



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



Doktora Tezi

**Harezmi Eğitim Modeli Uygulama Verisinin Metin Madenciliği
Teknikleriyle Değerlendirilmesi**

Ali ÇİMEN

Enformatik Anabilim Dalı

Enformatik Programı

**DANIŞMAN
Prof. Dr. Sevinç GÜLSEÇEN**

Aralık, 2022

İSTANBUL

Bu alıřma, 15.12.2022 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Prof. Dr. Sevin GÜLSEÇEN (Danışman)
İstanbul Üniversitesi
Fen Fakültesi

Prof. Dr. Şirin KARADENİZ
Baheşehir Üniversitesi
Eğitim Fakültesi

Do. Dr. Zümrüt SATI
İstanbul Üniversitesi
Siyasal Bilgiler Fakültesi

Do. Dr. Mihriban Betül YILMAZ
Yıldız Teknik Üniversitesi
Eğitim Fakültesi

Do. Dr. Çiğdem EROL
İstanbul Üniversitesi
Enformatik Bölümü

İntihal Programı Beyanı

20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

Tezden Üretilmiş Yayınların Künye Bilgileri

Çimen, A. and Gülseçen, S., 2022, Harezmi eğitim modeli verisindeki gizli desenlerin ortaya çıkarılması, *Turkish Studies - Education*, 17 (6), 1395-1419.



ÖNSÖZ

Doktora hayatımın başından beri desteğini her zaman üzerimde hissettiğim, her daim bana güvenen, akademik ve yönetici kişiliği, vatanseverliği ve akademik bakış açısıyla her daim kendime örnek aldığım danışman hocam Prof. Dr. Sevinç GÜLSEÇEN'e ve manevi tez danışmanım olarak mirasını her daim omzumda hissettiğim ve huzurunda saygıyla ve şerefle eğildiğim Şehit Prof. Dr. İlhan VARANK hocama teşekkürlerimi sunarım.

Kendi tezleri gibi tez sürecimi sonuna kadar büyük bir titizlikle takip eden, bilgi, birikim ve akademik gücünü benden esirgemeyen, hoca olmanın yanında arkadaş olan değerli hocam Doç. Dr. Zümrüt SATI'ya da ayrıca teşekkürü borç bilmekteyim.

Bahçeşehir Üniversitesi Rektörü değerli hocam Prof. Dr. Şirin KARADENİZ'e Harezmi Eğitim Modeli çalışmalarına ve şahsıma model çalışmalarının başlangıcından beri verdiği destek, her daim yanımda olması ile de motivasyon sağlaması, tez sürecimi de bir o kadar titizlikle takip etmesi adına teşekkür ederim.

Yüksek lisans tez danışmanım ve aynı zamanda tez savunma sınavı komisyon üyesi hocam Doç. Dr. M. Betül YILMAZ ve doktora sürecinde her daim yanımda olan değerli hocam Doç. Dr. Çiğdem EROL hocalarıma da teşekkürü borç bilirim.

Kekemeliğim sürecinde bana destek olan ve beni gelecek öğretim hayatına hazırlayan ilkokul öğretmenim hocam Zehra ERKAN'a, manevi desteklerini eğitim hayatım boyunca arkamda hissettiğim eşim sınıf öğretmeni Zeynep ÇİMEN'e, babam Mehmet ÇİMEN'e, annem Huriye ÇİMEN'e, kardeşim Meryem ÇİMEN KIVILCIM'a, yiğenim Doğa KIVILCIM'a, oğullarım Ahmet Arda ve Mehmet Mete'ye minnettarlığımı sunarım.

Aralık 2022

Ali ÇİMEN

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	viii
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ	xii
ÖZET	xiii
SUMMARY	xv
1 GİRİŞ.....	1
2 GENEL KISIMLAR.....	4
2.1 VERİ TABANLARINDAN BİLGİ KEŞFİ VE VERİ MADENCİLİĞİ	4
2.2 BİR VERİ OLARAK METİN	8
2.1.1 Metnin Kodlanması	9
2.3 METİN MADENCİLİĞİ NEDİR?	11
2.4 METİN MADENCİLİĞİ SÜRECİ.....	14
2.1.2 Çalışmanın Amacını Belirleme.....	15
2.1.3 Verinin Doğasını ve Erişilebilirliğini İnceleme	15
2.1.4 Veriyi Hazırlama	16
2.4.1.1 Veriyi Temizleme.....	16
2.4.1.2 Veriyi İndirgeme.....	21
2.4.1.3 Veriyi Dönüştürme	23
2.1.5 Modelleri Oluşturma	26
2.4.1.4 Gizli Anlamsal Analiz.....	27
2.4.1.5 Olasılıksal Gizli Anlamsal Analiz	28
2.4.1.6 Gizli Dirichlet Ataması	30
2.4.1.7 Denetimli ve Yarı Denetimli Eklentiler	31
2.4.1.8 Hiyerarşi Eklentileri.....	32
2.4.1.9 Zaman Serisi Eklentileri.....	33
2.4.1.10 Duygu Analizi.....	34
2.1.6 Sonuçların Değerlendirilmesi	35
2.1.7 Uygulama	39

2.5	TEMEL DUYGULAR VE DEĞERLER	39
2.1.8	<i>Temel Duygular</i>	39
2.1.9	<i>Kök Değerler</i>	41
2.6	5E EĞİTİM MODELİ	42
2.7	BİLGİ İŞLEMSEL DÜŞÜNME.....	44
2.8	HAREZMÎ EĞİTİM MODELİ.....	46
2.1.10	<i>Modelin Uygulama Süreci</i>	47
2.9	LİTERATÜR ANALİZİ.....	47
3	MALZEME VE YÖNTEM.....	54
3.1	KULLANILAN VE GELİŞTİRİLEN UYGULAMALAR.....	54
3.1.1	<i>Veri Yönetim Sistemi.....</i>	54
3.1.2	<i>Verinin İşlenmesi Süreci.....</i>	56
3.2	METİN MADENCİLİĞİ SÜRECİ.....	57
3.1.3	<i>Çalışmanın Amacı</i>	57
3.1.4	<i>Araştırmanın Kapsamı ve Sınırlıkları</i>	58
3.1.5	<i>Verinin Doğası ve Erişilebilirliği</i>	58
3.1.6	<i>Veriyi Hazırlama</i>	61
3.2.1.1	<i>Kaynaktan Veriyi Almak ve Ön İşlemden Geçirmek</i>	62
3.2.1.2	<i>Yazım Yanlışlarını Düzeltme</i>	62
3.2.1.3	<i>Normalleştirme.....</i>	64
3.2.1.4	<i>Terim Etiketleme</i>	65
3.2.1.5	<i>Morfolojik İndirgeme</i>	68
3.2.1.6	<i>Eşanlam İndirgemesi.....</i>	69
3.2.1.7	<i>Dolgu Kelimelerinin ve Belgelerinin Kaldırılması</i>	69
3.2.1.8	<i>Sürecin Genel Değerlendirmesi</i>	70
3.1.7	<i>Modelleme ve Değerlendirme.....</i>	73
3.2.1.9	<i>Konu Modelleme Algoritması</i>	74
3.2.1.10	<i>Değerlendirme Yöntemleri</i>	77
3.2.1.11	<i>Derlemin Oluşturulması.....</i>	81
3.2.1.12	<i>Modelleme Parametreleri</i>	83
3.2.1.13	<i>Konu Sayılarının Belirlenmesi</i>	83
3.2.1.14	<i>Konuların İsimlendirilmesi</i>	86
3.2.1.15	<i>Uzman Görüşü Alma Süreci.....</i>	87
4	BULGULAR.....	89

4.1	ARAŞTIRMA SORULARI.....	89
4.2	BİRİNCİ ARAŞTIRMA SORUSUNA YÖNELİK BULGULAR.....	91
4.3	İKİNCİ ARAŞTIRMA SORUSUNA YÖNELİK BULGULAR	92
4.4	ÜÇÜNCÜ ARAŞTIRMA SORUSUNA YÖNELİK BULGULAR.....	93
4.1.1	<i>S1 Derlemi</i>	93
4.1.2	<i>S2 Derlemi</i>	95
4.1.3	<i>S3 Derlemi</i>	97
4.5	DÖRDÜNCÜ ARAŞTIRMA SORUSUNA YÖNELİK BULGULAR.....	99
4.1.4	<i>S1 Derlemi</i>	99
4.1.5	<i>S2 Derlemi</i>	102
4.1.6	<i>S3 Derlemi</i>	109
5	TARTIŞMA VE SONUÇ	117
	KAYNAKLAR.....	121
	EKLER	131
	EK 1. S1 Sorusu Konu-Kelime Ağırlıkları Tablosu.	131
	EK 2. S2 Sorusu Konu Kelime Ağırlıkları Tablosu.....	137
	EK 3. S3 Sorusu Konu Kelime Ağırlıkları Tablosu.....	142
	EK 4. Uzman Görüşü Sorular Ve Örnek Cevaplar Tablosu.	146
	EK 5. Değerlendirme Formu.....	148
	EK 6. Araştırma İzni.	149
	ÖZGEÇMİŞ	150

ŞEKİL LİSTESİ

	Sayfa No
Şekil 2.1: Veri tabanlarından bilgi keşfi süreci.	5
Şekil 2.2: CRISP-DM modeli temel şeması [12].	6
Şekil 2.3: ASCII kodlama sistemi karakter tablosu [17].	10
Şekil 2.4: Karakter kodlama standartları kullanım oranları grafiği [22].	11
Şekil 2.5: Metin madenciliği süreç diyagramı [28].	14
Şekil 2.6: Örnek metin madenciliği süreç diyagramı [24].	14
Şekil 2.7: Metin madenciliği süreç modeli [27].	15
Şekil 2.8: NER kullanılarak yapılan bir parçalama işlemi [36].	18
Şekil 2.9: Normalizasyon işlemi örnek şeması [40].	20
Şekil 2.10: Vektör uzay ve LSA uzay gösterimleri [53].	28
Şekil 2.11: PLSA modelinin grafiksel gösterimi [56].	29
Şekil 2.12: LDA modelinin grafiksel gösterimi [59].	30
Şekil 2.13: LLDA algoritması grafiksel gösterimi [63].	32
Şekil 2.14: PA ile oluşturulmuş NIPS özet veri seti (NIPS00-12) konu hiyerarşisi [67].	32
Şekil 2.15: DTM'in grafiksel gösterimi (3 zaman parçası için) [68].	33
Şekil 2.16: NIPS veritabanı kullanılarak oluşturulan "Image Processing" konusu altındaki kelimelerin yıllara göre evrimi [69].	34
Şekil 2.17: Model değerlendirmesi çerçeve modeli [77].	36
Şekil 2.18: Duygunun oluşum süreci [82].	40
Şekil 2.19: Plutchik duygu çemberi [83].	41
Şekil 2.20: 5E eğitim modeli basamakları [87].	43
Şekil 2.21: Bilgi işlemsel düşünmenin alt boyutları [94].	45
Şekil 3.1: VYS form oluşturma ekranı örnek görüntüsü.	55

Şekil 3.2:	Normalizasyon araçlarının örnek kullanımı	64
Şekil 3.3:	Sözlük tabanlı terim etiketleme fonksiyonları.	66
Şekil 3.4:	NER araçları sorgu kodu ve cevapları.....	66
Şekil 3.5:	NER yardımıyla terim etiketleme Python fonksiyonları.	67
Şekil 3.6:	extract_ngrams fonksiyonuyla etiket adayları oluşturmak ve birleştirmek.	68
Şekil 3.7:	Morfolojik indirgeme fonksiyonunun ürettiği çıktılar.	68
Şekil 3.8:	Veritabanı sorgusu kod satırları ve alınan verinin şekli.	71
Şekil 3.9:	Süreç sonunda oluşan ön işleme süreci diyagramı.	72
Şekil 3.10:	Tomotopy kütüphanesi LDAModel sınıfı parametreleri.....	74
Şekil 3.11:	Tomotopy kütüphanesi Corpus sınıfı parametreleri.....	75
Şekil 3.12:	Tomotopy kütüphanesi train fonksiyonu ve parametreleri.	76
Şekil 3.13:	Tomotopy kütüphanesi Coherence sınıfı parametreleri	77
Şekil 3.14:	LDA modeli 2-200 konu aralığı, 40-2000 tekrar için başarımlı tutarlılık ölçümleri serpinti grafikleri.....	79
Şekil 3.15:	normalperplexity ve C_v değerleri ortak grafiği.	80
Şekil 3.16:	Derlem oluşturmak için kullanılan kod blokları.	81
Şekil 3.17:	Konu sayısı tespitine ait perplexity ve C_v değerleri.	84
Şekil 3.18:	Konu sayısı tespitine ait yakınlştırılmış perplexity ve C_v değerleri.	85
Şekil 3.19:	Konu modelleme iş akışı.	86
Şekil 3.20:	Konu-kelime ağırlığı görsel gösterimi.	87
Şekil 4.1:	S2 derlemi konu etiketlerinin periyotlar arasında dağılım grafikleri.	106
Şekil 4.2:	S2 derlemi konu etiketlerinin okul türleri arasında dağılım grafikleri.....	109
Şekil 4.3:	S3 derlemi konu etiketlerinin ders periyotları arasında değişimi gösterir grafikler.	112
Şekil 4.4:	S3 derlemi konu etiketlerinin okul dereceleri arasında değişimini gösterir grafikler.	114
Şekil 4.5:	S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri arasında dağılımlarına ait grafikler.....	116

TABLO LİSTESİ

Sayfa No

Tablo 2.1:	Yapılandırılmış, yarı yapılandırılmış ve yapılandırılmış metin [15].....	9
Tablo 2.2:	Örnek stemming ve lemmization işlemi girdi ve çıktıları.....	23
Tablo 2.3:	Örnek terim-belge matrisi.	24
Tablo 2.4:	Yerel ve genel terim ağırlıklandırma yöntemleri.	24
Tablo 2.5:	Kök değerler ve değerlerle ilişkili bazı tutum ve davranışlar [85].....	42
Tablo 3.1:	HEM 2018-2019 uygulamasına ait okul, öğrenci ve öğretmen sayıları.....	59
Tablo 3.2:	3N formu verilerinin uygulama haftası ve okul türüne göre dağılımı.....	60
Tablo 3.3:	Derlemede bulunan belge, kelime ve belge başına kelime ortalamasını gösterir tablo.....	61
Tablo 3.4:	Unicode dönüşümü ve ön temizlik işlemlerinde metin.	62
Tablo 3.5:	Metin ön işleme işlem basamakları,işlem öncesi ve sonrası verilere ait tablo.....	72
Tablo 3.6:	Terim adayları ve nPMI değerleri	73
Tablo 3.7:	Derlem bazlı karşılaştırma, temizleme alanları ve dolgu kelimeleri tablosu.	82
Tablo 3.8:	Derlemlerde bulunan belge sayısı, sözlük büyüklüğü ve kelime sayıları.	84
Tablo 4.1:	Derlemlere göre belge uzunlukları betimsel istatistikleri.....	91
Tablo 4.2:	Derlemlere göre belge uzunlukları ve okul grupları arasında Mann-Whitney-U testi sonuçları.....	91
Tablo 4.3:	Derlemlere göre belge uzunlukları ve okul grupları arasında Mann-Whitney U testi sonuçları.	92
Tablo 4.4:	S3 sorusu duygu kutbu analizi Ki-Kare testi sonuçları.	92
Tablo 4.5:	S1 derlemi LDA analizi özet bilgileri.	93
Tablo 4.6:	S1 derlemi konu-kelime dağılımları özet tablosu.	93
Tablo 4.7:	S1 derlemi konu-belge birikimi tablosu.	94

Tablo 4.8:	S1 derlemi 5E modeline göre konu etiketleri.....	95
Tablo 4.9:	S2 derlemi LDA analizi özet bilgileri.	95
Tablo 4.10:	S2 derlemi konu-kelime dağılımları özet tablosu.	96
Tablo 4.11:	S2 derlemi konu-belge birikimi tablosu.	96
Tablo 4.12:	S2 derlemi konularının öğretmenler tarafından isimlendirilmesi.	97
Tablo 4.13:	S3 derlemi LDA analizi özet bilgileri.	97
Tablo 4.14:	S3 derlemi konu-kelime dağılımları özet tablosu.	97
Tablo 4.15:	S3 derlemi konu-belge birikimi tablosu.	98
Tablo 4.16:	S3 derlemi duygu etiketleri.	99
Tablo 4.17:	S1 derlemi konu etiketleri frekans ve yüzde değerleri.....	99
Tablo 4.18:	S1 derlemi okul derecelerine göre ki-kare testi sonuçları.	100
Tablo 4.19:	S1 derlemi etiketleri ve ders periyotları ki-kare testi sonuçları.	101
Tablo 4.20:	S2 derlemi konu etiketleri frekans ve yüzde verileri.....	103
Tablo 4.21:	S2 derlemi konu etiketlerinin ders periyotları arasında dağılımı.	103
Tablo 4.22:	S2 derlemi konu etiketleri ki-kare testi sonuçları.....	105
Tablo 4.23:	S2 derlemi konu etiketlerinin okul dereceleri arasında dağılımı.....	107
Tablo 4.24:	S2 derlemi konu etiketleri ve okul dereceleri arasında ki-kare testi sonuçları.....	108
Tablo 4.25:	S3 derlemi konu etiketleri frekans ve yüzde verileri.....	110
Tablo 4.26:	S3 derlemi konu etiketleri ve ders periyotları arasında Ki-Kare testi sonuçları.....	111
Tablo 4.27:	S3 derlemi konu etiketlerinin okul dereceleri arasında ki-kare testi sonuçları.....	113
Tablo 4.28:	S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri arasında dağılımlarını gösterir ki-kare testi sonuçları.	115

SİMGE VE KISALTMA LİSTESİ

Simgeler Açıklama

w	: Kelime
d	: Belge
D	: Derlem
V	: Sözlük büyüklüğü
W	: Kelime kümesi
β	: Konu
θ	: Her konunun belge üzerindeki ağırlığı
z	: Her belge için konu ataması
α	: Dirichlet Alpha parametresi
μ	: Dirichlet Muon parametresi
k	: Konu Sayısı
s	: Saniye
C	: Başarım
f	: Frekans
ef:	: Beklenen frekans
%	: Yüzde
Sd	: Standart sapma

Kısaltmalar Açıklama

TDDİ	: Türkçe Doğal Dil İşleme
A12	: Anaokulu ve 12 yıllık zorunlu eğitim süreci
S1	: Soru 1: “Ne yaptın?”
S2	: Soru 2: “Ne öğrendin?”
S3	: Soru 3: “Ne hissettin?”
CRISP-DM	: Cross Industry Standard Process for Data Mining (Veri Madenciliği için Sektörler Arası Standart Süreç)
HEM	: Harezmi Eğitim Modeli
IDF	: Inverse Documant Frequency (Tersine Belge Frekansı)
TF	: Term Frequency – (Terim Frekansı)
PMI	: Pointwise Mutual Information (Noktasal Birliktelik Bilgisi)

ÖZET

DOKTORA TEZİ

HAREZMİ EĞİTİM MODELİ UYGULAMA VERİSİNİN METİN MADENCİLİĞİ TEKNİKLERİYLE DEĞERLENDİRİLMESİ

Ali ÇİMEN

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman: Prof. Dr. Sevinç GÜLSEÇEN

Harezmi Eğitim Modeli (HEM) 2019-2020 öğretim yılında İstanbul genelinde 411 okulda 4195 öğrencinin katılımıyla, 32 haftalık bir süreç boyunca uygulanmıştır. Öğrencilerin ders süreçlerini değerlendirmeleri amacıyla deneyimlerine yönelik farkındalıklarını açık uçlu sorulardan oluşan bir forma yansıtılmaları istenmiş, çalışma kapsamında ise bu verilerin anlamlandırılması amaçlanmıştır. Bu amaç doğrultusunda Gizli Dirichlet Ataması (Latent Dirichlet Allocation – LDA) yöntemi kullanılarak veri içerisindeki gizli desenler ortaya çıkarılmıştır. Ders ortamından alınan “Ne yaptın?” (S1), “Ne öğrendin?” (S2) ve “Ne hissettin?” (S3) başlıklarına sahip üç derleme ait konu sayıları Perplexity (şaşıklık) ve Röder vd. (2015) tarafından Cv olarak isimlendirilen başarımlarına göre tespit edilmiş ve veri modellenmiştir. S1 derleminde bulunan 63938 belge için model 54 konu sayısında 2000 tekrarda eğitilmiş ve 0.73 tutarlılık değerine ulaşılmıştır. S2 derleminde bulunan 69482 belge için ise model 47 konu sayısında 2000 tekrarla eğitildiğinde 0.79 tutarlılık değerine ulaşıldığı görülmüştür. S3 derleminde bulunan 36465 belge için model 54 konu sayısında 2000 tekrarda eğitilmiş ve 0.70 tutarlılık değerine ulaşılmıştır.

Yorum uzunluğuna göre yapılan analizlerde tüm derlemler için en uzun yorumları ortaokul öğrencilerinin yazdıkları, liselilerin sonraki en uzun yorumları yazdıkları ve ilkökul öğrencilerinin en kısa yorumları yazdıkları görülmüştür. S3 derlemi üzerinde yapılan sözlük tabanlı duygu kutbu analizlerinde ise ilkökul öğrencilerinin yorumlarının en yüksek seviyede

olumlu olduđu, bu durumun liseye dođru gidildikçe tersine döndüđu görölmüşür. S1 konu etiketleri incelendiğinde 5E öğretim modeli basamakları (dikkat çekme, keşfetme, açıklama, derinleştirme, değerlendirme) ve o basamaklara ek olarak “bilgi işlemsel düşünme (computational thinking)” basamağının da alan uzmanları tarafından etiketlendiğı görölmüşür. S2 sorusu konu etiketleri incelendiğinde derslerin genel olarak çevre, bilgi işlemsel düşünme ve problem çözme konularına odaklandığı değerlendirilebilir. S3 derlemi bulguları değerlendirildiğinde ise öğrencilerin temel duyguların dışında temel değerleri de duygu olarak yansıttıkları değerlendirilmiştir.

Tez kapsamında metin madenciliğı ön işlem ve modelleme metotları derinlemesine analiz edilmiş ve bir yol haritası ortaya koyulmaya çalışılmıştır. Ortaya çıkarılan veri setiyle gerçekleştirilen çalışma sonucunda gerek konu modelleme gerekse sonrasında yapılan keşifsel analizlerde ortaya çıkan bulgular öğrencilerin bakış açısıyla derste ne işlendiğini ve öğrencilerin duygularını ortaya koymakta olduğundan ortaya çıkarılan gizli örüntülerin bu alanda yapılacak gelecek çalışmalara temel olacağı düşünülmektedir.

Aralık 2022, 162 sayfa.

Anahtar kelimeler: Harezmi eğitim modeli, 5E modeli, öğrenme yaşantıları, konu modelleme, duygu madenciliğı.

SUMMARY

Ph.D. THESIS

EVALUATION OF HAREZMİ EDUCATION MODEL APPLICATION DATA WITH TEXT MINING TECHNIQUES

Ali ÇİMEN

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor: Prof. Dr. Sevinç GÜLSEÇEN

Lessons designed around a new education model called Harezmi Education Model (HEM) were implemented in 411 schools in Istanbul, involving 4195 students, for a period of 32 weeks during 2019-2020 academic year. Students were asked to reflect upon their experiences they had in these lessons on a form consisting of open-ended questions and the present study aims to examine and analyze the data in which students reflected their experiences of HEM practices. The hidden patterns acquired from data were revealed through Latent Dirichlet Allocation (LDA) method. The number of topics consisting of three corpuses determined around “What did you do? (Q1)”, “What did you learn?” (Q2) and “How did you feel? (Q3)” questions were extracted with the Perplexity and Cv, a coherence score announced by Röder et al. (2015). The corpus of Q1 which has 63,938 documents was modeled for 54 subjects with 2,000 iterations and reached a coherence value of 0.73. The corpus of Q2 which has 69,482 documents was modeled for 47 subjects with 2,000 iterations and reached a coherence value of 0.79. The corpus of Q3 which has 36,465 documents was modelled for 54 subjects with 2,000 iterations and reached a coherence value of 0.70.

Analyzing the length of the reflections of the students, it is revealed that the longest reflections were made by secondary school students and the second longest reflections were made by high

school students. The shortest comments, on the other hand, were made by primary school students. Based on dictionary-based sentiment polarity analysis on Q3 corpus, primary school students are the happiest group followed by middle and high school students. In addition, analyzing Q1 topic labels, it is acknowledged that 5E Instructional model phases (engagement, exploration, explanation, elaboration and evaluation) as well as Computational Thinking Phase were labeled by experts. Furthermore, analyzing Q2 topic labels, it is ascertained that the main focus points of the lessons were environment, computational thinking and problem solving. Finally, analyzing the findings of Q3, it is discovered that students reflected upon their feelings as well as social norms.

Within the scope of the current study, text mining and modeling methods were analyzed comprehensively in order to create a road map. By the revealed dataset which is the product of the study carried out and the patterns which are the reflection of students' perspective it can be conceivable that the study could be a basis for future studies in this field.

December 2022, 162 pages.

Keywords: Harezmi education model, 5E model, learning experiences, topic modeling, emotion mining.

1 GİRİŞ

Her alanda olduğu gibi eğitim alanında da devasa sayılabilecek ölçüde veri üretilmektedir. Bu veriler yapılandırılmış, yarı yapılandırılmış, ya da yapılandırılmamış formatlarda olabilmektedirler. Doğal dil şeklinde olan veriler, yapılandırılmamış formatta olan verilerin büyük bir kısmını oluşturmaktadır. Her ne kadar yapılandırılmış formatta olan verilerin işlenmesi nispeten kolay olsa da metin formatında olan verilerin işlenmesi zorlu bir süreç olabilmektedir.

Veri analizi, nitel araştırmacıların en fazla zorluk çektikleri aşamadır. Nitel araştırmalarda kısa olan yazılı materyallerde sözcüklere dayalı bu tür analiz elle yapılabilir, ancak uzun ve karmaşık yazılı materyallerde bu tür bir analizi yapmak, mutlaka bilgisayarlardan yararlanmayı gerektirecektir [1]. Bilgisayarlar yardımıyla metinlerin analiz edilmesinde doğal dil işleme alanı devreye girmektedir. Türkçe doğal dil işleme (TDDİ) 60M konuşanı olan Türkçe için oldukça geç kalınmıştır ve alanda yeterince araştırma bulunmamaktadır ve kullanılabilir kütüphaneler yetersizdir [2,3]. Ayrıca Türkçe dilinin kendine has morfolojik yapısı da doğal dil işleme alanında Türkçe'nin zorlukları arasında yer almaktadır. Sağlam [3]'ün aktardığına göre farklı diller üzerinde çalışan araştırmacıların dillerindeki metinleri İngilizce diline çevirdiği ve İngilizce üzerinde araştırmalarına devam ettikleri görülmektedir ancak çeviri sistematikinden kaynaklı olarak anlam kaymaları olacaktır.

Metinler üzerinde çalışmanın en büyük zorluğu metin ön işleme sürecidir. Araştırmayı temel olarak ön işleme süreci yönlendirmektedir. Yazım yanlışlarını düzeltmek, kök bulmak, kendi başına anlam ifade etmeyen kelimeleri/kavramları temizlemek vb. işlemler metne uygulandıktan sonra metin üzerinde nitel araştırma yöntemlerinde veri analizi çalışmaları yapılabilecek hale getirilmesi gerekmektedir. Metinler içlerinde ihtiva ettikleri kavramlara göre kategoriler (temalar) altında sınıflandırılabilir. Yıldırım ve Şimşek'e [1] göre bu süreç (1) verilerin kodlanması, (2) temaların bulunması, (3) kodların ve temaların düzenlenmesi, (4) bulguların tanımlanması ve yorumlanması şeklinde dört aşamada ele alınmıştır.

Metin üzerine; hangi alandan, nasıl ve hangi süreçle alındığı, veri türü (yapılandırılmamış, yarı yapılandırılmış), işin hedefi vb. özellikleri göz önünde bulundurularak kendine has bir süreç içinde madencilik işlemleri uygulanabilir. A12 eğitim ortamından ortaya çıkarılan öğrenci

görüşleri ile ilgili temalar eğitim alanı için oldukça önemlidir fakat bugüne kadar eğitim alanından alınan Türkçe metin verileri üzerinde konu modelleme teknikleri yardımıyla kategori ve temalar ortaya çıkarılmamıştır. Bu durum da tez kapsamında ortaya koyulan ana problemidir.

Bu sebeple tez kapsamında Harezmi Eğitim Modeli (HEM) uygulama ortamında öğrencilerin ders süreçlerini değerlendirdikleri verilerin metin madenciliği teknikleri kullanılarak anlamlandırılması amaçlanmaktadır. Bu bağlamda tez kapsamında aynı dersi işleyen öğrencilerin derste ne yaptıkları, ne öğrendikleri, hangi duyguyu hissettikleri ve bu durumların zamansal değişimleri incelenecektir. Ayrıca tez kapsamında metin verisinin madenciliği süreci boyunca yolculuğu adım adım incelenerek bir yol haritası ortaya konulmaya çalışılacaktır.

Çalışma “GENEL KISIMLAR”, “MALZEME VE YÖNTEM”, “BULGULAR”, “TARTIŞMA VE SONUÇ” olmak üzere dört ana bölümden oluşmaktadır. Birinci bölümde öncelikle veri madenciliği konusu üzerinde durulmuş ve metin madenciliği süreci ayrıntılarıyla işlenmiştir. Metin madenciliği süreci aşamaları olan; çalışmanın amacını belirleme, verinin doğasını ve erişilebilirliğini inceleme, veriyi modelleme sürecine hazırlama, modelleme, sonuçları değerlendirme ve uygulama aşamaları ayrıntılarıyla işlenmiştir. Duygu analizine temel olan duyguların oluşum süreci ve temel insani duygulardan bahsedilen bölümde ayrıca Millî Eğitim Bakanlığı tarafından müfredatta ayrıntılarıyla yer verilen kök değerlere, 5E Eğitim Modeline, Bilgi İşlemsel Düşünmeye, HEM’in genel yapısı ve uygulama sürecine yer verilmiştir. Bölüm, alanda bugüne kadar yapılan çalışmalara kategorize edilmiş şekilde yer verilerek tamamlanmıştır.

Tez çalışmasının “MALZEME VE YÖNTEM” bölümünde tez kapsamında kullanılan ve geliştirilen uygulamalara yer verilmiş ve verinin işlenmesi süreci ayrıntılarıyla anlatılmıştır. Verinin elde edilme süreci ve veriyi modellemeye hazır hale getirme (ön işlem) aşamalarına ayrıntılı olarak yer verilmiş, süreç karşılaştırmalı olarak ele alınarak bir yol haritası ortaya koyulmaya çalışılmıştır. Verinin modellenmesi aşamasında önemli yeri olan parametre seçimi, konu sayısı tespiti ve uzman görüşleri yardımıyla ortaya çıkan gizli örüntülerin anlamlandırılması aşamaları da bu bölümde bahsedilen diğer konu başlıklarıdır.

Tezin “BULGULAR” bölümünde, modelleme aşamasından sonra ortaya çıkarılan ve gizli örüntülerle daha zengin bir hale gelen veri seti üzerinden oluşturulan araştırma sorularına cevap aranmış ve ortaya çıkan bulgular paylaşılmıştır. Veri seti öncelikle yorum uzunluğuna, duygu kutuplarına göre analiz edilmiş, sonrasında ise konu analizinde ortaya çıkan örüntülere yer

verilmiştir. Keşifsel analiz kapsamında ise verilerin birbirleriyle ilişkileri araştırılarak ortaya çıkan gizli örüntüler paylaşılmıştır.

“TARTIŞMA VE SONUÇ” bölümünde ise elde edilen bulgular teorik çerçeveye karşılaştırılmış, ortaya çıkarılan örüntüler anlamlandırılmaya çalışılmıştır. Ortaya çıkarılan veri setiyle gerçekleştirilen çalışmanın sonuçlarına ve alana katkısına yer verilmiş, alanda bundan sonra yapılacak çalışmalara öneriler sunulmaya çalışılmıştır. Ortaya çıkarılan gizli örüntülerin bu alanda yapılacak gelecek çalışmalara temel olacağı düşünülmektedir.



2 GENEL KISIMLAR

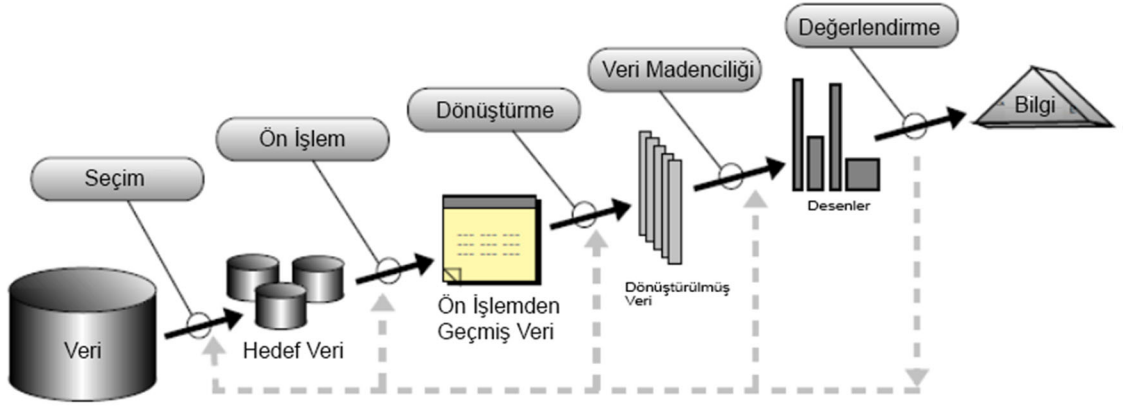
2.1 Veri Tabanlarından Bilgi Keşfi ve Veri Madenciliği

Coffman ve Odlyzko [4] Moore yasasına atıf yaparak yaptıkları araştırmada Internet trafiğinin her yıl ikiye katlandığını değerlendirmişlerdir. 2001-2021 yılları arasındaki zaman diliminde Internet trafiğindeki gelişim değerlendirildiğinde bu değerlendirmenin doğru olduğu kabul edilebilir. Aynı zamanda bilgisayarların, mobil cihazların ve bulut teknolojilerinin veri tutma kapasitelerinin bugün için terabaytlar seviyesine ulaşması, ayrıca kişisel kullanımda bile 1Gbps¹ indirme hızlarında Internet bağlantılarına ev kullanıcılarının bile sahip olabildiği düşünüldüğünde verinin üretim ve yayım hızıyla birlikte depolama miktarlarının da geçmişe nazaran oldukça üst seviyelerde olduğu değerlendirilebilir. Esas olarak bilgi çağında geleneksel veri işleme yöntemleri bu bilgi seliyle başa çıkmak için yeterli değildir [5].

Ekonomistler, istatistikçiler, tahminciler, iletişim mühendisleri vb. uzunca bir zaman veriler içindeki gizli özelliklerin otomatik olarak tanımlanacağı, doğrulanacağı ve tahmin için kullanılacağı yollar aramışlardır [6]. Örnek olarak mobil cihazınızda bulunan konum verileri anlamlandırılmadan önce sadece bir yığın halinde durmaktadır fakat bu veriler kişilerin hangi lokantalarda bulunduğu şeklinde değerlendirildiğinde yeme/içme alışkanlıklarınız tahmin edilebilir ve bulunduğunuz yere göre yeme/içme tavsiyelerinde bulunulabilir. Bu durumda **veri tabanlarından bilgi keşfi** süreci; depolanmış veri içerisindeki anlamlı örüntüleri ortaya çıkarma süreci olarak tanımlanabilir [7].

Fayyad vd. [8]'e göre veri tabanlarından bilgi keşfi süreci; verinin seçimi, ön işleme, indirgeme, veri madenciliği ve değerlendirme olmak üzere beş adımdan oluşan bir süreçtir. Veri tabanlarından veri keşfi süreci diyagramı Şekil 2.1'de gösterilmiştir.

¹ Turknet, 2022, *Fiber Internet kampanyaları*, <https://turk.net/fiber-internet>, [Ziyaret Tarihi: 15 Mayıs 2022]

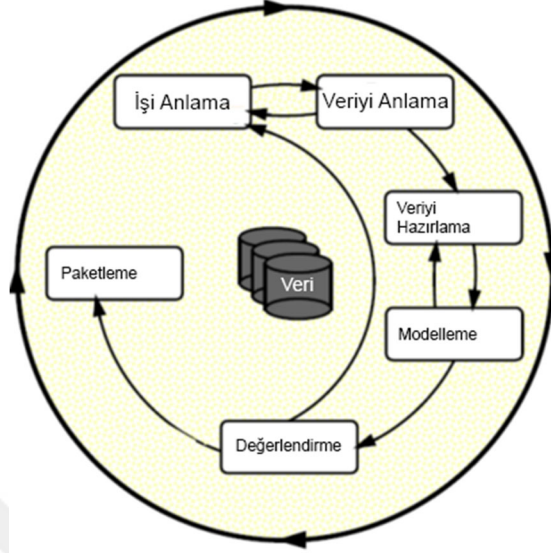


Şekil 2.1: Veri tabanlarından bilgi keşfi süreci.

Şekil 2.1’de görüldüğü üzere veri madenciliği süreci, veri tabanlarından bilgi keşfi süreci içinde bir adımdır. İşlenmemiş veri üzerinde birçok işlem yapılarak örüntüler ortaya çıkarılabilir. Ortalama, toplam, veri süzme vb. matematiksel işlemler yaparak ortaya çıkan verileri özetlemek, grafiklere dönüştürmek gibi istatistiksel işlemler veriyi anlamlandırmak için kullanılacak sayısız metottan bir kaçıdır. Temel istatistiksel yöntemlerin hedefimize hizmet edemediği durumlarda veriyi anlamlandırmak için farklı yollar aramak gerekmektedir. İşte bu noktada **veri madenciliği** kavramı veri tabanlarından bilgi keşfi sürecinin veri analizi ve keşfi algoritmaları yardımıyla anlamlı örüntüler ortaya koyulan bir aşaması olarak karşımıza çıkmaktadır [8]. Veri madenciliğinin geleneksel istatistiksel yöntemlerden ayrıldığı nokta, geleneksel istatistiksel yöntemlerin varsayıma dayalı olması ve hipotezlerini veriler üzerinde test etmesi, veri madenciliğinin ise modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı yaklaşımlar sunmasıdır [9]. Veri madenciliği problemlere çözüm yolları ararken istatistik, makine öğrenmesi, görselleştirme, veri bilimi ve diğer farklı disiplinlerin alandaki problem çözme metodlarının bir bütünlük içinde uygulanmasıyla gerçekleşen disiplinler arası bir süreçtir [10].

Veri madenciliği işlemlerini sistematik ve standartlara bağlı olarak yürütmek amacıyla genellikle standart modeller takip edilmektedir. Bu standart modellerden en bilineni ve endüstri tarafından kullanılan *Veri Madenciliği için Sektörler Arası Standart Süreç* (Cross Industry Standard Process for Data Mining - CRISP-DM) modelidir [11]. CRISP-DM modeli işi anlama,

veriyi anlama, veriyi hazırlama, modelleme, değerlendirme ve uygulama olmak üzere 6 iç içe geçmiş adımdan meydana gelmektedir (Şekil 2.2).



Şekil 2.2: CRISP-DM modeli temel şeması [12].

Şekil 2.2’de görüldüğü üzere CRISP-DM modeli veriyi merkeze alan ve birbiriyle ilişkili 6 basamaktan oluşan sarmal bir modeldir. Wirth ve Hipp [12] modelin aşamalarını şöyle açıklamışlardır.

İş Anlama

CRISP-DM modelinin ilk aşaması olan “İş Anlama” projenin hedeflerini, gereksinimleri iş perspektifinden anlamak için uğraş verildiği ve hedeflere ulaşmak için bir ön proje planının oluşturulduğu adımdır. Bu aşamada iş hedefleri belirlenir, durum değerlendirmesi yapılır, veri madenciliği hedefleri belirlenir ve proje planı oluşturulur.

Veriyi Anlama

Verileri anlama aşaması, veri toplama ile başlar ve veriyi tanımak, veri kalitesi sorunlarını belirlemek faaliyetleriyle başlar. Verilerle ilgili öngörü oluşturmak veya gizli örüntüleri tanımlamak amacıyla hipotezler oluşturmak için veriyi taramak gibi faaliyetlerle devam eder. Problemin ortaya koyulması ve proje planının oluşturulması için verinin yapısını anlamak gerekmektedir. Bu aşamada ön veriler toplanır, tanımlanır, keşfetmek için analizler yapılır ve veri kalitesi doğrulanır.

Veriyi Hazırlama

Veriyi hazırlama aşaması ham veriden modelleme araçlarında kullanılacak son veri setinin oluşturulmasına kadar geçen tüm aktiviteleri kapsayan aşamadır.

Bu aşamada;

- Tablo, kayıt ve özellik seçimi,
- Verinin temizlenmesi,
- Yeni özelliklerin oluşturulması,
- Verinin kullanılacak modelleme araçları için dönüştürülmesi gibi örnek aktiviteler takip edilebilir.

Fakat veriyi hazırlama aşamasındaki işlemler tekrar tekrar ve önceden öngörülen bir sıralama olmadan uygulanır. Veriyi modellemeye hazırlamak amacıyla oluşturulan bu yeni veri seti modelleme aşamasında kullanılır ve görülen eksikliklerin giderilmesi amacıyla tekrar veriyi hazırlama aşamasına geçilebilir.

Modelleme

Modelleme aşamasında farklı algoritmalar seçilerek veri üzerinde uygulanır ve parametre optimizasyonu yapılır. Veriyi hazırlama ve modelleme aşaması arasında sıkı bir bağlantının olmasının sebebi veri üzerindeki sorunların modelleme aşamasında fark edilmesi ve verinin tekrar hazırlanarak yeni bir veri kümesi oluşturulması için fikir edinilmesidir.

Bu aşamada modelleme tekniğinin belirlenmesi, başarımın nasıl test edileceğinin belirlenmesi, modelin oluşturulması ve değerlendirilmesi işlemleri yapılır.

Değerlendirme

Modelleme aşamasında aynı ya da farklı algoritmalar kullanılarak birden fazla model oluşturulabilir. Oluşturulan modellerin iş hedeflerine ne kadar ulaşıldığı ve ne kadar başarılı oldukları sorularına cevap aranır. Bu aşamada; oluşturulan modellerin tatmin edici düzeyde başarımlar elde ettikleri/etmedikleri sorularına verilecek cevapla bir sonraki aşama olan uygulama aşamasına geçilmesi ya da yeniden önceki basamaklara dönülmesi seçenekleri arasında seçim yapılması gerekmektedir.

Bu aşamada sonuçların değerlendirilmesi, sürecin incelenmesi ve sonraki süreçte hangi işlemlerin yapılacağı değerlendirilir. Değerlendirme aşaması; başarıyı sağlayan en iyi modelle uygulama aşamasına geçeceğinden veri madenciliği çalışmalarının amacına ulaşp/ulaşmadığı konusuna karar verilen kritik aşamadır. Veri madenciliği modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı bir yaklaşım olduğundan bu aşama yeni hipotezler kurularak tekrar işi anlama basamağına dönüşebilen kritik bir basamaktır.

Uygulama

Bir model oluşturma, projenin bittiği anlamına gelmeyeceği için oluşturulan modelden nasıl bir fayda üretileceği projenin her aşamasında düşünülmesi gerekli bir olgudur. Proje sonucunda bulunan bilginin müşteriyle paylaşılması ya da oluşturulan modelin bir sisteme entegre edilmesi gerekmektedir. Bir veri madenciliği projesinin ürünleri olan raporlar ve sistem entegrasyonları bu aşamada üretilerek paylaşılırlar.

Bu aşama projenin son aşaması olduğundan dağıtım planlanır, izleme ve bakım planı oluşturulur, final raporu üretilir ve proje genel olarak değerlendirilir.

2.2 Bir Veri Olarak Metin

Yeni teknolojiler, insan etkileşimi, iletişim ve kültürün giderek artan bir payını kaydeden büyük miktarlarda dijital metni kullanıma sunmaktadır [13]. İnternet ortamı yüksek oranda metinden meydana gelmektedir. İnternet sayfalarını gezmek için kullanılan protokol olan Köprü Metni Aktarım Protokolü (Hyper Text Transfer Protocol - HTTP) kavramı metinlerin transfer edilmesinden türetilmiştir. Dünya üzerinde verilerin 85% kadarı metin formatındadır [14]. Metinler yapılandırılmış, yarı yapılandırılmış ya da yapılandırılmamış formatta bulunabilirler. Bu durumda metinlerin hangi durumlarda yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış formatta olduklarının daha iyi anlaşılabilmesi için hangi formatlarda nasıl anlaşılabilceği, saklanabildiği ve karakteristik özelliklerinin neler olduğu hususu Tablo 2.1’de paylaşılmıştır.

Tablo 2.1: Yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış metin [15].

	Yapılandırılmış	Yarı Yapılandırılmış	Yapılandırılmamış
Nedir?	Yüksek derecede organize edilmiş ve veritabanı ve veri tabloları şeklinde saklanan metin.	İçinde belli bir organizasyona sahip fakat standart bir yapı oluşturulmamış metin.	İnsanlar tarafından yazıldığı halde tutulan metin.
Hangi dosya türlerinde saklanır?	Hesap Tablosu (XLS, XLSX, ODS vb.) CSV Veri Tabanı (MySQL, Mariadb, Oracle, MSSQL vb.) JSON XML	HTML JSON XML	Metin Dosyaları (TXT) Yazım programı dosyaları (ODT, DOC, DOCX vb.) Ses Dosyaları Video Dosyaları Resim dosyaları
Karakteristik özellikleri nelerdir?	Veri ilişkisel veri tabanı ya da hesap tablosu şeklinde organize edilmiştir. Veriler tablolar şeklinde organize edilmiştir, her alan belli bir formatta, uzunluktadır ve hep sırayı takip ederler. Okunabilir ve üzerinde ön işlem gerekmeden işlem yapılabilir.	Veriler dosyalar halinde saklanır ve az da olsa organize edilmiştir. Veriler etiketlenmiştir ya da karakterler yardımıyla birbirinden ayrılmıştır fakat uzunluk, farklılık, sıralama gibi faktörlerle birbirinden ayrılmamıştır. Metnin durumuna göre ön işlem gerektirebilir.	Veri herhangi bir formatta olabilir. Dosyalar içinde herhangi bir yapı ayrımı yoktur. Bilgisayarlar tarafından analiz edilmeden önce genellikle büyük bir ön işlem gerektirir. Genellikle insanlar tarafından okunarak, dinlenerek ve izlenerek anlaşılabilir.

Tablo 2.1’de işlendiği gibi metin verileri toplanma, işleme ve saklanma yöntemlerine göre üç farklı formatta da bulunabilirler. Yapılandırılmış metinler, halen metin formatında saklansalar da metnin hedeflediği niteliğe göre (“insan ismi”, ”adres”, ”şehir” vb.) veri tabanı normalizasyonu yapılabilecek (nitelikler sayısal olarak ifade edilebilecek) durumda bulunabilirler. Yarı yapılandırılmış metinler bağlama göre HTML, XML, JSON gibi metnin tabanlı iletişim formatlarında saklanan metinleri ifade etmektedir ve belirli (“başlık”, ”özet”, ”içerik” vb.) başlıklara ayrılmışlardır. Yapılandırılmamış metinler ise şu anda okuduğunuz tez metni, bir kitap, alınmış bir not gibi sizin okuyarak anlayacağınız formatlarda bulunabilir.

2.1.1 Metnin Kodlanması

Bilgisayar verileri sayısal olarak sakladığı ve işlediği için metin verisini de 0 ve 1’lerden oluşan elektrik sinyalleri şeklinde saklamaktadır. Bu yüzyılın en eski kodlama sistemi olan telgraftan başlayıp bir standart haline gelen MORS alfabesi, devamında verilerin delikli kartlarla bilgisayarlara aktarılması sırasında kullanılan bir çok farklı standarttan birisi olan EBCDIC (Extended Binary-Coded Decimal Interchange Code) [16] kodlama standardı ve sonrasında ise 8 bitlik bir karakter kodlaması standardı olan ve hatta Internet dünyasının üzerine inşa edildiği “American Standard Code for Information Interchange – ASCII” kodlama standardından oluşan geniş bir yelpazede metinler dijital ortamda saklanmış ya da dijital evrende haberleşme

için kullanılmışlardır. ASCII kodlama standardı karakter tablosunda karakterlerin desimal ve hegzadesimal karşılıkları Şekil 2.3'te paylaşılmıştır.

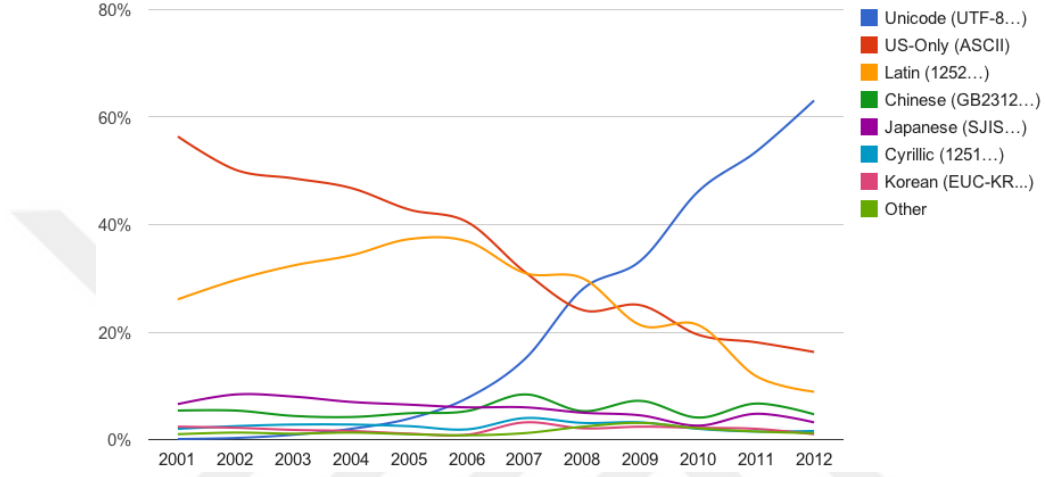
Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	@	96	60	`
1	1	[START OF HEADING]	33	21	!	65	41	A	97	61	a
2	2	[START OF TEXT]	34	22	"	66	42	B	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	C	99	63	c
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	e
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	'	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(	72	48	H	104	68	h
9	9	[HORIZONTAL TAB]	41	29	)	73	49	I	105	69	i
10	A	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	B	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	C	[FORM FEED]	44	2C	,	76	4C	L	108	6C	l
13	D	[CARRIAGE RETURN]	45	2D	-	77	4D	M	109	6D	m
14	E	[SHIFT OUT]	46	2E	.	78	4E	N	110	6E	n
15	F	[SHIFT IN]	47	2F	/	79	4F	O	111	6F	o
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	P	112	70	p
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	s
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	T	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	X	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Y	121	79	y
26	1A	[SUBSTITUTE]	58	3A	:	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[	123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	\	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	]	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	3F	?	95	5F	_	127	7F	[DEL]

Şekil 2.3: ASCII kodlama sistemi karakter tablosu [17].

ASCII kodlama sisteminde 33 adet basılamayan kontrol karakteri ve 95 adet de basılabilir karakter bulunmaktadır. Doğal dilin bilgisayar ortamına aktarılması ve anlamlandırılması işlemlerindeki en büyük sorun Dünya üzerinde 7139 dilin konuşulması ve bu dillerin 153 dil ailesinde toplanmış olmasından kaynaklanmaktadır [18]. Ortalama bir hesapla Dünya'nın 36% kadarı Latin alfabesini ve türevlerini, %18 Çin alfabesini, 14% kadarı Hint alfabesi türevlerini (Devanagari), 4% kadarı Kiril alfabesini ve 3.5% Güney Afrika (Dravidian) kullanmaktadır [19]. Bilgisayarlar ve Internet Latin alfabesinin standart bir türünün kullanıldığı Amerika Birleşik Devletleri'nde doğmuş olduğundan yazdırılabilir ASCII karakterlerini temel almaktadır. Internet'in Dünya'ya yaygınlaşması sırasında ise standart Latin alfabesinin bu dilleri dijital ortama aktarmaya yetmemesi sorunu ortaya çıkmıştır.

Türkçe söz konusu olduğunda standart Latin alfabesine eklenmiş olan “çÇşŞöÖğĞıİüÜ” harfleri temel Latin standardında bulunmazken, ISO8859-9 ve Unicode standartlarında bulunmaktadır. 1989 yılında İzlandaca karakterlerin bulunduğu ISO 8859-3 karakter setinden memnun olmayan Türkiye, Türkçe karakterler için yeni bir karakter seti oluşturulmasını istemiş ve ISO 8859-9 yada Latin 5 olarak bilinen karakter seti Türkçe için özel olarak oluşturulmuştur [20].

Unicode karakter kodlama sistemi şu anda ASCII'nin 95 adet basılabilir karakteri dahil olmak üzere 143.759 karaktere ev sahipliği yapmaktadır ve tek bir kodlama sistemi kullanılarak Dünya'da bulunan alfabelere ait karakterler kapsanabilmektedir [21]. ASCII karakter sisteminin 7 bitlik karakter kodlama yapısı 128 karakterin kodlanmasına izin verirken, Unicode karakter sistemi standardı 1.114.112 karakter pozisyonuna izin verdiğinden şu anda en çok tercih edilen karakter standardıdır (Şekil 2.4).



Şekil 2.4: Karakter kodlama standartları kullanım oranları grafiği[22].

Unicode standardı birçok farklı karakter seti yerine tek bir karakter setiyle Dünya üzerindeki dillerin tamamını bilgisayar ortamında saklayabilme ve paylaşabilme imkânı sağlar. Şekil 2.4 incelendiğinde UTF mimarisinin Dünya üzerindeki tüm alfabeleri kapsayabilecek yapıda tasarlanması ve bünyesine devamlı olarak yeni karakter setleri eklemesi sebebiyle ASCII ve Latin standartları sonrasında metin saklanması ve paylaşılması işlemlerinde en çok kullanılan mimari olmasının kaçınılmaz bir sonuç olduğu söylenebilir.

2.3 Metin Madenciliği Nedir?

Metnin kodlanması başlığı altında anlatıldığı gibi metinler bilgisayarlar için sayısal bir ifadeden farksızdır ve sayılarla ifade edilmektedirler. Fakat insanlar bir metni beyinlerinde oluşturdukları daha karmaşık kodlama sistemleriyle okuyarak karakterleri, kelimeleri, cümleleri, paragrafları ve onlar arasındaki karmaşık etkileşimi kavrayabilirler. Bir metni bilgisayarın anlamlandırması düşünüldüğünde ise benzer şekilde metnin anlamlı verilere dönüştürülmesi için karmaşık dilbilimsel yapıların ve kelimeler arasındaki etkileşimin hesaba katılması gerekmektedir [13]. Metinler herhangi bir dosyada, ses kayıtlarında, videolarda, veri tabanında, e-postalarda

kısacası karakterlerle sembolize edilmiş ya da halen ses, resim ya da video içerisinde olabilir. İnternet'in hızlı büyümesiyle dijital kütüphaneler, e-devlet ve sağlık uygulamaları gibi sistemler nedeniyle giderek artan miktarda metin ve multimedya verisi çevrimiçi olarak erişilebilir duruma gelmiş, birikmiş ve bu verilerin etkili bir şekilde aranması ve analiz edilmesi konusu ciddi zorlu sorunlar ortaya çıkarmıştır [23]. Bu noktada metin madenciliği ve multimedya veri madenciliği gibi alanların önemi gün geçtikte artmaktadır.

Metin madenciliği ifadesini ilk olarak tanımlayan Feldman ve Dagan'a göre [5] metin madenciliği ya da metinden bilgi çıkarma süreci metin içinden anlamlı bilgileri ortaya çıkarma süreci olarak tanımlanmaktadır. Veri tabanları, XML dosyaları gibi yapılandırılmış veriyle çalışılan veri madenciliği çalışmalarının aksine, metin madenciliği çalışmalarında yapılandırılmamış ya da yarı yapılandırılmış veri setleriyle (epostalar, metin dosyaları, HTML dosyaları gibi) çalışılmaktadır [24]. Metin madenciliği; enformasyona erişiminin daha ötesine geçerek, kullanıcıların analiz, bilgi çıkarımı çalışmalarına yardımcı olmak ve karar verme süreçlerini kolaylaştırmak olarak da tanımlanabilir [23]. En genel ifadeyle metin madenciliği, metin içinden anlamlı örüntüleri ortaya çıkarma süreci olarak tanımlanabilir.

Metin madenciliğiyle ilintili bir çalışma alanı olan Doğal Dil İşleme (NLP), bilgisayarların doğal dil metinlerini veya konuşmaları anlamak ve işlemek amacıyla nasıl kullanılabileceğini araştıran bir araştırma ve uygulama alanıdır [25]. Doğal dil işleme alanının amacı bilgisayar ve insan dilleri arasındaki boşluğu doldurarak insan dilini bilgisayarlar için anlaşılır hale getirmektir. Bu alanda doğal dilin bilgisayarlar tarafından analiz edilmesi, anlaşılması ve yeni metinler üretilmesi gibi teknolojiler üretilmesi amaçlanmaktadır [24]. NLP alanında yapılan çalışmalar neticesinde çeviri, konuşmadan metne (Speech to Text), metinden konuşmaya (Text to Speech) gibi teknolojiler üretilmektedir. NLP alanı doğal dilin; dil bilgisi kuralları, kelimelerin türleri (isim, sıfat, bağlaç, edat, fiil vb.), eş anlamlılık, eş seslilik gibi yapılarıyla ayrıntılı bir biçimde ilgilenmekte ve dil bilimi alanı ile bilgisayarlar arasında bağlantı kurmaktadır [26]. Bu açıdan bakıldığında metin madenciliği çalışmalarında yapılandırılmamış metnin işlenmesi sürecinde en büyük yardımı NLP'nin verebileceği değerlendirilebilir. Metin madenciliği uygulama süreci ele alınırken bu iş birliği daha derin bir şekilde ele alınacaktır.

Metin madenciliği alanında yapılan çalışmalar arama ve bilgi alma, belge kümeleme, belge sınıflama, web madenciliği, bilgi çıkarma, doğal dil işleme ve anlam çıkarma gibi 7 grupta ele alınabilir [24,27–29] .

1. **Arama ve bilgi alma (IR);** indeksleme, arama, büyük metinler arasından belgeleri bulma (arama motorları) gibi uygulamaları içeren çalışma alanıdır.
2. **Belge kümeleme;** veri madenciliği kümeleme metotlarını kullanarak terimleri, metin parçalarını, paragrafları ve belgeleri kümelere ayırarak etiketleme gibi işlemlerin yapıldığı çalışma alanıdır.
3. **Belge sınıflama;** etiketlenmiş metinler üzerinde eğitilmiş modellere dayalı olarak veri madenciliği sınıflandırma yöntemlerini kullanarak terimleri, paragrafları veya belgeleri gruplandırma ve kategorilere ayırma işlemlerinin yapıldığı çalışma alanıdır.
4. **Web madenciliği;** İnternet siteleri ve sayfalarının birbirleri ile olan bağlantılarını kullanarak belli bir alanda Web üzerinde veri ve metin madenciliği çalışmaları yapılmaktadır. İnternet sitelerinin transfer edilmesi için kullanılan HTML dili bu alanda yapılandırılmış metin ile çalışılması anlamında değerlendirilebilir.
5. **Bilgi çıkarma (IE);** alanında amaç yapılandırılmamış metinlerden yapılandırılmış verilerin üretilmesidir. Bu alandaki çalışmalar yüksek derecede bilgi, birikim ve uğraş gerektirmektedir.
6. **Doğal dil işleme (NLP)** alanı her ne kadar metin madenciliğine yardımcı olan ve dil bilimleriyle arasında bağlantı kuran bir alan olsa da Miner vd. [27] metin madenciliği alanında istatistiksel yaklaşımlara da ihtiyaç duyulduğu için metin madenciliği alanı altında çalışılabilecek ortak alanları olduğunu belirtmişlerdir.
7. **Anlam çıkarma;** Bilgisayarların bir metni anlaması zor bir işlem olsa da metinlerin insanlar ve bilgisayarların birlikte, bir metni yalnız yaptıklarından daha iyi bir performansla anlamlandırabilir. Örneğin Google firması, *Google Kitaplar*² üzerinde yaptığı bir çalışmayla İngilizce, Fransızca, Çince, Almanca, Rusça, İspanyolca gibi dillerde *n-gram* ($n=5$)³ veri tabanı oluşturmuş ve kullanıma açmıştır. Bu n-gramlar kullanılarak kümeleme algoritmalarıyla bir metnin hangi konuya ev sahipliği yaptığı bulunabilir.

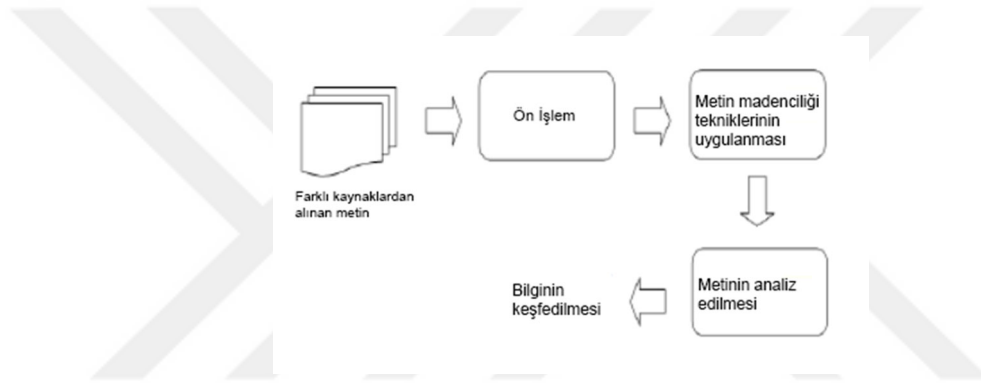
² <https://books.google.com.tr>

³ <https://books.google.com/ngrams>

2.4 Metin Madenciliği Süreci

Veri madenciliği; modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı bir yaklaşımdır [12]. Her veri türü hatta veri seti kendine özel bir süreçle analiz edilebilir. Metin üzerine de hangi alandan, nasıl ve hangi süreçle alındığı, veri türü (yapılandırılmamış, yarı yapılandırılmış), işin hedefi vb. özellikleri göz önünde bulundurularak kendine has bir süreç içinde madencilik işlemleri uygulanabilir.

Gaikwad vd. [28] göre metin madenciliği süreci farklı kaynaklarda bulunan metinlerle başlar, metin ön işleminden geçirilerek düzenlenir, veri madenciliği teknikleri uygulanır, metin analiz edilir ve bilgi keşfedilerek süreç tamamlanır (Şekil 2.5).



Şekil 2.5: Metin madenciliği süreç diyagramı [28].

Fan vd. [24] yaptıkları çalışmada doküman koleksiyonundan metnin alınması, ön işleminden geçirilmesi, analiz edilmesi, yönetim bilişim sistemlerine aktarılması ve bilginin elde edilmesi olarak sıralanan beş basamaklı bir model paylaşmışlardır (Şekil 2.6).



Şekil 2.6: Örnek metin madenciliği süreç diyagramı [24].

Gaikwad vd. [28] ve Fan vd. [24] tarafından sunulan modeller lineer modellerdir ve modellerin değerlendirilmesi aşamasına ve birbirleriyle ilişkilerine yer verilmemiştir. Miner vd. [27] ise CRISP-DM modelini kullanarak ürettikleri, süreci ve aşamaların birbirleriyle olan ilişkilerine

- Verilerin zenginliğinin araştırılması (örneğin; eldeki veri metin madenciliği çalışması için gereken bilgi içeriğine/zenginliğe sahip mi?),
- Verilerin nicelik ve niteliğinin değerlendirilmesi,
- Eğer değerlendirme olumlu sonuçlanırsa bu aşamadan sonraki iş; çeşitli kaynaklardan çalışmada kullanılacak olan büyük miktarda veriyi toplamak ve çalışmaya entegre etmek olabilir.

Bu aşamada “Eldeki veri madencilik çalışmaları yapmak için yeterli mi?” ya da “Hedefe ulaşmak hangi verileri nasıl toplamak gerekir?” soruları sorulabilir.

2.1.4 Veriyi Hazırlama

Yapılandırılmamış bir veri türü olan metin söz konusu olduğunda veri madenciliğinin en kritik aşaması, metnin doğası gereği veriyi hazırlama ya da “Metin Ön İşleme” aşamasıdır [30]. Metin ön işleme doğal dilin hayranlık uyandıran çeşitliliğini yönetilebilir bir dizi özelliğe indirgemek için kullanılan adımlardan meydana gelmektedir [31]. Başka bir deyişle metnin ön işlemde geçirilerek modellemeye hazırlanmasındaki amaç çalışılan verinin boyutunu azaltmak ve/veya modelimizin verimliliğini ve etkinliğini artırmaktır [32].

Veri madenciliği söz konusu olduğunda veriyi hazırlama aşaması temizleme, birleştirme (birden çok veri kaynağının birleştirilmesi), indirgeme ve dönüştürme olarak adlandırılabilir dört aşamadan oluşmaktadır [33]. Metin madenciliğinde, doğal dilin doğası gereği, bu aşamalar altında metni farklı madencilik algoritmalarıyla çalışmaya uygun hale getirmek üzere özel ön işleme metotlarına ihtiyaç duyulmaktadır [14].

2.4.1.1 Veriyi Temizleme

Metin temizleme adımları her metne göre farklı adımlar takip edilerek yapılabilir fakat genel olarak izlenen adımlar aynıdır [27]. Bu adımlar veriyi temizleme işleminde hedeflenen sonuca göre ön işlem, metni parçalama, yazım yanlışlarını düzeltme, normalizasyon gibi adımlar olabilir.

Ön İşlem/Hazırlık:

Daha önce bahsedildiği üzere metin verisi resmi ve resmi olmayan sayısız formatta bulunabilir. Veriyi temizleme işlemi anomalileri tespit etmek, bu anomalileri düzeltmek için dönüşüm

işlemlerinin seçilmesi ve dönüşümlerin veri kümesine uygulanması olarak üç aşamada gerçekleştirilebilir [34]. Metin verisini temizlemek ve işlenebilecek bir hale getirmek için yapılabilecek işlemler aşağıdaki gibi örneklenebilir.

- HTML etiketlerinin temizlenmesi,
- Metin içindeki standart ifadelerin çıkarılması ya da etiketlenmesi,
- Metin içinden harfler ve sayılar dışındaki karakterlerin çıkarılması,
- Metnin tüm karakterlerinin büyük ya da küçük harfe dönüştürülmesi,

Parçalama (Tokenization):

Metni temsil eden karakter dizileri parçalama işlemi yardımıyla kelimeler/terimlerden oluşan bir dizi olarak ifade edilir [35]. Metni parçalarına ayırma işleminde en çok kullanılan metot metni boşluk ya da herhangi bir işarete göre parçalarına ayırmaktır. Bu metodun sonunda metin parçalarına kelime-kelime ayrılacaktır. Bir başka problem de “U.N.” gibi kısaltmaların U ve N olarak iki jetona parçalanacak olması sonucunda oluşan kelime kayıplarıdır. Bu sorunu çözmek için çözüm olarak kısaltmaları algılayan akıllı bir parçalama işlemi yapılabilir [27].

Parçalama işlemi kelime-kelime yapılacağı gibi iki ya da daha fazla kelimenin bir araya gelmesiyle oluşan terimler, insan isimleri ya da diğer adlandırılmış varlıklar da parçalama işlemi aşamasında tanımlanarak bu kelimelerin birlikte parçalanması sağlanabilir. Terimlerin bulunmasında en çok kullanılan yöntem *n-gram* yöntemidir. *n-gram* yönteminde *n* dizi elemanının bir derlem içinde birlikte geçme sıklığı sorgulanır. *n* sayısı 2,3,4... ve daha fazla elemanı ifade edebilir.

Parçalama işleminde karşılaşılan bir başka sorun da insan isimleri, özel isimler, kuruluşlar vb. birden fazla kelimeyle yazılmış fakat *n-gram* yöntemiyle bulunamayan metin parçaları olabilir. Bu sebeple metni adlandırılmış varlıkları birbirinden ayırmadan parçalamak için *Adlandırılmış Varlık Tanıma (Named Entity Recognition - NER)* yöntemi de kullanılabilir. Örneğin; “Enerji Verimliliği Merkezi kurucu başkanı Bülent Yeşilata , Ankara'da bir toplantıya katıldı .” cümlesi bu işlemin ardından “[ORG Enerji Verimliliği Merkezi] kurucu başkanı[PER Bülent Yeşilata] ,[LOC Ankara'da] bir toplantıya katıldı .” şeklinde ifade edilir [36]. Gösterilen örnekle birlikte yapılan bir parçalama işlemine ait şema Şekil 2.8’de bulunmaktadır.



Şekil 2.8: NER kullanılarak yapılan bir parçalama işlemi [36].

Yazım Yanlışlarını Düzeltmek:

Resmi formdaki belgeler (resmi yazışmalar, kitaplar, makaleler vb.) yazmak için kullanılan programlarda yazım yanlışı düzeltme ve öneri sistemleri bulunduğu ve incelenirken yazım yanlışları düzeltildiği için yazım yanlışlarını düzeltmeye gerek kalmayabilir. Sosyal medya gibi resmi olmayan ortamlarda bulunan metinlerle çalışırken ise yanlış yazım, eksik veya parça cümleler, kısaltmalar fazlaca kullanıldığı için yazım yanlışlarını düzeltmek için oldukça büyük bir çaba harcamak gereklidir [2].

Yazım yanlışı düzeltme NLP altında bir çalışma alanıdır. En basit ifadeyle bir sözlük oluşturulur, yazım yanlışı düzeltilecek kelimeye sözlükteki en yakın kelime farklı algoritmalar yardımıyla bulunarak kullanıcıya önerilir. Bu alanda en çok kullanılan algoritmalarından birisi olan Levenshtein Uzaklığı (LD) algoritması; kaynak metin (s) ve hedef metin (t) arasındaki benzerliğin bir ölçüsüdür. Mesafe, s'yi t'ye dönüştürmek için silinmesi, eklenmesi veya değiştirilmesi gereken harf sayısıdır. Levenshtein mesafesi ne kadar büyük olursa, dizeler o kadar farklı demektir. Fakat Türkçe sondan eklemeli bir yapıya sahip olduğu için sonlu bir kelime listesine veya sabit sayıda ek içeren bir kök kelime listesine dayanan yöntemler, Türkçe bağlamında yazım yanlışlarını yakalamak için yetersizdir [2]. Özellikle Türkçe kendi morfolojik yapısına uygun olarak analiz edilmesi gereken bir dildir. Ş

Java dili ile yazılmış bir Türkçe NLP kütüphanesi olan Zemberek-NLP [37] yazım yanlışlarını düzeltmek için kullanılabilir fonksiyonlar içermektedir. *SpellCorrection* fonksiyona verilen kelimeye göre yazım yanlışı düzeltmek için öneriler üretmektedir.

“spellChecker.suggestForWord” fonksiyonuna “okuyablirim, tartışıyor, Ankar'ada, Ankara'na, knlıca” kelimelerini işlemesi için yazılan kod ve çıktısı aşağıda paylaşılmıştır.

JavanKodu

```
String[] words = {"okuyabilirim", "tartııyor", "Ankar'ada", "knlıca"};
for (String word : words) {
    System.out.println (word + " = " + spellChecker.suggestForWord (word));
}
```

Çıktı

```
okuyabilirim =[okuyabilirim]
tartııyor =[tartııyor, tartılıyor, tartınıyor]
Ankar'ada =[Ankara'da, Ankara'ya, Ankara'dan, Ankaray'da, Ankara'ma, Antara'da, Ankara'ca, Cankara'da, Anakarada,
Ankara'na, Angara'da, Ankara'mda, Ankara'nda, Ankaray'a]
knlıca =[Kanlıca, kanlıca, In'lıca, kılıca, anlıca, Kanlı'ca, kınlıca, kalıca]
```

Türkçe için başka bir sorun da Türkçe harflerin yerine Latin alfabesinde onlara yakın görünen harflerin kullanılmasıdır (ASCIIifying). Zemberek-NLP kütüphanesi yazım yanlışı düzeltirken deASCIIifying (ASCII olarak değiştirilerek yazılmış Türkçe karakterlerin yeniden Türkçe karakterlere çevrilmesi) problemini düzeltememektedir.

Normalleştirme (Normalization):

Yazım yanlışı düzeltmenin başka bir formu da normalleştirme işlemidir. Normalleştirme işlemi en geniş anlamıyla standart olmayan dil kullanılarak yazılan metinleri ya da tanınan konuşmayı standart dile çevirme işlemidir [38]. Cep telefonlarının kullanımının artması sonrasında kısa mesajların ve sosyal ağların kısıtlı yazım alanı (250 karakter) ve kullanıcıların mesajla duygu katmak istemeleri sebebiyle Hassan ve Menezes [39] tarafından örneklendirilen aşağıdaki durumlar ortaya çıkmıştır;

- Birden çok kelimeyi kısaltarak yeni kısaltılmış formlar oluşturmak. (örn., “LMS” --> “like my status” , “idk” --> “i do not know”, “rofl” --> “rolling on floor laughing”)
- Kelimelerin fonetiği sayıya benzeyen kısımlarını sayılarla ifade etmek. (örn., “4get” --> “forget” , “2morrow” --> “tomorrow”, “b4” --> “before”)
- Fonetik benzerliği olan harflerin birleştirilmesi; (örn., “fon” --> “phone”)
- Ünlü harfleri ortadan kaldırmak; (örn., “my lv” --> “my love”)
- Kelimeleri yerel şivelerle ifade etmek için kelimelerin fonetik karşılığını kullanmak. (örn; “wuz up bro” --> “what is up brother”)

- Kelimeleri yazarken duyguyu abartmak için çok fazla harf kullanmak. (örn., “Gooooooooo morniiiiiiiiiiiiiiiiing” --> “good morning”)

Türkçe dili için yapılan araştırmada aşağıdaki sistemler incelenmiştir;

- Eryiğit ve Torunoğlu-Selamet [40] tarafından geliştirilen sistem Internet sitesi ve API olarak hizmet vermektedir,
- Akın ve Akın [37] tarafından geliştirilen ve Java programlama dili ile yazılan kütüphane ise açık kaynak kodlu bir şekilde kullanımdadır.

Eryiğit ve Torunoğlu-Selamet [40] ‘e göre normalizasyon işleminde parçalayıcı, biçimsiz kelime ayrıştırıcı (IWD)’dan geçen metin büyük-küçük harf dönüştürücü, değiştirme kuralları ve sözlük araması, adlandırılmış varlık tanıma, aksan düzeltici, deASCIIfyer, sesli harf restorasyonu, yazım yanlışı düzeltici aşamalarından geçmektedir. İşlem basamakları Şekil 2.9’de paylaşılmıştır.



Şekil 2.9: Normalizasyon işlemi örnek şeması [40].

Normalizasyon işlemi Şekil 2.9’de görüldüğü üzere metin ön işleme alanında yapılan çalışmaları kapsayan ve birbirine girdi ve çıktı olarak bağlı bir sistemdir. Normalizasyon işlemi örnek girdi ve çıktıları aşağıdaki gibidir.

- Değiştirme kuralları ve sözlük araması (örn., mrb nbr? --> “Merhaba ne haber?”)
- Büyük- Küçük harf dönüşümü (örn., ALİ’NİN -> “Ali’nin”)
- Adlandırılmış varlık tanıma (örn; “Enerji Verimliliği Merkezi kurucu başkanı Bülent Yeşilata , Ankara'da bir toplantıya katıldı .”, “[ORG Enerji Verimliliği Merkezi] kurucu başkanı[PER Bülent Yeşilata] ,[LOC Ankara'da] bir toplantıya katıldı.”);
- deASCIIifying (örn; “sanliurfa” --> “Şanlıurfa”)
- Sesli harf restorasyonu; (örn., “çk krnm çkt” --> “çok karnım acıktı”)
- Yazım yanlışı düzeltimi (örn., “örenğin” --> “örneğin”)

2.4.1.2 Veriyi İndirgeme

Veriyi indirmek söz konusu olduğunda sözlük boyutunun nasıl azaltılacağı ve ne kadar az özellikle modelleme işlemine geçileceği akla gelmektedir. Sözlüğün boyutunun düşürülmesi amacıyla dolgu kelimelerinin kaldırılması, kelimelerin kök ve gövdelerine indirgenmesi ve dizin kelimesi seçimi metotları kullanılabilir.

Terim Etiketleme (Term Tagging):

Birden fazla kelimenin birlikte kullanıldığı ve birlikteyken anlam kazanan insan isimleri, yer isimleri, terimler, kısaltmalar vb. metin üzerinde işlem yaparken birbirinden ayrılmaması ya da tek bir terim olarak ifade edilmesi gereken yapılardır. Örneğin “Mustafa” kelimesi bir insan ismidir fakat 3 kelimeden oluşan “Mustafa Kemal ATATÜRK” bir terim olarak 1 kişiyi ifade eder. Bir başka örnek vermek gerekirse “Türkiye” kelimesi bir ülkeyi ifade ederken 4 kelimeden oluşan “Türkiye Büyük Millet Meclisi”, Türkiye Cumhuriyeti’nin yürütmeyi denetleme ve yasamayla yükümlü kurumudur.

Terim etiketleme amacıyla sözlük tabanlı yaklaşımlar, NER sistemleri ya da Noktasal Birliktelik Bilgisi (Pointwise Mutual Information – PMI) tabanlı yaklaşımlar kullanılabilir.

Dolgu kelimelerini kaldırma (Stopwords Removal):

Nasıl her sinyalde gürültü varsa, “Veriyi Temizleme” aşamasında anlatıldığı gibi *Dolgu kelimeleri*’de metin sinyallerinde de karşımıza çıkan bir gürültüdür. Dolgu kelimeleri dilbilimi açısından gerekli olan ve metinde çok yaygın olan ancak tek başına bir anlam ifade etmeyen

kelimelerdir [41]. Dolgu kelimeleri genel olarak derlemde yüksek frekanslara sahip kelimelerdir ve bu sebeple eğer ağırlıklandırma yöntemleri uygun seçilmemişse konu modellemesi çalışmalarında konuların anlamlandırılmasını olumsuz etkileyebilir. Özellikle konu analizi çalışmalarında dolgu kelimeleri birçok konu içinde kendine yer bulur ve her konuyu aynı şekilde etkileyen yüksek ağırlıklara sahip olurlar [23]. Dolgu kelimeleri kaldırmak sadece vektör uzayını küçültmekle kalmaz, aynı zamanda işlem hızını ve doğruluğu artırarak performansı iyileştirmeye de yardımcı olur [42].

Dolgu kelimesi kaldırma amacıyla dört metot öne çıkmaktadır. Kaur ve Buttar'e göre [42];

- **Klasik metot;** araştırmacının daha önce hazırladığı ya da standart kabul gören kelime listesindeki kelimeleri derlemde çıkarmak olarak tanımlanabilir. Python programlama dili üzerinde en çok kullanılan doğal dil işleme kütüphanelerinden biri olan NLTK [43] dolgu kelime sözlüğünde Türkçe bağlamı için şu anda kabul görmüş 53 kelime bulunmaktadır. (örn., 'acaba', 'ama', 'aslında', 'az', 'bazı', 'belki').
- **Zipf Yasasına dayalı yöntemler (Z-Yöntemleri):** Terimlerin derlemde geçme frekansına göre dolgu kelime listesi oluşturmak için üç adet metodu kapsamaktadır. Bu metotlarda kelime havuzu terim frekansı yüksek olanlar (TF-HIGH), derlemde en fazla n kadar geçen kelimeler (TF-n) ve düşük tersine belge frekansına (IDF) sahip olan kelimelerle oluşturulur.
- **Karşılıklı Bilgi (Mutual Information – MI):** Bu metotta öncelikle kelimelerin birbirleriyle ve belgeyle olan MI değerleri hesaplanır. Düşük MI bilgisi terimin ayırt etme gücünün düşük olduğunu ve sonuç olarak derlemde çıkarılması gerektiğini düşündürebilir.
- **Terim Tabanlı Rasgele Örnekleme (TBRs):** Metni oluşturan terimlerin derlem için önemini Kullback-Leibler sapma ölçüsü kullanılarak değerlendirildiği bir yaklaşımdır [44]. Verilen derlem içerisinde en düşük bilgiye sahip terimler dolgu kelimesi olarak işaretlenir ve derlemde çıkarılır.

Morfolojik İndirgeme (kelime köklerine veya gövdelerine indirgeme):

Sözlükteki kelime sayısını indirmek amacıyla kelimeler eklerinden kurtararak köklerine veya gövdelerine indirgenebilir. Bu durumda “Stemming” ve “Lemmatization” olmak üzere iki kavram öne çıkmaktadır. Bu kavramlar aşağıda tanımlanmıştır.

- *Stemming* bir kelimelerin eklerini keserek kelime kökünün bulunmaya çalışıldığı bir yöntemdir. Bu yöntemde köklere ek eklenirken oluşan ses dönüşümleri dikkate alınmaz.
- *Lemmatization* yönteminde ise kelimenin morfolojik analizini esas alınmakta ve bu işlem için detaylı sözlükler kullanılmaktadır. *Lemmatization*'un amacı ekler budandıktan sonra kalan kelime kökündeki ünsüz yumuşaması gibi değişimleri düzeltmek ve kelimenin sözlükte bulunan kökünü ya da gövdesini elde etmektir [45].

Türkçe sözlük kullanarak Stemming ve Lemmatization yapan sistemler araştırılmış ve aşağıda yer verilmiştir.

- Doğuç vd. [46] tarafından geliştirilen Lemmatizer (TDK sözlüğünü baz almakta ve ekleri tespit etmek amacıyla bir sonlu durum makinesi (SDM) kullanmaktadır. Python dili kullanılarak gerçekleştirilen algoritma Knime programı üzerinde test edilmiştir. Programa Internet aramasında ulaşamamıştır.
- Akın ve Akın [37] tarafından geliştirilen Zemberek programının Türkçe morfolojik analiz sınıfı kelimelerin morfolojik analizini yapabilmekte ve kelimelerin gövde ve köklerini çıktı olarak vermektedir. Zemberek NLP programı kullanılarak yapılan örnek analiz sonuçları Tablo 2.2'de paylaşılmıştır.

Tablo 2.2: Örnek stemming ve lemminization işlemi girdi ve çıktıları.

Girdi Kelime	Stemming	Lemmatization
kitabımızın	kitab	kitap
yağdanlığımız	yağdanlık	yağdanlık

2.4.1.3 Veriyi Dönüştürme

Metin verisi üzerinde modelleme yapmak amacıyla kelimelerin ve belgelerin sayısal olarak ifade edilmesi gerekmektedir. Belgeleri modellemenin en yaygın yolu onları sayısal vektörlere dönüştürmek amacıyla kelimeler/terimler üzerinde matematiksel işlemler yapmaktır. Belgelerin ve terimlerin sayısal olarak ifade edilmesi kavramı Vektör Uzay Modeli (Vector Space Model -VSM) olarak adlandırılabilir.

VSM oluşturmak için en standart yaklaşım “Kelime Torbası (Bag of Words – BoW)” yaklaşımıdır. BoW yaklaşımı terimlerin (*i*) ve belgelerin (*j*), her satırın bir *i* terimini, her sütunun bir *j* belgesini temsil ettiği bir matris biçiminde düzenlenebileceği anlayışına dayanır ve bu matristeki her (*i,j*) hücresi her terimin belgelerde geçme sıklığını ifade eder [47]. Bu

bağlamda karşımıza belge ve terim vektörleri adlarında iki kavram çıkmaktadır. Belge vektörleri, belgelerin içinde bulunan terimler o belgelere bağlı özellikler, terimlerin belgede geçme sıklıkları ise o niteliklerin değerleri olarak düşünülebilir; benzer şekilde, terim vektörü de niteliklere (belgeler) ve nitelik değerlerine sahiptir [47]. Terim Frekansı (TF) yönteminin baz alındığı BoW ile ilgili örnek bir Terim-Belge Matrisi (TBM) Tablo 2.3’de gösterilmiştir.

Tablo 2.3: Örnek terim-belge matrisi.

Terim	Belge 1 “the man walked the dog”	Belge 2 “the man carried the dog to the park”	Belge 3 “the dog walks to the park”	Toplam
The	2	3	2	7
Man	1	1	0	2
Walked	1	0	0	1
Dog	1	1	1	3
Carried	0	1	0	1
To	0	1	1	2
Park	0	1	1	2
Walks	0	0	1	1
Toplam	5	8	6	19

Terimlerin, belgeler içinde sayısal olarak ifade edilmeleri için birçok farklı yol takip edilerek her terimin derlemdeki önemini göstermek için bir genel ağırlık değeri atanabilir. Bu değerleri belge düzeyinde değerlere de dönüştürmek de mümkündür. Genel olarak TF’in aksine diğer terim ağırlıklandırma yöntemleri derlemde en az bulunan terimleri ödüllendirme üzerine kuruludur [48].

Tablo 2.4: Yerel ve genel terim ağırlıklandırma yöntemleri.

Yerel	Yok ya da Terim Frekansı (TF) Binary log (TF+1)
Genel	Yok Normal Genel frekans – Tersine Doküman Frekansı (IDF) IDF Entropi

Tablo 2.4’te görüldüğü üzere çok kullanılan yerel ağırlıklandırma yöntemleri arasında TF, Binary ve log (TF+1) yöntemleri bulunmaktadır. Bu yöntemleri açıklamak gerekirse;

- TF: Bir terimin bir belgede geçme sıklığını,

- Binary: Bir terimin bir belgede var olduğu durumda 1, var olmadığı durumda 0 olarak ifade edilmesini,
- $\log(TF+1)$: Ham TF'in logaritması ifade eder, böylece yüksek ve düşük TF'e sahip terimler arasında büyük frekans farklılıklarının olması önlenir [48].

En iyi bilinen genel terim ağırlıklandırma yöntemleri olan Normal, GfIdf, Idf ve Entropi ağırlıklandırma formüllerini açıklamadan önce bazı kavramları ele almak gerekmektedir. Terim frekansı (tf_{ij}) bir terimin (i) bir belge (j) içerisinde geçme sıklığını ifade eder. Belge frekansı (df_i) bir terimin (i) kaç belge içinde geçtiğini gösterir. Genel frekans (gf) ise bir terimin belgelerin tamamının bulunduğu derlem içinde geçme sıklığıdır.

Buradan yola çıkarak Dumais (1991) [48]'e göre;

- Normal ağırlıklandırma: $\sqrt{\frac{1}{\sum_j tf_{ij}^2}}$
- GfIdf: $\frac{gf_i}{df_i}$
- Idf: $\log_2\left(\frac{n-belge}{df_i}\right)$ olarak matematiksel olarak ifade edilebilen Idf yönteminde derlemde en az geçen kelimenin daha yüksek bir değer alması amaçlanmaktadır. $n-belge$; derlemde bulunan belge sayısını ifade eder.

- Entropi: $1 - \sum_j \frac{p_{ij} \cdot \log p_{ij}}{\log(n-belge)}$

$p_{ij} = \frac{tf_{ij}}{gf_i}$ ve $n-belge$ derlemde bulunan belge sayısını ifade eder.

Yerel ve genel ağırlıklandırma yöntemleri birbirleriyle kombine bir şekilde kullanılarak TBM üzerinde her terimin ham frekansı yukarıda anlatılan yerel ve genel ağırlıklandırma yöntemleri kullanılarak düzenlenebilir. Bu kombinasyon bir terimin (i), bir dökümandaki (j) ağırlığı terimin yerel ağırlığı $L_{(i,j)}$ ve genel ağırlığının G_i çarpımı olarak $w = L_{(i,j)} \times G_i$ şeklinde de ifade edilebilir [48]. Bu durumda yerel bir ağırlık yöntemi olan tf ve genel bir yöntem olan idf bir

terimi ağırlıklandırmak amacıyla kullanıldığında ortaya çıkacak formül Denklem (2.1)'de gösterilmiştir.

$$W_{(i,j)}(tf - idf) = tf_{i,j} \times \log_2 \left(\frac{n - belge}{df_i} \right) \quad (2.1)$$

Chew vd. [47] yaptıkları çalışmada yerel ve genel ağırlıklandırma metotlarının bir arada kullanılması yerine terimleri Noktasal Birliktelik Bilgisi (PMI) kullanılarak ağırlıklandırmışlardır. PMI algoritması olasılıksal bir algoritma olup derlemde iki terimin birlikte geçme sıklığını baz almaktadır. PMI'nın amacı terimlerin birlikte geçme olasılıklarını kullanarak terimleri bu duruma göre ağırlıklandırmaktır. PMI ağırlıklandırma formülü Denklem (2.2)'de gösterilmiştir.

$$pmi(w_i, w_j) = \log \left(\frac{P(w_i, w_j)}{P(w_i)P(w_j)} \right) \quad (2.2)$$

2.1.5 Modelleri Oluşturma

Metin belgesi koleksiyonlarına veri madenciliği yöntemlerinin uygulanmasının ana amaçlarından biri, onları yapılandırılmış bir formata kavuşturmadır [14]. Hangi makine öğrenmesi yönteminin seçileceğini ise verimizin doğası ve hedeflerimiz belirler.

Metin sınıflandırmasında amaç yeni belgeleri önceden tanımlanmış kategorilere atamaktır [29]. Örneğin; bir haber sitesinde daha önceden yazılmış haberler ve bunların kategorileri ışığında siteye yeni eklenen haberlere otomatik kategori ataması yapmak istenmektedir. Elde bulunan haber metinlerinin kategorileri belirli olduğu için modelimiz bu haberler ışığında oluşturulabilir ve siteye yeni eklenen haber metinlerinin hangi kategoriye (spor, siyaset, magazin ...) ait oldukları bu model ışığında tahmin edilebilir.

Metin sınıflandırma problemleri hastalık tahmini problemi üzerinde de örneklenebilir. Hastane ortamında hastaların semptomlarını dinleyen bir robotumuz olduğunu düşünelim. Bu robot hastalara belirli sorular sorarak hastanın konuştuğu metinleri kaydetmektedir, doktorlar ise hastalığı teşhis etmektedirler. Elde bulunan veriler bir sınıflandırma modelini eğitmeye yeterli seviyeye ulaştığında veri setimizde semptomlar ve o semptomların hedef hastalıkları yer

alacaktır. Böylece model eğitildiğinde yeni hastaları dinleyerek otomatik teşhis üreten (tahmin eden) bir modelimiz olacaktır.

Metin kümeleme problemi ise benzer belgeleri aynı grup altında toplamak amaçlanmaktadır [27,29]. Kümeleme problemlerinde karşılaşılan zorluk önceden belgelerin hangi kümede olduğunun bilinmemesidir. Aggarwal ve Zhai [23]'e göre belgeleri düzenleme, tarama, özetleme ve gruplama problemlerini çözmek amacıyla kullanılabilir. Örneğin Cutting vd. [49] kümeleme yöntemlerini kullanarak büyük bir metin yığını kümelemiş ve "İçindekiler" bölümünü oluşturmuşlardır. Kümeleme yöntemleri konu tespiti, duygu analizi gibi birçok probleme çözüm bulmak amacıyla kullanılmaktadır.

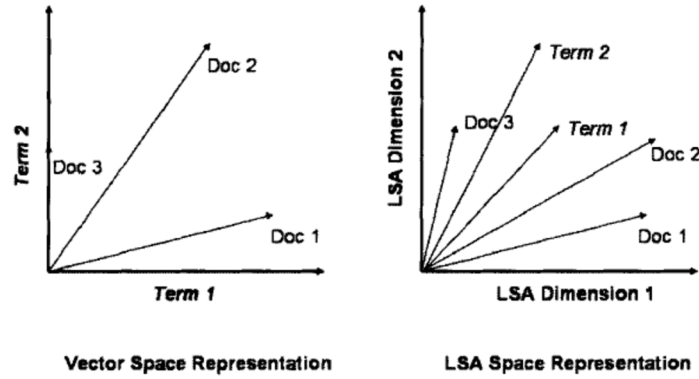
Yapılandırılmamış bir veri türü olan metinler üzerinde çalışırken, diğer veri türlerinde de kullanılan yöntemler metne özgü özellikler sebebiyle iyi çalışmazlar, bu sebeple metinler üzerinde işlem yaparken metne özgü algoritmaların kullanılması bir gerekliliktir [23]. Metinler üzerinde çalışmak için konu modelleme algoritmaları kullanılabilir. Konu modelleme algoritmaları, orijinal metinlerin kelimelerini, içindeki temaları, bu temaların birbirleriyle nasıl bağlantılı olduğunu ve zaman içinde nasıl değiştiklerini keşfetmeye çalışan istatistiksel yöntemlerdir [50]. Metin analizi alanındaki (metin sınıflandırması, kümeleme, bilgi erişimi, anlamsal analiz vb.) çok sayıda teori ve yöntem arasında, geniş metin koleksiyonlarındaki gizli tematik yapıyı keşfetme yöntemleri (konu modelleme yöntemleri), yalnızca ilgi çekmekle kalmamakta, ayrıca yalnızca bilim insanları arasında değil, aynı zamanda daha geniş bir metin analizi uygulayıcıları kitlesi tarafından kullanılmaktadır [51].

2.4.1.4 Gizli Anlamsal Analiz

Gizli Anlamsal İndeksleme (*Latent Semantic Indexing - LSI*) olarak da anılan Gizli Anlamsal Analiz (*Latent Semantic Analysis - LSA*) yöntemi Dumais vd. [52] tarafından geliştirilmiştir. Büyük kelime havuzlarına (belgelere) kelime tabanlı erişimde karşılaşılan problemlerin üstesinden gelmek için her kelimeyi daha büyük bir kelime havuzunun bir parçası olarak değerlendirmektedir. LSA, belgelerde, paragraflarda veya cümlelerde kullanım bağlamları aracılığıyla kelimeler arasındaki ilişkileri çıkarmaya yönelik tam otomatik bir istatistiksel yaklaşımdır [53].

Dumais [53]'e göre LSA yönteminin 4 ana basamağı bulunmaktadır. Bu basamaklardan ilk ikisi Veriyi Dönüştürme (2.4.1.3) başlığı altında anlatılmıştır. Bu basamaklar;

1. **Terim-Belge Matrisi:** Veriyi Dönüştürme (2.4.1.3) başlığı altında da anlatıldığı üzere, Terim Belge-Matrisi her terimin satırlarda (i) ve her bir belgenin, paragrafın ya da cümlelerin sütunlar (j) halinde gösterildiği ve her (i,j) hücresinin ise (i) teriminin, (j) belgesi içindeki frekansını gösterdiği bir tablodur.
2. **Dönüştürülmüş Terim-Belge Matrisi:** Saf frekanslarla uğraşmaktansa terim frekansları farklı ağırlıklandırma yöntemleri kullanılarak dönüştürülür.
3. **Boyut Azaltma:** Oluşturulan matrise Tekil Değer Ayrışımı (Singular Value Decomposition-SVD) tekniği uygulanarak her terim ve belge SVD tarafından oluşturulan k boyutlu uzayda vektörel olarak temsil edilir.
4. **İndirgenmiş Uzaya Alma:** Terimler ve belgeler arasındaki benzerlik ölçütleri TBM'den ziyade SVD tarafından boyutları azaltılmış uzayda hesaplanır. Böylece hem belgeler hem de terimler aynı uzayda vektörler olarak temsil edildiğinden; belge-belge, terim-terim ve terim-belge benzerliklerinin hesaplanmasında kolaylık sağlanmış olur (Şekil 2.10). Örnek olarak; yeni bir sorgu geldiğinde, oluşturulan modelin merkezinde yeni bir vektör oluşturularak bu vektörün diğer terim ve belgelere benzerlikleri hesaplanır.

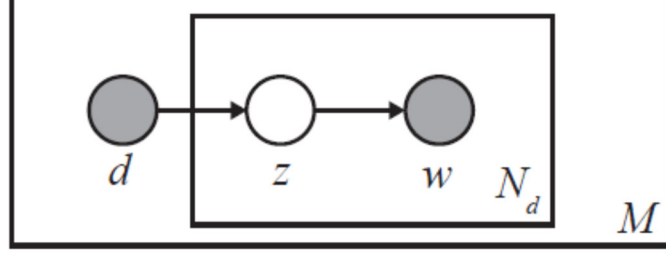


Şekil 2.10: Vektör uzay ve LSA uzay gösterimleri [53].

2.4.1.5 Olasılıksal Gizli Anlamsal Analiz

Her ne kadar LSA modeli metinleri uzayda temsil etmek ve metinler içinde arama yapmak için imkanlar sunsa da olasılıksal bir model değildir. Bu durum üzerine Hofmann [54] olasılıksal yöntemlere dayanan Olasılıksal Gizli Anlamsal Analiz (*Probabilistic Latent Semantic Analysis - PLSA*) modelini tanıtmıştır. PLSA'nın temel amacı, bir sözlüğe veya eş anlamlılar sözlüğüne

başvurmadan kelimelerin farklı kullanım bağlamlarını (eş-anlamlılık ve eş-seslilik) tanımlamak ve ayırt etmektir [55].



Şekil 2.11: PLSA modelinin grafiksel gösterimi [56].

PLSA modeli çekirdek olarak Saul ve Peveira [56] tarafından tanımlanan Görünüş Modeli (Aspect Model) yer almaktadır. Görünüş modeli, gözlemlenemeyen sınıf değişkenleri (konular) $z_k \in \{z_1, \dots, z_k\}$ ile gözlemlenebilen, kelimeler ($w \in N = \{w_1, \dots, w_k\}$) ve kelimelerin birlikte kullanımıyla ortaya çıkan belgeler $d \in M = \{d_1, \dots, d_k\}$, arasında gizli değişkenlerle gösterilebilecek bir ilişki olduğunu varsayan bir modeldir ve Şekil 2.11’de gösterilmiştir [57]. Bir belge içindeki her kelime, karışım modelinin bir örneği olarak kabul edilirken, bu karışımın bileşenleri ise gizli konuların çok terimli rastgele değişkenleri olarak nitelendirilebilir. Bir belge de konuların temsil ettiği karışım oranlarının bir listesidir. PLSA modelinde kelimelerin ve belgelerin birlikte ortaya çıkma sürecini tanımlayan üretken modeli aşağıdaki 3 aşamada tanımlanmaktadır;

1. Bir belgeye (d_i), $P(d_i)$ olasılığı atanır,
2. Bir konuya (z_k), $P(z_k|d_i)$ olasılığı atanır,
3. Kelimeler $P(w_j|z_i)$ olasılığıyla bir konuya atanır [55].

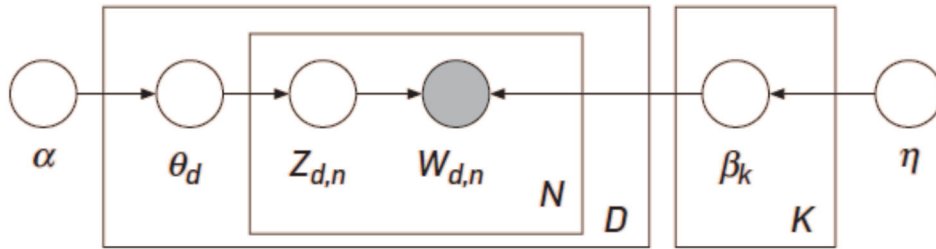
PLSA her ne kadar eş-anlamlılık ve eş-seslilik gibi sorunlara çözüm bulan olasılıksal bir model olsa da daha önce görülmemiş metinleri konu dağılımına dahil edecek kadar esnek bir yapıya sahip değildir. Ayrıca koleksiyondaki her belge için modelin tekrar eğitilmesi gerekmektedir, böylece belge sayısı arttıkça parametre sayısı artar, kısaca büyük belge koleksiyonlarıyla çalışmak imkânsız bir hale gelmektedir [58].

2.4.1.6 Gizli Dirichlet Ataması

Gizli Dirichlet Ataması (Latent Dirichlet Allocation – LDA) derlemde (**D**) bulunan gizli tematik yapıları ortaya çıkarma amaçlı *üretken* ve *olasılıksal* bir konu modelidir. Arkasında yatan ana düşünce belgelerin *gizli konuların* rastgele karışımları olarak temsil edildiği ve her konunun kelimelerin özel dağılımlarına göre oluştuğudur. Metinler için *üretken* bir model, gözlemlenen verileri (belgeler, kelimeler) açıklayabilecek en iyi gizli değişken kümesini bulmayı amaçlar. LDA modeli her ne kadar metin için geliştirilse de işbirlikçi filtreleme (bir kullanıcının aktivitelerine dayalı olarak örüntüler türetmek), içerik tabanlı görüntü işleme ve biyoenformatik alanlarında da kullanılmaktadır [59].

LDA kelime, belge ve derlem olarak üç temel kavram üzerine inşa edilmiştir. Bu temel kavramlar;

- Kelime (w), verinin en küçük yapı taşıdır ve her birisi sözlükte bir indekste temsil edilir $\{1, \dots, V\}$. Bir kelime bir bileşeni sözlükte 1 ve diğer bileşenleri 0'a eşit olan birim vektörler halinde temsil edilir.
- Bir belge (d), N elemanlı bir kelime dizisidir. $w = \{w_1, \dots, w_n\}$ şeklinde bir dizi olarak temsil edilir.
- Bir derlem (D), M belgeden oluşur ve $D = \{d_1, \dots, d_M\}$ şeklinde bir dizi olarak temsil edilir.



Şekil 2.12: LDA modelinin grafiksel gösterimi [59].

Şekil 2.12’de grafiksel gösterimi paylaşılan LDA Modeli, grafiksel gösteriminde kullanılan simgelerin anlamları şöyledir. Konular $\beta_{1:K}$ ile ifade edilir ve her β_k konuların sözlük üzerindeki dağılımını temsil eder. Her konunun *belge_d* üzerindeki ağırlığı θ_d ile simgelenir, böylece $\theta_{d,k}$, k konusunun d belgesi üzerindeki ağırlığıdır. Derlemdeki d ’inci belge için konu ataması z_d ile simgelenir, böylece $z_{d,n}$ d belgesindeki n ’inci kelimenin konu atamasıdır.

Gözlemlenen d belgesinin gözlemlenen kelimeleri w_d ile simgelenir, böylece $W_{d,n}$ d belgesinin n 'inci kelimesinin ağırlığıdır [50]. LDA modelinin tüm aşamalarında sisteme dahil olan gözlemlenen, gizli ve rastlantısal değişkenlere ait matematiksel gösterim Denklem (2.3)'te paylaşılmıştır.

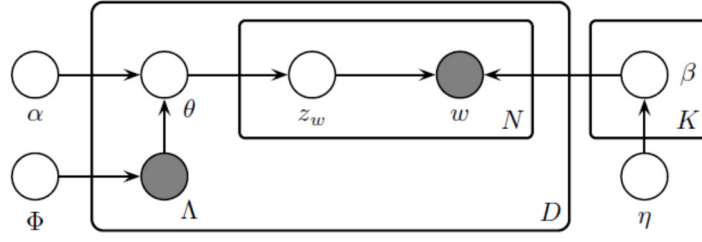
$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^K p(\beta_i) \prod_{d=1}^D p(\theta_d) \left(\prod_{n=1}^N p(z_{d,n} | \theta_d) p(w_{d,n} | \beta_{1:K}, z_{d,n}) \right) \quad (2.3)$$

Konular, ağırlıklar, konu atamaları ve kelimeler arasındaki olasılıksal dağılımın gösterildiği denklem bir dizi bağımlılığı içermektedir ve olasılık dağılımlarının β (konular), θ (oranlar) ve z (atamalar) arasında etkileşiminin hesaplanması Dirichlet dağılımıyla yapılır. LDA modeline göre konular, belgeler yazılmadan önce oluşturulur ve her belge konuları farklı oranlarda içerir. Konular, sözlükte bulunan kelimeler ve belgeler üzerinde bir olasılık dağılımına sahiptir[50].

Bir derlemi modellemek için denetimsiz bir üretken olasılık yöntemi olan LDA, en yaygın olarak kullanılan konu modelleme yöntemidir ve araştırmacılar LDA merkeze alınarak ve eklentiler geliştirerek denetimli, zaman serileri, hiyerarşik modeller gibi gruplanabilecek algoritmalar ortaya çıkarılmıştır [60]. Lee vd. [61] göre LDA modelinin en bilinen sınırlığı konular arasındaki bağlantıları ortaya koyamamasıdır. Bunun yanında LDA modelinin denetimsiz bir algoritma olması ve zaman serilerini ortaya koyamaması da farklı sınırlılıklardır.

2.4.1.7 Denetimli ve Yarı Denetimli Eklentiler

LDA Modelini denetimli öğrenme algoritması haline getirmek için geliştirilen algoritmalara Denetimli LDA (Supervised LDA -sLDA) [62], Etiketlenmiş LDA (Labeled LDA -LLDA) [63], Kısmen Etiketlenmiş LDA (Partially Labeled LDA – PLLDA) [64] algoritmaları örnek olarak verilebilir. Bu algoritmaların ortak noktaları LDA Modelini merkeze alarak her belgeye ya da belgelere kısmi olarak bir değişken eklemeleridir. Bu değişkenler bir bilmin beğenilme oranı, bir makalenin indirilme sayısı ya da kategorisi olabilir. Denetimli algoritmalar belgelerin gizli konularını bulmak için belgeler ve bu belgelere eklenen değişkenleri kullanabilir. Denetimli ve yarı denetimli eklentilerin geliştirilmesinin amacı konu modelleme algoritmalarının başarımlarını insanların gözle değerlendirmelerine yakın bir performansa ulaştırmaktır [64]. Denetimli eklentilere örnek olarak verilebilecek LLDA modelinin grafiksel gösterimi Şekil 2.13'te paylaşılmıştır.



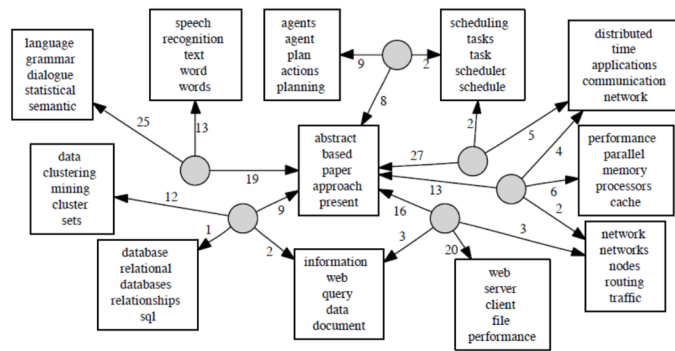
Şekil 2.13: LLDA algoritması grafiksel gösterimi [63].

LDA Modelinin grafiksel gösterimi Şekil 2.12’de paylaşılmıştı. Şekil 2.13’te gösterilen LLDA algoritmasının grafiksel yapısına göre algoritma çekirdeğine Λ ile ifade edilen bilinen belge etiketleri kümesi ve Φ ile ifade edilen bu etiketlerin konular üzerindeki ağırlıkları değişkenler olarak eklenmiştir.

2.4.1.8 Hiyerarşi Eklentileri

LDA modelinin bir sınırlılığı da belgelere konuları atayabilmesi fakat bu konular arasındaki ilişkileri ortaya koyamamasıdır. Bu sebeple konular arasındaki ilişkiyi de ortaya koymak adına İlişkili Konu Modeli (Correlated Topic Model – CTM) [65], Dirichlet Multinomial Regression (DMRModel) [66], Pachinko Allocation - PA [67] ile örneklenebilecek eklentiler geliştirilmiştir.

Örnek olarak CTM, LDA ile aynı üretken süreci kullanır fakat konuların birbirleriyle ilişkilerini ortaya koymak için Dirichlet dağılımı yerine lojistik normal dağılımı kullanır [61]. PA ise iç içe geçmiş konuların arasındaki bağlantıyı tespit etmek için yönlendirilmiş döngüsel olmayan grafik (Directed Acyclic Graph - DAG) kullanır [67].

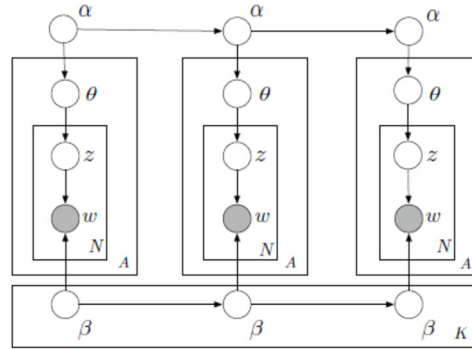


Şekil 2.14: PA ile oluşturulmuş NIPS özet veri seti (NIPS00-12) konu hiyerarşisi [67].

Şekil 2.14 incelendiğinde PA modelinin bir ana konu ve o konuyla ilişkilendirilmiş alt konulara ev sahipliği yapan konu ağaçlarını ortaya çıkarabildiği görülebilir. Ortada bulunan “özet, tabanlı, makale, yaklaşım ve sunmak” birçok bilimsel makalede bulunan ortak kelimelerdir ve bu sebeple her ana konu bu konuya bir alt konu olarak sahip çıkmaktadır. 1,2,9 ve 12 numaralı konuların bağlı olduğu ana konu “veri” ile ilişkili konulardır. Veri konusuyula ilişkili olan 12 numaralı konu “Veri Madenciliği”, 1 numaralı konu “Veritabanları”, 2 numaralı konunun “Enformasyon” olduğu görülebilir. Yine 25,13 ve 19 numaralı konuya ev sahipliği yapan ana konu “Dil” olabilir. Yine bu ana konuya bağlı 25-” Dilbilimi”, 13-”Konuşma Tanıma” gibi alt konular bulunmaktadır.

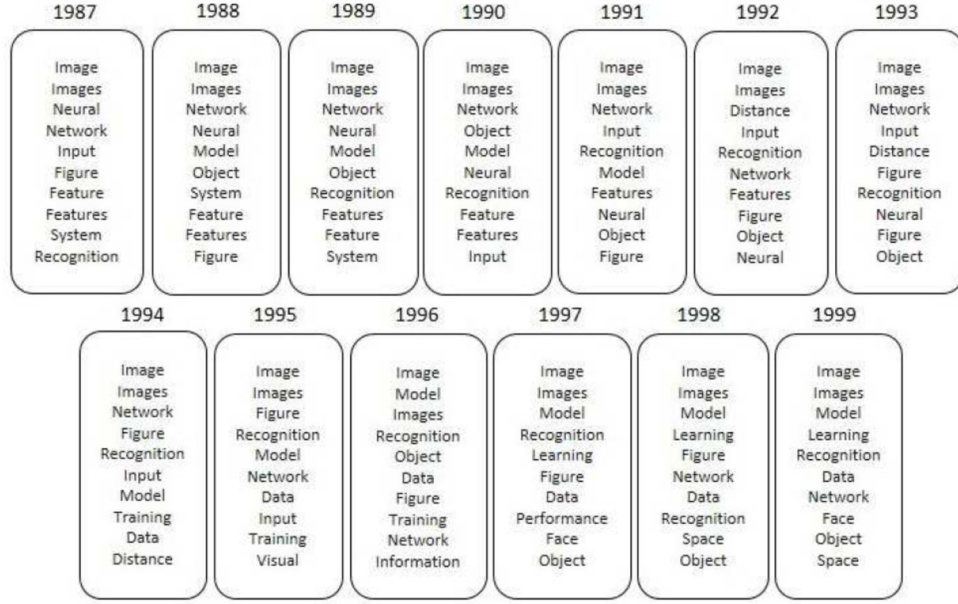
2.4.1.9 Zaman Serisi Eklentileri

Konuların zamanla farklılaşmalarını incelemek amacıyla da Dynamic Topic Models (DTM) [68] geliştirilmiştir. DTM derlemdeki her belgenin zaman dilimlerine göre organize edilmesi üzerine kurgulanmıştır. Dirichlet ataması ile zaman serilerini modellenememektedir, bu nedenle DTM dinamik modeli Gauss dağılımlarını zincirleyerek ve yayılan değerleri eşleştirerek bileşimsel rastgele değişken dizilerini modellemektir. DTM’in 3 zaman parçası için grafiksel gösterimi Şekil 2.15’te paylaşılmıştır.



Şekil 2.15: DTM'in grafiksel gösterimi (3 zaman parçası için) [68].

Şekil 2.15'te görüldüğü üzere DTM'in her zaman parçası için bağımsız bir konu modeli oluşturulduğu görülebilir. Zaman dinamiği göz önünde bulundurulduğunda (t) zaman parçasındaki (k) konusunun, ($t-1$) zamanındaki (k) konusundan evrimleştiği varsayılabilir. Bu değişimin bir örneği Şekil 2.16'da paylaşılmıştır [69].



Şekil 2.16: NIPS veritabanı kullanılarak oluşturulan “Image Processing” konusu altındaki kelimelerin yıllara göre evrimi [69].

Şekil 2.16’da görüldüğü gibi 1987-1999 yılları arasında akademik makalelerde “Image Processing” konusuna altında toplanan kelimelerin görüntü işleme alanındaki algoritmalar güçlendikçe konuya “Yüz Tanıma” ve “Nesne Tanıma” gibi alanların eklendiği görülebilir [69].

2.4.1.10 Duygu Analizi

Duygu analizi (Sentiment Analysis- SI) temel olarak bir metin işleme tekniği olup bir derlemin duygusal olarak ifade etmek istediği duyguyu belirlemek amaçlanır[70]. Örneğin; Bir e-ticaret sitemiz olsun, bu e-ticaret sitesinde sattığımız ürünlere müşterilerin yorumlarının olumlu ya da olumsuz olma durumunu merak etmekteyiz. Bu durumda metin işleme tekniklerinden yararlanarak müşterilerin kaçının olumlu, kaçının olumsuz yorum yazdığını bulmaya çalışırız.

Literatürde ilk ve halen üzerinde en çok çalışılan uygulama duygusal kutupları (sentimental polarity) araştırmaktır [70]. Hussein [71] yaptığı çalışmada SI alanında en çok kullanılan tekniklerin konuşma bölümü etiketleme (part of speech tagging) ve sözlük tabanlı yaklaşımlar (lexicon-based) olduğunu değerlendirmişlerdir. Bu durumun duygusal kutup analizinin bu alandaki en yoğun çalışma alanı olduğunu destekler nitelikte olduğu görülmektedir. Aynı şekilde Can ve Alataş [72] yaptıkları derleme çalışmasında pozitif ve negatif duyguların ortaya çıkarıldığı yaklaşımları incelemişlerdir. Literatürde bulunan çalışmaların haber siteleri,

alışveriş siteleri, mikroblog servisleri gibi Internet kaynaklarından alınan veriler üzerine kurgulandığı görülmektedir.

Sözlük tabanlı yaklaşımda, kelimelerin pozitif ve negatif olarak işaretlendiği ya da (-1,+1) aralığında bir duygu puanı verildiği sözlükler kullanılabilir. Sözlük tabanlı yaklaşımla belgelerde duygusal kutup analiziyle ilgili matematiksel ifade Denklem (2.4'te paylaşılmıştır.

$$pol(d) = \frac{\sum_{i=1}^n pol(w_i)}{n} \quad (2.4)$$

Denklem (2.4'te paylaşıldığı üzere, d belgesinin duygusal ağırlığı tüm kelimelerinin sözlükteki ağırlıklarının ortalamasıyla eşdeğerdir. Türkçe dili için yapılan araştırmada; Dehkharghani vd. (2016) [73] in oluşturdukları Sentiturknet⁴ ve Sağlam vd. [74] tarafından geliştirilen SWNetTR⁵ sözlükleri bulunmuştur. SWNetTR kütüphanesinde 49.000 kelime bulunmaktadır ve başarımı 84.42% olarak hesaplanmıştır.

2.1.6 Sonuçların Değerlendirilmesi

Konu modelleri tarafından keşfedilen gizli uzayın anlamlı ve faydalı olduğuna dair uzun süredir devam eden bir varsayım olsa da denetimsiz bir modeli değerlendirmek oldukça zordur. Tüm derlemler için bir standart konu listesi olmadığından farklı değerlendirme metotlarına ihtiyaç duyulmaktadır [75]. Konu modellerinin değerlendirilme aşamalarıyla ilgili grafiksel gösterim Şekil 2.17'de paylaşılmıştır.

Değerlendirme ölçümleri doğrudan ve dolaylı ölçümler olarak ikiye ayrılmaktadır. Doğrudan ölçümler genellikle model oluşturulurken hesaplanırken dolaylı ölçümler için ek hesaplamalara ihtiyaç duyulmaktadır. En çok kullanılan doğrudan ölçüm yöntemi olan Logaritmik Olasılık (*Log-Likelihood*)'a ait matematiksel gösterim aşağıda verilmiştir.

$$\log - \text{likelihood}(W', W^*) = \log \frac{P(W', W^*) + \epsilon}{P(W', \neg W^*) + \epsilon}; \epsilon = 10^{-12}; W', W^* \in W \quad (2.5)$$

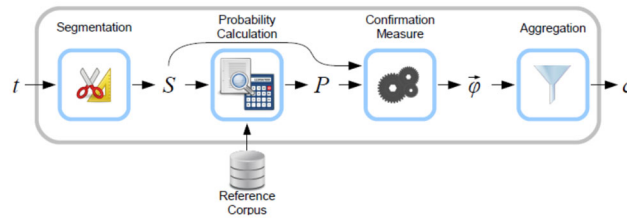
⁴ <https://github.com/akoksai/BERT-Sentiment-Analysis-Turkish/tree/master/Other%20Data/SentiTurkNet>

⁵ <https://github.com/swnettr/SWNetTR>

Başka bir doğrudan ölçüm yöntemi de Şaşkınlık (*Perplexity*) değeridir. Perplexity ya da başka bir deyişle “Tahminsel Olabilirlik – Predictive Likelihood” , test verilerinin olasılığında sabit olarak azalmaktadır ve matematiksel olarak kelime başına olasılığın (*likelihood*) geometrik ortalamasının tersi ile eşdeğerdir [59]. Düşük Perplexity ve artan Log-Likelihood değeri veri seti üzerinde daha yüksek bir modelleme doğruluğu anlamına gelir [76]. M belgeye sahip bir derlem için Perplexity değerine ait formül Denklem (2.6’te paylaşılmıştır.

$$perplexity(D_{test}) = \exp \left\{ \frac{\sum_{d=1}^M \log p(w_d)}{\sum_{d=1}^M N_d} \right\} \quad (2.6)$$

Perplexity değeri maksimum tekrar (iteration) ve konu sayısını belirlemede oldukça başarılı bir ölçüt olsa da konuların insanlar tarafından anlaşılabilirliği ve yorumlanabilirliğini sınayan ek ölçüm metotlarına ihtiyaç duyulmaktadır [77]. Doğrudan yöntemler model oluşturulurken otomatik olarak hesaplanırsa da dolaylı yöntemlerle tutarlılık hesabı için veriyi tekrar örneklemek ve değerlendirme yöntemini uygulamak gerekmektedir. Röder vd. [77] önerdikleri model değerlendirmesi çerçeve modeli Şekil 2.17’de paylaşılmıştır.



Şekil 2.17: Model değerlendirmesi çerçeve modeli [77].

Şekil 2.17’te göre bir değerlendirme yöntemi öncelikle derlemin parçalara ayrılması, olasılık hesabı yapılması, değerlendirme yönteminin uygulanması ve birleştirilmesi olmak üzere dört adımdan meydana gelmektedir [77].

Parçalama (S) aşamasında birden farklı metot kullanılabilir. Bu parçalama metotlarından bir-bir, bir-önce ve bir-sonra yöntemleri aşağıdaki matematiksel formüllerle açıklanabilir.

$$S_{bir}^{bir} = \{W', W^* | W' = \{w_i\}; W^* = \{w_j\}; w_i, w_j \in W; i \neq j\} \quad (2.7)$$

$$S_{önce}^{bir} = \{W', W^* | W' = \{w_i\}; W^* = \{w_j\}; w_i, w_j \in W; i > j\} \quad (2.8)$$

$$S_{sonra}^{bir} = \{W', W^* | W' = \{w_i\}; W^* = \{w_j\}; w_i, w_j \in W; i < j\} \quad (2.9)$$

Eğer elimizde sıralı bir kelime seti varsa, Denklem (2.7 kelime, başka bir kelimeyle, (2.8 ve (2.9 'de gösterildiği üzere de bir kelime kendinden önceki ya da sonraki kelime ile eşleştirilir. Douven ve Meijs (2007) [78] ve Röder vd. (2015) [77] bu parçalama yöntemlerine ek olarak aşağıdaki yöntemleri önermiştir.

$$S_{hepsi}^{bir} = \{W', W^* | W' = \{w_i\}; w_i \in W; W^* = W\{w_i\}\} \quad (2.10)$$

$$S_{herhangi}^{bir} = \{W', W^* | W' = \{w_i\}; w_i \in W; W^* \subseteq W\{w_i\}\} \quad (2.11)$$

$$S_{herhangi}^{herhangi} = \{W', W^* | W', W^* \subset W; W' \cap W^* = \emptyset\} \quad (2.12)$$

$$S_{küme}^{bir} = \{W', W^* | W' = \{w_i\}; w_i \in W; W^* = W\} \quad (2.13)$$

Denklem (2.10, (2.11, (2.12, (2.13'de görüldüğü üzere *Bir-hepsi* yönteminde bir kelime, sözlükteki diğer her kelimeyle eşleştirilir. *Bir-herhangi* sözlükteki kelime alt kümelerini kullanarak *bir-hepsi* yöntemini genişletir. *Herhangi-herhangi* ise her alt kümeyi diğer tüm ayrık alt kümelerle karşılaştıran başka bir varyasyondur.

Doğrulama ölçütleri veri setinden alınan iki örneğin birbirini ne kadar güçlü ifade ettiğini ortaya koymak için kullanılan ölçütlerdir. Araştırma kapsamında C_{UCI} , C_{UMASS} , C_{NMPI} ve C_V başarımlı ölçütleri kullanılacaktır. Bu bağlamda tez kapsamında bu ölçütlere kaynaklık eden Logaritmik Koşullu Olasılık (M_{lc}), Logaritmik Oran (M_{lr}), Normalize Edilmiş Logaritmik Oran (M_{nlr}) ve Kosinüs Benzerliği (M_{cos}) ölçütlerine ait denklemler aşağıda paylaşılmıştır.

$$S_i = (W', W^*) \text{ ve } \epsilon = 10^{-12} \text{ olmak üzere;} \quad (2.14)$$

$$M_{lc}(S_i) = \log \frac{P(W', W^*) + \epsilon}{P(W^*)}$$

$$M_{lr}(S_i) = \log \frac{P(W', W^*) + \epsilon}{P(W^*) * P(W')}$$

$$M_{nlr}(S_i) = \frac{M_{lr}(S_i)}{-\log (P(W', W^*) + \epsilon)}$$

$$\vec{u} = \vec{v} (W') \text{ ve } \vec{u} = \vec{v} (W^*) \text{ kelime setlerinin vektörize edilmiş halleri ve } \quad (2.15)$$

$$S_{\cos} \left(\vec{u}, \vec{v} \right) = \frac{\sum_{i=1}^{|W|} u_i \cdot v_i}{\|\vec{u}\|_2 \cdot \|\vec{v}\|_2} \text{ olmak üzere;}$$

$$M_{\cos (nlr,1)}(S_i) = S_{\cos} \left(\vec{v}_{nlr,1} (W'), \vec{v}_{nlr,1} (W^*) \right) \text{ olarak ifade edilebilir.}$$

Olasılık atamasındaki amaç, olasılıkların veri setinden nasıl türetileceğini tanımlar. Belge olasılığı (P_{bd}), tek bir kelimenin olasılığının, kelimenin geçtiği belge sayısının toplam belge sayısına bölünmesiyle tahmin edilmesi üzerine kuruludur. Aynı şekilde, iki kelimenin ortak olasılığı, her iki kelimeyi içeren belge sayısının toplam belge sayısına bölünmesiyle tahmin edilir. Kayan pencere yaklaşımı (P_{sw}) ise belirli bir kelimeyi gösteren (pencere büyüklüğü) bir pencerenin belge üzerinde adım adım kaymasını ifade eder. Örneğin $P_{sw(10)}$ pencere büyüklüğünün 10 kelime olduğunu ifade etmektedir [77].

Toplama işleminde tüm alt kümelerde hesaplanan tutarlılık puanları bir araya getirilir. σ_a aritmetik ortalama, σ_m ise medyan anlamında kullanılabilir [77].

Röder vd. [77] oluşturdukları çerçeveye göre şekillenen başarımlar ölçütlerinin genel formüllerini paylaşmıştır. Yukarıda anlatılan parçalama, doğrulama ölçütü, olasılık ataması ve toplama işlemlerinin birleştirilerek kullanıldığı C_{UCI} , C_{UMASS} , C_{NMPI} ve C_V başarımlar ölçütlerine ait genel formüller aşağıdaki gibidir.

$$C_{UMASS} = (P_{bd}, S_{\text{önce}}^{bir}, m_{lc}, \sigma_a) \quad (2.16)$$

$$C_{UCI} = (P_{sw(n)}, S_{\text{bir}}^{bir}, m_{lr}, \sigma_a) \quad (2.17)$$

$$C_{NMPI} = (P_{sw(n)}, S_{\text{bir}}^{bir}, m_{nlr}, \sigma_a) \quad (2.18)$$

$$C_V = (P_{sw(n)}, S_{\text{küme}}^{bir}, m_{\cos(nlr,1)}, \sigma_a) \quad (2.19)$$

Örnek olarak C_{UMASS} başarımlar ölçütünün belge olasılığı olasılık atamasına ve bir-önce parçalama düzenine sahip olduğu, logaritmik koşullu olasılık doğrulama ölçütüne göre ağırlıklandırma yaptığı ve tüm işlemleri aritmetik toplam ile topladığı değerlendirilebilir. Biraz daha karmaşık bir yapıya sahip olan C_V başarımlar ölçütü ise kayan pencere olasılık atamasına ve

bir-küme parçalama düzenine sahiptir. Kelimelerin normalize edilmiş logaritmik oranları kosinüs benzerliğine göre ağırlıklandırılır ve tüm işlemler aritmetik toplam ile birleştirilir.

2.1.7 Uygulama

Modellerin oluşturulması ve değerlendirilmesi projenin bittiği anlamına gelmez. Veriden alınan bilginin kullanıcılara sunulmak için organize edilmesi ya da farklı araçlarla kullanılması için hazırlanması gerekmektedir. Veri madenciliği süreci hangi amaçla yapılıyor olsun, oluşturulan modellerin hangi amaçlara hizmet edeceğinin iyi anlaşılması gerekmektedir [12].

Uygulama aşaması elde edilen bilgilerin paylaşılması dışında modelin farklı yazılım platformlarına da eklenmesini içerebilir. Örneğin 2.1.5 “Modelleri Oluşturma” başlığı altında yer alan haber sitesi örneğine tekrar dönersek, haberler kategorilendirilip model oluşturulduktan sonra haber sitesine entegre edilebilir. Böylece her yeni eklenen haber için kategori tahmini otomatik olarak yapılacak ve yeni eklenen haberlerin doğru kategorilere atanması işi kolaylaştırılacaktır.

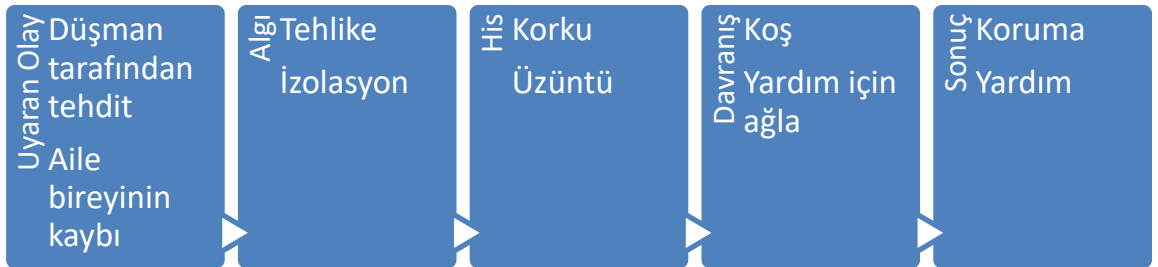
2.5 Temel Duygular ve Değerler

2.1.8 Temel Duygular

Temel duyguların sınıflandırılmasında araştırmacıların üzerinde hemfikir oldukları ortak bir nokta bulunmamaktadır. Duyguların en temel hali olarak olumlu-olumsuz ya da hoş-ağrı veren-gitmeyen şeklinde duygusal kutuplar yer almaktadır [79]. Literatürde farklı bakış açıları gruplanmış temel duygu teorileri bulunmaktadır. Cornelius [80] tarafından derlenen bu bakış açıları aşağıda örneklenmiş ve kısaca bilgi verilmiştir.

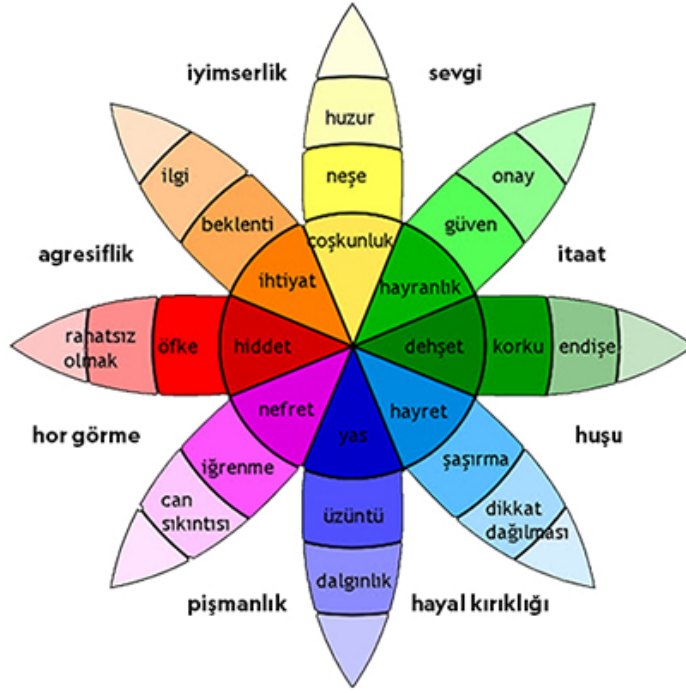
- **Darwin Perspektifi:** Darwin perspektifine göre duygular karşılaştığımız belirli sorunları çözdüğü için evrimsel süreçte seçilmiş, bireyin hayatta kalması için önemli işlevleri olan olgulardır. Bu perspektife göre temel 6 duygu *mutluluk, üzüntü, korku, iğrenme, öfke ve şaşırma*dır. Shaver vd. [81] yaptıkları bir araştırmada 3 farklı kültürden (İtalya, Çin ve Amerika) insanlara uzun bir duygu listesi vererek gruplamalarını istemişler, sonrasında ise elde ettikleri veriyi hiyerarşik kümeleme analizine tabi tutarak *sevgi, eğlence, şaşırma, sinirlenme, üzüntü ve korku* olmak üzere 6 temel duyguya ulaşmışlardır [80].

- **James Perspektifi:** James, Darwin'in evrimsel bakış açısı çerçevesinde duyguların bedensel değişikliklerle birlikte ortaya çıktığını savunmaktadır. Bu bakış açısı duygusal deneyimin doğasını açıklamakla ilgilenmektedir. James'e göre; duyguları, bedenlerimiz çevrenin bizim için hayatta kalma ile ilgili önemi olan özelliklerine otomatik olarak ve uyarlanabilir bir şekilde yanıt verecek şekilde evrimleştiğinden, deneyimlediğimizi savunmaktadır.
- **Bilişsel Perspektif:** Bilişsel bakış açısının temel varsayımı, düşünce ve duygunun birbirinden ayıramaz olduğudur. Arnold'a göre duygular çevredeki olayların bizim için iyi ya da kötü olarak yargıladığımız sürece bağlı olarak ortaya çıkar. Bilişsel perspektife göre duygular fizyolojik uyaranlara birlikte bu uyaranlara yanıt olarak gösterilen duygunun bilişsel yönüyle de ilgilenir [79]. Örneğin bazı psikolojik rahatsızlıklarda ilaç yerine ilaç görünümünde fakat hiçbir işlevi olmayan maddeler kullanılır. Bu ilaçların hiçbir fizyolojik etkisi olmamasına rağmen hasta bu ilaçları içtikten sonra rahatlamaktadır [79].
- **Sosyal Yapılandırmacı Perspektif:** Sosyal yapılandırmacılar, duyguların anlamlarını ve tutarlılıklarını öğrenilmiş sosyal kurallara borçlu olan kültürel ürünler olduğuna inanırlar. Kültür, duyguları üreten değerlendirmelerin içeriğini sağlar. Değerlendirme süreci biyolojik bir adaptasyon olsa da değerlendirmelerimizin içeriği kültürelidir [80].



Şekil 2.18: Duygunun oluşum süreci [82].

Temel duyguları ve bu duyguların birbiriyle karışımıyla ortaya çıkan duygu ve davranışları modelle yen Plutchik'e göre [82] bir duygu bir uyarıcı olayla oluşmaya başlar, olay algılanır, duygu hissedilir, duyguya cevap verilir ve bu cevabın sonuçları hissedilir. Duygunun oluşumuyla ilgili diyagram Şekil 2.18'da paylaşılmıştır. Plutchik duygu çemberi olarak da adlandırılan temel duygular ve o duyguların birbiriyle karışımıyla oluşan alt duyguları şematize etmiştir [83]. Bu duygu şeması Şekil 2.19'de paylaşılmıştır.



Şekil 2.19: Plutchik duygu çemberi [83].

Şekil 2.19’de paylaşılan duygu çemberinin merkezinde temel 8 duygu bulunmaktadır. Çemberin her kanadında dışa doğru gidildiğinde duygunun derecesi azalmaktadır. Temel bir duygunun karşısında yer alan duygu o duygunun tersini temsil etmektedir. Duygular arasında yer alan duygular ise o iki duygunun karışımı olan duygulardır [83].

2.1.9 Kök Değerler

Okullarda verilen değerler eğitimi, toplumun çoğunluğu tarafından benimsenmiş değerlerin yeni nesile kazandırılması süreci olarak tanımlanabilir. Özellikle ülkemiz göz önüne alındığında, toplumun idamesi için milli değerlerin kazandırılması daha sistematik hale getirilmiş ve dolaylı veya doğrudan öğretim programlarına işlenmiştir [84]. Türkiye Cumhuriyeti - Milli Eğitim Bakanlığı’da [85] toplumun hemfikir olduğu adalet, dostluk, dürüstlük, öz denetim, sabır, saygı, sevgi, sorumluluk, vatanseverlik ve yardımseverlikten oluşan 10 kök değerle ilgili bir liste yayımlamış ve eğitim programına dahil etmiştir. Kök değerlerle ilgili tutum ve davranış örnekleri Tablo 2.5’te paylaşılmıştır.

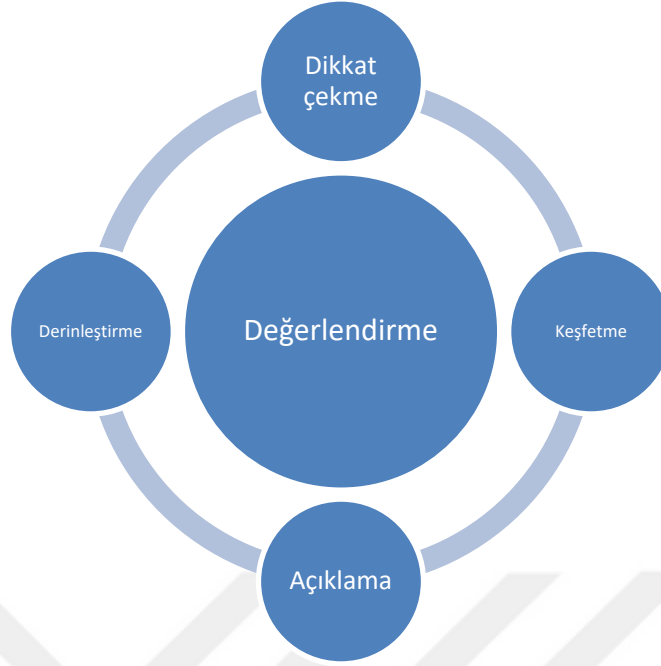
Tablo 2.5: Kök değerler ve değerlerle ilişkili bazı tutum ve davranışlar [85].

Değerler	Değerlerle İlişkili Bazı Tutum ve Davranışlar
Adalet	adil olma, eşit davranma, paylaşma...
Dostluk	diğerkâmlık, güven duyma, anlayışlı olma, dayanışma, sadık olma, vefalı olma, yardımlaşma...
Dürüstlük	açık ve anlaşılır olma, doğru sözlü olma, güvenilir olma, sözünde durma...
Öz denetim	davranışlarını kontrol etme, davranışlarının sorumluluğunu üstlenme, öz güven sahibi olma, gerektiğinde özür dileme...
Sabır	azimli olma, tahammül etme, beklemeyi bilme...
Saygı	alçakgönüllü olma, başkalarına kendine davranılmasını istediği şekilde davranma, diğer insanların kişiliklerine değer verme, muhatabının konumunu, özelliklerini ve durumunu gözetme...
Sevgi	aile birliğine önem verme, fedakârlık yapma, güven duyma, merhametli olma, vefalı olma...
Sorumluluk	kendine, çevresine, vatanına, ailesine karşı sorumlu olma; sözünde durma, tutarlı ve güvenilir olma, davranışlarının sonuçlarını üstlenme...
Vatanseverlik	çalışkan olma, dayanışma, kurallara ve kanunlara uyma, sadık olma, tarihsel ve doğal mirasa duyarlı olma, toplumu önemseme...
Yardımseverlik	cömert olma, iş birliği yapma, merhametli olma, misafirperver olma, paylaşma...

Sosyal yapılandırımcılık perspektifinde düşünüldüğünde nasıl kızdığımız veya korktuğumuz kültürel olarak belirlenmektedir [80]. Değerler; duygularımızın sosyal yapılandırımcılık perspektifiyle değerlendirildiğinde içinde yaşadığımız kültürle etkileşimiyle yaşadığı dönüşüm olarak da yorumlanabilir. Örnek olarak; öfke duygusunun hiddete dönüşmesi ya da daha ağır bir duygu olan, hiddet ve nefretin karışımı hor görme duygusunun öz denetim ve saygı değerleriyle engellenebileceği düşünülebilir.

2.6 5E Eğitim Modeli

5E Eğitim modeli ya da orijinal olarak “BSCS 5E” 1980’li yılların ortalarında IBM tarafından desteklenen BSCS (Biological Sciences Curriculum Study - Biyoloji Bilimi Müfredat Çalışması) araştırma grubu tarafından ortaokullar için yeni bir bilim ve sağlık müfredatı geliştirmek amacıyla ortaya koyulmuştur [86]. Şekil 2.20’de paylaşıldığı üzere model değerlendirme basamağı merkezde olmak üzere dikkat çekme, keşfetme, açıklama, derinleştirme basamaklarını içinde bulunduran sarmal bir modeldir. Model bilişsel psikoloji, yapılandırımcı öğrenme teorisi ve fen bilimleri eğitimi alanında deneyimlerin üzerine inşa edilmiştir [87].



Şekil 2.20: 5E eğitim modeli basamakları [87].

Şekil 2.20’de bulunan beş basamak incelendiğinde;

- **Dikkat çekme** aşamasında öğretmen öğrencilerin önceden getirdikleri bilgileri değerlendirir ve kavramlar hakkında olabilecek yanlışları kavramaya çalışır. Öğrencilerin öğrenme eylemine zihinsel anlamda odaklanmaları amaçlanır. Bu aşamada yapılacak öğrenme aktiviteleri öğrencilerin geçmiş deneyimleriyle yeni öğrenme yaşantıları arasında bağlantı kurmalarını sağlamaktadır.
- **Keşfetme** aşaması öğrenciyi merkeze alarak öğrencileri kavramları aktif olarak keşfetmelerine fırsat vermek için kurgulanan aktiviteler içermektedir. Bu aşamada öğretmenin rolü, direkt bilgiyi vermektense öğrencilerin bireysel ya da grup halinde çalışarak kavramları keşfettikleri bir ortam oluşturmak ve yardımcı rolüyle öğrencileri yüreklendirmektir. Bir kavramı öğrenmek ya da bir konuda deneyim sahibi olmak için zamana ihtiyaç duyulmakta olduğundan keşfetme basamağına yeterince zaman ayrılması gerekmektedir [88,89].
- **Açıklama** aşaması öğrencilerin keşfetme basamağında özgürce inceledikleri ve anlamaya çalıştıkları konuyu öğretmenin de yardımıyla açıkladıkları aşamadır. Öğrenciler, öğretmen konuyu açıklamadan önce konu hakkında kendi açıklamaları ve fikirlerini geliştirir ve sunarlar. Öğrenciler kendi açıklamalarını yapmadan önce

öğretmen bilimsel ve teknik kavramları sunarak öğrencilerin yanlış anlamalarını düzeltmeye ve konu hakkında derinlemesine düşünmeye teşvik eder.

- **Derinleştirme** aşaması kapsamında öğrenciler konular hakkında kendi anlayışlarını ortaya koydukları ve yeni becerilerini pekiştirdikleri aşamadır. Öğrenciler geliştirdikleri yeni beceriler ve zihinlerinde yeni oluşan kavramları deneyler ve modeller kurarak test ederler ve anlayışlarını derinleştirirler. Öğrenciler bu aşamada yeni araştırmalar yapabilir, ürünler geliştirirler. Fikirlerini, bilgi ve becerilerini farklı disiplinlere uygulamaya çalışırlar.
- **Değerlendirme** basamağı geleneksel bir fen bilimleri dersinin değerlendirme anlayışının dışında doğrudan ve dolaylı değerlendirme yöntemlerinin birlikte kullanıldığı ve modelin her aşamasında uygulanan bir yapıdadır. Portfolyolar, performans değerlendirmesi, kavram haritaları, araştırma günlükleri gibi farklı değerlendirme araçları kullanılabilir. Değerlendirme araçları derslerin her aşamasında uygulanmak, sonucun dışında öğrenme süreçlerinin de değerlendirildiği bir yapıda kurgulanır.

2.7 Bilgi İşlemsel Düşünme

Bilgi işlemsel düşünme kavramı yalnızca bilgisayar bilimciler için değil okuma, yazma ve aritmetik dışında herkesin analitik düşünme yeteneğine eklenmesi gereken bir kavramdır [90]. İnsan beyni, modern bilgi işlem cihazları gibi analitik ve bilgi işlemsel düşünme üzerine kuruludur [91]. Bilgisayarlar insan beyninin problem çözme süreçlerinden ilham alınarak geliştirilmiştir, bu sebeple bilgi işlemsel düşünme eğitimi beyinde var olan problem çözme becerilerini ortaya çıkarmayı amaçlar.

Bilgi işlemsel düşünme kavramı bilgisayar bilimleri alanının temel kavramlarından yararlanarak problem çözme, sistemler tasarlamayı ve insan davranışlarını anlamayı içeren bir kavram olarak karşımıza çıkmaktadır [90]. Denning ve Tedre [92]'ye göre bilgi işlemsel düşünme kavramı hesaplamaları bilgisayarların yaptıkları gibi yapmak için denklemler *tasarlayabilmek* ve Dünya'yı bilgi işlem alanının karmaşıklığıyla *açıklayıp yorumlayabilmek* olarak iki temel düşünce üzerinde tanımlamışlardır. Tasarlama kavramı insanlara yardımcı olmak amacıyla metotlar ve makineler tasarlamak amacıyla kullanılan mühendislik metotlarını kapsar [92]. Açıklayıp yorumlayabilmek ise oluşturulan mühendislik metotlarını açıklayabilmek ve yorumlayabilmek anlamında kullanılabildiğinden bu iki kavram birbiriyle iç

içe geçmiş şekilde karşımıza çıkmaktadır [92]. Başka bir deyişle matematiksel düşünme ve mühendislik düşüncesi kavramlarının birlikte kullanılarak gerçek dünyada çalışan karmaşık bir sistemi tasarlayıp değerlendirebilmeyi kapsar ve analitik düşünce kavramının bir türüdür [90].

ISTE ve CTSA [93]'ya göre bilgi işlemsel düşünme kavramı;

- Problemleri çözmek için bilgisayar gibi teknolojik ekipmanları kullanabilmeyi,
- Verileri mantıksal olarak organize ve analiz edebilmeyi,
- Modeller ve simülasyonlar gibi kavramlarla veriyi somutlaştırarak gösterebilmeyi,
- Algoritmik düşünme yardımıyla problem çözümlerini otomatikleştirebilmeyi,
- Olası çözümlerin en etkili kombinasyonunu bulmak amacıyla çözümleri tanımlama, analiz etme ve uygulamayı
- Bir problem çözme sürecini farklı birçok duruma adapte edebilmeyi kapsayan fakat bu yeterliliklerle sınırlı olmayan karakteristik yeterlikleri kapsamaktadır.

Bu karakteristik yeterliklerin yanında bilgi işlemsel düşünme süreciyle ilişkili kavramlar bulunmaktadır. Bilgi işlemsel düşünme süreciyle ilişkili kavramlar Şekil 2.21 paylaşılmıştır.



Şekil 2.21: Bilgi işlemsel düşünmenin alt boyutları [94].

Şekil 2.21’de bulunan kavramlar incelendiğinde soyutlama algoritma tasarımı, otomasyon, veri toplama, veri çözümleme, veri sunma, ayırıştırma, eş zamanlı çalışma, örüntü tanıma, örüntü genelleştirme ve modelleme olarak sıralandığı görülebilir. Soyutlama kavramı bir problemi çözmek amacıyla veriye odaklanma ve ilgili olmayan verileri kapsam dışında bırakmak olarak tanımlanabilir [94]. Algoritma tasarımı kavramı ise herhangi bir problemi çözmek amacıyla akış şemaları oluşturmak, oluşturulan süreci izlemek ve değerlendirmekle ilgili bir kavram olarak değerlendirilebilir. Otomasyon kavramı algoritma tasarımıyla ilişkili olsa da sürekli tekrarlayan iş ve işlemleri yapabilmek için bilgisayar ve makineleri kullanmak ve bu cihazları belirli bir akış şeması çerçevesinde bu makineleri programlayabilmek olarak tanımlanabilir.

2.8 Harezmi Eğitim Modeli

Harezmi Eğitim Modeli (HEM)’ne ait çalışmalar “Zihinden Makineye Bilgisayar Bilimleri ve Disiplinler Arası Eğitim Çalıştayı” ile başlamıştır. Çalıştay kapsamında;

- Konusunda uzman akademisyen, yönetici ve öğretmenlerin katılımıyla bir eğitim programı yazma çalıştayı düzenlenmesi ve program yazma sürecinde uzman akademisyenlerden danışmanlık alınması,
- 5 pilot okulun belirlenmesi ve bu okullarda farklı branşlarda 5-6 öğretmen görevlendirilerek pilot uygulamaya geçilmesi,
- “Bilgi Temelli Hayat Problemi” doğrultusunda hazırlanacak 9 haftalık ders programının uygulanması,
- Pilot uygulama dersleri sonrası öğrenci ve öğretmen süreç kazanımları, deneyimleri ve ürünlerini paylaşmak üzere İstanbul genelinde bir etkinlik düzenlenmesine karar verilmiştir [95].

Harezmi Eğitim Modeli, A12 düzeyinde tüm okul türlerini ve sınıfları kapsayan geniş bir öğrenci kitlesine yönelik olarak uygulanmaktadır. İstanbul ilinden uygulanmaya başlayan model sonrasında 58 ilde uygulanmaya başlanmıştır. 1. pilot uygulama dönemi olan 2016-2017 öğretim yılında 5, 2017-2018 öğretim yılında 50, 2017-2018 öğretim yılında 50, 2018-2019 öğretim yılında 196, 2019-2020 öğretim yılında 411 okulda uygulanmıştır [96]. Öğretmenlerin akademik olarak yüzyüze ve uzaktan hizmetiçi eğitimlerle desteklendiği uygulama kapsamında yüzyüze 7088 öğretmene yüzyüze, 2328 öğretmene de uzaktan eğitim yoluyla eğitim verilmiştir.

2.1.10 Modelin Uygulama Süreci

HEM, öğrencilerin günlük, gerçek yaşam problemlerini tanımlayıp, çözmek üzere algoritma tasarlayan, adım adım bu sorunun nasıl çözüleceğini belirleyip programlamanın gücünden faydalanarak yenilikçi fikirler üretme sürecidir [97]. HEM uygulaması en az 3 disiplinden (Teknoloji, Fen ve Matematik, Sosyal Bilimler-Sanat ve Spor) birer öğretmenin birlikte planlama, uygulama ve değerlendirme yaptığı bir süreçtir. Haftalık 4 saat şeklinde uygulanan programdan önce öğretmenler toplantı yaparak günlük eğitimi planlar, 2 saat Harezmi uygulaması yapılır ve ders sonunda öğretmenler süreci değerlendirdikleri bir toplantı daha gerçekleştirirler.

Bilimsel araştırma yöntemleri kılavuzluğunda yürütülen süreçte öğrenci ve öğretmenlerin birlikteliğinde “Hayatın İçinden Sorunlar – HİS” belirlenir ve uygulama sürecine başlanır. HEM’in 5 zemini olan;

1. Disiplinlerarası yaklaşımın yeniden yorumlanarak farklı disiplinlerin eşdeğerde yer alması,
2. Bilgi işlemsel düşünme becerilerinin problem çözme sürecinde kullanımı,
3. Programlama becerilerinin öğrenme ortamlarında etkin kullanımı,
4. Robotik ve oyun tasarımının öğrenme ortamlarında etkin kullanımı,
5. Problem çözme sürecinde yaşam becerilerinin desteklenmesi zeminleri HEM’in merkezinde yer almaktadır.

Öğrenciler ve öğretmenler bu 5 zeminin esas alındığı dersler ile uygulama süreci boyunca HİS’lere çözüm yolu aramaktadırlar.

2.9 Literatür Analizi

Bu bölümde kategoriler halinde; duygu analizi ve metin madenciliği hakkında ortaya koyulan algoritmalar, eğitim alanında duygu analizi çalışmaları, konu analizi çalışmaları ve HEM özelinde yapılan çalışmalara yer verilmiştir.

Yadollahi vd. [98] duygu madenciliği alanında yapılan çalışmaları derledikleri çalışmalar incelendiğinde genel olarak kural tabanlı yaklaşımların ve denetimli öğrenme algoritmalarının kullanıldığı (SVM, K- En Yakın Komşu vb.) görülmüştür. Fikir madenciliği konusunda zengin

bir araştırma grubunun varlığını fakat metinden duygu madenciliği konusunun halen başlangıç aşamasında bir konu olarak karşımızda durduğunu belirtmişlerdir.

Duygu analizi ya da madenciliği hakkında ortaya koyulan algoritmalar;

Mahadevan ve Arock [99] yaptıkları çalışmada filmlere verilen yıldız sayısını tahmin eden bir yaklaşım geliştirmişlerdir. Yaklaşımına göre öncelikle film yorumları veri setinden bir yorum alınmakta ve ön işlemden geçirilmektedir, sonrasında ise 1-5 yıldız arasında bir kategoriyle etiketlenmektedir. Yorumlarla LDA modeli eğitilmektedir. Sosyal medya metinleri için kural tabanlı bir model olan Vader [100] kullanılarak da duygu etiketi verilen yorumlar LDA modelinin türettiği konularla bağlanmaktadır. Bu etiketlemeler sonucunda regresyon analizi yapılan yorumlar doğru yıldız etiketlerine atanmaktadır. Deneysel çalışmalar sonucunda ürettikleri yaklaşımın aynı veri setleri üzerinde çalışılmış diğer yaklaşımlardan daha iyi performans sergilediğini göstermişlerdir.

Li vd. [101] sosyal duygu analizi üzerine yaptıkları çalışmada bir bayezyen modeli (WMCM) ve bir sinir ağı (PCNN) olmak üzere geliştirdikleri 2 modeli Çince bağlamında daha önce etiketlenmiş 2 veri setinde (SemEval (nefret, iğrenme, korku, mutluluk, üzgünlük ve şaşırma) ve Sinanews (dokunma, empati, sıkılma, nefret eğlence, hüznün, sürpriz ve sıcaklık) kategorilendirme başarımını ölçümlemişlerdir. Araştırmada yapay sinir ağı temelli PCNN modelinin daha yüksek başarıya ulaştığı görülmüştür.

Liang vd. [102] yaptıkları çalışmada kısa metinler için duygu analizinde kullanılması amacıyla Evrensel Duygu Modeli (Universal Affective Model – UAM) modelini önermişlerdir. Çalışmalarında klasik metin sınıflandırma modelinden farklı olarak, yapısal olarak konu düzeyinde ve terim düzeyinde alt modellerden oluşan UAM'ın sosyal medyada okuyucuların bakış açısıyla sosyal duyguları algıladığını öngörmüşlerdir. UAM modelinin diğer modellerden farkı kısa metinlerin doğasından kaynaklanan seyreklik olgusunu denetimli öğrenme ve kelime tabanlı duygu sözlüğü ile aşmaktadır.

Rao vd. [103] kısa metinlerden konu konu düzeyinde maksimum entropi modelini (topic based maximum entropy model – TME) önermişlerdir. TME, gizli konuları, çoklu duygu etiketlerini ve çok sayıda okuyucu tarafından ortaklaşa puanlanan puanları modelleyerek konu düzeyinde özellikler üretmektedir. Gerçek verilerle yapılan deneysel çalışma TME modelin başarımını doğrular nitelikte sonuçlar üretmiştir.

Li vd. [104] yaptıkları çalışmada konu modelleme ile duygu analizini birleştirerek Sentiment-LDA ve Dependency-Sentiment-LDA adında iki algoritma önermişlerdir. Algoritmalar olumlu-olumsuz şeklinde 2 konunun oluşturulduğu farklı veri setleri üzerinde denenmiş ve ortalama 60%~65% aralığında başarımlar göstermişlerdir. Yaptıkları çalışma konu modelleme ve duygu analizinin bir araya getirildiği ilk çalışmadır.

Onan vd. [105] LDA ve diğer makine öğrenmesi algoritmalarının duygu veri setlerindeki başarımlarını inceledikleri çalışmalarında olumlu-olumsuz yada nötr olarak sınıflandırılan 4 veri setinde LDA ile boyutu azaltılan veri setlerinin topluluk (*ensemble*) öğrenme (*Bagging*, *AdaBoost*, *Random Subspace* vb.) ve diğer sınıflandırma algoritmalarının (Naive Bayes, SVM, KNN vb.) başarımlarını karşılaştırmışlardır. Sınıflandırma algoritmaları ile bağlantılı olarak topluluk öğrenme algoritmalarının deneysel analizde kullanılan veri kümeleri için önemli performans iyileştirmeleri sağladığı sonucuna ulaşmışlardır.

Eğitim alanında duygu analizi çalışmaları;

Lopez vd. [106] yaptıkları çalışmada bir e-öğrenme ortamında öğrencilerin içeriklere yaptıkları yorumları olumlu, olumsuz ve nötr olarak sınıflandırmışlardır. Toplam 1378 adet yorumla gerçekleştirilen çalışmada yorumlar gözlem yoluyla işaretlenmiş, sonrasında ise etiketlenen yorumlar sınıflandırılmıştır. Sınıflandırma başarımları 81.54% olarak ölçümlenmiştir.

de Paula Santos vd. [107] yaptıkları çalışmada bir yükseköğretim kurumunda öğrencilerin bakış açılarından olumlu veya olumsuz öğretim uygulamalarını belirlemek için duygu analizini kullanan yeni bir kurumsal öğretim değerlendirme yöntemi sunmuşlardır.

Chauhan vd. [108] öğrenci görüşlerini kullanarak öğretim ve içerik kalitesini ölçümledikleri çalışmalarında 1000 öğrenci görüşünü olumlu, olumsuz ve nötr olarak etiketlemişler, sonrasında ise başarımlar ölçümü yapmışlardır. İnsanlar tarafından etiketlenmiş veri ve otomatik işaretleme arasında 0.81 oranında bir başarımlar elde etmişlerdir.

Göker ve Tekedere [109] Fatih projesine yönelik görüşleri değerlendirmek amacıyla bir forum sitesi üzerinden 444 yorumu analiz etmişlerdir. Çalışma kapsamında yorumlar pozitif ve negatif olarak işaretlenerek farklı algoritmaların başarımlarını karşılaştırmışlardır. Çalışma kapsamında ön işlemden geçen metin üzerinde Naive Bayes, k-NN, J48, SMO ve RBF Network

sınıflandırma algoritmalarının başarımlarını karşılaştırdıkları çalışmada 84.90% başarımla RBF Network'ün en başarılı sınıflandırıcı olduğu sonucuna ulaşmışlardır.

Konu analizi çalışmaları;

Gürcan [110] doktora tezi kapsamında yaptığı çalışmada; yazılım endüstrisinde ortaya çıkan güncel talep ve eğilimleri belirlemek amacıyla yazılım odaklı iş ilanları (Stack Overflow iş platformu Internet sitesi verileri) üzerinde LDA yöntemiyle olasılıksal konu modellemeye dayalı bir anlamsal analiz gerçekleştirilmiştir. Ocak-Kasım 2016 tarihleri arasındaki 11 aylık dönemde yayınlanan İngilizce iş ilanları ile oluşturulan 4128 belgeden oluşan derlemi 52 konuya indirgemıştır. İngilizce metinlerle sınırlı olan çalışmasında deneysel veri setindeki metinlerin miktarı ve çeşitliliğinin keşfedilen konuların anlam düzeyini yansıtması açısından analizin başarısını yakından etkilediğini değerlendirmiştir. Bir başka değindiği nokta da optimal parametrelerin seçiminin deneysel analiz gerektirmesi ve en anlamlı düzeyde konu sayısının (K) belirlenmesi için çok sayıda test amaçlı analiz gerektirdiğidir.

Caner [111] yüksek lisans tezi kapsamında yaptığı çalışmada; otel yorumlarını konu ve duygu analizi yöntemlerini kullanarak yorumlamıştır. Derleminde tripadvisor.com sitesinden elde edilen, İstanbul, Roma, Barselona, New York, Atina, Berlin ve Paris'te bulunan 142 otel için yapılan 36.241 adet yorum bulunmaktadır. Çalışma İngilizce dili üzerinde yapıldığından farklı dillerdeki yorumları modellememiştir.

Harezmi Eğitim Modeli üzerine yapılan çalışmalar;

Koçoğlu [112] yaptığı çalışmada HEM'in alan uzmanlarının görüşleri doğrultusunda incelenmesini araştırma, nitel araştırma yaklaşımına dayalı betimsel tarama yöntemiyle gerçekleştirilmiştir. Araştırmanın çalışma grubunu, İnönü, Fırat, Yıldız Teknik, Akdeniz, Gazi, Atatürk, Gazi Osman Paşa, ve İbrahim Çeçen Üniversitelerine bağlı Eğitim Fakültelerinde yer alan Bilgisayar Öğretmenliği (BÖTE), Sosyal Bilgiler Eğitimi, PDR Ana Bilim Dallarında görev yapan, HEM ve bu modelin Türkiye'deki uygulama düzeyine ilişkin gerekli yeterliliklere yeterliliğe sahip 50 öğretim elemanı oluşturmaktadır. Çalışma kapsamında HEM'in en dikkat çekici özelliği, eğitsel faaliyetlere katkısı ve uygulama sürecinde öğretmenlere düşen sorumluluklar altında 3 soruya cevap aranmıştır.

Ceylan vd. [113] yaptıkları araştırmada İstanbul ili kapsamında 12 ilçe ve 25 okuldan toplam 482 öğrenciye “Benim için Harezmi Eğitim Modeli gibidir/ benzemektedir çünkü.....” cümlesini içeren formlar dağıtmışlar ve cevapları içerik analizi yöntemiyle çözümlemişlerdir. Çözümleme aşamasında Microsoft Excel programı kullanılmıştır. Öğrencilerin sorulara verdikleri cevaplar kategorilendirilerek “Öğrencilerin Harezmi Eğitim Modeli’ne yönelik ürettikleri metaforlar nelerdir?” ve “ortak özellikleri bakımından hangi kategoriler altında toplanmaktadır? Sorularına cevap aranmıştır.

Yavuz vd. [114] yaptıkları çalışmada HEM’in özel yetenekli öğrencilerin yaratıcılık becerilerine etkilerini incelemişlerdir. 10-12 yaş aralığında 26 öğrenci ile gerçekleştirilen çalışmada öğrenciler deney ve kontrol gruplarına dağıtılarak deneysel bir çalışma yapılmıştır. Torrance Yaratıcı Düşünme Testi öğrencilere ön test ve son test olarak uygulanmıştır. Deney grubundaki öğrencilerin yaratıcılık becerileri alt boyutları (akıcılık, orijinallik, başlıkların soyutluluğu, kapamaya karşı direnç, zenginleştirme) ve toplam yaratıcılık puanlarında istatistiki olarak anlamlı bir artış olduğu görülmüştür.

Öğrencilere S3 sorusunun sorulduğu nitel araştırmalar;

Aydemir vd. [115] yaptıkları nitel araştırmada kademeli sorumluluk aktarımı modeli kullanılarak öğretilen internette arama-anlama stratejilerine yönelik ilkökul dördüncü sınıf öğrencilerinin yansıtıcı düşünme eylemlerinin incelemeyi amaçlamışlardır. Araştırma gruplarında 10 ilkökul 4. sınıf öğrencisi bulunmaktadır. Öğrencilere yansıtıcı günlük şeklinde (1) ”Ders dinledikten sonra kendini nasıl hissettin, derste neler yaptın ve öğrendin? (2) Becerilerle ve stratejilerle ilgili özdeğerlendirmeni yapar mısın? (3) Rubrikte yer alan soruları cevaplandırırken neler yaptın? Aşama aşama anlatır mısın? (4) Soruları cevaplarken hangi stratejileri kullandın? soruları sorulmuştur. Ortaya çıkan veriler 4 uzman görüşüyle analiz edilmiştir.

Taş [116] yaptığı araştırmada karikatür destekli fen öğretiminin ilköğretim düzeyindeki öğrencilerin akademik başarı ve bilginin kalıcılığına etkisi ve uygulama hakkında öğrencilerin görüşlerini incelemiştir. Çalışma grubunda 21 uygulama ve 20 kontrol grubu öğrencisi olmak üzere 41 öğrenci bulunmaktadır. Öğrenci görüşlerini ölçmek için araştırmacı tarafından hazırlanan yarı yapılandırılmış açık uçlu 6 soruya yer verilmiştir. Öğrencilere sorulan S3 sorusuna verilen cevaplar olumlu ve olumsuz olarak etiketlenmiştir. Olumlu 85.71% ve olumsuz 14.28% görüş bulunduğu görülmüştür.

Dilmaç ve Dilmaç [117] ortaöğretim öğrencilerinin görsel sanatlar dersine ait otantik değerlendirme yaklaşımlarına yönelik görüşlerini inceledikleri nitel çalışmalarının çalışma grubunu 11 kız ve 9 erkek toplam 20 kişi öğrenci oluşturmaktadır. Araştırma kapsamında yarı yapılandırılmış bir görüşme formu kullanılmıştır. Öğrencilerin verdikleri cevaplar uzman değerlendirmesiyle kodlanarak sonuçlar ortaya koyulmuştur.

Dere vd. [118] Ankara’da özel bir okulun ilköğretim bölümünden toplam 182 öğrencinin yer aldığı çalışmalarında öğrencilerin Edmodo platformuna ilişkin görüşlerini araştırmışlardır. Nicel ve nitel veriler kullanılan çalışmalarında nitel değerlendirme enstrümanında “Edmodo’yu kullanırken neler hissettin (Başarılı, güvenli, özel, sosyal, mutlu, kaygılı, özgür vb.)? Ayrıntılarıyla açıklar mısın?” sorusu yöneltilmiştir. Nitel veriler üzerinde içerik analizi yapılarak kategorize edilmiştir.

Türkçe Metin Ön İşleme Alanında Yapılan Çalışmalar

Hickman vd. [119] yaptıkları derleme çalışmasında farklı dillerde ön işleme metotlarının kullanıldığı metin madenciliği çalışmalarını derlemiş ve önerilerde bulunmuşlardır. Türkçe metin ön işleme yöntemleri harf eşleme, alfabetik olmayan karakterlerin kaldırılması (karakter kodlamasının düzeltilmesi), dolgu kelimelerinin kaldırılması, gövdeleme gibi alanlarda kullanılmakta ve dolgu kelimelerinin kaldırılması haricinde diğer yöntemler başarıyı yükseltmektedir. Bir diğer değerlendirdikleri nokta da çalışmalarda kullanılan ön işleme yöntemlerinin araştırmacılar tarafından paylaşılmasıdır. İnceledikleri çalışmaların 3 adedinde araştırmacılar ön işleme yöntemlerini paylaşmamış, 3 araştırmada ön işleme yöntemlerinin tamamı değil belli bir kısmına atıf yapılmış, 3 araştırmada araştırmacılar gövdeleme genel olarak kullanıldığı için gövdeleme yaptıklarını, 2 araştırmada ise standart ön işleme tekniklerini kullandıklarını paylaşmışlardır.

Tunalı ve Bilgin [120] yaptıkları çalışmada Türkçe metin ön işleme amacıyla kullanılması önerilen bir araç geliştirmişlerdir. DeASCIIfication, terim filtreleme, gövdeleme gibi temel özellikleri içeren araç metin dosyaları üzerinde işlem yapabilen bir yapıya sahiptir.

Salih ve Günel [121] yaptıkları çalışmada Türkçe metin ön işlemenin doküman sınıflandırma üzerinde etkilerini araştırmışlardır. Jetonlama, büyük-küçük karakter eşleme, gövdeleme ve dolgu kelimelerini kaldırma tekniklerinin etkilerini araştırdıkları çalışmalarında ön işleme yöntemlerinin karışımlarının sınıflama performansını artırdığını bulgulamışlardır. Bir başka

değerlendirmeleri de en iyi özetleme performansına ulaşmak için ise metnin bulunduğu bağlamın ve kullanılan değerlendirme yönteminin önemli olduğunu bulgulamışlardır.



3 MALZEME VE YÖNTEM

Daha önce de bahsedildiği üzere veri madenciliği; modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı bir yaklaşımdır [12]. Bu bağlamda tezin malzeme ve yöntem kısmında öncelikle metnin bilgi haline getirilmesi için kullanılan ve geliştirilen uygulamalara yer verilecek, sonrasında ise metin madenciliği süreci adımları Gaikwad vd. [28] tarafından CRISP-DM modeli kullanılarak oluşturulan Metin Madenciliği Süreç Modeli izlenerek işlenecektir.

3.1 Kullanılan ve Geliştirilen Uygulamalar

Araştırma kapsamında verinin toplanması ve işlenmesini kolaylaştırmak amacıyla farklı uygulamalar kullanılmış ve geliştirilmiştir.

3.1.1 Veri Yönetim Sistemi

HEM uygulama sürecinde uygulama okullarından gelen verileri sistematik bir şekilde toplamak, veri güvenliğini sağlamak, sürecin değerlendirilmesine katkı sunmak ve toplanan verileri araştırmacıların kullanımına açmak amacıyla araştırmacı tarafından Çimen ve Dede [122]'de anlatılan FABSİS sisteminin baz alındığı bir Veri Yönetim Sistemi (VYS)⁶ oluşturulmuştur. FABSİS sisteminin temel amaçları;

- Yazılım kodu değiştirilmeden formlar oluşturmak ve doldurulmak üzere kurumlara göndermek,
- Kurum verilerinin kurumsal veri kartları desteği ile kolayca incelenmesini sağlamak,
- Kişisel bilgisayarlarda saklanan eski verilerin kolaylıkla sisteme entegre edilebilmesini,
- Verilerin sistematik bir biçimde sistemden alınabilmesini ve
- Kurumsal verilerin tek bir platformda yönetilmesini sağlamaktır.

⁶ <http://istanbul.meb.gov.tr/harezmi/>

Program, PHP programlama dili kullanılarak geliştirilmiştir. Programın oluşturulmasında Model-Denetçi-Görünüm (Model-Controller-View – MVC) tasarım modeli kullanılmıştır. MVC'nin kullanılması, bir uygulamanın geliştirilmesini ve bakımını kolaylaştırır çünkü; uygulamanın “görünümü”, veri yapıları ve iş mantığını değiştirmeden büyük ölçüde değiştirilebilir [123]. Uygulama PHP programlama dili MVC çerçeveleri arasından kolaylık, hız, güvenlik, gelişmiş dokümantasyon gibi özelliklerle öne çıkan[124,125] Codeigniter Framework [126] seçilerek geliştirilmiştir.

Uygulama kapsamında veriyi saklamak ve transfer etmek amacıyla JavaScript Nesne Gösterimi (Javascript Object Notation – JSON) veri alışveriş yapısı kullanılmıştır. Bu tercihin sebebi JSON yapısının MySQL, PHP ve Javascript tarafından kolayca çözülebilir ve kodlanabilir olmasından dolayı sistemin arkaplanı ve arayüzü arasında iletişimi kolaylaştırmaktır. Ayrıca MySQL veritabanı sistemi JSON veri formatını desteklemektedir ve depolama, doğrulama, veri içinde arama gibi işlemler için fonksiyonlar barındırmaktadır [127].

Sistemin ön yüzünde ise sistemin modern Internet tarayıcıların çoğunda hatasız bir şekilde çalışabilmesini temin etmek amacıyla JQuery Javascript Framework [128] kullanılmıştır. Yine sistemin mobil cihazlar gibi günümüzde Internet'e erişmek için sıkça kullanılan cihazlarda hatasız bir şekilde açılmasını sağlamak amacıyla da Bootstrap [129] kullanılmıştır.

Şekil 3.1: VYS form oluşturma ekranı örnek görüntüsü.

VYS'nin form oluşturma ekranına ait örnek bir görüntü Şekil 3.1'de görülmektedir. Sistemde bulunan formlar proje yöneticileri tarafından oluşturulabilmekte, kimin dolduracağı, kimin bu verileri inceleyebileceği ve raporlayabileceği belirlenebilmektedir. Formun sistemde hangi

tarihler arasında görünür olacağı ve doldurma periyotları da ayarlanabilecek değişkenler arasındadır.

3.1.2 Verinin İşlenmesi Süreci

Veri madenciliği ve makine öğrenmesi alanında en çok kullanılan programlama dilleri sırasıyla Python, SQL ve R dilleridir [130]. Python dilinin bu denli fazla kullanılması ile paket geliştiricilerinin de Python programlama diline ilgisi neticesinde Python paket indeksi (pypi.org) topluluğuna kayıtlı 524.631 kişi ve halihazırda kullanıma hazır 317.147 paket olduğu [131] değerlendirildiğinde Python dilinin güçlü bir altyapı sunduğu değerlendirilebilir. Bu sebeple veri işleme sürecinde tüm uygulamalar Python programlama diliyle geliştirilmiştir. Kullanılan uygulamalar ile ilgili bilgiler aşağıda belirtilmektedir.

- Python [132]: Güçlü paket ekosistemi nedeniyle veri madenciliği ve makine öğrenmesi uygulamaları bağlamında en çok tercih edilen programlama dillerinden birisidir.
- Anaconda [133]: Paket yönetimi ve dağıtımını basitleştirmeyi amaçlayan, bilimsel hesaplama (veri madenciliği ve makine öğrenmesi, büyük veri vb.) için oluşturulmuş Python ve R programlama dillerinin kullanılabildiği bir çerçevedir.
- PyCharm: Python ve R gibi programlama dilleriyle geliştirme ve bilimsel hesaplama yapmayı kolaylaştırmak amacıyla JetBrains firması tarafından geliştirilen bir yazılım geliştirme ortamıdır. Öğrenciler için ücretsiz kullanım lisansı sunmaktadır.
- Zemberek-NLP: Akın ve Akın [37] tarafından geliştirilmiş ve açık kaynak olarak paylaşılmış Türkçe NLP aracıdır. Java diliyle yazıldığından Python üzerinde kullanılması için bir iletişim katmanına ihtiyaç duyulmaktadır.
- Jpype [134]: Python üzerinden, Java uygulamalarına tam erişim için geliştirilen bir pakettir. Zemberek-NLP ile Python arasında iletişim kurmak amacıyla kullanılmıştır.
- Pandas: Pandas paketi Python programlama dili ortamında kullanılmak üzere Reback vd. [135] tarafından geliştirilen bir veri analitiği kütüphanesidir. Pandas

kütüphanesi proje kapsamında; Dataframe veri yapısı sayesinde CSV, MySQL gibi kaynaklardan alınan verilerin tutulması ve işlenmesi kapsamında kullanılmıştır.

- PyMySQL [136]: Python programlama dili için bir MySQL bağlantı sağlayıcısıdır.
- Pickle [137]: İşlenen verilerin bilgisayar ortamında saklanması ve geri çağırılması işlemleri için kullanılmaktadır.
- Tomotopy[138]: Konu modelleme algoritmalarından LDA, sLDA, LLDA, PLLDA, CTM, DMR, PA, DTM vb. algoritmalarla birlikte C_{uci} , C_{u_mass} , C_{nmpi} ve C_v gibi değerlendirme yöntemlerini kapsayan Python programlama dili için hazırlanmış bir konu modelleme paketidir.

Kullanılan paketler dışında, araştırma kapsamında *TurkishTopicModel* isminde bir Python kütüphanesi oluşturularak;

- Başka dillerde yazılmış (Java, C/C++) uygulamalar ve Web servisleri ile Python ortamı arasında bir iletişim katmanı oluşturmak,
- Metin işleme sürecini sistematik bir biçimde yürütmek,
- Bu araştırmadan sonra yapılacak Türkçe NLP uygulamaları için de kolaylık sağlamak amaçlanmıştır.

3.2 Metin Madenciliği Süreci

3.1.3 Çalışmanın Amacı

Verinin işlenmesi nitel araştırma desenlerinde ortak olarak takip edilen bir süreçtir ve “Deşifre”, “Kodlama”, “Araştırmacı Günlüğü Yazımı” ve “Kategori Oluşturma” basamaklığına ev sahipliğin yapmaktadır [139]. Deşifre; görüşme, gözlem ve ses kayıtları veya alan notlarını metne dönüştürme işidir. Deşifre işleminden sonra gelen aşamalar nitel araştırmalar; alan

yazında görüldüğü üzere küçük gruplarla yapıldığında elle veya MS Excel programıyla, büyük gruplarda yapıldığında ise Nvivo⁷ programıyla yapıldığı görülmüştür [112,113,115–118].

Büyük çalışma gruplarında gözle kodlama ve kategori oluşturmak oldukça zor bir süreçtir ve metinler üzerinde otomatik kategori oluşturan algoritmalar bu analiz işini kolaylaştırıcı bir etkiye sahip olabilir. Bu sebeple çalışmanın amacı HEM veri yönetim sisteminde bulunan formlar kullanılarak A12 bağlamından alınan metin verilerinin kategorilendirilmesini sağlayacak modeller oluşturmak olarak belirlenmiştir.

3.1.4 Araştırmanın Kapsamı ve Sınırlıkları

Araştırma kapsamında kullanılan veri seti 2018-2019 yılında İstanbul ili genelinde Harezmi Eğitim Modelini uygulayan 208 okuldan 4195 öğrenciden alınan verilerden oluşmaktadır.

Dolgu kelimeleri ve belgeleri derlemlerden çıkarılmıştır. Tablo 3.7’de paylaşıldığı üzere derlemlerden 2 kelimeden daha kısa ve 10 kereden fazla tekrar eden yorumlar temizlenmiştir. Ayrıca bütün konularda kendisine ağırlık bulan kelimeler derlemlerden çıkarılmıştır. Bu durum tüm belgelerin derlemlerde temsil edilememesi yönünden araştırmanın sınırlılığıdır.

3.1.5 Verinin Doğası ve Erişilebilirliği

Analizlerde kullanılacak veri seti HEM’in 2018-2019 yılı uygulaması süresince doldurulan ve VYS’den elde edilen 3N formlarıdır ve “Ne yaptın?” (S1), “Ne Öğrendin?” (S2) ve “Ne Hissettin?” (S3) olmak üzere 3 soru içermektedirler. HEM uygulamasının yapıldığı okul ve o okullarda oluşturulan gruplar, öğrenci ve öğretmen sayılarını içeren veriler Tablo 3.1’de paylaşılmıştır.

⁷ <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>

Tablo 3.1: HEM 2018-2019 uygulamasına ait okul, öğrenci ve öğretmen sayıları.

Kurum Türü	N	Öğrenci	Öğretmen
Anaokulu	9	186	36
İlkokul	49	1081	238
Ortaokul	68	1347	347
Özel Türk Ortaokulu	4	26	21
İmam Hatip Ortaokulu	21	433	102
Anadolu Lisesi	17	367	83
Anadolu Meslek Programı	23	463	113
İmam Hatip Lisesi	4	77	21
Anadolu İmam Hatip Lisesi	5	101	24
Özel Anadolu Lisesi	2	23	10
Fen Lisesi	3	36	11
Üstün veya Özel Yetenekliler	3	55	16
TOPLAM	208	4195	1022

Tablo 3.1’de bulunan veriler ışığında özet olarak sistemde 208 okuldan toplam 4195 öğrenci ve 1022 öğretmen bulunmaktadır. Sistemde anaokulu düzeyinde 9 okuldan 186 öğrenci ve 36 öğretmen, ilkokul düzeyinde 49 okuldan 1081 öğrenci ve 238 öğretmen, ortaokul düzeyinde 93 okuldan, 1806 öğrenci ve 470 öğretmen, lise düzeyinde 54 okuldan toplam 1067 öğrenci ve 262 öğretmen, üstün ve özel yetenekli ise 3 adet okuldan toplam 55 öğrenci ile 16 öğretmenin doldurduğu form verisi bulunmaktadır.

Uygulama 32 hafta sürmüştür fakat form okul uygulama planlamalarının 32 (± 2) hafta sürebileceği düşünülerek 34 hafta üzerinden tanımlanmıştır. Uygulama süreci boyunca doldurulan 3N formlarının okul türüne ve haftalara göre dağılımı Tablo 3.2’de paylaşılmıştır.

Tablo 3.2: 3N formu verilerinin uygulama haftası ve okul türüne göre dağılımı.

Hafta	Anaokulu	İlkokul	Ortaokul	Anadolu İmam Hatip Lisesi	İmam Hatip Ortaokulu	Anadolu Lisesi	Anadolu Meslek Lisesi	Fen Lisesi	İmam Hatip Lisesi	Özel Anadolu Lisesi	Üstün veya Özel Yetenekliler	Genel Toplam	Öğrenci/Form (%)
1	146	942	1188	88	371	286	357	31	70	20	23	3522	83,96
2	134	892	1221	91	375	285	312	14	69	13	30	3436	81,91
3	139	907	1202	90	389	285	369	14	68	8	28	3499	83,41
4	98	790	1134	42	331	238	317	13	71	21	48	3103	73,97
5	147	870	1196	91	388	285	368	14	68	19	45	3491	83,22
6	135	874	1117	69	389	267	349	6	71	15	46	3338	79,57
7	42	373	536	12	185	68	165		32	22	30	1465	34,92
8	149	792	1144	89	389	216	325	25	54	3	41	3227	76,92
9	135	879	1181	84	366	266	321	14	68	5	49	3368	80,29
10	164	871	1161	75	395	267	287	15	71	4	25	3335	79,50
11	154	799	1121	87	398	276	341	13	55		42	3286	78,33
12	153	863	1053	87	395	246	341	14	35		45	3232	77,04
13	151	737	922	43	334	164	268	23	36		31	2709	64,58
14	110	723	973	63	381	261	308	21	67	13	39	2959	70,54
15	49	257	337	36	114	66	116	27			13	1015	24,20
16	110	700	973	59	358	235	253		42		31	2761	65,82
17	129	754	1054	84	365	239	266	12	44		40	2987	71,20
18	138	752	1091	89	367	260	230	10	42		16	2995	71,39
19	125	696	965	37	334	207	201	14	36		30	2645	63,05
20	125	729	1081	67	374	249	280	13	41	12	28	2999	71,49
21	144	765	1078	92	374	217	241	13	41	2	29	2996	71,42
22	150	699	1004	59	374	193	197	12	39	13	23	2763	65,86
23	136	717	1033	77	322	205	204	14	36		13	2757	65,72
24	95	616	900	18	337	136	168	14	42	13	24	2363	56,33
25	137	749	986	56	350	192	200	14	36		21	2741	65,34
26	144	684	947	67	343	217	195	13	44		22	2676	63,79
27	132	487	720	30	284	147	179	13	29			2021	48,18
28	66	554	621	69	267	130	168	7	7		26	1915	45,65
29	126	531	803	76	282	154	161	7	35		32	2207	52,61
30	136	648	791	76	260	161	148	11	41		17	2289	54,56
31	112	487	574	48	241	150	115		36		33	1796	42,81
32	46	217	254	32	45	131	43		2		13	783	18,67
33			12	3	13						13	41	0,98
34			4	6			10				13	33	0,79
Toplam	3957	22354	30377	2092	10490	6699	7803	411	1428	183	959	86753	64,63

Tablo 3.2 incelendiğinde farklı okul türlerinden uygulama öğrencilerinin ortalama 64,63%'ünün 3N formuna yanıt verdiği ve derlemde 86753 form verisi bulunduğu görülmektedir. İlk haftalarda uygulama öğrencilerinin 83,96 gibi yüksek bir yoğunluğu formlara cevap verirken, bu oran son haftalara doğru 45-50% bandına doğru gerilemiştir.

Formlar herhangi bir alan doldurulduğunda veri kaybına sebebiyet vermemek için otomatik kaydedilmiştir. Kullanıcı, formu doldurmayı yarıda bırakabilir ya da bir alana hiç veri girişi yapmayabilir. Bu sebeple derlem büyüklüğünü belirlemek amacıyla veri içeren belgeler (d), derlemdeki kelime sayısı (W) ve belgelerin içerdiği ortalama kelime sayılarının ($\bar{x}(d,w)$) analizi yapılarak Tablo 3.3'te paylaşılmıştır.

Tablo 3.3: Derlemde bulunan belge, kelime ve belge başına kelime ortalamasını gösterir tablo.

Soru	d	W	$\bar{x}(d,w)$
S1	85986	37155	6.82
S2	84703	38722	5.66
S3	84848	22382	4.33

Tablo 3.3 incelendiğinde S1 sorusuna verilen toplam 85986 cevap bulunduğu, toplam 37155 kelime kullanıldığı ve belge başına ortalama 6,82 kelime bulunduğu görülmüştür. S2 sorusu için toplam 84703 cevap bulunduğu, toplam 38722 kelime kullanıldığı ve belge başına ortalama 5.66 kelime bulunduğu ve S3 sorusu için ise 84848 cevap, 22382 kelime ve belge başına ortalama 4.33 kelime bulunduğu görülmüştür.

3.1.6 Veriyi Hazırlama

Metin belgeleri için veri hazırlama aşaması “Metin Ön İşleme” olarak da adlandırılabilir. Bu aşamada Bölüm 2.1.4'te anlatıldığı üzere işlem adımları, metnin bulunduğu bağlam, kodlama standardı, temizlik durumu vb. değişkenlere bağlı olarak değişkenlik gösterir. Modelleme ve değerlendirme aşamalarıyla bağlantılı olarak da veriyi modelleme için hazırlama aşaması tekrar tekrar yapılabilir. Deneysel süreçte veri modelleme için tekrar tekrar hazırlansa da bu başlık altında sürecin özetlenmesi ve edinilen deneyimin paylaşılması amaçlanmıştır.

Veriyi hazırlama sürecinde aşağıdaki sorulara cevap aranmıştır.

1. Veri ön işlem yöntemlerinin derleme etkileri nelerdir?

2. Veriyi en az hatayla ön işlemden geçirmek için hangi yöntemler hangi sırayla uygulanmalıdır?

Bu araştırma soruları ışığında veri ön işleme yöntemleri ile ilgili olarak farklı yöntemler veri seti üzerinde denenmiş ve bir yol haritası ortaya koyulmaya çalışılmıştır.

3.2.1.1 Kaynaktan Veriyi Almak ve Ön İşlemden Geçirmek

Metin verisi belgelerde, Internet üzerinde bir kaynakta, veri tabanında, ses ve video kaydı gibi birçok farklı ortamda bulunabilir. Araştırma kapsamında kullanılan MySQL veri tabanı ortamında JSON formatında saklanmaktadır. JSON formatında belgeler transfer edilirken ASCII formatı kullanıldığından, ASCII dışındaki karakterler karakter kodu şeklinde etiketlenir. Örneğin “ı” karakterinin karakter kodu “\u0131” şeklindedir. Verinin kaynaktan alınması, Unicode karakter dönüşümü ve ön temizlik işlemleri sonrasında durumları Tablo 3.4’te paylaşılmıştır.

Tablo 3.4: Unicode dönüşümü ve ön temizlik işlemlerinde metin.

	S1	S2	S3
Sorgu sonrası	“Oyun tasar\u0131m\u0131 yapt\u0131k.”	“\u015eehirlerimiz hakk\u0131nda bilgi.”	“İyi.”
Unicode karakter dönüşümü	“Oyun tasarımı yaptık.”	“Şehirlerimiz hakkında bilgi.”	“İyi.”
Ön temizlik sonrası	['oyun', 'tasarımı', 'yaptık']	['şehirlerimiz', 'hakkında', 'bilgi']	['iyi']

Tablo 3.4 incelendiğinde veri tabanından gelen veride bulunan ve Unicode karakter kodları (\u0131, \u015e) karakter karşılıklarına dönüştürülmüş, karakterler eşlenmiş (büyük-küçük karakter dönüşümü) sonrasında ise metin parçalanmıştır. Zemberek-NLP; boşluk, yeni satır, kelimeler, simge içeren kelimeler, kısaltmalar, noktalama işaretleri, sayısal ifadeler, tarih, saat, e-posta, URL, emoji vb. ifadeleri yakalayabilmektedir. Bu sebeple parçalama işlemi metin temizliğiyle birlikte Zemberek-NLP Tokenization sınıfıyla yapılmıştır.

3.2.1.2 Yazım Yanlışlarını Düzeltme

Derlem içinde yazım yanlışlarını düzeltmek amacıyla normalleştirilmiş genel frekans (*ngf*) ve Jaccard Benzerliği (*Jaccard Confiencty -JaccardC*) tabanlı bir algoritma oluşturulmuştur. Öncelikle derlemde (*D*) bulunan her kelime (*w*) için *ngf* ve Jaccard değerleri hesaplanarak bir

sözlük (V) oluşturulur. Sözlükte bulunan kelimeler Zemberek-NLP aracında bulunan *SpellCorrection* (SC) sınıfına sorulur ve her kelime için öneriler alınır. Bu önerilere sorulan kelime de eklenir ve $ngfJ$ skoru en yüksek öneri hatalı yazılması muhtemel kelimeyle değiştirilir. $ngfJ$ ağırlıklandırma yöntemi aşağıdaki gibi ifade edilebilir.

1. Her $W \in D = \{w_1, \dots, w_n\}$ için her öge kendi gf değerini göstermek üzere ($V: w \rightarrow gf(w)$) genel frekans sözlüğü ($V[w \rightarrow gf(w)] = \{(w_1 \rightarrow gf(w_1)), \dots, (w_n \rightarrow gf(w_n))\}$) oluşturulur.
2. Her $w \in V$ için $SC(w)$ fonksiyonundan öneri (s) alınır ve sorulan kelime (w) de önerilere eklenerek öneri kümesi (OK) oluşturulur. Her $OK_w \in OK$ için $OK_w = \{s_1, \dots, s_n, w\} \leftarrow SC(w) + w$ olarak tanımlanır.
3. Her öneri (s) için sözlükten (V), ngf ve *JaccardC* değeri alınır ve öneri sözlüğü (OS) oluşturulur. Her $OK_w \in OK = \{OK_1, \dots, OK_V\}$ için $OS_w \in OS = \{OS_1, \dots, OS_V\}$ oluşturulur. Her OK_w için;
 - a. $JaccardC(w, s_n) = \frac{|w_1 \cap s_n|}{|w_1 \cup s_n|}$ ve $ngf(s_n) = \frac{f(s_n) - \min(f(OK))}{\max(f(OK)) - \min(f(OK))}$ ise
 $ngfJ(w, s_n) = \sqrt{ngf(w, s_n) * JaccardC(w, s_n)}$ olarak ifade edilir.
 - b. ($OS_w: w \rightarrow ngfJ(w)$) olmak üzere ($OS_w[S \rightarrow ngfJ(S)] = \{(s_1 \rightarrow ngfJ(s_1)), \dots, (s_n \rightarrow ngfJ(s_n)), : w \in V : eğer s \notin V: ngfJ(s) = \bar{x}(ngfJ(V))\}$) olarak tanımlanır.
4. $ngfJ$ değeri en yüksek öneriler alınarak değişim sözlüğü (DS) oluşturulur. $argmax OS_w: \{w \in V: ngfJ(w) < ngfJ(s): s \in S\}$ ise ($DS[w \rightarrow argmax OS_w] = \{(w_1 \rightarrow argmax OS_{w1}, \dots, (w_n \rightarrow argmax_{wn} OS_{wn} | w \in W)\}$)
5. Derlemde (D) bulunan her belgede (d) bulunan her kelimenin (w) DS 'de bulunan karşılığı ($DS(w)$) alınıp derlemde bulunan kelimenin yerine konularak ($w \leftarrow DS(w)$) düzeltme işlemi tamamlanır.

Otomatik yazım yanlış düzeltiminde $ngfJ$ kullanılmasının ve SC sınıfından gelen önerilere sorulan kelimenin de eklenmesinin sebebi öneri üretilmeyen yabancı kelimeler, hatalı öneri üretilen kelimeler vb. için derlemde de öneriler üretilmesini sağlamak ve hatalı kelimenin daha önce derlemdeki diğer belgelerde doğru yazılmış olma ihtimalini göz önünde bulundurmaktır.

3.2.1.3 Normalleştirme

Daha önce Bölüm 2.4.1.1’de anlatıldığı üzere normalleştirme birden fazla metodu (yazım yanlışı düzeltimi, değiştirme kuralları, deASCIIfying, sesli harf restorasyonu vb.) bir araya getirerek metnin standart yazıma en yakın haline getirmeye çalışıldığı bir işlem kümesidir. Örneğin “Merhabaaaa bu gün çooooook heyecanlıyım çnkü Yrn havyanat bahçsine gidicem” cümlesi aşağıdaki problem durumlarını içermektedir.

- Kelimeleri yazarken duyguyu abartmak için çok fazla harf kullanmak (merhabaaaa, çooooook)
- ASCIIfying (heyecanlıyım)
- Sesli harf restorasyonu (Yrn)
- Yazım yanlışı (çnkü, havyanat, bahçsine)
- Kelimelerin yerel şive ya da fonetik karşılığı (gidicem)

Yukarıda problem durumları açıklanan cümlelerin ITUNLPTools ve Zemberek-NLP’ye sorulması için kullanılan Python kodu ve araçların cevapları Şekil 3.2’te paylaşılmıştır.

```

from TurkishTopicModel import Preprocess
"""ITU NLP Tools"""
PreProcess.Main.ITUNLPTools.ask ("normalize"," Merhabaaaa bu gün çooooook heyecanlıyım çnkü Yrn havyanat
bahçsine gidicem")
'Merhaba bu gün çok heyecanlıyım çünkü yarın hayvanat bahçesine gideceğim'

"""Zemberek-NLP"""
PreProcess.Main.Zemberek.TurkishNormalizer ("Merhabaaaa bu gün çooooook heyecanlıyım çnkü Yrn havyanat
bahçsine gidicem")
'merhaba bu gün çok heyecanlıyım çünkü yarın hayvanat bahçesine gideceğim'

```

Şekil 3.2: Normalizasyon araçlarının örnek kullanımı

Şekil 3.2’te bulunan cevaplar incelendiğinde her iki aracın da tüm problemleri çözerek cümleleri standart forma çevirebildiği görülebilir. Normalleştirme yöntemi yazım yanlışı düzeltme dahil olmak üzere birçok probleme tek bir sorguyla çözüm bulduğundan “Veri Hazırlama” aşamasında kullanılabilecek en etkili araçlardan birisidir.

3.2.1.4 Terim Etiketleme

Araştırma kapsamında terim etiketleme amacıyla sözlük tabanlı, NER ve PMI tabanlı yaklaşımlardan yararlanılmıştır. Terim etiketleme konusunda karşılaşılabilecek problemler aşağıda belirtilmiştir.

1. Derlemde terimlerin kısaltmalarıyla birlikte uzun açılımlarının da kullanılması. Örnek: (“Hayatın İçinden Sorular”, “HİS”), (“Türkiye Büyük Millet Meclisi”, “TBMM”)
2. Terimleri oluşturan kelimelerin eksik, fazla ya da hatalı kullanılması. Örneğin: (“Harezmi Eğitim Modeli”, “Harezmi Modeli”, “Harezmi Öğretim Modeli”), (“Atatürk”, “Kemal Atatürk”, “Mustafa Kemal Atatürk”)
3. Terimleri oluşturan kelimelerde yazım yanlışı bulunması. Örneğin: “Hazerim Modeli”, “Hayat İinden Sorular”

Sözlük tabanlı yöntemin uygulanmasına tespit edilerek birleştirilecek terimlerin bulunduğu bir sözlük oluşturularak başlanır. Bu sözlük yalnızca terimleri ifade edebileceği gibi, yukarıda bahsedilen problemlerden dolayı terimlerin dönüştürüleceği kelime de sözlüğe eklenebilir. En uzun terimin kelime uzunluğunda bir pencere oluşturulur ve bu pencere belgeler üzerinde birer birer kaydırılarak terimler etiketlenir. Sözlük tabanlı terim etiketleme ile ilgili oluşturulan Python kodu Şekil 3.3’te paylaşılmıştır.

```
def tagTerms (self,area,newarea=False,tagList=[]):
    if (newarea == False): newarea = area
    if (area != newarea): self.Data.datafrm[newarea] = pd.Series ()
    tags = {}
    maxsize = 0
    for tag in tagList:
        tagtoken = self.Zemberek.TurkishTextToken (tag)
        if len (tagtoken) in tags.keys ():
            tags[len (tagtoken)].append (tagtoken)
        else:
            tags[len (tagtoken)] =[tagtoken]
    self.Data.datafrm[newarea] = self.Data.datafrm[area].apply (lambda x: self.__tagtermworker (x,tags))

def __tagtermworker (self,x,tags={}):
    if isinstance (x,str):
        tokens=self.Zemberek.TurkishTextToken (x)
    else:
        tokens=x
    match = False
    for i in range (0,len(tokens)):
        for k in range (2,max (tags.keys ())+1):
```

```

if (i-k)>=0:
    query=tokens[i-k:i]
    if query in tags[k]:
        match=True
        tokens[i-k]=“_”.join (query)
        for j in range (i-k+1,i):
            tokens[j]=False
if False in tokens:
    tokens.remove (False)
return tokens

```

Şekil 3.3: Sözlük tabanlı terim etiketleme fonksiyonları.

Python kodu Şekil 3.3’te paylaşılan Python kodunda ana fonksiyon olan tagTerms () terimleri parçalama işlemine tabi tutar, aranan terimleri uzunluklarına göre ayırır ve derlemdeki her belge için __tagtermworker() fonksiyonunu çağırarak her belge için her terim uzunluğunda pencereler kurulmasını, terimlerin belgeler içinde aranarak etiketlenmesini sağlar. Araştırma kapsamında birleştirme karakteri olarak “_” kullanılmıştır.

Terim etiketleme için takip edilebilecek bir başka yaklaşım da NER yöntemidir. Bu yöntemde belgeler bir NER aracına sorularak ürettiği sonuçlar üzerinden etiketleme işlemi yapılır. Bu yöntemde karşılaşılabilecek problem NER araçlarının kural, sözlük ve makine öğrenmesi tabanlı yaklaşımları birlikte kullanarak etiket üretmeleridir. Zemberek-NLP ve İTÜ NLP Toolkit, NER büyük-küçük harf kurallarına uyularak yazılmış “Mustafa Kemal Atatürk 19 Mayıs 1919’da Bandırma Vapuru ile Samsun’a çıktı. 23 Nisan 1920’de Türkiye Büyük Millet Meclisi’ni kurdu” cümlesi sorulmuş ve sonrasında aynı cümlenin tamamı küçük harfe dönüştürülerek tekrar sorulmuştur. Sorgu için kullanılan Python kodu ve araçların cevapları Şekil 3.4’te paylaşılmıştır.

```

from TurkishTopicModel import PreProcess
sentence=“““Mustafa Kemal Atatürk 19 Mayıs 1919’da Bandırma Vapuru ile Samsun’a çıktı. 23 Nisan 1920’de Türkiye Büyük Millet Meclisi’ni kurdu”““
PreProcess.Zemberek.TurkishNER (sentence)
[["PERSON",["Mustafa", 'Kemal', 'Atatürk']],["LOCATION",["Bandırma"]],["LOCATION",["Samsun'a"]],["LOCATION",["Türkiye"]]
morph= PreProcess.ITUNLPTools.ask (“pipelineSSMorph”, sentence)
morph=“<DOC> <DOC>+BDTag\n”+morph+”\n<DOC> <DOC>+EDTag”
print (PreProcess.ITUNLPTools.ask (“ner”, morph))
Mustafa B-PERSON, Kemal I-PERSON, Atatürk I-PERSON, 19 B-DATEMayıs I-DATE, 1919’da I-DATE, Samsun’a B-PERSON, 23 B-DATE, Nisan I-DATE, 1920’de I-DATE, Türkiye B-LOCATION
Sentence=sentence.lower ()
PreProcess.Zemberek.TurkishNER (sentence)
[]
morph= PreProcess.ITUNLPTools.ask (“pipelineSSMorph”, sentence)
morph=“<DOC> <DOC>+BDTag\n”+morph+”\n<DOC> <DOC>+EDTag”
print (PreProcess.ITUNLPTools.ask (“ner”, morph))
[]

```

Şekil 3.4: NER araçları sorgu kodu ve cevapları.

Şekil 3.4'te görüldüğü gibi büyük-küçük harf kurallarına uygun olarak yazılan cümle için Zemberek-NLP insan ve yer isimlerini yakalamış, İTÜ NLP Toolkit ise insan isimleri ve tarihleri yakalarken yer isimlerini yakalayamamış, "Samsun'a" kelimesini de yanlış bir şekilde insan ismi olarak etiketlemiştir. Metnin tamamı küçük harfe çevrilip sorgulandığında ise iki araç da herhangi bir etiket üretememiştir. Bu durum iki aracın da NER için kural tabanlı yöntemlerden yardım aldığı (Örn: Büyük harfle başlayan kelimeler insan ya da yer isimleri olabilir.) yönünde değerlendirilebilir. NER tabanlı etiketleme fonksiyonları Şekil 3.5'te paylaşılmıştır.

```
def tagNER (self,area,newarea=False,tagList=[]):
    if (newarea == False): newarea = area
    if (area != newarea): self.Data.datafrm[newarea] = pd.Series ()
    self.Data.datafrm[newarea] = self.Data.datafrm[area].apply (lambda x: self.__tagnerworker (x,tagList))

def __tagnerworker (self,x,tagList=[]):
    tags=self.Zemberek.TurkishNER (x)
    newtext=[]
    for tag in tags:
        if (tag[0] in tagList) or len (tagList)==0:
            newtext.append ("_" .join (tag[1]))
        else:
            newtext+=tag[1]
    return newtext
```

Şekil 3.5: NER yardımıyla terim etiketleme Python fonksiyonları.

Şekil 3.5'da paylaşılan fonksiyonlara göre tagNER() fonksiyonu tüm derlem için __tagnerworker() fonksiyonunun çalışmasını sağlar. __tagnerworker() fonksiyonu ise Zemberek-NLP'ye sorduğu metnin içinde bulunan etiketleri "_" karakterini kullanarak birleştirir.

Terim etiketleme konusunda kullanılan bir diğer yaklaşım ise PMI yöntemidir. n adet kelimenin PMI skoru bir terim olabilme ihtimallerini belirtir. Bu skorlar incelenerek belirli bir seviyenin üzerinde bulunan PMI skorları terim olması muhtemel kelime birliktelikleri olarak değerlendirilebilir. Bu yöntem için tomotopy[138] kütüphanesinin derlem sınıfında yer alan bir fonksiyon olan *extract_ngrams* kullanılabilir. Örnek olması amacıyla tomotopy kütüphanesi GitHub hesabında bulunan örnek değerlendirme kodlarının⁸ kısaltılmış versiyonu Şekil 3.6'de paylaşılmıştır.

⁸ https://github.com/bab2min/tomotopy/blob/main/examples/extract_ngram.py

```

import tomotopy as tp
corpus = tp.utils.Corpus (tokenizer=tp.utils.SimpleTokenizer (),
stopwords=lambdax: len (x) <= 2 or x in stops)
# data_feeder yields a tuple of (raw string, user data) or a str (raw string)
corpus.process (open ('enwiki-1000.txt', encoding='utf-8'))
cands = corpus.extract_ngrams (min_cf=20, min_df=10, max_len=5, max_cand=1000, normalized=True, workers=0)
for cand in cands:
    print (cand)
>>>tomotopy.label.Candidate (words=["buenos","aires"], name="", score=0.996445)
>>>tomotopy.label.Candidate (words=["los","angeles"], name="", score=0.988719)
>>>tomotopy.label.Candidate (words=["las","vegas"], name="", score=0.982273)
>>>tomotopy.label.Candidate (words=["hong","kong"], name="", score=0.978606)
>>>tomotopy.label.Candidate (words=["hip","hop"], name="", score=0.965971)
>>>tomotopy.label.Candidate (words=["nova","scotia"], name="", score=0.957440)
>>>tomotopy.label.Candidate (words=["ice","hockey"], name="", score=0.932300)
>>>tomotopy.label.Candidate (words=["nobel","prize","laureate"], name="", score=0.927281)
>>>tomotopy.label.Candidate (words=["sri","lankan"], name="", score=0.925504)
>>>tomotopy.label.Candidate (words=["ann","arbor"], name="", score=0.921129)
corpus.concat_ngrams (cands, delimiter='_')

```

Şekil 3.6: `extract_ngrams` fonksiyonuyla etiket adayları oluşturmak ve birleştirmek.

Şekil 3.6’de `extract_ngrams` fonksiyonunun çalışması gösterilmektedir. Fonksiyon parametre olarak `min_cf` (n-gram’ların minimum *gf* değeri), `min_df` (n-gram’ların minimum *df* değeri), `max_len` (maksimum n-gram uzunluğu), `max_cand` (bulunacak en fazla terim adayı), `normalized` (Skor üretiminde PMI değerlerinin saf hallerinin ya da normalize edilmiş hallerinin mi kullanılacağı), `workers` (işlemci paralelleştirmesi yapılacaksa işlemin kaç parçaya bölüneceği) parametrelerini kullanmaktadır. Fonksiyon, hesaplamaları yaptıktan sonra terim adayları üretir ve istenirse bu terim adayları herhangi bir karakter yardımıyla birleştirilerek tek bir terim haline getirilebilir.

3.2.1.5 Morfolojik İndirgeme

Araştırma kapsamında Lemmatization yöntemi kullanılmış ve Zemberek-NLP aracından faydalanılmıştır. Zemberek-NLP aracına sorulan örnek soru ve cevabı Şekil 3.7’de gösterilmiştir.

```

from TurkishTopicModel import PreProcess

PreProcess.Zemberek.Morphology (text="Ayakkabın eskimiş. Ayakkabıcıya götür de tamir ettir", type="lemma", length="max")
>>> {'Ayakkabın': 'ayakkabı', 'eskimiş': 'eskimiş', 'Ayakkabıcıya': 'ayakkabıcı', 'götür': 'götür', 'de': 'de', 'tamir': 'tamir', 'ettir': 'ettir'}

PreProcess.Zemberek.Morphology ("Gülmekten daha güzel bir eylem varsa o da gülmektir", type="lemma", length="max")
>>> {'Gülmekten': 'gülmek', 'daha': 'daha', 'güzel': 'güzel', 'bir': 'bir', 'eylem': 'eylem', 'varsa': 'var', 'o': 'o', 'da': 'da', 'gülmektir': 'gülmek'}

```

Şekil 3.7: Morfolojik indirgeme fonksiyonunun ürettiği çıktılar.

Geliştirilen morfolojik analiz fonksiyonu `text` (morfolojik analizi yapılacak metin), `type` (lemmatization veya stemming), `length` (analizden gelen sonuçlardan maksimum ya da

minimum uzunluktakinin gönderileceği) parametrelerini almaktadır. Kelimelerden daha az ek budayarak kelimenin sözlükteki en yakın karşılığını almak amaçlandığından lemmatization metoduyla birlikte maksimum uzunluk parametreleri tercih edilmiştir. Böylece örnek olarak verilebilecek “Ayakkabıcılar” kelimesinin en kısa kökü olan “ayak” yerine en az ek çıkarıldıktan sonra elde kalan “ayakkabıcı” kelimesine indirgeme yapılması sağlanmaktadır.

3.2.1.6 Eşanlam İndirgemesi

Eşanlam indirgemesi için sözlük tabanlı bir metot kullanılmıştır. Sözlük olarak Yıldız Teknik Üniversitesi Kemik Doğal Dil İşleme Grubu⁹ tarafından oluşturulan “Türkçe Anlamsal İlişkiler Veri Kümesi- TAİVK” ve Dokuz Eylül Üniversitesi – “Eş ve Yakın Anlamli Kelimeler Sözlüğü - EYAKS¹⁰” kullanılmıştır. TAİVK indirilebilir bir sözlüktür ve bilgisayar üzerinde bulundurulabilir fakat EYAKS Internet ortamından sorgu yapılabilen bir sözlüktür ve bilgisayara indirilememektedir. Internet tabanlı sözlüklerin ya da araçlar küçük sözlüklerde kullanılabilir fakat büyük sözlüklerde sorgular yüksek miktarda zaman almaktadır. Bu durumda EYAKS’ta yapılacak sorgularda yerel bir sözlük oluşturularak sorgulanan kelimelerin sonraki sorgularda tekrar sorgulanmaması sağlanmıştır. Sözlük tabanlı yaklaşımın diğer bir zorluğu bu tür sözlüklerde kelimelerin sözlük karşılıklarının bulunuyor olmasıdır. Sorgu yapılmadan önce kelimelere morfolojik analiz yapılarak sözlük karşılıkları alınmalı ve sorgular bu karşılıklarla yapılmalıdır.

Eşanlam indirgemesi amacıyla kullanılan algoritma Bölüm 3.2.1.2’de anlatılan *gfJ* tabanlı “Otomatik Yazım Yanlışı Düzeltme Algoritması” kullanır. Öneri üretimi için kullanılan fonksiyon *SC (w)* değiştirilerek TAİVK ve EYAKS fonksiyonlarının ürettikleri öneriler derlemde en büyük *gfJ* skoruna sahip kelimeyle değiştirilir.

3.2.1.7 Dolgu Kelimelerinin ve Belgelerinin Kaldırılması

Daha önce de bahsedildiği üzere dolgu kelimelerini kaldırmak vektör uzayını küçültmenin yanında, modelleme aşamasında işlem hızını ve doğruluğunu artırarak performansı

⁹ <http://www.kemik.yildiz.edu.tr>

¹⁰ <http://nlpapps.cs.deu.edu.tr/esveyakin>

iyileştirmeye yardımcı olan bir işlemdir. Bu tür kelimeleri tespit ederek derlemden çıkarmak amacıyla sözlük tabanlı yaklaşım olarak;

1. NLTK dolgu kelime sözlüğü,
2. Modelleme denemeleri sırasında inceleme yapıldığında her konuda kendine yer bulan kelimelerden oluşturulan sözlük,
3. Derlemde en yüksek ve en düşük *gf* skoruna sahip kelimelerden oluşturulan sözlük ve
4. VYS’ye kayıtlı kullanıcı (öğretmen, öğrenci vb.) adları ve soyadları kullanılarak oluşturulan isim sözlüğü kullanılmıştır.

Araştırmada yapılan modelleme denemeleri sırasında da “*Dolgu Belgeleri*” ismi verilebilecek ve öğretmenler tarafından doldurulduğu görülen ya da yeterince özellik içermeyen belgeler bağlamdan çıkarılmıştır. “*Dolgu Belgeleri*” ismi altında sınıflandırılan belgelere ait örnekler aşağıda paylaşılmıştır.

1. Tüm alanları aynı metinlerle doldurulmuş belge kümeleri (Örn: “Kar Tatili”, “Kar Tatili”, “Kar Tatili”)
2. Tek alanı doldurulmuş belge kümeleri (Örn: “Ders yapılmadı”, “”, “”)
3. Belirli bir kelime sayısından daha az kelime içeren belge kümeleri (Örn: “Mutlu”)
4. Belirli sayıdan fazla tekrar eden belgeler (Örn: 15 x (“Şenlik Hazırlığı”, “Şenlik Hazırlığı”, “Şenlik Hazırlığı”))

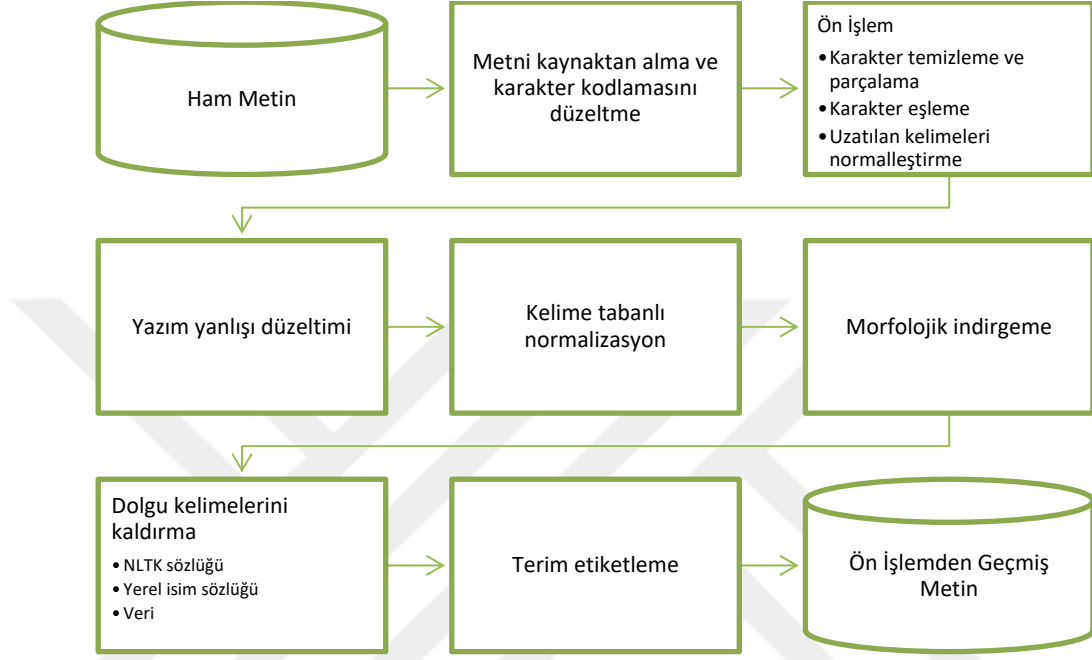
3.2.1.8 Sürecin Genel Değerlendirmesi

Bölüm 3.1.6 “Veriyi Hazırlama” genelinde veriyi modellemeye hazırlamak amacıyla hangi işlemlerin, hangi amaçla tercih edildiği, bu işlemleri gerçekleştirmek için hangi araçlardan nasıl yararlanıldığı ve hangi algoritmaların geliştirildiği anlatılmıştır. Seçilen ve geliştirilen yöntemlerin veriye uygulanması ve veriye etkileri bu başlık altında işlenecektir.

- Duygu ifade etmek için kelimelerin uzatılması:Örn: çok eğlendimmm çok eğlendimm
- Noktalama işaretlerinin kullanılmaması ya da gereksiz kullanılması: Örn: "Anneler günü için mblockta kod yazdım Servo motor yaptım", "seslere sinir oldum. ama grupça calistigimdan mutluyum."
- ASCCIifying: Örn. "seslere sinir oldum. ama grupça calistigimdan mutluyum.", "kulaktan kulağa oynadık sonra karikatürü inceledik ondan sonra www.wordart.com a girdik kelimelerden şekil tasarladık”
- İngilizce terimlerin yazım yanlışları, hatalı kullanımları: arduino->ardunio, MBlock->mblok, Kahoot->kahot vb.

- Büyük-küçük harf kurallarına uyulmaması: Örn: "Bahçede Maydanoz ektik, bilgisayardan quizlet yarışması yaptık."

Bu sebeple süreç içinde ön işleme süreci yeniden ele alınarak dönüştürülmüş ve ortaya çıkan süreç diyagramı Şekil 3.9’da paylaşılmıştır.



Şekil 3.9: Süreç sonunda oluşan ön işleme süreci diyagramı.

Şekil 3.9’da paylaşılan Ön İşlem Sürecinin basamaklarının işletilmesi sonucunda oluşan metin dönüşümü Tablo 3.5’te paylaşılmıştır.

Tablo 3.5: Metin ön işleme işlem basamakları, işlem öncesi ve sonrası verilere ait tablo.

Yapılan İşlem	İşlem Öncesi Veri	İşlem Sonrası Veri
Karakter kodu düzeltme	"G\u00f6zel hissettim."	"Güzel hissettim."
Karakter temizleme ve parçalama	Bugün çok güzeldi. Bir sürü şey yaptık. Mutluydum."	['Bugün', 'çok', 'güzeldi', 'Bir', 'sürü', 'şey', 'yaptık', 'Mutluydum']
Karakter eşleme	['GüzeeeeeeEEEEel']	['güzeeeeeeeeeeel']
Uzatılan kelimeleri normalleştirme	['güzeeeeeeeeeeel']	['güzeel']
Yazım yanlışları düzeltimi	['mutlu', 'heyecanlı', 've', 'hırslı']	['mutlu', 'heyecanlı', 've', 'hırslı']
Kelime tabanlı normalizasyon	['çevre', 'kirliliği', 'hayvanların', 'ölmesi', 'gibi', 'şeyler', 'için', 'üzüldüm']	['çevre', 'kirliliği', 'hayvanların', 'ölmesi', 'gibi', 'şeyler', 'için', 'üzüldüm']
Morfolojik indirgeme	['eğlenceliydi', 'mutlu', 'hissettim']	['eğlenceli', 'mutlu', 'hisset']
Dolgu kelimelerini kaldırma	['ibrahim', 'müdür', 've', 'yunus', 'abi', 'ile', 'plickers', 'yaptık', 'ağaç', 'kuş', 'astık', 'arada', 'anne', 've', 'mehmet', 'ali', 'gördüm']	['abi', 'plickers', 'yaptık', 'ağaç', 'kuş', 'astık', 'arada', 'anne', 'gördüm']

Ön işlem sürecinde yaşanan en büyük problemlerden birisi normalizasyon işleminin doğru kelimeleri de değiştirmesi ve küfürlü kelimeler üretebilmesidir [36]. Bu sebeple cümle bazında yapılması gereken normalizasyon işlemi kelime bazına indirilmiş ve doğru yazılan kelimeler normalizasyon işlemine sokulmamıştır.

Terim etiketleme işlemi için ise konu modelleme için derlem oluşturduktan sonra *nPMI* ağırlıklandırma yöntemi kullanılarak 1000 adet terim adayı için analiz yapılmış ve en yüksek skora sahip terim adayları ve *nPMI* değerleri Tablo 3.6’da paylaşılmıştır.

Tablo 3.6: Terim adayları ve *nPMI* değerleri.

S1	<i>nPMI</i>	S2	<i>nPMI</i>	S3	<i>nPMI</i>	
1	['atsushi', 'miyazaki']	1,00	['power', 'point']	1,00	['beyin', 'fırtına']	0,93
2	['nezahat', 'gökyiğit']	1,00	['abur', 'cubur']	1,00	['zihin', 'harita']	0,92
3	['stemist', 'space']	1,00	['stop', 'motion']	1,00	['bahk', 'kılçık']	0,92
4	['ping', 'pong']	1,00	['alış', 'veriş']	0,99	['mide', 'bulan']	0,89
5	['abur', 'cubur']	1,00	['donuk', 'imge']	0,99	['iple', 'çek']	0,88
6	['just', 'dance']	1,00	['chrome', 'music']	0,99	['birlik', 'beraberlik']	0,88
7	['kuşçu', 'gözlemevi']	0,99	['fast', 'food']	0,98	['kırıklık', 'uğra']	0,88
8	['kamu', 'spot']	0,99	['action', 'sound']	0,98	['ilçe', 'milli']	0,87
9	['deve', 'cüce']	0,98	['kamu', 'spot']	0,98	['milli', 'eğitim']	0,86
10	['rob', 'reşat']	0,98	['movie', 'maker']	0,98	['geri', 'dönüşüm']	0,85
11	['zip', 'zap']	0,97	['ayak', 'iz']	0,98	['milli', 'eğitim', 'müdür']	0,84
12	['code', 'org']	0,97	['green', 'screen']	0,97	['şube', 'müdür']	0,84
13	['bakırköy', 'gençlik']	0,97	['meşe', 'palamut']	0,97	['hoş', 'git']	0,84
14	['prens', 'prensler']	0,97	['my', 'wall']	0,97	['hayal', 'güç']	0,84
15	['meşe', 'palamut']	0,97	['sketch', 'up']	0,97	['hayal', 'kırıklık']	0,83
16	['power', 'point']	0,97	['code', 'org']	0,97	['ilçe', 'milli', 'eğitim']	0,83
18	['living', 'lab']	0,97	['küresel', 'ısınma']	0,96	['gurur', 'duy']	0,82
19	['my', 'wall']	0,97	['yanıp', 'sönme']	0,96	['ilçe', 'milli', 'eğitim', 'müdür']	0,82
20	['tik', 'tak']	0,96	['sadberk', 'hanım']	0,96	['eh', 'işte']	0,82

Tablo 3.6’da bulunan veriler incelendiğinde terim adayı oluşturmada *nPMI* ağırlıklandırma yönteminin mantıklı terim adayları ürettiği görülebilir. Terimler, terim adayları arasından göz incelemesiyle 106 aday seçilmiş ve işaretlenmiştir. Örneğin işaretleme öncesinde “['balık', 'kılçık', 'yaptık', 'oyun', 'oynadık', 'yazdık']” şeklindeki bir belge işaretleme sonrasında “['balıkkılçık', 'yaptık', 'oyun', 'oynadık', 'yazdık']” haline dönüştürülmüştür.

3.1.7 Modelleme ve Değerlendirme

LDA’in arkasında yatan ana düşünce belgelerin *gizli konuların* rastgele karışımları olarak temsil edildiği ve her konunun kelimelerin özel dağılımlarına göre oluştuğudur. Metin madenciliği sürecinin işletildiği derlemde bulunan veriler daha önce üzerinde işlem

yapılmamıştır ve ihtiva ettiği desenler bilinmemektedir. Bu sebeple tez kapsamında keşif odaklı bir süreç takip edilmiş ve veri içindeki gizli desenleri ortaya çıkarmak amaçlanmıştır.

Veri madenciliği; modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı bir yaklaşımdır [12]. Bu sebeple modelleme aşamasında aşağıdaki sorulara cevap aranmıştır.

1. Verinin kapsamında bulunan 3 derlemin de (S1, S2, S3) optimum konu sayıları nasıl hesaplanır?
 - a. Derlemden dolgu belgeleri nasıl çıkarılabilir?
 - b. Her konuda yer bulan dolgu kelimeleri nelerdir?
 - c. Modelleme parametreleri nasıl seçilmelidir ve modeller nasıl optimize edilmelidir?
 - d. Derlemde bulunan konuları ortaya çıkarmak için hangi başarımlar ölçütleri nasıl uygulanmalı ve nasıl değerlendirilmelidir?
 - e. Konu sayısı nasıl belirlenmelidir?

3.2.1.9 Konu Modelleme Algoritması

Konu modelleme amacıyla LDA algoritması seçilmiş ve bu algoritma veriye *tomotopy* kütüphanesiyle uygulanmıştır. *tomotopy* kütüphanesinde bulunan `LDAModel` sınıfı parametreleri Şekil 3.10'da gösterilmiştir.

```
class LDAModel (tw=TermWeight.ONE, min_cf=0, min_df=0, rm_top=0, k=1, alpha=0.1, eta=0.01,
seed=None, corpus=None, transform=None)
```

Şekil 3.10: Tomotopy kütüphanesi `LDAModel` sınıfı parametreleri.

Şekil 3.10'da bulunan sınıf gösterimine göre sınıfın aldığı parametreler şöyle açıklanabilir.

- *tw* (Terim Ağırlıklandırma Şeması- *Term Weighting Scheme*): Kullanılacak terim ağırlıklandırma algoritmasıdır. Her terimin modellemeye aynı ağırlıkla başladığı (*ONE*), derlemde en az geçen terimin en değerli olduğu (*IDF*) ve terimlerin birlikte geçme oranlarına göre ağırlıklandırma uygulanan (*PMI*) ağırlıklandırma fonksiyonları ağırlıklandırma yöntemi olarak kullanılabilir.

- *min_cf* (*Minimum Derlem Frekansı – Minimum Collection Frequency*): Kelimelerin derlemde geçme sıklığına göre hesaplamadan çıkarılması için kullanılan parametredir. Öntanımlı değeri 0'dır.
- *min_df* (*Minimum Belge Frekansı – Minimum Document Frequency*): Kelimelerin belgelerde geçme sıklığına göre hesaplamadan çıkarılması için kullanılan parametredir. Öntanımlı değeri 0'dır.
- *rm_top* (*Yükseği Kaldır – Remove Top*): Derlemde en çok geçen kelimelerin derlemde çıkarılması için kullanılır. Öntanımlı değeri 0'dır.
- *k* (*Konu Sayısı – Number of Topics*): Derlemde kaç konu çıkarılacağını belirleyen parametredir.
- *alpha* (α): Belge-Konu dağılımı için Dirichlet dağılımının hiper parametresidir.
- *eta* (η): Konu-Terim dağılımı için Dirichlet dağılımının hiper parametresidir.
- *seed* (*Rasgelelik Sabiti – Random Seed*): Algoritma kapsamında üretilen rasgele değerlerin sabitlenmesi için kullanılır. Rasgele değerler sabitlenmediğinde algoritma her işlediğinde farklı konu dağılımları çıkarma ihtimali yüksektir.
- *corpus* (*Derlem – Corpus*): Modellemeye girecek derlemdir ve belgelerden oluşur.
- *Transform* (*Dönüşüm – Transform*): İsteğe bağlı anahtar kelimeleri işleme katmak için kullanılan bir parametredir.

LDAModel sınıfına kaynak olan derlemi üreten Derlem (*Corpus*) ve modeli belli bir tekrarda eğitmek amacıyla kullanılan Eğitim (*Train*) sınıfları da yardımcı sınıflar olarak kullanılmaktadır. Büyük derlemleri yönetebilmek için oluşturulmuş olan *Corpus* sınıfı metin parçalama, terim adayı üretimi ve etiketlemesi, dolgu kelimelerinin kaldırılması gibi özellikler de sunmaktadır. *Corpus* sınıfı ve parametreleri Şekil 3.11'de paylaşılmıştır.

```
class Corpus (tokenizer=None, batch_size=64, stopwords=None)
```

Şekil 3.11: Tomotopy kütüphanesi Corpus sınıfı parametreleri.

Şekil 3.11'de bulunan parametreler şöyle açıklanabilir.

- *tokenizer (Parçalayıcı – Tokenizer)*: Belgeleri parçalamak için hangi parçalama algoritmasının kullanılacağını ifade eder. *tomotopy* dışındaki kütüphanelerin de parçalayıcıları, kök-gövde üreticileri bu nesne için veri üretebilir.
- *batch_size (Demet Büyüklüğü – Batch Size)*: Corpus sınıfının derlem işleme fonksiyonu bu parametrede belirtilen sayıda belgeyi aynı anda okuyarak işler ve derleme ekler.
- *stopwords (Dolgu Kelimeleri – Stop Words)*: Derleme eklenen belgeler işlenirken kelimeler arasından çıkarılacak kelimeleri ifade eder.

Train fonksiyonu ise modelin kaç tekrarda eğitileceğini, eğitim işlemi için hangi paralelleştirme şemasının kullanılacağını, veri setinin hangi algoritma kullanılarak örneklenip bu paralel işlemlere gönderileceğini tanımlayan parametreler almaktadır. *Train* fonksiyonu ve parametrelerinin gösterimi Şekil 3.12’de paylaşılmıştır.

```
Def train(self, iter=10, workers=0, parallel=0, freeze_topics=False)
```

Şekil 3.12: Tomotopy kütüphanesi train fonksiyonu ve parametreleri.

Şekil 3.12’de bulunan parametreler şöyle açıklanabilir;

- *iter (Terkar – Iteration)*: Gibbs örneklemesinin yineleme sayısı
- *workers (İşçi Sayısı – Workers)*: Model eğitimi için kaç işlemci çekirdeğinin kullanılacağını belirtildiği parametredir. Öntanımlı olarak tüm işlemci çekirdekleri tekrar için kullanılır.
- *parallel (Paralelleştirme Şeması – Parallel Scheme)*: Tekrarlama işlemlerinin hangi yöntem kullanılarak paralelleştirileceğinin belirtildiği parametredir. İşlem paralelleştirme seçildiğinde kullanılan hafıza ve işlem yükü artacağından modellemenin yapıldığı bilgisayarın işlem gücünün ve hafızasının yüksek olması önerilmektedir.
- *freze_topics (Konuları Dondur – Freeze Topics)*: hLDA gibi hiyerarşik eklentilerin eğitim sırasında yeni konu üretmelerini engellemek için kullanılır.

3.2.1.10 Değerlendirme Yöntemleri

Daha önce Bölüm 2.1.6’te bahsedildiği üzere oluşturulan modelin etkililiğini değerlendirmek üzere doğrudan ve dolaylı yöntemler kullanılmaktadır. Doğrudan yöntemler olan *log-likelihood* ve *Perplexity* değerleri model eğitilirken otomatik olarak hesaplanır. Daha önce de bahsedildiği üzere *Perplexity* ya da başka bir deyişle “Tahminsel Olabilirlik – Predictive Likelihood” , test verilerinin olasılığında sabit olarak azalmaktadır ve matematiksel olarak kelime başına olasılığın (*likelihood*) geometrik ortalamasının tersi ile eşdeğerdir [59]. Düşük *Perplexity* ve artan Log-Likelihood değeri veri seti üzerinde daha yüksek bir modelleme doğruluğu anlamına gelir [76].

Matematiksel gösterimleri Bölüm 2.1.6’te gösterilen ve dolaylı yöntemler olan C_{uci} , C_{u_mass} , C_{nmpi} ve C_v başarımlar ölçütleri ise model eğitildikten sonra ayrı bir hesaplama gerektirmektedir. *Tomotopy* kütüphanesinin bir sınıfı olan Tutarlılık (*Coherence*) modelin başarımlarını ölçmek için kullanılabilir. *Coherence* sınıfına ait parametreler Şekil 3.13’te paylaşılmıştır.

```
class Coherence (corpus, coherence='u_mass', window_size=0, targets=None, top_n=10,
eps=1e-12, gamma=1.0)
```

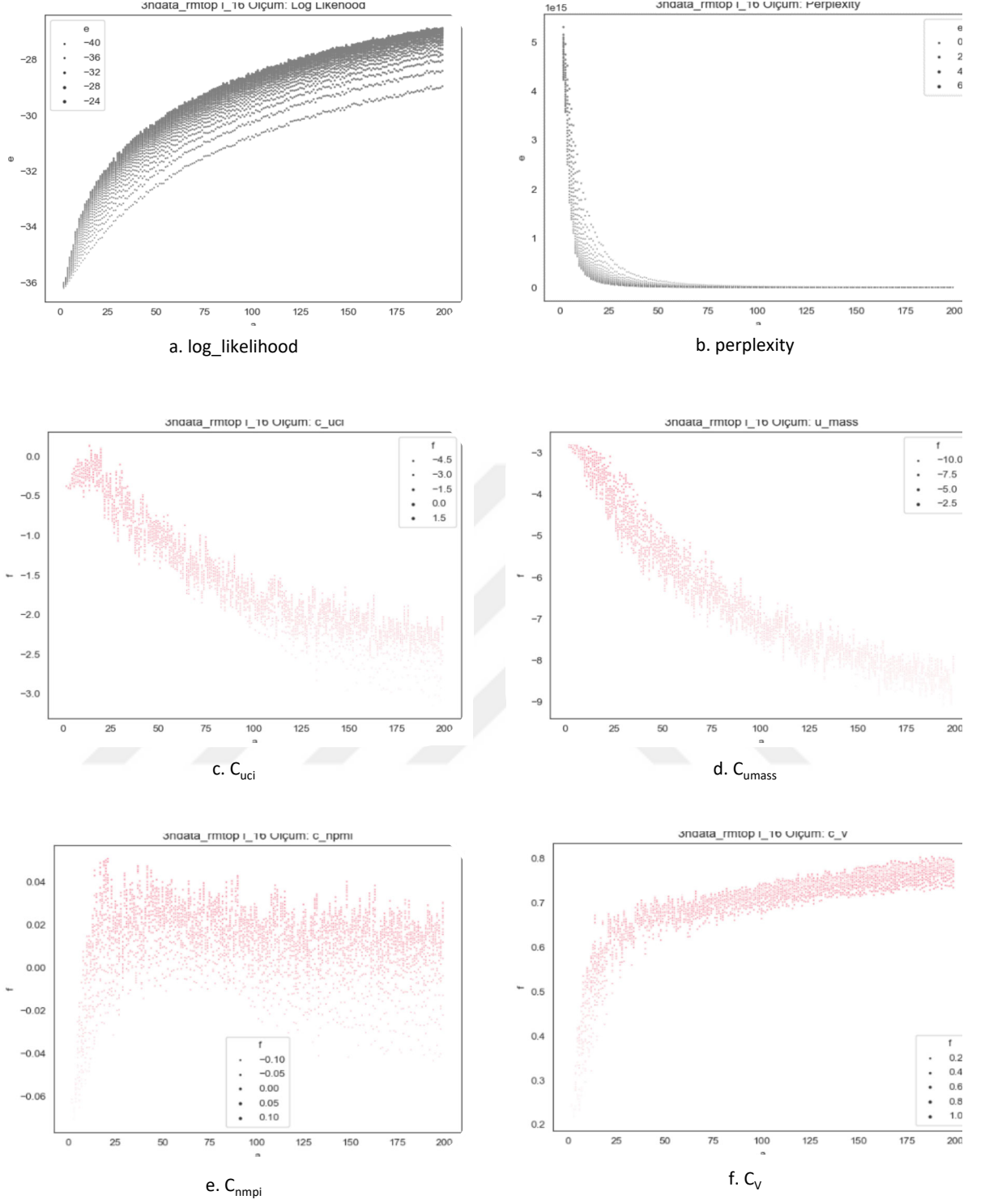
Şekil 3.13: Tomotopy kütüphanesi Coherence sınıfı parametreleri

Şekil 3.13’te paylaşılan *Coherence* sınıfı parametreleri şöyle açıklanabilir;

- *corpus* (*Derlem – Corpus*): Başarımı ölçülecek derlem ya da LDAModel tarafından oluşturulmuş bir model bu parametreye eklenerek başarımlar ölçülebilir. Eğer model yerine Derlem kullanılırsa *targets* parametresinin de kullanılması gerekmektedir.
- *coherence* (*Tutarlılık – Coherence*): Derlem yada model üzerinde hangi başarımlar ölçümü algoritmasının işletileceğinin belirtildiği parametredir. C_{uci} , C_{u_mass} , C_{nmpi} ve C_v tutarlılık ölçümlerini desteklemektedir.
- *window_size* (*Pencere Büyüklüğü, Window Size*): Kayan pencere yaklaşımı (P_{sw}) belirli bir kelimeyi gösteren (pencere büyüklüğü) bir pencerenin belge üzerinde adım adım kaymasını ifade eder. C_{uci} ve C_{nmpi} için varsayılan değer $P_{sw}(10)$ ve C_v için $P_{sw}(110)$ olarak alınır.

- *targets (Hedefler – Targets)*: Her konu için en yüksek olasılıklı kelimelerin bir listesidir. Eğer *corpus* parametresine oluşturulmuş ve eğitilmiş bir model verilirse *targets* parametresine ihtiyaç olmamaktadır.
- *eps (ϵ)*: Logaritmik hesaplamalarda sıfırla bölünmenin önüne geçmek için hesaplamalara eklenen sayıdır.
- *gamma (Γ)*: Dolaylı yöntemler için kullanılan giriş parametresidir ve ön-tanımlı olarak 1.0'dir.

Dolaylı yöntemler modelin eğitiminden sonra ayrıca hesaplandığı için işlem süresini uzatmaktadır, bu sebeple doğrudan yöntemlerle birlikte kullanılması hız açısından önemlidir. Örnek bir form alanında belli bir konu aralığı (2-200) ve belli bir tekrarlama aralığı (40-2000) ile yapılan model eğitimi ile elde edilen *log-likelihood*, *perplexity*, C_{uci} , C_{u_mass} , C_{nmpi} ve C_v tutarlılık ölçümlerine ait grafikler Şekil 3.14'te paylaşılmıştır.



Şekil 3.14: LDA modeli 2-200 konu aralığı, 40-2000 tekrar için başarımlı tutarlılık ölçümleri serpinti grafikleri.

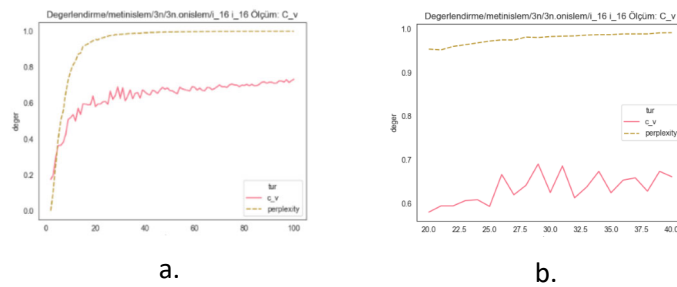
Şekil 3.14'te bulunan grafiklerde a. *Log-Likelihood*, b. *Perplexity*, c. C_{uci} , d. $C_{u_{mass}}$, e. C_{nmpi} ve f. C_v tutarlılık ölçümlerine ait grafikler paylaşılmıştır. Grafiklerde yatay eksen konu sayısını, dikey eksen ise o konu sayısında alınan ölçüm değerini ifade eder. Her sütunda bulunan noktalar

40 tekrardan başlamak üzere 2000 tekrara kadar her 40 tekrarda bir alınan ölçüm değerlerini gösterir. Örneğin: Şekil 3.14.a 25-50 konu değerleri arasında *perplexity* değerlerinin maksimum değerlerine ulaştığı ve 50-75 konu aralığında tekrarlar arasındaki değer farkı azalmıştır şeklinde yorumlanabilir. Dolaylı ölçüm yöntemlerine ait grafikler (c., d., e. ve f.) birlikte değerlendirildiğinde ise C_v ölçüm yöntemine ait grafiğin diğer grafiklerden daha kolay değerlendirildiği görülebilir. Farklı bir yorum olarak; “Şekil 3.14.f’te paylaşılan C_v grafiğine göre 25-50 konu aralığında ideal değerlere ulaşılmış ve 50-75 konu aralığında maksimum değerlere yakın değerler görülmüştür” diye değerlendirme yapılabilir.

Yukarıdaki sebeplerden dolayı doğrudan yöntemlerden *perplexity*, dolaylı yöntemlerden ise C_v yönteminin muhtemel konu sayılarının belirlenmesinde tutarlılık ölçümleri olarak kullanılmasının uygun olacağı değerlendirilmiştir. Bu iki yöntemin ürettiği sonuçların aynı grafikte gösterilmesini engelleyen sorun *perplexity* değerlerinin (0-1e15) aralığında değer alması fakat C_v değerlerinin ise (0-1) aralığında bulunmasıdır. Bu sebeple *perplexity* değerleri (0-1) aralığına indirgenmiş ve ters çevrilmiştir. Normalleştirilmiş *perplexity* (*normalperplexity*) değerini hesaplamak için kullanılan formül Denklem (3.1’de paylaşılmıştır.

$$normalperplexity = 1 - \frac{perplexity - \min(perplexity)}{\max(perplexity) - \min(perplexity)} \quad (3.1)$$

Denklem (3.1’de paylaşılan *normalperplexity* ve C_v değerlerinin birlikte bulunduğu grafikler Şekil 3.15’te paylaşılmıştır.



Şekil 3.15: normalperplexity ve C_v değerleri ortak grafiği.

Örnek olarak Şekil 3.15’de paylaşılan grafikleri (a. ve b.) birlikte inceleyerek yorumlamak gerekirse “*perplexity* değerlerinin 20-40 konu aralığında sabitlendiği ve C_v değerlerinin de aynı değerler arasında maksimum değerlerine yakın sonuçlar ürettiği değerlendirilebilir. Değerler

20-40 konu aralığında yaklaştırıldığında ise 29 konu sayısının *perplexity* ve C_v değerleri bakımından muhtemel konu sayısı olduğu değerlendirilebilir” şeklinde yorum yapılabilir.

3.2.1.11 Derlemin Oluşturulması

Derlemlerin oluşturulmasında daha önce Şekil 3.11’de gösterilen Corpus sınıfı kullanılmıştır. Derlem oluşturmak için kullanılan fonksiyonlar Şekil 3.16’da paylaşılmıştır.

```
TurkishTopicModel.transferData ()

TurkishTopicModel.TopicModel.createCorpus
(column="neyaptin8",compare="neogrendin8",stops=['yap','yaptık'],clean=["neogrendin8","nehissetin8"],df=10)

TurkishTopicModel.TopicModel.createCorpus (column="neogrendin8", stops=['öğren','öğrendik'],
compare="neyaptin8", clean=["neyaptin8","nehissetin8"], df=10)

TurkishTopicModel.TopicModel.createCorpus
(column="nehissetin8",stops=['hisset','duygu'],compare="neogrendin8",clean=["neyaptin8","neogrendin8"],df=10)
```

Şekil 3.16: Derlem oluşturmak için kullanılan kod blokları.

Şekil 3.16’de gösterilen kod blokları incelendiğinde öncelikle belgelerin TopicModel sınıfına transfer edildiği, sonrasında ise her alan için bir derlem oluşturulduğu görülebilir. Fonksiyonun aldığı parametrelerden *column* derlemin hangi form alanından oluşturulacağı, *compare* alanında hangi alanla karşılaştırılarak derleme eklenip/eklenmeyeceğine karar verileceği, *clean* ise hangi form alanlarında olan kelimelerin bu alandan kaldırılacağını belirlemektedir. *df* parametresi ise belge frekansı belirli bir sayının üzerinde olan belgelerin derleme alınmayacağını gösterir. Çoklu alanlar kullanılarak okullardan toplanan ve araştırma kapsamında işlenen belgelerde derlem oluşturma konusunda yaşanan sorunlardan birisi Bölüm 3.2.1.7’de bahsedildiği gibi dolgu belgeleri olarak isimlendirilebilecek ve genel olarak formlar doldurulurken yer tutucu olarak eklenmiş belgelerdir. Bu belgeler aşağıda örneklenmiştir.

- Formun doldurulmadığını gösterir belgeler (“Kar tatili”, “Öğrenci Harezmi Modeli dersine katılım sağlayamamıştır”, “Toplantı yapıldı”)
- Çok az kelime içeren ya da yer tutucu kullanılan belgeler (“Mutlu”, “fgfdgwtr”)
- Tüm alanları aynı doldurulmuş belgeler ([“Ders yaptım”, “Ders yaptım”, “Ders yaptım”])

Dolgu belgelerini derlemden ayırmak aşağıdaki metotlar takip edilmiştir.

- Derlemde belirli bir belge frekansının üstünde bulunan belgeler dolgu belgesi olarak kabul edilmiştir. Bu tip belgeler bir ekibin belirli haftalarda her üyesi için aynı içerikte doldurulmuştur. Bu tip belgelerin kaldırılması için belge frekansı tüm alanlar için $df=10$ olarak kabul edilmiştir.
- Tüm alanları aynı olan belgeler derlemden çıkarılmıştır. Bu belgeler de çoğunlukla dolgu belgesi özelliği taşımaktadır.

Dolgu belgelerinin dışında; derlem oluşturulurken form alanlarını kelime düzeyinde birbirinden ayırmak için derleme eklenen alandaki kelimeler formun diğer alanlarıyla karşılaştırılmış, form girdisinde diğer alanlarda da bulunan kelimeler derlemden çıkarılmıştır. Bu temizlik yöntemindeki varsayım “Bir öğrenci eğer tüm alanlarda aynı kelimeleri kullanmışsa S1, “Ne düşündün?” ve S3 sorularından herhangi birine cevap vermiş fakat bu cevabı diğer alanlara kopyalayıp yapıştırmıştır ya da cevap verdiği alanla anlattığı konu farklıdır” şeklinde kurulmuştur.

Konu sayıları belirlenirken oluşturulan modellerde her konuya ağırlık olarak katkı sunan kelimeler de tespit edilerek derlemden çıkarılmıştır. Dolgu kelimesi bakımından en zorlu derlem S3 derlemidir. Bu derlemde bulunan sorun; başlangıçta 84.848 belge bulunmaktayken, 36.842 belgede “mutlu” kelimesinin bulunması ve 26.866 belgenin ise tek kelime yazılmasıyla özetlenebilir. Tek kelimeyle yazılan belgeler “mutlu, iyi, eğlen, hiçbir, normal, güzel” kelimeleriyle örneklenebilirler. Böylece derlem bazında oluşturulan karşılaştırma, temizleme derlemleri ve dolgu kelimeleri ve kabul edilen minimum derlem uzunluğu (kelime) Tablo 3.7’de paylaşılmıştır.

Tablo 3.7: Derlem bazlı karşılaştırma, temizleme alanları ve dolgu kelimeleri tablosu.

Derlem	Karşılaştırma	Temizleme	Dolgu Kelimeleri	Minimum Belge Uzunluğu	Maksimum Belge Tekrarı
S1	S2	S2 S3	'yap','yaptık'	2	10
S2	S1	S2 S2	'olduk','bilgi','ilgili','edin','öğren','öğrendik'	2	10
S3	S2	S1 S2	'güzel', 'mutlu', 'mutluluk', 'happy', 'good', 'very', 'feel', 'glen', 'harezmimodeli', 'harezmi', 'model', 'modeli', 'hisset', 'hissettir', 'duygu'	2	10

3.2.1.12 Modelleme Parametreleri

LDA modeli modelleme parametrelerinin en önemlileri α ve η ile belirtilen konu-belge dağılımı ve konu-terim dağılımı Dirichlet parametreleridir. Griffiths ve Steyvers (2004) [140], çalışmalarında $\alpha=50/k$ ve $\eta=0.01$ olarak belirlemişlerdir. Deneyler için kullandıkları derlem 28.000 belgeden ve 20.000 kelimeden oluşmaktadır. Rieger vd. (2020) [141] ise deneysel çalışmalarında $\alpha= \eta=1/k$ olarak hesaplayarak deneysel incelemelerini bu değerlerle yapmışlardır. Wood vd. (2017) [142] ise yaptıkları çalışmada $\alpha=50/k$ ve $\eta=200/V$ değerlerini seçmişlerdir. Yapılan Internet aramalarında da Wood vd. (2017) tarafından kullanılan değerlerin seçilen ve tavsiye edilen değerler olduğu değerlendirilmiştir. Böylece deneysel işlemler için $\alpha=50/k$ ve $\eta=200/V$ olarak belirlenmiştir.

Terim ağırlıklandırma yöntemi olarak ONE, TF-IDF ve PMI ağırlıklandırma yöntemleri arasından kelimelerin derlemde birlikte geçmelerini skorlayan PMI yaklaşımı benimsenmiştir. Yapılan ön değerlendirme işlemlerinde 3 ağırlıklandırma yöntemiyle de 50 konu dağılımında denemeler yapılmış, PMI yönteminin gözle incelemede kelime/konu dağılımlarında daha mantıklı sonuçlar verdiği görülmüştür. Modelin eğitilmesinde tekrar sayısı da önemli bir değişkendir. Şekil 3.14.b.'de bulunan **2-200 konu, 40-2000 tekrar**la oluşturulan *perplexity* grafiği incelendiğinde konu sayısı düşükken tekrar sayısının *perplexity* skorundaki dağılımı ciddi bir şekilde etkilediği görülebilir fakat konu sayıları arttığında tekrarlar arasında *perplexity* skor farkı da azalmaktadır. Konu sayısı seçimi yapılırken sabit tekrar sayısı 500 olarak belirlenmiştir.

3.2.1.13 Konu Sayılarının Belirlenmesi

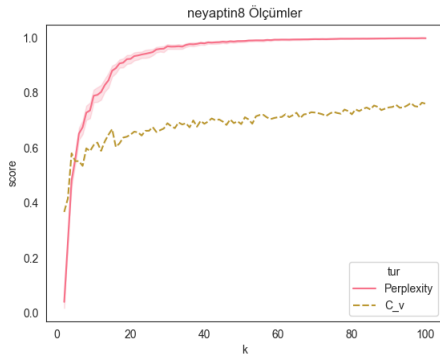
Tez kapsamında çalışılan derlemde daha önce çalışılmadığından konu sayısı ve konularda bulunan kelimelerin ağırlıkları daha önceden bilinmemektedir. Farklı bir ifadeyle ön tanımlı bir konu dermesi üzerinden çalışılmıyor olması sebebiyle konu sayısı ve bu konuların kapsamı belli değildir. Bu sebeple farklı ön işleme yöntemleriyle ve farklı parametrelerle birçok yöntem denenerek optimum konu sayısına ulaşmak amaçlanmıştır. Araştırma verileri 3. kişilerle paylaşılamadığı için ise analizler kişisel bilgisayarda gerçekleştirilmiştir. Konu sayısının belirlenmesi işi kapsamında model 2-100 konu aralığında 500 tekrarla eğitilmiş ve *perplexity* ve C_v değerlerinin birlikte değerlendirilebileceği grafikler elde edilmiştir. *Perplexity* değerleri her 100 tekrarda bir, C_v değerleri ise tekrarlar bitip model oluşunca (500 tekrar) toplanmıştır. Gerçekleştirilen yöntemle değerlerin hesaplanması işlemi Intel I7-10510U işlemci ve 16GB

sistem belleğine sahip bir kişisel bilgisayarda gerçekleştirilmiştir. Bahsedilen sistemde model eğitimi ve grafiklerin elde edilme süreci derlem büyüklüğü, konu sayısı gibi değişkenlere göre değişmekle birlikte bir derlem için ortalama 1 saat süre almaktadır. Oluşturulan derlemde bulunan belge sayısı, sözlük büyüklüğü ve kelime sayılarını gösteren veriler Tablo 3.8’de paylaşılmıştır.

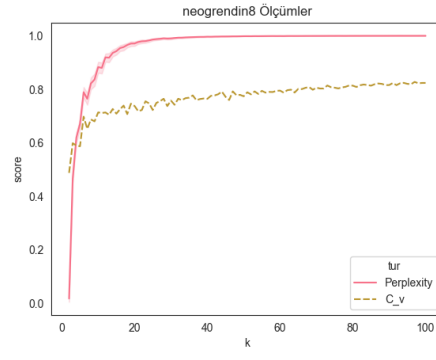
Tablo 3.8: Derlemlerde bulunan belge sayısı, sözlük büyüklüğü ve kelime sayıları.

Form Alanı	D	V	W
S1	65524	2860	310565
S2	56538	2467	188296
S3	36069	1458	158112

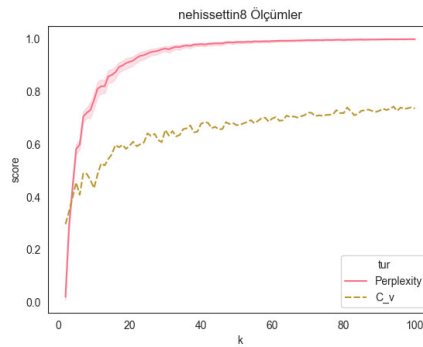
Tablo 3.8’de paylaşılan verilerin daha önce paylaşılan form verilerinden farklılaşmasının sebebi, dolgu belgelerinin derlemde çıkarılması, “1” kelime içeren belgelerin derleme alınmaması ve form alanlarının diğer form alanlarıyla karşılaştırılması faktörleridir. LDA modelinin üç alan üzerinde de eğitilmesiyle ortaya çıkan perplexity ve Cv değerlerine ait grafikler Şekil 3.17’de paylaşılmıştır.



a. S1



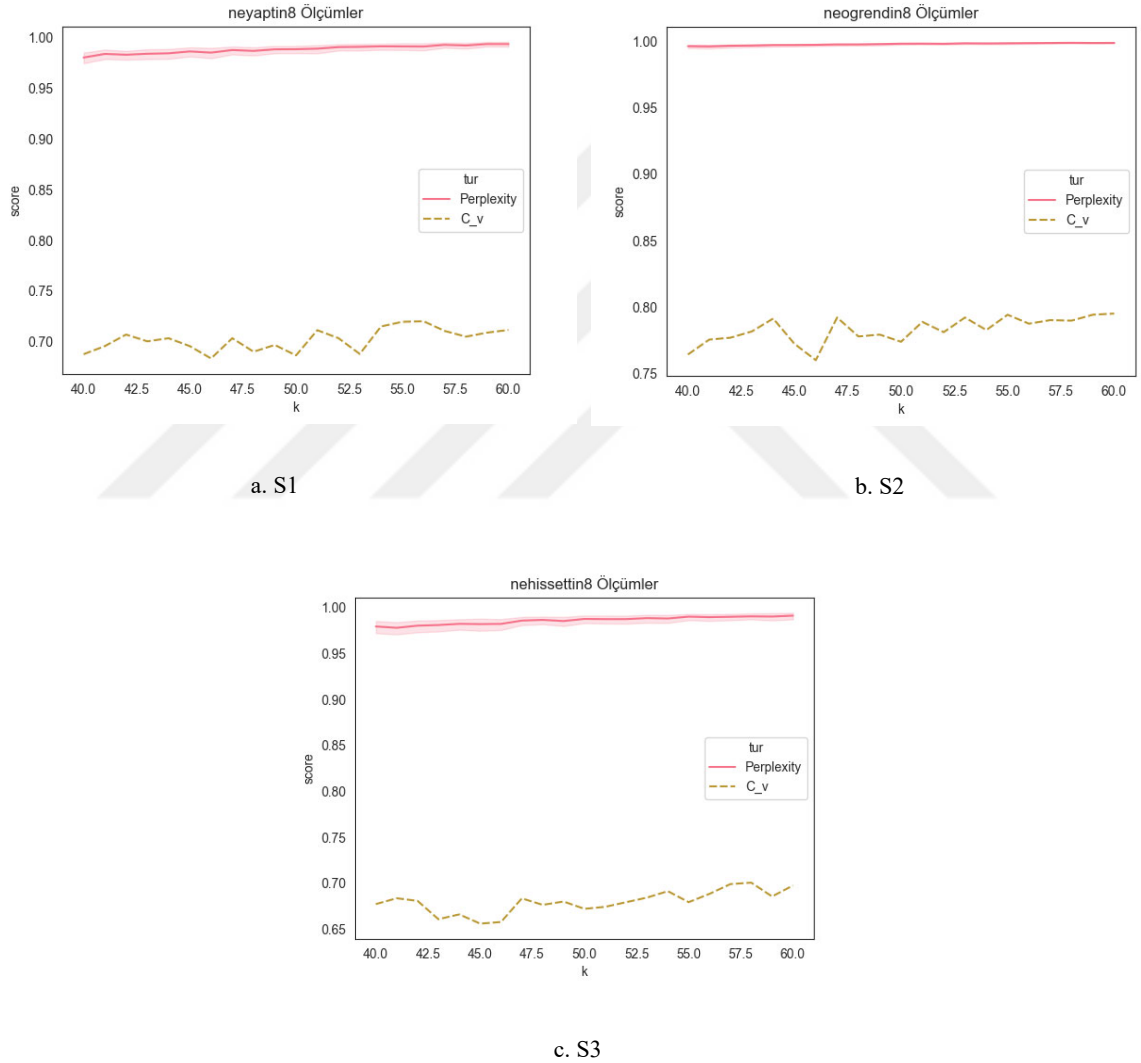
b. S2



c. S3

Şekil 3.17: Konu sayısı tespitine ait perplexity ve C_v değerleri.

Şekil 3.17’de paylaşılan grafiklerden; (a.) S1, (b.) “Ne öğrendin” ve (c.) S3 derlemlerine ait grafiklerdir. Grafiklerde kırmızı çizgiler *perplexity* değerlerini ifade etmektedir. *Perplexity* çizgilerinin etraflarında bulunan hareler 100-500 tekrar arasındaki değer değişimini göstermektedir. Sarı kesikli çizgiler ise C_v değerlerini ifade etmektedir. Grafikler *perplexity* değerlerinin düzleştiği ve 1’e yaklaştığı noktadan itibaren 20 konuyu kapsayacak şekilde yakınlaştırılmış ve C_v değerlerinin yüksek olduğu bir nokta muhtemel konu sayısı olarak belirlenmiştir. Belirli aralıklara yakınlaştırılan grafikler Şekil 3.18’de paylaşılmıştır.



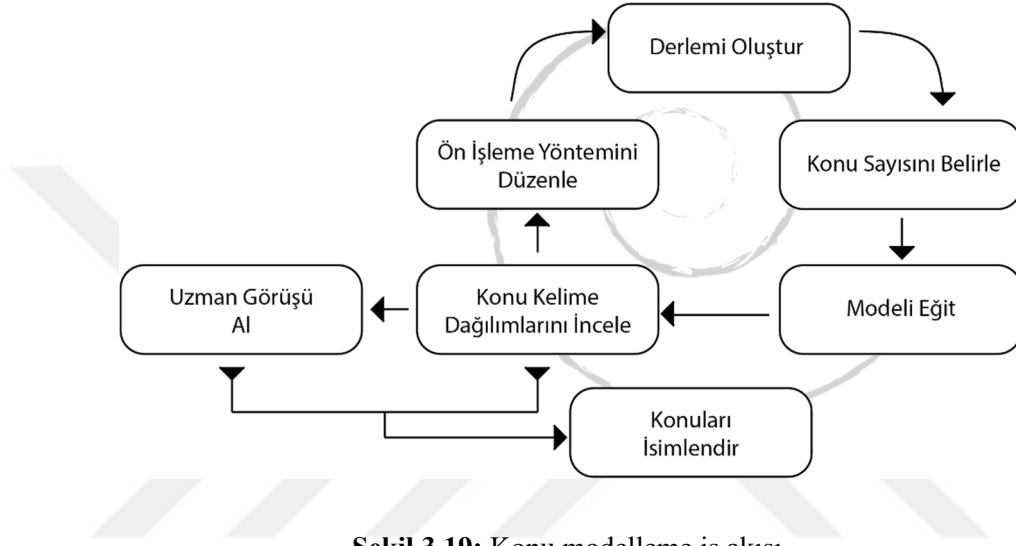
Şekil 3.18: Konu sayısı tespitine ait yakınlaştırılmış perplexity ve C_v değerleri.

Şekil 3.18’de paylaşılan grafiklerde perplexity değerinin 1’e yakın olduğu ve C_v değerlerinin de birbirine yakın olduğu ve dalgalandığı görülebilir. Bu değerlerden;

- a. S1 derlemi için C_v tepe değerinin $C_v(54) = 0.71$ olduğu ,

- b. S2 derlemi için C_v tepe değerinin $C_v(47) = 0.79$ olduğu,
- c. S3 derlemi için C_v tepe değerinin $C_v(54) = 0.69$ olduğu değerlendirilmiştir.

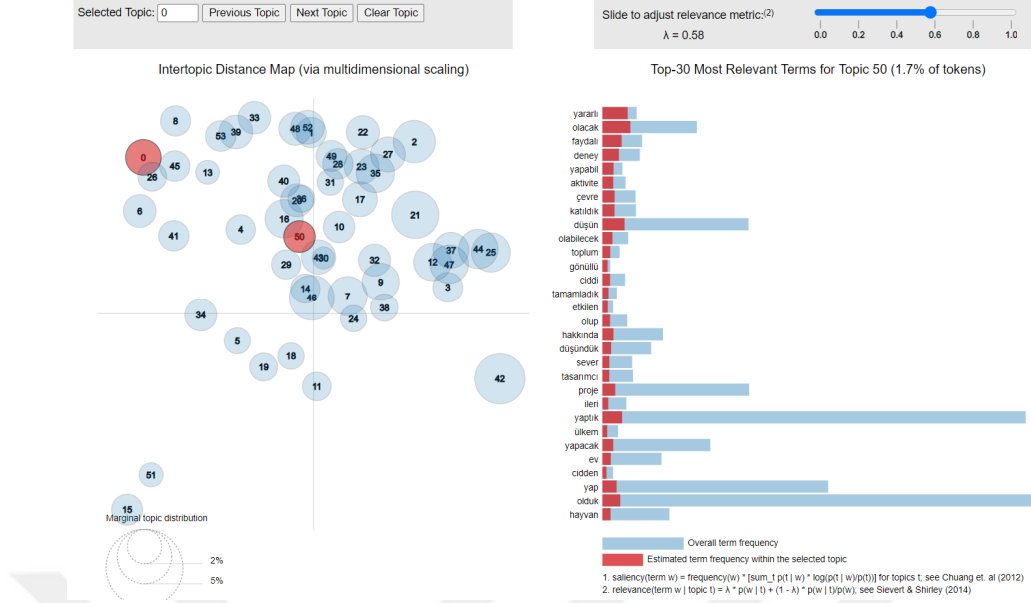
Bu değerler modelin eğitimi, konu-kelime dağılımlarının incelenmesi ve yeni dolgu kelimelerinin belirlenmesi, bir anlam verilemeyen kelimelerin belirli bir konuya birikmesiyle o konunun sistemden çıkarılması gibi faktörlerle değişiklik gösterebilir. Konuların belirlenmesi işlemleri kapsamında uygulanan iş akışı Şekil 3.19’da paylaşılmıştır.



Şekil 3.19: Konu modelleme iş akışı.

3.2.1.14 Konuların İsimlendirilmesi

Konu sayıları tespit edilerek belirlenen konu sayısında bir LDA modeli oluşturulduğunda kelimelerin konular üzerinde ağırlıkları elde edilir. Bu ağırlıkların isimlendirilmesi hususunda insan yardımına gerek duyulmaktadır. Bu sebeple S1 ve S2 derlemlerinin konularının isimlendirilmesinde Harezmi Eğitim Modeli derslerine girmiş öğretmenlerden, S3 derleminde bulunan konularının isimlendirilmesinde ise “Rehberlik ve Psikolojik Danışmanlık” branş öğretmenlerinden yardım alınmıştır. Öğretmenlere *pyLDavis* [143] kütüphanesi kullanılarak oluşturulan LDA modeli konu dağılımı görselleri gösterilerek her konu-kelime ağırlıklandırması için bir isim tahmini yapmaları istenmiş ve konular öğretmen görüşlerine göre isimlendirilmiştir. Oluşturulan konu-kelime ağırlığı gösterimi örneği Şekil 3.20’de paylaşılmıştır.



Şekil 3.20: Konu-kelime ağırlığı görsel gösterimi.

Şekil 3.20’de gösterilen konu-kelime ağırlığı görselleri ile öğretmenlerle birlikte konular tespit edilmiş ve birleştirilmesi gereken konular işaretlenmiştir.

3.2.1.15 Uzman Görüşü Alma Süreci

Her ne kadar başarımların değerleri yüksek de olsa konu modelleme algoritmaları semantik olarak anlamsız konular da ortaya koyabilir [144] bu sebeple uzman görüşleri denetimsiz bir metot olan LDA ile yapılan konu çıkarımlarının insanlar tarafından da onaylanması için oldukça önemlidir. İyi bilinen uzman görüşü alma teknikleri arasında Delphi tekniği, küçük grup tekniği (uzman paneli) ve fikir birliği geliştirme konferansları bulunmaktadır [145]. Delphi tekniği oturumlar halinde her aşamada bir önceki basamakta ulaşılan fikirlerin derinleştirildiği bir yapıda planlanarak uygulanır. Delphi Tekniği kesin kanıtın olmadığı veya çok az olduğu konularda uzmanlardan toplu bir görüş elde etme yöntemi olarak nitelendirilebilir [146]. Küçük grup tekniği ise iki aşama halinde panelistlerin belli öğeleri ya da soruları puanladığı, tartıştığı ve yeniden puanladığı bir yapıda sürdürülmektedir [145]. Araştırma kapsamında modellerin oluşturulması ve bu modellerin araştırmacı ve uzmanlar tarafından değerlendirme süreci farklı aşamalar içerdiği için karma bir metot kullanmaya gerek duyulmuştur. Bu bağlamda araştırma kapsamında uzman görüşleri alma süreci Delphi ve küçük grup teknikleri baz alınarak sürdürülmüş ve Şekil 3.19’de gösterildiği gibi modelleme süreciyle iç içe geçmiş bir şekilde uygulanmıştır. Modelleme sürecinde de anlatıldığı üzere süreç boyunca;

1. Metin ön işleme yöntemi bir önceki hatalar incelenerek yeniden düzenlenmiş,
2. Derlem oluşturulmuş,
3. Modelin eğitileceği konu sayısı belirlenmiş ve model eğitilmiş,
4. Konu kelime dağılımları incelenmiş ve
5. Uzman görüşleriyle birlikte konular isimlendirilmiştir.

Her uzmana öncelikle konu-kelime ağırlığı görsel gösterimleri hakkında bilgi verilmiştir. Sonrasında ise uzmanlara derlemlere özel sorular yöneltilmiştir. Oturumlarda ilerledikçe araştırmacı ve uzmanların fikirleri olgunlaşmış, konular isimlendirilmiş, isimler puanlanarak konu isimlerine netlik kazandırılmış ve uzman görüşü alma basamağı noktalanmıştır. Derlemlere göre uzman görüşleri için hazırlanan sorular ve alınan cevaplarla ilgili tablo EK-4'te paylaşılmıştır.

Uzman görüşü alma sürecinde;

- Öğrencilerin ders deneyimlerini yansıttıkları S1 ve S2 sorusuna yönelik olarak HEM derslerine giren ve ilçe koordinatörlüğü yapan öğretmenlerin görüşlerinden yararlanılmıştır.
- Öğrencilerin duygularını yansıttıkları S3 sorusuna yönelik olarak İstanbul ili Avcılar ilçesinde görev yapan PDR branşı öğretmenlerinden görüş alınmıştır.

4 BULGULAR

Bu aşamaya kadar veri seti üzerinde doğal dil işleme tekniklerinden nasıl yararlanılabileceği incelenmiş ve veri setinin ön işlemden nasıl geçirileceği, konu analizi yöntemlerinin nasıl uygulanacağı gibi sorulara cevap aranmıştı. İlk veri seti üzerinde daha önceki aşamalarda uygulanan işlemler sonucunda yeni bir veri seti ortaya çıkarılmıştır. Bu bölümde metin madenciliği teknikleri kullanılarak oluşturulan yeni veri seti üzerinde oluşturulan araştırma sorularına bu araştırma sorularına yönelik bulgulara yer verilmiştir.

4.1 Araştırma Soruları

Zhang ve Zhang [9] tarafından veri madenciliğinin geleneksel istatistiksel yöntemlerden ayrıldığı noktanın, geleneksel istatistiksel yöntemlerin varsayıma dayalı olması ve hipotezlerini veriler üzerinde test etmesi, veri madenciliğinin ise modellerin ve hipotezlerin veriden çıkarılarak test edildiği keşif odaklı yaklaşımlar sunması olarak değerlendirilmektedir. Başka bir deyişle istatistik insan odaklıyken, veri madenciliği veri odaklıdır [9].

Bu sebeple yeni veri seti üzerinde bilginin keşfedilmesi amacıyla yeni araştırma soruları oluşturulmuştur. Bu araştırma soruları aşağıda paylaşılmıştır.

1. Yorum uzunluğuna göre;
 - a. İlk, orta ve lise kademesi öğrencilerinin oluşturdukları belgeler uzunluk bakımından istatistiksel olarak birbirlerinden ayrılmakta mıdır?
2. Duygu kutuplarına göre;
 - a. S3 derleminden oluşturulan duygu kutuplarına göre ilk, orta ve lise kademesi öğrencilerinin duygu kutupları istatistiksel olarak birbirinden ayrılmakta mıdır?
3. Konu analizine göre;
 - a. S1 derleminde bulunan belgeler konulara ayrılabilir mi?
 - b. S1 derleminde bulunan konular belirli özelliklerine göre temalar altında birleştirilebilir mi?
 - c. S2 derleminde bulunan belgeler konulara ayrılabilir mi?

- d. S2 derleminde bulunan konular belirli özelliklerine göre temalar altında birleştirilebilir mi?
 - e. S3 derleminde bulunan belgeler konulara ayrılabilir mi?
 - f. S3 derleminde bulunan konular belirli özelliklerine göre temalar altında birleştirilebilir mi?
4. Konu analizi sonucunda çıkan konu etiketleri üzerinden keşifsel analiz yapıldığında;
- a. S1 derlemi için;
 - i. Üretilen konu etiketlerinin okul dereceleri arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - ii. Üretilen konu etiketlerinin ders periyotları arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - b. S2 derlemi için;
 - i. Üretilen konu etiketlerinin ders periyotları arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - ii. Üretilen konu etiketlerinin okul dereceleri arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - c. S3 derlemi için;
 - i. Üretilen konu etiketlerinin ders periyotları arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - ii. Üretilen konu etiketlerinin okul dereceleri arasında dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?
 - iii. Üretilen konu etiketleriyle S1 derlemi konu etiketleri içinde dağılımında istatistiksel olarak anlamlı ilişkiler var mıdır?

4.2 Birinci Araştırma Sorusuna Yönelik Bulgular

Öğrencilerin derlemlerde bulunan yorumları oluşturmak için kullandıkları kelimelerin sayıları karşılaştırılmıştır. Derlemlere göre oluşturulan belge-kelime uzunlukları betimsel istatistikleri Tablo 4.1’de paylaşılmıştır.

Tablo 4.1: Derlemlere göre belge uzunlukları betimsel istatistikleri.

Derlem	Okul Derecesi	Min	Maks	$\bar{x}$	Ss	Kolmogorov Smirnow
S1	İlk	0	92	3,87	3,21	,200*
	Orta	0	104	5,39	4,511	,174*
	Lise	0	47	4,64	3,661	,179*
	Genel	0	104	4,77	4,031	,183*
S2	İlk	0	41	4,03	2,768	,213*
	Orta	0	41	4,91	3,437	,176*
	Lise	0	40	4,32	3,140	,191*
	Genel	0	41	4,52	3,208	,194*
S3	İlk	0	29	2,80	2,462	,266*
	Orta	0	49	3,82	3,469	,209*
	Lise	0	45	3,01	2,990	,248*
	Genel	0	49	3,34	3,128	,230*

Tablo 4.1’de paylaşılan veriler ışığında S1 derlemi için ortalama 4,77 kelime, S2 derlemi için ortalama 4,52 ve S3 derlemi için ortalama 3,34 kelime kullanıldığı görülebilir. Kolmogorov-Smirnow normallik testi sonuçlarına göre veriler normal dağılım göstermemektedir ($p < 0.05$). Bu sebeple bu veri seti üzerinde yapılacak istatistiksel analizler parametrik olmayan testler kullanılarak gerçekleştirilmiştir.

Okul derecelerine göre yorum uzunluklarının istatistiksel olarak anlamlı farklılık gösterip göstermediğini ortaya koymak amacıyla Mann-Whitney U-Testi testleri gerçekleştirilmiştir. Testlere yönelik istatistiksel veriler Tablo 4.2’de paylaşılmıştır.

Tablo 4.2: Derlemlere göre belge uzunlukları ve okul grupları arasında Mann-Whitney-U testi sonuçları.

Derlemler	Gruplar	N	Sıralar Ortalaması	Sıralar Toplamı
S1	İlk	26311	29217,73	768747784,00
	Orta	41826	37120,73	1552611669,00
	Lise	18616	24336,14	453041536,00
S2	İlk	26311	30676,09	807118578,00
	Orta	41826	36203,34	1514240875,00
	Lise	18616	23011,38	428379861,50
S3	İlk	26311	30467,44	801628858,50
	Orta	41826	36334,59	1519730594,50
	Lise	18616	22109,75	411595163,50

Derlemlerin kelime uzunlukları ve okul dereceleri arasında yapılan Mann-Whitney U testi sonuçları Tablo 4.3’de paylaşılmıştır.

Tablo 4.3: Derlemlere göre belge uzunlukları ve okul grupları arasında Mann-Whitney U testi sonuçları.

Derlemler	Gruplar	U	Z	p
S1	İlk-Orta	422600268,000	-51,509	,000
	İlk-Lise	210051076,000	-26,044	,000
	Orta-Lise	352175091,000	-18,884	,000
S2	İlk-Orta	460971062,000	-36,106	,000
	İlk-Lise	234712750,500	-7,627	,000
	Orta-Lise	344635550,500	-22,771	,000
S3	İlk-Orta	455481342,500	-38,733	,000
	İlk-Lise	238308127,500	-5,035	,000
	Orta-Lise	320564753,000	-35,431	,000

Tablo 4.3’de bulunan verilere göre tüm gruplar arasında istatistiksel olarak anlamlı farklar bulunduğu görülmektedir ($p < 0.05$). Sıra ortalamaları dikkate alındığında tüm derlemler için en uzun yorumların orta okul grubu öğrencileri tarafından yazıldığı, en kısa yorumların ise lise grubunda bulunan öğrenciler tarafından yazıldığı sonucuna ulaşılmıştır.

4.3 İkinci Araştırma Sorusuna Yönelik Bulgular

S3 derlemi üzerinde duygu kutbu analizi SWNet duygu kutbu sözlüğüne göre yapılmıştır. Duygu kutbu analizi sonucunda ortaya çıkan duygu kutuplarına ait betimsel istatistikler ve Ki-Kare testi sonuçları Tablo 4.4’da paylaşılmıştır.

Tablo 4.4: S3 sorusu duygu kutbu analizi Ki-Kare testi sonuçları.

Duygu Kutbu	İlk			Orta			Lise			TOPLAM	
	f	ef	%	f	ef	%	f	ef	%	N	%
Olumsuz	1892	2504	7,34%	4104	3924	10,16%	2139	1707	12,17%	8135	9,71%
Olumlu	23882	23270	92,66%	36293	36473	89,84%	15433	15865	87,83%	75608	90,29%
Toplam	25774		100%	40397		100%	17572		100%	83743	100%
$\chi^2=295,771, p=0.000$											

Tablo 4.4’da paylaşılan veriler incelendiğinde olumlu duygu kutbuna sahip yorumlar ilk derece okullarda 92% yoğunluğa sahipken, orta dereceli okullarda 89%, lise seviyesinde ise 87% seviyesinde olduğu, olumsuz duygu yüküne sahip yorumların ise ilk derece okullarında 7%, orta derece okullarında 10% ve lise dereceli okullarda 12% olduğu görülebilir. Ki-Kare testi

sonucunda ilk, orta ve lise kademeleri arasında olumlu ve olumsuz duygu yükleri bakımından istatistiksel olarak anlamlı bir farklılık ($X^2=295,771$, $p<0.05$) olduğundan söz edilebilir.

4.4 Üçüncü Araştırma Sorusuna Yönelik Bulgular

4.1.1 S1 Derlemi

Derlem malzeme ve yöntem bölümünde bahsedilen aşamalardan geçirilerek modellemeye hazırlanmış ve LDA algoritması işletilerek modellenmiştir. Modelleme ile ilgili değişkenler ve modelin başarımlar değeri Tablo 4.5’de paylaşılmıştır.

Tablo 4.5: S1 derlemi LDA analizi özet bilgileri.

Derlem	S1
k	54
Tekrar	2000
Dirichlet α (50/k)	.89
Dirichlet η (200/V)	,0699
Örnekleme Yöntemi	Gibbs
D	63938
V	2860
W	307986
Cv	.73

Modelleme işleminden sonra ortaya çıkan konu-kelime dağılımlarına ait ilk 2 kelime ve ağırlıkları Tablo 4.6’de, tablonun tamamı ise Ek 1’de paylaşılmıştır.

Tablo 4.6: S1 derlemi konu-kelime dağılımları özet tablosu.

t	Kelime	Ağırlık	Kelime	Ağırlık	t	Kelime	Ağırlık	Kelime	Ağırlık
0	anket	0,057	sonuç	0,030	27	koyduk	0,022	elma	0,020
1	problem	0,042	alt	0,031	28	devam	0,089	ettik	0,066
2	oynadık	0,051	oyun	0,049	29	yedik	0,040	yemek	0,030
3	bilgi	0,024	hakkında	0,017	30	fikir	0,047	arkadaş	0,036
4	ektik	0,060	toprak	0,038	31	led	0,046	yaktık	0,042
5	gittik	0,050	park	0,025	32	izledik	0,075	video	0,064
6	dans	0,036	oynadık	0,029	33	ilgili	0,026	sağlıklı	0,021
7	oynadık	0,038	balon	0,034	34	geçen	0,026	kalan	0,023
8	gittik	0,051	bahçe	0,026	35	ders	0,035	ilk	0,026
9	şenlik	0,058	hakkında	0,043	36	sistem	0,022	site	0,018
10	hikaye	0,055	yazdık	0,049	37	gittik	0,064	git	0,023
11	hafta	0,021	yapacak	0,017	38	tiyatro	0,064	rol	0,033
12	bahçe	0,046	kuş	0,025	39	sihirli	0,039	kule	0,030
13	çözüm	0,047	problem	0,036	40	boya	0,055	boyadık	0,039
14	göz	0,028	oynadık	0,026	41	soru	0,046	sorduk	0,026

Tablo 4.6 (devam): S1 derlemi konu-kelime dağılımları özet tablosu.

t	Kelime	Ağırlık	Kelime	Ağırlık	t	Kelime	Ağırlık	Kelime	Ağırlık
15	ilgili	0,035	bağımlılık	0,028	42	oynadık	0,040	oyun	0,032
16	hoca	0,039	bize	0,025	43	günü	0,029	oyna	0,025
17	oynadık	0,072	oyun	0,065	44	afiş	0,045	hazırladık	0,035
18	soru	0,054	çözdük	0,048	45	oda	0,021	bişim	0,021
19	ders	0,056	form	0,041	46	arduino	0,042	devre	0,029
20	grup	0,025	ayır	0,014	47	düğüm	0,023	oynadık	0,021
21	çizgi	0,025	film	0,024	48	fotoğraf	0,026	çekildik	0,018
22	araştırma	0,028	sunum	0,026	49	ağaç	0,027	yapıştırdık	0,013
23	kitap	0,075	ayraç	0,028	50	kahoot	0,035	uygulama	0,031
24	kutu	0,041	geri dönüşüm	0,028	51	konuştuk	0,022	hakkında	0,019
25	oynadık	0,051	kulak	0,048	52	oyun	0,039	oynadık	0,029
26	müze	0,072	gittik	0,061	53	kurt	0,044	oynadık	0,027

Ortaya çıkan modele göre modelin tüm belgeler için bir sonuç üretimi (*infering*) sağlanmış ve belge-konu birikimi ile ilgili istatistiki bilgiler Tablo 4.7’de paylaşılmıştır.

Tablo 4.7: S1 derlemi konu-belge birikimi tablosu.

k	N	k	N	k	N
0	997	18	1470	36	834
1	1842	19	868	37	721
2	888	20	1755	38	1379
3	795	21	1154	39	184
4	1087	22	1101	40	925
5	542	23	382	41	961
6	907	24	1422	42	1055
7	1002	25	694	43	869
8	999	26	1049	44	2624
9	1147	27	184	45	1076
10	3923	28	1057	46	950
11	1684	29	942	47	947
12	883	30	1385	48	809
13	1271	31	1655	49	791
14	1095	32	3542	50	927
15	1396	33	954	51	1194
16	1396	34	844	52	1692
17	1388	35	1764	53	548
				Toplam	63950

Konu-kelime dağılımları öğretmenlerle birlikte incelenerek kodlanmıştır. Öğretmenler 5E öğretim modeli basamaklarına “Bilgi işlemsel düşünme” basamağının da eklendiğinde konuları açıklayabileceğini dile getirmişlerdir. Öğretmenlerle birlikte oluşturulan etiketler Tablo 4.8’de paylaşılmıştır

Tablo 4.8: S1 derlemi 5E modeline göre konu etiketleri.

Konu	Basamak	Konu	Basamak	Konu	Basamak
0	Keşfetme	18	Dikkat çekme	36	Bilgi işlemsel düşünme
1	Keşfetme	19	Değerlendirme	37	Dikkat çekme
2	Dikkat çekme	20	Keşfetme	38	Derinleştirme
3	Keşfetme	21	Açıklama	39	Dikkat çekme
4	Derinleştirme	22	Açıklama	40	Derinleştirme
5	Dikkat çekme	23	Keşfetme	41	Dikkat çekme
6	Dikkat çekme	24	Derinleştirme	42	Bilgi işlemsel düşünme
7	Dikkat çekme	25	Dikkat çekme	43	Anneler günü
8	Dikkat çekme	26	Dikkat çekme	44	Derinleştirme
9	Değerlendirme	27	Keşfetme	45	Bilgi işlemsel düşünme
10	Keşfetme	28	Derinleştirme	46	Bilgi işlemsel düşünme
11	Değerlendirme	29	Dikkat çekme	47	Dikkat çekme
12	Derinleştirme	30	Açıklama	48	Derinleştirme
13	Keşfetme	31	Bilgi işlemsel düşünme	49	Derinleştirme
14	Dikkat çekme	32	Dikkat çekme	50	Dikkat çekme
15	Keşfetme	33	Dikkat çekme	51	Keşfetme
16	Dikkat çekme	34	Derinleştirme	52	Dikkat çekme
17	Dikkat çekme	35	Derinleştirme	53	Bilgi işlemsel düşünme

4.1.2 S2 Derlemi

Derlem malzeme ve yöntem bölümünde bahsedilen aşamalardan geçirilerek modellemeye hazırlanmış ve LDA algoritması işletilerek modellenmiştir. Modelleme ile ilgili değişkenler modelin başarımlar değeri Tablo 4.9’te paylaşılmıştır.

Tablo 4.9: S2 derlemi LDA analizi özet bilgileri.

Derlem	S2
k	47
Dirichlet α (50/k)	1.06
Dirichlet η (200/V)	.0676
Örnekleme Yöntemi	Gibbs
D	69482
V	2958
W	256683
C _v	.79

Modelleme işleminden sonra ortaya çıkan konu-kelime dağılımlarına ait ilk 2 kelime ve ağırlıkları Tablo 4.10’da, tablonun tamamı ise Ek 2’de paylaşılmıştır.

Tablo 4.10: S2 derlemi konu-kelime dağılımları özet tablosu.

k	kelime	Ağırlık	kelime	Ağırlık	k	kelime	Ağırlık	kelime	Ağırlık
0	kullanma	0,03	program	0,03	24	balıkkılçık	0,06	trafik	0,06
1	soru	0,02	yeni	0,02	25	farklı	0,03	herkes	0,02
2	engelli	0,06	sorun	0,03	26	led	0,07	devre	0,05
3	afiş	0,05	hazırlama	0,05	27	yapma	0,04	boyama	0,04
4	oyun	0,10	oynama	0,04	28	cezeri	0,03	çocuk	0,02
5	kirlilik	0,05	hayvan	0,02	29	işe	0,05	yaradık	0,05
6	renk	0,08	kirlilik	0,03	30	kelime	0,03	ingilizce	0,03
7	atık	0,04	geri	0,03	31	hesaplama	0,03	şekil	0,03
8	sorun	0,04	problem	0,02	32	grafik	0,05	yapacak	0,04
9	akım	0,03	direnç	0,03	33	robot	0,06	yapma	0,03
10	ısınma	0,06	küresel	0,05	34	internet	0,03	teknoloji	0,03
11	yapma	0,04	kodlama	0,04	35	çalışma	0,09	grup	0,06
12	sensör	0,10	mesafe	0,05	36	zarar	0,05	doğay	0,03
13	problem	0,09	sorun	0,05	37	çöp	0,04	yer	0,02
14	hayvan	0,06	tarihi	0,04	38	teknoloji	0,05	zarar	0,04
15	bir	0,02	hafta	0,02	39	bitki	0,06	tohum	0,04
16	geri_dönüşüm	0,04	atık	0,03	40	ağaç	0,03	yapma	0,03
17	besin	0,10	piramit	0,06	41	yapma	0,08	yapım	0,03
18	yapma	0,04	uçak	0,03	42	meyve	0,05	film	0,05
19	yapma	0,03	kunduz	0,02	43	meslek	0,02	görev	0,01
20	sağlıklı	0,04	sağlıksız	0,03	44	problem	0,03	sorun	0,03
21	deprem	0,05	gerektik	0,04	45	enerji	0,09	rüzgar	0,04
22	fikir	0,04	teknik	0,03	46	slogan	0,06	hikaye	0,06
23	israf	0,05	zorbalık	0,04					

Ortaya çıkan modele göre modelin tüm belgeler için bir sonuç üretimi (*infering*) sağlanmış ve belge-konu birikimi ile ilgili istatistiki bilgiler Tablo 4.11’de paylaşılmıştır.

Tablo 4.11: S2 derlemi konu-belge birikimi tablosu.

k	N	k	N	k	N
0	1813	16	1678	32	1341
1	1208	17	238	33	1680
2	1166	18	1289	34	1568
3	2687	19	965	35	2394
4	1895	20	1122	36	1462
5	1167	21	1287	37	816
6	944	22	1189	38	914
7	1797	23	1161	39	1352
8	2470	24	1168	40	1269
9	1039	25	2291	41	1307
10	1312	26	1786	42	1007
11	1535	27	1303	43	1448
12	1535	28	1185	44	1249
13	4596	29	1178	45	992
14	1323	30	1363	46	2177
15	1972	31	940		
				Toplam	69578

Konu-kelime dağılımları öğretmenlerle birlikte incelenerek isimlendirilmiş ve Tablo 4.12’de paylaşılmıştır.

Tablo 4.12: S2 derlemi konularının öğretmenler tarafından isimlendirilmesi.

k	Konu	k	Konu
0	Bilgi işlemsel düşünme	24	Trafik kuralları
1	Tanışma etkinliği	25	Farklı bakış açıları
2	Engellilerin sorunları	26	Bilgi işlemsel düşünme
3	Tasarım	27	Boyama etkinliği
4	Zeka oyunları	28	Cezeri'nin icatları
5	Çevre kirliliği	29	Malzeme bilgisi
6	Işık kirliliği ve renkler	30	İngilizce kelimeler
7	Geri dönüşüm	31	Geometri
8	Problem çözme	32	İstatistik
9	Elektrik	33	Bilgi işlemsel düşünme
10	İklim değişikliği	34	Teknoloji bağımlılığı
11	Bilgi işlemsel düşünme	35	İşbirlikli çalışma
12	Bilgi işlemsel düşünme	36	Çevreyi koruma
13	Problem çözme	37	Çevreyi koruma
14	Doğal ve kültürel miras	38	Teknoloji bağımlılığı
15	Geçen dersin tekrarı	39	Tarım
16	Geri dönüşüm	40	Doğa
17	Besin zinciri	41	Üretim faaliyetleri
18	Uygulama	42	Sağlıklı beslenme
19	Bilgi işlemsel düşünme	43	Meslek seçimi
20	Sağlıklı beslenme	44	Problem çözme
21	Deprem	45	Yenilenebilir enerji kaynakları
22	Hikaye yazma	46	HİS cümlesi ve sloganı oluşturma
23	Olumsuz davranışlar		

4.1.3 S3 Derlemi

Derlem malzeme ve yöntem bölümünde bahsedilen aşamalardan geçirilerek modellemeye hazırlanmış ve LDA algoritması işletilerek modellenmiştir. Modelleme ile ilgili değişkenler ve modelin başarımlar değeri Tablo 4.13’de paylaşılmıştır.

Tablo 4.13: S3 derlemi LDA analizi özet bilgileri.

Derlem	S3
k	54
Dirichlet α (50/k)	.92
Dirichlet η (200/V)	.1295
Örnekleme Yöntemi	Gibbs
D	36465
V	1544
W	139101
Cv	.70

Modelleme işleminden sonra ortaya çıkan konu-kelime dağılımlarına ait ilk 2 kelime ve ağırlıkları Tablo 4.14’te, tablonun tamamı ise Ek 2’de paylaşılmıştır.

Tablo 4.14: S3 derlemi konu-kelime dağılımları özet tablosu.

t	Kelime	Ağırlık	Kelime	Ağırlık	t	Kelime	Ağırlık	Kelime	Ağırlık
0	yanlış	0,039	kork	0,037	27	sorun	0,029	problem	0,027
1	anne	0,040	insan	0,026	28	duy	0,075	gurur	0,070
2	süper	0,091	harika	0,081	29	heyecan	0,082	öğrenme	0,071
3	öğretmen	0,054	teşekkür	0,043	30	bazen	0,047	biraz	0,042
4	sunum	0,033	heyecanlan	0,030	31	eğlen	0,047	yaparken	0,037
5	bilim	0,107	kendi	0,088	32	yapacak	0,042	heyecanlan	0,033
6	kendi	0,065	mühendis	0,057	33	bir	0,040	değişik	0,032
7	gururlu	0,080	heyecanlı	0,066	34	kötü	0,023	iyi	0,023
8	ettik	0,047	etmek	0,039	35	hikaye	0,098	yazarken	0,063
9	biraz	0,060	sıkıcı	0,043	36	kötü	0,020	ben	0,018
10	ders	0,088	eğlenceli	0,078	37	fikir	0,025	kendi	0,022
11	oyun	0,060	oynadık	0,038	38	biraz	0,059	film	0,046
12	kendi	0,106	özel	0,056	39	olma	0,040	müze	0,025
13	üzül	0,045	hayvan	0,030	40	bilgi	0,073	sahip	0,058
14	geçti	0,093	eğlenceli	0,083	41	ilk	0,048	sonra	0,043
15	maket	0,043	yaparken	0,035	42	yeni	0,036	bilgi	0,035
16	etkinlik	0,039	bir	0,029	43	yapma	0,027	bir	0,023
17	sorun	0,066	çözüm	0,056	44	heyecanlı	0,092	meraklı	0,065
18	kedi	0,040	yaptık	0,027	45	oyun	0,051	gezi	0,043
19	afiş	0,056	soru	0,041	46	video	0,058	oyun	0,025
20	hayal	0,046	kırıklık	0,037	47	eğlendik	0,045	hoca	0,036
21	heyecan	0,063	dolu	0,062	48	form	0,050	yavaş	0,040
22	hafta	0,044	ders	0,034	49	çalışma	0,077	grup	0,063
23	oyun	0,059	öğrendik	0,033	50	geçir	0,096	vakit	0,085
24	şarkı	0,034	gel	0,029	51	ders	0,033	keşke	0,025
25	şaşır	0,042	üzül	0,033	52	hoş	0,053	git	0,042
26	yeni	0,138	öğren	0,087	53	heyecan	0,192	sevinç	0,127

Ortaya çıkan modele göre modelin tüm belgeler için bir sonuç üretimi (*infering*) sağlanmış ve belge-konu birikimi ile ilgili istatistiki bilgiler Tablo 4.15'te paylaşılmıştır

Tablo 4.15: S3 derlemi konu-belge birikimi tablosu.

k	N	k	N	k	N
0	527	18	484	36	694
1	463	19	551	37	426
2	422	20	337	38	534
3	646	21	567	39	411
4	693	22	824	40	461
5	614	23	557	41	653
6	726	24	560	42	647
7	1096	25	414	43	396
8	584	26	1282	44	567
9	1379	27	524	45	544
10	1480	28	607	46	427
11	1478	29	374	47	331
12	393	30	501	48	323
13	1064	31	517	49	991
14	2736	32	669	50	726
15	781	33	538	51	1101
16	871	34	291	52	337
17	1029	35	554		
				Toplam	36702

Konu-kelime dağılımları “Rehberlik ve Psikolojik Danışmanlık” alanı öğretmenleriyle birlikte incelenerek kodlanmış ve Tablo 4.16’de paylaşılmıştır.

Tablo 4.16: S3 derlemi duygu etiketleri.

Konu	Basamak	Konu	Basamak	Konu	Basamak
0	ihtiyat	18	empati	36	iğrenme
1	empati	19	heyecanlı	37	özgüven
2	coşkunluk	20	hayal kırıklığına uğramış	38	sıkılmış
3	müteşekkir	21	enerji dolu	39	ilgili
4	heyecanlı	22	sıkılmış	40	sorumluluk
5	başarılı	23	şaşırmış	41	sıkılmış
6	başarılı	24	sıkılmış	42	meraklı
7	gururlu	25	üzgün	43	meraklı
8	özgür	26	memnun	44	güçlü
9	sıkılmış	27	empati	45	istekli
10	neşeli	28	gururlu	46	heyecanlı
11	eğlenmiş	29	kafası karışmış	47	eğlenmiş
12	özgüven	30	kızgın	48	sıkılmış
13	empati	31	eğlenmiş	49	aidiyet
14	eğlenmiş	32	heyecanlı	50	eğlenmiş
15	heyecanlı	33	şaşırmış	51	üzgün
16	sevecen	34	kötü	52	memnun
17	yaratıcı	35	eğlenmiş	53	---

4.5 Dördüncü Araştırma Sorusuna Yönelik Bulgular

Keşifsel analiz başlığı altında, üretilen konu etiketleri ile kurum dereceleri ve derslerin işlendiği haftalar arasında anlamlı istatistiksel ilişkiler olup-olmaması durumu araştırılmıştır.

4.1.4 S1 Derlemi

Konu etiketlerinin frekansları ve yüzdelik değerleri Tablo 4.17’te paylaşılmıştır.

Tablo 4.17: S1 derlemi konu etiketleri frekans ve yüzde değerleri.

Etkinlik Türü	<i>f</i>	%
Dikkat çekme	23169	26,7
Keşfetme	13739	15,8
Açıklama	3640	4,2
Derinleştirme	13585	15,7
Değerlendirme	3699	4,3
Bilgi işlemsel düşünme	6118	7,1
Etiketlenemeyen	22803	26,3
Toplam	63950	100,0

Tablo 4.17’de paylaşılan verilere göre 5E modeline göre üretilen etiketlerden dikkat çekme basamağının 26,7%, keşfetme basamağının 15,8%, derinleştirme basamağının 15,7%, açıklama

basamağının 4,2% ve değerlendirme basamağının ise 4,3%'ünü kapsadıkları görülebilir. Harezmi modelinin bir ayağı olan bilgi işlemsel düşünme basamağı da yıl içindeki derslerde 7,1% yer almıştır.

Tablo 4.18: S1 derlemi okul derecelerine göre ki-kare testi sonuçları.

Etiket	Ölçüm	İlk	Orta	Lise	Genel
Dikkat çekme	f	7484 _a	11658 _b	4027 _c	23169
	ef	5862,5	11986,2	5320,3	23169,0
	%	17,3%	13,2%	10,2%	13,5%
	ar	26,4	-4,6	-21,7	
Keşfetme	f	5190 _a	14246 _b	8042 _c	27478
	ef	6952,8	14215,5	6309,8	27478,0
	%	12,0%	16,1%	20,5%	16,1%
	ar	-26,7	0,4	27,1	
Açıklama	f	2571 _a	5184 _a	3165 _b	10920
	ef	2763,1	5649,3	2507,6	10920,0
	%	5,9%	5,9%	8,1%	6,4%
	ar	-4,4	-9,2	15,5	
Derinleştirme	f	15380 _a	28156 _b	10804 _c	54340
	ef	13749,7	28112,2	12478,1	54340,0
	%	35,5%	31,8%	27,5%	31,8%
	ar	19,5	0,5	-20,7	
Değerlendirme	f	3575 _a	8860 _b	6060 _c	18495
	ef	4679,8	9568,2	4247,0	18495,0
	%	8,3%	10,0%	15,4%	10,8%
	ar	-19,8	-11,0	33,6	
Bilgi işlemsel düşünme	f	9096 _a	20418 _b	7194 _c	36708
	ef	9288,2	18990,5	8429,3	36708,0
	%	21,0%	23,1%	18,3%	21,5%
	ar	-2,6	16,8	-17,3	
$X^2=3715,943$, $sd=10$, $p<0.05$					

Ki-Kare testi sonuçlarına göre okul dereceleri ile konu etiketleri arasında anlamlı bir ilişki olduğundan söz edilebilir ($X^2= 3715,943$, $p<0.05$).

Tablo 4.18'de bulunan tablo üzerinden inceleme yapıldığında;

- Dikkat çekme ile etiketlenen derslerin ilk seviye okullarda 17,3, orta seviye okullarda 13,2% ve lise seviyesindeki okullarda ise 10,2% seviyesinde olduğu ve tüm okul seviyeleri değerlerinin birbirinden anlamlı düzeyde farklı olduğu,
- Keşfetme ile etiketlenen derslerin ilk seviye okullarda 12,0%, orta seviye okullarda 16,1% ve lise seviyesindeki okullarda ise 20,5% seviyesinde yer bulduğu ve tüm değerlerin birbirinden anlamlı düzeyde farklı olduğu,

- Derslerin işlendiği 2 ders dönemi ve 32 hafta süren süreç 8'er haftalık 4 periyota ayrılmıştır. Etiketlerin bu periyotlar arasındaki değişimi ile ilgili Ki-Kare testi sonuçları Tablo 4.19'te paylaşılmıştır.

[illegible][illegible]

Tablo 4.19’da bulunan veriler analiz edildiğinde konu etiketlerinin periyotlardaki dağılımları ile ilgili anlamlı ilişkilerden söz edilebilir ($X^2=21357,997$, $p<0.05$). Değerler üzerinden inceleme yapıldığında;

- Dikkat çekme ile etiketlenen belgelerin frekanslarının tüm periyotlarda birbirinden ayrıştığı ve 2. Periyotta en yüksek seviyesine ulaştığı,
- Keşfetme ile etiketlenen belgelerin frekanslarının tüm periyotlarda birbirinden ayrıştığı ve 1. periyotta 40,9% ile en yüksek seviyesine ulaştığı,
- Açıklama ile etiketlenen belgelerin frekanslarının tüm periyotlarda birbirinden ayrıştığı; 1. Periyotta en yüksek değerinde (9,8%) olan oranın her periyotta düşerek 4. Periyotta en düşük seviyesine ulaştığı (3,4%),
- Derinleştirme ile etiketlenen belgelerin frekanslarının tüm periyotlar için birbirinden ayrıştığı 1. periyotta 23,5%, 2. periyotta 30,6%, 3. periyotta 34,8% ve 4. periyotta 39,9% ile en yüksek seviyesine ulaştığı,
- Değerlendirme ile etiketlenen belgelerin frekanslarının tüm periyotlar için birbirinden ayrıştığı 2. Periyot (11,6%) ve 4. Periyotta (16,5%) yüksek seviyede olduğu,
- Bilgi işlemsel düşünme eğitimi ile etiketlenen belgelerin frekanslarının 1. periyotta 13,5%, 2. periyotta 21,7%, 3. periyotta 28,2% olduğu ve 4. periyotta düşüş yaşayarak 22,7% seviyesinde yer buldukları değerlendirilebilir.

4.1.5 S2 Derlemi

S2 derlemi konu etiketlere ait frekans ve yüzde verileri Tablo 4.20’de paylaşılmıştır

Tablo 4.20: S2 derlemi konu etiketleri frekans ve yüzde verileri.

Etiket	f	%	Etiket	f	%
0 Çevre	14047	20,19	19 Üretim faaliyetleri	1307	1,9
1 Bilgi işlemsel düşünme	9314	13,4	20 Boyama etkinliği	1303	1,9
2 Problem çözme	8315	12,0	21 Uygulama	1289	1,9
3 Geri dönüşüm	3475	5,0	22 Deprem	1287	1,8
4 Tasarım	2687	3,9	23 Doğa	1269	1,8
5 Teknoloji bağımlılığı	2482	3,6	24 Tanışma etkinliği	1208	1,7
6 İşbirlikli çalışma	2394	3,4	25 Hikaye yazma	1189	1,7
7 Farklı bakış açıları	2291	3,3	26 Cezeri'nin icatları	1185	1,7
8 Çevreyi koruma	2278	3,3	27 Malzeme bilgisi	1178	1,7
9 HİS cümlesi ve sloganı oluşturma	2177	3,1	28 Trafik kuralları	1168	1,7
10 Sağlıklı beslenme	2129	3,1	29 Çevre kirliliği	1167	1,7
11 Geçen dersin tekrarı	1972	2,8	30 Engellilerin sorunları	1166	1,7
12 Zeka oyunları	1895	2,7	31 Olumsuz davranışlar	1161	1,7
13 Meslek seçimi	1448	2,1	32 Elektrik	1039	1,5
14 İngilizce kelimeler	1363	2,0	33 Yenilenebilir enerji kaynakları	992	1,4
15 Tarım	1352	1,9	34 Işık kirliliği	944	1,4
16 İstatistik	1341	1,9	35 Geometri	940	1,4
17 Doğal ve kültürel miras	1323	1,9	36 Besin zinciri	238	0,3
18 İklim değişikliği	1312	1,9			

Tablo 4.20’da bulunan verilere göre öğrenciler en fazla “Bilgi işlemsel düşünme” alanında öğrenme gerçekleştirdiklerini (13,4%), sonrasında ise “Problem Çözme” (12,0%) ve “Geri dönüşüm” (5%) alanlarını öğrendiklerini dile getirmişlerdir.

Çevre ile ilgili konu etiketleri birleştirilmiş, sonrasında etiketler ders periyotlarına göre gruplanmıştır. Konu etiketlerinin ders periyotları arasındaki dağılımını gösterir veriler Tablo 4.21’de paylaşılmıştır.

Tablo 4.21: S2 derlemi konu etiketlerinin ders periyotları arasında dağılımı.

Etiket	1.Periyot	2.Periyot	3.Periyot	4.Periyot	Toplam
Çevre	2835	4277	4272	2663	14047
Bilgi işlemsel düşünme	1094	2215	3531	2474	9314
Problem çözme	5371	1425	860	659	8315
Tasarım	734	890	589	474	2687
Teknoloji bağımlılığı	742	827	626	287	2482
İşbirlikli çalışma	1095	409	470	420	2394
Farklı bakış açıları	990	609	401	291	2291
HİS cümlesi ve sloganı oluşturma	1354	346	308	169	2177
Sağlıklı beslenme	268	836	675	350	2129
Geçen dersin tekrarı	474	527	516	455	1972
Zeka oyunları	369	484	602	440	1895
Meslek seçimi	308	420	360	360	1448
İngilizce kelimeler	254	455	333	321	1363
Tarım	97	431	499	325	1352
İstatistik	155	430	304	452	1341
Üretim faaliyetleri	144	392	469	302	1307

Tablo 4.21 (devam): S2 derlemi konu etiketlerinin ders periyotları arasında dağılımı.

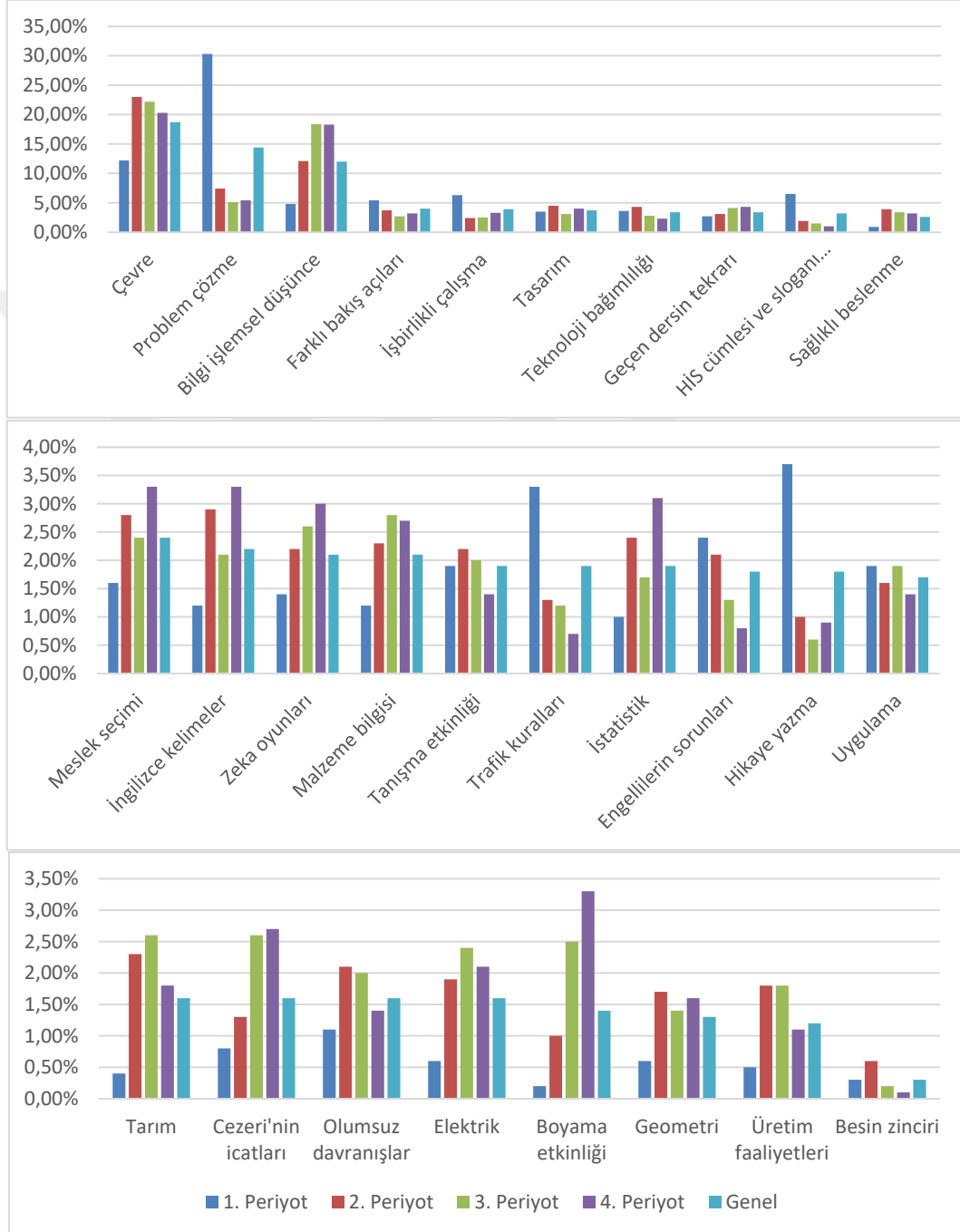
Etiket	1.Periyot	2.Periyot	3.Periyot	4.Periyot	Toplam
Uygulama	389	365	350	185	1289
Tanışma etkinliği	388	361	279	180	1208
Hikaye yazma	737	191	150	111	1189
Cezeri'nin icatları	151	270	450	314	1185
Malzeme bilgisi	184	321	436	237	1178
Trafik kuralları	613	252	204	99	1168
Engellilerin sorunları	458	348	255	105	1166
Olumsuz davranışlar	242	441	341	137	1161
Elektrik	138	300	367	234	1039
Geometri	108	359	264	209	940
Besin zinciri	36	107	60	35	238

Tablo 4.21'e göre 1. Periyot için en ağır 5 etiket sırasıyla "Problem çözme", "Çevre", "HİS cümlesi ve sloganı oluşturma", "İşbirlikli çalışma" ve "Bilgi işlemsel düşünme" olmuştur. 2. Periyot için en ağır 5 etiket sırayla "Çevre", "Bilgi işlemsel düşünme", "Problem çözme", "Tasarım" ve "Sağlıklı beslenme" etiketlerinden oluşmaktadır. 3. Periyotta en ağır 5 etiket sırasıyla "Çevre", "Bilgi işlemsel düşünme", "Problem çözme", "Sağlıklı beslenme" ve "Teknoloji bağımlılığı" etiketlerinden oluşmaktadır. 4. Periyotta ise en ağır etiketler "Çevre", "Bilgi işlemsel düşünme", "Problem çözme", "Boyama etkinliği" ve "Tasarım" olarak sıralanabilir. Genel ağırlıklara göre sıralandığında ise en ağır etiketler "Çevre", "Bilgi işlemsel düşünme", "Problem çözme", "Tasarım" ve "Teknoloji bağımlılığı" olarak sıralanabilir. Her etiketin periyotlar içinde dağılımlarını incelemek amacıyla Ki-Kare testi yapılmış ve test sonuçları Tablo 4.22'de paylaşılmıştır.

Tablo 4.22: S2 derlemi konu etiketleri ki-kare testi sonuçları.

Etiket	1. Periyot				2. Periyot				3. Periyot				4. Periyot				Genel		
	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%
Çevre	21381a	32892	12,2%	-86,82	32243b	26250	23,0%	48,12	27671c	23297	22,2%	36,52	14746d	13601	20,3%	11,75	96041	96041	18,7%
Problem çözme	53286a	25335	30,3%	234,09	10417b	20219	7,4%	-87,41	6327c	17944	5,1%	-107,71	3944d	10476	5,4%	-74,46	73974	73974	14,4%
Bilgi işlemsel düşünme	8503a	21128	4,8%	-114,20	16932b	16862	12,1%	0,68	22961c	14964	18,4%	80,08	13295c	8737	18,3%	56,12	61691	61691	12,0%
Farklı bakış açıları	9448a	6987	5,4%	37,05	5236b	5576	3,7%	-5,45	3390c	4949	2,7%	-25,98	2327d	2889	3,2%	-11,52	20401	20401	4,0%
İşbirlikli çalışma	11075a	6837	6,3%	64,48	3345b	5456	2,4%	-34,20	3141b	4842	2,5%	-28,66	2402c	2827	3,3%	-8,80	19963	19963	3,9%
Tasarım	6068a	6589	3,5%	-8,06	6345b	5258	4,5%	17,92	3907c	4667	3,1%	-13,02	2918d	2724	4,0%	4,08	19238	19238	3,7%
Teknoloji bağımlılığı	6332a	5999	3,6%	5,39	5986b	4788	4,3%	20,67	3526c	4249	2,8%	-12,97	1673d	2481	2,3%	-17,81	17517	17517	3,4%
Geçen dersin tekrarı	4670a	5916	2,7%	-20,32	4349b	4721	3,1%	-6,47	5113c	4190	4,1%	16,66	3142c	2446	4,3%	15,44	17274	17274	3,4%
HİS cümlesi ve 105logan oluşturma	11426a	5692	6,5%	95,29	2616b	4542	1,9%	-34,09	1880c	4031	1,5%	-39,58	697d	2354	1,0%	-37,47	16619	16619	3,2%
Sağlıklı beslenme	1512a	4630	0,9%	-57,27	5460b	3695	3,9%	34,51	4207c	3280	3,4%	18,86	2341c	1915	3,2%	10,66	13520	13520	2,6%
Meslek seçimi	2876a	4212	1,6%	-25,69	3985b	3361	2,8%	12,78	3046c	2983	2,4%	1,34	2390d	1742	3,3%	16,98	12297	12297	2,4%
İngilizce kelimeler	2130a	3836	1,2%	-34,35	4062b	3062	2,9%	21,44	2625c	2717	2,1%	-2,06	2385d	1586	3,3%	21,88	11202	11202	2,2%
Zeka oyunları	2399a	3747	1,4%	-27,45	3113b	2990	2,2%	2,66	3211c	2654	2,6%	12,56	2217d	1549	3,0%	18,51	10940	10940	2,1%
Malzeme bilgisi	2117a	3698	1,2%	-32,41	3236b	2951	2,3%	6,21	3450c	2619	2,8%	18,85	1995c	1529	2,7%	12,99	10798	10798	2,1%
Tanışma etkinliği	3344a	3409	1,9%	-1,38	3116b	2720	2,2%	8,98	2489a	2414	2,0%	1,76	1004c	1410	1,4%	-11,77	9953	9953	1,9%
Trafik kuralları	5823a	3333	3,3%	53,70	1876b	2660	1,3%	-18	1504c	2361	1,2%	-20,45	529d	1378	0,7%	-24,93	9732	9732	1,9%
İstatistik	1714a	3256	1,0%	-33,64	3428b	2599	2,4%	19,26	2087c	2306	1,7%	-5,30	2279d	1347	3,1%	27,69	9508	9508	1,9%
Engellilerin sorunları	4153a	3199	2,4%	20,99	3002b	2553	2,1%	10,52	1611c	2266	1,3%	-15,95	575d	1323	0,8%	-22,40	9341	9341	1,8%
Hikaye yazma	6422a	3172	3,7%	71,81	1465b	2532	1,0%	-25,09	721c	2247	0,6%	-37,32	654d	1312	0,9%	-19,78	9262	9262	1,8%
Uygulama	3337a	3028	1,9%	6,99	2194b	2416	1,6%	-5,36	2316a	2145	1,9%	4,29	994c	1252	1,4%	-7,94	8841	8841	1,7%
Tarım	718a	2891	0,4%	-50,26	3163b	2307	2,3%	21,07	3233c	2048	2,6%	30,34	1328d	1196	1,8%	4,17	8442	8442	1,6%
Cezeri'nin icatları	1332a	2871	0,8%	-35,72	1855b	2291	1,3%	-10,78	3259c	2033	2,6%	31,49	1937c	1187	2,7%	23,68	8383	8383	1,6%
Olumsuz davranışlar	1878a	2853	1,1%	-22,69	2944b	2277	2,1%	16,54	2474b	2021	2,0%	11,68	1034c	1180	1,4%	-4,62	8330	8330	1,6%
Elektrik	992a	2787	0,6%	-42,26	2638b	2224	1,9%	10,38	3010c	1974	2,4%	27,01	1497d	1152	2,1%	11,05	8137	8137	1,6%
Boyama etkinliği	397a	2499	0,2%	-52,23	1438b	1995	1,0%	-14,73	3096c	1770	2,5%	36,46	2367d	1034	3,3%	45,09	7298	7298	1,4%
Geometri	1050a	2198	0,6%	-30,38	2454b	1754	1,7%	19,73	1735c	1557	1,4%	5,23	1178b	909	1,6%	9,70	6417	6417	1,3%
Üretim faaliyetleri	867a	2186	0,5%	-35,01	2487b	1745	1,8%	20,98	2258b	1548	1,8%	20,85	771c	904	1,1%	-4,80	6383	6383	1,2%
Besin zinciri	535a	604	0,3%	-3,49	904b	482	0,6%	22,56	256c	428	0,2%	-9,58	70d	250	0,1%	-12,31	1765	1765	0,3%
X2=117524,195, Sd=81, p<0.05																			

Tablo 4.22’de bulunan veriler incelendiğinde her etiketin periyotlardaki değerleri arasında anlamlı farklılıklar olduğu değerlendirilebilir ($X^2=117524,195$, $p<0.05$). Konu etiketlerinin periyotlar arasında dağılımını görsel olarak incelemek amacıyla periyotlar arası frekansları ile ilgili grafikler oluşturularak Şekil 4.1’de paylaşılmıştır.



Şekil 4.1: S2 derlemi konu etiketlerinin periyotlar arasında dağılım grafikleri.

S2 derlemi konu etiketlerinin okul kademeleri arasında dağılımını incelemek amacıyla etiketler okul türlerine göre gruplanmıştır. Konu etiketlerinin okul kademeleri arasındaki dağılımını gösterir veriler Tablo 4.23'te paylaşılmıştır.

Tablo 4.23: S2 derlemi konu etiketlerinin okul dereceleri arasında dağılımı.

Etiket	İlk		Orta		Lise		Genel	
	f	%	f	%	f	%	f	%
Çevre	4819	24,34%	6356	18,22%	2872	19,27%	14047	20,19%
Bilgi işlemsel düşünme	2646	13,36%	4943	14,17%	1725	11,58%	9314	13,39%
Problem çözme	1779	8,98%	4314	12,37%	2222	14,91%	8315	11,95%
Tasarım	566	2,86%	1437	4,12%	684	4,59%	2687	3,86%
Teknoloji bağımlılığı	575	2,90%	1183	3,39%	724	4,86%	2482	3,57%
İşbirlikli çalışma	413	2,09%	1460	4,19%	521	3,50%	2394	3,44%
Farklı bakış açıları	542	2,74%	1206	3,46%	543	3,64%	2291	3,29%
HİS cümlesi ve sloganı oluşturma	539	2,72%	1179	3,38%	459	3,08%	2177	3,13%
Sağlıklı beslenme	1000	5,05%	789	2,26%	340	2,28%	2129	3,06%
Geçen dersin tekrarı	462	2,33%	1041	2,98%	469	3,15%	1972	2,83%
Zeka oyunları	970	4,90%	737	2,11%	188	1,26%	1895	2,72%
Meslek seçimi	306	1,55%	725	2,08%	417	2,80%	1448	2,08%
İngilizce kelimeler	389	1,96%	756	2,17%	218	1,46%	1363	1,96%
Tarım	429	2,17%	560	1,61%	363	2,44%	1352	1,94%
İstatistik	288	1,45%	655	1,88%	398	2,67%	1341	1,93%
Üretim faaliyetleri	545	2,75%	537	1,54%	225	1,51%	1307	1,88%
Boyama etkinliği	425	2,15%	688	1,97%	190	1,28%	1303	1,87%
Uygulama	280	1,41%	703	2,02%	306	2,05%	1289	1,85%
Tanışma etkinliği	376	1,90%	593	1,70%	239	1,60%	1208	1,74%
Hikaye yazma	261	1,32%	662	1,90%	266	1,79%	1189	1,71%
Cezeri'nin icatları	497	2,51%	520	1,49%	168	1,13%	1185	1,70%
Malzeme bilgisi	133	0,67%	694	1,99%	351	2,36%	1178	1,69%
Trafik kuralları	314	1,59%	648	1,86%	206	1,38%	1168	1,68%
Engellilerin sorunları	194	0,98%	715	2,05%	257	1,72%	1166	1,68%
Olumsuz davranışlar	329	1,66%	612	1,75%	220	1,48%	1161	1,67%
Elektrik	345	1,74%	551	1,58%	143	0,96%	1039	1,49%
Geometri	298	1,51%	476	1,36%	166	1,11%	940	1,35%
Besin zinciri	80	0,40%	137	0,39%	21	0,14%	238	0,34%

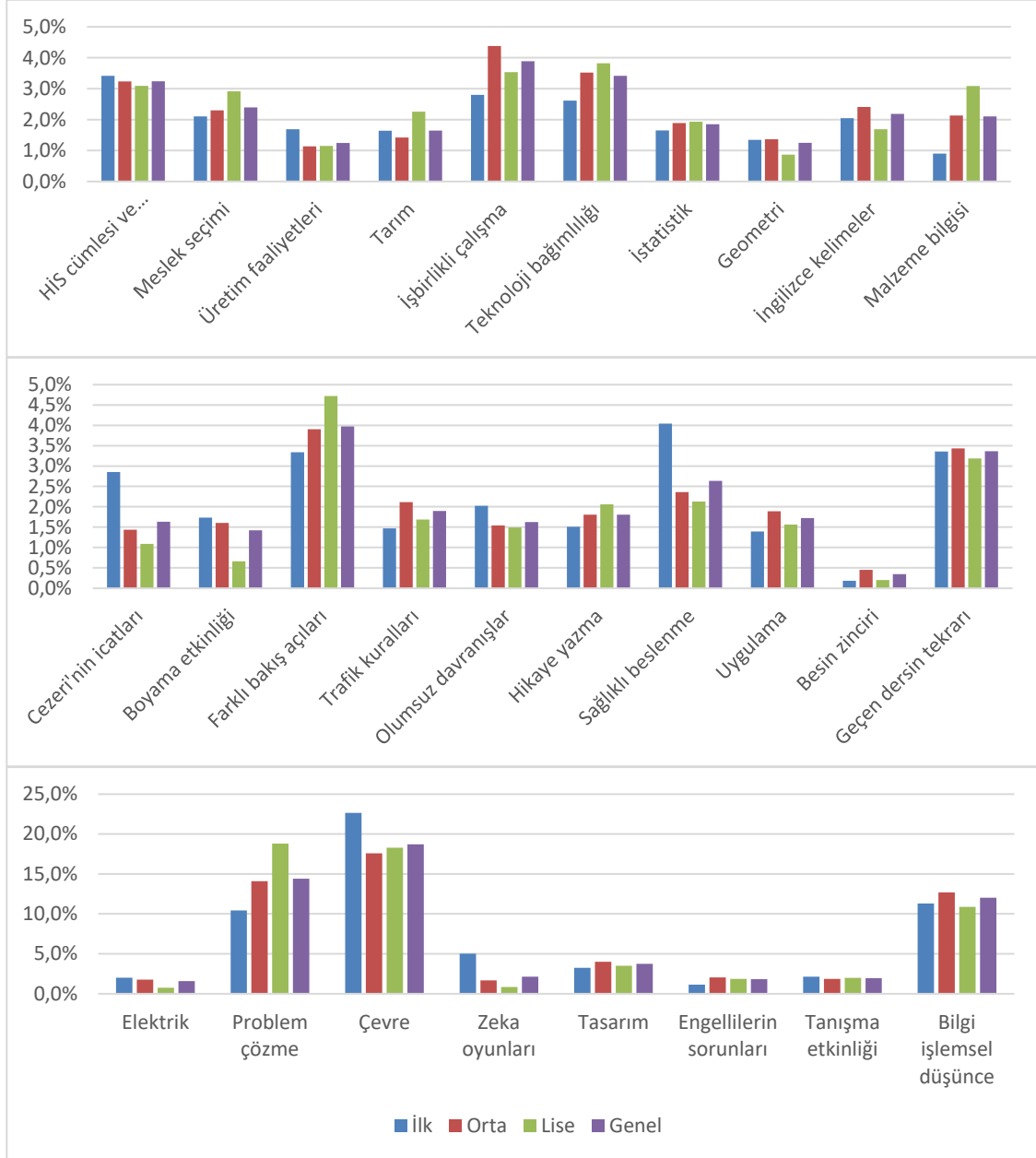
Tablo 4.23'te paylaşılan verilere göre ilk derece okullarda en baskın etiketlerin sırasıyla “Çevre”, “Bilgi işlemsel düşünme”, “Problem çözme”, “Sağlıklı beslenme” ve “Zeka oyunları” olduğu değerlendirilebilir. Orta dereceli okullarda “Çevre”, “Bilgi işlemsel düşünme”, “Problem çözme”, “İşbirlikli çalışma” ve “Tasarım” etiketlerinin ağırlıklı olduğu değerlendirilebilir. Lise seviyesi okullarda ise “Çevre”, “Problem çözme”, “Bilgi işlemsel düşünme”, “Teknoloji bağımlılığı” ve “Tasarım” konularının ağırlıklı olduğu değerlendirilebilir. S2 derlemi konu etiketlerinin okul derecelerine göre dağılımını incelemek amacıyla Ki-Kare testi yapılmıştır. Ki-Kare testi sonuçları Tablo 4.24'de paylaşılmıştır.

Tablo 4.24: S2 derlemi konu etiketleri ve okul dereceleri arasında ki-kare testi sonuçları.

Etiket	İlk				Orta				Lise				Genel	
	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	%
HİS cümlesi ve 108logan oluşturma	3368 _a	3195	3,4%	3,46	9790 _b	9800	3,2%	-0,16	3461 _b	3624	3,1%	-3,11	16619	3,2%
Meslek seçimi	2074 _a	2364	2,1%	-6,72	6963 _b	7251	2,3%	-5,35	3260 _c	2681	2,9%	12,79	12297	2,4%
Üretim faaliyetleri	1664 _a	1227	1,7%	13,96	3434 _b	3764	1,1%	-8,45	1285 _b	1392	1,1%	-3,26	6383	1,2%
Tarım	1621 _a	1623	1,6%	-0,06	4298 _b	4978	1,4%	-15,18	2523 _c	1841	2,3%	18,13	8442	1,6%
İşbirlikli çalışma	2761 _a	3838	2,8%	-19,73	13248 _b	11772	4,4%	21,66	3954 _c	4353	3,5%	-6,98	19963	3,9%
Teknoloji bağımlılığı	2580 _a	3368	2,6%	-15,37	10661 _b	10330	3,5%	5,18	4276 _c	3820	3,8%	8,50	17517	3,4%
İstatistik	1626 _a	1828	1,6%	-5,30	5719 _b	5607	1,9%	2,36	2163 _b	2073	1,9%	2,25	9508	1,9%
Geometri	1326 _a	1234	1,3%	2,94	4122 _a	3784	1,4%	8,63	969 _b	1399	0,9%	-13,09	6417	1,3%
İngilizce kelimeler	2019 _a	2154	2,0%	-3,26	7294 _b	6606	2,4%	13,37	1889 _c	2443	1,7%	-12,81	11202	2,2%
Malzeme bilgisi	887 _a	2076	0,9%	-29,34	6457 _b	6368	2,1%	1,77	3454 _c	2355	3,1%	25,90	10798	2,1%
Cezeri'nin icatları	2817 _a	1612	2,9%	33,68	4349 _b	4943	1,4%	-13,31	1217 _c	1828	1,1%	-16,29	8383	1,6%
Boyama etkinliği	1710 _a	1403	1,7%	9,18	4849 _b	4304	1,6%	13,07	739 _c	1591	0,7%	-24,34	7298	1,4%
Farklı bakış açıları	3295 _a	3922	3,3%	-11,37	11822 _b	12030	3,9%	-3,03	5284 _c	4449	4,7%	14,46	20401	4,0%
Trafik kuralları	1449 _a	1871	1,5%	-10,96	6396 _b	5739	2,1%	13,67	1887 _c	2122	1,7%	-5,83	9732	1,9%
Olumsuz davranışlar	1996 _a	1601	2,0%	11,06	4666 _b	4912	1,5%	-5,53	1668 _b	1816	1,5%	-3,97	8330	1,6%
Hikaye yazma	1487 _a	1781	1,5%	-7,81	5470 _b	5462	1,8%	0,18	2305 _c	2020	2,1%	7,25	9262	1,8%
Sağlıklı beslenme	3988 _a	2599	4,0%	30,72	7149 _b	7973	2,4%	-14,59	2383 _c	2948	2,1%	-11,93	13520	2,6%
Uygulama	1376 _a	1700	1,4%	-8,81	5715 _b	5213	1,9%	10,94	1750 _c	1928	1,6%	-4,62	8841	1,7%
Besin zinciri	181 _a	339	0,2%	-9,58	1359 _b	1041	0,4%	15,42	225 _a	385	0,2%	-9,23	1765	0,3%
Geçen dersin tekrarı	3313 _{a, b}	3321	3,4%	-0,16	10394 _b	10186	3,4%	3,27	3567 _a	3767	3,2%	-3,74	17274	3,4%
Elektrik	1981 _a	1564	2,0%	11,82	5323 _b	4798	1,8%	11,92	833 _c	1774	0,7%	-25,47	8137	1,6%
Problem çözme	10293 _a	14222	10,4%	-39,62	42628 _b	43622	14,1%	-8,03	21053 _c	16130	18,8%	47,38	73974	14,4%
Çevre	22342 _a	18464	22,6%	35,22	53227 _b	56635	17,6%	-24,79	20472 _c	20942	18,3%	-4,08	96041	18,7%
Zeka oyunları	4949 _a	2103	5,0%	69,79	5061 _b	6451	1,7%	-27,31	930 _c	2386	0,8%	-34,07	10940	2,1%
Tasarım	3190 _a	3699	3,2%	-9,48	12123 _b	11345	4,0%	11,63	3925 _c	4195	3,5%	-4,80	19238	3,7%
Engellilerin sorunları	1116 _a	1796	1,1%	-18,01	6158 _b	5508	2,0%	13,79	2067 _c	2037	1,8%	0,76	9341	1,8%
Tanışma etkinliği	2112 _a	1913	2,1%	5,10	5627 _b	5869	1,9%	-4,98	2214 _c	2170	2,0%	1,07	9953	1,9%
Bilgi işlemel düşünme	11155 _a	11860	11,3%	-7,68	38368 _b	36379	12,7%	17,36	12168 _c	13452	10,9%	-13,35	61691	12,0%

X²= 16976,939, Sd=54, p<0.05

Tablo 4.24’de bulunan veriler ışığında konu etiketleri ile okul kademelerinde dağılımında anlamlı ilişkiler olduğu değerlendirilebilir ($X^2=16976,939$, $Sd=54$, $p<0.05$). S3 derlemi konu etiketlerinin okul kademeleri içinde dağılımını incelemek amacıyla oluşturulan grafikler Şekil 4.2’de paylaşılmıştır.



Şekil 4.2: S2 derlemi konu etiketlerinin okul türleri arasında dağılım grafikleri.

4.1.6 S3 Derlemi

S3 derlemi, daha önce bahsedildiği üzere etiketleme işlemi için en zorlu derlemidir. Bu derleme ait verilerden 50051 adedi vektör bakımından zengin olmamaları dolayısıyla işleme

alınamamıştır. Etiketlenebilen 36702 adet belgeye ait frekans ve yüzde değerlerini gösterir veriler Tablo 4.25’da paylaşılmıştır.

Tablo 4.25: S3 derlemi konu etiketleri frekans ve yüzde verileri.

k	Etiket	f	%	k	Etiket	f	%
1	eğlenmiş	6342	17,3	16	iğrenme	694	1,9
2	sıkılmış	4273	11,6	17	müteşekkik	646	1,8
3	heyecanlı	3121	8,5	18	özgür	584	1,6
4	empati	2535	6,9	19	enerji dolu	567	1,5
5	gururlu	1703	4,6	20	güçlü	567	1,5
6	memnun	1619	4,4	21	istekli	544	1,5
7	üzgün	1515	4,1	22	ihtiyat	527	1,4
8	neşeli	1480	4,0	23	kızgın	501	1,4
9	başarılı	1340	3,7	24	sorumluluk	461	1,3
10	şaşırmış	1095	3,0	25	coşkunluk	422	1,1
11	meraklı	1043	2,8	26	ilgili	411	1,1
12	yaratıcı	1029	2,8	27	kafası karışmış	374	1,0
13	aidiyet	991	2,7	28	hayal kırıklığına uğramış	337	0,9
14	sevecen	871	2,4	29	kötü	291	0,8
15	özgüven	819	2,2				
Toplam						36702	

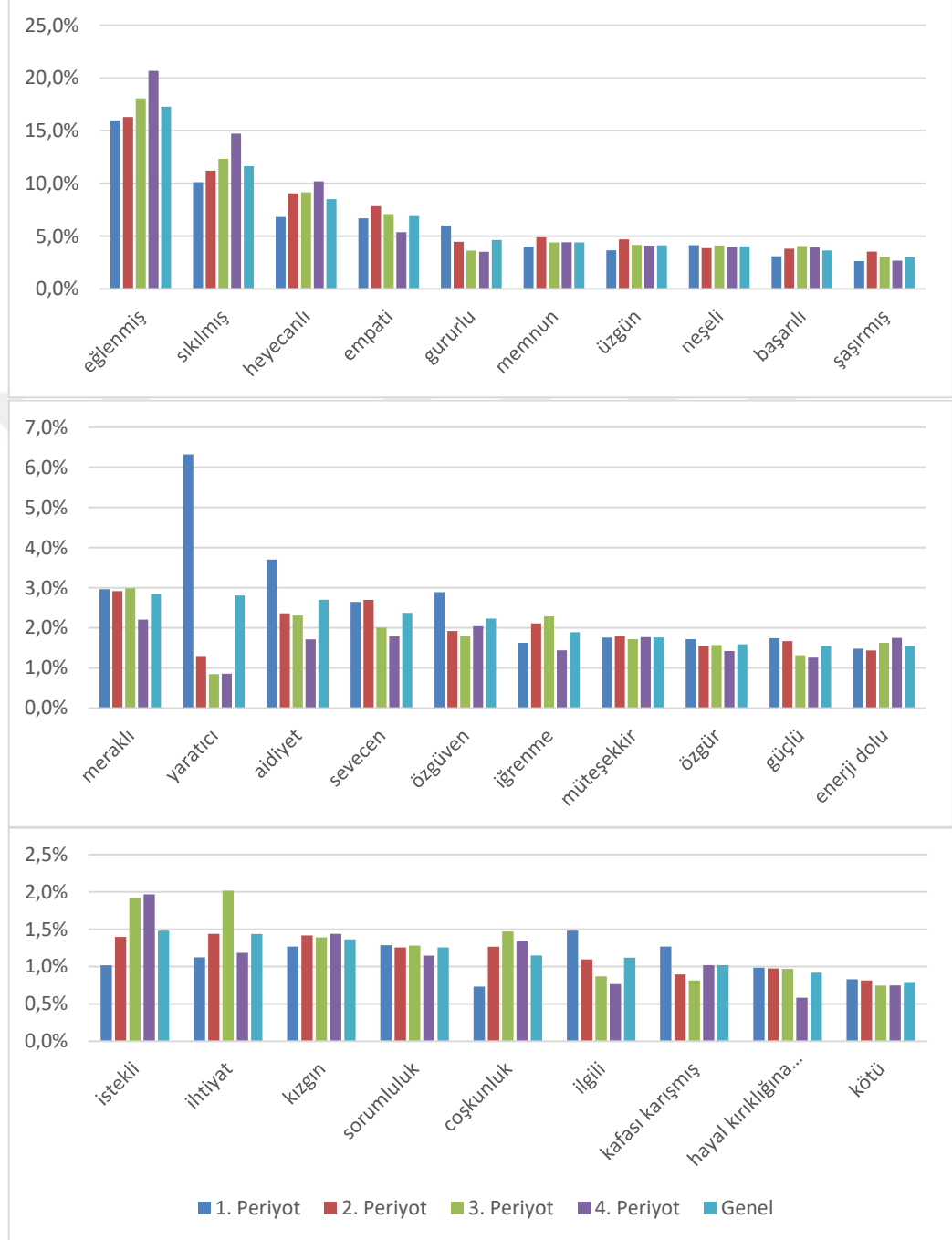
Tablo 4.25’da bulunan veriler incelendiğinde en ağır duygunun “eğlenmiş” (17,3%) olduğu, “sıkılmış” (11,6%) etiketine sahip belgelerin derlemin ikinci en ağır konusu olduğu, sonrasında ise “heyecanlı” (8,5%) ve “empati” (6,9%) duygularının bulunduğu görülebilir.

S3 derlemi konu etiketlerinin ders periyotları arasında değişimini incelemek amacıyla Ki-Kare testi gerçekleştirilmiş ve sonuçları Tablo 4.26’da paylaşılmıştır.

Tablo 4.26: S3 derlemi konu etiketleri ve ders periyotları arasında Ki-Kare testi sonuçları.

Etiket	1. Periyot				2. Periyot				3. Periyot				4. Periyot				Genel	
	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	%
eğlenmiş	21604 _a	24396	10,9%	-23,7	17831 _b	18839	11,7%	-9,2	17842 _c	16717	13,2%	10,7	12485 _d	9810	15,7%	31,1	69762	12,3%
aidiyet	22295 _a	16981	11,3%	52,9	11515 _b	13113	7,5%	-17,1	10143 _b	11636	7,5%	-16,6	4606 _c	6829	5,8%	-30,3	48559	8,6%
meraklı	15288 _a	15319	7,7%	-0,3	12180 _b	11829	8,0%	3,9	11256 _c	10497	8,3%	8,8	5082 _d	6160	6,4%	-15,4	43806	7,7%
memnun	12844 _a	14720	6,5%	-19,9	12662 _b	11367	8,3%	14,8	10270 _c	10087	7,6%	2,2	6318 _d	5920	7,9%	5,8	42094	7,4%
sıkılmış	11187 _a	13449	5,7%	-25,0	10035 _b	10385	6,6%	-4,2	9963 _c	9215	7,3%	9,3	7272 _d	5408	9,1%	28,3	38457	6,8%
üzgün	11225 _a	13245	5,7%	-22,5	11675 _b	10228	7,6%	17,3	9350 _c	9076	6,9%	3,4	5625 _c	5326	7,1%	4,6	37875	6,7%
şaşırmış	7452 _a	8807	3,8%	-18,3	8096 _b	6801	5,3%	18,8	6256 _c	6035	4,6%	3,3	3381 _d	3542	4,2%	-3,0	25185	4,4%
iğrenme	7200 _a	8737	3,6%	-20,9	7560 _b	6747	4,9%	11,9	7380 _c	5987	5,4%	21,1	2844 _a	3513	3,6%	-12,5	24984	4,4%
güçlü	9416 _a	8724	4,8%	9,4	7304 _a	6737	4,8%	8,3	5192 _b	5978	3,8%	-11,9	3036 _b	3508	3,8%	-8,8	24948	4,4%
istekli	5625 _a	8561	2,8%	-40,2	6255 _b	6611	4,1%	-5,2	7740 _c	5866	5,7%	28,7	4860 _d	3443	6,1%	26,6	24480	4,3%
sorumluluk	6320 _a	6449	3,2%	-2,0	5000 _a	4980	3,3%	0,3	4600 _b	4419	3,4%	3,2	2520 _a	2593	3,2%	-1,6	18440	3,3%
b																		
yaratıcı	13209 _a	6117	6,7%	114,2	2193 _b	4724	1,4%	-43,8	1292 _c	4192	1,0%	-52,2	799 _c	2460	1,0%	-36,7	17493	3,1%
ilgili	7098 _a	5605	3,6%	25,1	4251 _b	4328	2,8%	-1,4	3042 _c	3841	2,2%	-15,0	1638 _d	2254	2,1%	-14,2	16029	2,8%
kızgın	4680 _a	5256	2,4%	-10,0	4230 _b	4059	2,8%	3,2	3750 _b	3602	2,8%	2,9	2370 _c	2114	3,0%	6,1	15030	2,7%
neşeli	5090 _a	5176	2,6%	-1,5	3840 _b	3997	2,5%	-2,9	3700 _c	3547	2,7%	3,0	2170 _a	2081	2,7%	2,1	14800	2,6%
b																		
sevecen	5200 _a	4873	2,6%	5,9	4288 _b	3763	2,8%	10,1	2880 _c	3339	2,1%	-9,2	1568 _c	1960	2,0%	-9,7	13936	2,5%
heyecanlı	3352 _a	4366	1,7%	-19,2	3604 _b	3371	2,4%	4,7	3288 _b	2992	2,4%	6,3	2240 _c	1756	2,8%	12,6	12484	2,2%
gururlu	5180 _a	4169	2,6%	19,6	3108 _b	3219	2,0%	-2,3	2282 _c	2857	1,7%	-12,5	1351 _c	1676	1,7%	-8,7	11921	2,1%
enerji dolu	3822 _a	4164	1,9%	-6,6	3003 _a	3215	2,0%	-4,4	3066 _b	2853	2,3%	4,6	2016 _c	1674	2,5%	9,1	11907	2,1%
kafası karışmış	4524 _a	3793	2,3%	14,9	2581 _b	2929	1,7%	-7,6	2117 _c	2599	1,6%	-10,9	1624 _d	1525	2,0%	2,8	10846	1,9%
kötü	3468 _a	3460	1,8%	0,2	2754 _a	2672	1,8%	1,9	2278 _a	2371	1,7%	-2,2	1394 _a	1391	1,8%	0,1	9894	1,7%
özgüven	4260 _a	3437	2,2%	17,6	2292 _b	2654	1,5%	-8,3	1932 _b	2355	1,4%	-10,1	1344 _c	1382	1,7%	-1,1	9828	1,7%
hayal kırıklığına uğramış	2420 _a	2357	1,2%	1,6	1940 _a	1820	1,3%	3,3	1740 _a	1615	1,3%	3,6	640 _b	948	0,8%	-10,9	6740	1,2%
başarılı	1900 _a	2343	1,0%	-11,4	1895 _b	1809	1,2%	2,4	1825 _b	1606	1,3%	6,3	1080 _b	942	1,4%	4,9	6700	1,2%
özgür	1688 _a	1634	0,9%	1,7	1232 _a	1262	0,8%	-1,0	1128 _a	1120	0,8%	0,3	624 _a	657	0,8%	-1,4	4672	0,8%
empati	823 _{a, b}	887	0,4%	-2,7	780 _c	685	0,5%	4,3	637 _{b, c}	607	0,5%	1,4	295 _a	356	0,4%	-3,5	2535	0,4%
müteşekkik	648 _a	678	0,3%	-1,4	537 _a	523	0,4%	0,7	462 _a	464	0,3%	-0,1	291 _a	273	0,4%	1,2	1938	0,3%
coşkunluk	180 _a	295	0,1%	-8,3	252 _b	228	0,2%	1,9	264 _b	202	0,2%	5,0	148 _b	119	0,2%	2,9	844	0,1%
X ² = 25780.936, Sd=81, p<0.05																		

Tablo 4.26’da paylaşılan test sonuçlarını incelemek amacıyla konu etiketleri 10’arlı gruplara ayrılarak grafikler oluşturulmuş ve Şekil 4.3’de paylaşılmıştır.



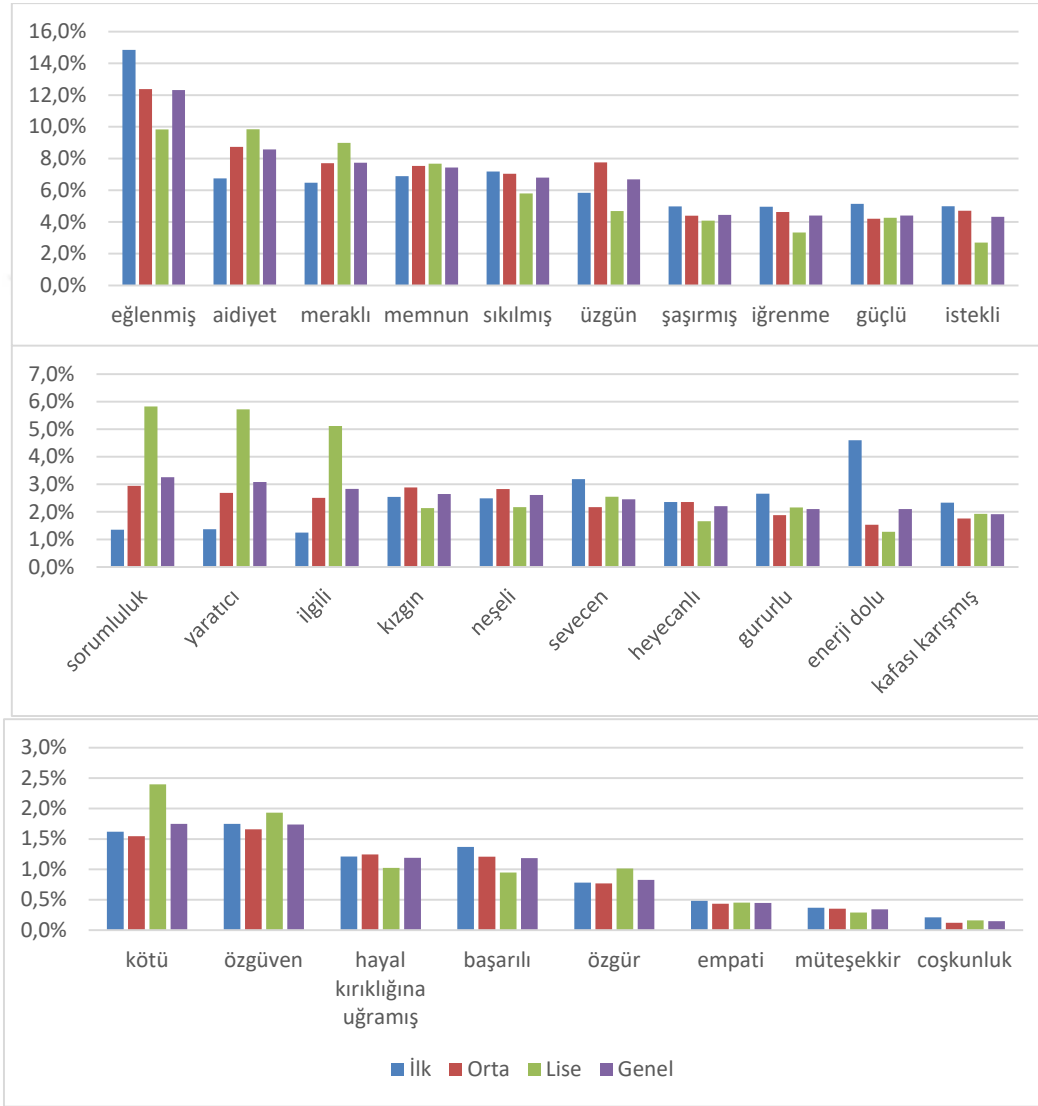
Şekil 4.3: S3 derlemi konu etiketlerinin ders periyotları arasında değişimi gösterir grafikler.

S3 derlemi konu etiketlerinin okul dereceleri arasındaki ilişkiyi ortaya çıkarmak amacıyla yapılan Ki-Kare testi sonuçları Tablo 4.27’de paylaşılmıştır.

Tablo 4.27: S3 derlemi konu etiketlerinin okul dereceleri arasında ki-kare testi sonuçları.

Etiket	İlk				Orta				Lise				Genel			
	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	
eğlenmiş	17149 _a	14228	14,9%	29,3	40326 _b	40145	12,4%	1,5	12287 _c	15388	9,8%	-30,2	69762	69762	12,3%	
aidiyet	7791 _a	9904	6,7%	-24,9	28469 _b	27944	8,7%	5,0	12299 _c	10711	9,8%	18,2	48559	48559	8,6%	
meraklı	7476 _a	8935	6,5%	-18,0	25116 _b	25209	7,7%	-0,9	11214 _c	9663	9,0%	18,6	43806	43806	7,7%	
memnun	7956 _a	8585	6,9%	-7,9	24544 _b	24223	7,5%	3,3	9594 _b	9285	7,7%	3,8	42094	42094	7,4%	
sıkılmış	8289 _a	7844	7,2%	5,8	22923 _a	22130	7,0%	8,5	7245 _b	8483	5,8%	-15,8	38457	38457	6,8%	
üzgün	6750 _a	7725	5,8%	-12,9	25275 _b	21795	7,8%	37,4	5850 _c	8355	4,7%	-32,1	37875	37875	6,7%	
şaşırmış	5750 _a	5137	5,0%	9,8	14329 _b	14493	4,4%	-2,1	5106 _c	5555	4,1%	-7,0	25185	25185	4,4%	
iğrenme	5724 _a	5096	5,0%	10,1	15084 _b	14377	4,6%	9,3	4176 _c	5511	3,3%	-20,8	24984	24984	4,4%	
güçlü	5940 _a	5088	5,1%	13,7	13684 _b	14357	4,2%	-8,8	5324 _b	5503	4,3%	-2,8	24948	24948	4,4%	
istekli	5760 _a	4993	5,0%	12,4	15345 _b	14087	4,7%	16,6	3375 _c	5400	2,7%	-31,9	24480	24480	4,3%	
sorumluluk	1560 _a	3761	1,4%	-40,9	9600 _b	10611	2,9%	-15,3	7280 _c	4068	5,8%	58,0	18440	18440	3,3%	
yaratıcı	1581 _a	3568	1,4%	-37,9	8772 _b	10066	2,7%	-20,1	7140 _c	3859	5,7%	60,8	17493	17493	3,1%	
ilgili	1443 _a	3269	1,2%	-36,3	8190 _b	9224	2,5%	-16,8	6396 _c	3536	5,1%	55,3	16029	16029	2,8%	
kızgın	2940 _a	3065	2,5%	-2,6	9420 _b	8649	2,9%	12,9	2670 _c	3315	2,1%	-12,9	15030	15030	2,7%	
neşeli	2880 _a	3019	2,5%	-2,9	9210 _b	8517	2,8%	11,7	2710 _c	3265	2,2%	-11,1	14800	14800	2,6%	
sevecen	3680 _a	2842	3,2%	17,8	7072 _b	8020	2,2%	-16,4	3184 _c	3074	2,5%	2,3	13936	13936	2,5%	
heyecanlı	2724 _a	2546	2,4%	4,0	7684 _a	7184	2,4%	9,2	2076 _b	2754	1,7%	-14,8	12484	12484	2,2%	
gururlu	3073 _a	2431	2,7%	14,7	6146 _b	6860	1,9%	-13,4	2702 _c	2630	2,2%	1,6	11921	11921	2,1%	
enerji dolu	5313 _a	2429	4,6%	66,3	4998 _b	6852	1,5%	-34,7	1596 _c	2626	1,3%	-23,0	11907	11907	2,1%	
kafası karışmış	2697 _a	2212	2,3%	11,7	5742 _b	6241	1,8%	-9,8	2407 _c	2392	1,9%	0,3	10846	10846	1,9%	
kötü	1870 _a	2018	1,6%	-3,7	5032 _a	5694	1,5%	-13,6	2992 _b	2182	2,4%	19,8	9894	9894	1,7%	
özgüven	2016 _a	2004	1,7%	0,3	5400 _a	5656	1,7%	-5,3	2412 _b	2168	1,9%	6,0	9828	9828	1,7%	
hayal kırıklığına uğramış	1400 _a	1375	1,2%	0,8	4060 _a	3879	1,2%	4,5	1280 _b	1487	1,0%	-6,1	6740	6740	1,2%	
başarılı	1580 _a	1367	1,4%	6,5	3935 _b	3856	1,2%	2,0	1185 _c	1478	0,9%	-8,7	6700	6700	1,2%	
özgür	904 _a	953	0,8%	-1,8	2504 _a	2689	0,8%	-5,5	1264 _b	1031	1,0%	8,3	4672	4672	0,8%	
empati	557 _a	517	0,5%	2,0	1413 _a	1459	0,4%	-1,8	565 _a	559	0,5%	0,3	2535	2535	0,4%	
müteşekkiri	429 _a	395	0,4%	1,9	1146 _a	1115	0,4%	1,4	363 _b	427	0,3%	-3,5	1938	1938	0,3%	
çoşkunluk	246 _a	172	0,2%	6,3	398 _b	486	0,1%	-6,1	200 _c	186	0,2%	1,1	844	844	0,1%	
X ² = 23633,939, Sd=54, p<0.05 * Her alanda istatistiksel olarak farklı özellik gösteren değerler koyu 113ankle işaretlenmiştir.																

Tablo 4.27’de bulunan veriler değerlendirildiğinde veri setinde gruplar arasında anlamlı farklılıklardan söz edilebilir ($X^2 = 23633,939$, $Sd=54$, $p<0.05$). Test sonuçlarını incelemek amacıyla konu etiketleri 10’arlı gruplara ayrılarak grafikler oluşturulmuş ve Şekil 4.4’te paylaşılmıştır.



Şekil 4.4: S3 derlemi konu etiketlerinin okul dereceleri arasında değişimini gösterir grafikler.

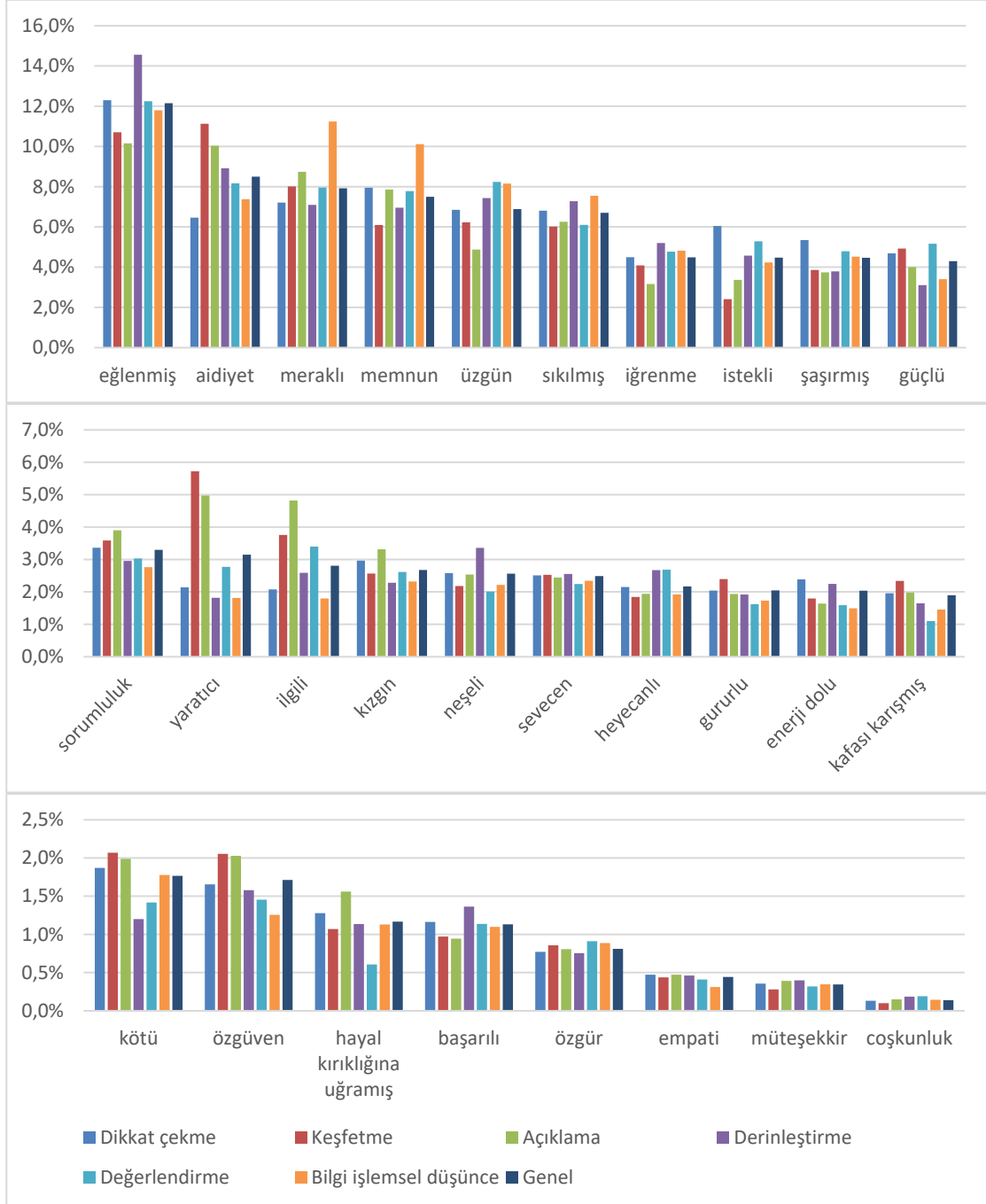
S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri arasında dağılımlarını incelemek amacıyla Ki-Kare testi yapılarak sonuçları Tablo 4.28’de paylaşılmıştır.

Tablo 4.28: S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri arasında dağılımlarını gösterir ki-kare testi sonuçları.

Etiket	Dikkat çekme				Keşfetme				Açıklama				Derinleştirme				Değerlendirme				Bilgi işlemsel düşünme				Genel	
	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	ef	%	ar	f	%
eğlenmiş	20790 _a	20534	12,3%	2,4	13013 _b	14768	10,7%	-17,8	3124 _b	3737	10,2%	-11,1	13607 _c	11355	14,6%	25,1	3234 _{a,d}	3207	12,3%	0,5	5643 _d	5809	11,8%	-2,5	59411	12,1%
aidiyet	10927 _a	14361	6,5%	-37,0	13524 _b	10329	11,1%	37,9	3087 _c	2614	10,0%	10,0	8330 _d	7942	8,9%	5,1	2156 _e	2243	8,2%	-2,0	3528 _f	4063	7,4%	-9,2	41552	8,5%
meraklı	12180 _a	13384	7,2%	-13,4	9744 _b	9626	8,0%	1,4	2688 _c	2436	8,7%	5,5	6636 _c	7401	7,1%	-10,3	2100 _b	2090	8,0%	0,2	5376 _d	3787	11,2%	28,3	38724	7,9%
memnun	13442 _a	12671	8,0%	8,8	7410 _b	9113	6,1%	-21,4	2418 _a	2306	7,9%	2,5	6500 _c	7007	7,0%	-7,0	2054 _a	1979	7,8%	1,8	4836 _d	3585	10,1%	22,9	36660	7,5%
üzgün	11575 _a	11639	6,8%	-0,8	7575 _b	8371	6,2%	-10,4	1500 _c	2118	4,9%	-14,4	6950 _d	6436	7,4%	7,4	2175 _e	1818	8,2%	8,9	3900 _e	3293	8,2%	11,5	33675	6,9%
sıkılmış	11502 _a	11326	6,8%	2,1	7317 _b	8146	6,0%	-11,0	1926 _b	2061	6,3%	-3,2	6804 _c	6263	7,3%	7,9	1611 _b	1769	6,1%	-4,0	3609 _c	3204	7,5%	7,8	32769	6,7%
iğrenme	7596 _a	7590	4,5%	0,1	4968 _b	5459	4,1%	-7,8	972 _c	1381	3,2%	-11,6	4860 _d	4197	5,2%	11,6	1260 _{a,d,e}	1185	4,8%	2,3	2304 _c	2147	4,8%	3,6	21960	4,5%
istekli	10215 _a	7559	6,0%	38,6	2925 _b	5436	2,4%	-40,2	1035 _c	1376	3,4%	-9,7	4275 _d	4180	4,6%	1,7	1395 _e	1181	5,3%	6,6	2025 _d	2138	4,2%	-2,6	21870	4,5%
şaşırmış	9039 _a	7552	5,3%	21,6	4692 _b	5431	3,9%	-11,8	1150 _b	1374	3,7%	-6,4	3542 _b	4176	3,8%	-11,2	1265 _c	1179	4,8%	2,6	2162 _c	2137	4,5%	0,6	21850	4,5%
güçlü	7920 _a	7269	4,7%	9,6	5984 _b	5228	4,9%	12,3	1232 _c	1323	4,0%	-2,6	2904 _d	4020	3,1%	-20,0	1364 _b	1135	5,2%	7,1	1628 _c	2057	3,4%	-10,2	21032	4,3%
sorumluluk	5680 _a	5571	3,4%	1,8	4360 _b	4007	3,6%	6,5	1200 _b	1014	3,9%	6,1	2760 _c	3081	3,0%	-6,5	800 _{a,c}	870	3,0%	-2,5	1320 _c	1576	2,8%	-6,9	16120	3,3%
yaratıcı	3621 _a	5323	2,1%	-29,3	6953 _b	3829	5,7%	59,2	1530 _c	969	5,0%	18,9	1700 _d	2944	1,8%	-25,9	731 _c	831	2,8%	-3,6	867 _d	1506	1,8%	-17,6	15402	3,1%
ilgili	3510 _a	4745	2,1%	-22,5	4563 _b	3412	3,8%	23,0	1482 _c	864	4,8%	22,1	2418 _d	2624	2,6%	-4,5	897 _b	741	3,4%	6,0	858 _c	1342	1,8%	-14,1	13728	2,8%
kızgın	5010 _a	4521	3,0%	9,1	3120 _b	3251	2,6%	-2,7	1020 _c	823	3,3%	7,2	2130 _d	2500	2,3%	-8,3	690 _b	706	2,6%	-0,6	1110 _{b,d}	1279	2,3%	-5,0	13080	2,7%
neşeli	4360 _a	4327	2,6%	0,6	2650 _b	3112	2,2%	-9,7	780 _{a,c}	788	2,5%	-0,3	3140 _d	2393	3,4%	17,2	530 _b	676	2,0%	-5,8	1060 _{b,c}	1224	2,2%	-5,0	12520	2,6%
sevecen	4240 _a	4203	2,5%	0,7	3072 _a	3023	2,5%	1,0	752 _a	765	2,4%	-0,5	2384 _a	2324	2,6%	1,4	592 _a	656	2,2%	-2,6	1120 _a	1189	2,3%	-2,1	12160	2,5%
heyecanlı	3632 _a	3661	2,1%	-0,6	2240 _b	2633	1,8%	-8,9	596 _{a,b}	666	1,9%	-2,8	2496 _c	2024	2,7%	11,8	708 _c	572	2,7%	5,9	920 _b	1036	1,9%	-3,8	10592	2,2%
gururlu	3451 _a	3457	2,0%	-0,1	2912 _b	2487	2,4%	9,9	595 _{a,c,d}	629	1,9%	-1,4	1792 _{a,d}	1912	1,9%	-3,1	427 _c	540	1,6%	-5,0	826 _{c,d}	978	1,7%	-5,2	10003	2,0%
enerji dolu	4032 _a	3440	2,4%	12,6	2184 _b	2474	1,8%	-6,8	504 _{b,c}	626	1,6%	-5,1	2100 _a	1903	2,2%	5,1	420 _{b,c}	537	1,6%	-5,3	714 _c	973	1,5%	-8,8	9954	2,0%
kafası karışmış	3306 _a	3207	2,0%	2,2	2842 _b	2307	2,3%	13,0	609 _a	584	2,0%	1,1	1537 _c	1774	1,6%	-6,3	290 _d	501	1,1%	-9,8	696 _c	907	1,5%	-7,5	9280	1,9%
kötü	3162 _a	2985	1,9%	4,0	2516 _b	2147	2,1%	9,3	612 _{a,b}	543	2,0%	3,1	1122 _c	1651	1,2%	-14,6	374 _c	466	1,4%	-4,4	850 _a	844	1,8%	0,2	8636	1,8%
özgüven	2796 _a	2895	1,7%	-2,3	2496 _b	2082	2,1%	10,6	624 _b	527	2,0%	4,4	1476 _c	1601	1,6%	-3,5	384 _{a,c}	452	1,5%	-3,3	600 _c	819	1,3%	-8,1	8376	1,7%
hayal kırıklığına uğramış	2160 _a	1970	1,3%	5,3	1300 _b	1417	1,1%	-3,6	480 _c	359	1,6%	6,7	1060 _b	1089	1,1%	-1,0	160 _d	308	0,6%	-8,7	540 _{a,b}	557	1,1%	-0,8	5700	1,2%
başarılı	1965 _a	1913	1,2%	1,5	1180 _b	1376	1,0%	-6,1	290 _b	348	0,9%	-3,2	1275 _c	1058	1,4%	7,5	300 _{a,b,c}	299	1,1%	0,1	525 _{a,b}	541	1,1%	-0,7	5535	1,1%
özgür	1304 _a	1369	0,8%	-2,2	1040 _a	984	0,9%	2,1	248 _a	249	0,8%	-0,1	704 _a	757	0,8%	-2,1	240 _a	214	0,9%	1,9	424 _a	387	0,9%	2,0	3960	0,8%
empati	799 _a	747	0,5%	2,4	531 _a	537	0,4%	-0,3	145 _a	136	0,5%	0,8	430 _a	413	0,5%	0,9	108 _{a,b}	117	0,4%	-0,8	148 _b	211	0,3%	-4,6	2161	0,4%
müteşekkik	600 _a	581	0,4%	1,0	339 _b	418	0,3%	-4,4	120 _a	106	0,4%	1,4	372 _a	321	0,4%	3,2	84 _{a,b}	91	0,3%	-0,7	165 _{a,b}	164	0,3%	0,1	1680	0,3%
çoşkunluk	220 _{a,b}	234	0,1%	-1,1	120 _b	168	0,1%	-4,3	46 _{a,b,c}	43	0,1%	0,6	172 _c	129	0,2%	4,2	50 _{a,c}	36	0,2%	2,3	68 _{a,b,c}	66	0,1%	0,2	676	0,1%

X²= 16838,534, Sd= 135, p<0.05

S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri içindeki dağılımının istatistiksel olarak anlamlı düzeyde farklılaştığı değerlendirilebilir ($X^2= 16838,534$, $Sd=135$, $p<0.05$). Değerlerin gruplar arasında değişimini incelemek amacıyla oluşturulan grafikler Şekil 4.5'te paylaşılmıştır.



Şekil 4.5: S3 derlemi konu etiketlerinin S1 derlemi konu etiketleri arasında dağılımlarına ait grafikler.

5 TARTIŞMA VE SONUÇ

Bu çalışmada HEM veri yönetim sistemi üzerinde bulunan ve A12 bağlamından alınan metin verilerinin kategorilendirilmesini sağlayacak modeller oluşturmak amaçlanmıştır. Bu amaçla veri hazırlanmış, işlenmiş ve üzerinde metin madenciliği çalışmaları yapılmıştır. Çalışma kapsamında üzerinde işlem yapılan veriler A12 bağlamından alınan ham veriler olduğundan bu çalışma aynı zamanda konu analizi süreci yol haritasının da ele alındığı bir çalışma olmuştur. Tez kapsamında TDDİ alanında kullanılacak bir ön işlem kütüphanesi oluşturulması, konu modelleme yöntemlerinin incelenmesi amaçlanmıştır. Bu kapsamda mevcut ön işlem yöntemleri analiz edilerek bir yol haritası ortaya koyulmuş ve bir kütüphane geliştirilmiştir. Sonrasında ise konu modelleme yöntemleri, değerlendirme yöntemleri ve bu yöntemlerin kullanımları incelenmiş ve sürecin yol haritası oluşturulmaya çalışılmıştır. Metin madenciliği sürecinin sonunda ise keşifsel analizler yapılmış ve veriden ortaya çıkan örüntüler incelenmiştir.

Araştırma kapsamında öncelikli olarak belge uzunlukları karşılaştırılmıştır. Tüm derlemeler için en uzun belgeleri ortaokul öğrencilerinin oluşturduğu, ortaokul öğrencilerini ise sırasıyla lise ve ilkokul öğrencilerinin takip ettiği ve bu farkların istatistiksel olarak anlamlı olduğu görülmüştür.

Metin ön işleme yöntemlerinin sınıflandırma başarısına etkisi üzerine yapılan çalışmalar incelendiğinde, farklı ön işleme yöntemlerinin farklı sınıflandırma ve kümeleme algoritmalarında sınıflandırma başarısını etkileyip etkilemediği ile ilgili bulgular ve değerlendirmeler bulunduğu tespit edilmiştir [30,40,40,147–150]. Bu durumun dışında Hickman vd. [119] yaptıkları derleme çalışmasında farklı ön işleme yöntemlerinin farklı dillere uygulandığı araştırmaları incelemişler ve metin ön işlem için bir öneri akış diyagramı oluşturmuşlardır. Literatürde ön işleme yöntemlerinin paylaşılmadığı, kısmen paylaşıldığı çalışmalar bulunmaktadır. Bu çalışma ön işleme tekniklerinin derleme etkileri konusunda karşılaştırmalı değerlendirmeleri ve semantik değerlendirmeler sonucunda çalışmada kullanılan derlemeler için en optimal yol haritasını ortaya koymaktadır.

Bu çalışmada ham bir veri seti kullanılmış ve veri denetimsiz öğrenme algoritması olan LDA ile modellenmiştir. Bu durum, başarımın istatistiksel olarak değerlendirilmesini zorlaştırmaktadır. Chang vd. [75] göre konuları keşfetmek denetimsiz bir süreç olduğundan ve altın standartta bir konu listesi olmadığından semantik tutarlılığın insan tarafından

değerlendirilmesine ihtiyaç duyulmaktadır. Bu sebeple metin ön işlem ve modelleme aşamalarında pyLDavis ile oluşturulan konu-kelime dağılımı görselleri uzman değerlendirmesine tabi tutulmuştur. Uzmanlardan alınan dönütler sonucunda süreç düzenlenerek son halini almıştır.

S1 sorusu konu etiketleri incelendiğinde 5E eğitim modelinin basamakları (dikkat çekme, keşfetme, açıklama, derinleştirme, değerlendirme) ve o basamaklara ek olarak “bilgi işlemsel düşünme” basamağının da alan uzmanları tarafından etiketlendiği görülebilir. Bilgi işlemsel düşünme Harezmi Eğitim Modeli’nin 5 zemininden birisidir. 5E modeli STEM eğitimine entegre edilebilecek bir model olduğundan 5E modeli basamaklarına ek olarak “Bilgi İşlemsel Düşünce” basamağının da uzmanlar tarafından işaretlenmesinin Harezmi Eğitim Modelinin uygulama basamaklarını yansıttığı düşünülebilir. Nitekim Tablo 4.19’de bulunan Harezmi Eğitim Modeli aşamalarının 4 ders periyodu arasındaki Ki-Kare testi sonuçları ise modelin hangi basamağının hangi ders periyodunda yoğunlaştığını ve kendine derslerde ne kadar yer bulduğunu gösterir bir veri olduğu sonucuna ulaşılabilir. “Derinleştirme” basamağının her ders periyotunda en baskın olarak karşımıza çıkması durumu derslerin öğrencilerin derse etkin katılımı gözetilerek kurgulandığını gösterebilir. Diğer taraftan; “Bilgi işlemsel düşünme” basamağının dersler içinde 2. Periyottan itibaren baskın bir şekilde yer almasıyla, HEM’in bilgi işlemsel düşünme, programlama becerileri, robotik ve oyun tasarımı gibi zeminlerinin dersler içinde kendine yoğun bir şekilde yer bulduğu sonucuna varılabilir. S1 derlemi etiketlerinin okul dereceleri arasında dağılımlarına göre öğrencilerin yaş seviyeleri düştükçe “Dikkat çekme” ve “Derinleştirme” etkinliklerinin daha yoğun olduğu, “Keşfetme” basamağı etkinliklerinin de öğrencilerin yaş seviyelerine paralel olarak yükseldiği bulgulanmıştır. Bu durum küçük yaş gruplarında ders etkinliklerini yönlendirmede öğretmenin daha etkili olduğu şeklinde yorumlanabilir. “Keşfetme” basamağında baskın bir şekilde ders işlenmiş olması ise bir kavramı öğrenmek ya da bir konuda deneyim sahibi olmak için zamana ihtiyaç duyulduğu durumuyla açıklanabilir [88,89].

S2 sorusu konu etiketleri incelendiğinde derslerin genel olarak çevre, bilgi işlemsel düşünme ve problem çözme konularına odaklandığı değerlendirilebilir. Bu durum Harezmi Eğitim Modeli’nde HİS cümlelerinin “Çevre” alanına yoğunlaştığı şeklinde de yorumlanabilir.

Öğrencilere S3 sorusunun sorularak olumlu-olumsuz olarak değerlendirildiği çalışmalar incelendiğinde 10 ila 182 öğrenci arasında çalışma grupları bulunduğu ve nitel analiz

yöntemleri kullanıldığı görülmektedir [115,117,118,151]. Dere vd. [118] yaptığı çalışmada “Edmodo’yu kullanırken neler hissettin (Başarılı, güvenli, özel, sosyal, mutlu, kaygılı, özgür vb.)? Ayrıntılarıyla açıklar mısın?” sorusunu yöneltilmiş ve öğrenci cevapları uzman tarafından yorumlanmıştır. Yine yapılan duygu analizi çalışmalarında duygu kutbu analizlerine yer verildiği ve verinin gözle işaretlenerek farklı sınıflandırma algoritmaların başarımları karşılaştırılmıştır [109]. Tez kapsamında 4195 öğrencinin 86753 cevabı analiz edildiğinden S3 sorusuna öğrencilerin verdiği cevaplar diğer çalışmalardan farklı olarak LDA algoritmasının otomatik konu-kelime dağılımları çıkarması sağlanmıştır. Ortaya çıkan konu-kelime dağılımları PDR uzmanları tarafından değerlendirilmiş ve konular isimlendirilmiştir. Bulgular incelendiğinde öğrencilerin temel duyguların dışında temel değerleri de duygu olarak yansıttıkları sonucuna ulaşılabilir. Temel değerlerin duygu olarak etiketlenmesi durumu Harezmi modelinin uygulama amaçlarından olan “güvenli, etik ve ahlaki değerleri özümsemek [96]” amacına ulaştığını manidar bir şekilde gösterebildiği bulgulanmıştır. Farklı bir bakış açısından ise bu çalışmaya ait bulgular ışığında insan duygularının duygu kutbu veya temel duygulara ek olarak temel değerleri de içine katan kompleks yapılar içerdiği sonucuna ulaşılabilir. Öğrencilerin büyük bir çoğunluğunun derslerde eğlenmiş olmasının, ayrıca sadece “mutlu” gibi vektör bakımından fakir yorumların derlemeden çıkarılması durumunun ilk olarak öğrencilerin derslerde eğlenmeleri ve mutlu ayrılmaları ile eşleştiği düşünülebilir fakat bulgular incelendiğinde sıkılmışlık duygusunu yansıtan öğrencilerin ders periyotları ilerledikçe artıyor olmasının ve ikinci en yüksek frekanstaki duygu olması durumunun Birgin [152]’in bulgularında da belirttiği üzere öğrencilerin form doldurmaktan sıkıldıkları şeklinde de yorumlanabilir.

S3 sorusu duygu bakımından sözlük tabanlı yöntemle analiz edilmiş ve olumlu yorumların okul türü bazında birbirinden istatistiksel olarak anlamlı bir şekilde ayrıldığı, ilkokuldan liseye doğru gidildikçe olumlu duygu kutbuna sahip belgelerin ağırlığının düştüğü görüşmüştür. Bu durum konu modellemesi sonucunda ortaya çıkan “eğlenmiş” duygu etiketine sahip yorumların ilkokulda en yüksek seviyedeyken liseye doğru azalması ile paralellik göstermektedir.

Tez kapsamında metin madenciliği ön işlem ve modelleme metotları derinlemesine analiz edilmiş ve bir yol haritası ortaya koyulmaya çalışılmıştır. Ayrıca ortaya çıkarılan veri setiyle gerçekleştirilen çalışma sonucunda gerek konu modelleme gerekse sonrasında yapılan keşifsel analizlerde ortaya çıkan bulgular öğrencilerin gözünden derste ne işlendiğini, öğrencilerin duygularını, modelin basamaklarının zamana göre uygulanma biçimini de ortaya koymaktadır.

Bu sebeple ortaya ıkarılan gizli rntlerin bu alanda yapılacak gelecek alıřmalara temel olacađı dřnlmektedir.

Gelecek alıřmalarda arařtırmacılar ve alan uzmanlarıyla birlikte yapılacak derinlemesine analiz ve deđerlendirmelerle halihazırda bulunan veritabanından veya eđitim ortamından alınacak yeni veri kaynaklarından anlamlandırılmıř rntlerin ortaya ıkarılması ve derinlemesine analiz edilerek eđitim alanına katkı sađlanması amalanmaktadır. Ayrıca oluřturulan TurkishTopicModel ktphanesinin geliřtirilmeye devam edilmesi dřnlmektedir.



KAYNAKLAR

- [1] Yıldırım, A. and Şimşek, H., 2013, *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*, Seçkin, Ankara.
- [2] Oflazer, K., 2018, *Turkish Natural Language Processing*, Springer Berlin Heidelberg, New York, NY.
- [3] Sağlam, F., 2019, *Otomatik Duygu Sözlüğü Geliştirilmesi ve Haberlerin Duygu Analizi*, Thesis (PhD.), Hacettepe Üniversitesi.
- [4] Coffman, K. G. and Odlyzko, A. M., 2001, *Internet Growth: Is There A 'Moore's Law' For Data Traffic?*, P. 45.
- [5] Feldman, R. and Dagan, I., *Knowledge Discovery In Textual Databases (KDT)*, P. 7.
- [6] Witten, I. H., Witten, I. H., Frank, E., Hall, M. A and Pall, C. D., Eds., 2017, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, Amsterdam.
- [7] Berry, M. J. A and Linoff, G., 2004, *Data Mining Techniques: For Marketing, Sales and Customer Relationship Management*, Wiley Pub, Indianapolis, Ind.
- [8] Fayyad, U., Piatetsky-Shapiro, G and Smyth, P., 1996, *From Data Mining To Knowledge Discovery In Databases*, AI Magazine, 17(3), P. 37.
- [9] Zhang, C and Zhang, S., 2003, *Association Rule Mining: Models and Algorithms*, Springer.
- [10] Gülseçen, S., Ed., 2015, *Bilgi Yönetimi: Bilgi Türeticileri, Büyük Veri, İnovasyon ve Kurumsal Zeka*, Papatya Yayıncılık, İstanbul
- [11] Olson, D. L and Delen, D., 2008, *Advanced Data Mining Techniques*, Springer Science & Business Media.
- [12] Wirth, R and Hipp, J., 2000, *CRISP-DM: Towards A Standard Process Model For Data Mining*, Proceedings Of The 4th International Conference On The Practical Applications Of Knowledge Discovery and Data Mining, P. 11.
- [13] Gentzkow, M., Kelly, B and Taddy, M., 2019, *Text As Data*, Journal Of Economic Literature, 57(3), 535–574.
- [14] Hotho, A., Nurnberger, A., Paaß, G and Augustin, S., *A Brief Survey Of Text Mining*, 37.
- [15] Gramlich, M., 2020, What Is Structured, Semi Structured and Unstructured Data? › Michael Gramlich., <https://www.michael-gramlich.com/what-is-structured-semi-structured-and-unstructured-data/#:~:text=Semi-structured%20data%20is%20data,structured%20or%20semi-structured%20data.>, [Ziyaret Tarihi: 21-May-2021]
- [16] Savard, J. J. G., *The Punched Card*, <http://www.quadibloc.com/comp/cardint.htm>. [Ziyaret Tarihi: 21-May-2021].
- [17] Wikimedia Commons, 2021, *File:ASCII-Table-Wide.Svg*, <https://commons.wikimedia.org/wiki/File:ASCII-Table-wide.svg>, [Erişim Tarihi: 12-May-2021]

- [18] Ethnologue, 2016, *How Many Languages Are There In The World?*, , <https://www.ethnologue.com/guides/how-many-languages/>, [Ziyaret Tarihi: 17-May-2021].
- [19] Worldstandards, 2020, *The World's Scripts and Alphabets*, Worldstandards, <https://www.worldstandards.eu/other/alphabets/>, [Ziyaret Tarihi: 21-May-2020].
- [20] Haralambous, Y and Horne, P. S., 2007, *Fonts & Encodings*, O'Reilly Media, Beijing ; Sebastopol, Calif.
- [21] Unicode Inc., 2020, *The Unicode Standard, Version 13.0.*, <https://www.unicode.org/versions/components-13.0.0.html>, [Ziyaret Tarihi: 20-May-2020]
- [22] Swofford, M., 2020, *UTF-8 Unicode Vs. Other Encodings Over Time | Pinyin News.*, <http://pinyin.info/news/2015/utf-8-unicode-vs-other-encodings-over-time/>, [Ziyaret Tarihi: 20-May-2020]
- [23] Aggarwal, C. C and Zhai, C., Eds., 2012, *Mining Text Data*, Springer US, Boston, MA.
- [24] Fan, W., Wallace, L., Rich, S and Zhang, Z., 2006, *Tapping The Power Of Text Mining*, Communications Of The ACM, 49, 76–82.
- [25] Chowdhury, G. G., 2003, *Natural Language Processing*, Annual Review Of Information Science and Technology, 37(1), 51–89.
- [26] Kao, A and Poteet, S. R., Eds., 2007, *Natural Language Processing and Text Mining*, Springer, London.
- [27] Miner, G. D., Delen, D., Elder, J. F., Fast, A., Hill, T and Nisbet, R. A., 2012, *Practical Text Mining and Statistical Analysis For Non-Structured Text Data Applications*, Elsevier.
- [28] Gaikwad, V. S., Chaugule, A and Patil, P., 2014, *Text Mining Methods and Techniques*, IJCA, 85(17), 42–45.
- [29] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B and Kochut, K., 2017, *A Brief Survey Of Text Mining: Classification, Clustering and Extraction Techniques*, Arxiv:1707.02919[Cs].
- [30] Torunoglu, D., Cakirman, E., Ganiz, M. C., Akyokus, S and Gurbuz, M. Z., 2011, *Analysis Of Preprocessing Methods On Classification Of Turkish Texts*, 2011 International Symposium On Innovations In Intelligent Systems and Applications, IEEE, Istanbul, Turkey, 112–117.
- [31] Grimmer, J and Stewart, B. M., 2013, *Text As Data: The Promise and Pitfalls Of Automatic Content Analysis Methods For Political Texts*, Polit. Anal., 21(3), 267–297.
- [32] Kannan, D. S and Gurusamy, V., *Preprocessing Techniques For Text Mining.*, International Journal of Computer Science & Communication Networks, 5(1), 7-16.
- [33] Han, J., Kamber, M and Pei, J., 2011, *Data Mining Concepts and Techniques Third Edition*, The Morgan Kaufmann Series in Data Management Systems, 5(4), 83–124.
- [34] Raman, V and Hellerstein, J. M., 2001, *Potter's Wheel: An Interactive Data Cleaning System*, Proceedings Of The 27th VLDB Conference, Rome, Italy, 10.
- [35] Gupta, V and Lehal, G. S., 2009, *A Survey Of Text Mining Techniques and Applications*, JETWI, 1(1), 60–76.

- [36] Akın, A. A., 2021, *Ahmetaa/Zemberek-Nlp*, <http://zembereknlp.blogspot.com>, [Ziyaret Tarihi: 21-May-2021].
- [37] Akın, A. A and Akın, M. D., 2007, *Zemberek, An Open Source NLP Framework For Turkic Languages*, Structure, P. 8.
- [38] Lusetti, M., Ruzsics, T., Göhring, A., Samardžić, T and Stark, E., 2018, *Encoder-Decoder Methods For Text Normalization*.
- [39] Hassan, H and Menezes, A., 2013, *Social Text Normalization Using Contextual Graph Random Walks*, *Proceedings Of The 51st Annual Meeting Of The Association For Computational Linguistics (Volume 1: Long Papers)*, Association For Computational Linguistics, Sofia, Bulgaria, 1577–1586.
- [40] Eryiğit, G and Torunoğlu-Selamet, D., 2017, *Social Media Text Normalization For Turkish*, *Natural Language Engineering*, 23(6), 835–875.
- [41] Gülseçen, S., Sharma, S and Akadal, E., 2020, *WHO RUNS THE WORLD: DATA*, Istanbul University, İstanbul.
- [42] Kaur, J and Buttar, P. K., 2018, *A Systematic Review On Stopword Removal Algorithms*, *International Journal On Future Revolution In Computer Science & Communication Engineering*, 4(4), 207–210.
- [43] Loper, E and Bird, S., 2002, *NLTK: The Natural Language Toolkit*, In *Proceedings Of The ACL Workshop On Effective Tools and Methodologies For Teaching Natural Language Processing and Computational Linguistics*. Philadelphia: Association For Computational Linguistics.
- [44] Lo, R. T.-W., He, B and Ounis, I., 2005, *Automatically Building A Stopword List For An Information Retrieval System*, *Journal On Digital Information Management: Special Issue On The 5th Dutch-Belgian Information Retrieval Workshop (DIR)*, 17–24.
- [45] Balakrishnan, V and Lloyd-Yemoh, E., *Stemming and Lemmatization: A Comparison Of Retrieval Performances*, P. 6.
- [46] Doğuç, Ö., Aytaç, Ö. B and Silahtaroglu, G., 2020, *Lemmatizer: Akıllı Türkçe Kök Bulma Yöntemi*, *Turkishstudies*, 15(3), 289–299.
- [47] Chew, P. A., Bader, B. W., Helmreich, S., Abdelali, A and Verzi, S. J., 2011, *An Information-Theoretic, Vector-Space-Model Approach To Cross-Language Information Retrieval*, *Nat. Lang. Eng.*, 17(1), 37–70.
- [48] Dumais, S. T., 1991, *Improving The Retrieval Of Information From External Sources*, *Behavior Research Methods, Instruments, & Computers*, 23(2), 229–236.
- [49] Cutting, D. R., Karger, D. R and Pedersen, J. O., 1993, *Constant Interaction-Time Scatter/Gather Browsing Of Very Large Document Collections*, *Proceedings Of The 16th Annual International ACM SIGIR Conference On Research and Development In Information Retrieval*, Association For Computing Machinery, New York, NY, USA, 126–134.
- [50] Blei, D. M., 2012, *Probabilistic Topic Models*, *Commun. ACM*, 55(4), 77–84.
- [51] Sbalchiero, S and Eder, M., 2020, *Topic Modeling, Long Texts and The Best Number Of Topics. Some Problems and Solutions*, *Qual Quant*, 54(4), 1095–1108.

- [52] Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S and Harshman, R., 1988, *Using Latent Semantic Analysis To Improve Access To Textual Information*, *Proceedings Of The SIGCHI Conference On Human Factors In Computing Systems*, Association For Computing Machinery, Washington, D.C., USA, 281–285.
- [53] Dumais, S. T., 2004, *Latent Semantic Analysis*, *Annual Review Of Information Science and Technology*, 38(1), 188–230.
- [54] Hofmann, T., 1999, *Probabilistic Latent Semantic Indexing*, *Proceedings Of The 22nd Annual International ACM SIGIR Conference On Research and Development In Information Retrieval*, 50–57.
- [55] Hofmann, T., 2001, *Unsupervised Learning By Probabilistic Latent Semantic Analysis*, *Machine Learning*, 42, 177–196.
- [56] Saul, L and Pereira, F., 1997, *Aggregate and Mixed-Order Markov Models For Statistical Language Processing*, *Arxiv Preprint Cmp-Lg/9706007*.
- [57] Hennig, L., 2009, *Topic-Based Multi-Document Summarization With Probabilistic Latent Semantic Analysis*, *Proceedings Of The International Conference RANLP-2009*, Association For Computational Linguistics, Borovets, Bulgaria, 144–149.
- [58] Gimpel, K., 2006, *Modeling Topics*, *Inform. Retrieval*, 5, 1–23.
- [59] Blei, D. M., Ng, A. Y and Jordan, M. I., 2003, *Latent Dirichlet Allocation*, *Journal Of Machine Learning Research*, 993–1022.
- [60] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y and Zhao, L., 2019, *Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, A Survey*, *Multimed Tools Appl*, 78(11), 15169–15211.
- [61] Lee, S., Baker, J., Song, J and Wetherbe, J. C., 2010, *An Empirical Comparison Of Four Text Mining Methods*, *2010 43rd Hawaii International Conference On System Sciences*, 1–10.
- [62] Blei, D. M and Mcauliffe, J. D., 2010, *Supervised Topic Models*, *Arxiv Preprint Arxiv:1003.0783*.
- [63] Ramage, D., Hall, D., Nallapati, R and Manning, C. D., 2009, *Labeled LDA: A Supervised Topic Model For Credit Attribution In Multi-Labeled Corpora*, *Proceedings Of The 2009 Conference On Empirical Methods In Natural Language Processing*, Association For Computational Linguistics, Singapore, 248–256.
- [64] Ramage, D., Manning, C and Dumais, S., 2011, *Partially Labeled Topic Models For Interpretable Text Mining*, 457–465.
- [65] Blei, D and Lafferty, J., 2006, *Correlated Topic Models*, *Advances In Neural Information Processing Systems*, 18, P. 147.
- [66] Mimno, D. M and McCallum, A., 2008, *Topic Models Conditioned On Arbitrary Features With Dirichlet-Multinomial Regression*, *UAI, Citeseer*, 411–418.
- [67] Li, W and McCallum, A., 2006, *Pachinko Allocation: DAG-Structured Mixture Models Of Topic Correlations*, *Proceedings Of The 23rd International Conference On Machine Learning*, 577–584.
- [68] Blei, D. M and Lafferty, J. D., 2006, *Dynamic Topic Models*, *Proceedings Of The 23rd International Conference On Machine Learning*, 113–120.

- [69] Bhadury, A., Chen, J., Zhu, J and Liu, S., 2016, *Scaling Up Dynamic Topic Models*, Arxiv:1602.06049[Stat].
- [70] Seker, S. E., 2016, *Duygu Analizi (Sentimental Analysis)*, YBS Ansiklopedi, 3, 3.
- [71] Hussein, D. M. E.-D. M., 2018, *A Survey On Sentiment Analysis Challenges*, Journal Of King Saud University-Engineering Sciences, 30(4), 330–338.
- [72] Can, U and Alatas, B., 2017, *Duygu Analizi Ve Fikir Madenciliği Algoritmalarının İncelenmesi*, P. 37.
- [73] Dehkharghani, R., Saygin, Y., Yanikoglu, B and Oflazer, K., 2016, *Sentitürknet: A Turkish Polarity Lexicon For Sentiment Analysis*, Lang Resources & Evaluation, 50(3), 667–685.
- [74] Sağlam, F., Genç, B and Sever, H., *Extending A Sentiment Lexicon With Synonym–Antonym Datasets: Swnettr++*, P. 15.
- [75] Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L and Blei, D. M., *Reading Tea Leaves: How Humans Interpret Topic Models*, P. 9.
- [76] Wang, Y., Zhao, X., Sun, Z., Yan, H., Wang, L., Jin, Z., Wang, L., Gao, Y., Zeng, J., Yang, Q and Law, C., 2014, *Towards Topic Modeling For Big Data*.
- [77] Röder, M., Both, A and Hinneburg, A., 2015, *Exploring The Space Of Topic Coherence Measures*, *Proceedings Of The Eighth ACM International Conference On Web Search and Data Mining*, 399–408.
- [78] Douven, I and Meijs, W., 2007, *Measuring Coherence*, Synthese, 156(3), 405–425.
- [79] Tuna, Y. and Kayaoğlu, A., 2018, *BİREY VE DAVRANIŞ*, T.C. ANADOLU ÜNİVERSİTESİ YAYINI, Eskişehir.
- [80] Cornelius, R. R., 2000, *THEORETICAL APPROACHES TO EMOTION*, *Workshop (ITRW) On Speech and Emotion*, P. 8.
- [81] Shaver, P. R., Wu, S and Schwartz, J. C., 1992, *Cross-Cultural Similarities and Differences In Emotion and Its Representation*, *Emotion*, Sage Publications, Inc, Thousand Oaks, CA, US, 175–212.
- [82] Plutchik, R., 1980, *Chapter 1 - A GENERAL PSYCHOEVOLUTIONARY THEORY OF EMOTION*, *Theories Of Emotion*, R. Plutchik and H. Kellerman, Eds., Academic Press, 3–33.
- [83] Plutchik, R., 2001, *The Nature Of Emotions: Human Emotions Have Deep Evolutionary Roots, A Fact That May Explain Their Complexity and Provide Tools For Clinical Practice*, *American Scientist*, 89(4), 344–350.
- [84] Yıldırım, S. G and Demirel, M., 2019, *Türk Tarihinde Değerler Ve Değerler Eğitimi Üzerine Bir İnceleme*, *Eğitim Kuram ve Uygulama Araştırmaları Dergisi*, 5(1), 92-99.
- [85] MEB, 2017, *Müfredatta Yenileme ve Değişiklik Çalışmalarımız Üzerine*, <https://ttkb.meb.gov.tr>, [Erişim Tarihi: 10-May-2021]
- [86] Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A and Landes, N., 2006, *The BSCS 5E Instructional Model: Origins and Effectiveness*, Colorado Springs, Co: BSCS, 5, 88–98.

- [87] Duran, L. B and Duran, E., 2004, *The 5E Instructional Model: A Learning Cycle Approach For Inquiry-Based Science Teaching*, The Science Education Review, 3(2), 49–58.
- [88] Bybee, R. W and Landes, N. M., 1990, *Science For Life & Living: An Elementary School Science Program From Biological Sciences Curriculum Study*, The American Biology Teacher, 52(2), 92–98.
- [89] Eisenkraft, A., 2003, *Expanding The 5E Model: A Proposed 7E Model Emphasizes 'Transfer Of Learning' and The Importance Of Eliciting Prior Understanding*, The Science Teacher, 70(6), 56–59.
- [90] Wing, J. M., 2006, *Computational Thinking*, Communications Of The ACM, 49(3), 33–35.
- [91] Henderson, P. B., Cortina, T. J and Wing, J. M., 2007, *Computational Thinking, Proceedings Of The 38th SIGCSE Technical Symposium On Computer Science Education*, 195–196.
- [92] Denning, P. J and Tedre, M., 2019, *Computational Thinking*, MIT Press.
- [93] ISTEand CSTA, 2011, *Computational Thinking Operational Definition ISTE.Pdf*.
- [94] Gülbahar, Y., Kert, S. B and Kalelioğlu, F., 2019, *Bilgi İşlemsel Düşünme Becerisine Yönelik Öz Yeterlik Algısı Ölçeği: Geçerlik Ve Güvenirlik Çalışması*, Turkish Journal Of Computer and Mathematics Education (TURCOMAT), 10(1), 1–29.
- [95] Yelkenci, F., Karadeniz, Ş and Allsop, Y., 2016, *Zihinden Makineye Bilgisayar Bilimleri ve Disiplinlerarası Eğitim Çalıştayı Raporu*, İstanbul İl Milli Eğitim Müdürlüğü, İstanbul.
- [96] İstanbul İl Milli Eğitim Müdürlüğü, 2020, *Harezmi Eğitim Modeli Nedir?* <https://harezmi.meb.gov.tr>[Ziyaret Tarihi: 10-Jul-2021]
- [97] MEB, *EBA Harezmi | Eğitim Bilişim Ağı*, <https://www.eba.gov.tr/harezmi>. [Ziyaret Tarihi: 15-Jul-2021].
- [98] Yadollahi, A., Shahraki, A. G and Zaiane, O. R., 2017, *Current State Of Text Sentiment Analysis From Opinion To Emotion Mining*, ACM Comput. Surv., 50(2), 1–33.
- [99] Mahadevan, A and Arock, M., *Integrated Topic Modeling and Sentiment Analysis: A Review Rating Prediction Approach For Recommender Systems*, 17.
- [100] Hutto, C and Gilbert, E., 2014, *Vader: A Parsimonious Rule-Based Model For Sentiment Analysis Of Social Media Text*, Proceedings Of The International AAAI Conference On Web and Social Media.
- [101] Li, X., Rao, Y., Xie, H., Liu, X., Wong, T.-L and Wang, F. L., 2019, *Social Emotion Classification Based On Noise-Aware Training*, Data & Knowledge Engineering, 123, P. 101605.
- [102] Liang, W., Xie, H., Rao, Y., Lau, R. Y. K and Wang, F. L., 2018, *Universal Affective Model For Readers' Emotion Classification Over Short Texts*, Expert Systems With Applications, 114, 322–333.
- [103] Rao, Y., Xie, H., Li, J., Jin, F., Wang, F. L and Li, Q., 2016, *Social Emotion Classification Of Short Text Via Topic-Level Maximum Entropy Model*, Information & Management, 53(8), 978–986.

- [104] Li, F., Huang, M and Zhu, X., 2010, *Sentiment Analysis With Global Topics and Local Dependency, Proceedings Of The AAAI Conference On Artificial Intelligence*.
- [105] Onan, A., Korukoğlu, S and Bulut, H., *LDA-Based Topic Modelling In Text Sentiment Classification: An Empirical Analysis*, P. 19.
- [106] Lopez, M. B., Alor-Hernández, G and Sánchez-Cervantes, J. L., 2018, *Eduarp: An Educational Resources Platform Based On Opinion Mining and Semantic Web*, Journal Of Universal Computer Science, 24(11), 1515–1535.
- [107] De Paula Santos, F., Lechugo, C. P and Silveira-Mackenzie, I. F., 2016, 'Speak Well' Or 'Complain' About Your Teacher: A Contribution Of Education Data Mining In The Evaluation Of Teaching Practices, 2016 International Symposium On Computers In Education (SIIE), 1–4.
- [108] Chauhan, G. S., Agrawal, P and Meena, Y. K., 2019, *Aspect-Based Sentiment Analysis Of Students' Feedback To Improve Teaching–Learning Process, Information and Communication Technology For Intelligent Systems*, S.C. Satapathy and A. Joshi, Eds., Springer, Singapore, 259–266.
- [109] Göker, H and Tekedere, H., 2017, *FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi*, Bilişim Teknolojileri Dergisi, 10(3), 291–299.
- [110] Gürçan, F., 2017, *Yeni Nesil Yazılım Geliştirme Eğilimlerine Yönelik Uzman Bilgi Ve Becerilerinin Olasılıksal Konu Modelleme Yordamıyla Belirlenmesi.*, Thesis (PhD), Karadeniz Teknik Üniversitesi.
- [111] Caner, F., 2019, *Topic Modeling and Sentiment Analysis Of Hotel Reviews.*, Thesis (Ms.), Bahcesehir University.
- [112] Koçoğlu, E., 2018, *Türkiye’de Pilot Uygulama Sürecinde Olan Harezmi Eğitim Modelinin Alan Uzmanlarının Görüşleri Doğrultusunda Analizi*, Electronic Turkish Studies, 13(19), 1187–1200.
- [113] Ceylan, Ö., Öğten, M., Tüfekçi, V. And Yurttaş, M. Ö., *Öğrencilerin Harezmi Eğitim Modeline Yönelik Metaforik Algılarının Belirlenmesi*, Milli Eğitim Dergisi, 49(225), 227–251.
- [114] Yavuz, O., Yavuz, Y., Özyürek, H., Boral, D., 2019, *Harezmi Eğitim Modeli’nin Özel Yetenekli Öğrencilerin Yaratıcılık Becerilerine Etkisi*, *International Congress on Gifted and Talented Education*, 1-2-3 November 2019, 234.
- [115] Aydemir, Z., Sakiz, G., Doğan, M. C and Aşıcı, M., 2017, *Kademeli Sorumluluk Aktarımı Modeli Kullanılarak Öğretilen İnternette Arama-Anlama Stratejilerine Yönelik Öğrencilerin Yansitici Düşünme Eylemleri*, Journal of International Social Research, 10(51), 629–640.
- [116] Taş, M., 2014, *Karikatür Destekli Fen Öğretimine İlişkin Bir Araştırma: İlköğretim 6. Sınıf Yaşamımızdaki Elektrik Ünitesi Örneği*, Uludağ Üniversitesi Eğitim Fakültesi, 26(2), 473–500.
- [117] Dilmaç, O and Dilmaç, S., *Ortaöğretim Öğrencilerinin Otantik Değerlendirme Yaklaşımlarına Yönelik Görüşleri: Nitel Bir Araştırma*, Uşak Üniversitesi Sosyal Bilimler Dergisi, 7(4), 275–298.

- [118] Dere, E., Yücel, Ü. A and Yalçınalp, S., 2016, *İlköğretim Öğrencilerinin Eğitsel Bir Çevrimiçi Sosyal Öğrenme Ortamı Olan Edmodo'ya İlişkin Görüşleri* *Opinions Of K-12 Students About An Online Social Learning Environment: Edmodo*, *Elementary Education Online*, 15(3), 804–819.
- [119] Hickman, L., Thapa, S., Tay, L., Cao, M and Srinivasan, P., 2022, *Text Preprocessing For Text Mining In Organizational Research: Review and Recommendations*, *Organizational Research Methods*, 25(1), 114–146.
- [120] Tunalı, V. and Bilgin, T. T., 2012, *PRETO: A High-Performance Text Mining Tool For Preprocessing Turkish Texts*, *Proceedings Of The 13th International Conference On Computer Systems and Technologies*, 134–140.
- [121] Salih, B and Günel, E. S., 2022, *The Impact Of Features and Preprocessing On Automatic Text Summarization*, *SCIENCE AND TECHNOLOGY*, 25(2), 117–132.
- [122] Çimen, A and Dede, D., 2018, *Fabsis: A Data Management System Ineducational Administration*, *Conference Proceedings Of »Elearning and Software For Education« (ELSE)*, Bucharest, 382-388.
- [123] Leff, A and Rayfield, J. T., 2001, *Web-Application Development Using The Model/View/Controller Design Pattern*, *Proceedings Fifth Ieee International Enterprise Distributed Object Computing Conference*, IEEE, 118–127.
- [124] Patel, J., 2021, *PHP Framework Comparison: What Is Best PHP Framework To Use?*, <https://www.monocubed.com/blog/php-framework-comparison/>, [Erişim Tarihi: 10-Dec-2021]
- [125] Singh, S., 2019, *10 Best PHP Frameworks For Web Development in 2021 [Latest Updated]*, <https://enlear.academy/top-10-php-frameworks-for-web-development-in-2021-d99696a92f05>, [Ziyaret Tarihi : 27-Ekm-2021]
- [126] Codeigniter, 2021, *Codeigniter Web Framework*, <https://codeigniter.com/>, [Ziyaret Tarihi: 22-Jul-2021].
- [127] Oracle, *Mysql :: Mysql 8.0 Reference Manual :: 11.5 The JSON Data Type*, <https://dev.mysql.com/doc/refman/8.0/en/json.html>, [Ziyaret Tarihi: 22-Jul-2021].
- [128] Js.Foundation, 2021, *Jquery.*, <https://jquery.com/>, [Ziyaret Tarihi: 22-Jul-2021].
- [129] Thorntonand, J., 2021, *Bootstrap*, <https://getbootstrap.com/>, [Ziyaret Tarihi: 22-Jul-2021].
- [130] Anaconda Inc, 2021, *Anaconda | State Of Data Science 2021*, Anaconda, <https://www.anaconda.com/state-of-data-science-2021>, [Ziyaret Tarihi: 22-Jul-2021].
- [131] Python Software Foundation, 2021, *PyPI · The Python Package Index*, <https://pypi.org/>, [Ziyaret Tarihi: 17-Feb-2021].
- [132] Python Software Foundation, *Welcome To Python.Org.*, <https://www.python.org/>, [Ziyaret Tarihi: 18-Jul-2021].
- [133] Anaconda Inc, *Anaconda | The World's Most Popular Data Science Platform*, <https://www.anaconda.com/>, [Ziyaret Tarihi: 22-Jul-2021].
- [134] Menard, S., 2021, *Jpype1: A Python To Java Bridge*, *Jpype1*: <https://pypi.org/project/Jpype1/>, [Ziyaret Tarihi: 22-Jul-2021].

- [135] Reback, J., Jbrockmendel, Mckinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Hawkins, S., Gfyoung, Sinhrks, Roeschke, M., Klein, A., Terji Petersen, Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., Hayden, A., Saxton, D., Gorelli, M. E., Shadrach, R., Jancauskas, V., McMaster, A., Fangchen Li, Battiston, P., Skipper Seabold, Attack68and Kaiqi Dong, 2021, *Pandas-Dev/Pandas: Pandas 1.3.0*, Zenodo.
- [136] Yutaka.Matsubara, *Pymysql: Pure Python Mysql Driver*.
- [137] Pitrou, A., *Pickle5 :: Anaconda.Org* <https://anaconda.org/conda-forge/pickle5>. [Ziyaret Tarihi: 22-Jul-2021].
- [138] Lee, M., *Tomotopy API Documentation (V0.12.1)*, <https://bab2min.github.io/tomotopy/v0.12.1/en/>. [Ziyaret Tarihi: 23-Jul-2021].
- [139] Çelik, H., Baykal, N. B and Memur, H. N. K., 2020, *Nitel Veri Analizi Ve Temel İlkeleri*, Eğitimde Nitel Araştırmalar Dergisi, 8(1), 379–406.
- [140] Griffiths, T. L and Steyvers, M., 2004, *Finding Scientific Topics*, Proceedings Of The National Academy Of Sciences, 101 (Supplement 1), 5228–5235.
- [141] Rieger, J., Rahnenführer, J and Jentsch, C., 2020, *Improving Latent Dirichlet Allocation: On Reliability Of The Novel Method Ldaprototype*, *Natural Language Processing and Information Systems*, E. Métais, F. Meziane, H. Horacekand P. Cimiano, Eds., Springer International Publishing, Cham, 118–125.
- [142] Wood, J., Tan, P., Wang, W and Arnold, C., 2017, *Source-LDA: Enhancing Probabilistic Topic Models Using Prior Knowledge Sources*, Arxiv:1606.00577[Cs].
- [143] Carson S. and Kenny S., 2021, *Pyldavis — Pyldavis 2.1.2 Documentation* <https://pyldavis.readthedocs.io/en/latest/readme.html>. [Ziyaret Tarihi: 05-Aug-2021].
- [144] Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L and Blei, D. M., 2009, *Reading Tea Leaves: How Humans Interpret Topic Models*, *Advances In Neural Information Processing Systems*, P. 9.
- [145] Jones, J and Hunter, D., 1995, *Qualitative Research: Consensus Methods For Medical and Health Services Research*, BMJ, 311(7001), 376–380.
- [146] Thangaratinam, S and Redman, C. W., 2005, *The Delphi Technique*, *The Obstetrician & Gynaecologist*, 7(2), 120–125.
- [147] Salur, M. U and Aydın, İ., 2018, *The Impact Of Preprocessing On Classification Performance In Convolutional Neural Networks For Turkish Text*, *2018 International Conference On Artificial Intelligence and Data Processing (IDAP)*, Zonguldak, TURKEY, 1–4.
- [148] Açıkalın, B and Bayazıt, N. G., 2016, *Türkçe Metinlerin Sınıflandırılmasında Önışlemenin Önemi*, *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2053–2056.
- [149] Mayda, I and Korkmaz, M., 2018, *Sentiment Analysis With Turkish Adjective Dictionary*, *2018 Innovations In Intelligent Systems and Applications Conference (ASYU)*, IEEE, Adana, 1–6.
- [150] Yıldırım, S and Yıldız, T., 2018, *A Comparative Analysis Of Text Classification For Turkish Language*, *Pamukkale J Eng Sci*, 24(5), 879–886.

- [151] Taş, M., 2013, *A Research On Science Teaching Aided Cartoons: Is An Example Lives Of Electrical Unit Primary 6th Class*, Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 26(2), 473–500.
- [152] Birgin, O., 2008, *Alternatif Bir Değerlendirme Yöntemi Olarak Portfolyo Değerlendirme Uygulamasına İlişkin Öğrenci Görüşleri*, Türk Eğitim Bilimleri Dergisi, 6(1), 1–24.



EKLER

EK 1. S1 Sorusu Konu-Kelime Ağırlıkları Tablosu.

k	Kelimeler
0	anket: 0.05694776028394699, sonuç: 0.029566138982772827, grafik: 0.028642699122428894, değerlendirdik: 0.018019292503595352, bilgisayar: 0.01652228832244873, tablo: 0.012895622290670872, soru: 0.012311632744967937, oluşturduk: 0.011976806446909904, oyun: 0.01172555424273014, hazırladık: 0.01147541031241417
1	problem: 0.04209676757454872, alt: 0.031245093792676926, sorun: 0.030216913670301437, başlık: 0.028188616037368774, bulduk: 0.02134213224053383, belirledik: 0.01994926854968071, çözüm: 0.01698066107928753, seçtik: 0.01585943065583706, balıkkılçık: 0.015187191776931286, yazdık: 0.01506807841360569
2	oynadık: 0.05140218883752823, oyun: 0.04921185225248337, okuduk: 0.0285414457321167, kitap: 0.023061946034431458, hikaye: 0.01994924433529377, lego: 0.018329352140426636, testi: 0.014873120002448559, sandalye: 0.014853250235319138, oyna: 0.014809386804699898, ders: 0.012609257362782955
3	bilgi: 0.02435719594359398, hakkında: 0.017061643302440643, hayat: 0.015580818057060242, boy: 0.013144143857061863, aldık: 0.010472046211361885, öğretmen: 0.010066010989248753, günlük: 0.009837106801569462, verdi: 0.009746192023158073, yeni: 0.009121675044298172, ilgili: 0.008667565882205963
4	ektik: 0.060016658157110214, toprak: 0.03767864406108856, tohum: 0.030971236526966095, diktik: 0.02223789691925049, saksı: 0.021361278370022774, suladık: 0.021008459851145744, koyduk: 0.01749950461089611, bahçe: 0.016923177987337112, bitki: 0.01629938744008541, çiçek: 0.012873499654233456
5	gittik: 0.049637001007795334, park: 0.025221548974514008, arıtma_tesis: 0.02508656680583954, otobüs: 0.023182667791843414, bindik: 0.019244275987148285, gezdik: 0.01835443079471588, kara: 0.017967013642191887, şimşek: 0.017922349274158478, yemek: 0.015388657338917255, gezi: 0.014325420372188091
6	dans: 0.03589189425110817, oynadık: 0.02935994230210781, oyun: 0.028648993000388145, etkinlik: 0.02501250058412552, ettik: 0.024481123313307762, şarkı: 0.019419271498918533, hareket: 0.017860939726233482, istasyon: 0.01561258640140295, söyledik: 0.015382985584437847, ısınma: 0.013974161818623543
7	oynadık: 0.03808159381151199, balon: 0.03366945683956146, oyun: 0.03220273554325104, top: 0.03057698905467987, oyna: 0.014628211967647076, harf: 0.012784549966454506, işaret: 0.012726624496281147, isim: 0.012003805488348007, etkinlik: 0.011765491217374802, bardak: 0.00988777820020914
8	gittik: 0.05134712532162666, bahçe: 0.025908179581165314, gezdik: 0.025292186066508293, git: 0.021352745592594147, gezi: 0.02043592557311058, botanik: 0.020031563937664032, fuar: 0.017093120142817497, spor: 0.016489656642079353, bilim: 0.016438592225313187, atölye: 0.01492013968527317
9	şenlik: 0.05775674432516098, hakkında: 0.04312000796198845, harezmi: 0.04147554561495781, konuştuk: 0.04023836925625801, hazırlık:

	0.037629444152116776, proje: 0.030300810933113098, yapacak: 0.029652871191501617, bilgi: 0.027076169848442078, ilgili: 0.018878430128097534, planlama: 0.01870819181203842
10	hikaye: 0.0548919215798378, yazdık: 0.04859797656536102, slogan: 0.040309567004442215, oluşturdık: 0.030606580898165703, hiscümlesi: 0.027909306809306145, his: 0.023753609508275986, balıkkılçık: 0.022923452779650688, grup: 0.021460993215441704, bulduk: 0.01932908408343792, afiş: 0.014573420397937298
11	hafta: 0.021350476890802383, yapacak: 0.016831286251544952, konuştuk: 0.015667805448174477, neler: 0.01292545348405838, görev: 0.011918486095964909, verdik: 0.011300020851194859, hakkında: 0.010442213155329227, karar: 0.009963742457330227, ders: 0.009937562979757786, geçen: 0.009808406233787537
12	bahçe: 0.04610133916139603, kuş: 0.024618767201900482, boyadık: 0.023936042562127113, duvar: 0.02381882444024086, yem: 0.0183041263371706, ağaç: 0.018291259184479713, astık: 0.016474315896630287, alan: 0.016105391085147858, temizledik: 0.014539514668285847, okul: 0.014097846113145351
13	çözüm: 0.0467151515185833, problem: 0.036365509033203125, sorun: 0.03433821350336075, bulduk: 0.0263331588357687, kılçık: 0.023403607308864594, balıkkılçık: 0.02131395973265171, bulma: 0.019094830378890038, çözme: 0.018971776589751244, çalıştık: 0.018475065007805824, balki: 0.017242515459656715
14	göz: 0.027917496860027313, oynadık: 0.02572464756667614, satar: 0.018780801445245743, oyun: 0.018206961452960968, kapmaca: 0.016762161627411842, güneş: 0.01655140332877636, kapalı: 0.014301961287856102, sırt: 0.013964914716780186, şekil: 0.012997761368751526, panel: 0.012730750255286694
15	ilgili: 0.034647513180971146, bağımlılık: 0.02783435769379139, teknoloji: 0.026659658178687096, kirlilik: 0.02408420480787754, izledik: 0.021612253040075302, konuştuk: 0.018969222903251648, video: 0.01791015826165676, çevre: 0.017094198614358902, hakkında: 0.014439052902162075, zarar: 0.012591439299285412
16	hoca: 0.03896601125597954, bize: 0.025352923199534416, sonra: 0.019586486741900444, anlat: 0.019124196842312813, öğretmen: 0.01637645997107029, ilk: 0.01151579711586237, ders: 0.011063607409596443, bir: 0.009531264193356037, ben: 0.009070390835404396, öğret: 0.008807393722236156
17	oynadık: 0.07247406989336014, oyun: 0.0652904063463211, kelime: 0.025467857718467712, oyna: 0.02294616773724556, tabu: 0.01961144618690014, körebe: 0.013334978371858597, bulut: 0.012625370174646378, dışarı: 0.012274147011339664, taş: 0.010294599458575249, kodlama: 0.00994686409831047
18	soru: 0.0540572814643383, çözdük: 0.047611866146326065, test: 0.0250952560454607, kahoot: 0.02138347551226616, uygulama: 0.017334764823317528, plickers: 0.01628449372947216, ilgili: 0.014945365488529205, video: 0.013997162692248821, izledik: 0.01325752679258585, gel: 0.013085048645734787
19	ders: 0.056082770228385925, form: 0.041036903858184814, değerlendirme: 0.034731440246105194, öğrenci: 0.031672850251197815, doldur: 0.027746230363845825, harezmi: 0.020813995972275734, doldurdük: 0.018813997507095337, hafta: 0.01757889799773693, dönem: 0.014558137394487858, katıl: 0.014277508482336998

20	grup: 0.024760261178016663, ayır: 0.014335584826767445, belirledik: 0.01367343869060278, konu: 0.012897554785013199, problem: 0.012001849710941315, sorun: 0.010808768682181835, bir: 0.010448296554386616, seçtik: 0.01007779873907566, hal: 0.008793096989393234, gruplar: 0.00874929316341877
21	çizgi: 0.02513173222541809, film: 0.024353094398975372, yazdık: 0.019732169806957245, izledik: 0.01791626587510109, doldurdük: 0.013122313655912876, resim: 0.012998087331652641, bir: 0.011995609849691391, grup: 0.011883554980158806, kağıt: 0.011719372123479843, bulduk: 0.011445373296737671
22	araştırma: 0.027716679498553276, sunum: 0.025603845715522766, ilgili: 0.023450862616300583, sunduk: 0.022771647199988365, araştırdık: 0.019762007519602776, israf: 0.01608329638838768, hakkında: 0.01509473193436861, bilgi: 0.013856451958417892, makale: 0.012985187582671642, hazırladık: 0.012166783213615417
23	kitap: 0.07529585063457489, ayraç: 0.028438016772270203, yuva: 0.02843681536614895, tavşan: 0.027565736323595047, okuma: 0.018574444577097893, kuş: 0.017934732139110565, kumbara: 0.015189087018370628, çerçeve: 0.015002784319221973, ölçtük: 0.014988082461059093, okuduk: 0.014003931544721127
24	kutu: 0.041166771203279495, geri_dönüşüm: 0.027541741728782654, malzeme: 0.015563732013106346, maket: 0.01414455659687519, getirdik: 0.013713502325117588, atık: 0.01242637075483799, çöp: 0.011705916374921799, eşya: 0.011125863529741764, hale: 0.010978342965245247, kağıt: 0.01087543461471796
25	oynadık: 0.05064567178487778, kulak: 0.04790511727333069, kulaktan: 0.04726424813270569, oyun: 0.03162314370274544, resim: 0.023238999769091606, park: 0.02309962548315525, çizdik: 0.020017825067043304, arap: 0.018108125776052475, oyna: 0.017986230552196503, tasarladık: 0.01448818203061819
26	müze: 0.07195886969566345, gittik: 0.06076326593756676, gezdik: 0.026536157354712486, git: 0.024808639660477638, cezri: 0.02085636556148529, gezi: 0.02060834690928459, istanbul: 0.0201510451734066, üniversitesi: 0.01761964149773121, diyalog: 0.01685377210378647, ora: 0.016205303370952606
27	koyduk: 0.021516457200050354, elma: 0.020079242065548897, ali: 0.01780601218342781, kuşçu: 0.01749904826283455, boş: 0.017073770985007286, tuz: 0.016943758353590965, gözlemevi: 0.016558606177568436, yerine: 0.015372983179986477, kap: 0.015128057450056076, koyup: 0.014696958474814892
28	devam: 0.08911653608083725, ettik: 0.06605947762727737, yapma: 0.03572413697838783, maket: 0.03478372097015381, başladı: 0.02580033615231514, çalışma: 0.02262682281434536, çalıştık: 0.018747100606560707, tasarım: 0.013643057085573673, proje: 0.011520326137542725, tamamlama: 0.011342106387019157
29	yedik: 0.03996149078011513, yemek: 0.029706837609410286, yedi: 0.020436955615878105, izledik: 0.01795324683189392, şarkı: 0.016300726681947708, dans: 0.015676293522119522, oynadık: 0.01452627032995224, gül: 0.012122618965804577, ettik: 0.010642463341355324, sağlıklı: 0.010519683361053467
30	fikir: 0.04668396711349487, arkadaş: 0.03612940013408661, dinle: 0.02492028847336769, sun: 0.017242534086108208, öğretmen: 0.01696019619703293, paylaş: 0.015630023553967476, söyle:

	0.014005366712808609, üret: 0.012685098685324192, hakkında: 0.012648004107177258, düşünce: 0.012347670271992683
31	led: 0.04640316218137741, yaktık: 0.042374107986688614, arduino: 0.027651457116007805, ışık: 0.02529945783317089, sensör: 0.0232862439006567, devre: 0.020207643508911133, kod: 0.01977071352303028, lamba: 0.016313156113028526, çalıştırdık: 0.015881411731243134, arduino: 0.01421435084193945
32	izledik: 0.07542121410369873, video: 0.06431646645069122, izle: 0.03922928869724274, ilgili: 0.03691449761390686, film: 0.03430517762899399, slayt: 0.019775381311774254, sunum: 0.017921360209584236, kısa: 0.015629570931196213, bir: 0.013628985732793808, arkadaş: 0.012561957351863384
33	ilgili: 0.025512441992759705, sağlıklı: 0.020686930045485497, bulmaca: 0.020624669268727303, beslenme: 0.01949531026184559, oynadık: 0.01609458588063717, işledik: 0.015314193442463875, çözdük: 0.013994586654007435, konu: 0.013622594997286797, atık: 0.012938734143972397, oyun: 0.012808933854103088
34	geçen: 0.025770574808120728, kalan: 0.02304878458380699, hafta: 0.022903814911842346, tamamladık: 0.022101767361164093, yarım: 0.019011199474334717, tişört: 0.01596130058169365, maket: 0.014891192317008972, boyadık: 0.013135493732988834, proje: 0.012489666230976582, robot: 0.012140890583395958
35	ders: 0.034870970994234085, ilk: 0.02627597376704216, sonra: 0.022961027920246124, ikinci: 0.014192849397659302, bir: 0.013536231592297554, önce: 0.012388369999825954, oynadık: 0.009702077135443687, oyun: 0.007819043472409248, kağıt: 0.007129736244678497, olarak: 0.0061814491637051105
36	sistem: 0.02168719470500946, site: 0.018499255180358887, akıllı: 0.016162147745490074, yapım: 0.013127530924975872, form: 0.01178017072379589, öğren: 0.011195030063390732, code_org: 0.011014708317816257, üst: 0.010780470445752144, maket: 0.010618636384606361, oluşturduk: 0.009734565392136574
37	gittik: 0.06403873860836029, git: 0.022976793348789215, müdürlük: 0.018567165359854698, ziyaret: 0.018224725499749184, eğitim: 0.017124207690358162, gezi: 0.015259455889463425, ilçe: 0.014395547099411488, tesis: 0.013512278907001019, fabrika: 0.013479884713888168, etik: 0.012780454009771347
38	tiyatro: 0.06373559683561325, rol: 0.03323104605078697, senaryo: 0.029271282255649567, prova: 0.025874193757772446, çalıştık: 0.01869737170636654, drama: 0.017736954614520073, metin: 0.0171988382935524, skeç: 0.016879256814718246, oynadık: 0.013273362070322037, gösteri: 0.013165702112019062
39	sihirli: 0.0389738529920578, kule: 0.030138926580548286, arap: 0.022927554324269295, sirke: 0.02159407176077366, görev: 0.021152755245566368, deterjan: 0.016553718596696854, kaldır: 0.016390157863497734, parmak: 0.015994222834706306, dağılım: 0.01580720767378807, kimya: 0.015636002644896507
40	boya: 0.055222753435373306, boyadık: 0.0387420579791069, masa: 0.016153547912836075, resim: 0.014948450960218906, boyama: 0.012554922141134739, meyve: 0.01214753556996584, köpek: 0.011958310380578041, yapıştır: 0.011906235478818417, hamur: 0.011862696148455143, sprey: 0.011788027361035347

41	soru: 0.045975785702466965, sorduk: 0.02643072046339512, dinledik: 0.0213889479637146, dinle: 0.02036423049867153, seminer: 0.019821777939796448, cevap: 0.017267346382141113, sunum: 0.016128042712807655, sor: 0.016017770394682884, sınıf: 0.015817243605852127, veteriner: 0.014612136408686638
42	oynadık: 0.040182262659072876, oyun: 0.03240193799138069, elektrik: 0.030850213021039963, harita: 0.023147566244006157, akım: 0.021125320345163345, devre: 0.019494229927659035, kurduk: 0.01702921837568283, beyin: 0.013568279333412647, kavram: 0.012724512256681919, oyna: 0.012159176170825958
43	günü: 0.028841184452176094, oyna: 0.025000981986522675, oyun: 0.021250741556286812, oynadık: 0.020668385550379753, yarışma: 0.019494226202368736, anne: 0.016279686242341995, ilgili: 0.016201253980398178, resim: 0.014964153990149498, kitap: 0.014155014418065548, yaz: 0.013509471900761127
44	afiş: 0.04516854137182236, hazırladık: 0.034996699541807175, slogan: 0.03161981701850891, hikaye: 0.030571451410651207, yaz: 0.02894054353237152, yazdık: 0.026831908151507378, hazırla: 0.02384420856833458, ilgili: 0.02037663944065571, çanta: 0.017589516937732697, poster: 0.017262229695916176
45	oda: 0.0214504636824131, bilişim: 0.02144346944987774, sınıf: 0.019286662340164185, sonra: 0.017921075224876404, bilgisayar: 0.01161813735961914, gittik: 0.011510116048157215, ben: 0.010237440466880798, git: 0.010152310132980347, ilk: 0.009783382527530193, bir: 0.009481329470872879
46	arduino: 0.042331937700510025, devre: 0.028504839166998863, block: 0.017981018871068954, uygulama: 0.017613723874092102, yapma: 0.016397587954998016, program: 0.016212059184908867, enerji: 0.015494566410779953, öğrendik: 0.014775725081562996, kurduk: 0.01427379995584488, çalışma: 0.014172251336276531
47	düğüm: 0.022932330146431923, oynadık: 0.020864352583885193, doldurdık: 0.019873397424817085, form: 0.018261387944221497, oyun: 0.017830513417720795, deney: 0.01640477403998375, kahoot: 0.014260200783610344, etkinlik: 0.013965134508907795, bir: 0.013234145008027554, video: 0.013082416728138924
48	fotoğraf: 0.026395881548523903, çekildik: 0.0177619606256485, yemek: 0.016636576503515244, çöp: 0.013782563619315624, aldık: 0.013768495991826057, süsledik: 0.012794127687811852, topladık: 0.012253777123987675, yedik: 0.012188924476504326, bez: 0.011751811020076275, topla: 0.011169081553816795
49	ağaç: 0.026698878034949303, yapıştırdık: 0.01260237954556942, atık: 0.012512221932411194, malzeme: 0.012381977401673794, maket: 0.012271522544324398, hazırladık: 0.009550013579428196, hayvan: 0.008342631161212921, atladık: 0.008250842802226543, köprü: 0.008034389466047287, karton: 0.007776238489896059
50	kahoot: 0.035476282238960266, uygulama: 0.03136591985821724, oynadık: 0.02901162952184677, oyun: 0.026153860613703728, hayvan: 0.02567487396299839, yarışma: 0.019342033192515373, soru: 0.01815878413617611, hazırladık: 0.013110427185893059, oyna: 0.013008259236812592, çanta: 0.012864836491644382
51	konuştuk: 0.022137081250548363, hakkında: 0.019192850217223167, fikir: 0.01669757440686226, proje: 0.016628187149763107, tartıştık: 0.01593051664531231, problem: 0.015379960648715496, neler:

	0.014733774587512016, belirledik: 0.014609132893383503, ürettik: 0.014261584728956223, düşündük: 0.014120078645646572
52	oyun: 0.03875765949487686, oynadık: 0.029224418103694916, etkinlik: 0.024502847343683243, oyna: 0.02059059776365757, birbiri: 0.017631199210882187, tanışma: 0.017188027501106262, bir: 0.014740575104951859, arkadaş: 0.01461884006857872, tanıştık: 0.014177919365465641, eğlenceli: 0.013619961217045784
53	kurt: 0.044420916587114334, oynadık: 0.02651125192642212, oyun: 0.024297652766108513, program: 0.019174089655280113, kodlama: 0.018156955018639565, kurduk: 0.017771221697330475, piyano: 0.017239095643162727, kuzu: 0.016669444739818573, fil: 0.01594764180481434, scratch: 0.015111333690583706



EK 2. S2 Sorusu Konu Kelime Ağırlıkları Tablosu.

k	Kelimeler
0	program:0,0300162266939878 kod:0,0270918663591146 uygulama:0,0268208067864179 kodlama:0,0242700614035129 ekran:0,0226221084594726 block:0,0219761319458484 lcd:0,0219538435339927 arduino:0,0214307028800249 bilgisayar:0,0209271889179945
1	yeni:0,0230950415134429 arkadaş:0,0213593933731317 öğretmen:0,0190049596130847 oyun:0,0168465860188007 cevap:0,0164370220154523 isim:0,0149987582117319 harezmi:0,014220543205738 bir:0,0124736921861767 ders:0,00983862485736608
2	sorun:0,0348888263106346 zorluk:0,0347497388720512 yaşadık:0,0344262421131134 hayat:0,0224378053098917 görme:0,0215652994811534 engel:0,0141845103353261 çektik:0,0138766225427389 çözüm:0,0138061959296464 günlük:0,0126727754250168
3	hazırlama:0,0462375059723854 anket:0,0417690314352512 yapma:0,0336961708962917 çanta:0,0327724851667881 program:0,0272387191653251 kullanma:0,0218909587711095 poster:0,020200353115797 soru:0,019281655550003 uygulama:0,0179874412715435
4	oynama:0,0434557758271694 yeni:0,0299391020089387 oyundık:0,0274445116519927 yapma:0,0207867827266454 oynadık:0,015380092896521 kitap:0,0152623821049928 bir:0,0135184675455093 okuma:0,0133876986801624 taş:0,0124475359916687
5	hayvan:0,0245659817010164 ses:0,0187445953488349 çevre:0,0183638669550418 yıl:0,016604846343398 atık:0,0160121992230415 kalori:0,0145146828144788 kaç:0,013320236466825 doğa:0,0129850069060921 hava:0,012580232694745
6	kirlilik:0,029243966564536 çevre:0,021937184035778 ana:0,0214535407721996 ışık:0,0187036264687776 temiz:0,0185191947966814 ara:0,0178031362593173 tutma:0,0139377377927303 renkler:0,0127103757113218 toprak:0,0124533548951148
7	geri:0,0326460488140583 kağıt:0,0322334431111812 gübre:0,0249357968568801 yapım:0,0224661342799663 organik:0,0196129940450191 tarım:0,0196095071732997 yapıldık:0,0174666699022054 yapma:0,0156486928462982 madde:0,0149522172287106
8	problem:0,0232583489269018 bir:0,0213763657957315 arkadaş:0,0194905046373605 çözüm:0,0173985939472913 sürü:0,0105548668652772 farklı:0,00851429626345634 fark:0,00848770327866077 yeni:0,0084401099011302 insan:0,00839587487280368
9	direnç:0,0274765510112047 mimik:0,0263019129633903 dans:0,0243201814591884 oyun:0,020628735423088 elektrik:0,0185382533818483 jest:0,0171874854713678 yapma:0,0127047076821327 tiyatro:0,0125974640250206 rol:0,0120981317013502
10	küresel:0,0529374778270721 dış:0,0288120508193969 sağ:0,0260564070194959 yapma:0,0240246821194887 kuş:0,0228053722530603 sol:0,0189351364970207 yuva:0,0186039209365844 kuşlar:0,0171067025512456 öğrenme:0,0167094208300113

11	kodlama:0,0374403558671474 robotik:0,0226832889020442 yaptık:0,0180327072739601 alet:0,015436322428286 ses:0,0142938923090696 rol:0,0140954591333866	müzik:0,0240499936044216 tiyatro:0,0196052081882953 drama:0,0145144471898674
12	mesafe:0,0491679087281227 yapma:0,0265944320708513 nem:0,0236159861087799 arduino:0,0177572630345821 ultrasonik:0,0152339152991771	motor:0,027433056384325 sıcaklık:0,0239373017102479 sistem:0,018556572496891 servi_motor:0,0167138464748859
13	sorun:0,0511622913181781 çözüm:0,0453718230128288 alt:0,0287960711866617 çözme:0,0247512273490428 bir:0,0197962280362844 algoritma:0,0155845005065202 başlık:0,0146667119115591	bulma:0,0220465455204248 yol:0,0161136034876108
14	tarihi:0,0438107103109359 eser:0,0283775385469198 bölge:0,0250049885362386 günü:0,0175382085144519 hakkında:0,0157464109361171 sokak:0,012950737029314 hak:0,0127716390416026	yer:0,016974251717329 plaka:0,0148312821984291
15	hafta:0,018704256042838 harezmi:0,0158120598644018 tekrar:0,00903317052870988 bildik:0,00888435635715723 bugün:0,00813944730907678	yaptık:0,0186891872435808 bile:0,00990537833422422 yeni:0,00890565384179353 ders:0,00859748478978872
16	atık:0,0334897823631763 malzeme:0,0246706213802099 kutu:0,0171854961663484 çizgi:0,0156133892014622 eşya:0,0148770725354552 ayraç:0,0142075112089514 kağıt:0,0134978359565138	yapma:0,0265662241727113 kitap:0,0232535116374492
17	piramit:0,0566381141543388 kilo:0,0297548063099384 element:0,0260563492774963 prizma:0,0229557044804096 yağ:0,0195785481482744 boy:0,0187657680362463	zincir:0,0322109758853912 kelebekler:0,028863763436675 gerçeklik:0,0241723954677581
18	uçak:0,0327128209173679 web:0,0315288715064525 harita:0,0309423077851533 yapım:0,0239485297352075 araç:0,0191084705293178 site:0,0174975488334894 balıkkılçık:0,0137133067473769	uygulama:0,0224961396306753 kavram:0,0178991444408893
19	kunduz:0,0170942954719066 site:0,0168432332575321 çanta:0,0141086736693978 oyun:0,0115696163848042 oluşturma:0,0109910769388079	bilge:0,0170486643910408 karikatür:0,015632027760148 uygulama:0,0128500647842884 birlik:0,0115424552932381
20	sağlıksız:0,0265181101858615 zararlı:0,0216289609670639 zarar:0,0204091519117355 besin:0,0170353613793849 gıda:0,0167783536016941 ürün:0,0152192115783691	yiyecek:0,0236256457865238 şeker:0,021460510790348 beslenme:0,0174125880002975
21	gerektik:0,0410540886223316 neler:0,036351915448904 yapılma:0,0228125844150781 yaparken:0,0172600671648979 maket:0,011837787926197	dikkat:0,0394167862832546 gereken:0,0231095105409622 etme:0,0219889525324106 ölçü:0,0164433158934116
22	teknik:0,034550841897726 çalışma:0,0228794813156127 yazma:0,0209552366286516 parlak:0,0167515762150287 arkadaş:0,0149864554405212	hikaye:0,0326566062867641 düşünme:0,0221720933914184 grup:0,0189183801412582 yapma:0,0158662144094705

23	zorbalık:0,0361211150884628 kötü:0,0195163432508707 yapmama:0,0170121788978576 güvenlik:0,015803288668394 yangın:0,0151592325419187	oran:0,0253940168768167 orantı:0,0190715342760086 orman:0,0161014590412378 siber:0,0154214035719633
24	trafik:0,0564239397644996 kural:0,0391353406012058 yapma:0,0226091537624597 balki:0,0213146731257438 akış:0,0190941616892814	zihin_harita:0,0540463775396347 kılçık:0,0286811646074056 şema:0,0213887635618448 teknik:0,0212753992527723
25	herkes:0,0215164776891469 insan:0,0141346435993909 düşünce:0,0117293773218989 bakma:0,010777092538774	açı:0,0182047840207815 bakış:0,0155386757105588 bir:0,0122425202280282 arkadaş:0,0115863708779215 gereklik:0,00934894196689128
26	devre:0,0500877201557159 arduino:0,0434694811701774 ışık:0,0225224755704402 ardunio:0,017287915572524 elektrik:0,0155524080619215	yakma:0,046901486814022 kurmay:0,0231532156467437 buton:0,0184766147285699 yapma:0,0157652255147695
27	boyama:0,0363710150122642 tişört:0,0224979929625988 boyan:0,0145597727969288 katlama:0,0135485883802175 kelime:0,012574641034007	boya:0,0357692763209342 bulut:0,0161047726869583 çanta:0,014188633300364 seramik:0,0128885516896843
28	çocuk:0,0239739101380109 oyun:0,020891373977065 makine:0,0172575898468494 bir:0,0135692423209548	icat:0,0229897964745759 hak:0,0217941869050264 eski:0,0205739438533782 yaptık:0,0182529930025339 araba:0,0171591602265834
29	yaradık:0,0482898652553558 ses:0,0259621553122997 parça:0,0173066034913063 kullanıldık:0,0134930340573191 arduinonun:0,0122480355203151	malzeme:0,0284619349986314 yalıtım:0,0216283723711967 arduino:0,0157136134803295 alet:0,0127969160676002
30	ingilizce:0,0306389983743429 tekrar:0,0255081951618194 anlam:0,0145054748281836 ettik:0,0122326025739312	şarkı:0,0282801706343889 yazma:0,0153227783739566 yeni:0,0140966288745403 söz:0,0123778516426682 sistem:0,0107768354937434
31	şekil:0,031066108494997 dikdörtgen:0,0215216260403394 prizma:0,0203179139643907 ölçme:0,0184215381741523 özellik:0,0151150356978178	geometrik:0,026136202737689 hacim:0,0204132050275802 alan:0,0196647364646196 kare:0,0153216067701578
32	yapacak:0,0444718226790428 yapma:0,0263671223074197 anket:0,0212144162505865 pasta:0,0197370294481515 excel:0,0171933341771364	şenlik:0,0426235273480415 neler:0,0224488954991102 tablo:0,0202487744390964 harezmi:0,0191995166242122
33	yapma:0,0316775105893611 hareket:0,0140450671315193 sanat:0,0138657670468091 arap:0,0106106083840131 plan:0,0100109325721859	kodlama:0,0150411445647478 oyun:0,0138680562376976 temizlik:0,0116523671895265 gereklik:0,0100309196859598

34	teknoloji:0,0304040592163801 kötü:0,0188097469508647 şiddet:0,0174325313419103 hayvan:0,0165756568312644 bir:0,013938239775598 bağımlı:0,0127602946013212 sağlık:0,0117291323840618	bağımlılık:0,0207055993378162 güvenli:0,0180102996528148
35	grup:0,061147142201662 takım:0,0509481728076934 olma:0,0360864959657192 birlikte:0,0249959211796522 birlik:0,021237002685666 önem:0,0205649323761463 hal:0,0168766174465417 yapma:0,0168105959892272 arkadaş:0,0150504745543003	
36	doğay:0,0332804322242736 hayvan:0,0258887819945812 verdik:0,0215851329267025 yaşam:0,015341811813414 kadar:0,0138939525932073	çevre:0,029444545507431 alan:0,0225090123713016 koruma:0,0208192355930805 vermeme:0,0152055658400058
37	yer:0,0210056323558092 adım:0,0181441847234964 balon:0,0168265104293823 boşa:0,016097892075777 deniz:0,0131413266062736 atma:0,0121538722887635	atmama:0,0184586141258478 oyun:0,0173583757132291 uzay:0,014298690482974
38	zarar:0,0449294261634349 bağımlılık:0,0355978198349475 obezite:0,0203697867691516 olumsuz:0,0190837252885103 diyafram:0,0130644710734486	etki:0,0372167155146598 sosyal_medya:0,0315714590251445 yarar:0,0192002523690462 insan:0,0179859045892953
39	tohum:0,0447523109614849 toprak:0,0372585281729698 dikme:0,0211384184658527 ekildik:0,0170296970754861 ekilir:0,015401535667479	ekme:0,0398088246583938 çiçek:0,0333800837397575 limon:0,0188848096877336 fidan:0,0166664626449346
40	yapma:0,0259826462715864 kelime:0,0246644672006368 dikme:0,0233513452112674 çiçek:0,0216655544936656 dil:0,0216420982033014 altın:0,0204680822789669	yeni:0,0255770981311798 ingilizce:0,0235163681209087 oran:0,0230762083083391
41	yapım:0,030545350164175 yoğurt:0,0244072210043668 hamur:0,0218771249055862 yapılır:0,0181753933429718 origami:0,016894694417715	turşu:0,0283672623336315 kuş:0,0236914940178394 yapıldık:0,0211538430303335 sabun:0,0176660902798175
42	film:0,0463717840611934 sebze:0,0357981659471988 beslenme:0,0209425650537014 ekmek:0,0194069631397724 dengeli:0,0181597955524921	çekme:0,037557877600193 sağlıklı:0,0247261803597211 kısa:0,0203168410807847 yapma:0,0183853451162576
43	görev:0,0130488444119691 kadar:0,010765777900815 engelli:0,00906291790306568 rol:0,00804606173187494 ders:0,00758417695760726	zor:0,0116315046325325 bir:0,0113148195669054 seçim:0,00908105727285146 önemli:0,00905587896704673
44	sorun:0,0279865954071283 harezmi:0,0184826869517564 adam:0,0140071492642164 önemli:0,0115450592711567 algoritma:0,0104045569896698	hayat:0,027293162420392 bilim:0,0182052943855524 fikir:0,0119266510009765 bir:0,0111361993476748
45	rüzgar:0,0358495078980922 kaynak:0,0311884079128503 gezegen:0,0246928259730339	güneş:0,0342472493648529 elektrik:0,0310575179755687 panel:0,0178052335977554

	yenilenebilir:0,0165180657058954 çeşit:0,0108351400122046	uzay:0,0109249362722039
46	hikaye:0,0580571442842483 hiscümlesi:0,0477473884820938 afiş:0,0351505391299724 oluşturma:0,0212091021239757 kurmay:0,017478097230196	yazma:0,0508424453437328 his:0,0392267368733882 yapma:0,0215859860181808 hazırlama:0,0209219120442867



EK 3. S3 Sorusu Konu Kelime Ağırlıkları Tablosu.

k	Kelimeler
0	yanlış:0,0390363112092018 kork:0,0371930189430713 yaparken:0,0288017950952053 biraz:0,0248696487396955 heyecanlan:0,023687595501542 elim:0,0217117704451084 acı:0,0190440248697996 heyecanlı:0,0177715960890054 şaşkın:0,0143415043130517 cevap:0,0134329106658697
1	anne:0,0402626730501651 insan:0,0255301427096128 yardım:0,0248834192752838 hayvan:0,0214844234287738 limon:0,0197409484535455 üzü:0,018893539905548 doğay:0,0188083834946155 çevre:0,0167225915938615 doğa:0,0161841455847024 temiz:0,0157782025635242
2	süper:0,0913840904831886 harika:0,0813495144248008 muhteşem:0,0695410445332527 iyi:0,0660620406270027 meraklı:0,0586296468973159 mükemmel:0,0585735216736793 heyecanlı:0,0530591569840908 dolu:0,0412596315145492 yardımsever:0,0374801307916641 sevgi:0,0372616462409496
3	öğretmen:0,0535687543451786 teşekkür:0,0432077459990978 bütün:0,0330056622624397 eğlenceli:0,0295916814357042 iyi:0,0291978456079959 sev:0,0284315887838602 yaptık:0,0232598204165697 eğlen:0,0229224171489477 geçti:0,0212700404226779 sevdi:0,0203752741217613
4	sunum:0,032596006989479 heyecanlan:0,0295758470892906 yaparken:0,0289234090596437 yaptık:0,0235261972993612 eğlen:0,0230964682996273 drama:0,0176886394619941 biraz:0,0165383890271186 eğlenceli:0,0159481819719076 zorlan:0,0147371031343936 sıkıl:0,0132654281333088
5	bilim:0,107384473085403 kendi:0,0879328921437263 adamı:0,053571555763483 bir:0,0529255792498588 insan:0,0467699468135833 tok:0,0406089015305042 yemek:0,0363913625478744 mucit:0,026683572679758 tasarımcı:0,0248799119144678 değişik:0,0229610074311494
6	kendi:0,0654799789190292 mühendis:0,056697554886341 bilgisayar:0,0474943481385707 bir:0,0326743237674236 sağlıklı:0,0315608754754066 başarılı:0,0300621837377548 teknoloji:0,0225785300135612 alan:0,0222595147788524 ressam:0,0215234812349081 uğraşmak:0,0209283847361803
7	gururlu:0,0798914209008216 heyecanlı:0,0656619817018508 kendi:0,0651286467909812 başarılı:0,0576708056032657 neşeli:0,0576475709676742 sevinçli:0,0569082647562027 çalışkan:0,0502690598368644 eğlenmiş:0,0490478686988353 huzurlu:0,047794371843338 zeki:0,0453999489545822
8	etik:0,0466141328215599 etmek:0,0390599928796291 dans:0,0337844900786876 ifade:0,0305532179772853 düşünce:0,029295191168785 hayal_güç:0,0271126814186573 öğrendik:0,0253159869462251 sohbet:0,0228799656033515 kendi:0,0224308464676141 yardım:0,0214006528258323
9	biraz:0,0599243193864822 sıkıcı:0,0433332473039627 normal:0,0344715118408203 kötü:0,0266619585454463 mutsuz:0,0242950450628995 iyi:0,023344837129116 üzgün:0,0229148361831903 yorgun:0,0221169311553239 uyuklu:0,0187977850437164 sıkıl:0,0181714706122875
10	ders:0,0882615447044372 eğlenceli:0,0775806158781051 komik:0,0676680877804756 heyecanlı:0,049545120447874 yorucu:0,0402927994728088 eğlen:0,037517812103033 bir:0,0341782048344612 boyunca:0,0320419073104858 keyifli:0,0299872234463691 neşeli:0,0295153502374887
11	oyun:0,0598322078585624 oynadık:0,0380208641290664 eğlen:0,0322036668658256 eğlenceli:0,0311784129589796 yaptık:0,0242132525891065 yap:0,0226548723876476 resim:0,0223068743944168 oynarken:0,0200450550764799 oyna:0,0192784704267978 eğlendik:0,019180042669177
12	kendi:0,106235019862651 özel:0,0558331534266471 huzurlu:0,0464681237936019 mimar:0,0420555286109447 garip:0,0386727936565876 güven:0,0385592766106128 yazar:0,0361717939376831 heyecanlı:0,0303959809243679 art:0,0288940351456403 önemli:0,0284130107611417

k	Kelimeler
13	üzül:0,0452433973550796 hayvan:0,030314901843667 engelli:0,0298924650996923 insan:0,0267736315727233 kötü:0,0166850313544273 zorluk:0,0160977821797132 empati:0,0155892288312315 olduk:0,0153288412839174 görünce:0,0123317399993538 durum:0,0118800476193428
14	geçti:0,092643491923809 eğlenceli:0,083493061363697 iyi:0,0768465623259544 ders:0,046449389308691 keyifli:0,0456560589373111 heyecanlı:0,0429171472787857 bugün:0,0381176508963108 zamanki:0,0380475297570228 eğlen:0,0374150462448596 kendi:0,0333634503185749
15	maket:0,0434817411005497 yaparken:0,0348566435277462 heyecanlan:0,0308530982583761 eğlen:0,0302799865603446 yaptık:0,0297589711844921 eğlenceli:0,0216245763003826 oyun:0,0208394639194011 oyun:0,0208379682153463 yarışma:0,0205579400062561 yapmak:0,0193519350141286
16	etkinlik:0,0394908152520656 bir:0,0291622597724199 böyle:0,027420425787568 yaptık:0,0229611881077289 ısınma:0,0201112385839223 ders:0,0199698749929666 proje:0,0185802914202213 sevdi:0,0184895396232604 ben:0,0182719752192497 olduk:0,0175095926970243
17	sorun:0,066463716328144 çözüm:0,0557550452649593 problem:0,0467514880001544 bulduk:0,0237003844231367 bulma:0,0186820998787879 fikir:0,0167533177882432 bulmak:0,015128337778151 olduk:0,0140901366248726 çözme:0,0139679992571473 bir:0,0129478303715586
18	kedi:0,0401391088962554 yaptık:0,0272261202335357 kuş:0,0245966725051403 bahçıvan:0,0203873720020055 yemek:0,0199659895151853 kendi:0,0199210084974765 yuva:0,0174003448337316 köpek:0,0164084397256374 üzül:0,0163729935884475 boya:0,0151761332526803
19	afiş:0,0562770031392574 soru:0,0405541025102138 anket:0,0378153882920742 yaparken:0,0377874225378036 eğlen:0,0337055139243602 yaptık:0,0316668413579463 heyecanlan:0,0283485241234302 hazırlarken:0,0206550154834985 hazırladık:0,0203668624162673 çözerken:0,0202786494046449
20	hayal:0,0461028143763542 kırıklık:0,0369715131819248 his:0,0368081405758857 turşu:0,0364285930991172 sınır:0,0312737338244915 umutlu:0,0305185597389936 değişik:0,0293986126780509 merak:0,0283905826508998 heyecan:0,0282857157289981 önce:0,0226282160729169
21	heyecan:0,0634830147027969 dolu:0,0620055086910724 sevinç:0,0464069694280624 korku:0,0404443070292472 sevgi:0,0388479232788085 kıpır:0,0377260260283947 aşırı:0,0328933894634246 derece:0,0327985547482967 enerjik:0,0313248075544834 içim:0,0290221404284238
22	hafta:0,0436225645244121 ders:0,0343412309885025 geçen:0,0233087055385112 son:0,0222999341785907 göre:0,0210273191332817 geçti:0,0209366679191589 sıkıcı:0,0205591917037963 eğlenceli:0,0188893321901559 sıkıl:0,0186190884560346 biraz:0,0179852955043315
23	oyun:0,058858685195446 öğrendik:0,0328944958746433 şaşır:0,0278455205261707 oynadık:0,0273927506059408 oynarken:0,0269264075905084 öğren:0,0267164651304483 yeni:0,0263393633067607 bilgi:0,0225823540240526 kelime:0,0214339345693588 eğlen:0,0202803798019886
24	şarkı:0,0342014729976654 gel:0,0289293713867664 biraz:0,0257710963487625 sıkıl:0,0240497626364231 söyledik:0,0217392984777688 heyecanlan:0,0177560690790414 sıra:0,0153309814631938 fikir:0,0148019948974251 uyku:0,0146185932680964 ders:0,0143883004784584
25	şaşır:0,0424707010388374 üzül:0,0330165587365627 zarar:0,0291956067085266 görünce:0,028401406481862 insan:0,0241192206740379 kadar:0,0206585321575403 şeker:0,0182054806500673 öğrenince:0,0180165972560644 olduk:0,0179886668920516 üzgün:0,0151061797514557
26	yeni:0,137731447815895 öğren:0,0874007269740104 öğrendik:0,081406146287918 bilgi:0,0784229487180709 öğrenmek:0,0568233132362365 bir:0,032608076930046 ben:0,0301792044192552 iyi:0,0255523286759853 eğlen:0,0214286316186189 heyecanlı:0,0155581291764974

k	Kelimeler
27	sorun:0,0291794650256633 problem:0,0274421405047178 kadar:0,0254689306020736 olduk:0,0250155609101057 hayat:0,0242862831801176 fark:0,0173898637294769 kendi:0,0146536584943532 günlük:0,0145719032734632 ben:0,0139731857925653 üzül:0,0136758713051676
28	duy:0,0745214521884918 gurur:0,0699127316474914 kendi:0,0445843338966369 olmak:0,0350234657526016 bura:0,0340077094733715 olduk:0,0330137908458709 şanslı:0,0245980452746152 çalışma:0,0241609308868646 bir:0,0237655322998762 proje:0,0226802099496126
29	heyecan:0,0818749442696571 öğrenme:0,0710358694195747 sevinç:0,0518585033714771 yaşa:0,0504066944122314 kafa:0,0388262197375297 utanç:0,0339091345667839 karışık:0,033868808299303 korku:0,0329434126615524 verdik:0,0318243242800235 karış:0,0307225789874792
30	bazen:0,0467654317617416 biraz:0,0416819229722023 sinirlen:0,0220287628471851 genel:0,0200953539460897 eğlen:0,0199472326785326 zorlan:0,0196371767669916 sıklı:0,0193181280046701 olarak:0,0189731065183877 üzü:0,0176629982888698 kaybettik:0,0157476738095283
31	eğlen:0,0467564724385738 yaparken:0,0373866222798824 biraz:0,0366762951016426 heyecanlan:0,030368372797966 etkinlik:0,0283526554703712 stresli:0,0279009118676185 gül:0,0271402299404144 yorul:0,0268287789076566 heyecanlı:0,0236696302890777 eylen:0,0232530292123556
32	yapacak:0,0421210564672946 heyecanlan:0,0334278978407382 heyecanlı:0,0331061296164989 şenlik:0,0318710282444953 merak:0,0243192389607429 olacak:0,0226748324930667 ilk:0,0201639756560325 meraklı:0,0155599033460021 proje:0,0143118910491466 sınıf:0,0137000801041722
33	bir:0,0404387563467025 değişik:0,0317926593124866 garip:0,028008023276925 normal:0,027945015579462 olduk:0,0252072494477033 his:0,0228823795914649 birinci:0,0218495465815067 kahootta:0,0183696281164884 farklı:0,0176497232168912 oyun:0,017463345080614
34	kötü:0,0231759622693061 iyi:0,0230569187551736 değil:0,0226783696562051 fazla:0,0221835765987634 yan:0,0212372746318578 fena:0,0199776161462068 bir:0,0186538118869066 kendi:0,0180579610168933 bilim:0,0156304705888032 küçük:0,0154951112344861
35	hikaye:0,0983151420950889 yazarken:0,0631414204835891 yazdık:0,0500675179064273 eğlen:0,0318887941539287 heyecanlan:0,0268904957920312 okurken:0,024920780211687 slogan:0,023612966760993 gül:0,0213291700929403 komik:0,0211151000112295 yazmak:0,017148008570075
36	kötü:0,019620232284069 ben:0,01754667237401 gel:0,017478359863162 mide:0,0154702980071306 bir:0,0154693042859435 bulan:0,0139958802610635 bura:0,0132598737254738 okul:0,0124246478080749 iste:0,0121388733386993 kişi:0,0108527019619941
37	fikir:0,0252071507275104 kendi:0,0218349769711494 bir:0,0209821574389934 zaman:0,0175963137298822 düşünce:0,017313377931714 rahat:0,0157085862010717 aynı:0,0154355512931942 özgür:0,0153219383209943 sıcak:0,0144668174907565 ortam:0,0143707105889916
38	biraz:0,0592908337712287 film:0,0456922501325607 ağrı:0,0408149659633636 sıklı:0,0403223112225532 yorul:0,0401757657527923 yorgun:0,0331260114908218 acı:0,0280778855085372 kötü:0,0262695886194705 baş:0,0227338261902332 hüzünlü:0,0219968892633914
39	olma:0,0400369241833686 müze:0,025130508467555 bir:0,0249320790171623 olduk:0,0224312096834182 gezi:0,0210064519196748 gerektik:0,0183282848447561 kendi:0,0177302956581115 proje:0,0166008658707141 güçlü:0,0159419924020767 grup:0,0140357874333858
40	bilgi:0,0729987919330596 sahip:0,0578685440123081 yeni:0,0557033009827137 sorumluluk:0,0520913712680339 hoş:0,0416534058749675 edindik:0,0361594334244728 git:0,0358723551034927 öğrenmek:0,0305497907102108 öğrenmiş:0,0205317996442317 edin:0,0182426832616329

k	Kelimeler
41	ilk:0,047662578523159 sonra:0,042691845446825 biraz:0,0367796756327152 başta:0,0347747653722763 sıkı:0,0290689393877983 heyecanlan:0,0217166431248188 ders:0,0214734468609094 önce:0,0208340361714363 eğlen:0,0207520145922899 kork:0,0189258120954036
42	yeni:0,0364947691559791 bilgi:0,0352124758064746 öğrendik:0,0301469098776578 kodlama:0,0230398662388324 bir:0,0224068462848663 merak:0,0192416962236166 arkadaş:0,0182002037763595 öğrenme:0,0177499037235975 heyecan:0,0165873225778341 hafta:0,0155756883323192
43	yapma:0,0267424322664737 bir:0,0229143071919679 öğren:0,0228215437382459 led:0,0221613105386495 iyi:0,0218263305723667 ilginç:0,0216243788599967 eser:0,0209417585283517 yeni:0,0191188603639602 tarihi:0,0190455988049507 öğrendik:0,0190285481512546
44	heyecanlı:0,0918632969260215 meraklı:0,0650405287742614 düşünceli:0,063565544784069 bilgili:0,0567056760191917 neşeli:0,051520999521017 sevinçli:0,0433962978422641 zeki:0,036953043192625 şaşkın:0,0345178209245204 akıllı:0,0324331521987915 kendi:0,0303276665508747
45	oyun:0,0511870719492435 gezi:0,0432775951921939 oynarken:0,0361321419477462 eğlen:0,0263702124357223 gittik:0,0252830721437931 heyecanlan:0,0209255516529083 bekle:0,0187141671776771 hafta:0,0183032676577568 git:0,0165562033653259 gelecek:0,0164976064115762
46	video:0,0581634454429149 oyun:0,024681681767106 izlerken:0,024232005700469 izledik:0,0210906956344842 biraz:0,0177173838019371 oynarken:0,0169747043401002 heyecanlı:0,0148451020941138 üzül:0,0145312407985329 kazandık:0,0143397348001599 izle:0,0142542775720357
47	eğlendik:0,0446082018315792 hoca:0,0358522608876228 güldük:0,0353055968880653 yemek:0,0254689920693635 eğlen:0,0230065435171127 yazmak:0,0189682617783546 harika:0,0186064764857292 hafta:0,0179643891751766 yedik:0,0174420997500419 günü:0,017364501953125
48	form:0,0498358719050884 yavaş:0,0402372553944587 ver:0,0321901962161064 artık:0,0270676743239164 yaz:0,0233102142810821 uzun:0,0229198224842548 süre:0,019383069127798 iste:0,0167383272200822 kağıt:0,0151135809719562 doldururken:0,0142190316691994
49	çalışma:0,0774467960000038 grup:0,0626607164740562 takım:0,0484142005443573 yaptık:0,0449505113065242 arkadaş:0,0356395170092582 eğlen:0,0282924249768257 birlikte:0,023245109245181 ruh:0,0202397108078002 çalıştık:0,0179278757423162 yapmak:0,0167463440448045
50	geçir:0,0958107933402061 vakit:0,0846149176359176 gün:0,0636138021945953 geçirdik:0,0559419356286525 bir:0,0557514056563377 eğlenceli:0,0521278567612171 keyifli:0,0421575419604778 iyi:0,0343259908258914 zaman:0,031486764550209 ders:0,0287388917058706
51	ders:0,0332473888993263 keşke:0,0246692616492509 zaman:0,0207693707197904 gün:0,0182236265391111 hafta:0,0181091465055942 bit:0,0165097713470459 geçtik:0,0159640405327081 üzgün:0,0150635344907641 geçti:0,0133201349526643 bir:0,0130977947264909
52	hoş:0,0525705218315124 git:0,0424959175288677 ben:0,0256393775343894 özal:0,0253922995179891 iyi:0,019572626799345 yapmak:0,0195573102682828 ister:0,0161920171231031 iste:0,0160591695457696 zor:0,0159860905259847 eğlenceli:0,0159523915499448
53	heyecan:0,192071959376335 sevinç:0,12689208984375 eğlence:0,10002027451992 neşe:0,0698072090744972 merak:0,0688540488481521 verici:0,0587822161614894 üzüntü:0,0566402822732925 şaşkınlık:0,0484815165400505 telaş:0,0282175038009881 keyif:0,024436192587018

EK 4. Uzman Görüşü Sorular ve Örnek Cevaplar Tablosu.

Derlem	Örnek Soru	Örnek Cevap
S1	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne yaptığıyla ilgili görüşlerini içermektedir. Bu konu-kelime dağılımlarını inceleyerek anlamlı gördüğünüz konuları isimlendirebilir misiniz?	Konu-kelime dağılımları bana bir anlam ifade etmedi
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne yaptığıyla ilgili görüşlerini içermektedir. Konu-kelime dağılımları ile ilgili bir cümle kurup bu cümleyi bir kavramla ifade edebilir misiniz?	. “Anket sonuçlarını bilgisayarla değerlendirdik” olabilir. “Araştırma” da kavram olabilir. . Fakat konuların tamamını incelediğimde bir ders sürecinin basamakları gibi geldi bana 5E modeli kullanılarak kurslar verilmişti. 5E modelinin basamaklarına indirgeyebiliriz.
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne yaptığıyla ilgili görüşlerini içermektedir. Konu-kelime dağılımları ile ilgili bir cümle kurup bu cümleyi 5E modelinin basamaklarıyla ifade edebilir misiniz?	. Anket geliştirmek “Keşfetme” basamağında yer alabilir. . Fakat burada 5E modelinin basamaklarına oturmayan konular da var. Hiçbir basamağa uymayan konuları ne yapalım? “led yaktık” gibi birden fazla konu var “Bilgi işlemsel düşünme” gibi görünen. Bu basamağı da ekleyelim.
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne yaptığıyla ilgili görüşlerini içermektedir. Konu-kelime dağılımları ile ilgili bir cümle kurup bu cümleyi 5E modeli basamakları ve “Bilgi İşlemsel Düşünme” ile ifade edebilir misiniz? Hiçbir basamağa uymayan konuları kendiniz isimlendirebilirsiniz.	. Bu kez dağılımlar doğru isimlendirilmiş gibi görünüyor.
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne yaptığıyla ilgili görüşlerini içermektedir. Konu-kelime dağılımları 5E modeli basamakları ve “Bilgi İşlemsel Düşünme” ile etiketlenmiştir. Bu isimlendirmede size uygun gelmeyen bir yer bulunmakta mıdır?	
S2	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne öğrendiğiyle ilgili görüşlerini içermektedir. Bu konu-kelime dağılımlarını inceleyerek anlamlı gördüğünüz konuları isimlendirebilir misiniz?	Konu-kelime dağılımları bana bir anlam ifade etmedi
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne öğrendiğiyle ilgili görüşlerini içermektedir. Konu-kelime dağılımları ile ilgili bir cümle kurup bu cümleyi bir kavramla ifade edebilir misiniz?	Engelli, sorun mesela “engellilerin sorunları” olabilir.
S3	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne hissettiğiyle ilgili görüşlerini içermektedir. Bu konu-kelime dağılımlarını inceleyerek anlamlı gördüğünüz konuları isimlendirebilir misiniz?	. “Yanlış-kork”, “Yanlış yapacağımdan korktum” - “sunum heyecanlı”, “Sunum yaparken heyecanlanmış” . Hocam “Plutchnic Duygu Çemberi” var hiç duydun mu? Bence bu hisleri çemberdeki duygularla ifade edebiliriz.

Derlem	Örnek Soru	Örnek Cevap
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne hissettiğiyle ilgili görüşlerini içermektedir. Bu konu-kelime dağılımlarını Plutchnic duygu çemberindeki bir duyguya yerleştirebilir misiniz?	Hocam birçok duygu çembere yerleşiyor fakat 1 numaralı konu mesela “anne, insan, yardım...” “empati”. Bu konuyu “Kök Değerler” ile isimlendirebiliriz.
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne hissettiğiyle ilgili görüşlerini içermektedir. Bu konu-kelime dağılımlarını Plutchnic duygu çemberindeki veya “Kök Değerler” kapsamındaki bir değere yerleştirebilir misiniz?	
	Gördüğünüz konu-kelime dağılımında her konu öğrencinin derste ne hissettiğiyle ilgili görüşlerini içermektedir. Bu konu-kelime dağılımları Plutchnic duygu çemberi veya “Kök Değerler” kapsamında etiketlenmiştir. Bu isimlendirmede size uygun gelmeyen bir yer bulunmakta mıdır?	

EK 5. Değerlendirme Formu.

İSTANBUL MİLLİ EĞİTİM MÜDÜRLÜĞÜ ARGE BİRİMİ PROJE DEĞERLENDİRME FORMU	
PROJE ADI/ÇEŞİDİ	Harezmi Modeli Uygulama Verisinin Metin Madenciliği Teknikleriyle Değerlendirilmesi
GELDİĞİ BİRİM	Strateji Geliştirme Şubesi
GÖNDEREN KURUM/Tarih -Sayı	Ali ÇİMEN 04.02.2019 2338603
KURUMSAL İLETİŞİM /İRTİBAT KİŞİSİ	0505 430 57 23
PROJE GENEL İÇERİĞİ	Ekte belirtilen doktora tez çalışmasının kapsamında metin madenciliği yöntemleri kullanarak analiz etme talebi.
ANA FAALİYETLER	Ekte belirtilen doktora tez çalışmasının kapsamında metin madenciliği yöntemleri kullanarak analiz etme talebi.
KURUMUN PAYDAŞ OLARAK ROLÜ	İzin ve duyuru
GENEL DEĞERLENDİRME	Yapmış olduğunuz başvuru komisyonumuzca incelenmiş olup, Harezmi projesi Bilgi İşlem Teknolojileri Şubesi tarafından yürütüldüğünden gerekli inceleme ve değerlendirmenin Bilgi İşlem Teknolojileri Şubesi tarafından yapılması komisyonumuzca uygundur.
KOMİSYON ÜYELERİ	Komisyon Teslim Etme Tarihi: 05.02.2019

İLGİLİ BİRİM DEĞERLENDİRMESİ	
SONUÇ	OLUMLU
SORUMLU KİŞİ /BÖLÜM	
İZLEME DEĞERLENDİRME/ FAALİYET TARİHLERİ	
SONUÇ BİLDİRİMİ	EMAIL / RESMİ YAZI
ONAY	
SONUÇ RAPORU	
ULAŞILAN ÖĞRENCİ SAYISI	
ULAŞILAN ÖĞRETMEN SAYISI	
ULAŞILAN OKUL SAYISI	
SONUÇ:	

EK 6. Araştırma İzni.



T.C.
İSTANBUL VALİLİĞİ
İl Millî Eğitim Müdürlüğü

Sayı : 82973634-604.01.01-E.3309374
Konu : Harezmi Projesi (Ali ÇİMEN)

14.02.2019

MÜDÜRLÜK MAKAMINA

- İlgi : a) Valilik Makamının 29/06/2018 tarih ve 12520577 sayılı oluru.
b) Ali ÇİMEN'in 04.02.2019 tarihli ve 2338603 sayılı yazısı.
c) Müdürlüğümüz Strateji Geliştirme Şube Müdürlüğünün 08/02/2019 tarihli ve 2802135 sayılı yazısı.

İlimiz Avcılar ilçesi FATİH Projesi Eğitmeni olarak görev yapan Ali ÇİMEN; Müdürlüğümüz Bilgi İşlem ve Eğitim Teknolojileri Şubesince yürütülen "**Harezmi Eğitim Modeli**" **Veri Giriş ve Raporlama Modülü** çalışmalarını yürütmek üzere **28 Haziran 2018-28 Eylül 2018 tarihleri arasında** ilgi (a) Valilik Oluru ile görevlendirilmiştir.

Ali ÇİMEN'in Müdürlüğümüz Strateji Geliştirme Şube Müdürlüğüne yapmış olduğu "Harezmi Projesi" ile ilgili doktora tez araştırması talebi, İnceleme Komisyonunca olumlu değerlendirildiği ve hazırlanan bilgi formu çerçevesinde gerekli iş ve işlemlerin Şubemizce takip edilmesi ilgi (c) yazı ile Birimimize gönderilmiştir.

Ali ÇİMEN'in doktora tez çalışması kapsamında yürüteceği Harezmi Eğitim Modeli Projesinin metin madenciliği yöntemleri kullanarak analiz edilmesi konusunda, kullanacağı verilerin hiçbir koşulda 3. kişi ve kurumlarca paylaşılmaması ve araştırma neticesinde ortaya çıkan tez, bildiri veya makalelerin bir örneğinin Müdürlüğümüz Bilgi İşlem ve Eğitim Teknolojileri Şube Müdürlüğü ve Strateji Geliştirme Şube Müdürlüğüne gönderilmesi şartıyla uygun görülmektedir.

Makamınızca da uygun görüldüğü takdirde olurlarınıza arz ederim.

Murat ALTINÖZ
İl Müdür Yardımcısı

Ek: İlgi (b) Yazı ve Ekleri

O L U R
<...>

Levent YAZICI
İl Millî Eğitim Müdürü

İstanbul İl MEM (Bilgi İşlem ve Eğitim Teknolojileri Şubesi)
İmran Öktem C.No1 Eski Adliye Dinası Sultanahmet/İSTANBUL
e-posta: bikoordinator34@meb.gov.tr

Bilgi İçin: N.ATAERİ(Şef)

Telefon: 0212 455 04 00/204

Bu evrak güvenli elektronik imza ile imzalanmıştır. <http://evraksorgu.meb.gov.tr> adresinden d710-20fe-3c08-8c10-2f43 kodu ile teyit edilebilir.

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Ali ÇİMEN
Doğum Yeri	
Doğum Tarihi	Tarih girmek için tıklayın veya dokununuz.
Uyruğu	☒ T.C. ☐ Diğer:
E-Posta Adresi	
Web Adresi	

Eğitim Bilgileri	
Lisans	
Üniversite	Çanakkale Onsekiz Mart Üniversitesi
Fakülte	Eğitim Fakültesi
Bölümü	Bilgisayar ve Öğretim Teknolojileri Eğitimi
Mezuniyet Yılı	2006

Yüksek Lisans	
Üniversite	Yıldız Teknik Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Bilgisayar ve Öğretim Teknolojileri Eğitimi
Programı	Bilgisayar ve Öğretim Teknolojileri Eğitimi

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik Anabilim Dalı
Programı	Enformatik Programı

Makale ve Bildiriler	
Çimen, A., Yılmaz, M. B., 2014, Which content is appropriate for instructional based social network? Opinions of K12 teachers in Turkey, <i>Procedia-Social and Behavioral Sciences</i> , 116, 2338-2343.	
Çimen, A., Yılmaz, M B., 2017, How Should We Blend? The Impact of Blending Social Networks on High School Students' Achievement and Social Networking Behaviors, <i>Eğitim ve Bilim-Education and Science</i> , 42.	
Çimen, A., and Dede, D., 2018, Fabsis: A Data Management System in Educational Administration., <i>eLearning & Software for Education</i> , 4.	
Çimen, A. and Gülseçen, S., 2022, Harezmi eğitim modeli verisindeki gizli desenlerin ortaya çıkarılması, <i>Turkish Studies - Education</i> , 17(6), 1395-1419.	