educational Data Mining Society*, 2015.
- [44] C. Tucker, B. K. Pursel, and A. Divinsky, "Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes," *The ASEE Computers in Education (CoED) Journal*, vol. 5, no. 4, p. 84, 2014.
- [45] J. He, J. Bailey, B. I. Rubinstein, and R. Zhang, "Identifying at-risk students in Massive Open Online Courses," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [46] S. I. El Ahrache, H. Badir, Y. Tabaa, and A. Medouri, "Massive Open Online Courses: A new dawn for higher education?" *International Journal on Computer Science and Engineering*, vol. 5, no. 5, p. 323, 2013.
- [47] M. Blagojević and D. Milošević, "Massive Open Online Courses: Edx vs Moodle MOOC," in *Proc. 5th International Conference on Information Society and Technology, Kopaonik, Serbia*, 2015, pp. 346–351.
- [48] L. Pappano, "The year of the MOOC," *The New York Times*, vol. 2, no. 12, p. 2012, 2012.
- [49] J. Kerr, S. Houston, L. Marks, and A. Richford, "Building and Executing MOOCs: A practical review of Glasgow's first two MOOCs (Massive Open Online Courses)," 2015.
- [50] R. Ferguson and M. Sharples, "Innovative pedagogy at massive scale: Teaching and learning in MOOCs," in *European Conference on Technology Enhanced Learning*, Springer, 2014, pp. 98–111.
- [51] M. A. Chatti, V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti, and U. Schroeder, "Learning analytics: Challenges and future research directions," *eled*, vol. 10, no. 1, 2014.
- [52] M. M. Ashenafi, G. Riccardi, and M. Ronchetti, "Predicting students' final exam scores from their course activities," in *2015 IEEE Frontiers in Education Conference (FIE)*, Oct. 2015, pp. 1–9. DOI: 10.1109/FIE.2015.7344081.
- [53] A. Ramesh, D. Goldwasser, B. Huang, H. Daume III, and L. Getoor, "Uncovering hidden engagement patterns for predicting learner performance in MOOCs," in *Proceedings of the first ACM conference on Learning@ scale conference*, ACM, 2014, pp. 157–158.
- [54] S. Papadakis, M. Kalogiannakis, E. Sifaki, and N. Vidakis, "Access Moodle using smart mobile phones. A case study in a Greek University," in *Interactivity, Game Creation, Design, Learning, and Innovation*, Springer, 2017, pp. 376–385.

- [55] A. Wolff, Z. Zdrahal, D. Herrmannova, and P. Knuth, "Predicting student performance from combined data sources," in *Educational data mining*, Springer, 2014, pp. 175–202.
- [56] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong, "Learning about social learning in MOOCs: From statistical analysis to generative model," *IEEE Transactions on Learning Technologies*, vol. 7, no. 4, pp. 346–359, Oct. 2014, ISSN: 2372-0050. DOI: 10.1109/TLT.2014.2337900.
- [57] J. DeBoer and L. Breslow, "Tracking progress: Predictors of students' weekly achievement during a circuits and electronics MOOC," in *Proceedings of the first ACM conference on Learning@ scale conference*, ACM, 2014, pp. 169–170.
- [58] J. Liang, C. Li, and L. Zheng, "Machine learning application in MOOCs: Dropout prediction," in *2016 11th International Conference on Computer Science Education (ICCSE)*, Aug. 2016, pp. 52–57. DOI: 10.1109/ICCSE.2016.7581554.
- [59] E. Paradis, J. Claude, and K. Strimmer, "APE: Analyses of phylogenetics and evolution in R language," *Bioinformatics*, vol. 20, no. 2, pp. 289–290, 2004.
- [60] T. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, "Behavior-based grade prediction for MOOC via Time Series Neural Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, Aug. 2017, ISSN: 1941-0484. DOI: 10.1109/JSTSP.2017.2700227.
- [61] E. Młynarska, D. Greene, and P. Cunningham, "Indicators of good student performance in Moodle activity data," *arXiv preprint arXiv:1601.02975*, 2016.
- [62] P. De Barba, G. E. Kennedy, and M. Ainley, "The role of students' motivation and participation in predicting performance in a MOOC," *Journal of Computer Assisted Learning*, vol. 32, no. 3, pp. 218–231, 2016.
- [63] C. Alario-Hoyos, P. J. Munoz-Merino, M. Perez Sanagustin, C. Delgado Kloos, and H. A. Parada G, "Who are the top contributors in a MOOC? Relating participants' performance and contributions," *Journal of Computer Assisted Learning*, vol. 32, no. 3, pp. 232–243, 2016.
- [64] G. Hughes and C. Dobbins, "The utilization of data analysis techniques in predicting student performance in Massive Open Online Courses (moocs)," *Research and practice in technology enhanced learning*, vol. 10, no. 1, p. 10, 2015.
- [65] W. Xing and D. Du, "Dropout prediction in MOOCs: Using Deep learning for personalized intervention," *Journal of Educational Computing Research*, vol. 57, no. 3, pp. 547–570, 2019.
- [66] W. Wang, H. Yu, and C. Miao, "Deep model for dropout prediction in MOOCs," in *Proceedings of the 2nd International Conference on Crowd Science and Engineering*, ACM, 2017, pp. 26–32.
- [67] S. Tang and Z. A. Pardos, "Personalized behavior recommendation: A case study of applicability to 13 courses on edX," in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, ACM, 2017, pp. 165–170.

- [68] O. Almatrafi, A. Johri, and H. Rangwala, "Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums," *Computers & Education*, vol. 118, pp. 1–9, 2018.
- [69] D. S. Chaplot, E. Rhim, and J. Kim, "Predicting student attrition in MOOCs using sentiment analysis and neural networks," in *AIED Workshops*, vol. 53, 2015, pp. 54–57.
- [70] F. Brouns, N. Serrano Mart'inez-Santos, J. Civera, M. Kalz, and A. Juan, "Supporting language diversity of European MOOCs with the EMMA platform," 2015.
- [71] W. Wu and Q. Bai, "Why do the MOOC learners drop out of the school? – based on the investigation of MOOC learners on some Chinese MOOC platforms," in *2018 1st International Cognitive Cities Conference (IC3)*, Aug. 2018, pp. 299–304. DOI: 10.1109/IC3.2018.00039.
- [72] Q. Wang, B. Chen, Y. Fan, and G. Zhang, "MOOCs as an alternative for teacher professional development. examining learner persistence in one Chinese MOOC," 2018.
- [73] T. Beaven, T. Codreanu, and A. Creuzé, "Motivation in a language MOOC: Issues for course designers," in *Language MOOCs: Providing Learning, Transcending Boundaries*, De Gruyter Open, 2014, pp. 48–66.
- [74] S. Türkay, H. Eidelman, Y. Rosen, D. Seaton, G. Lopez, and J. Whitehill, "Getting to know English language learners in MOOCs: Their motivations, behaviors, and outcomes," in *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, ACM, 2017, pp. 209–212.
- [75] J. O. Uchidiuno, A. Ogan, E. Yarzebinski, and J. Hammer, "Going global: Understanding English language learners' student motivation in English-language MOOCs," *International Journal of Artificial Intelligence in Education*, pp. 1–25, 2017.
- [76] J. Uchidiuno, K. Koedinger, J. Hammer, E. Yarzebinski, and A. Ogan, "How do English language learners interact with different content types in MOOC videos?" *International Journal of Artificial Intelligence in Education*, vol. 28, no. 4, pp. 508–527, 2018.
- [77] H. Obari and S. Lambacher, "Impact of a blended environment with m-learning on EFL skills.," in *Proceedings of the 2014 EUROCALL Conference, Groningen, the Netherlands*, doi:10.14705/rpnet.2014.000229, 2014, pp. 267–272.
- [78] T. Nie and J. Hu, "EFL students' satisfaction with the college English education in the MOOC: An empirical study," *atmosphere*, vol. 3, p. 92, 2018.
- [79] C. Fuchs, "The Structural and dialogic aspects of language Massive Open Online Courses (LMOOCs): A case study," in *Computer-Assisted Language Learning: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2019, pp. 1540–1562.
- [80] E. Martín-Monje, M. D. Castrillo, and J. Mañana-Rodríguez, "Understanding online interaction in language MOOCs through learning analytics," *Computer Assisted Language Learning*, vol. 31, no. 3, pp. 251–272, 2018.

- [81] K. Qian and S. Bax, "Beyond the language classroom: Researching MOOCs and other innovations," Research-publishing. net, <https://doi.org/10.14705/rpnet.2017.mooc2016.668>, 2017, pp. 15–27.
- [82] Y. Tan, X. Zhang, H. Luo, Y. Sun, and S. Xu, "Learning profiles, behaviors and outcomes: Investigating international students' learning experience in an English MOOC," in *2018 International Symposium on Educational Technology (ISET)*, Jul. 2018, pp. 214–218. DOI: 10.1109/ISET.2018.00055.
- [83] R. Hemavathy and S. Harshini, "Adaptive Learning in computing for non-English speakers," *J Comput Sci Syst Biol*, vol. 10, pp. 061–63, 2017.
- [84] P. J. Guo, "Non-native English speakers learning computer programming: Barriers, desires, and design opportunities," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, p. 396.
- [85] J. Uchidiuno, A. Ogan, E. Yarzebinski, and J. Hammer, "Understanding ESL students' motivations to increase MOOC accessibility," in *Proceedings of the third (2016) ACM conference on Learning@ Scale*, ACM, 2016, pp. 169–172.
- [86] J.-F. Colas, P. B. Sloep, and M. Garreta-Domingo, "The effect of multilingual facilitation on active participation in MOOCs," *The International Review of Research in Open and Distributed Learning*, vol. 17, no. 4, 2016.
- [87] O. A. Aboshady, A. E. Radwan, A. R. Eltaweel, A. Azzam, A. A. Aboelnaga, H. A. Hashem, S. Y. Darwish, R. Salah, O. N. Kotb, A. M. Afifi, *et al.*, "Perception and use of Massive Open Online Courses among medical students in a developing country: Multicentre cross-sectional study," *BMJ open*, vol. 5, no. 1, e006804, 2015.
- [88] Y. Rimbaud, T. McEwan, A. Lawson, and S. Cairncross, "Adaptive learning in computing for non-native speakers," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, Oct. 2014, pp. 1–4. DOI: 10.1109/FIE.2014.7044142.
- [89] I. De Waard and K. Demeulenaere, "The MOOC-CLIL project: Using MOOCs to increase language, and social and online learning skills for 5th grade K-12 students," *Beyond the language classroom: researching MOOCs and other innovations*, pp. 29–42, 2017.
- [90] E. D. Reilly, K. M. Williams, R. E. Stafford, S. B. Corliss, J. C. Walkow, and D. K. Kidwell, "Global times call for global measures: Investigating automated essay scoring in linguistically-diverse MOOCs," *Online Learning*, vol. 20, no. 2, pp. 217–229, 2016.
- [91] S. Calvo, A. Morales, and J. Wade, "The use of MOOCs in social enterprise education: An evaluation of a North–South collaborative FutureLearn program," *Journal of Small Business & Entrepreneurship*, vol. 31, no. 3, pp. 201–223, 2019.
- [92] N. M. Castillo, J. Lee, F. T. Zahra, and D. A. Wagner, "MOOCs for development: Trends, challenges, and opportunities," *International Technologies & International Development*, vol. 11, no. 2, p. 35, 2015.
- [93] I. Duru, G. Dogan, and D. Banu, "EdX kullanıcılarının aktivite Kullanımlarının ve cinsiyetlerinin performansa etkilerinin incelenmesi," in *Uluslararası Bilgisayar Bilimleri ve Mühendisliği*, 2016.

- [94] I. Duru, G. Dogan, and B. Diri, "An overview of studies about students' performance analysis and learning analytics in MOOCs," in *2016 IEEE International Conference on Big Data (Big Data)*, Dec. 2016, pp. 1719–1723. DOI: 10.1109/BigData.2016.7840786.
- [95] G. Dogan, A. S. Sunar, I. Duru, and S. White, "Who is the English as a second language speaker in this MOOC?" *International Journal of Information and Education Technology*, vol. 8, no. 3, 2018.
- [96] I. Duru, A. S. Sunar, G. Dogan, and B. Diri, "Investigation of social contributions of language oriented MOOC learner groups," in *2017 International Conference on Computer Science and Engineering (UBMK)*, Oct. 2017, pp. 667–670. DOI: 10.1109/UBMK.2017.8093493.
- [97] I. Duru, B. Diri, M. E. Özçevik, K. Ataseven, G. Dogan, and S. White, "Analysis of English language groups with regular expressions," in *2018 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2018, pp. 1–5. DOI: 10.1109/ASYU.2018.8554018.
- [98] I. Duru, G. Dogan, A. Demir, H. Ata, B. Diri, and S. White, "English language proficiency level classification of students in a Massive Open Online Course," in *7th International Conference on Advanced Technologies (ICAT'18)*, pp. 424–432.
- [99] B. D. Ripley, "The R project in statistical computing," *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network*, vol. 1, no. 1, pp. 23–25, 2001.
- [100] A. G. Picciano, "The evolution of big data and learning analytics in American higher education," *Journal of Asynchronous Learning Networks*, vol. 16, no. 3, pp. 9–20, 2012.
- [101] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile networks and applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [102] G. Siemens and P. Long, "Penetrating the fog: Analytics in learning and education," *EDUCAUSE review*, vol. 46, no. 5, p. 30, 2011.
- [103] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 2014, pp. 60–65.
- [104] J. E. Friedl, *Mastering regular expressions*. "O'Reilly Media, Inc.", 2002.
- [105] E. S. Acosta and J. J. E. Otero, "Automated assessment of free text questions for MOOC using regular expressions," *Information Resources Management Journal (IRMJ)*, vol. 27, no. 2, pp. 1–13, 2014.
- [106] H. Shukla and M. Kakkar, "Keyword extraction from educational video transcripts using NLP techniques," in *2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)*, Jan. 2016, pp. 105–108. DOI: 10.1109/CONFLUENCE.2016.7508096.
- [107] Y.-H. An, M. K. Chandrasekaran, M.-Y. Kan, and Y. Fu, "The MUIR Framework: Cross-linking MOOC resources to enhance discussion forums," in *International Conference on Theory and Practice of Digital Libraries*, Springer, 2018, pp. 208–219.

- [108] K. Thiyagarajan, “Higher education and practice of English in India,” *Language in India*, vol. 8, no. 8, 2008.
- [109] P. Liao, Y. Sun, S. Ye, X. Li, G. Su, and Y. Sun, “Predicting learners’ multi-question performance based on neural networks,” in *2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC)*, Oct. 2017, pp. 1–6. DOI: 10.1109/BESC.2017.8357663.
- [110] Z. Liu, F. Xiong, K. Zou, and H. Wang, “Predicting learning status in MOOCs using LSTM,” *arXiv preprint arXiv:1808.01616*, 2018.
- [111] I. Duru, A. Sunar, S. White, B. Diri, and G. Dogan, “A case study on English as a second language speakers for sustainable MOOC study,” *Sustainability*, vol. 11, no. 10, p. 2808, 2019.
- [112] J. Pennington, R. Socher, and C. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [113] J. Li, X. Chen, E. Hovy, and D. Jurafsky, “Visualizing and understanding neural models in NLP,” *arXiv preprint arXiv:1506.01066*, 2015.
- [114] D. Clow, “MOOCs and the funnel of participation,” in *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM, 2013, pp. 185–189.
- [115] S. Jiang, A. Williams, K. Schenke, M. Warschauer, and D. O’Dowd, “Predicting MOOC performance with week 1 behavior,” in *Proceedings of the 7th International Conference on Educational Data Mining 2014*, 2014.

İletişim Bilgileri: ismailduru57@gmail.com

Makale

1. I. Duru, A. Sunar, S. White, B. Diri, and G. Dogan, "A case study on English as a second language speakers for sustainable MOOC study," Sustainability, vol. 11, no. 10, p. 2808, 2019.
2. G. Dogan, A.S. Sunar, I. Duru, S. White, "Who is the English as a second language speaker in this MOOC?," International Journal of Information and Education Technology, vol.8, no. 3, 2018

Konferans Bildirisi

1. I. Duru, B. Diri, M. E. Özçevik, K. Ataseven, G. Dogan, and S. White, "Analysis of English language groups with regular expressions," in 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), Oct. 2018, pp. 1–5. DOI: 10.1109/ASYU.2018.8554018.
2. I. Duru, A. S. Sunar, G. Dogan, and S. White, "Challenges of identifying second language English speakers in MOOCs," in European Conference on Massive Open Online Courses, Springer, 2017, pp. 188–196.
3. I. Duru, A. S. Sunar, G. Dogan, and B. Diri, "Investigation of social contributions of language oriented MOOC learner groups," in 2017 International Conference on Computer Science and Engineering (UBMK), Oct. 2017, pp. 667–670. DOI: 10.1109/UBMK.2017.8093493.
4. I. Duru, G. Dogan, and B. Diri, "An overview of studies about students' performance analysis and learning analytics in MOOCs," in 2016 IEEE International Conference on Big Data (Big Data), Dec. 2016, pp. 1719–1723. DOI: 10.1109/BigData.2016.7840786.
5. I. Duru, G. Dogan, and B. Diri, "EdX kullanıcılarının aktivite kullanımlarının ve cinsiyetlerinin performans etkilerinin incelenmesi" in 2016 International Conference on Computer Science and Engineering (UBMK). IEEE, 2016.
6. I. Duru, G. Dogan, A. Demir, H. Ata, B. Diri, S. White, "English language proficiency level classification Of students in a Massive Open Online Course", 7th International Conference on Advanced Technologies (ICAT), 2018

Proje

1. 2016-2019, “Uzaktan Eğitim Platformlarında Öğrenci Performans Tahmininin Ve Ölçümünün Gerçekleştirilmesi“, '01/04/2016 DOP05' nolu Yıldız Teknik Üniversitesi BAP (DOP) Projesi, Araştırmacı
2. 2018-2019, “1059B141601346” nolu Tübitak 2214-A projesi, Bursiyer