



T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANA BİLİM DALI

YÜKSEK
LİSANS
TEZİ

EĞİTİMDE METİN MADENCİLİĞİ:
TÜRKÇE METİNLERDE
SÖZLÜK TABANLI DUYGU ANALİZİ

LEYLA KARAKUŞ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

Antalya, 2021

T.C.
AKDENİZ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANA BİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

EĞİTİMDE METİN MADENCİLİĞİ: TÜRKÇE METİNLERDE
SÖZLÜK TABANLI DUYGU ANALİZİ

YÜKSEK LİSANS TEZİ

Leyla Karakuş

Danışman

Doç. Dr. Bilal Barış Alkan

Antalya, 2021

DOĞRULUK BEYANI

Yüksek lisans tezi olarak sunduğum bu çalışmayı, bilimsel ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalardan gösterilenlerden oluştuğunu ve bu eserleri her kullanışında alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

25/06/2021

Leyla Karakuş

TEŞEKKÜR

Türkçe metinler üzerinde bir metin madenciliği yöntemi olan sözlük tabanlı duygu analizinin uygulanması ve bu uygulamanın eğitimde ne şekilde uygulanabilir olduğunun ortaya konulması amacını taşıyan bu tez çalışmamda; öncelikle engin bilgi ve görüşlerini, sabırlı ve anlayışlı bir tutumla paylaşan, desteğini esirgemeyen çok değerli tez danışmanım Doç. Dr. Bilal Barış ALKAN' a sonsuz saygı ve teşekkürlerimi sunarım.

Değerli fikir ve görüşleri ile eğitim hayatıma yön veren, desteğini ve ilgisini esirgemeyen değerli hocam Doç. Dr. Bekir DİREKÇİ' ye ise ayrıca teşekkürü borç bilirim.

Tüm tez yazım sürecinde destekleri ile yanımda olan sevgili arkadaşlarım Ainur JINALYS, Aslı UYSAL ve Selinay PULAT' a çokça teşekkür ederim.

Başta, tüm eğitim hayatım boyunca sevgi ve emeğiyle başarılarımın kaynağı olan canım annem Fadime KARAKUŞ olmak üzere değerli aileme sonsuz teşekkür ederim.

ÖZET

EĞİTİMDE METİN MADENCİLİĞİ: TÜRKÇE METİNLERDE

SÖZLÜK TABANLI DUYGU ANALİZİ

KARAKUŞ, Leyla

Yüksek Lisans, Eğitim Bilimleri Ana Bilim Dalı,

Tez Danışmanı: Doç. Dr. Bilal Barış Alkan

Haziran 2021, 153 sayfa

Serüvenine metinlerin taşıyıcılığında başlayan bilgi, teknoloji alanındaki gelişmelerin de etkisiyle muazzam boyutlara ulaşmıştır. Bu nedenle günümüzde bilgiyi işleme ve değerlendirme yeteneği, bilgiye ulaşma yeteneğinden çok daha önemli hale gelmiştir. Bilginin çeşitli kaynaklarından biri olması ve eğitim aracı olarak sıklıkla tercih edilmesi özelliği ile metinler; belirli amaçlar doğrultusunda keşfedilmesi, incelenmesi ve analiz edilmesi gereken temel materyallerdir. Bu amaçların gerçekleştirilmesi noktasında ise devreye gelişmekte olan Eğitimde Metin Madenciliği alanı girmektedir.

Yapılan tez çalışmasının amacı da bir metin madenciliği yöntemi olan sözlük tabanlı duygu analizini, Türkçe metinler üzerinde uygulamak ve içerisinde bulunan duygu kelimelerini baz alarak baskın olan duygu kutuplarını belirlemektir. Betimleyici nitel bir yaklaşımla ele alınan araştırmanın materyalini ise Ömer Seyfettin'in hikâye türündeki 165 eseri oluşturmaktadır. Verilerin analizi, bir istatistiksel hesaplama ve grafik yazılımı olan R programında, sözlük tabanlı duygu analizine ilişkin kodlamaların yapılmasıyla gerçekleştirilmiştir. İki aşamalı şekilde gerçekleştirilen çalışmada, öncelikle ilgili analiz daha dengeli ve tutarlı bir karşılaştırma yapılabilmesi için belirli kriterlere göre seçilen 51 hikâye metni üzerinde uygulanmıştır. Devamında, analiz ilk aşamada dahil edilmeyen hikâye metinleri de eklenerek toplamda 165 hikâye üzerinde yeniden uygulanmıştır. Analizin temel bileşenlerini oluşturan duygu sözlüklerinin boyut ve kelime çeşitliliği gibi özelliklerinin, analiz sonucuna etkisinin ortaya konulması amacıyla iki farklı duygu sözlüğü oluşturulmuş ve her iki çalışma grubunda ayrı ayrı kullanılmıştır.

Yapılan sözlük tabanlı duygu analizi uygulaması neticesinde; Ömer Seyfettin'e ait metinlerdeki polarite temelli genel duygu durumunun pozitif, duygu-his kategorisi temelli genel duygu durumunun ise coşkulu olduğu sonucuna ulaşılmıştır. Bu doğrultuda, duyguların temel

bir eğitim materyali olan metinlerdeki yoğunluğunun, yapılan eğitim ve öğretimin kalıcılığını, etkinliğini ve başarısını artırdığı hususu göz önüne alındığında, Ömer Seyfettin'in hikâye türündeki eserlerinin önemli bir eğitim ve aktarım kaynağı olduğu yargısına varılmıştır.

Anahtar Sözcükler: Metin madenciliği, Duygu analizi, Duygu sözlüğü, Metin analizi, R



ABSTRACT

TEXT MINING IN EDUCATION: DICTIONARY-BASED IN TURKISH TEXTS SENTIMENT ANALYSIS

Karakuş, Leyla

Master of Arts, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Bilal Barış Alkan

June 2021, 153 pages

Information, whose adventure started with transferring texts, has reached enormous dimensions with the effect of developments in the field of technology. Therefore, nowadays, the ability to process and evaluate information has become much more important than the ability to access information. Being one of the various sources of information and frequently preferred as an educational tool, texts are fundamental materials that need to be discovered, examined and analyzed for specific purposes. The field of Text Mining in Education proves helpful in realising these objectives.

This study aims to apply dictionary-based sentiment analysis, which is a method of text mining, on Turkish texts and determine dominant sentiment poles based on emotion words on its content. The material of this descriptive qualitative study consists of 165 short stories of Ömer Seyfettin. Analysis of data is performed by a dictionary-based sentiment analysis program which is developed on R, which is a statistical computation and graphics software, with means of programming dictionary-based sentiment analysis. In this two-staged research, the analysis is firstly applied on the text of 51 short stories to ensure a more balanced and consistent comparison. Afterwards, the short stories that were not included are also added, and the analysis is reapplied on a total of 165 short stories. In order to determine the effects of the fundamental components of the analysis, such as the size and word diversity of sentiment dictionaries, on the result of analysis two sentiment dictionaries are found; and these dictionaries are separately used on two study groups.

As a result of the dictionary-based sentiment analysis application, it is concluded that the polarity-based general sentiment state is positive on the texts of Ömer Seyfettin; and emotion-category-based general sentiment state on the same is excited. Accordingly, a judgement is reached that Ömer Seyfettin's works in short story genre constitute an important source of

education and transfer, considering that they improve the permanency, efficacy, and success of education and teaching.

Keywords: Text mining, Sentiment analysis, Sentiment dictionary, Text analysis, R



İÇİNDEKİLER

TEŞEKKÜR.....	i
ÖZET	ii
ABSTRACT	iv
İÇİNDEKİLER.....	vi
TABLOLAR LİSTESİ	xiii
ŞEKİLLER LİSTESİ	xv
KISALTMALAR LİSTESİ	xvii

BÖLÜM I

GİRİŞ

1.1. Problem Durumu	1
1.2. Araştırmanın Amacı ve Problemleri.....	3
1.2.1. Araştırmanın Alt Problemleri	3
1.3. Araştırmanın Önemi	4
1.4. Araştırmanın Sayıltıları	5
1.5. Araştırmanın Sınırlılıkları	6
1.6. Tanımlar	6

BÖLÜM II

KURAMSAL ÇERÇEVE İLGİLİ ARAŞTIRMALAR

2.1. Eğitimde Veri Madenciliği	7
2.2. Eğitimde Metin Madenciliği.....	8
2.2.1. Metin Verisi.....	8
2.2.2. Metin Madenciliğinin Tanımı.....	8
2.2.3. Metin Madenciliği ile İlişkili Diğer Dallar ve Alt Disiplinler.....	9
2.2.3.1. Arama ve Bilgi Alma.....	9

2.2.3.2. Belge Kümeleme	10
2.2.3.3. Belge Sınıflama	10
2.2.3.4. Web Madenciliği	10
2.2.3.5. Bilgi Çıkarma	10
2.2.3.6. Doğal Dil İşleme.....	11
2.2.3.7. Kavram Çıkarma.....	11
2.2.4. Metin Madenciliği Uygulamaları	12
2.2.5. Metin Madenciliğinin Genel Adımları	13
2.2.5.1. Veri Setini Oluşturma.....	13
2.2.5.2. Metin Ön İşleme	13
2.2.5.3. Metin Madenciliği Yönteminin Uygulanması.....	14
2.2.5.4. Değerlendirme ve Yorumlama	14
2.2.6. Eğitimde Metin Madenciliği için Kullanılan Yöntem ve Teknikler.....	14
2.2.6.1. Metin Sınıflandırma ve Kümeleme	15
2.2.6.2. Bilgi Çıkarma	16
2.2.6.3. Metin Özetleme	16
2.2.6.4. Doğal Dil İşleme.....	17
2.2.7. Eğitimde Metin Madenciliği Kaynakları.....	17
2.2.7.1. Forum	17
2.2.7.2. Çevrimiçi Değerlendirme	18
2.2.7.3. Sohbet	19
2.2.7.4. Belge, Sosyal Ağ, Blog ve E-posta.....	19
2.2.8. Eğitim Hedefleri	20
2.2.8.1. Değerlendirme	20
2.2.8.2. Destekleme	21
2.2.8.3. Analiz	21
2.2.8.4. Soru-İçerik Oluşturma	22
2.2.8.5. Geri Bildirim	22
2.2.8.6. Tavsiye Sistemi	23
2.3. Duygu Analizi	23
2.3.1. Duygu Analizinin Tanımı ve Temel Kavramları.....	23
2.3.2. Duygu Analizinin Görevleri	25

2.3.2.1. Duygu Sınıflandırma	25
2.3.2.2. Öznellik Sınıflandırma	25
2.3.2.3. Görüş Özetleme	25
2.3.2.4. Görüş Alma	26
2.3.2.5. Alay ve İroni	26
2.3.2.6. Diğer Duygu Analizi Görevleri	26
2.3.3. Duygu Analizinin Düzeyleri	27
2.3.3.1. Belge Düzeyinde Duygu Analizi	27
2.3.3.2. Cümle Düzeyinde Duygu Analizi	27
2.3.3.3. Özellik Düzeyinde Duygu Analizi	27
2.3.4. Duygu Analizinde Kullanılan Yöntemler	28
2.3.4.1. Makine Öğrenmesi Tabanlı Yaklaşım	28
2.3.4.2. Sözlük Tabanlı Yaklaşım	28
2.3.5. Duygu Sözlüğü	29
2.3.5.1. Duygu Kuramları	29
2.3.5.2. Duygu Sözlüğü Türleri	30
2.3.5.2.1. Polarite Temelli Sözlükler	30
2.3.5.2.1.1. Sözlükler	31
2.3.5.2.1.1.1. Harvard General Inquirer	31
2.3.5.2.1.1.2. Opinion Lexicon	31
2.3.5.2.1.1.3. Çok Perspektifli Soru Cevaplama (MPQA)	32
2.3.5.2.1.1.4. WPARD	32
2.3.5.2.1.1.5. SentiWordNet 3.0	32
2.3.5.2.1.1.6. NRC	33
2.3.5.2.1.2. Veri Kaynakları	33
2.3.5.2.1.2.1. Amazon	33
2.3.5.2.1.2.2. Film Veri Kümeleri	34
2.3.5.2.1.2.3. Bloglar	35
2.3.5.2.2. Duygu Temelli sözlükler	35
2.3.5.2.2.1. Sözlükler	35
2.3.5.2.2.1.1. Wordnet Affect	36
2.3.5.2.2.1.2. LIWC	36

2.3.5.2.2.1.3. NRC	37
2.3.5.2.2.1.4. NRC Hashtag.....	37
2.3.5.2.2.1.5. Açık Dengeli Duygusal Tweetler (CBET)	37
2.3.5.2.2.2. Veri Kaynakları	37
2.3.5.2.2.2.1. ISEAR.....	38
2.3.5.2.2.2.2. Peri Masalları.....	38
2.3.5.2.2.2.3. SemEval 2007	39
2.3.5.2.2.2.4. TEC.....	39
2.3.5.2.2.2.5. CBET	39
2.3.6. Duygu Analizi ile İlgili Çalışmalar	40
2.3.6.1. Eğitimde Duygu Analizi Uygulamalarının Hedefleri ve Yurt Dışında-Eğitim Alanında Yapılan Duygu Analizi Çalışmaları.....	40
2.3.6.1.1. Duygu Analizi Yöntemlerinin veya Sistemlerinin Uygulanması	40
2.3.6.1.1.1. Sistem Geliştirme	40
2.3.6.1.1.2. Öğrenme Müdahalesi.....	40
2.3.6.1.1.3. Görsel Geri Bildirim.....	41
2.3.6.1.2. Öğrenme Davranışını ve Performansını Anlama	41
2.3.6.1.2.1. Davranış ve Duygu Etkileşimi.....	41
2.3.6.1.2.2. Duygu ve Akademik Performans İlişkisi	42
2.3.6.1.2.3. Duyguların Zamansal Özelliklerinin Önemi	42
2.3.6.1.3. Öğretme ve Öğrenme Sürecini İyileştirme.....	42
2.3.6.1.3.1. Kurs Memnuniyeti.....	42
2.3.6.1.3.2. Platformdaki Deneyim.....	43
2.3.6.1.3.3. Kuruma Karşı Tutumu Tespit Etme	43
2.3.6.1.3.4. Öğretmen Performansının Değerlendirilmesi.....	43
2.3.6.2. Yurtdışında Hikâye ve Roman Türündeki Metinler Üzerinde Gerçekleştirilen Duygu Analizi Çalışmaları	44
2.3.6.3. Türkçe Metinlerde Duygu Analizi Çalışmaları	46
2.4. R Yazılımı	49
2.4.1. Avantajları	49

BÖLÜM III

YÖNTEM

3.1. Araştırmanın Modeli	51
3.2. Araştırmanın Materyali.....	51
3.3. Veri Analizi Aracı	52
3.4. Veri Analizi Süreci	52
3.4.1. Metin Verilerinin Elde Edilmesi	52
3.4.2. Metin Verilerinin Analize Hazırlanması	52
3.4.3. Türk Diline Özgü Morfolojik Düzenlemeler.....	54
3.4.4. Duygu Analizi Süreci	55
3.4.4.1. Analize İlişkin R Paketlerinin Kurulması.....	55
3.4.4.2. Çalışma Metninin R'ye Aktarılması.....	56
3.4.4.3. Çalışma Metni ile Korpus Oluşturulması	56
3.4.4.4. Metin Ön İşleme Adımlarının Uygulanması	58
3.4.4.5. Metin İstatistiklerinin Elde Edilmesi.....	60
3.4.4.6. Tip-Token Oranına İlişkin Regresyon Analizi	60
3.4.4.7. Duygu Sözlüklerinin Oluşturulması ve R' ye Aktarılması.....	61
3.4.4.8. Terim-Duygu Matrisinin Oluşturulması.....	63
3.4.4.9. Bölümlerin Daha Küçük Parçalara Ayrılması	63
3.4.4.10. Duygu Sözlüklerine Göre Metinlerin Duygu Skorlarının Hesaplanması.....	64
3.4.4.11. Duygu Analizi Sonuçlarının Görselleştirilmesi.....	64
3.4.4.11.1. Duygu Analizi Sonuç Grafiğinin Çizilmesi	64
3.4.4.11.2. Duygu Analizi Sonucuna İlişkin Kelime Bulutunun Oluşturulması	64
3.4.4.12. Ömer Seyfettin'e Ait 51 Hikâye ile Gerçekleştirilen Sözlük Tabanlı Duygu Analizinin, Yazarın Daha Önce Analize Dahil Edilmemiş Diğer Hikâyelerinin De Eklenerek Tekrar Uygulanması	65

BÖLÜM IV

BULGULAR

4.1. Metinlere İlişkin İstatistik Bulgular.....	66
4.2. Birinci Duygu Sözlüğü ile Yapılan Duygu Analizine İlişkin İstatistik Bulgular	69
4.2.1. Analiz Sonucuna İlişkin Frekans Poligonu	69
4.2.2. Analiz Sonucuna İlişkin Kelime Bulutu.....	71
4.2.3. Yıllara Göre Metinlerde En Sık Kullanılan Duygu Kelimelerinin Belirlenmesi	72
4.2.4. En Sık Kullanılan Duygu Kelimelerinin Doküman İçerisindeki Yerlerinin Belirlenmesi.....	72
4.3. İkinci Duygu Sözlüğü ile Yapılan Duygu Analizine İlişkin İstatistik Bulgular.....	77
4.3.1. Analiz Sonucuna İlişkin Frekans Poligonu	77
4.3.2. Analiz Sonucuna İlişkin Kelime Bulutu.....	79
4.3.3. Yıllara Göre Metinlerde En Çok Kullanılan Duygu Kelimelerinin Belirlenmesi	79
4.3.4. En Sık Kullanılan Duygu Kelimelerinin Metin İçerisindeki Yerlerinin Belirlenmesi.....	80
4.4. Birinci ve İkinci Sözlüğü ile Yapılan Duygu Analizi Sonuçlarının Karşılaştırılması.....	85
4.5. Ömer Seyfettin’e Ait 165 Hikâye Üzerinde Yeniden Uygulanan Sözlük Tabanlı Duygu Analizine İlişkin Kelime Bulutunun Yorumlanması	87

BÖLÜM IV

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma	88
5.2. Öneriler.....	93
KAYNAKÇA.....	96
EKLER	125
Ek-1: Duygu Analizine İlişkin R Kodları.....	125
Ek-2: Etkisiz İfadeler Listesi	128
ÖZGEÇMİŞ	130
BİLDİRİM.....	131



TABLÖLAR LİSTESİ

Tablo 2.1. Duygu Kavramı ile İlgili Teoriler	30
Tablo 2.2. Polarite Temelli Sözlükler.....	31
Tablo 2.3. Polarite Temelli Veri Kaynakları	33
Tablo 2.4. Duygu Temelli Sözlükler	36
Tablo 2.5. Duygu Temelli Veri Kaynakları.....	38
Tablo 2.6. Türkçe Metinlerde Duygu Analizi ile İlgili Çalışmalar	46
Tablo 3.1. Hikâyelerin Künyeleri ve Hikâyelere Ait Düzenlenmemiş Metin Parçaları.....	52
Tablo 3.2. Analizde Kullanılan R Kütüphaneleri	56
Tablo 3.3. Corpus_frame() Komutu Kullanılarak Oluşturulan Korpusa Ait Çıktı.....	56
Tablo 3.4. Metin Ön İşleme Adımlarının Ardından Korpus İçerisinde Yer alan Metinlerin Analiz Öncesindeki Son Halleri	59
Tablo 3.5. Tip-Token Oranının Test Edilmesine İlişkin Regresyon Sonuçları	61
Tablo 3.6. Birinci Duygu Sözlüğünün İçeriğine İlişkin Örnek Kelimeler ve Duygu Kutupları	62
Tablo 3.7. İkinci Duygu Sözlüğünün İçeriğine İlişkin Örnek Kelime ve Duygu Kutupları	62
Tablo 3.8. Birinci Duygu Sözlüğü İçerisinde Yer Alan Terimler ve Duygu Kutupları	63
Tablo 3.9. İkinci Duygu Sözlüğü İçerisinde Yer Alan Terimler ve Duygu Kutupları	63
Tablo 4.1. Hikâyelerdeki Token ve Tip Sayıları	66
Tablo 4.2. Hikâyelerdeki Token ve Tip Sayılarına İlişkin İstatistiksel Bulgular	68
Tablo 4.3. Yazıldıkları Yıllara Göre Gruplanan Metinlerde Sıkça Kullanılan Duygu Kelimeleri	72
Tablo 4.4. Yazıldıkları Yıllara Göre Gruplanan Metinlerde Sıkça Kullanılan Duygu Kelimeleri	80

Tablo 4.5. Birinci ve İkinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonuçları	85
--	----



ŞEKİLLER LİSTESİ

Şekil 2.1 Metin Madenciliğini Alt Disiplinleri ve Bağlantılı Uygulamalar	12
Şekil 4.1. Birinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonucuna İlişkin Frekans Poligonu	70
Şekil 4.2. Birinci Duygu Sözlüğü ile Yapılan Analiz Sonucuna İlişkin Kelime Bulutu.....	71
Şekil 4.3. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Gör-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	73
Şekil 4.4. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Başla-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	74
Şekil 4.5. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Bul-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	75
Şekil 4.6. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Büyük” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	76
Şekil 4.7. İkinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonucuna İlişkin Frekans Poligonu	78
Şekil 4.8. İkinci Duygu Sözlüğü Göre Yapılan Duygu Analizi Sonucuna İlişkin Kelime Bulutu	79
Şekil 4.9. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Gül-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	81
Şekil 4.10. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Sev-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi	82

Şekil 4.11. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Aşk” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi.....	83
Şekil 4.12. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Korku” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi.....	84
Şekil 4.13. Ömer Seyfettin’e Ait 165 Hikâye Üzerinde Gerçekleştirilen Duygu Analizine İlişkin Kelime Bulutu	87



KISALTMALAR LİSTESİ

DT: Duygu Tespiti
DS: Duygu Seçimi
DA: Duygu Analizi
YK: Yöntem Karşılaştırma
MÖ: Makine Öğrenmesi
ST: Sözlük Temelli
DDİ: Doğal Dil İşleme
P: Pozitif
N: Negatif



BÖLÜM I

GİRİŞ

1.1. Problem Durumu

Dil içerisinde kendi kurallarını barındıran, düşünmeyi gerçekleştiren ve düşünceyi ileten en temel iletişim aracıdır. Yazılı veya sözlü olarak ortaya konabilen dilin yapı taşı ise belirli seslerden ya da harflerden oluşan sözcüklerdir. Sözcüklerin anlamlı birliktelikleri bilginin, tarihin, uygarlığın devamlılığını ve kalıcılığını sağlayan metinleri oluşturmaktadır. Bu doğrultuda, toplumların ve toplumları oluşturan bireylerin bilgi birikimi artmakta; her geçen gün fikir ve duyguları değişmektedir (Hacızade, 2012).

Anlaşılmasının ve aktarılmasının zorluğu nedeniyle duygular, insanların sahip olduğu en karmaşık yapılardır. Felsefe, psikoloji, sosyoloji, tıp, eğitim, dil bilimi gibi çeşitli alanlarda farklı biçimlerde ele alınan duygu kavramının tanımı konusunda kesin bir yargı yoktur (Varışoğlu, 2017). Her disiplin, kavramı kendi penceresinden görmüş ve açıklamıştır. Türk Dil Kurumunun Güncel Türkçe Sözlüğünde “duyularla algılama, his; belirli nesne, olay veya bireylerin insanın iç dünyasında uyandırdığı izlenim, önsezi; nesneleri veya olayları ahlaki ve estetik yönden değerlendirme yeteneği; kendine özgü bir ruhsal hareket ve hareketlilik” olarak açıklanmaktadır. Buna göre duygular insanın fizyolojik ve psikolojik durumunun bir yansımasıdır.

Duygu ve düşünceyi anlama ve anlatma aracı olarak tanımlanan metinler, eğitimde öğrenmeyi ve öğretmeyi gerçekleştirmek amacıyla kullanılan en temel materyallerdir (Aktaş, 2009). Sadece eğitim alanındaki metin varlığı dikkate alındığında dahi tarif edilemez büyüklükte bir veri yığınıyla karşılaşilmektedir. Bu doğrultuda büyük miktardaki metin verisinden birtakım istatistiksel yöntemler yardımıyla anlamlı ve nitelikli bilginin elde edilmesini amaçlayan Metin Madenciliği alanı devreye girmektedir (Feldman ve Sanger, 2007).

Metin madenciliği, veri madenciliğinin bir alt dalı olarak gelişmekte olan disiplinlerarası bir alandır. Metin madenciliği yöntemlerine, eğitimde, eğitim ortamlarından gelen bilgilerle öğrenme hakkında fikir edinmek ve eğitim çıktılarını iyileştirmek amacıyla sıklıkla başvurulmaktadır (Baker ve Yacef, 2009). Özellikle, duygu analizi, bir olay, doküman veya etkileşim için ortaya konan duygusal kutbu ya da bazı konularla ilgili yazarın, okuyucunun

vb., tutumunu belirlemek amacıyla eğitimde yaygın biçimde kullanılan bir tekniktir (Medhat ve diğerleri, 2014).

Duygu analizinde temel amaç sözcükler aracılığıyla bir metindeki genel duyguyu ortaya çıkarmaktır. Sözcüklerin barındırdığı duygular kişiden kişiye değişebilecek bir öznellikte olabilmelerinin yanı sıra, aynı dil birliğine sahip kişilerde ortak duygular uyandırabilecek güce sahiptir. Örneğin ölüm, hastalık, sürgün, verem, sefalet gibi sözcükler olumsuz; aydınlık, başarı, diriliş, bayrak, kavuşmak, neşe gibi sözcükler de olumlu bir duygu yönüne sahiptir. Eğitimde yaygın olarak kullanılan materyallerden biri olan metinler (roman, hikâye) üzerinde sözcük tabanlı duygu analizi gerçekleştirilerek metnin genel duygu kutbunun belirlendiği ve duygu kutbu belirlenmiş diğer metinlerle birlikte karşılaştırmanın yapıldığı çeşitli uygulamalar bulunmaktadır (Liu, 2012).

Türkiye’de metin madenciliği alanında yapılan çalışmaların sayısında son yıllarda genel bir artış görülmektedir. Ancak yurtdışında yapılan çalışmaların sayısı ve uygulama alanlarının genişliği dikkate alındığında, bu artışın kayda değer nitelikte olmadığı belirlenmiştir. Türkiye’de metin madenciliği alanında duygu analizi ile ilgili yapılan az sayıdaki çalışma incelendiğinde ise genellikle mühendislik temelli araştırmacılar tarafından yapılan çalışmalar olduğu görülmektedir. Çalışmalarda çoğunlukla bir durum üzerinde çok sayıda birey tarafından belirtilen görüşlerin (film yorumu vb.) sınıflandırılması ya da aynı veri seti üzerinde birtakım sınıflandırıcıların performanslarının değerlendirilmesi amaçlanmıştır (Vural ve diğerleri, 2012; Türkmenoğlu, 2015; Özdeş, 2017; Albayrak, 2018; Toçoğlu, 2018; Sevindi, 2019; Abdoulaye, 2019; Karamanlı, 2019). Öte yandan ilgili alan yazın çalışmaları, uygulama hedefleri, uygulama ortamları ve veri kaynakları bakımından detaylı olarak incelendiğinde ise çalışmaların eğitim alanıyla herhangi bir ilişkisinin bulunmadığı açıkça görülmektedir. Bu genellemenin dışında kalan ve Boynukalın (2012) tarafından ortaya konan tez çalışmasında ise eğitimle ilişkilendirilebilme ve eğitimde uygulanabilme olanağı öngörülen ikinci bir veri seti olarak Türk masalları kullanılmıştır; ancak bu veri seti Boynukalın’ın araştırmasında eğitimde metin analizinin gerçekleştirilmesi gibi temel bir hedefe ulaşmak amacıyla ele alınmamıştır. Bir yan uygulama örneği niteliği taşımaktadır.

Alan yazının incelenmesiyle varılan bu sonuç üzerinde, eğitim bilimcilerin, duygu analizini barındırdığı tekniklerin yoğun yazılım bilgisi gerektirdiği algısı nedeniyle kullanmamalarının etkisi oldukça büyüktür (Akın ve Şimşek, 2018). Dolayısıyla ilgili analizin,

analize dair tekniklerin ve uygulama alanlarının eğitim bilminde tanıtılması alan yazına katkı sağlanması açısından önem arz etmektedir.

Bu araştırmada R yazılımı aracılığıyla Ömer Seyfettin'in hikâyeleri üzerinde bir metin madenciliği yöntemi olan sözlük tabanlı duygu analizi uygulanarak baskın olan duygu kutbunun belirlenmesi amaçlanmaktadır. Alan yazındaki çalışmaların aksine, tek bir yazarın çok sayıdaki metni üzerinde duygu analizinin yapılması ve metinlerin duygu kutupları bakımından karşılaştırılması bir diğer amacı oluşmaktadır. Bu bağlamda araştırmanın problem cümlesi; içerisinde kullanılan kelimelerin duygu kutupları aracılığıyla Ömer Seyfettin'in hikâyelerindeki duygu durumunu incelemektir.

1.2. Araştırmanın Amacı ve Problemleri

Bu çalışmanın genel amacı R yazılımı ile Türkçe bir metin üzerinde sözlük tabanlı duygu analizinin nasıl gerçekleştirildiğini ve ilgili analizin eğitim bilminde ne şekilde uygulanabilir olduğunu ortaya koymaktır.

1.2.1. Araştırmanın Alt Problemleri

Araştırmanın amacı doğrultusunda cevap aranacak olan araştırma problemleri şunlardır:

- i. R'de sözlük tabanlı duygu analizi için Ömer Seyfettin'in hikâyelerinde metin ön işleme adımları nasıl gerçekleştirilir?
- ii. R'de sözlük tabanlı duygu analizi için Ömer Seyfettin'in hikâyelerinde yer alan kelimelere yönelik duygu sözlüğü nasıl oluşturulur?
- iii. R'de sözlük tabanlı duygu analizi için Ömer Seyfettin'in hikâyelerinde yer alan kelimelerin duygu sözlüğüne göre duygu kutupları nasıl belirlenir?
- iv. R'de Ömer Seyfettin'in hikâyelerinin sözlük tabanlı duygu analizi ile belirlenen duygu kutupları oranları nasıl görselleştirilir?
- v. R'de Ömer Seyfettin'in, sözlük tabanlı duygu analizi ile incelenen hikâyelerinde en çok kullanılan duygu kelimeleri hangileridir?
- vi. R'de Ömer Seyfettin'in, sözlük tabanlı duygu analizi ile incelenen hikâyelerindeki baskın duygu durumu nedir?

1.3. Araştırmanın Önemi

Tezin problem durumunun ortaya konduğu kısımda da kısmen açıklandığı üzere alan yazında, eğitimde metin madenciliği ile ilgili çalışmaların azlığı dikkat çekmektedir. Çalışmalar, bu tez çalışmasının analiz yöntemini oluşturan ve bir metin madenciliği uygulaması olan duygu analizinin eğitim alanındaki kullanım durumu ele alınarak incelendiğinde ise, ilgili analizin kullanımında bu azlığın daha şiddetli olduğu görülmektedir. Ayrıca çalışmaların, genel olarak ekonomi, ticaret, sosyoloji vb. alanlarda, mühendislik temelli araştırmacılar tarafından uygulandığı açıkça görülmektedir. Akın ve Şimşek (2018)'e göre eğitimde kullanımı büyük fayda sağlayacağı öngörülen ve yurtdışında çokça örneği bulunmakta olan duygu analizinin, eğitim alanındaki araştırmacılar tarafından henüz uygulanmamış olmasının temel nedeni; ilgili tekniğin yoğun yazılım bilgisi gerektirdiği algısıdır. Bu doğrultuda tez, duygu analizinin eğitim alanında nasıl kullanılabileceğine ve özellikle Türkçe metinler üzerinde nasıl uygulanabileceğine ilişkin örnek bir kaynak niteliği taşımaktadır.

Tezin önemi, çalışmanın materyalini oluşturan, Ömer Seyfettin'e ait 165 hikâye açısından ele alındığında ise; öncelikle Bozkaplan, Kubilay ve Karademirlidağ (2016) tarafından hazırlanan, "Mukaddime ve Piyale Önsözü'nden Yola Çıkarak Türkçede Duygu Sözlüğü Oluşturulması Üzerine Metin Tabanlı Bir Deneme" adlı sempozyum bildirisinde hem duygu analizi yönteminin temel bir bileşeni olan duygu sözlüğünün ilk kez telaffuz edilmiş olduğu hem de ilgili analizin eğitim ve edebiyat alanında uygulanması gerektiğinin ilk kez dile getirilmiş olduğu vurgulanmalıdır. Ancak devamında herhangi bir çalışmaya rastlanmamış olduğu görülmektedir. Dolayısıyla yapılan bu tez çalışması, daha önce problem durumu olarak zihinlerde oluşmuş; ancak, çözümlenmemiş temel bir soruna/duruma açıklık getirmektedir.

Eğitimde belirli bir yazarın eserlerinin belirli kavramlar açısından incelendiği, içerik analizi niteliği taşıyan ve kullanılan yöntem bakımından basit bir mantığa sahip fakat uğraştırıcı bir süreç barındıran çok çeşitli çalışmalar bulunmaktadır. Örneğin Fırat ve Coşkun (2018)'un çalışmasında, Mustafa Ruhi Şirin'e ait çocuk şiirlerinde, duygu değeri belirli olan kavramların sıklığı incelenmiş ve Türkçe derslerinde içerdiği kavramlar açısından bu niteliğe sahip metinlerin kullanılmasının; çocukların sözcük dağarcığı, kavramlararası ilişkilendirme ve soyut düşünme becerilerinin gelişmesine yardımcı olacağı sonucuna ulaşılmıştır. Benzer şekilde eğitimde kullanılan içeriklerin incelenmesi bakımından, Geçgel ve Sarıçan (2011) tarafından hazırlanan çalışmada da Ömer Seyfettin'in hikâyeleri, çocuk eğitimine uygunluğu açısından,

içerisinde yer alan kavramlar baz alınarak incelenmiştir. Bir diğer yandan Turna (2020), Ömer Seyfettin'in hikâye türündeki eserlerini psikobiyografik açıdan inceleyen ve bu doğrultuda, yazarın biyografik, psikolojik ve bakış açısına yönelik özelliklerinin eserlerine yansıdığı sonucuna ulaşan bir çalışma ortaya koymuştur.

İlgili incelemeler ışığında bir yazarın tüm eserlerinin hem çeşitli kavramlar açısından boylamsal olarak incelenebildiği hem de psikanaliz yöntemi kullanılarak değerlendirilebildiği görülmektedir. Dolayısıyla bu tez çalışması, yazarlara ve ilgili eserlerine farklı açıdan bakılan bu çalışmaları birleştirmekle kalmamakta; aynı zamanda bu çalışmalarda daha hızlı, etkili ve nesnel bir analiz yöntemi olan duygu analizinin kullanılabileceğini ortaya koymaktadır.

Araştırmanın önemi, kullanılan analiz aracı açısından ele alındığında; bu tez çalışması eğitim alanında R yazılımının kolaylıkla uygulanabilir olduğunu ve uygulanması neticesinde sağlayacağı avantajları ortaya koyan örnek bir kaynak niteliği taşımaktadır. Ayrıca doküman üzerinde çalışılan çeşitli araştırmalarda, sıklıkla kullanılan Nvivo gibi analiz araçlarına yeni ve güçlü bir alternatif oluşturmaktadır.

Araştırmanın önemi kullanılan araştırma modeli açısından ele alındığında ise öncelikle Şenyurt ve Özkan (2017)'in çalışmalarında ortaya koydukları sonuçlara göre Eğitimde Ölçme ve Değerlendirme alanında yapılan yüksek lisans tezlerinde kullanılan araştırma modeli yaklaşımlarının %25'ini betimsel tarama modeli oluşturmaktadır. Ayrıca doküman incelemesine bu tezlerin sadece % 0.60'ında rastlandığı bilgisine yer verilmektedir. Dolayısıyla kullanılan yöntem ve analiz aracı birlikte ele alındığında; bu tez çalışmasının, nitel araştırma yöntemlerine nicel bir bakış açısı kazandıracağı öne sürülmektedir.

1.4. Araştırmanın Sayıltıları

Dil uzmanları eşliğinde oluşturulan duygu sözlüğündeki kelimelerin duygu kutuplarının, alan yazında kabul gören duygu sınıflamalarını içeren benzer çalışmalar baz alınarak doğru bir şekilde belirlendiği düşünülmektedir.

1.5. Araştırmanın Sınırlılıkları

Dil bilgisine özgü zorluklar nedeniyle Türkçe metinlerde duygu analizinin gerçekleştirilebilmesi için ilgili metin üzerinde birtakım morfolojik düzenlemelerin yapılması gerekmektedir. Bu işlem için Doğal Dil İşleme alanı ile ilişkili olarak ortaya konan çeşitli morfolojik düzenleme araçları bulunmaktadır. Ancak araştırmanın materyalini oluşturan Ömer Seyfettin'e ait 165 hikâyenin yoğun şekilde, Osmanlı Türkçesine özgü tamlama ve kelime grubu içermesi nedeniyle alan yazındaki morfolojik çözümleme araçları yetersiz kalmaktadır. Dolayısıyla metin ön işleme adımlarından biri olması sebebiyle metinler üzerinde gerçekleştirilen morfolojik düzenleme, araştırmacının kendisi tarafından yapılmıştır.

1.6. Tanımlar

Metin madenciliği: Birtakım istatistiksel yöntemlerle yapılandırılmamış büyük miktardaki metinlerden, anlamlı ve kullanışlı bilginin çıkarılmasını amaçlayan disiplinlerarası bir alandır.

Duygu analizi: Duygu analizi, veri madenciliği ve hesaplamalı dilbilim alanındaki bir alt disiplindir. Duygu analizinin temel amacı metin verileri üzerinden insanların ruh hallerini, davranışlarını ve fikirlerini keşfetmektir.

Sözlük tabanlı duygu analizi: Kelime veya kelime gruplarının içerdikleri duyguya göre (pozitif/negatif/nötr) sınıflandırıldığı bir duygu sözlüğü aracılığıyla metinlerin duygu durumlarının belirlenmesidir. Bu doğrultuda metindeki pozitif, negatif ve nötr sözcük sayısı incelenir ve çoğunluğu oluşturan duygu, metnin duygu durumu olarak atanır.

Duygu kutbu: Sözcüğün okunduğunda ya da duyulduğunda dili konuşan bireyler üzerinde yoğun bir şekilde oluşturduğu ilk histir. Örneğin, çiçek pozitif, mezar negatif, sandalye ise nötr duygu kutbuna sahiptir.

Duygu sözlüğü: Kelime veya kelime gruplarının içerdikleri duygu kutbuna (pozitif, negatif, nötr) göre karşılıklarının belirlendiği sözlüktür.

Metin ön işleme: Analizde kullanılacak olan metnin analize hazır hale getirilmesi işlemidir.

R: Veri madenciliğinde sıklıkla kullanılan R, istatistiksel hesaplama ve görselleştirme için oluşturulmuş, açık kaynak kodlu bir yazılımdır.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1. Eğitimde Veri Madenciliği

Eğitimde veri madenciliği, eğitimsel veri madenciliği uygulaması olarak da bilinen, gelişmekte olan disiplinler arası bir araştırma alanıdır. Eğitim ortamlarından gelen benzersiz biçimdeki veriyi keşfetmek ve veride saklı kalmış yapıları ortaya koymak için yöntem geliştirmekle ilgilidir. Amaç; eğitim ortamlarını açıklamak, öğrenme hakkında fikir edinmek, eğitim çıktılarını iyileştirmek için öğrencilerin nasıl öğrendiklerini daha iyi anlamaktır (Romero ve Ventura, 2013).

Eğitimde internet kullanımının, e-öğrenme kaynaklarının ve eğitim yazılımlarının artması ile birlikte öğrenci bilgilerini depolayan veri tabanları kurulmuştur. Eğitimde bilgi sistemleri, farklı formatlarda ve ayrıntı düzeylerinde, çoklu kaynaklardan gelen büyük miktarda potansiyel veri depolamaktadır. Bu doğrultuda büyük miktarda dinamik eğitim verisi elde edilmiştir (Koedinger, Cunningham, Skogsholm ve Leber, 2008). Dolayısıyla eğitim araştırmacıları için öğrenmenin nasıl gerçekleştiğinin anlaşılması ve eğitim çıktılarının keşfedilmesi adına oldukça önemlidir (Mostow ve Beck, 2006).

Eğitimde veri madenciliği, içinde bulundukları büyük miktardaki veri nedeniyle analizi imkansız ya da çok zor olan eğitim verilerindeki modelleri belirlemek amacıyla hesaplama metotlarının uygulanması, araştırılması ve geliştirilmesi ile ilgilidir (Romero, Ventura, Pechenizky ve Baker, 2010).

Günümüzde geleneksel eğitim ya da web tabanlı eğitim veren çok çeşitli eğitim ortamları ve bilgi sistemleri bulunmaktadır. Bunların her biri mevcut verinin doğasına bağlı olarak farklı veri madenciliği teknikleri ile çözülebilen özel problemleri ve ön işleme süreci gerektiren farklı veri kaynaklarını sağlamaktadır (Romero ve Ventura, 2007). Geleneksel veri madenciliği tekniklerinin çoğunluğu eğitim alanında uzun süredir başarılı bir şekilde uygulanmakta olan sınıflama, kümeleme, ilişkilendirme gibi analiz tekniklerini içermektedir fakat bunlarla sınırlı değildir. EVM içerisinde çok sayıda popüler yöntem bulunmaktadır (Baker, 2010).

2.2. Eğitimde Metin Madenciliği

2.2.1. Metin Verisi

Tipik veri madenciliği uygulamaları, çoğunlukla, yapılandırılmış verilerin kullanımına olanak sağlamaktadır. Yapılandırılmış veri, belirli bir ön tasarıma dayalı olarak elde edilmiş, sınırları belirli olan, kategorik ya da sıralı-sayısal niteliğe sahip ve elektronik ortamda tablo biçiminde tanımlanabilen verilerdir. Dolayısıyla belirli amaçlar doğrultusunda organize edilmiş bu tip veride matematiksel işlemlerin ve çok çeşitli analizlerin yapılması yapılandırılmamış verilerle karşılaştırıldığında oldukça kolaydır (Weiss, Indurkya ve Zhang, 2015).

Yapılandırılmamış ya da yarı-yapılandırılmış veriler, dünyadaki verilerin %80'e yaklaşan büyük bir kısmını oluşturmaktadır. Yapılandırılmamış ya da diğer bir deyişle metin türündeki verilerin mevcut hacmi göz önüne alındığında, insanların metinlerden nitelikli bilgileri elde edebilmeleri ve belirli hedefler doğrultusunda kullanabilmeleri için gerekli yöntemlere ya da araçlara başvurmaları gerekmektedir (Nisbet, Miner, Elder, Hill, Fast ve Delen, 2012).

2.2.2. Metin Madenciliğinin Tanımı

Metin madenciliği, veri madenciliğinin içerisindeki diğer tekniklere nazaran madencilik metaforuyla en çok uyuşan tekniktir. Madencilik kavramının değerli metali bulunduğu kaynaktaki öğelerden ayırma sürecini ifade etmesi doğrultusunda; metin madenciliği, belge veya belge koleksiyonları içerisinde bulunan değerli anahtar kelimelerin, bir yığın diğer kelimeden ayrılması ve bunun için modellerin oluşturulması ya da tahminlerde bulunulması şeklinde tanımlanmaktadır (Kotu ve Deshpande, 2019).

Metin madenciliği ya da diğer adıyla metin analizi, yarı yapılandırılmış ve yapılandırılmamış metin verilerinin analiz edilmesi ve işlenmesi için bir dizi teknolojiyi içerisinde barındıran geniş kapsamlı bir terimdir. Bu teknolojilerinin ardında bulunan ortak tema, güçlü algoritmaların devasa ölçüde bulunan belge türündeki veri setlerine uygulanabilmesi için, metinlerin sayılara dönüştürülmesi ihtiyacıdır. Metni, yapılandırılmış sayısal bir biçime dönüştürmek ve analitik algoritmaları uygulamak; sözcüklerden başlayarak

büyük metin veri setlerine varana kadar metin işleme tekniklerinin nasıl uygulanacağını bilinmesini gerektirir (Miner ve diğerleri, 2012).

Metin madenciliğinde kullanılan veriler, kayıtlı ya da resmi veri tabanlarında açık bir şekilde bulunan verilerden değildir. Metin madenciliğin amacını oluşturan ve ulaşılması gereken nitelikli bilgiler ya da ilginç kalıplar, yapılandırılmamış metin türündeki veriler içerisinde gizlidir. Metin madenciliği, genel olarak, kullanıcının bir dizi analiz aracı yardımıyla belge koleksiyonları ile etkileşime girdiği yoğun bilgi keşfi süreci olarak tanımlanmaktadır. Metin madenciliğinde, veri madenciliğine benzer şekilde, veri kaynaklarının incelenmesi ve ilginç örüntülerin tanımlanması yoluyla yararlı bilgilerin elde edilmesi amaçlanmaktadır (Feldman ve Sanger, 2007).

Metin madenciliği kaynaklarını oluşturan metin türündeki verilerin, veri tabanlarında depolanan veriler ile karşılaştırıldığında, şekilsiz ve algoritmik olarak ele alınmalarının zor oldukları görülmektedir. Metin madenciliği, doğal dil metninden anlamlı bilgileri çıkarmak amacıyla sözcükleri ya da dile özgü kalıpları tespit eden sistemleri barındırmaktadır (Witten, 2004).

2.2.3. Metin Madenciliği ile İlişkili Diğer Dallar ve Alt Disiplinler

Metin madenciliği, birbirleriyle sıkı bir ilişki içerisinde olan ve kendilerine has özelliklere sahip yedi uygulama alanına sahiptir. Bu yedi uygulama alanı, metin madenciliğine önemli ölçüde katkıda bulunan Veri Madenciliği, Yapay Zeka ve Makine Öğrenmesi, Hesaplamalı Dilbilim, İstatistik, Veri Tabanları, Kütüphane ve Bilgi Bilimleri (Bilgi Yönetimi) adlarıyla bilinen diğer altı ana alanın kesişim noktaları üzerine kurulmuştur (Miner ve diğerleri, 2012).

2.2.3.1. Arama ve Bilgi Alma

Arama ve bilgi alma, anahtar kelime sorguları ile büyük metin veritabanları içerisindeki belgeleri aramayı, elde etmeyi ve indekslemeyi kapsamaktadır. Google, Yahoo! ve Bing gibi güçlü internet arama motorlarının kullanımındaki artış ile birlikte arama ve bilgi erişimi çoğu insan için alışlagelmiş bir durum haline gelmiştir. E-postadan kelime işlemeye kadar neredeyse her bilgisayar uygulaması bir arama işlevi içermektedir (Manning, Raghavan ve Schütze, 2008).

2.2.3.2. Belge Kümeleme

Belge kümeleme, belgeleri benzerliklerine göre kümeler halinde gruplamak için veri madenciliğine dayanan algoritmaları kullanmaktadır. Metin madenciliğinin son 20 yıldır etkinlik gösteren bir alan olmasının aksine kümeleme algoritmalarının geçmişi daha eskidir. Bu nedenle kümeleme algoritmaları, birçok ticari veri ve metin madenciliği yazılım paketinde yaygın olarak bulunmaktadır (Nisbet, Elder ve Miner, 2009).

2.2.3.3. Belge Sınıflama

Veri madenciliği temelli sınıflandırma yöntemlerini kullanarak terimleri, paragrafları veya belgeleri, etiketli örnekler üzerinde eğitilmiş modellere dayalı olarak gruplama ve kategorilere ayırma işlemidir. Belge sınıflama, bilinen etiketli belgelerden öğrenilen bir metin modelini kullanarak etiketsiz belgelere bilinen bir etiket kümesinin atanmasıdır. Metin madenciliğinde kullanılan en önemli tekniklerden biridir (Nisbet ve diğerleri, 2018).

2.2.3.4. Web Madenciliği

Web madenciliği, benzersiz yapısı ve muazzam veri hacmiyle kendi uygulama alanına sahip bir metin madenciliği alanıdır. Web belgeleri, tipik olarak, köprüler yardımıyla birbirine bağlı olan sayfalar şeklinde yapılandırılmış bir metin biçiminde sunulmaktadır. Standart bir metinden farklı olması, beraberinde birçok zorluğu ve fırsatı meydana getirmektedir. İnternet, Facebook, Twitter ve diğer sosyal medya kanallarının yükselişiyle birlikte popüler kültürümüzde daha da yerleşik hale gelen web madenciliği, hızla gelişmekte olan bir alandır (Chakrabarti, 2002).

2.2.3.5. Bilgi Çıkarma

Yapılandırılmamış metinden ilgili gerçeklerin, ilişkilerin belirlenmesi ve çıkarılması ya da diğer bir deyişle yapılandırılmamış ve yarı yapılandırılmış metinden yapılandırılmış veri oluşturma süreci olarak tanımlanmaktadır. Bilgi çıkarma, metin madenciliğindeki en olgun alanlardan biridir. Bir bilgi çıkarma sisteminin eğitilmesi veya ayarlanması büyük miktarda çaba gerektirmektedir (Nisbet ve diğerleri, 2018).

2.2.3.6. Doğal Dil İşleme

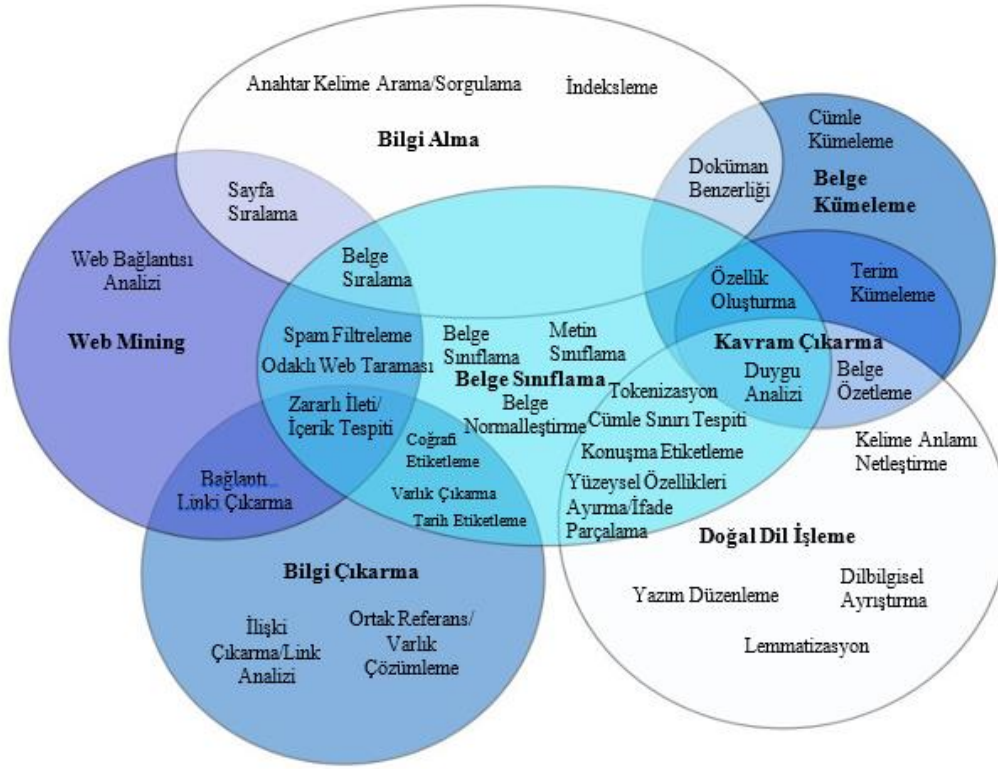
Genellikle hesaplamalı dilbilim kavramı ile eş anlamlı olarak anılmaktadır. Doğal dil işleme, hem dilbilimde hem de bilgisayar bilimlerinde uzun ve etkin bir geçmişe sahiptir. Son zamanlarda, DDI' nin odak noktası, istatistiksel yaklaşımları göz önünde bulunduran metin madenciliği alanıdır. DDI, konuşma etiketlerinin görevleri ve ifade sınırları gibi metin madenciliği için yararlı girdi değişkenlerini sağlamak için güçlü bir araçtır (Manning ve Schütze, 1999).

2.2.3.7. Kavram Çıkarma

Kavram çıkarma, kısaca, kelimelerin veya cümlelerin anlamsal olarak benzer gruplar halinde gruplanması işlemidir. Kavram çıkarma, bazı yönlerden uygulama yapılması hem en kolay hem de en zor alanlarından biridir. Metnin anlamını, otomatik sistemlerin "anlaması" herkesin bildiği gibi zorlu bir işlemdir. Bununla birlikte, insan anlayışıyla birleştirilen otomatikleşme çalışmaları, bir makinenin veya tek başına bir insanın performansına bağlı olan çalışmalara nazaran önemli ölçüde yol etmiştir (Croft, Metzler ve Strohman, 2009).

2.2.4. Metin Madenciliği Uygulamaları

Metin madenciliğinin üzerine kurulu olduğu yedi uygulama alanı ve bu uygulama alanlarının etkileşimleri neticesinde ortaya konan yöntem, teknik ve uygulamalar Şekil 2.1’de gösterilmiştir.



Şekil 2.1 Metin Madenciliğini Alt Disiplinleri ve Bağlantılı Uygulamalar

Metin madenciliğinde birbirleri ile etkileşim içerisinde olan bu yedi uygulama alanı birtakım ortak yöntem ve tekniklere sahiptir. Şekil 2.1, aynı zamanda araştırmada kullanılan ve eğitimde metin madenciliği yöntemi olan duygu analizi ele alınarak incelendiğinde, kavram çıkarma alanı temelinde tanımlanan duygu analizinin DDİ alanıyla yakından ilişkili olduğu görülmektedir. Bu doğrultuda duygu analizinin temel adımları olan tokenizasyon, cümle sınırı tespiti, konuşma etiketleme, yüzeysel özellikleri ayırma ve ifade parçalama gibi adımların aynı zamanda DDİ alanına dahil olan ortak teknikler olduğu görülmektedir (Nisbet ve diğerleri, 2018).

2.2.5. Metin Madenciliğinin Genel Adımları

Metin Madenciliği, yapılandırılmamış metin biçimindeki veri kaynaklarında keşfedilmemiş bilgileri, ilişkileri ve kalıpları tespit etme sürecidir (Younis, 2015).

Bir metin madenciliği projesine başlamadan önce, öncelikle metin madenciliğinin kavramsal temellerinin anlaşılması ve ardından karar vermeyi yönlendirmek için verilerin gücünden nasıl yararlanılacağına iyi bilinmesi gerekmektedir. Herhangi bir modelleme ya da analiz faaliyetine harcanan zamanın önemli bir kısmını verilere erişim, veri ön işleme ve veri temizleme adımları kapsamaktadır. Dolayısıyla çeşitli kaynaklardan gelen verilere erişme, bunları istatistiksel analiz ortamına yükleme ve analize elverişli bir şekilde işleme yeteneğine sahip olması gerekmektedir (Kumar ve Paul, 2016).

2.2.5.1. Veri Setini Oluşturma

Metin madenciliği süreci, çeşitli kaynaklardan bir dizi yapılandırılmamış ya da yarı yapılandırılmış metnin toplanması ve belge koleksiyonlarının oluşturulmasıyla başlamaktadır (Miner, 2012).

2.2.5.2. Metin Ön İşleme

Metin madenciliği faaliyetinin öncesinde metin üzerinde duraklama ifadelerini kaldırma, yaygın olarak kullanılan ve anlama katkısı olmayan kelimeleri ayıklama, kelimeleri köklerine ayırma gibi ön işleme teknikleri uygulanmaktadır. Bu süreç sonunda satır ve sütunların belge ve belge içerisindeki terimleri temsil ettiği, belge-terim matrisi olarak bilinen yapısal tablolar elde edilir (Kumar ve Paul, 2016). Bu prosedür ile, daha küçük veri kümesi boyutu, daha az hesaplama ve arama alanının daraltılması gibi önemli avantajlar elde edilmektedir (Patel ve Soni, 2012). Metin ön işlemenin kendi içerisindeki olası adımları, tüm metin madenciliği görevleri için aynıdır, ancak hangi işleme adımlarının kullanılacağı belirlenen göreve bağlıdır (Miner, 2012).

2.2.5.3. Metin Madenciliği Yönteminin Uygulanması

Metin madenciliğinin bir diğer adımı olan yöntem uygulama adımı ise metindeki terim ilişkilerini ve ilginç kalıpları keşfetmek için kümeleme, sınıflandırma, ilişkilendirme gibi temeli veri madenciliğine dayanan metin madenciliği yöntemleri uygulanmaktadır (Kumar ve Paul, 2016).

2.2.5.4. Değerlendirme ve Yorumlama

Metin madenciliği sürecinin son adımı ise doğruluk, f ölçümü vb. hesaplama yöntemleri yardımıyla sonuçların değerlendirilmesi ve yorumlanması anlamına gelmektedir. Bu amaçla kelime bulutu, frekans poligonları gibi çeşitli görselleştirme araçlarından yararlanılmaktadır (Kumar ve Paul, 2016).

2.2.6. Eğitimde Metin Madenciliği için Kullanılan Yöntem ve Teknikler

Metin Madenciliği, genellikle yapılandırılmamış metinden ilginç ve önemsiz olmayan bilgilerin çıkarılması sürecini ifade etmektedir. Finans, işletme, klinik, biyoloji, biyomedikal gibi farklı alanlarda çeşitli uygulamalarına rastlamak mümkündür (Luque, Luna, Luque ve Ventura, 2019). Berry ve Castellanos (2004) tarafından yapılandırılmamış metinden yüksek nitelikteki bilgiyi çıkarma işlemi olarak tanımlanan metin madenciliği, eğitim alanındaki araştırmalarda, eğitimde metin madenciliği adı ile ortaya konmaktadır.

Öğrenme sürecinde eğitimcileri ve öğrencileri desteklemeyi amaçlayan ve geleneksel eğitim için yeni bir alternatif haline gelen Web tabanlı Uyarlanabilir ve Akıllı Eğitim Sistemleri, Web tabanlı Akıllı Öğretim Sistemi, Öğrenme Yönetim Sistemi, Devasa Açık Çevrimiçi Kurslar (MOOC) gibi çevrimiçi eğitim platformları, karmaşık yönetim ve analiz yöntemleri gerektiren büyük miktarda yapılandırılmış ve yapılandırılmamış veri üreten ve etkileşim içerisinde olan yüzlerce veya binlerce öğrenci bilgisini bir arada tutmaktadır (Romero ve Ventura, 2013). Özellikle, tartışma forumu, sohbet, wiki, blog gibi farklı kaynaklardan, farklı formatlarda üretilen büyük miktarda yapılandırılmamış metin verisi genel olarak işlenmek için uygun değildir ve bu durum, öğrenme sürecinde eğitimciler ve öğrencilere yardımcı olacak olan bilginin kullanımı ile ilgili sorunlara yol açmaktadır (Valjataga, Põldoja ve Laanpere, 2011). Bu doğrultuda eğitimde veri madenciliği, metin madenciliği ve öğrenme analitiği teknikleri,

öğrencilerin öğrenmelerini geliştirmek ve öğrenme sürecinin geliştirilmesinde eğitimcilere yardımcı olmak amacıyla başarıyla uygulanmaktadır (Romero ve Ventura, 2017).

Eğitim alanında başlıca kullanılan metin madenciliği yöntemlerinin genel olarak metin sınıflandırması ve doğal dil işleme olduğu görülmektedir. Bu durum, genel uygulamalarda kullanılan metin madenciliği teknikleri ile benzer eğilimdedir (Hirschberg ve Manning, 2015).

Yeni nesil çevrimiçi platformlarda, doğal dil işleme, metin sınıflandırma ve kümeleme, bilgi alma ve metin özetleme gibi farklı metin madenciliği tekniklerinden sıklıkla yararlanılmaktadır. Eğitimde Metin Madenciliği tekniklerinin uygulanması ile, özellikle çevrimiçi değerlendirme ve deneme, forum ve sohbet analizi, akademik metin üretimi, sosyal ağ ve blog gibi eğitim ortamlarında önemli sonuçlar elde edilmiştir (Romero ve Ventura, 2013).

2.2.6.1. Metin Sınıflandırma ve Kümeleme

Sınıflandırma, nesneleri önceden tanımlanmış kategoriler içerisine özelliklerine göre yerleştirmektir. Kümeleme ise nesneleri veya bireyleri, kendi aralarındaki benzerliğe göre kategorize etmektir (Aggarwal ve Zhai, 2012).

Metin sınıflandırma, eğitim ortamlarında farklı hedefler için kullanılmıştır. Bunlar eğitici desteği/dersi, sınıf tartışması ve öğrenci grup çalışması gibi genel söylevlerdeki faaliyetlerin otomatik olarak sınıflandırılması forum tartışmalarının konulara ve türlere göre sınıflandırılması, görüş (duygu) analizi, soru sınıflandırma ve otomatik puanlamadır (Lin, Hsieh, ve Chuang, 2009; Azevedo, Behar, ve Reategui, 2011; Wang, Pan, Miller, ve Cortina, 2014; Tobarra, Robles-Gómez, Ros, Hernández, ve Caminero, 2014; Ruseti ve diğerleri, 2018; Newman ve Joyner, 2018; Yoo ve Kim, 2014). Özellikle, Görüş Analizi, bir olay, doküman veya etkileşim için ortaya konan duygusal göstergeyi/ bağlamsal kutbu ya da bazı konularla ilgili yazarın, okuyucunun vb., tutumunu belirlemek amacıyla eğitimde yaygın biçimde kullanılan bir tekniktir (Mello, Andre, Pinheiro, Costa ve Romero, 2019).

Metin kümelemeye ilişkin olarak, literatürde bulunan eğitim uygulamaları öğrenci grupları ve eğitim kaynakları ile ilgilidir. Öğrenci kümeleme uygulamaları, metin tahminlerini iyileştirmek, müfredatı uyarlamak, katılımı ya da etkileşimi ölçmek, öğrenmeyi tanımlamak için kullanılmaktadır (Trivedi, Pardos, ve Heffernan, 2011; Shi, Larson, ve Jonker, 2015; Liu, Calvo, ve Pardo, 2013; Cobo ve diğerleri, 2010; Mansur ve Yusof, 2013).

2.2.6.2. Bilgi Çıkarma

Bilgi alma, geniş bir metin koleksiyonu içerisindeki belgeleri araştırma bilimi olarak tanımlanmaktadır (Manning, Raghavan, Schütze ve diğerleri, 2008). Eğitim alanında, farklı bir uygulama olarak, kütüphanelerde konuyla ilgili belgeleri bulmak amacıyla bilgi kurtarma yöntemi uygulanmıştır (Chinkina, Oswal ve Meurers, 2018). Bu uygulamaların temel amacı öğrencinin ilgili kitapları bulmasına yardımcı olmaktır (Chen ve diğerleri, 2008). Ayrıca, e-öğrenmede çevrimiçi tartışmayı ve iş birliğini geliştirmek için kullanılmaktadır (Distante, Fernandez, Cerulo ve Visaggio, 2014). Bir diğer görev olarak, bilgi kurtarma teknikleri farklı metinler arasında gezinmeyi kolaylaştırmaktadır. Eğitim bağlamında, öğrencilerin denemelerinde ve yazma ödevlerinde farklı konuları göz önüne getirmek için de kullanılmaktadır (O'Rourke, Calvo ve McNamara, 2011; Villalón ve Calvo, 2011).

2.2.6.3. Metin Özetleme

Otomatik metin özetleme, temel bilgileri içeren bir veya birkaç belgenin kısa bir versiyonunu oluşturmaktır. Çoğunlukla dijital kütüphaneler, web haberleri, bilimsel makaleler gibi metin-veri setlerindeki aşırı bilgi yükü ile başa çıkmak için kullanılmaktadır. Otomatik metin özetleme teknikleri, çıkarıcı ve özetleyici olarak iki şekilde sınıflandırılmıştır. Çıkarıcı sistemler, özeti oluşturmak için tam olarak görüldüğü şekilde, belgedeki en önemli cümleleri seçmektedir. Özetleyici sistemler ise, metindeki fazlalıkları eleyerek ve bağlamları netleştirerek özetteki cümleler arasındaki tutarlılığı geliştirmeye çalışmaktadır. Otomatik metin özetleme, eğitim ortamlarında temel olarak içeriği özetlemek, anahtar kelimeyi çıkarmak, öğrenci metinlerinden görsel modeller oluşturmak amacıyla uygulanmaktadır (Jorge-Botana, Luzón, Gómez-Veiga, ve Martín-Cordero, 2015; Sung, Liao, Chang, Chen, ve Chang, 2016).

Söz konusu uygulamaların yanı sıra, otomatik metin özetleme teknikleri, forum yazılarının otomatik olarak değerlendirilmesi, öğrencilere akademik metinler yazmalarında yardımcı olunması, işbirlikçi öğrenmenin geliştirilmesi, öğrenci geri bildirimini değerlendirilmesi gibi faaliyetlerde geleneksel eğitim kaynaklarının dezavantajlarıyla başa çıkmak amacıyla kullanılmaktadır (Kang ve diğerleri, 2008; Azevedo, Behar ve Reategui, 2011; Whitelock, Twiner, Richardson, Field, ve Pulman, 2015; Nitin, Swapna, ve Shankararaman, 2015).

2.2.6.4. Doğal Dil İşleme

Doğal dil işleme, bilgisayar biliminde doğal dil metnini veya konuşmayı manipüle etmek için kullanılan bir alandır (Chowdhury, 2003). Anlamsal ve sözdizimsel analiz için algoritmalar aracılığıyla çok miktarda doğal dil verisi işlenebilmektedir (Litman, 2016).

Eğitimde doğal dil işleme uygulamalarının büyük çoğunluğu, denemelerin ve açık uçlu soruların otomatik olarak değerlendirilmesi ile ilgilidir. Denemelerle ilgili olarak doğal dil işleme, temelde metin uyumunun ve yazılı tartışmanın analizinde kullanılmıştır (Balyan, McCarthy ve McNamara, 2017; Elouazizi ve diğerleri, 2017). Açık uçlu sorulara gelince; esas araştırma hedefi, öğrencilerin cevaplarını anlamsal olarak değerlendirmek ve yeni sorular oluşturmaktır (Flor ve Riordan, 2018; Cutrone ve Chang, 2010). Ayrıca, öğrencilerin temel çevrimiçi kurslardaki etkileşimlerine dayalı olarak akıllı özel dersler için geri bildirim sağlamak amacıyla farklı doğal dil işleme teknikleri benimsenmiştir (Dzikovska, Steinhauer, Farrow, Moore, ve Campbell, 2014).

Etkileşimleri artırmak için doğal dil işlemenin benimsemesi fikrini takiben, sohbetlerde iş birliğine dayalı çalışmaları desteklemek, grup proje ve geçmiş tartışmalara bağlı olarak etkileşimleri tahmin etmek amaçlı çalışmalar ortaya konmuştur (Trausan-Matu, Dascalu, ve Rebedea, 2012; Kim ve Shaw, 2014).

2.2.7. Eğitimde Metin Madenciliği Kaynakları

Eğitimde metin madenciliği yapmak için mevcut ana eğitim kaynaklarında çoğunluğu forumlar, çevrimiçi değerlendirmeler ve denemeler oluşturmaktadır. Ayrıca, sosyal ağlar olarak adlandırılan topluluk uygulamaları ve özellikle dijital kütüphanelerdeki belgeler gibi diğer kaynaklara da yönelik artan bir ilgi bulunmaktadır (Romero ve Ventura, 2013).

2.2.7.1. Forum

Forum veya çevrimiçi tartışma, birden fazla kullanıcının eş zamansız olarak etkileşime girdiği bir iletişim aracıdır. Eğitim bağlamında forum, öğrenciler ve öğretmenler arasında daha fazla etkileşime izin veren bir kaynaktır (Caspi, Gorsky ve Chajut, 2003). Ayrıca, öğretmenlerin sınıfla verimli bir şekilde etkileşime girmeleri için çeşitli olanaklar sunmaktadır (De Wever ve diğerleri, 2006). Forumun eğitim ortamlarında bir iletişim kaynağı olarak benimsenmesiyle elde edilen avantajlara rağmen bu durum, aşırı bilgi sorununu da beraberinde getirmiştir (Wulf,

Blohm, Leimeister ve Brenner, 2014). Bu nedenle, gerekli ve önemli bilgileri elde etmek için otomatik yöntemlerin kullanılması önem arz etmektedir.

Forumlarda, kullanıcılara ya da öğrencilere otomatik geri bildirim sağlamak ve öğrencilerle etkileşimde bulunmak amacıyla sıklıkla kullanılan metin madenciliği tekniklerinin doğal dil işleme ve metin sınıflandırma olduğu görülmektedir (da Costa Pinho, Epstein, Reategui, Correa ve Polonia, 2013; Dzikovska ve diğerleri, 2014). Metin sınıflandırma, öğrencilerin çevrimiçi katılımının ve katkısının değerlendirilmesi, bir araştırma projesi tarafından başlatılan çevrimiçi öğrenmenin çeşitli yönlerinin araştırılması, karşılaştırmalı olarak öğrenme performansının değerlendirilmesi ve izlenmesi için kullanılmaktadır (Lui, Li ve Choy, 2007).

2.2.7.2. Çevrimiçi Değerlendirme

Çevrimiçi değerlendirme, deneme çalışmaları gibi öğrencilerin performanslarını değerlendirmek amacıyla kullanılan bir kaynaktır. En popüler iki değerlendirme türü olarak seçenekler arasında seçim yapılması gereken çoktan seçmeli ve öğrencinin cevap metni yazması gereken açık uçlu sorular öne çıkmaktadır (Wang ve diğerleri, 2014).

Metin madenciliği teknikleri, soruların hem oluşturulması hem de değerlendirmesi konusunda önemli ölçüde fayda sağlayabilmektedir. Bu doğrultuda Araki ve diğerleri (2016) tarafından soruların ve çoktan seçmeli cevapların otomatik olarak oluşturulması için doğal dil işlemeye ve şablon tabanlı algoritmalara dayanan metin madenciliği yöntemleri önerilmiştir. Soru oluşturma yanısıra, daha önce formüle edilmiş soruların analiz edilebilmesi metin madenciliğinin sağladığı bir diğer kolaylıktır.

Çevrimiçi değerlendirme platformlarında metin kümeleme ve konu modelleme gibi metin madenciliği teknikleri, veri tabanındaki soruları nitel olarak değerlendirmek amacıyla uygulanmaktadır (Nagashree ve Pujari, 2016). Ayrıca, öğrenmenin herhangi bir aşaması için hangi sorunun belirleyici olduğunu bilmek önemlidir (Sachan ve Xing, 2016). Soru değerlendirmesi ile ilgili olarak, açık uçlu soruların değerlendirilmesinde eğiticiye yardımcı olmak amacıyla araştırmacılar tarafından farklı öneriler ortaya konmuştur (Cutrone ve Chang, 2010; Rahimi ve diğerleri, 2014; Rus, Banjade, Lintean, Niraula, ve Stefanescu, 2013).

2.2.7.3. Sohbet

Sohbet, eşzamanlı iletişimi sağlayan bir kaynaktır. Eğitim ortamlarında, sohbetler, bir konuyu tartışmak, grup etkinliği gerçekleştirmek ve soru sormak için kullanılabilir. (Ferguson ve Shum, 2011). Sohbetler üzerinde metin madenciliği, genel olarak konuşma konusunun özetlenmesi, kullanıcı davranışının anlaşılması, kullanıcı özelliklerinin araştırılması, sosyal ve anlamsal etkileşimleri anlaşılması, sohbet ortamı konuşmalarının izlenmesi, ilgili bilgilerin çıkarılması, yazarlık niteliğinin değerlendirilmesi şeklindeki amaçlarla uygulanmaktadır (Leiyue ve Jianying, 2012).

Eğitim sohbetleri üzerinde metin madenciliği yardımıyla iş birliğini tanımlayan ve değerlendiren birçok çalışma bulunmaktadır. İş birliği, söylem uyum analizi kullanılarak otomatik olarak değerlendirilebilmektedir (Dascalu, Trausan-Matu ve diğerleri, 2015). İş birliğini araştıran ve daCosta Pinho ve diğerleri (2013) tarafından ortaya koyulan bir başka çalışmada da eşzamanlı çevrimiçi tartışmaların sürdürülmesi için pedagojik bir araç önerilmektedir. Trausan-Matu ve diğerleri (2012), Bakhtin'in diyalogculuğuna dayalı olarak konuları ve kullanıcı katkısını tanımlamak için ifadeler ve kelimeler arasında grafik/diyagram kullanan bir yöntem önermiştir (Koschmann, 1999). Scheuer ve McLaren (2008), eğitimcilerin sohbet verilerinde ilgili bilgileri bulmalarına yardımcı olmak amacıyla metin madenciliğini benimsemiştir. Reimann, Frerejean ve Thompson (2009) da sohbetlerde tartışılan olayların sırasını sunan bir model oluşturarak eğitime yardımcı olunabileceği bilgisini ortaya koymaktadır. Trausan-Matu ve diğerleri (2012) öğrencinin tartışmalarındaki önemli noktaları bulmak için doğal dil işleme tekniklerinin benimsenmesi gerektiğini öne sürmüştür. Ayrıca birçok çalışma eş zamansız tartışmaya uygulanan duygu analizini geniş ölçüde tanıtmaktadır (Coutinho, Moreira, Paillard ve de Lima, 2016). Son olarak, birçok çalışmada, öğretici bir diyalog sohbetinde ikinci dil ediniminin analizi üzerinde durulmuştur (Nayak ve Rao, 2018; Xu, Chen ve Qin, 2018).

2.2.7.4. Belge, Sosyal Ağ, Blog ve E-posta

Eğitim ortamlarında kullanıcı etkileşimi ile ilgili bilgi sağlayan diğer kaynaklar sosyal ağlar, bloglar ve e-postalardır. Bu kaynaklarda kullanan metin madenciliği uygulamaları genellikle metin sınıflandırması ile ilgilidir ve temel olarak duygu analizine odaklanmaktadır. Bu bağlamda, eğitim ortamıyla ilgili görüş bildirimi, öğrenci katılımının artması için eğitim ortamını geliştirilmesinde eğitime yardımcı olunması, uyarlanabilir bir eğitim ortamı

oluşturulması, sosyal ağdaki kullanıcı etkileşimine dayanan geri bildirimin sağlanması ve değerlendirme öğretiminde eğiticiye yardımcı olunması gibi farklı problemleri çözmek amacıyla görüş analizi uygulanmaktadır (Kechaou, Ammar, ve Alimi, 2011; Ortigosa, Martín, ve Carro, 2014; Leong, Lee, ve Mak, 2012).

Duygu analizine ek olarak, metin sınıflandırma da öğrenci davranışlarını hakkında bilgi edinmek, eğiticilere geri bildirimi geliştirmeleri ve öğrenci kaybını önlemeleri konusunda yardımcı olmak için kullanılmaktadır. Bu doğrultuda Mansur ve Yusof (2013), kullanıcı davranışlarını elde etmek için eğitim ortamlarından alınan günlüklerin ve sosyal ağ analizinin bir kombinasyonuna dayanan hibrit (karma) bir yöntemin benimsenmesini önermiştir. Benzer şekilde, Tobarra ve diğerleri (2014), öğrenci modellerini oluşturmak için sosyal ağ ve forum etkileşimlerini birleştirmektedir. Metin madenciliği uygulamasının bir başka örneği ise e-posta analizi ile ilgilidir (Aghaee, 2015). Son olarak, kurs blogosferi (blog alemi) ile etkileşimi temelinde her öğrencinin notunun otomatik olarak tahmin edilmesi mümkündür (Gunnarsson ve Alterman, 2012).

2.2.8. Eğitim Hedefleri

Eğitimde metin madenciliği uygulamalarının mevcut ana eğitim hedefi, öğrencilerin performanslarının değerlendirilmesinde öğreticilere yardımcı olmaktır. Öte yandan, öğrencilerin desteklenmesi ve eğiticiler için analiz sağlanması amacıyla otomatik yöntemlerin geliştirilmesi temel eğitim hedefidir.

2.2.8.1. Değerlendirme

Metin madenciliği, özellikle denemelerin ve çevrimiçi ödevleri değerlendirilmesi şeklinde, farklı bağlamlardaki öğrenci performansını incelemek amacıyla geniş ölçüde kullanılmıştır (Ranathunga, ve Dias, 2017). Denemeleri değerlendirmek için, birçok çalışmada, sözcük sayımı ile yüzeysel özelliklerin elde edilmesi önerilmiştir (Dikli, 2006). Ancak, bu analizin ötesine geçmek önemlidir ve yazma stili, tartışma gibi semantik yöntemlerin benimsenmesine odaklanan çok sayıda çalışma bulunmaktadır (Hughes, Hastings, Magliano, Goldman ve Lawless, 2012; Şimsek ve diğerleri, 2015; Snow, Allen, Jacovina, Perret, ve McNamara, 2015). Benzer doğrultuda, çevrimiçi ödevlerin değerlendirilmesinde sözcüksel ve anlamsal bir yaklaşım benimsenmektedir (Ramachandran ve Gehring, 2011). Bununla birlikte, çalışmalar intihal, kısa cevap analizi ve soru sınıflama gibi belirli sorunların çözümüne

daha fazla odaklanma eğilimindedir (Saha, Dhamecha, Marvaniya, Sindhgatta ve Sengupta, 2018). Son olarak çeşitli metin madenciliği yöntemleri, biçimlendirici değerlendirmede, eğitime yönelik çevrimiçi tartışmalardaki etkileşimleri değerlendirmek ve bu ortamı sürdürmek amaçlı pedagojik temelli kararların oluşturulması konusunda eğitimcileri desteklemek amacıyla uygulanmaktadır (Gibson ve diğerleri, 2017).

2.2.8.2. Destekleme

Çevrimiçi platformlarda öğrencilerin devamlılığını sağlamak çok önemlidir. Özellikle uzaktan eğitim derslerinde, pedagojik başarı için öğrenciler arasında iş birliği şarttır. Bu nedenle, metin madenciliği öğrencilere destek sağlamak amacıyla kullanılmaktadır (Liu, Pardo ve Liu, 2017). Benzer şekilde, bu hedefe ilişkin yayınların çoğu yazma etkinliğine de sahiptir. Birçok çalışma, geleneksel makaleler/dememeler, akademik belgeler ve e kitaplar gibi farklı metinlerin üretiminde öğrencilere yardımcı olmaya odaklanmıştır. Ayrıca, bu amaç doğrultusunda, forumlar, sohbetler ve bloglar gibi etkileşimli kaynakların analizi üzerinde durulmuştur. Bu bağlamda, öğrencilerin duygu ve düşüncelerini elde etmek için görüş ya da diğer bir deyişle duygu analizi yöntemleri üzerinde çalışılmış; iş birliğini teşvik etmek, kayıpları önlemek ve öğrenci motivasyonunu korumak amaçlanmıştır (Sinclair ve diğerleri, 2018). Ayrıca, bu araçlar içerisinde öğrencileri izlenmesi konusunda eğitimcilere yardımcı olmayı amaçlayan birkaç çalışma bulunmaktadır (Scheuer ve McLaren, 2008).

2.2.8.3. Analiz

Eğiticilere, eğitim ortamını iyileştirmeleri ve öğrencilere sürekli geri bildirim vermeleri amacıyla analiz sağlanması hayati önem taşımaktadır (Lyons, Aksayli ve Brewer, 2018). Bu bağlamda, eğitimde metin madenciliği yöntemleri, yazı üretimi ve forum / sohbet / e-posta etkileşimleri gibi faaliyetlerden bilgi elde edilmesi için kullanılmaktadır. Uygulamaların çoğu, öğrencinin performansının ve davranışının analizinde eğiticiye yardımcı olabilecek bilgileri çıkarmayı amaçlamaktadır (Nunes ve diğerleri, 2014; Joksimovic ve diğerleri, 2014; Aghaee, 2015).

2.2.8.4. Soru-İçerik Oluřturma

Otomatik soru oluřturma verilen metne baėlı olarak ierikle ilgili sorular oluřturmaya odaklanmaktadır. Otomatik ierik oluřturma ile, verilen herhangi bir konuda, ierik oluřturabilmektedir. Metin madenciliėi, eėitim ortamlarında eėiticilere yardımcı olmak amacıyla otomatik olarak soru ve ierik uřetimi iin uygulanmaktadır (Araki ve diėerleri, 2016).

En geleneksel yaklařım, soruyu, ders kitapları ve eėitim ortamları benzeri bařvuru kaynaklarından oluřturmaktır; ancak web uřerinden de kullanıřlı bilgiyi elde etmek mmkndr (Mazidi ve Tarau, 2016). Dersin ieriėine gre soruların deėerlendirilmesi ve temel yapının kullanılması, farklı metin madenciliėi grevleridir. rneėin, Sachan ve Xing (2016), eėiticilere, bařlangıta daha basit sorular gnderildiėinde ėrencilerin daha istekli bir řekilde derse devam ettikleri ynndeki analiz iin bir vaka alıřması nermiřtir.

2.2.8.5. Geri Bildirim

Geri bildirimin amacı, ėrencilere performanslarını artırmalarına yardımcı olacak bilgileri vermektir. Eėitimde metin madenciliėi, ėrencilere, etkileřimlerine ve gerekleřtirilen etkinliklere gre otomatik olarak geri bildirim saėlamak amacıyla kullanılmaktadır (Woods ve diėerleri, 2017). Genel olarak, literatrde bu amacı gerekleřtirmek zere iki farklı yaklařım ele alınmıřtır. İlki, ėrenci davranıřlarını veya performanslarını analiz etmek iin otomatik bir ynteme baėlı olarak, bir gsterge tablosu veya tavsiye gibi doėrudan ėrencilere geri bildirim gndermeye odaklanmaktadır (Lewkow ve diėerleri, 2016). Bu baėlamda, uygulamanın amacı, iřbirliki ortamlarda ėrencileri desteklemek ve gerek zamanlı olarak geri bildirim gndermek iin soru cevaplama uygulamasında ėrencilere yardımcı olmaktır. İkinci yaklařım, farklı kaynaklardan gelen geri bildirimlerin deėerlendirilmesinde eėiticilere yardımcı olacak bilginin toplanması iin eėitimde metin madenciliėi kullanılmasıdır.

2.2.8.6. Tavsiye Sistemi

İnternet ve web tabanlı eğitim ortamlarındaki, mevcut kaynakların çokluğu nedeniyle, her bir ders ve öğrenci için en iyi içeriğin bulunması zordur (Hsu, Hwang ve Chang, 2010). Dolayısıyla tavsiye sistemlerinin benimsenmesi ile, eğitimcilere ve öğrencilere en alakalı bilgileri bulma sürecinde yardımcı olunması sağlanmıştır. Tavsiyeler, kitaplar gibi birleştirilmiş eğitim materyallerine ve web gibi açık ortamlara dayandırılabilir (Garrido, Pera ve Ilarri, 2014).

Acosta ve diğerleri (2014), web' de ilgili içerikleri aramak için öğrencilerin yazılarındaki alakalı terimleri ve anahtar kelimeleri çıkaran bir sistem önermiştir. Bu bağlamda, tavsiye sistemi farklı öğrenciler için en alakalı içeriği filtrelemektedir. Figueira (2008b) daha fazla öneri geliştirmek amacıyla öğrenme nesnelerini anlamsal kategorilere göre sınıflandırmak için metin madenciliği tekniklerini uygulamıştır.

2.3. Duygu Analizi

2.3.1. Duygu Analizinin Tanımı ve Temel Kavramları

Duygu analizi, veri madenciliği, web madenciliği ve sosyal medya analitiği gibi alanlarda yaygın olarak kullanılan temel bir metin madenciliği uygulamasıdır. Bir "Duygusal bilgi işlem" alanı olan duygu analizi, insanların farklı olaylara, sorunlara, hizmetlere veya başka herhangi bir ilgi alanına yönelik ruh halini algılanması, analiz edilmesi ve değerlendirilmesini ifade eder (Yadollahi, Shahraki ve Zaiane, 2017). Davranışların değerlendirilmesi konusunda önemli bir özellik olan duygu kavramıyla ilgilenmektedir. Bu doğrultuda duyguları barındırma ve iletme işlevi ile metinler, duygu analizinin temel veri kaynaklarını oluşturmaktadır (Chakraborty, Bhattacharyya, Bag ve Hassanien, 2019).

Duygu analizinin amacı, metin türündeki çeşitli kaynaklarda bulunan öznel bilgileri tespit etmek veya bir yazarın konu ya da belge üzerindeki genel duygusal eğilimini belirlemektir (Kumar ve Sebastian, 2012). Wilson, Wiebe, Bruce, Bell ve Martin (2004) öznelliği, bireylerin duygularının, düşüncelerinin, değerlendirmelerinin, inançlarının dilsel ifadesi olarak tanımlamaktadır.

Görüş, duygu, kanı ve inanç kelimeleri birbirinin yerine kullanılır ancak aralarında ince bir fark bulunmaktadır (Pang ve Lee, 2008). Bu doğrultuda Pang ve Lee (2008) tarafından;

Görüş, düşünülmüş ancak tartışmaya açık olan sonuç;

Kanı, öznel görüş;

İnanç, kasıtlı kabul ve entelektüel onay;

Duygu, kişinin duygularını yansıtan yerleşik bir fikir olarak tanımlanmaktadır.

Duygu analizi, genel olarak kullanıcı tarafından oluşturulan ve fikirleri, düşünceleri ya da görüşleri barındıran içerikler üzerinde yapılmaktadır. Bu içerik, bir nesneyi değerlendiren ürün incelemesi, bir forum gönderisi, bir blog veya bir tweet olabilmektedir. Belirtilen görüşler; ürünler, sorunlar, kuruluşlar veya bir hizmet gibi herhangi bir nesne, konu veya kişi hakkında olabilmektedir (Kumar ve Sebastian, 2012).

Liu (2010) matematiksel olarak bir görüşü ya da duyguyu beş temel ögeye ayırarak (n, ö, y, k, z) tanımlamaktadır. Burada n, nesneyi; ö, n nesnesinin özelliğini; y, n nesnesinin ö özelliği ile ilgili görüşün/duygunun yönü veya kutbunu; k, görüş veya duygunun sahibini ve z, görüşün ifade edildiği zamanı temsil etmektedir. Liu (2010)'un daha detaylı tanımlamalarına göre;

Nesne: Bir ürün, kişi, etkinlik, organizasyon, kurum veya konu olabilen bir varlıktır. Kendisiyle ilişkili niteliklere, özelliklere veya bileşenlere sahiptir.

Özellik: Değerlendirmenin yapıldığı nesnenin bir niteliğidir.

Duygu yönü veya kutbu: Bir özelliğe ilişkin bir fikrin yönelimini olumlu, olumsuz veya nötr olup olmadığını gösterir. Çoğu çalışma ikili sınıflandırma üzerinde, pozitif veya negatif olarak yapılmıştır. Ancak duyguların yoğunluğu çok güçlüden zayıfa şeklinde de derecelendirilebilmektedir. Uygulamaya bağlı olarak seviye sayısına karar verilebilmektedir.

Duygu veya görüş sahibi: Görüşü ifade eden kişi veya kuruluştur. Web 2.0 uygulamaları tarafından sağlanan zengin ve çeşitli veriler sayesinde metin madenciliği ve duygu analizi alanı hızla ilerlemektedir. Bloglar, forumlar, wiki'ler ve sosyal ağların tümü, duygu analizi için kullanılan verilere farklı boyutlar sağlamıştır.

2.3.2. Duygu Analizinin Görevleri

Duygu Analizi birbiri ile bağlantılı birçok göreve ya da hedefe sahiptir. Yakından ilişkili olmaları ve birçok temel hedefi paylaşımları bakımından net bir şekilde ayrılmaları zordur.

2.3.2.1. Duygu Sınıflandırma.

Duygu yönelimi, fikir yönelimi, anlamsal yönelim veya duygu kutupluluğunun belirlenmesi olarak da adlandırılmaktadır (Yu, Wu, Chann, Chu, 2013). Bir belgede ya da metinde, belirli bir varlık hakkındaki görüşün ifade edilmesi ile ve görüş sahibinin varlığa olan duygusal görüşünün ölçülmeye çalışılması fikrine dayanmaktadır. Temelde görüşlerin/duyguların olumlu, olumsuz veya tarafsız gibi çeşitli kategorilere ayrılması işleminden ibarettir. Basit bir görev gibi görünür; ancak, görüşlerin birden çok alan veya dilden geldiği dikkate alındığında karmaşık ve zorlu bir süreç barındırdığı görülmektedir (He, Lin ve Alani, 2011; Boiy ve Moens, 2008).

2.3.2.2. Öznellik Sınıflandırma

Temelde, belirli bir cümlemin öznel olup olmadığının tespit edilmesidir. Bu doğrultuda nesnel cümleler olgusal bilgileri; öznel cümleler ise görüşler, değerlendirmeler, duygular ve inançlar gibi diğer kişisel bilgi türlerini ifade etmektedir. Ek olarak, öznel cümleler olumlu veya olumsuz duyguları da belirtebilmektedir. Bu görev, duygu sınıflandırma görevinden önceki adım olarak görülmektedir. İyi bir öznellik sınıflandırması, daha iyi bir duyarlılık sınıflandırması anlamına gelmektedir. Ayrıca olumlu, olumsuz veya tarafsız duyguları birbirinden ayırmaktan daha zor bir süreç olarak kabul edilmektedir. (Barbosa ve Feng, 2010; Raaijmakers ve Kraaij, 2008; Sarvabhotla, Pingali ve Varma, 2011; Esuli ve Sebastiani, 2006; Montoyo, Martínez-Barco ve Balahur, 2012; Maks ve Vossen, 2012).

2.3.2.3. Görüş Özetleme

Özellikle bir veya birkaç belgede yer alan varlığın, temel özelliklerinin ve bu özelliklere ilişkin duyguların çıkarılması anlamına gelmektedir (Wang, Zhu, Li, 2013). Bu görev, tek belgeli ve çok belgeli özetleme şeklinde iki açıdan ele alınmaktadır. Tek belgeli özetleme, analiz edilen belgedeki mevcut iç gerçeklerin analiz edilmesinden oluşmaktadır. Bu doğrultuda varlığı en iyi tanımlayan metin parçalarının elde edilmesi amacıyla belge boyunca duygu yönelimindeki değişiklikler veya bulunan farklı varlık ya da özellikler arasındaki bağlantılar

tespit edilmektedir. Öte yandan, çok belgeli özetlemede ise, özellikler ve varlıklar tespit edildiğinde, sistem bu varlıklar veya özellikler ile ilgili duyguları ifade eden farklı cümleleri gruplandırmakta ya da sıralamaktadır. Devamında nihai özet, ana özellikleri ya da varlıkları gösteren ve her biri ile ilgili duyguyu nicelleştiren bir grafik veya metin olarak sunulabilmektedir. Örneğin, duyguların yoğunluğunu ortaya konulması veya olumlu-olumsuz cümlelerin sayısını belirlenmesi şeklinde görselleştirme yöntemlerinden yararlanılmaktadır (Pang, Lee, 2004; Park, Park, Kim ve Ko, 2012; Nishikawa, Hasegawa, Matsuo ve Kikui, 2010; Ganesan, Zhai ve Han, 2010).

2.3.2.4. Görüş Alma

Belirli bir sorgu hakkında fikir ifade eden belgelerin elde edilmesi anlamına gelmektedir. Bu tür sistemlerde, her belge için sorguya göre alaka düzeyi puanı ve sorgu hakkındaki görüş puanı şeklinde iki tür puan belirlenmesi gerekmektedir. Ayrıca bu puanlar aracılığıyla belgelerin amaca yönelik olarak sıralanması sağlanmaktadır (Lee, Song, Lee, Han ve Rim, 2011; Guo ve Wan, 2012).

2.3.2.5. Alay ve İroni

İronik ve alaycı içerik barındıran ifadelerin tespit edilmesi anlamına gelmektedir. İroni veya alaycılığın resmi olarak nasıl tanımlanabileceği konusunda araştırmacılar arasında fikir birliği olmaması nedeniyle bu alandaki en karmaşık görevlerden biridir (Filatova, 2012; Reyes ve Rosso, 2012).

2.3.2.6. Diğer Duygu Analizi Görevleri

Duygu Analizi; tür veya yazar tespiti, bir metin veya görüşü (Savoy, 2012; Montesi ve Navarrete, 2008) belirten kişinin tespit edilmesi veya kamuoyunun kişilere, şirketlere veya ürünlere karşı (Jindal ve Liu, 2007; Mukherjee, Liu ve Glance, 2012; Xie, Wang, Lin ve Yu, 2012) görüşlerinin çarpıtılmasına neden olan ve güvenilmeyen içerikler içeren görüşlerin tespit edilmesi şeklindeki çeşitli görevleri de barındırmaktadır.

2.3.3. Duygu Analizinin Düzeyleri

Amaçlarına ve ayrıntı düzeylerine göre duygu analizi çeşitli kategorilere ayrılmaktadır.

2.3.3.1. Belge Düzeyinde Duygu Analizi

Bu kategoride, doküman, uzunluğu dikkate alınmaksızın problemin temel birimini oluşturmaktadır. Dokümanın bütünü duygu yönünün belirlenmesi, analizin temel amacıdır. Belge düzeyinde duygu kutbu sınıflandırması, bu alandaki çalışmaların çoğunu kapsamaktadır ve en temel duygu analizi görevi olarak kabul edilmektedir (Liu 2015). Aynı zamanda, çevrimiçi verilerde inceleme, blog gönderisi ve yorum gibi metin belgelerinin çoğunluğu oluşturmaları ile kullanım alanı olarak oldukça geniştir. Belge düzeyinde duygu analizi, sosyal ağ üzerinde yapılan tüketici memnuniyeti ve tıbbi ortamlarda hastaların analizi gibi toplumsal ve psikolojik çalışmalar için temel bir gerekliliktir (Kang ve Park, 2014; Ortigosa ve diğerleri, 2014; Denecke ve Deng, 2015; Gao ve diğerleri 2015a).

2.3.3.2. Cümle Düzeyinde Duygu Analizi

Temelde cümlelerin kutupluluğunun belirlenmesidir. Neviarouskaya ve diğerlerine (2007) göre bu seviyedeki esas zorluk, cümleyi çevreleyen bağlamın cümle üzerindeki etkisidir. Örneğin, hangi bağlamda kullanıldığına bağlı olarak, "Bu ürünü bundan daha iyi tarif edemem" cümlesi hem olumlu hem de olumsuz görüş olarak ifade edilebilmektedir. Son yıllarda yoğun olarak incelenen tweetlerin duygu kutbu bakımında sınıflandırılması, cümle düzeyinde duygu analizinin en ilginç uygulamasıdır.

2.3.3.3. Özellik Düzeyinde Duygu Analizi

Özellik tabanlı fikir madenciliği olarak da bilinen bu kategori, bir ürün veya hizmetin belirli bir yönü hakkındaki fikir kutuplarını keşfedilmesi esasına dayanmaktadır. Örneğin, restoranlar hakkındaki görüşler, yemeklerin lezzeti ve restoranın temizliği şeklinde kalitenin iki yönü hakkında olabilir. Belge veya cümle düzeyinde sınıflandırmaların yeterli olmadığı, ürün ve hizmetlerinin çeşitli özelliklerine ilişkin fikirlerinin bir araya getirilmesi gerektiği durumlarda oldukça kullanışlıdır. Yönlerin metinden çıkarılması ve çıkarılan yönlerin kutupluluk sınıflandırması, özellik düzeyinde kutupluluk sınıflandırmasının iki ana bileşenidir. Hu ve Liu' nun (2004) çalışması, bu alandaki en eski çalışmalardan biridir.

2.3.4. Duygu Analizinde Kullanılan Yöntemler

2.3.4.1. Makine Öğrenmesi Tabanlı Yaklaşım

Makine öğrenmesi tabanlı yaklaşım denetimli ve denetimsiz olmak üzere iki ana kategoriye ayrılmaktadır. Her ikisinin de başarısı temelde duyguları tespit etmek için kullanılan uygun özelliklerin seçilmesine ve çıkarılmasına dayanmaktadır. Bu görevde DDİ teknikleri çok önemli bir rol oynamaktadır. Çünkü terimler (kelimeler veya n gram) ve sıklıkları, herhangi bir cümlemin anlamını değiştiren olumsuzluk birimleri ve cümlemin anlamını belirleyen bağlamlar uygulamanın niteliğini belirleyen temel noktalar arasındadır. Denetimli tekniklerle ilgili olarak, Destek Vektör Makineleri, Naive Bayes, Maximum Entropy, en yaygın tekniklerden bazılarıdır. Bu yaklaşımın kategorilerinden biri olan yarı denetimli ve denetimsiz teknikler ise, sınıflandırma için bir etiketli belge / görüş kümesine sahip olmanın mümkün olmadığı durumlarda önerilmektedir. Ayrıca, duyguları sınıflandırmak için denetimli ve denetimsiz teknikleri birleştiren hibrit yaklaşımlar kullanılabilmektedir (Bhonde, Bhagwat, Ingulka ve Pande; 2015).

2.3.4.2. Sözlük Tabanlı Yaklaşım

Sözlük tabanlı yaklaşım, temelde genel amaçlar için geliştirilmiş ve önceden derlenmiş duygu terimlerini içeren bir duygu sözlüğüne dayanmaktadır. Sözlük tabanlı yaklaşımda, sözlüğün kullanım amacı, oluşturma yöntemi ve içerdiği sözlük miktarı gibi özellikler ele alınarak belirlenen sözlüğe dayalı ve derleme dayalı yaklaşımlar şeklinde iki alt sınıflandırma bulunmaktadır. İlki, genellikle manuel bir şekilde toplanan ve açıklama eklenen duygu kelimelerinden oluşmaktadır. Bu yaklaşımda kullanılan duygu kelimesi kümesi, sözcüklerin eş ve zıt anlamlıları eklenerek genişletilebilmektedir. Derleme dayalı yaklaşımlar ise, belirli bir alanla ilgili duygu sözcüklerinin elde edilmesi amacı ile oluşturulmuştur. Bu yaklaşımlarda sözlükler genel olarak, istatistiksel veya anlamsal tekniklerin kullanılması birlikte ilgili kelimelerin aranması yoluyla büyüyen bir dizi temel terimden üretilmektedir (Yadollahi, Sahahraki ve Zaiane; 2017).

2.3.5. Duygu Sözlüğü

2.3.5.1. Duygu Kuramları

Duygu ile ilgili herhangi bir araştırmada cevaplanması gereken ilk soru, duygunun gerçekte ne olduğudur. Tezin bu kısmında, duyguyu tanımlayan ve bazı temel duygu setlerini öneren teoriler tanıtılmaktadır.

Bilgisayar biliminde duygular üzerine yapılan araştırmaların çoğunda duygu, ruh hali ve his terimleri birbirlerinin yerine kullanılsa da bu terimler aynı anlamı taşımamaktadır. Fox'a (2008) göre, afektif (duygusal) sinirbilimde terimler şu şekilde tanımlanmaktadır:

Duygu: Organizma için belirli bir önemi olan iç veya dış olaylara karşı ayrı ve tutarlı yanıtlardır. Ayrıca duygu durumu kısa vadelidir.

His: Duyguların, onları deneyimleyen bireye özel öznel bir temsildir.

Ruh hali: Duyguyla karşılaştırıldığında, genellikle daha az yoğunlukta ancak daha uzun süren yaygın bir duygusal durumdur.

Etki: Duygu, hisler ve ruh halleri konularını birlikte tanımlamak için kullanılan kapsayıcı bir terimdir.

Duyguların sınıflandırılmasına ilişkin bilimsel çalışmalar 1960'lara kadar uzanmaktadır. Ayrık duygu teorisi ve boyutsal model şeklinde bu alanda yaygın olan iki ayrı teori bulunmaktadır. Ayrık duygu teorisi, farklı duyguların farklı sinir sistemlerinden kaynaklandığını belirtirken; boyutsal model, tüm duygusal durumlardan ortak ve birbirine bağlı bir nörofizyolojik sistemin sorumlu olduğunu belirtmektedir. Ayrıca boyutsal model, duyguları bir veya daha fazla boyuta göre tanımlamakta ve bunlardan birinin genellikle duyguların yoğunluğuyla ilgili olduğunu ortaya koymaktadır (Yadollahi, Sahahraki ve Zaiane; 2017).

Temel duygular, içerisinde başka bir duyguya ait izlerin bulunmadığı ana durumlar olarak ifade edilmektedir. Ek olarak, duyguların ırk, kültür ve dil farkı gözetmeksizin dünyadaki her insan ortak olduğu belirtilmektedir. Duygunun tanımı ve oluşumu ile ilgili farklı görüşlere sahip olan bazı kuramcılar, çeşitli duygu kümelerini önermişlerdir.

Tablo 2.1’de, ilgili kuramcılar tarafından ortaya konan temel duygulara ilişkin sınıflandırmalara yer verilmiştir.

Tablo 2.1. Duygu Kavramı ile İlgili Teoriler

Teorinin Sahibi	Yıl	Temel duygular	Teori
Ekman	1972	öfke, tiksinti, korku, neşe, üzüntü, şaşkınlık	Ayrık
Plutchik	1986	öfke, beklenti, tiksinti, korku, neşe, üzüntü, şaşkınlık, güven	Boyutsal
Shaver	1987	öfke, korku, neşe, aşk, üzüntü, şaşkınlık	Ayrık
Lovheim	2011	öfke, tiksinti, sıkıntı, korku, neşe, ilgi, utanç, şaşkınlık	Boyutsal

En eski duygu kuramcılarında biri olan Ekman, evrensel olarak tanımladığı öfke, tiksinti, korku, neşe, üzüntü, şaşkınlık şeklindeki duyguların, temel duygular kümesini oluşturduğunu öne sürmüştür. Daha sonra olumlu ve olumsuz yöne sahip 12 yeni duygu ekleyerek temel duygu kümesini genişletmiştir (Ekman 1992). Plutchik ve Kellerman'ın (1986) ortaya koyduğu boyutsal modele göre ise duygular; neşe ve üzüntü, öfke ve korku, güven ve tiksinti, şaşkınlık ve beklenti şeklinde dört karşıt ekseninde bulunmaktadır. Ek olarak; öfke ve korku duygularındaki karşıt duygu zıtlığına ilişkin keskinliğin, sevinç ve üzüntü gibi karşıt duyguların zıtlığındaki keskinlik kadar sezgisel olmadığını öne sürmüştür. Öte yandan Shaver ve diğerleri (1987), duygu kategorilerini bir ağaç yapısına benzer şekilde modellemiştir. Öyle ki temel duygular ana dallardır ve her dalın kendi içerisinde kategorileri bulunmaktadır. Lovheim (2012) ise Plutchikte’den farklı bir boyutsal bir model önermektedir. Lovheim’in tanımladığı duygular üç temel hormon (serotonin, dopamin ve noradrenalinin) üzerine kurulur.

2.3.5.2. Duygu Sözlüğü Türleri

2.3.5.2.1. Polarite Temelli Sözlükler

Kavramların, kelimelerin veya görüşlerin temelde pozitif, negatif ya da nötr şeklinde duygu kutbuna sahip olduğu fikri ile oluşturulmuş sözlüklerdir.

2.3.5.2.1.1. Sözlükler

Sözlük oluşturma ya da önceki duygu analizi çalışmalarının genişletilmesi amacıyla ortaya konmuş olan halka açık birçok sözlük bulunmaktadır. Bu sözlükler arasında, literatürde en sık kullanılan ve en etkili olduğu bilinen sözlükler Tablo 2.2’de gösterilmektedir.

Tablo 2.2. Polarite Temelli Sözlükler

Sözlük	Yazar	Yıl	Sözcük Boyutu	Duygu-görüş kutupları
Harvard General Inquirer	P. Stone	1968	11.790	pozitif, negatif and nötr
Opinion Lexicon	B. Liu	2005	6.786	pozitif, negatif
MPQA	T. Wilson	2005	8.222	pozitif, negatif, nötr
WPARD	D. A. Medler	2005	1.400	pozitif , negatif
SentiWordNet 3.0	S. Baccianella	2010	155.287	dereceli polarite
NRC	S. M. Mohammad	2009+	çeşitli boyutlarda	pozitif, negatif

2.3.5.2.1.1.1. Harvard General Inquirer

Bu sözlük, duygu analizi için kelime listesi derlenmesi ile ilgili bir deneme çalışmasının sonucudur (Stone ve diğerleri, 1968). Sözlük, kelimelerin sözdizimine, anlambilimine özgü bilgilerini içermektedir. Her kelime için sağlanan bilgiler arasında, dikkat çeken bilgi "olumlu" ve "olumsuz" şeklindeki duygu kutbu bilgisidir. Sözlük, 11.790 kelime içermektedir. Ayrıca bu sözlükteki her kelimenin puanı 1, 0 veya -1 şeklinde belirtilmiştir. Bu da kelimenin sırasıyla pozitif, nötr veya negatif olduğu anlamına gelmektedir.

2.3.5.2.1.1.2. Opinion Lexicon

Bing Liu’nun duygu analizi çalışmalarının bir ürünü olan bu sözlük 2.009’ u olumlu, geri kalanı olumsuz olmak üzere 6.786 kelimedenden oluşmaktadır. Kelimelerin çıkarıldığı külliyat, müşterilerin ürünlerin çeşitli özellikleri hakkındaki görüşlerini içeren bir veri kaynağıdır. Bir ürünün sık kullanılan bir özelliğini içeren cümleleri bulunması ve bu cümlelerden sıfatların çekilmesi yoluyla ilgili duygu kelimeleri elde edilmiştir. Daha sonra, bu kelimeler, bir sözlük yardımıyla eş anlamlı ve zıt anlamlı kelimeler şeklinde ayrı ayrı belirtilerek puanlarına göre pozitif ve negatif olarak ikiye ayrılmıştır. Her kelimenin puanı, “Harvard General Inquirer” sözlüğünün puanlamasına benzer şekilde tanımlanmıştır (Hu ve Liu 2004; Liu ve diğerleri 2005).

2.3.5.2.1.1.3.Çok Perspektifli Soru Cevaplama (MPQA)

MPQA sözlüğü, her bir kelimenin ne kadar öznel olduğu ya da öznelliğinin ne kadar güçlü olduğu; pozitif, negatif veya nötr kutuplu olma durumu ve kelimenin kök olup olmadığı gibi bir dizi bilgiyi içerdiği, 8.222 kelimeden oluşmaktadır. Bu sözlük, aynı ekibin önceki çalışmalarından ortaya çıkan bir öznellik sözlüğü üzerine inşa edilmiştir (Wilson ve diğerleri, 2005a). İlk adımda, ifadeler ve kelimelere olumlu, olumsuz veya nötr olacak şekilde açıklama eklemek için açıklayıcılara bir dizi talimat ve açıklama şeması verilmiştir. İkinci adımda, sözlüğü değerlendirmek için açıklamalar arasındaki uyum ölçülmüştür. MPQA sözlüğü hem duygu sınıflandırması hem de kutupluluk(polarite) sınıflandırması için yararlı bir sözlüktür (Wilson ve diğerleri, 2005b).

2.3.5.2.1.1.4. WPARD

Medler ve diğerleri (2005), çevrimiçi bir form kullanarak, 342 lisans öğrencisinden bilgi toplamıştır. Katılımcılardan, -6 (çok olumsuz) ve +6 (çok olumlu) aralıklı bir ölçek kullanarak her kelimeyle ilişkilendirdikleri duygunun ne kadar olumsuz ya da olumlu olduğunu değerlendirmeleri istenmiştir. Sözlük, her bir sözcüğe karşılık gelen kutupluluğu ve bu kutupluluğun gücünü gösteren gerçek bir sayıya sahiptir. Ayrıca bu verilerden faydalanarak Wisconsin Algısal Nitelik Derecelendirme Veritabanı (WPARD) oluşturulmuştur.

2.3.5.2.1.1.5.SentiWordNet 3.0

SentiWordNet 3.0, Baccianella ve diğerleri (2010) tarafından oluşturulan bir sözcük kaynağıdır. Bu sözlük, önceki sürümü olan SentiWordNet 1.0 üzerine inşa edilmiştir. 155.287 kelime içermektedir. Ayrıca her kelimeye ondalıklı bir polarite derecesi eklenmiştir. SentiWordNet 3.0'ı oluşturma yöntemleri beş adımdan oluşmaktadır. Bunlar, arasında pozitif ve negatif kelimelerden oluşan bir kelime kökü setinden başlayarak sözlüğü genişletmek için eş ve zıt anlamlı kelimelerin elde edilmesi, yeni bir küme olarak nesnel kelimeler eklenmesi, sınıflandırıcı eğitimlerinin gerçekleştirilmesi, sınıflandırıcılarla kelime kümelerinin sınıflandırılması ve son olarak, kelime grafiği üzerinde değerlendirme yapılması şeklindeki adımlardır. Son zamanlarda, SentiWordNet' in diğer dillere uyarlanması ile ilgili çeşitli çalışmalar ortaya konmuştur. Örneğin, Das ve Bandyopadhyay (2010) üç Hint dili (Bengalce, Hintçe ve Telugu) için SentiWordNet'i, Vu ve Park (2014) SentiWordNet'in Vietnam dilindeki

sürümünü ve Dehkhargani ve diğerleri, (2015) ise Türk dili üzerine SentiTürkNet adlı çalışmayı yapmıştır.

2.3.5.2.1.1.6. NRC

S. M. Mohammad, 2009'dan bu yana; Twitter, Yelp ve Amazon müşteri incelemeleri de dahil olmak üzere farklı kurumları ele alan birçok kelime duyarlılığı sözlüğü derlemiştir. Etiketlemenin manuel olarak yapıldığı bazı durumlar dışında, duygu etiketlemesi olarak tweet'in hashtag' lerinin kullanılması şeklinde otomatik işlemler de yapılmaktadır (Svetlana Kiritchenko ve Mohammad, 2014; Mohammad ve diğerleri ,2013; Zhu ve diğerleri, 2014 ve Kiritchenko ve diğerleri, 2014).

2.3.5.2.1.2. Veri Kaynakları

Duygu sınıflandırması da dahil olmak üzere diğer metin sınıflandırma alanlarıyla karşılaştırıldığında, kutupluluk sınıflandırması daha fazla sayıda açıklanmış veri kümesinden yararlanmaktadır. Tablo 2.3'te de gösterildiği üzere literatürde sıkça kullanılan çeşitli veri setleri bulunmaktadır.

Tablo 2.3. Polarite Temelli Veri Kaynakları

Kaynak	Yazar	Yıl	Boyut	Veri Tipi
Movie	B. Pang	2004	29.419	yapılandırılmış IMDB film yorumları (belge düzeyinde)
		2004	2.000	ham/etiketsiz IMDB film yorumları (belde düzeyinde)
		2005	10.662	yapılandırılmış film yorumları (cümle düzeyinde)
Amazon	J. Blitzer	2007	8000	ürün yorumları
Blogs	P. Melville	2009	252	ürün yorumları ve politik gönderiler

2.3.5.2.1.2.1. Amazon

Blitzer ve diğerlerinin (2007) ortaya koyduğu çalışmaların bir sonucudur. Amazon web sitesinden elde edilen ürün incelemelerinden oluşan bir veri kümesidir. Her biri 2.000 inceleme içeren dört farklı (DVD, kitap, elektronik ve mutfak) inceleme kategorisi içermektedir. Her bir alana ilişkin incelemeler yarı pozitif ve yarı negatiftir. Dolayısıyla bu durum veri kümesini daha dengeli hale getirmektedir. Bu veri kümesindeki her bir örnek, 0 ile 5 yıldız arasında değişen derecelendirme, inceleme başlığı ve inceleme tarihi gibi kategorilerden oluşan bir incelemeye ilişkin ayrıntılı bilgileri içermektedir. Araştırmacılar, Amazon web sitesindeki verileri

taradıktan sonra; yorumları, derecelendirmeleri 3 yıldızdan yüksek puanları olumlu ve 3 yıldızdan düşük olanlar olumsuz olacak şekilde yorumlamışlardır. Etiketli verilere ek olarak; bu veri kümesi DVD alanında 3,685 etiketlenmemiş örnek; mutfak alanında ise 5.945 etiketlenmemiş örnek içermektedir.

2.3.5.2.1.2. Film Veri Kümeleri

Bu veri kümelerinin çeşitli sürümleri, Pang ve diğerleri (2008) tarafından oluşturulan ünlü çevrimiçi film veritabanlarındaki film incelemelerinden elde edilmiştir.

HTML dosyaları havuzu: Bu veriler, işlenmemiş ve etiketlenmemiş 27.886 HTML dosyasından oluşmaktadır. Dosyalar, "İnternet Film Veritabanı" (IMDB) adlı çevrimiçi bir veritabanından elde edilen incelemelerden oluşmaktadır. Bu havuz, polarite veri kümesi adındaki etiketli kümenin ham sürümüdür.

Polarite veri kümesi: Verinin bu sürümü dört farklı alt işlemi içerir. 0.9. ve 1.0 şeklindeki alt versiyonları 700 olumlu ve 700 olumsuz yapılandırılmış incelemeden oluşmaktadır (Pang ve diğerleri; 2002). 1.1 adlı alt sürümü, İngilizce olmayan ya da eksik olan birkaç incelemenin kaldırılması ve bazı yanlış etiketlenmiş incelemelerin düzeltilmesiyle yeniden oluşturulmuştur. Son olarak, sürüm 2, her bir kutup sınıfı için 1000 incelemeden oluşmaktadır (Pang ve Lee 2004). Ham sürümdeki tüm incelemeler aynı derecelendirme biçimine sahip olmadığından, derecelendirme biçimine bağlı olarak bunların etiketlenmesi işlemi farklı şekilde yapılmıştır. İlk olarak, yalnızca yazarı derecelendirmeyi açıkça beyan eden incelemeler sınıflandırılmıştır. 5 yıldızlı bir sistemde; 3.5 yıldızlı ve üzeri olan yorumlar olumlu olarak etiketlenmiş ve 2.5 yıldız ve altı olan yorumlar negatif olarak etiketlenmiştir. Dört yıldızlı bir sistemde ise, 3 yıldız ve üstü yorumlar pozitif olarak, 1.5 yıldız ve altı yorumlar ise negatif olarak etiketlenmiştir. Son olarak, bir harf notu sistemi ile belirlenmiş değerlendirmelerde, B ve üstü pozitif, C ve altı negatif kabul edilmiştir.

Cümle polaritesi veri kümesi: Verilerin bu versiyonu, Pang ve Lee (2005) tarafından hazırlanan 5,331 olumlu ve 5,331 olumsuz işlenmiş cümle veya cümle parçalarını içermektedir. Tüm örneklerle ilişkin incelemeler, yeni (olumlu anlamında) veya eski (olumsuz anlamında) olarak sınıflandıran "rottentomatoes" adlı bir film inceleme veri tabanından indirilmiştir.

2.3.5.2.1.2.3. Bloglar

Melville (2009) ve diğerleri, tarafından hazırlanan bu veri kümesi, biri teknoloji bloglarıyla, diğeri ise siyasi bloglarla ilgili olmak üzere iki farklı blog yazı grubunu içermektedir. “Lotus blogları” olarak adlandırılan ilk set, 14 blogdan oluşan ve IBM Lotus yazılımına karşılık gelen bir dize veri setidir. Veriler, her blogger’ın Zengin Site Özeti (RSS) eklemesi ve en son gönderilerin ya da blogun arşivlenmiş gönderilerinin indirilmesiyle sağlanmıştır. Daha sonra, etiketlerin kelimeler ilgili HTML dosyalarındaki metinlerden çıkarılmıştır. Daha sonra tüm gönderiler manuel olarak okunmuş ve olumlu, olumsuz, tarafsız veya alakasız olacak şekilde etiketlenmiştir. Bu sette 34 pozitif ve 111 negatif örnek vardır. Bu veri setinin ikinci kısmı, 2008'deki Amerika Birleşik Devletleri başkanlık seçimlerinin iki adayı olan "Barak Obama" ve "Hillary Clinton" ile ilgili siyasi paylaşımlardan oluşmaktadır. Gönderiler 16.741 blog verisinden alınmış; “Obama” ve “Clinton” kelimelerinin bulunup bulunmadığına göre filtrelenmiş ve manuel etiketleme için rastgele şekilde seçilmiştir. Melville ve diğerleri (2009) göre politik gönderileri etiketlemek, ürün incelemelerini etiketlemekten çok daha zordur, çünkü politik gönderiler daha duygusaldır ve çoğunlukla aday hakkında üstü kapalı yorumlardan ya da yargılardan bahseder. Bu nedenle, iki adaydan biri hakkında açıkça bir görüş bildiren gönderiler olumlu veya olumsuz olarak etiketlenmiştir. Dolayısıyla bu sette tarafsız veya alakasız gönderi bulunmamaktadır. Ayrıca politik veri kümesi 49 olumlu ve 58 olumsuz gönderi içermektedir.

2.3.5.2.2. Duygu Temelli sözlükler

2.3.5.2.2.1. Sözlükler

Neredeyse tüm duygu madenciliği çalışmaları bir sözlüğe dayanmaktadır. Bu tür sözlükler, her kelime veya kelime grubunun taşıdığı duygu türü/gücü hakkında daha detaylı bir bilgi verdikleri için oldukça kullanışlıdır.

Bu doğrultuda Tablo 2.4’de duygu madenciliği görevi için yararlı olan bazı sözlüklere yer verilmiştir.

Tablo 2.4. Duygu Temelli Sözlükler

Sözlük	Yazar	Yıl	Sözcük Boyutu	Duygu Kutupları
Wordnet Affect	C. Strapparava	2004	4.787	çeşitli duygu kategorileri
LIWC	J. W. Pennebaker	2007	5.000	duygusal/duygusal olmayan; olumlu/ olumsuz; kaygı, öfke, üzüntü
NRC	S. M. Mohammad	2010	14.182	öfke, korku, beklenti, güven, şaşkınlık, üzüntü, sevinç, iğrenme
NRC hashtag	S. M. Mohammad	2013	32.400	öfke, korku, beklenti, güven, şaşkınlık, üzüntü, sevinç, iğrenme
CBET	A. Gholipour Shahraki	2015	24.000	öfke, korku, sevinç, aşk, üzüntü, şaşkınlık, minnettarlık, iğrenme, suçluluk

2.3.5.2.2.1.1. Wordnet Affect

Wordnet Affect, “synsets” olarak adlandırılan eşanlamlı kelimelerin bir listesini içeren duygu temelli bir sözlük kaynağıdır. Bu sözlükteki duygu kümesi, hiyerarşik olarak düzenlenmiştir. Strapparava ve Valitutti (2004), bu sözlüğü önceki sözlükleri Wordnet’ in üzerine inşa etmişlerdir. El ile tür ayrımını dikkate alarak (isim, sıfat, zarf) 1.903 duygu kelimesinden oluşan bir başlangıç seti oluşturmuşlardır. Daha sonra, bu duygu kelimelerinden en az birini içeren bir “Wordnet synset “alt kümesi seçilmiş ve geri kalanı reddedilmiştir. Bu durum, sözlüğün özünü oluşturmaktadır. Daha sonra, bu çekirdek sözlüğün sentezleri ile Wordnet’in diğer sentezleri arasındaki sözcüksel ve anlamsal ilişkiler, bu çekirdek sentezler tarafından temsil edilen duyuşsal anlamı koruyup korumadıkları bakımından incelenmiştir. Yeni “synsets” ekledikten sonra, Wordnet Affect 2.874 synset ve 4.787 kelime içermiştir.

2.3.5.2.2.1.2. LIWC

Dilbilimsel Sorgulama ve Kelime Sayımı (LIWC), Pennebaker ve diğerleri (2007) tarafından geliştirilen sözlüktür. Bu sözlüğün oluşturulmasının ilk adımında, psikolojik bir süreç içerisinde öncelikle başlangıç kategori ölçekleri üretilmiş ve ardından beyin fırtınası seansları ile başlangıç listelerine çeşitli ölçekler eklenmiştir. Bir sonraki adımda, üç bağımsız karar verici, kelimeleri iki aşamada derecelendirmiştir. İlk LIWC değerlendirmesi 1992’de gerçekleştirilmiş ve o zamandan beri sürekli güncellenerek büyük ölçüde genişletilmiştir.

2.3.5.2.2.1.3.NRC

Mohammad ve Turney (2010) tarafından geliştirilmiştir. Bireylerden belirli olmayan alanlardan oluşan kelimelere uyandırdıkları duyguya göre açıklama eklemeleri istenmiştir. Bu süreçteki önemli bir zorluk, farklı anlamlardaki kelimelerin farklı duygular uyandırdığı durumlarda ortaya çıkabilecek yanlış açıklamalardır. Bu sorununun çözülmesi için, hedef duyusunun açıklayıcılara iletilmesi gerekmiştir. Bu nedenle, bireylerden, açıklayanın hedef terime aşına olmayabileceği durumları belirlemeye yardımcı olan kelime seçimi soruları da dahil olmak üzere ek sorular sorulmuştur. Bir sözlük oluşturmanın yanı sıra, düzenli bir kalabalığın uygun yönergeler göz önüne alındığında güvenilir duygu ek açıklamaları üretebileceği sonucuna varılmıştır.

2.3.5.2.2.1.4. NRC Hashtag

NRC sözlüğünün ana yazarı S. M. Mohammad, "NRC hashtag duygu sözlüğü" olarak adlandırılan yeni bir sözlük geliştirmiştir. 21.000 tweet içeren bir külliyat üzerinde, çeşitli istatistiksel hesaplamalar kullanılarak oluşturulmuştur.

2.3.5.2.2.1.5. Açık Dengeli Duygusal Tweetler (CBET)

Bu sözlük, Gholipour Shahraki (2015) tarafından aynı ada sahip bir veri kümesinin tek etiketli bölümünden derlenmiştir. Bu veri kümesi, her biri tek bir duygu ile etiketlenmiş, çok sayıda tweet içermektedir. Sözlük aslında bir $V \times E$ boyutlu bir matrisdir ve burada V, CBET veri kümesinde bulunan tüm tek birimli kelimelerin (unigram) kümesidir. E ise duygu kümesidir. CBET sözlüğü en yeni duygu sözlüklerindendir, halka açıktır ve öncekilere kıyasla daha fazla duyguyu kapsamaktadır.

2.3.5.2.2.2. Veri Kaynakları

Çoğu makine öğrenmesi çalışmasındaki en önemli zorluk; verilerin, özellikle de etiketli verilerin, elde edilmesi işlemidir. Manuel etiketlemelerin dışında; belirli duygu açıklama problemlerinde sonuçların niteliği, genellikle yanlış anlamalara, açıklayıcıların öznel yorumlarına, kişiliklerine, içeriğin analiz edildiği bakış açısına bağlıdır (Alm 2008).

Bu doğrultuda Tablo 2.5’te, güvenilir bir etiketleme sürecine sahip olan ve yaygın olarak kullanılan yararlı birkaç veri kümesine yer verilmiştir.

Tablo 2.5. Duygu Temelli Veri Kaynakları

Kaynak	Yazar	Yıl	Boyut	Veri tipi
ISEAR	K. R. Scherer	1997	7.666	Yoğun içerikli paragraflar
Peri Masalları	C. Ovesdotter Alm	2005	15.000	Çocuk hikâyelerinden cümleler
SemEval	C. Strapparava	2007	1.250	Haber başlıkları
TEC	S. M. Mohammad	2012	21.000	Tweetler
CBET	A. Gholipour Shahraki	2015	81.163	Tweetler

2.3.5.2.2.2.1. ISEAR

Ücretsiz olarak indirilebilen ve duygu etiketli en eski veri kümelerinden biri olan ISEAR, Scherer ve Wallbott (1994) tarafından ortaya konulmuştur. Veriler, 1990'larda, ISEAR üzerinde çalışan büyük bir psikolog grubu tarafından toplanmıştır. Bu çalışmada, beş kıtadaki 37 ülkeden, psikolog olan ya da olmayan üzere 3.000 öğrenciden, sevinç, korku, öfke, üzüntü, tiksinti, utanç ve suçluluk şeklindeki şu yedi ana duyguyu kullanarak yaşadıkları durumları bildirmeleri istenmiştir. Katılımcılar, yazdıkları ilgili durumları nasıl değerlendirdiklerini ve nasıl tepki verdiklerini açıklamışlardır. İngilizce konuşmayanlar için metin İngilizceye çevrilmiştir. Dolayısıyla, veriler tam olarak tek bir ifadeyle etiketlenmiş bir cümle veya paragraftan oluşmaktadır. Bu veri kümesi, yazarların kendi metinlerine kendileri tarafından açıklama eklemeleri sebebiyle etiketleme bakımından güvenilirdir. Ancak, diğer dillerden İngilizceye çeviri yapmak duyguları değiştirebilmektedir. Şaşırtıcı bir şekilde ISEAR, 2008 yılına kadar duygu madenciliği amacıyla kullanılmamıştır.

2.3.5.2.2.2.2. Peri Masalları

Alm ve Sproat (2005) tarafından geliştirilen başka bir veri seti ise bir dizi masaldır. Bu veri seti, Beatrix Potter, Grimm Kardeşler ve Hans Christian Andersen tarafından yazılan; öfke, tiksinti, korku, mutluluk, üzüntü, olumlu şaşkınlık, olumsuz şaşkınlık ve nötr şeklindeki duygulardan biri ile etiketlenmiş yaklaşık 15.000 cümle içeren toplam 185 çocuk öyküsünü içermektedir. Etiketleme anadili İngilizce olan altı kadın tarafından manuel olarak yapılmıştır.

Metinlerin belge düzeyinde açıklandığı ISEAR veri kümesinin aksine, masal veri kümesinde cümle düzeyinde açıklama yapılmıştır.

2.3.5.2.2.2.3. *SemEval 2007*

Strapparava ve Mihalcea (2007) tarafından geliştirilmiştir. Veriler, New York Times, CNN ve BBC News gibi büyük gazetelerin haber başlıklarının yanı sıra Google Haberlerdeki başlıkları da içermektedir. Etiketleme, altı uzman değerlendirmeci tarafından manuel olarak yapılmıştır. Etiket duygu kümesi, öfke, tiksinti, korku, neşe, üzüntü ve şaşkınlık şeklindeki altı duyguyu içermektedir. Sıradan 0-1 ikili açıklama yerine, daha ayrıntılı bir etiketleme işlemi yürütülmüştür. Her duygu için bir aralık (0, 100) belirlenmiş ve açıklayıcı, her başlığı bu aralıkta değerlendirmiştir. Bu nedenle, bir başlık, her biri farklı dereceye sahip birden fazla duyguya sahip olabilmektedir.

2.3.5.2.2.2.4. *TEC*

Mohammad (2012) tarafından Twitter'daki tweetlerden yararlanılarak bir duygusal tweet külliyatı oluşturulmuştur. Araştırmacı, Ekman ve diğerleri (1972) tarafından önerilen öfke, tiksinti, korku, neşe, üzüntü ve şaşkınlık şeklindeki altı temel duyguyu hedef almıştır. Devamında Twitter arayüz uygulamasını kullanarak uygun tweetleri aramak için bu duyguları ele alan altı hashtag (örneğin, #öfke, #korku, vb.) seçilmiştir. Çok kısa, kötü bir yazım dili ile yazılmış ve başka bir tweet'in retweetleri olan tweetleri atılmıştır. Ayrıca, mesajın sonunda duygusal hashtag içermeyen tweetler de kaldırılmıştır. Bu işlem sonrasında, külliyatta 21.051 tweet'e yer verilmiştir. Külliyatın sırasıyla %7.4'ü, % 3.6'sı, % 13.4'ü, % 39.1'i, %18.2'si ve %18.3'ü; öfke, tiksinti, korku, neşe, üzüntü ve sürpriz etiketlerine sahiptir. Bu durum, veri kümesinin ne kadar dengesiz bir veri kümesi olduğunu ortaya koymaktadır.

2.3.5.2.2.2.5. *CBET*

2015 yılında Gholipour Shahraki (2015), öfke, korku, neşe, aşk, üzüntü, sürpriz, minnettarlık, iğrenme ve suçluluk şeklindeki dokuz duygudan en az birine sahip tweetleri aramak için Twitter üzerindeki hashtag'leri kullanarak, Temizlenmiş Dengeli Duygusal Tweetler (CBET) veri setini derlemiştir. CBET' in iki bölümü vardır: Büyük bölüm, tek etiketli örnekler olarak adlandırılan, tek bir etikete sahip tweet'leri içermektedir. Bu bölüm, her duygu için 8.540 olmak üzere 76.860 etiketli tweet içermektedir ve mükemmel bir şekilde dengelidir.

Küçük bölüm ise çift etiketli tweetler, yani aynı anda iki duyguyu ifade eden tweetleri içermektedir. Bu bölümün boyutu 4.303'tür ve dengesiz olduğu açıkça görülmektedir. Çünkü tüm duygu kombinasyonları eşit sıklıkta gerçekleşmemiştir. En sık kullanılan eşleştirilmiş etiket sevinç-sevgidir. Öfke-minnettarlık gibi duygu çiftleri ise çok nadirdir. Bu veri kümesindeki toplam tweet sayısı (81.163), veri kümesini duygu madenciliği araştırmaları için mevcut en büyük külliyat haline getirmektedir.

2.3.6. Duygu Analizi ile İlgili Çalışmalar

2.3.6.1. Eğitimde Duygu Analizi Uygulamalarının Hedefleri ve Yurt Dışında-Eğitim Alanında Yapılan Duygu Analizi Çalışmaları

Tezin bu kısmında yurtdışı temelli alan yazında eğitimde duygu analizi çalışmaları, belirli bir duygu analizi yöntemi gözetmeksizin araştırma hedefi ve ulaşılan sonuçlar ele alınarak incelenmiştir.

2.3.6.1.1. Duygu Analizi Yöntemlerinin veya Sistemlerinin Uygulanması

2.3.6.1.1.1. Sistem Geliştirme

DA, öğrenci geri bildirimlerinin gerçek zamanlı analizini gerçekleştirmeye katkıda bulunan çeşitli sistemlere (yönetim sistemleri, çevrimiçi öğrenme sistemleri, değerlendirme sistemleri gibi) dahil edilebilmektedir (Elia ve diğerleri, 2019; Lin ve diğerleri, 2019; Oramas Bustillos ve diğerleri, 2019; Rani ve Kumar, 2017; Tseng ve diğerleri, 2018). Sistem geliştirmede DA, arayüz tasarımına ve öğrenme sürecine daha iyi odaklanılmasına yardımcı olmaktadır (Lin ve diğerleri, 2014). Bu doğrultuda, DA öğretim etkinliğini artırılması bakımından büyük bir potansiyele sahiptir. Ayrıca DA aracılığıyla elde edilen potansiyel bilgiler eğitim ve sistem uygulayıcıları tarafından platform yapımının iyileştirilmesi konusunda sıklıkla kullanılmaktadır (Rani ve Kumar, 2017; Liu ve diğerleri, 2019a).

2.3.6.1.1.2. Öğrenme Müdahalesi

DA, öğrenme sürecindeki öğrencilerin öğretim ile ilgili memnuniyetini araştırmak amacıyla kullanılabilmektedir (Ortigosa ve diğerleri, 2014). DA sonuçları ile tespit edilen ve öğrencilere ait olumsuz duygular, öğretmenlerin ve yöneticilerin zamanında öğretim

müdahalelerini gerçekleştirmeleri konusunda önemli bir noktadır (Liu ve diğerleri, 2019a). Li (2018), öğrenciler için duygu yoğunluğu ve duygu süresine temel alan duygusal müdahale veya öğrenme süreci müdahalesi adı verilen uygulamayı kullanmıştır. Tian ve diğerleri (2014), duygu düzenleme stratejileri için bir veri tabanı oluşturmuş ve öğrenmeye müdahale etmek için vaka temelli muhakeme tekniğini uygulamıştır. Sonuçlar, bu yöntemin etkileşimli metin tabanlı uygulamalarda duygu düzenlemede olumlu bir rolü olduğunu göstermiştir.

2.3.6.1.1.3. Görsel Geri Bildirim

Görsel geri bildirim, DA sonuçlarının okunabilirliğini artırmaktadır (Pong-inwong ve Songpan, 2019). Ayrıca görsel geri bildirim, öğrencilerin kendi duygularını algılamalarına (Arguedas ve diğerleri, 2018) ve duygularını kontrol etme becerilerini geliştirmelerine (Yu ve diğerleri, 2018) yardımcı olmaktadır. Öte yandan, Görsel geri bildirim ile öğrenci memnuniyetini sezgisel olarak öğretmenlere sunarak müfredatı geliştirmeye yönelik öneri ve rehberlik sağlanabilmektedir (Cunningham ve diğerleri, 2019). Sanal araçlar, kelime bulutları ve kontrol panelleri şeklinde üç ana görsel geribildirim türü bulunmaktadır. Sanal araçlar, öğrenme sisteminin çekiciliğini artırmak için farklı duygusal durumları ifade edebilmekte ve geri bildirimde bulunabilmektedir (Lin ve diğerleri, 2014). Dolayısıyla öğrencilerin öğrenme performansı üzerinde önemli bir olumlu etkiye sahiptir (Arguedas ve diğerleri, 2018). Bazı araştırmalar, DA sonuçlarını sunmak amacıyla kullanılan kelime bulutu veya grafiksel gösterim yöntemlerinin öğretmenlerin öğrencilerin öğrenme ilgisini () ve memnuniyetini analiz etmeleri için uygun olduğunu ortaya koymaktadır (Ortigosa ve diğerleri, 2014; Rani ve Kumar, 2017; Santos ve diğerleri, 2018). Ayrıca çeşitli çalışmalarda DA sonuçlarını görüntülemek için gösterge tabloları sıklıkla kullanılmaktadır (Elia ve diğerleri, 2019; Yu ve diğerleri, 2018).

2.3.6.1.2. Öğrenme Davranışını ve Performansını Anlama

Duygu, davranış, performans ve başarı arasındaki ilişkiyi inceleyen eğitimde duygu analizi araştırmaları, araştırma hedefleri ve sonuçları ele alınarak incelenmiştir.

2.3.6.1.2.1. Davranış ve Duygu Etkileşimi

Öğrencilerin etkileşimli davranışları, çevrimiçi tartışmalarda farklı öğrenme duygularını tetiklemekte ve dolayısıyla öğrenciler grup odaklı öğrenme görevlerinde daha fazla duygu yoğunluğuna sahip ve daha etkileşimli davranışlar sergilemektedirler (Huang ve

diğerleri, 2019). Örneğin MOOC forumunda, öğrenciler öğretmenlerin bu gibi analiz eylemlerinden sonra daha aktif olma eğilimindedirler (Moreno-Marcos ve diğerleri, 2019).

2.3.6.1.2.2. Duygu ve Akademik Performans İlişkisi

Öğrencinin coşkusu akademik başarısı ile olumlu bir şekilde ilişkilidir (Liu ve diğerleri, 2018). Bununla birlikte, olumsuz duygular ile yüksek, orta ve düşük başarılı öğrenciler arasında anlamlı bir fark yoktur (Liu ve diğerleri, 2018). Diğer yandan ortaya konulan birçok çalışma duygusal farkındalık ve geri bildirimin öğrencilerin öğrenme performansı üzerinde olumlu bir etkisinin olduğunu ve DA kullanımının akademik başarısızlığın erken süreçte tahmini ile ilgili uygulamaların başarısını artırabileceğini ortaya koymuştur (Yu ve diğerleri, 2018; Arguedas ve diğerleri, 2018).

2.3.6.1.2.3. Duyguların Zamansal Özelliklerinin Önemi

Duyguların zamansal özelliklerini dikkate alınması, risk altındaki öğrencilerin belirlenmesine ve müdahale edilmesi açısından önemlidir (Liu ve diğerleri, 2019b). Bu nedenle, dinamik bir duygu analizi gerçekleştirilmesi için zamansal özelliklerin dikkate alınması önerilmektedir (Leong ve diğerleri, 2012). Bir dönemin başında ve sonunda duygu ifade edilmesi olasılığı daha yüksektir (Liu ve diğerleri, 2019b). Deneyim zaman içinde artmaya devam ettikçe öğrencinin duyguları değişmektedir ve bu potansiyel bilginin belirli amaçlar doğrultusunda kullanılması önemlidir (Hixson, 2020). Öte yandan, bir çalışma çevrimiçi öğrenme sürecinde öğrenme duygularının periyodik bir gösterge sunduğunu ortaya koymaktadır (Huang ve diğerleri, 2019).

2.3.6.1.3. Öğretme ve Öğrenme Sürecini İyileştirme

2.3.6.1.3.1. Kurs Memnuniyeti

DA teknikleri, kursların niteliğini iyileştirmeyi kolaylaştırmak ve kurs ile ilgili değerlendirmelerin genel izlenimini ortaya koymak için kullanılmaktadır (Leong ve diğerleri, 2012; Cunningham-Nelson ve diğerleri, 2019; Nimala ve Jebakumar, 2019). Aynı zamanda ilgili araştırmalar, öğrencinin; kurs eğitmeni, eğitim içeriği ve eğitim programı ile ilgili değerlendirmelerinin öğrenci memnuniyetinin keşfedilmesi konusunda önemli derecede

öngörücü etkiye sahip olduğunu ortaya koymaktadır. Ayrıca öğrenciler, olumlu duygular içeren ders içeriğinin temasına daha fazla önem vermişlerdir (Liu ve diğerleri, 2019a).

2.3.6.1.3.2. Platformdaki Deneyim

DA yazılımı, eğitim içeriğini ya da platformunu ilk kez veya acemi olarak kullanma tutumunu anlamak ve öğrencilerin platformdaki öğrenme deneyimlerini keşfetmek amacıyla uygulanmaktadır. Örneğin, bu doğrultuda yapılan çeşitli çalışmalarda öğrenciler çevrimiçi bir öğrenme platformundaki öğrenme deneyimini olumlu olarak değerlendirmişlerdir (Suwal ve Singh, 2018; Hixson, 2020).

2.3.6.1.3.3. Kuruma Karşı Tutumu Tespit Etme

Santos ve diğerleri (2018), yüksek öğretim kurumlarının geçim masraflarını finansal olarak desteklemeleri, İngilizce kursları vermeleri ve uluslararası bir ortamı teşvik etmeleri halinde çevrimiçi olarak daha çekici hale gelebileceğini ortaya koymuştur. DA kullanılan bir diğer araştırma da üniversite seçimini etkileyen ana faktörlerin eğitim fırsatları, fiziksel yapı, iş fırsatları, itibar, iletişim, organizasyonel ve çevresel sürdürülebilirlik olduğunu ileri sürmüştür (Troisi ve diğerleri, 2018).

2.3.6.1.3.4. Öğretmen Performansının Değerlendirilmesi

DA kullanımı ile nitelikli öğretmenleri belirlemek amacıyla öğretmenlerin öğretim performansı değerlendirilebilmektedir (Nimala ve Jebakumar, 2019; Pong-inwong ve Songpan, 2019). Bu doğrultuda Sindhu ve diğerleri (2019), öğretmenlerin performansının pedagoji, bilgi, değerlendirme, deneyim, davranış ve genel durum şeklindeki altı boyutta değerlendirildiği bir çalışma ortaya koymuştur.

2.3.6.2. Yurtdışında Hikâye ve Roman Türündeki Metinler Üzerinde Gerçekleştirilen Duygu Analizi Çalışmaları

İnsan beyni, duyuşsal ve bilişsel işlevlerin ayrılmaz bir şekilde birbiriyle bütünleştiği karmaşık bir sistemdir. Bu sistem eğitim bakımından ele alındığında, öğrenme sürecinin içeriğiyle ilgilenen biliş ve öğrenme işlemi için gerekli olan enerjiyi sağlayan duygu arasındaki etkileşimin, öğrenme sürecine ilişkin her aşamada önem arz ettiği temel bir gerçektir. Bu doğrultuda öğrenme ve öğretme süreci hakkındaki faydalı ve potansiyel bilginin, aynı şekilde faydalı amaçlar doğrultusunda kullanılmak üzere keşfedilmesi, ortaya konması gerekmektedir. Teknolojinin gelişmesiyle birlikte bu hedeflerin gerçekleştirilmesini sağlayan çeşitli uygulamalar ve uygulama alanları ortaya çıkmıştır. Alandaki genel tanımıyla, amaçları metin analizi olarak bilinen metin madenciliği uygulamaları içerisinde, öğrenme ve öğretme sürecindeki duygusal izlerin ortaya konulması bakımından en ilginç uygulama, duygu analizi uygulamasıdır. Duygu analizi uygulaması, temelde bir konu hakkında gönderici (yazar, öğretici) tarafından alıcıya (okuyucu, öğrenci) gönderilen ya da geribildirim olarak alıcı konumundaki taraftan gönderici konumundaki tarafa gönderilen iletideki(metin) duygusal izleri ya da kalıpları ortaya koymaktır. Dolayısıyla bu uygulama, eğitim alanındaki çeşitli çalışmalarda, öğrenme ve öğretme süreci içerisinde etkin rol oynayan (öğretici-yazar, öğrenen-okuyucu) taraflar açısından, belirli amaçlar doğrultusunda sıklıkla kullanılmaktadır. Örneğin, Rajput, Haider ve Ghani (2016), eğitim kurumunun ve dahilindeki eğitimcilerin performansını değerlendirmek adına öğrenci geri bildirimlerini içeren metinleri, sözlük tabanlı duygu analizi yöntemi ile incelemiştir. Aung ve Myo (2017) da çalışmalarında, öğrenci yorumlarını öğretimin kalitesini değerlendirmek, öğrenci tutumlarını tespit etmek gibi benzer amaçlarla sözlük tabanlı duygu analizi yöntemini kullanarak incelemiştir. Eğitim alanında farklı bir uygulama olarak Sims ve Bokhave (2020) ise 17.000 belge boyutuna sahip okul teftiş raporlarını benzer şekilde sözlük tabanlı duygu analizi ile incelenmiştir. Bu çalışmalarda kullanılan metin verileri; genel olarak bir konu, kurum veya görüş hakkında çok sayıda birey tarafından ortaya konulan yorum, tweet, e-posta gibi çeşitli kısa metin türündeki duygu-görüş bildiren iletilerden oluşmaktadır. Temel bir varsayıma göre bu tür kısa metinler sabit bir duygu değerine sahiptir. Örneğin bir görüş, polarite bakımından pozitif, negatif ve nötr şeklindeki duygu kutuplarından sadece birini içermektedir. Duygu yönü bakımında ele alındığında ise ya öfkeli ya da mutludur. Ancak roman ve hikâye türündeki büyük boyutlu/uzun metinlerde genel duygu durumunun tespit

edilmesi haricinde duygu durumundaki zamansal/çizgisel değişiminin incelenmesi de mümkündür. Bu doğrultuda;

Jockers (2014)'ın metinlerde duygu analizi uygulaması için geliştirdiği, “syuzhet” adlı R paketi yardımıyla; ünlü “Moby Dick” adlı romanı çeşitli açılardan incelediği duygu analizi çalışması temel bir örnek olarak verilmektedir. Büyük boyutlu metin türlerinde doğrudan tespit için sözlük tabanlı duygu analizi yönteminin uygulanabilecek en uygun duygu analizi yöntemi olduğunu öne sürdüğü bu çalışmasında; öncelikle romanı bölümlere ayırmıştır. Bölümler içerisinde sıkça kullanılan duygu kelimelerini ve en sık geçen karakter isimlerini baz alarak, bölümlerdeki duygu değişimlerini ve karakter etrafında şekillenen duyguları ortaya koymuştur. Çalışmasında elde ettiği sonuçlara göre eserin duygu değişimi grafiğindeki inişli-çıkışlı değişimler, eserin olay örgüsü hakkında bilgi vermektedir. Belirli duygular etrafında gelişen bu değişim aynı zamanda eserin durağan ya da sürükleyici olma durumu hakkında bilgi vermektedir.

Elkins ve John (2019), Jockers'ın geliştirdiği paket (syuzhet) yardımıyla Virginia Wolf'a ait “Deniz Feneri” adlı eseri duygu analizi ile incelemiştir. Çalışmada uyguladığı analiz neticesinde romandaki olayların, klasik bir şekilde tek bir kahraman üzerinde gelişmediği ve romanın karakterler arasında dengeli bir biçimde dağılmış duygu yapısına sahip olduğu yargısına varmıştır. Ayrıca konusu açıkça belirtilmeyen bir eserin konusunun, duygu analizi ile kolaylıkla tespit edilebileceği sonucuna varmıştır.

Patel (2018), J.K. Rowling tarafından yazılan “Harry Potter” serisini; NRC, BİNG ve AFİNN sözlüklerini kullanarak sözlük tabanlı duygu analizi ile incelemiştir. Çalışma sonucunda seri içerisinde en sık geçen kelimeyi tespit etmiş ve eserlerin genel duygu durumunu ortaya koymuştur.

Jacobs (2019) da J.K. Rowling tarafından yazılan “Harry Potter” serisindeki metin bölümlerini “SentiArt” adlı duygu analizi aracı ile incelemiştir. Çalışmasında genel olarak serideki ana karakterlerin içerisinde bulunduğu bölümlerin duygu yönlerini tespit etmiş ve ilgili karakterleri çeşitli duygusal kişilik profillerine göre sınıflandırmıştır. Örneğin “Voldemort” karakterinin sıklıkla yer aldığı bölümlerin genel duygu durumu negatif olarak belirlemiştir.

Büyük boyutlu metinler üzerinde gerçekleştirilen zamansal duygu analizi çalışmalarını yönlendiren, daha eski bir metin analizi çalışmalarına örnek olarak Alm ve Sproat (2005)'un

“Masallarda duygusal sıralama ve gelişim” adlı çalışması verilebilmektedir. Alm ve Sproat, bu çalışmada 22 Grimm Kardeşler masalını, çocuk edebiyatına uygun metinler olup olmadığı yönünde incelemiştir.

Geliştirdiği duygu sözlükleri ile metinlerde duygu analizi alanındaki varlığını etkin bir şekilde ortaya koyan Mohammad (2011) ise bir e-posta koleksiyonunu ve “Google Kitaplar” içerisinde yer alan eserleri sözlük tabanlı duygu analizi ile incelemiştir. Kitaplar bazında ulaştığı sonuca göre hikâyelerin duygu yoğunlukları romanlardan daha yüksektir.

Reagan (2017) tarafından ortaya konulan doktora tezi çalışmasında ise tezin çeşitli amaçlarından sadece biri olan hikâye türündeki metinlerde duygusal değişimin tespit edilmesi noktasında, “Gutenberg Project” içerisinden elde edilen 4,803 hikâye üzerinde duygu analizi uygulaması gerçekleştirilmiştir.

2.3.6.3. Türkçe Metinlerde Duygu Analizi Çalışmaları

Türkçe metinler üzerinde gerçekleştirilen duygu analizi çalışmalarının belirlenmesi için, “duygu analizi”, “eğitimde metin madenciliği”, “eğitimde duygu analizi”, “sözlük tabanlı duygu analizi” ve “Türkçe metinlerde duygu analizi” şeklindeki anahtar kelime grupları temel alınarak, belirli bir araştırma alanı ve duygu sınıflandırma yaklaşımı gözetmeksizin yapılan geniş kapsamlı alan yazın incelemesi neticesinde Tablo 2.6.’da yer almak üzere 42 tez çalışması elde edilmiştir.

Tablo 2.6. *Türkçe Metinlerde Duygu Analizi ile İlgili Çalışmalar*

Yazar	Yıl	Amaç	Sınıflandırma yaklaşımı	Duygu Polaritesi	Veri Kaynağı/Seti	Analiz Aracı
Albayrak,N	2011	DT	MÖ	4 ruhsal durum (aksiyete,depresyon)	Psikolojik Yazılar	WEKA
Boynukalın	2012	DT	MÖ	Çoşku, korku, öfke, üzgünlük	ISEAR, Türk masalları	WEKA
Vural ve diğerleri	2012	DT	ST	P/N	Film yorumları	
Sevindi	2013	DS yöntemlerinin karşılaştırma	MÖ;ST	P/N	Twitter	
Çelik	2013	DT	MÖ	P/N	Kitap,film ve alışveriş yorumları	WEKA
Demirci	2014	DT(emotion)	MÖ	Öfke, Üzgünlük, Korku, İğrenme, Şaşırma, Coşku	Twitter	WEKA

Yazar	Yıl	Amaç	Sınıflandırma yaklaşımı	Duygu Polaritesi	Veri Kaynağı/Seti	Analiz Aracı
Uçan	2014	Otomatik duygu sözlüğü çevirme	MÖ;ST	P/N	Film ve otel yorumları	WEKA
Akba	2014	Öznitelik seçme	MÖ	P/N;P/Nötr	Film yorumları	WEKA
Türkmenoğlu	2015	DT ve yöntem karşılaştırma	MÖ, ST	P/N	Twitter gönderileri, Film yorumları	WEKA
Yurt	2015	DT	MÖ	P/N	Film yorumları	WEKA, Preto
Dekhargani	2015	Duygu sözlüğü üretimi	ST, ML	NLP		WEKA
Pak	2015	Alan bazında DT	MÖ	P/N	Ürün yorumları	
Kaplan	2016	DT	MÖ+	Mutluluk, kızgınlık, üzüntü, şaşkınlık	Twitter	KNİME
Yelmen	2016	DY/Öznitelik Seçimi	MÖ	P/N	Twitter(GSM)	
Kama	2016	DT	ÖT	P/N	Forum gönderileri	NLP Tool
Evirgen	2016	DT	MÖ	P/N(5 kategori)	Twitter	R
Önal	2017	DT(android tabanlı araç geliştirme)	ST	P/N/Nötr	Müşteri yorumları	Java
Amanet	2017	DT/DS	MÖ	Plutnik'in 10 duygusu	Twitter (gsm)	Rapid Miner
Özdeş	2017	DT ve Yöntem karşılaştırma	MÖ	P/N	İngilizce şarkı sözü	R
Çetin	2017	DT	HT	NLP, CRF	Sosyal medya	Python
Gümüş	2017	DT	MÖ	P/N/Nötr	Twitter	Java
Değer	2017	DT	MÖ	P/N/Nötr	Twitter (şirket tweetler)	Rapid Miner (VMAK)
Toçoğlu	2018	DT ve YK	MÖ	Mutluluk, korku, öfke üzüntü, tikslenme şaşırma	TREMO veri seti	WEKA
Albayrak, A	2018	DT ve mö/yk	MÖ	P/N	İngilizce ve Türkçe film yorumları	
Karaöz	2018	DT	ST	P/N	Twitter	R
Tan	2018	DT	ST	P/N;P/N;Nötr	Foursquare youmları	WEKA
Kızılkaya	2018	DT	MÖ, ST	P/N	Twitter (Siyasi parti oyları hakkındaki tweetler)	R
Ergüllü	2018	DT	ML, ST	P/N	Twitter	Python
Yurtöz	2019	DT(emoji)	MÖ	P/N sinir, mutluluk, aşk, üzüntü	Twitter	
Sağlam	2019	Otomatik Duygu Sözlüğü geliştirme ve DT	DDİ	P/N, Nötr	Haber metinleri	
Akbudak	2019	DS/DT	MÖ	A, N, L, F	Video, ses, müzik verileri	Python
Hasanlı	2019	DT/DS	MÖ;ST	P/N	Twitter(Azerbaycan Türkçesi)	
Yıldırım	2019	DT	ST	P/N(puanlı)	Twitter	Python

Yazar	Yıl	Amaç	Sınıflandırma yaklaşımı	Duygu Polaritesi	Veri Kaynağı/Seti	Analiz Aracı
Shehu	2019	DT	ML ve ST(daha iyi bulmuş)	P/N,Nötr	Twitter	R
Akısoğlu	2019	DT	MÖ	P/N/Nötr	Twitter	
Sheeha	2019	DT(emoji) ve Sözlük karşılaştırma	ST	P/N (?)	Twitter	R
Karamanlı	2019	DT	MÖ	P/N,Nötr	Müşteri yorumları	R
Abdoulaye	2019	DT	ST	P/N	Müşteri yorumları	R
Özkul	2019	DT		P/N	Twitter	Rapid Miner
İbrahim	2019	DT	MÖ	P/N	Sosyal medya	Matlab
Giraz	2020	DT ve Duygu sözlüğü oluşturma	MÖ	Mutluluk, korku, öfke, keder, tikslenme, şaşırma	Tremo, wikipedia	WEKA
Osmanoğlu ve diğerleri	2020	DT ve öğrenci hisleri	MÖ	iyi, süper, faydalı, güzeli başarılı	LMS(uzaktan eğitim sistemi)	

Tabloda görüldüğü üzere kronolojik olarak listelenen çalışmalar; amaçlarına (DT, DS, YK, ve diğerleri), kullanılan duygu sınıflandırma yaklaşımlarına (MÖ, ST), kelimelerin duygu kutuplarına yönelik yapılan kategorilere (P/N,nötr ve çeşitli duygu durumları), araştırmada kullanılan veri kaynağına(sosyal medya gönderileri, film yorumları vb.) ve kullanılan analiz aracına göre sınıflandırılmıştır.

Tablodaki çalışmalar ortaya konulduğu yıl açısından incelendiğinde, 2019 yılında diğer yıllara göre nispeten belirgin bir artış görülmektedir.

Amaçları bakımından incelendiğinde, çalışmaların genel olarak kullanılan veri seti üzerinde duygu tespiti amacı taşıdığı görülmektedir. Amacı, duygu sözlüğü oluşturma, duygu sözlüğü karşılaştırma gibi duygu sözlüğü ile doğrudan ilişkili olan beş çalışma bulunmaktadır (Uçan, 2014; Dekhargani, 2015; Sağlam, 2019; Sheeha, 2019; Giraz, 2020).

Kullanılan duygu sınıflandırma yaklaşımına göre incelendiğinde, her iki yaklaşımın birlikte kullanıldığı çalışma sayısının sadece bir yaklaşım kullanılan çalışma sayısından daha az olduğu görülmektedir. Aynı zamanda bu tez çalışmasının amacını oluşturmasıyla birlikte, içerisinde esas olarak sözlük tabanlı yaklaşımı barındıran yedi çalışma bulunmaktadır (Vural, 2012; Önal, 2017; Karaöz ve Akın, 2018; Tan, 2018; Yıldırım, 2019; Sheeha, 2019; Abdoulaye, 2019).

Çalışmalar kelimelerin sınıflandırıldığı duygu kutbu kategorilerine göre incelendiğinde, çalışmalarda daha az ayrıntı düzeyinde olan, polarite temelli (P/N) kategorilerin daha sık kullanıldığı görülmektedir.

Çalışmalar kullanılan veri kaynağına ya da veri setine göre incelendiğinde, çalışmalarda genellikle film yorumları, sosyal medya gönderileri gibi bir ürün veya konu hakkında çok sayıda kullanıcı ya da birey tarafından ortaya konan verilerin kullanıldığı görülmektedir. Eğitim alanıyla ilişkili olduğu görülen çalışmalar ise sadece iki tane olmak üzere, Türk masalları (Boynukalın, 2012) ve uzaktan eğitim sistemi verileri (Osmanoğlu ve diğerleri, 2020) üzerinde uygulanmış olan çalışmalardır.

Çalışmalar kullanılan analiz aracı bakımından incelendiğinde, WEKA programının daha ağırlıklı olarak kullanıldığı görülmektedir. Aynı zamanda bu tez çalışmasının analiz aracını oluşturan R yazılımının kullanımında ise son yıllarda kısmen belirgin bir artış görülmektedir.

Tabloda verilen bilgilerin haricinde alan yazın çalışmaları, araştırma hedefleri ve konuları bakımından incelendiğinde ise genel olarak ekonomi, ticaret ve sosyoloji vb. alanlar ile ilgili olarak ortaya konulan çalışmalar oldukları görülmektedir.

2.4. R Yazılımı

R, Ross Ihaka ve Robert Gentleman tarafından oluşturulan bir istatistiksel yazılım programıdır. Temelde, 1988 yılında “AT&T Bell Laboratories” tarafından geliştirilen ve veri işleme, analiz etme, sonuç grafiği çizme gibi işlemleri gerçekleştirmek için kullanılan nesne yönelimli bir yazılım olan S-Plus’ın geliştirilmiş halidir (Ihaka ve Gentleman, 1996). R, hem bir programlama dilinin hem de bir yazılım ortamının genel adıdır. Bu yazılımda komutlar tanımlayıcı kod satırları kullanılarak yürütülmektedir (Crawley, 2007).

2.4.1. Avantajları

R, tüm işletim sistemleri ile uyumlu, ücretsiz ve açık kaynak kodludur. İlk geliştiricileri tarafından ortaya konan “özgür/bedava yazılım hareketi” doğrultusunda, yazılım; CRAN (Comprehensive R Archive Network/Geniş Kapsamlı R Arşivi Ağı) adı verilen sistem içeriğine herkesin katkıda bulunabileceği şekilde geliştirilmiştir. Bu nedenle hızlı ve sürekli gelişen bir

yapıya sahiptir. Tüm bu özelliklerinin yanı sıra asıl R'nin asıl avantajı, SPSS, SAS veya Minitab gibi diğer istatistiksel yazılımlardaki bütün istatistiksel işlemleri içerisinde barındırmasıdır (de Micheaux, Drouilhet ve Liquet, 2013). Bu işlemler, ilgili diğer yazılımlardaki gibi birkaç fare tıklaması ile gerçekleştirilemez. Bu doğrultuda öğrenme eğrisinin daha dik olduğu öne sürülebilmektedir. Kullanımı için öncelikle ilgili istatistiksel yöntemin iyice anlaşılması gerekmektedir. R, yazılımın ve programlama dilinin prensibi kavrandıktan sonra araştırmacılara, diğer yazılımların aksine kendi analiz araçlarını, tekniklerini ve yöntemlerini geliştirebilme imkânı tanımaktadır. De Micheaux ve diğerlerine (2013) göre özellikle veri işleme, istatistiksel hesaplama ve görselleştirme bağlamında güçlü bir performansa sahip olan R'nin kayda değer diğer avantajları şunlardır:

- ❖ Çok iyi entegre edilmiş bir dokümantasyon sistemine,
- ❖ Veri işleme ve depolama için etkili fonksiyonlara,
- ❖ Özellikle matrisler olmak üzere tablolar üzerinde kolaylıkla işlem yapılmasını sağlayan bir dizi operatöre,
- ❖ Veri analizine ilişkin tüm adımları kolaylıkla gerçekleştirebilecek yüzlerce pakete,
- ❖ İleri düzey görselleştirme yeteneğine,
- ❖ Koşullar, döngüler gibi girdi-çıkış prensibine sahip çeşitli fonksiyonları içerisinde barındıran basit ve etkili bir programlama diline sahiptir.

R'de birçok klasik ve modern istatistiksel teknik uygulanmaktadır. Tanımlayıcı istatistikler, hipotez testleri, varyans analizi, doğrusal regresyon vb. istatistiksel analiz için yaygın olarak kullanılan yöntemler doğrudan sistem içerisinde yer almaktadır. Bu istatistiksel yöntemler harici paketler aracılığıyla da uygulanabilmektedir. CRAN (<http://cran.r-project.org>) içerisinde gruplandırılmış olan bu paketlere doğrudan ulaşılabilir. Ek olarak, CRAN'da her paket için ayrıntılı açıklama içeren belgeler mevcuttur. Dolayısıyla yazılım ve içerisindeki programlar amaca yönelik olarak kolayca uygulanabilmektedir (Pathak, 2014).

BÖLÜM III

YÖNTEM

3.1. Araştırmanın Modeli

Araştırmada metinlerden elde edilen sözcük birimlerin duygu niteliğinin incelenmesi ve metnin geneli hakkında yargıda bulunulması bakımından betimleyici nitel bir yaklaşım benimsenmiştir. Nitel yöntemle kurgulanmış araştırmaların temelinde incelenen olay veya durum hakkında derin bir algıya ulaşma gayreti bulunmaktadır (Morgan, 1996). Nitel araştırmalarda veri toplama teknikleri bakımından genellikle görüşme, gözlem, söylev, doküman ve analizi kullanılmaktadır (Baltacı, 2019). Ayrıca çalışmada nicel araştırma yaklaşımlarının aksine kesin bir nedensellik bağı kurulmaması bakımında nitel yaklaşımın hâkim olduğu belirtilmiştir.

3.2. Araştırmanın Materyali

Araştırmanın materyalini Ömer Seyfettin'in nispeten sadeleştirilmiş hikâye türündeki 165 eseri oluşturmaktadır. İki aşamalı şekilde gerçekleştirilen araştırmanın, esas kısmında belirli kriterlere göre seçilen 51 hikâye analiz edilmiş; devamında ise ilgili analiz tüm eserler üzerinde uygulanmıştır.

Ömer Seyfettin, edebiyatımızda modern hikâyeciliğin öncüsüdür. Kendisinden sonra gelenleri etkisinde bırakan Ömer Seyfettin, çağdaşı birçok yazarın unutulmasına rağmen edebiyatımızın daima gündeminde olan önder bir yazardır. Kişiliği ve yaşadığı dönemin olaylarından esinlenerek yazdığı eserleri, özellikle çocuk edebiyatında önemli bir yere sahiptir (Enginün,1985). Günümüzde eserleri en çok okunan yazarlardan olmasının yanı sıra, en çok eleştirilen yazarların da arasındadır. Hikâyelerinin içeriği ve çocuk okuyucuya uygunluğu konusunda çeşitli tartışmalar bulunmaktadır Bu durum, yazarın eserlerinin analize dahil edilmesinin temel nedenlerinden biridir.

Araştırmada incelenen doküman, Ömer Seyfettin'in tüm hikâyelerini içeren ve en çok tercih edilen bir yayınevine ait son yayımdır.

3.3. Veri Analizi Aracı

Araştırmada dokümanlar; veri ve metin madenciliği alanlarında sıklıkla kullanılan açık kaynak kodlu R (versiyon: 4.0.5) yazılımında, bir metin madenciliği yöntemi olan sözlük tabanlı duygu analizi ile incelenmiştir.

3.4. Veri Analizi Süreci

3.4.1. Metin Verilerinin Elde Edilmesi

Ömer Seyfettin’in kısa ve uzun türdeki 165 hikâyesinin kronolojik ve tematik şekilde Nazım Hikmet Polat tarafından düzenlenerek bir araya getirildiği, Yapı Kredi yayınlarına ait 1431 sayfadan oluşan “Bütün Hikâyeleri / Ömer Seyfettin“ adlı e-kitap, pdf formatında elde edilmiştir.

3.4.2. Metin Verilerinin Analize Hazırlanması

Elde edilen “.pdf” formatındaki belgede kapak, önsöz ve kaynakça gibi analize dahil edilmeyecek olan bölümler hariç tutularak hikâyelerin tamamı kopyalanmış ve “.txt” uzantılı bir metin belgesi uygulamasına kaydedilmiştir. Analiz süresince metinlerin uğradığı değişikliklerin ortaya konulması adına, hikayelerde ilk satırı oluşturan metin parçalarına hikayelere ilişkin başlık ve yazım yılı gibi künye bilgileri ile birlikte Tablo 3.1’de yer verilmiştir.

Tablo 3.1. *Hikâyelerin Künyeleri ve Hikâyelere Ait Düzenlenmemiş Metin Parçaları*

No	Hikâye Adı	Yıl	Metin
1	İlk Namaz	1905	Oh, bu sabah ne kadar soğuktu. Yatağımın hararetlerini terk ettiğim vakit,
2	Sahire Karşı	1905	Bazen muvaffak olamamış bir sanatkar hicran ve ıstırapıyla mantıksız ve
3	Erkek Mektubu	1907	Muzaez Sermetim, İşte nihayet tevhül ettim. Bekarlığa an-ı vedaımdan
4	Pervanelerin Ölümü	1909	Dışarıda, bütün nîlî-şeffaf serairiyle mahmul ve nuşin yaz gecesi...İçerde
5	Aşk ve Ayak Parmakları	1910	Evvelâ beni sen sevdin, yalvardın, yakardın; benim aşkım adeta senin galeyan
6	Koleksiyon	1914	Tokatlıyan’ın sırasındaki yirmi beş numaralı apartmanı arıyordum. Louis Durant
7	Gayet Büyük Bir Adam	1914	Bu roman hem hakikat, hem de hayaldir. Ömer Seyfettin hakiki simaların bazı
8	Sivrisinek	1917	Bilsen Efruzcuğum, kırk gündür burada ne rahat yaşıyordum. Ses yok, seda yok,
9	Doğuran Ev	1914	Ben Kadıköyü’nde büyüdüm. O vakitler mahalle bile yoktu! Tek tük evler..
10	Çanakale’den Sonra	1917	Ona akrabaları “meraklı”, uzaktan tanıyanlar “deli” derlerdi. Yaşı kırk beşi geçi-
11	Yeni bir hediye	1917	Yemekten kalkalı belki bir saat olmuştu.Karı koca, kahvelerini, her vakitki gibi

No	Hikâye Adı	Yıl	Metin
12	Binecek Şey	1917	Derviş Hasan birdenbire durdu. Kirli, yırtık yenleriyle alının terlerini sildi.
13	Çakmak	1917	“Ulan İboş, sen be!” “Vay Mıstık, sen ha?” “Ben ya...” İki hemşeri hemen
14	Hafiften Bir Seda	1918	Mahallenin yeni yetişen gençleri Hacı İmadettin Efendi'nin yalnız ismini bilir-
15	Velinimet	1918	Geçen gün hava ne güzeldi! Logaritmacı Hasan'la Hürriyet Tepesi'ne gittik.
16	Nadan	1918	İstanbul üç gündür sis içindeydi. Topkapı Sarayı'nı, açık, kül rengi kalın bir bulut
17	Bir Hatıra	1918	Ah gençlik!.. Tıpkı ezeli bir baharın ilk çiçekli günlerine benzer. Yeşil kırlar
18	Namus	1918	Daha geceydi, Çoban yıldızı derinlerden dünya üzerindeki uyumuş sürülere
19	Yemin	1918	Ah, on beş, yirmi sene evvel... hayat ne tatlıydı! Yaşamının, eğlenmenin hoş
20	Ezeli Bir roman	1918	Âdem Bey büyük ıhlamur ağaçlarının altında yapayalnız geziniyordu. Ay, mavi,
21	Türkçe Reçete	1918	Belkıs, geniş yatağında, mavi ipek kaplı yorganının altında sıkışmış bir yumruk
22	Nişanlılar	1919	Ruhumda ilahi bir ateş olmadığı için mi, yoksa müspet ilimlerle pek gençken
23	Deve	1919	Mestan Ağa doğduğu büyüdüğü Edirne'de "Çingane" olarak yaşamağa tahammül
24	Yüz Akı	1919	Mehmet Efendi on senedir kasabada oturuyordu. Köydeki tarlaları, bağları,
25	Forsa	1919	Akdeniz'in esatir yuvası nihayetsiz ufuklarına bakan küçük tepe minimini bir
26	Türbe	1919	Sıcak bir mayıs günüydü. Küçük bahçeyi gölge içinde bırakan kocaman tek dut
27	Memlekete Mektup	1919	“Sevgili Celil, İşte bir haftadan beri her sabah sana mektup yazmak azmiyle
28	Korkunç Bir Ceza	1919	Hasan Ağa başka hemşehrileri gibi yapmamış, İstanbul'a karısıyla beraber gel-
29	Bit	1919	Biz, bu son asrın muharrirleri en ehemmiyetsiz, en adi şeylere kıymet verir,
30	Perili Köşk	1919	Sermet Bey döndü, arkasındaki bekçiye, "İşte bir boş köşk daha!" dedi. Küçük
31	Nasıl Kurtarmış	1919	Kasaba içinde Kadı Mustafa Efendi'den hazzeden kimse yoktu! Dört parmak
32	Mürebbiye	1919	Çocukluğumuzda büyük çam ağaçları altına kurulan salıncakları her türlü ders
33	Meh-ma-em-ken	1919	On beş, yirmi sene evvel ordunun manzarası ne latifti! Taburlarda mektepli
34	Tam bir görüş	1919	Geçen sene bir gün arkadaşım Sermet'le Şişli'ye doğru yürüyorduk. Birdenbire
35	Heykel	1919	Bu kocaman, harap İstanbul'un içinde yaşayan hangi sanatkar "güzelliğin"
36	Havyar	1919	Hamdune Hanım eski bir şeyhülislam haremidir. Yaşı bugün altmışı aşmak üzere..
37	Pireler	1919	Aşk filan değil..Hani şu "tesadüf" dediğimiz tarihi yapan, saadetleri ibda eden
38	Kaşağı	1919	Ahırın avlusunda oynarken aşağıdaki, gümüş söğütler altında görünmeyen dere
39	Bir Hayır	1919	Yatağında kımıldayamayan Durmuş Ağa gözlerini basık tavana dikmişti. Sanki
40	Yuf Borusu	1919	Aksaraylı Câbir Paşa'yı bir zamanlar tanımayan yoktu; zevcesi cennetmekanın
41	Horoz	1919	Yine bir gün derin bir ıstırap içinde kıvranıyordum! Annem, babam, kardeşlerim
42	Dünyanın Nizamı	1919	Bir ay geçmeden fikrim değişti! Ama öyle yavaş yavaş değil... Birdenbire! Bugün
43	Uçurumun Kenarında	1919	Dün seni gördüm. Gayr-i ihtiyari avazım çıktığı kadar " A, a, aaa.."diye haykırdım.
44	Kurbağa Dansı	1920	Taşra âlemi... Yani İstanbul'un dışında geçen hayat, ne hoştur! Bunu ancak yaşayan
45	Şefkate iman	1920	Açtı. Ahmet Reşit iki günden beri kırk sekiz saattir, yerinden kovulmuş ihtiyar
46	Lokanta Esrarı	1920	Sevgili Sermet, dün akşam mektubunu aldım. Bir sene içinde o ne dehşetli
47	Kurumuş Ağaç	1920	Deli Murat, memleketin en azılı derbederiydi! Fakat yaşlandıkça aklı başına geldi.
48	½	1927	Bu sabah gazeteye mutlaka bir hikâye yetiştirmek icap ediyordu. Fakat canım
49	Kazın Ayağı	1947	Mahzufızade Hacı Afif Efendi öyle yeni tacirlerden değildi. Babası, dedesi, silsilesi
50	Sultanlığın Sonundan	1947	Sultanlığın Başı..Büyük zelzele İstanbul'u alt üst etmişti! Eski duvarlar yıkılmış
51	Foya'dan	1947	Ertesi günü gittim. Bu konak hakikaten küçük bir saraydı. Geniş mermer merdivenli

3.4.3. Türk Diline Özgü Morfolojik Düzenlemeler

R yazılımının Türkçe karakterlerin kullanımına yönelik hassasiyet göstermesi, hazırlanan programın hata bildiriminde bulunması ve çıktıların istenilen sonucu vermemesi gibi nedenlerden dolayı, analiz işlemi gerçekleştirilecek olan metin üzerinde birtakım düzenlemelerin yapılması gerekmektedir. Bu doğrultuda metin belgesi programının kısayolları kullanılarak ç, ğ, ı, ş, ö, ü harfleri c, g, s, o, u harflerine dönüştürülmüştür. Örneğin boyun atkısı ya da baş örtüsü anlamına gelen “şal” kelimesi, bir suda ulaşım aracının adı olan “sal” kelimesi olarak değiştirilmiştir. Bu tür kelime değişimleri, değişen kelimenin analizi etkileyecek ölçüde duygu içermemesi veya belirsiz bir duygu yönüne sahip olması bakımından herhangi bir sorun oluşturmamıştır.

Metin madenciliğinin önemli bir adımı olan metin ön işlemede; metnin kelime birimlerine ayrılması, anlama katkısı olmayan eklerin çıkarılması gerekmektedir. Duygu analizi sırasında metindeki her bir kelimenin, duygu sözlüğü içerisinde yer alan ve duygu yönü pozitif veya negatif olarak etiketlenmiş kelimelerden biri ile eşleşip eşleşmemesi durumu, kelimelerin ve dolayısıyla hikâyelerin duygu kutbunun belirlenmesinin temel mantığı oluşturmaktadır. Örneğin metinde yer alan “korku” ve “korkular” kelimeleri için duygu sözlüğünde sadece duygu yönü negatif olarak belirlenmiş “korku” kelimesine yer verilmiştir. Bu doğrultuda araştırmada korkular kelimesindeki “-ler” eki atılarak her iki kelimenin de duygu sözlüğündeki “korku” kelimesi ile eşleşmesi sağlanmıştır. R yazılımında Türk dilinin morfolojik yapısına uyan veya morfolojik yapının çözümlenmesini sağlayacak paket programlar bulunmamaktadır. İngilizce ve Almanca gibi çekimli ya da Çince gibi tek heceli dillerde yapılacak işlemler için ise çokça yazılım/kodlama çalışması bulunmaktadır. İngilizcede de “-s” ve “-ies” gibi çoğul eklerinin Türk diline benzer şekilde sonda olması ve yazılımdaki paketlerin uyum sağlaması neticesinde bu eklerin kelimelerden ayrıştırılması işlemi kolaydır. Ancak sevgi kelimesinden iyelik ekleri ile türetilen “sevgim” ve “sevgisi” gibi kelimelerin R’de yapı bakımından çözümlenmesi ve sevgi kelimesine indirgenmesi kritik bir zorluğu beraberinde getirmektedir. R yazılımının haricinde Türk diline özgü morfolojik analiz için geliştirilen birtakım programlar bulunmaktadır. Ancak bu programlar, araştırmanın nesnesini oluşturan, içerisinde Osmanlı Türkçesine özgü kelimelerin ve kelime gruplarının sıklıkla yer aldığı metin verisine uygulamak için yeterli hassasiyete sahip değildir. Dolayısıyla .txt formatındaki metin üzerinde çeşitli morfolojik düzenlemeler, bu tezin aynı zamanda Türk dili uzmanlık bilgisine sahip olan yazarı tarafından yapılmış ve kelimeler yalın hale getirilmiştir. Bu çözümlemelerde anlamı

farklılaştırmayacak olan yalınlığa ulaşılması, dikkat edilmesi gereken bir noktayı oluşturmaktadır. Örneğin “sevgim” ve “sevgisi” kelimelerinin sev- köküne indirgenmesi uygun bir çözümleme değildir. “Sevmek” kelimesi bir eylemi, “sevgi” kelimesi ise bu eylem sonucunda ortaya konan olguyu bildirmektedir. Araştırmada analiz öncesinde metin verisinde yapılan morfolojik düzenlemede genel olarak “korktu, korktuğu, korktukça, korktukları, korkmuştur vb.” gibi aynı kökten ve aynı anlamdan türeyen kelimeler “kork- “şeklinde sadeleştirilmiştir. Aksi halde on binlerce kelimedenden oluşan bir metinden, içerisinde aynı sayıda kelimenin bulunduğu bir duygu sözlüğünün oluşturulması gibi zorlu ve gereksiz bir işlem ortaya çıkabilmektedir.

Morfolojik düzenleme adımları içerisinde yer alan ve Türk diline özgü özellikler nedeniyle dikkat edilen olumsuzluk ekleri (-ma, -me, -siz, -sız vb.) bu çalışmada göz ardı edilmiştir. Çünkü bu tez çalışmasında duygular, kök haline indirilmiş ve yazarın hangi duygu köklerini kullanma eğiliminde olduğu incelenmiştir. Bu analizde bir ürünle alakalı olarak alıcının kararını etkileyen görüş türünde bir metin kullanımı bulunmamaktadır. Ayrıca olumsuzluk ekinin metnin anlamını etkilemesi ile ilgilenen çalışmalar Doğal Dil İşleme alanıyla daha yakından ilişkili çalışmalardır.

3.4.4. Duygu Analizi Süreci

3.4.4.1. Analize İlişkin R Paketlerinin Kurulması

Metin verisindeki dil yapısının, analize uygunluğu ile ilgili yapılan ön çalışmanın ardından araştırmanın esas noktası olan sözlük tabanlı duygu analizinin yapılması için R yazılımında gerekli paketlerin kurulması gerekmektedir. R programlama dilinde; paket, belirli amaçlar doğrultusunda oluşturulmuş çeşitli fonksiyonları içerisinde barındıran koleksiyon anlamına gelmektedir.

Tez çalışmasında kullanılan ilgili R kütüphaneleri Tablo 3.2’ de belirtilmiştir.

Tablo 3.2. *Analizde Kullanılan R Kütüphaneleri*

Paket	Kullanım amacı
Tidyttext	Metinlerin düzenli formatlara dönüştürülmesi için gerekli olan fonksiyonları barındırmaktadır.
Readtext	Çeşitli formatlardaki metin dosyalarının R ortamına aktarılmasını sağlayan fonksiyonları içermektedir.
Quanteda	Çeşitli metin analizi ve metin düzenleme fonksiyonlarını barındırmaktadır.
Corpus	Belge koleksiyonlarının oluşturulması ve çeşitli metin analizi işlemlerini yapılmasını sağlayan fonksiyonları barındırmaktadır.
Foreign	Farklı istatistik programlarına ait dosyaların dönüştürülmesini ve R ortamına aktarılmasını sağlayan fonksiyonları içermektedir.
Tm	Veri aktarma, veri yönetimi, metin ön işleme, belge terim matrisinin oluşturulması gibi çeşitli metin madenciliği işlerinin gerçekleştirilmesini sağlayan fonksiyonları barındırmaktadır.

<https://cran.r-project.org/>

3.4.4.2. Çalışma Metninin R’ye Aktarılması

Metin madenciliği uygulamalarının tamamında verinin R’ye aktarılması başlangıç aşamasını oluşturmaktadır. Bu adımda **readlines()** ve **file()** komutları kullanılarak .txt formatındaki yapılandırılmamış metin verisi R programına aktarılmıştır.

3.4.4.3. Çalışma Metni ile Korpus Oluşturulması

Korpus, içerisinde sadece metin türündeki verileri barındıran bir çeşit veri seti nesnesidir. Bu tez çalışmasında Ömer Seyfettin’in R ortamına aktararak analize dahil edilen yapılandırılmamış veri türündeki hikâyelerinin tamamı korpusu; verideki her bir hikâye metni ise korpusun elemanlarını oluşturmaktadır. R ortamına aktarılan ana metin dosyası üzerinden hikâyelerin başlıkları, başlangıç ve bitiş noktaları esas alınarak gerekli kodlamalar yapılmış ve yapılandırılmamış ana metin, yazılıma korpus olarak tanıtılmıştır. İşlem sonucuna ilişkin çıktı Tablo 3.3’te verilmiştir.

Tablo 3.3. *Corpus_frame() Komutu Kullanılarak Oluşturulan Korpusa Ait Çıktı*

Başlık	Metin
1. hikâye	Oh, bu sabah ne kadar soğuktu. Yatağımın hararetlerini terk ettiğim vakit,
2. hikâye	Bazen muvaffak olamamış bir sanatkar hicran ve ıstırapıyla mantıksız ve
3. hikâye	Muzzaez Sermetim, İşte nihayet tevhül ettim. Bekarlığa an-ı vedaımdan
4. hikâye	Dışarıda, bütün nîli-şeffâf serairiyle mahmul ve nuşın yaz gecesi...İçerde..
5. hikâye	Evvelâ beni sen sevdin, yalvardın, yakardın; benim aşkım adeta senin galeyan

Baslık	Metin
6. hikâye	Tokatlıyan'ın sırasındaki yirmi beş numaralı apartmanı arıyordum. Louis Durant
7. hikâye	Bu roman hem hakikat,hem de hayaldir. Ömer seyfettin hakiki simaların bazı
8. hikâye	Bilsen Efruzcuğum, kırk gündür burada ne rahat yaşıyordum. Ses yok, seda yok,
9. hikâye	Ben Kadıköyü'nde büyüdüm. O vakitler mahalle bile yoktu! Tek tük evler..
10. hikâye	Ona akrabaları "meraklı", uzaktan tanıyanlar "deli" derlerdi. Yaşı kırk beşi geçi-
11. hikâye	Yemekten kalkalı belki bir saat olmuştu. Karı koca, kahvelerini, her vakitki gibi
12. hikâye	Derviş Hasan birdenbire durdu. Kirli, yırtık yenleriyle alnının terlerini sildi.
13. hikâye	"Ulan İboş, sen be!" "Vay Mıstık, sen ha?" "Ben ya..." İki hemşeri hemen
14. hikâye	Mahallenin yeni yetişen gençleri Hacı İmadettin Efendi'nin yalnız ismini bilir-
15. hikâye	Geçen gün hava ne güzeldi! Logaritmacı Hasan'la Hürriyet Tepesi'ne gittik.
16. hikâye	İstanbul üç gündür sis içindeydi. Topkapı Sarayı'nı, açık, kül rengi kalın bir bulut
17. hikâye	Ah gençlik!.. Tıpkı ezeli bir baharın ilk çiçekli günlerine benzer. Yeşil kırlar
18. hikâye	Daha geceydi, Çoban yıldızı derinlerden dünya üzerindeki uyumuş sürülere
19. hikâye	Ah, on beş, yirmi sene evvel... hayat ne tatlıydı! Yaşamamın, eğlenmenin hoş
20. hikâye	Âdem Bey büyük ıhlamur ağaçlarının altında yapayalnız geziniyordu. Ay, mavi,
21. hikâye	Belkıs, geniş yatağında, mavi ipek kaplı yorganının altında sıkışmış bir yumruk
22. hikâye	Ruhumda ilahi bir ateş olmadığı için mi, yoksa müspet ilimlerle pek gençken
23. hikâye	Mestan Ağa doğduğu büyüdüğü Edirne'de "Çingane" olarak yaşamağa tahammül
24. hikâye	Mehmet Efendi on senedir kasabada oturuyordu. Köydeki tarlaları, bağları,
25. hikâye	Akdeniz'in esatir yuvası nihayetsiz ufuklarına bakan küçük tepe minimini bir
26. hikâye	Sıcak bir mayıs günüydü. Küçük bahçeyi gölge içinde bırakan kocaman tek dut
27. hikâye	"Sevgili Celil, İşte bir haftadan beri her sabah sana mektup yazmak azmiyle
28. hikâye	Hasan Ağa başka hemşehrileri gibi yapmamış, İstanbul'a karısıyla beraber gel-
29. hikâye	Biz, bu son asrın muharrirleri en ehemmiyetsiz, en adi şeylere kıymet verir,
30. hikâye	Sermet Bey döndü, arkasındaki bekçiye, "İşte bir boş köşk daha!" dedi. Küçük
31. hikâye	Kasaba içinde Kadı Mustafa Efendi'den hazzeden kimse yoktu! Dört parmak
32. hikâye	Çocukluğumuzda büyük çam ağaçları altına kurulan salıncakları her türlü ders
33. hikâye	On beş, yirmi sene evvel ordunun manzarası ne latif! Taburlarda mektepli
34. hikâye	Geçen sene bir gün arkadaşım Sermet'le Şişli'ye doğru yürüyorduk. Birdenbire
35. hikâye	Bu kocaman, harap İstanbul'un içinde yaşayan hangi sanatkar "güzelliğin"
36. hikâye	Hamdune Hanım eski bir şeyhülislam haemidir. Yaşı bugün altmışı aşmak üzere.
37. hikâye	Aşk filan değil..Hani şu "tesadüf" dediğimiz tarihi yapan, saadetleri ibda eden
38. hikâye	Ahırın avlusunda oynarken aşağıdaki, gümüş söğütler altında görünmeyen dere
39. hikâye	Yatağında kımıldayamayan Durmuş Ağa gözlerini basık tavana dikmişti. Sanki
40. hikâye	Aksaraylı Câbir Paşa'yı bir zamanlar tanımayan yoktu; zevcesi cennetmekanın
41. hikâye	Yine bir gün derin bir ısırtıp içinde kıvranıyordum! Annem, babam, kardeşlerim
42. hikâye	Bir ay geçmeden fikrim değişti! Ama öyle yavaş yavaş değil... Birdenbire! Bugün
43. hikâye	Dün seni gördüm. Gayr-i ihtiyari avazım çıktığı kadar " A, a, aaa.."diye haykırdım.
44. hikâye	Taşra âlemi... Yani İstanbul'un dışında geçen hayat, ne hoştur! Bunu ancak yaşayan
45. hikâye	Açtı. Ahmet Reşit iki günden beri kırk sekiz saattir, yerinden kovulmuş ihtiyar
46. hikâye	Sevgili Sermet, dün akşam mektubunu aldım. Bir sene içinde o ne dehşetli
47. hikâye	Deli Murat, memleketin en azılı derbederiydi! Fakat yaşlandıkça aklı başına geldi.
48. hikâye	Bu sabah gazeteye mutlaka bir hikâye yetiştirmek icap ediyordu. Fakat canım
49. hikâye	Mahzufzade Hacı Afif Efendi öyle yeni tacirlerden değildi. Babası, dedesi, silsilesi
50. hikâye	Sultanlığın Başlı..Büyük zelzele İstanbul'u alt üst etmişti! Eski duvarlar yıkılmış
51. hikâye	Ertesi günü gittim. Bu konak hakikaten küçük bir saraydı. Geniş mermer merdivenli

3.4.4.4. Metin Ön İşleme Adımlarının Uygulanması

Metin verilerinin analize hazırlanması, analizin kendisi kadar önemli ve analizin sonuçlarının seyri açısından oldukça etkili bir metin madenciliği aşamasıdır. Tokenlerine ayırma; noktalama işaretlerini, sembolleri ve boşlukları kaldırma; harflerin boyutunu normalleştirme (büyük-küçük), sayısal ifadeleri kaldırma, etkisiz kelimeleri çıkarma ve kelimeleri kök haline getirme gibi birtakım işlemler temel metin ön işleme adımlarıdır. Farklı metin madenciliği uygulamalarında bu adımlara eklemelerde bulunmaktadır.

Bu doğrultuda korpustaki metinler, noktalama işaretlerinin, sembollerin ve boşlukların kaldırılması; harflerin boyutunu normalleştirilmesi, sayısal ifadeleri kaldırılması şeklindeki işlemleri içerisinde barındıran **text_filter()** komutu kullanılarak, analize herhangi bir katkısı bulunmayan ifadelerden temizlenmiştir.

Metnin noktalama işaretleri de dahil olmak üzere içerisinde boşluklarla birbirinden ayrılan her bir parçasına token ya da dizge adı verilmektedir. Tokenlere ya da birimlere ayırma işlemi, istatistiksel bir metin analizi gerçekleştirilmesi amacıyla yapılandırılmamış ham veriden yapılandırılmış bir veri seti oluşturulması için uygulanması gereken en temel metin madenciliği adımıdır. Kelime ya da token sıklığı belirleme, özellik çıkarımı gibi çeşitli metin madenciliği uygulamaları için kullanılan metin istatistikleri, tokenlerine ayırma işleminin ardından elde edilmektedir. Bu doğrultuda tez çalışmasında **text_token()** komutu kullanılarak corpus içerisindeki metinler tokenlere/birimlere ayrılmıştır.

Veriyi temizleme, normalleştirme ve birimlere ayırma gibi işlemlerin ardından tez çalışmasında uygulanan son metin madenciliği ön işleme adımı; metinlerde anlama katkısı olmayan belki, birkaç, tabii, zaten gibi ifadelerin kaldırılmasıdır. Bu doğrultuda belirli kodlamalar ile kaynağı tanımlanan etkisiz ifadeler, **stopwords()** komutu kullanılarak metin verileri içerisinden çıkarılmıştır. Ek-2’ de 471 etkisiz ifadenin bulunduğu listeye yer verilmiştir.

Metin ön işleme adımlarından biri olan kelimelerin kök haline getirilmesi işlemi dile özgü farklılıklar ve yazılımdaki uyumsuzlar nedeniyle araştırmacı tarafından Metin Verilerinin Analize Hazırlanması başlığı altında gerçekleştirilmiştir. Metin ön işleme adımlarının ardından Tablo 3.4’te korpusun ve içerisindeki metinlerin analiz öncesindeki son haline yer verilmiştir.

Tablo 3.4. *Metin Ön İşleme Adımlarının Ardından Korpus İçerisinde Yer alan Metinlerin Analiz Öncesindeki Son Halleri*

Başlık	Metin
1. hikâye	oh bu sabah ne kadar soguk yatak hararet terk et vakit cilgin firtina haykir tehdit
2. hikâye	bazen muvaffak ol bir sanat hicran ve istirab mantik ve serseri bir tugiyen heves at
3. hikâye	muazzez sermet iste nihayet teehhul et bekarlik ani veda bir sene sonra garp ve
4. hikâye	disari butun nili seffaf serair mahmul ve nusin yaz gece iceri lamba yesil abajur
5. hikâye	evvela ben sen sev yalvar yakar ben ask adeta sen galeyan son bir cevap nihayet ben
6. hikâye	tokatliyan sira yirmi bes numara apartman ara lousi durant gayet parla ve kibar bir
7. hikâye	bu roman hem hakikat hem de hayal omer seyfettin hakiki sima bazi garabet tuhaf
8. hikâye	bil efruz kirk gun burada ne rahat yas ses yok seda yok dost yok dusman yok gurultu
9. hikâye	ben kadikoy buyu o vakit mahalle bile yok tek tuk ev kal ve hali arsa sokak yalniz
10. hikâye	ona akraba merak uzak tani deli de yas kirk bes gec henuz evlen yaz hep kir gez
11. hikâye	yemek kalk belki bir saat ol kari koca kahve her vakit gibi yali balkon ic ic sey
12. hikâye	dervis hasan birdenbire dur kirli yirt yen alin ter sil sicak bir haziran gunes dünya
13. hikâye	ulan ibos sen be vay mistik sen ha ben ya iki hemseri hemen kucak makedonya cik
14. hikâye	mahalle yeni yetis genc haci imadettin efendi yalniz isim ic daha yuz gor yok orta
15. hikâye	gecen gun hava ne guzel logartimaci hasanla hurriyet tepe git bomonti kadar uzan
16. hikâye	istanbul uc gun sis topkapi saray acik kul renk kalın bir bulut sar sanki butun dünya
17. hikâye	ah genc tipki ezeli bir bahar ilk cicek gun benze yesil kir kelebek dolu bahce güzel
18. hikâye	daha gece coban yildiz der dünya üzeri uyu suru bak tek bir goz gibi parla koyu
19. hikâye	ah on bes yirmi sene evvel hayat ne tat yasa eglen hos bir zevk guzel bir sekil var
20. hikâye	adem bey buyuk ihlamur agac alt yapayalniz gez ay mavi seffaf bir tul ortu dargin
21. hikâye	belkis genis yatak mavi ipek kapli yorgan alt sikil bir yumruk gibi yusyumu yat
22. hikâye	ruh ilahi bir ates ol icin yoksa muspet ilim pek genc ugras basla her neden ben ask
23. hikâye	mestan aga dog buyu edirne cingene ol yasa tahammul et gurbet diyar arap gibi
24. hikâye	mehmet efendi on sene kasaba otur koy tarla bag bahce ortak el kal aziz ahabab muftu
25. hikâye	akdeniz esatir yuva nihayetsiz ufuk bak kucuk tepe minimini bir cicek orman gibi
26. hikâye	sicak bir mayis gun kucuk bahce golge birak kocaman tek dut agac alt eski temiz bir
27. hikâye	sevgili celil iste bir hafta beri her sabah sana mektup yaz aziz kalk nihayet bugun
28. hikâye	hasan aga baska hemsehir gibi yap istanbul karisi beraber gel para kazan iyi sey
29. hikâye	biz bu son asir muharrir en ehemmiyet en adi sey kiymet ver el bir fırsat gec pire
30. hikâye	sermet bey don arka bekci iste bir bos kosk daha de kucuk bir cam orman beyaz
31. hikâye	kasaba kadi mustafa efendi haz kimse yok dort parmak siyah cat kas kucuk parla goz
32. hikâye	cocuk buyuk cam agac alt kur salincak her turlu ders ve eglence kitap tercih et
33. hikâye	on bes yirmi sene evvel ordu manzara ne latif tabur mektep zabıt tek tuk bul ekseriyet
34. hikâye	gecen sene bir gun arkadas sermet sisli dogru git birdenbire karsi efruz bey cik
35. hikâye	bu kocaman harap istanbul ici yasa hangi sanat guzel matem duy tabiat el sanki dünya
36. hikâye	hamdune hanim eski bir seyhulislam harem yas bugun altmis gec uzere gam kasvet
37. hikâye	ask filan degil han su tesaduf deniz tarih yap saadet ibda et yuva kur mechul el yok
38. hikâye	ahir avlu oyna asagi gumus sogut alt gorun dere hazin sirilti isit ev ic ic buyuk kestane
39. hikâye	yatak kimilda durmus aga goz bas tavan sanki saat eski sarar hatili sayi yuz toprak
40. hikâye	aksaray cabir pasa bir zaman tani yok zevce cennet mekan goz haznedar usta is ic
41. hikâye	yine bugun derin bir istirap ic kivran anne baba kardes akraba hepsi hepsi ben cilgin

Baslık	Metin
42. hikâye	bir ay gec fikir degis ama oyle yavas yavas degil birdenbire bugun anne hak ver
43. hikâye	dun sen gor gayriihtiyari avaz cik kadar diye haykir eger araba hizla uzak mutlaka
44. hikâye	tasra alem yan istanbul disinda gec hayat ne hos bu ancak yasa bir taraf esraf ulema
45. hikâye	ac ahmet resit ber kirk sekiz saat yer kov git serseri bir kopek gibi tani sokak bil
46. hikâye	sevgili sermet dun aksam mektup al bir sene o ne dehset daussilade bekar alem goz
47. hikâye	deli murat memleket en azili bir derebey fakat yaslan akil bas gel iyi kotu fark var
48. hikâye	bu sabah gazete mutlaka bir hikâye yetis cap et fakat can yatak kalk degil boy geri
49. hikâye	mahzufizade haci efendi oyle yeni tacir degil baba dede silsile tuccar kendi mahalle
50. hikâye	sultan basi buyuk zelzele istanbul alt ust et eski duvar yik curu cati cok sancaktar
51. hikâye	ertesi gun git bu konak hakikaten kucuk bir saray genis mermer merdiven ceviz

3.4.4.5. Metin İstatistiklerinin Elde Edilmesi

Metinler, token ya da dizgecik adı verilen parçaların bir araya gelmesiyle oluşan ana bir dizgidir. Doküman düzeyinde duygu analizi yapılan bu tez çalışmasında da korpus içerisindeki satırlarda yer alan metinler, tokenlerin bir araya gelmesiyle oluşmuş devasa uzunlukta bir cümleyi temsil etmektedir.

Dolayısıyla korpustaki metinlerin tokenlerine ilişkin istatistiksel bilgiler bölümlerin uzunlukları ve içerisindeki sözcük çeşitliliği hakkında bilgi edinmek için kullanılabilir. Bu doğrultuda **text_ntoken()** ve **text_notype()** komutu kullanılarak ilgili metin istatistikleri elde edilmiştir.

3.4.4.6. Tip-Token Oranına İlişkin Regresyon Analizi

Stanley'e (1978) göre, benzersiz token tiplerinin logaritması token sayısının doğrusal bir fonksiyonudur. Bu yasanın test edilmesi ile bölümlerdeki sözcük çeşitliliği (tip-token oranı) hakkında bilgi edinilmektedir. Farklı tokenlerin sayısı, aynı zamanda analizin gücünü/hassaslığını ortaya koymaktadır. Dolayısıyla güçlü ve yakın tip-token oranına sahip bölümlerin karşılaştırmalı analizinde daha sağlıklı sonuçlar elde edilmektedir.

Bu doğrultuda Ömer Seyfettin’e ait 51 hikâye metni üzerinde test edilen tip-token oranına ilişkin olarak R programı içerisinde lm() fonksiyonu yardımıyla yapılan basit doğrusal regresyon analizi ile elde edilen sonuçlar Tablo 3.5’de gösterilmiştir.

Tablo 3.5. Tip-Token Oranının Test Edilmesine İlişkin Regresyon Sonuçları

Katsayılar	Tahmin	t değeri	Pr(> t)	R-Kare
Sabit	0.65286	2.492	0.0161	0.9009
log(tokens)	0.79817	21.109	2e-16	

Tablo 3.5’de sunulan bulgular incelendiğinde, gözlenen ve kestirilen değerler arasındaki karesel korelasyon anlamına gelen R-kare değerinin 0.9009 olduğu görülmektedir. Bu değer, token tipleri içindeki varyasyonun %90.09’unun token sayısı ile açıklanabileceğini ifade etmektedir. Regresyon modelindeki sabit terim, açıklayıcı değişkenin sıfır olduğu durumda açıklanan değişkenin alacağı değeri temsil etmektedir. Tablo 3.5’de sabit terim için, p değeri=0.0161<0.05 olduğundan bu sabitin istatistiksel olarak önemli olduğu bilgisine ulaşılmaktadır. Ayrıca, çıktıda token sayısının; p=2e-16<0.05 olduğundan istatistiksel olarak önemli olduğu görülmektedir. Bu doğrultuda tip-token oranlarının birbirine yakın olduğu belirlenen metinler üzerinde duygu analizinin uygulanabilir olduğu ve ilgili analizin sonuçlarının güvenilir olacağı yargısına varılmaktadır.

3.4.4.7. Duygu Sözlüklerinin Oluşturulması ve R’ ye Aktarılması

Bu tez çalışmasında, metin verisi içerisinde token haline getirilmiş kelimelerden yararlanılarak iki adet duygu sözlüğü oluşturulması amaçlanmıştır.

Araştırmacı tarafından hazırlanan birinci sözlükteki duygu sözcükleri, korpusta yer alan bütün kelimelerin, türü (fiil, isim, sıfat) ve anlamı fark etmeksizin, duygu yönleri (pozitif-negatif) bakımından değerlendirilmesiyle oluşturulmuştur (toplam duygu kelimesi sayısı: 1343). Örneğin cinayet kelimesi, şaşkın kelimesi gibi bir duyguyu karşılamak için kullanılan sıfat türünde bir duygu ifadesi değildir. Bununla birlikte cinayet kelimesinin görevi bir duyguyu nitелеmek değil; isim türünde bir kelime işlevi görerek, bir durumu veya fiili adlandırmaktır. Ancak, kelimenin dili kullanan insanların zihninde oluşturduğu ilk ve baskın olan hissin olumsuz olduğu dikkate alınarak, bu kelimenin bir duygu içerdiği için duygu yönü bakımından sınıflandırılabilceği yargısına varılmaktadır.

Araştırmacı tarafından hazırlanan ikinci sözlükte ise metin verileri içerisinde, korku kelimesi gibi sadece bir duyguyu ifade etmek amacıyla kullanılan veya bu kelimeden türetilen (korkunç) duygu kelimelerine yer verilmiştir (toplam duygu kelimesi sayısı: 140). Araştırmada aynı metin verisi üzerinde iki farklı düzeyde duygu sözlüğü kullanılarak, duygu sözlüğünün hassasiyetinin analizin gücüne etkisinin incelenmesi amaçlanmıştır. Bu doğrultuda R ortamı dışında bir veri analizi aracı kullanılarak iki ayrı kelime ve duygu kutbu tablosu oluşturulmuştur. Tablo 3.6’ da birinci duygu sözlüğü için yapılan kelime duygusu etiketleme işlemine ilişkin örnek olarak çeşitli kelime ve duygu kutuplarına yer verilmiştir.

Tablo 3.6. *Birinci Duygu Sözlüğünün İçeriğine İlişkin Örnek Kelimeler ve Duygu Kutupları*

Kelime	Duygu kutbu
Üzüntü	Negatif
Heyecan	Pozitif
Gazap	Negatif
Şaşkın	Negatif
Endişe	Negatif
Ağlamak	Negatif
Gülmek	Pozitif
Mezar	Negatif

Tablo 3.7’de ise ikinci duygu sözlüğü için yapılan kelime duygusu etiketleme işlemine ilişkin örnek olarak çeşitli kelime ve duygu kutuplarına yer verilmiştir.

Tablo 3.7. *İkinci Duygu Sözlüğünün İçeriğine İlişkin Örnek Kelime ve Duygu Kutupları*

Kelime	Duygu kutbu
Üzüntü	keder
Heyecan	coşku
Gazap	öfke
Şaşkın	hayret
Endişe	korku

Devamında **read.csv()** komutu kullanılarak duygu sözlükleri R yazılımına aktarılmıştır.

3.4.4.8. Terim-Duygu Matrisinin Oluşturulması

Bu adımda R ortamına aktarılan veri tablosu formatındaki duygu sözlüğünün, yapılandırılarak analize tanıtılması amaçlanmıştır. Bu doğrultuda **with()** komutu yardımıyla satırlarda duygu terimi, sütunlarda kelimelerin sayısallaştırılmış duygu yönleri yer alacak şekilde terim-duygu matrisi oluşturulmuştur. Burada 1 ve 0 değerleri, işlem yapılabilen sayısal bir özellik yerine kelimelerin duygu yönleri bakımından sınıflandırılması görevini üstlenmektedir. Dolayısıyla 1 değeri kelimenin ait olduğu sınıfı belirtmektedir. Bu doğrultuda Tablo 3.8’de birinci duygu sözlüğü; Tablo 3.9’da ikinci duygu sözlüğü ile oluşturulan terim-duygu matrislerine ilişkin örnek kelime ve kelime kutbu değerlerine yer verilmiştir.

Tablo 3.8. Birinci Duygu Sözlüğü İçerisinde Yer Alan Terimler ve Duygu Kutupları

Kelime	Duygu kutbu	
	Pozitif	Negatif
Uzuntu	0	1
Heyecan	1	0
Gazap	0	1
Saskin	0	1
Endişe	0	1
Ağlamak	0	1
Gülmek	1	0
Pınar	1	0
Mezar	1	1

Tablo 3.9. İkinci Duygu Sözlüğü İçerisinde Yer Alan Terimler ve Duygu Kutupları

Kelime	Duygu Kutbu				
	Pozitif		Negatif		
	coşku	keder	öfke	hayret	Korku
gazap	0	0	1	0	0
Saskin	0	0	0	1	0
Endişe	0	0	0	0	1
Üzüntü	0	1	0	0	0
Heyecan	1	0	0	0	0

3.4.4.9. Bölümlerin Daha Küçük Parçalara Ayrılması

Bu adımda her bölümün kendi içerisindeki duygu durumu değişiminin ortaya konması ve karşılaştırmanın daha dengeli şekilde yapılması amacıyla, bölümler **text_split()** komutu ile en düşük token sayısı (25.bölüm:511) dikkate alınarak 500 tokenlik parçalara ayrılmıştır.

3.4.4.10. Duygu Sözlüklerine Göre Metinlerin Duygu Skorlarının Hesaplanması

Analizin esas uygulamasını oluşturan bu adımda, öncelikle duygu sözlüğündeki duygu terimleri ve 500 tokenlik bölümlerdeki kelimeler kullanılarak `term_matrix()` komutu ile bir terim-duygu puanı matrisi oluşturulmuştur. Devamında duygu oluşum oranının hesaplanması için duygu puanları token boyutuna bölündükten sonra 1000 ile çarpılmıştır. **List()** komutu kullanılarak yapılan bu hesaplama neticesinde parçalara ayrılmış her bir bölümde 1000 token başına düşen duygu terimi ortaya konmuştur.

3.4.4.11. Duygu Analizi Sonuçlarının Görselleştirilmesi

Metin madenciliği genel adımlarından biri olan değerlendirme ve yorumlama aşamasında analiz sonuçların görselleştirilmesi, yapılan analizin daha anlaşılır bir şekilde sunulması ve sonuçların daha iyi bir şekilde yorumlanması açısından önemlidir. Bu doğrultuda tez çalışmasında, Türkçe metinlerde sözlük tabanlı duygu analizi uygulaması sonuçlarının ortaya konması için frekans poligonu ve kelime bulutu oluşturulmuştur.

3.4.4.11.1. Duygu Analizi Sonuç Grafiğinin Çizilmesi

Sonuçların frekans poligonu ile çizilmesi amacıyla, temel olarak, R’de grafik çizimleri için sıkça kullanılan **plot()** komutundan yararlanılmıştır. Analizin esas kısımları ile birlikte bu adımı gerçekleştirmek için yazılan program kodları Ek-1’de verilmiştir.

3.4.4.11.2. Duygu Analizi Sonucuna İlişkin Kelime Bulutunun Oluşturulması

Metin madenciliğinde kelime bulutları genel olarak frekansı yüksek olan terimlerin ya da kelimelerin görsel açıdan daha anlaşılır ve ilgi çekici şekilde sunulması amacıyla kullanılmaktadır. Bulut içerisindeki kelimelerin fiziksel büyüklükleri, sıklıkları hakkında bilgi vermektedir. Bu adımı gerçekleştirmek için **wordcloud()** paketinden yararlanılmıştır.

3.4.4.12. Ömer Seyfettin'e Ait 51 Hikâye ile Gerçekleştirilen Sözlük Tabanlı Duygu Analizinin, Yazarın Daha Önce Analize Dahil Edilmemiş Diğer Hikâyelerinin De Eklenecek Tekrar Uygulanması

Ömer Seyfettin'e ait ve daha dengeli ve tutarlı bir karşılaştırma yapılması amacıyla özellikle seçilmiş 51 hikâye metni üzerinde gerçekleştirilen sözlük tabanlı duygu analizi, analize daha önce dahil edilmemiş diğer hikâyelerinde de eklenmesiyle toplamda 165 hikâye üzerinde yeniden uygulanmıştır.



BÖLÜM IV

BULGULAR

Tezin bu bölümünde öncelikle, araştırmanın amaçları doğrultusunda, R’de sözlük tabanlı duygu analizi ile incelenen Ömer Seyfettin’e ait 51 eserin analiz sonuçlarına ilişkin bulgulara yer verilmiştir. Bu doğrultuda, öncelikle metinlerin genelini hedef alan analiz sonuçları yorumlanmış; devamında ise 51 hikâye metni üzerinde farklı özelliklere sahip iki ayrı duygu sözlüğü ile gerçekleştirilen analiz sonuçları karşılaştırılmıştır. Daha sonra, 165 hikâye üzerinde yeniden uygulanan sözlük tabanlı duygu analizi sonuçlarına değinilmiştir.

4.1. Metinlere İlişkin İstatistik Bulgular

Bu başlık altında Ömer Seyfettin’in analize dahil edilen 51 hikâye metni içerdikleri token sayıları ve token çeşitlilikleri bakımından incelenerek duygu analizine uygunlukları açısından değerlendirilmiştir. Metinlere ilişkin genel istatistik bulgular verilmeden önce, Tablo 4.1’te araştırmanın materyallerini oluşturan her bir metnin sahip olduğu token sayısı ve token tipi açıkça gösterilmiştir.

Tablo 4.1. *Hikâyelerdeki Token ve Tip Sayıları*

Hikâye	Token Sayısı	Tip Sayısı	Cümle
1. hikâye	1104	517	1
2. hikâye	1458	701	1
3. hikâye	1216	562	1
4. hikâye	1207	469	1
5. hikâye	982	440	1
6. hikâye	887	446	1
7. hikâye	1316	615	1
8. hikâye	1208	535	1
9. hikâye	933	411	1
10. hikâye	1363	669	1
11. hikâye	1086	483	1
12. hikâye	1318	600	1
13. hikâye	918	458	1
14. hikâye	1372	672	1
15. hikâye	1266	570	1
16. hikâye	1132	657	1

Hikâye	Token Sayısı	Tip Sayısı	Cümle
17. hikâye	1155	580	1
18. hikâye	817	398	1
19. hikâye	1061	503	1
20. hikâye	633	298	1
21. hikâye	763	399	1
22. hikâye	763	384	1
23. hikâye	777	434	1
24. hikâye	511	260	1
25. hikâye	1015	499	1
26. hikâye	1578	679	1
27. hikâye	1580	743	1
28. hikâye	870	405	1
29. hikâye	850	488	1
30. hikâye	1477	552	1
31. hikâye	622	317	1
32. hikâye	791	365	1
33. hikâye	759	378	1
34. hikâye	1292	569	1
35. hikâye	591	362	1
36. hikâye	836	448	1
37. hikâye	1327	557	1
38. hikâye	863	377	1
39. hikâye	823	397	1
40. hikâye	964	531	1
41. hikâye	1548	646	1
42. hikâye	1044	461	1
43. hikâye	925	427	1
44. hikâye	1335	597	1
45. hikâye	720	377	1
46. hikâye	1325	592	1
47. hikâye	806	383	1
48. hikâye	1381	608	1
49. hikâye	969	488	1
50. hikâye	994	493	1
51. hikâye	1151	527	1

Tablo 4.1’deki verilere göre, sayfa sayıları (4, 5 ve 6) dikkate alınarak analiz için seçilen metinlerin token ve tip sayıları bakımından genel olarak yakın değerlere sahip oldukları görülmektedir. Araştırmanın doküman bazında gerçekleştirilen bir analiz olması doğrultusunda

cümle değerlerinin tamamının 1 olduğu görülmektedir. Bu her bir metnin devasa bir cümle olarak incelendiği anlamına gelmektedir.

Veriler aracılığıyla ulaşılan token ve tip sayıları ışığında analize uygunluklarına ilişkin elde edilen istatistik bulgulara Tablo 4.2’te verilmiştir.

Tablo 4.2. *Hikâyelerdeki Token ve Tip Sayılarına İlişkin İstatistiksel Bulgular*

	Token Sayısı	Tip Sayısı	Cümle
En Küçük	511.0	260.0	1
1. Çeyrek	829.5	402.0	1
Ortanca	1015.0	488.0	1
Ortalama	1052.6	494.8	1
3.Çeyrek	1304.0	569.5	1
En Büyük	1580.0	743.0	1

Tablo 4.2’de de görüldüğü üzere, metinlerde ortalama token sayısı 1052.6; ortalama tip sayısı ise 494.8’dir. Bu bilgiler ışığında metinlerdeki duygu skoru oranlamasının 1000 paydasında yapılabileceği söylenebilmektedir. Buna ek olarak daha eşit bölümler içinde analiz yapılması ve daha dengeli analiz sonuçlarına ulaşılması bakımından en az olan token sayısı dikkate alınarak, 51 metnin tamamının 500 tokenlik bölümlere ayrılması gerektiği bilgisine ulaşılmıştır. Tip başlığı altında ele alınan token çeşitliliği ortalamaları incelendiğinde ise sayfa sayılarına bağlı olarak token sayıları ile aralarında doğru bir orantı olduğu görülmektedir. Dolayısıyla bu başlık altında analiz sırasında gerçekleştirilen daha küçük bölümlere ayırma ve duygu skorlarını oranlama gibi adımlarda, yapılan işlemlerin uygunluğunun bir kez daha doğrulandığı görülmektedir.

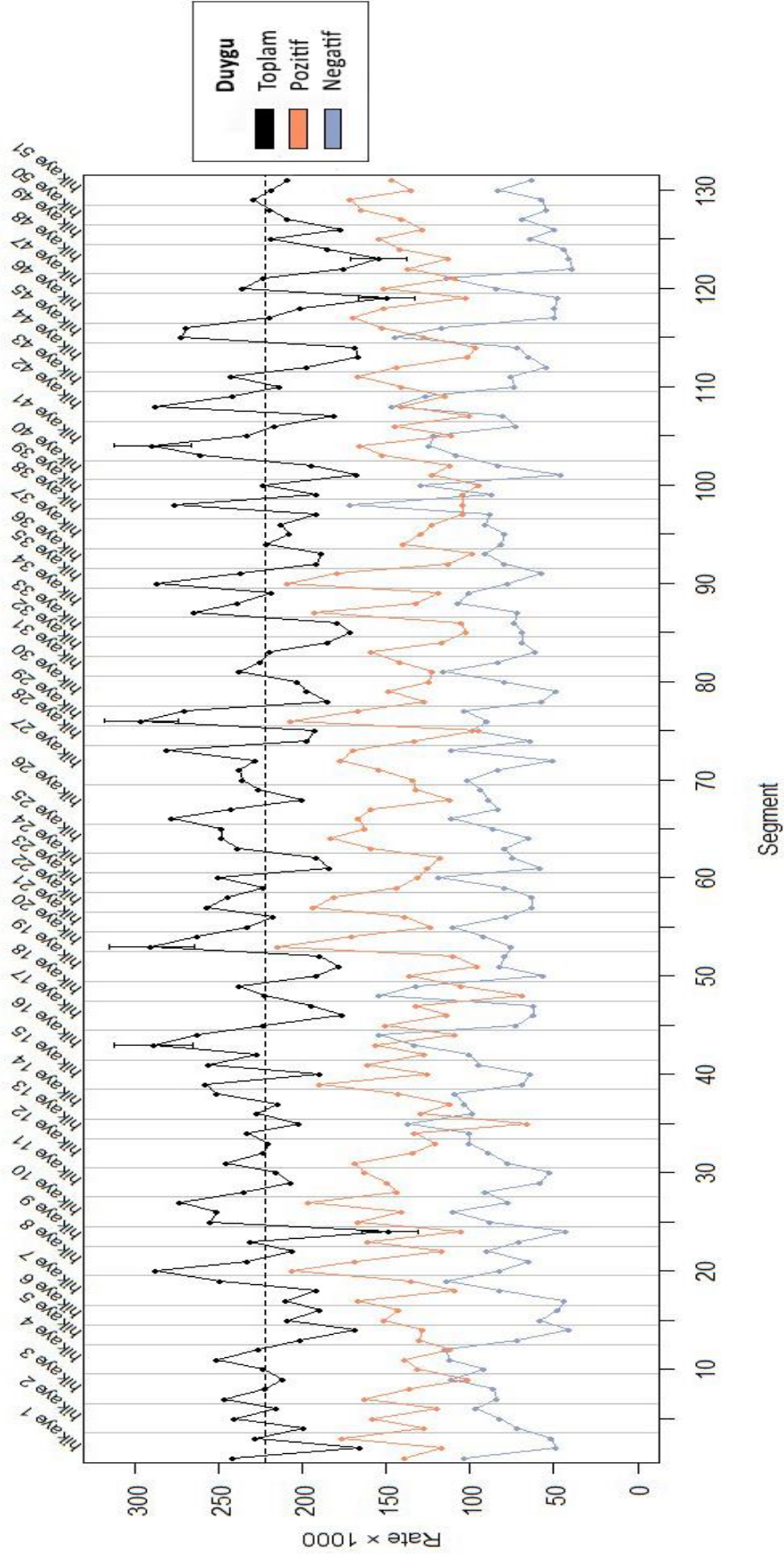
4.2. Birinci Duygu Sözlüğü ile Yapılan Duygu Analizine İlişkin İstatistik Bulgular

Bu başlık altında, metinler üzerinde birinci duygu sözlüğü kullanılarak gerçekleştirilen sözlük tabanlı duygu analizi sonuçlarına yer verilmiştir.

4.2.1. Analiz Sonucuna İlişkin Frekans Poligonu

R yazılımında birinci duygu sözlüğü temelinde ilgili kodlamalar yapılarak elde edilen sözlük tabanlı duygu analizi sonuçlarına ilişkin Şekil 4.1’de yer verilmiştir.

Grafikte, metinlerin 500 tokenlik parçalara ayrılması işleminin ardından toplamda 130 bölüm elde edildiği görülmektedir. Ayrıca total-toplam kategorisindeki eğriler ve eğri üzerindeki noktalar incelendiğinde; ortalamanın üzerinde olan ve duygu içeren kelime sayısı toplamı bakımından diğerlerine göre belirgin bir farklılık ortaya koyan ve 9, 16., 20., 29., 41., 46. ve 48. hikâyelerde yer almak üzere toplamda 7 hikâye bölümünün tespit edildiği görülmektedir.



Şekil 4.1. Birinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonucuna İlişkin Frekans Poligonu

4.2.3. Yıllara Göre Metinlerde En Sık Kullanılan Duygu Kelimelerinin Belirlenmesi

Bu başlık altında Ömer Seyfettin'in 51 hikâye metnini ortaya koyduğu yıllar ve metinler içerisinde birinci duygu sözlüğüne göre tanımlanarak en sık kullandığı duygu kelimeleri olduğu kanısına varılan kelimeler incelenmiştir. Bu doğrultuda Tablo 4.3'te aynı yıl içerisinde yazılan metinler bir grubu oluşturacak şekilde; metinlere, yazıldığı yıla ve sırasıyla sıkça kullanılan duygu kelimelerine yer verilmiştir.

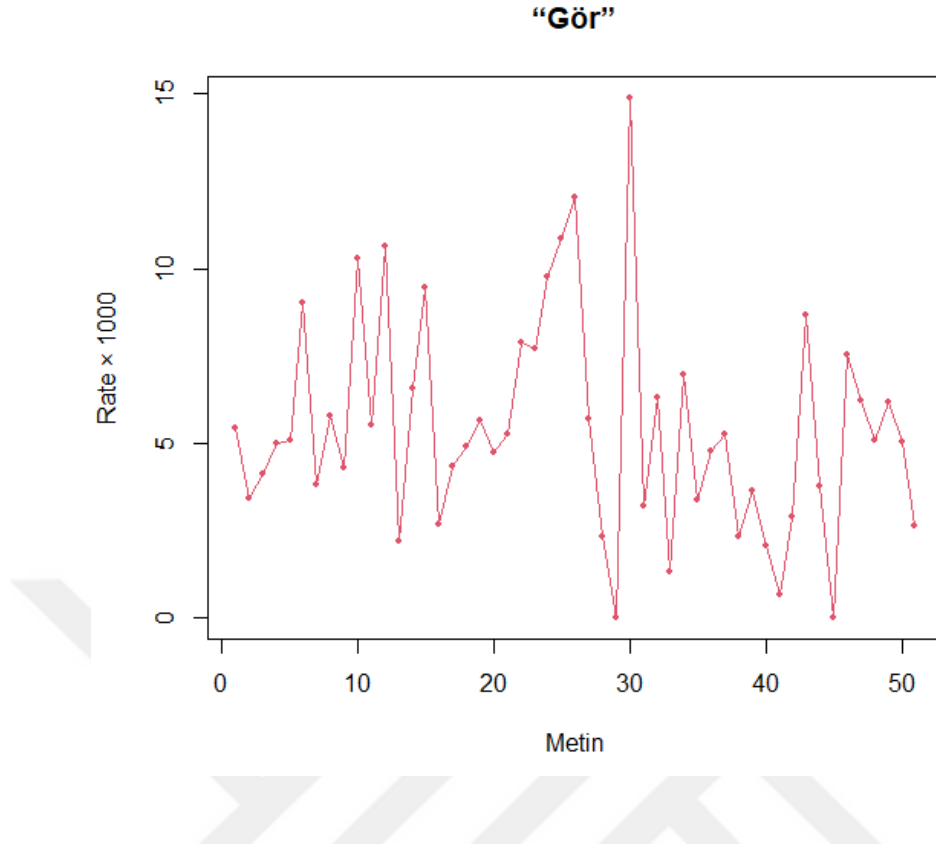
Tablo 4.3. *Yazıldıkları Yıllara Göre Gruplanan Metinlerde Sıkça Kullanılan Duygu Kelimleri*

Metin	Yıl	Metinlerde en sık kullanılan duygu kelimesi
1, 2	1905	his, gör, oku, kitap, düşün, ask, güzel
3	1907	his, bul, hayal, basla, gör, güzel, son, ay, bedbaht
4	1909	kitap, basla, hücum, siyah, karanlık, matem, buyu-, emel, gör, hırs
5	1910	bul, güzel, sev, gör, basla, ask, tekamül, zengin, asla, birinci
6-9	1914	gör, büyük, bul, gül, basla, güzel, çok, kir, cevap
10-13	1917	gör, düşün, basla, çok, su, büyük, genç, ay, gül
14- 21	1918	gör, gül, bul, büyük, basla, üst, genç, can beyaz ölüm bırak akıl
22-43	1919	basla, gör, bul, , büyük, çok, doğru, gül, su, son, bırak, akıl
44- 47	1920	bul, düşün, başla, ölüm, çok, açıl, doğru, gül, uzak, beraber, yaşa, gör, hayat
48	1927	muallim, mektep, bul, çocuk, gör, üst, can, hayat, bekle, , büyük
49-51	1947	büyük, gör, genç, kazan, basla, oku, bul, gul, yeni, beraber, mektep, üst

Tablo incelendiğinde, grup içerisindeki eleman sayısının en yüksek olduğu 1918 ve 1919 yıllarında, sıkça kullanılan kelimelerin (gör, bul, başla, büyük ve diğerleri) kelime bulutu yardımıyla elde edilen ve metinlerin genelinde frekansı yüksek olan duygu kelimeleri ile uyduğu görülmektedir.

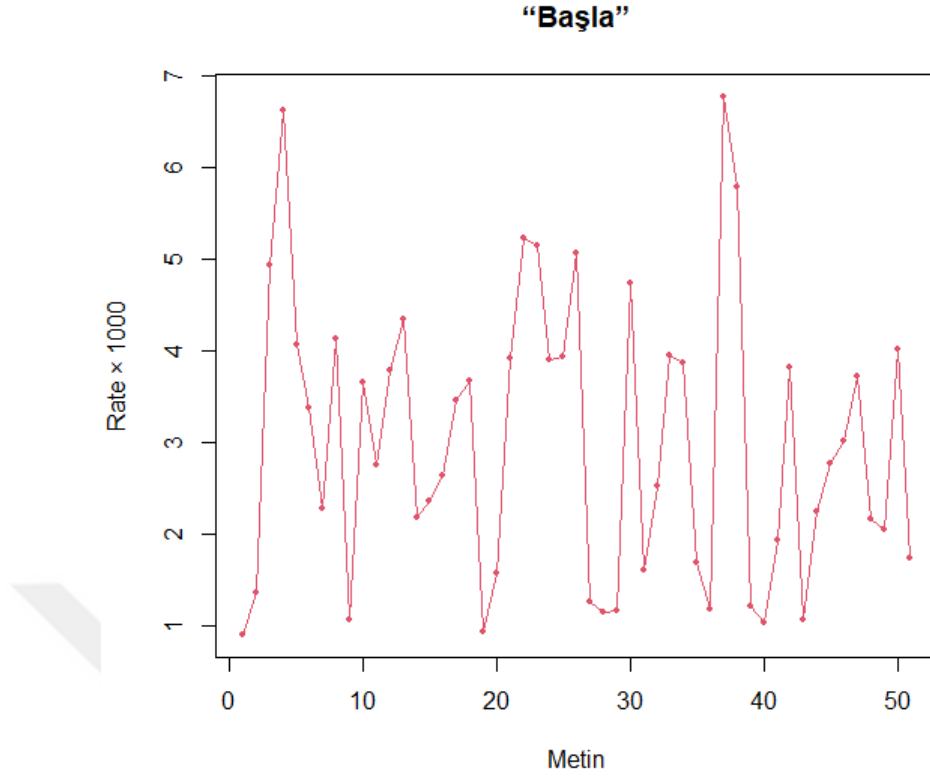
4.2.4. En Sık Kullanılan Duygu Kelimelerinin Doküman İçerisindeki Yerlerinin Belirlenmesi

Bu başlık altında, birinci duygu sözlüğüyle tanımlanmış olan ve metinlerde en sık kullanılan duygu kelimelerinin metinler arasındaki değişimi incelenmiştir.



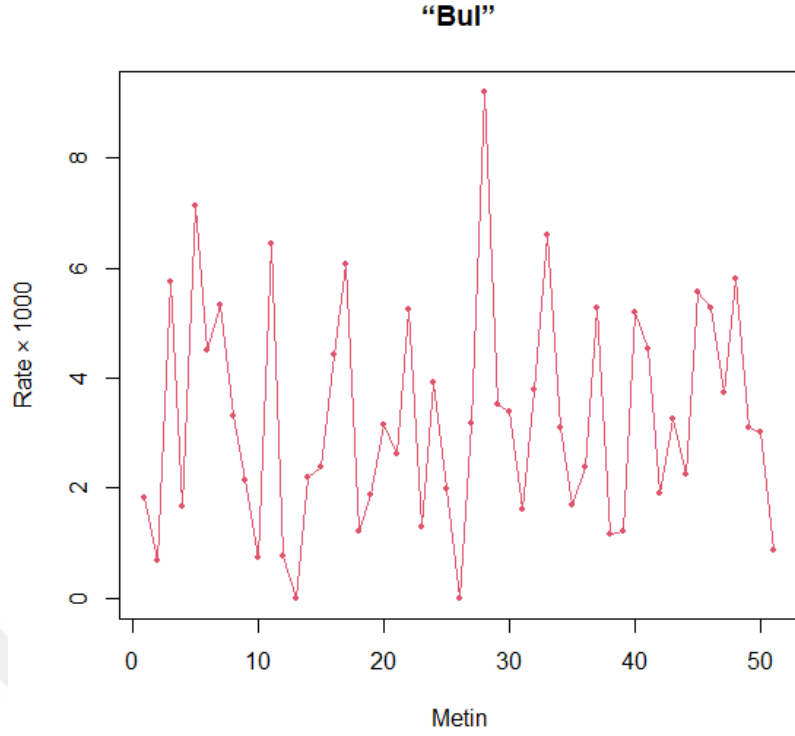
Şekil 4.3. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Gör-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.3’te “gör” kelimesinin 29. ve 45. hikâye dışındaki bütün hikâyelerde kullanıldığı görülmektedir. Ek olarak kelimenin 30. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



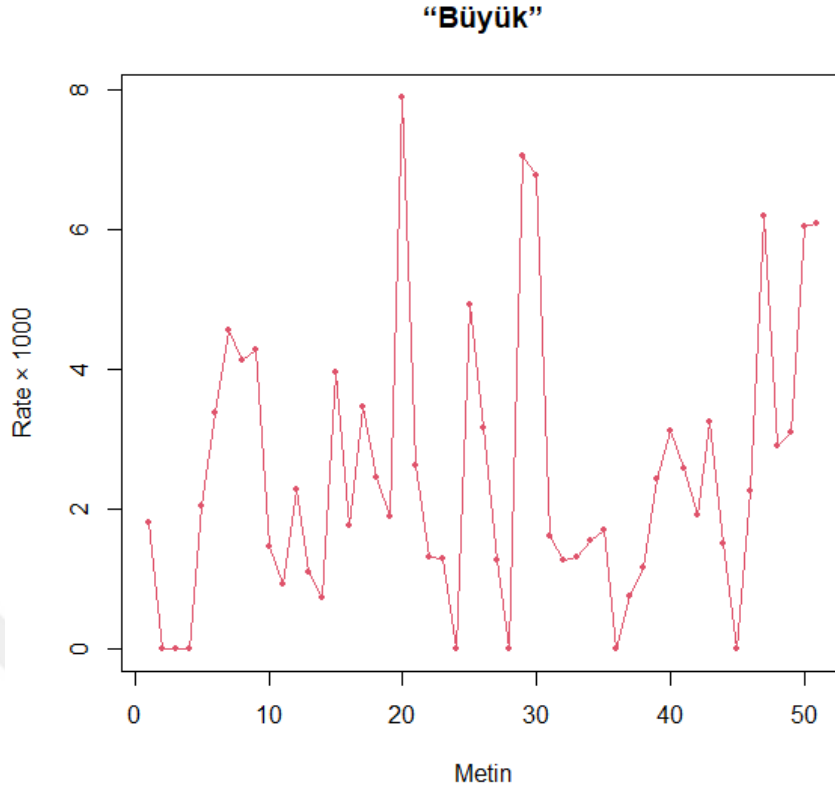
Şekil 4.4. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Başla-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.4’te “başla” kelimesinin 1. ve 19. hikâye dışındaki bütün hikâyelerde kullanıldığı görülmektedir. Ek olarak kelimenin 37. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



Şekil 4.5. Birinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Bul-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.5’te “bul” kelimesinin 13. ve 26. hikâye dışındaki bütün hikâyelerde kullanıldığı görülmektedir. Ek olarak kelimenin 28. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



 ekil 4.6. Birinci Duygu S zl    Kullanılarak Yapılan Analiz Sonucuna G re Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “B y k” Kelimesinin Metinlerin Her Biri  cerisindeki Kullanım Sıklı ının Belirlenmesi

 ekil 4.6’da “b y k” kelimesinin 2., 3., 4., 24., 28., 36. ve 45., hik ye dı ındaki b t n hik yelerde kullanıldı ı g r lmektedir. Ek olarak kelimenin 20. hik yede en y ksek frekansa sahip oldu u belirgin bir  ekilde ortaya konmu tur.

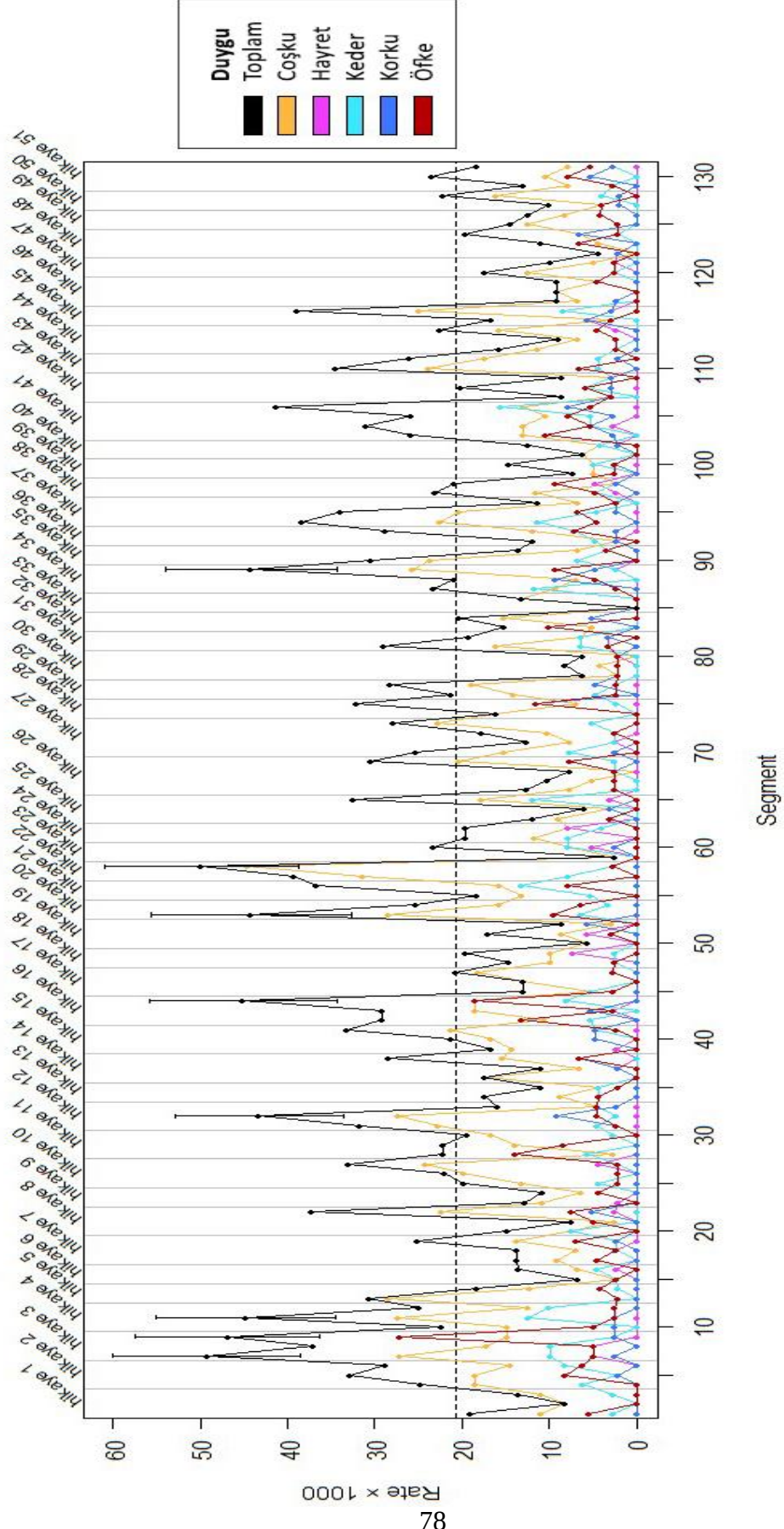
4.3. İkinci Duygu Sözlüğü ile Yapılan Duygu Analizine İlişkin İstatistik Bulgular

Bu başlık altında, metinler üzerinde ikinci duygu sözlüğü kullanılarak gerçekleştirilen sözlük tabanlı duygu analizi sonuçlarına yer verilmiştir.

4.3.1. Analiz Sonucuna İlişkin Frekans Poligonu

R yazılımında ikinci duygu sözlüğü temelinde ilgili kodlamalar yapılarak elde edilen sözlük tabanlı duygu analizi sonuçlarına ilişkin frekans poligonuna Şekil 4.7’ de yer verilmiştir.

Grafikte, metinlerin 500 tokenlik parçalara ayrılması işleminin ardından toplamda 130 bölüm elde edildiği görülmektedir. Ayrıca total-toplam kategorisindeki eğriler ve eğri üzerindeki noktalar incelendiğinde; ortalamanın üzerinde olan ve duygu içeren kelime sayısı toplamı bakımından diğerlerine göre belirgin bir farklılık ortaya koyan ve 3., 4., 12., 16., 20., 22., ve 34. hikâyelerde yer almak üzere toplamda 8 hikâye bölümü tespit edilmiştir.



Şekil 4.7. İkinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonucuna İlişkin Frekans Poligonu

4.3.2. Analiz Sonucuna İlişkin Kelime Bulutu

Bu başlık altında, R yazılımında ikinci duygu sözlüğü temelinde ilgili kodlamalar yapılarak elde edilen sözlük tabanlı duygu analizi sonuçlarına ilişkin kelime bulutuna yer verilmiştir.



Şekil 4.8. İkinci Duygu Sözlüğü Göre Yapılan Duygu Analizi Sonucuna İlişkin Kelime Bulutu

İkinci duygu sözlüğünde tanımlanan duygu kelimelerine göre belirlenen ve Ömer Seyfettin'in analize dahil edilen 51 hikâye metninin tamamında sıkça geçen ilk 50 duygu kelimesi Şekil 4.8'de verilen kelime bulutunda gösterilmiştir. Kelime bulutu incelendiğinde en yüksek frekansa sahip kelimenin “gül” fiili olduğu görülmektedir. “Gül” fiilini sırasıyla; “ask”, “sev” ve “acı” kelimeleri takip etmektedir.

4.3.3. Yıllara Göre Metinlerde En Çok Kullanılan Duygu Kelimelerinin Belirlenmesi

Bu başlık altında Ömer Seyfettin’in 51 hikâye metnini ortaya koyduğu yıllar ve metinler içerisinde ikinci duygu sözlüğüne göre tanımlanarak en sık kullandığı duygu kelimeleri olduğu kanısına varılan kelimeler incelenmiştir.

Bu doğrultuda Tablo 4.4’de aynı yıl içerisinde yazılan metinler bir grubu oluşturacak şekilde; metinlere, yazıldığı yıla ve sırasıyla sıkça kullanılan duygu kelimelerine yer verilmiştir.

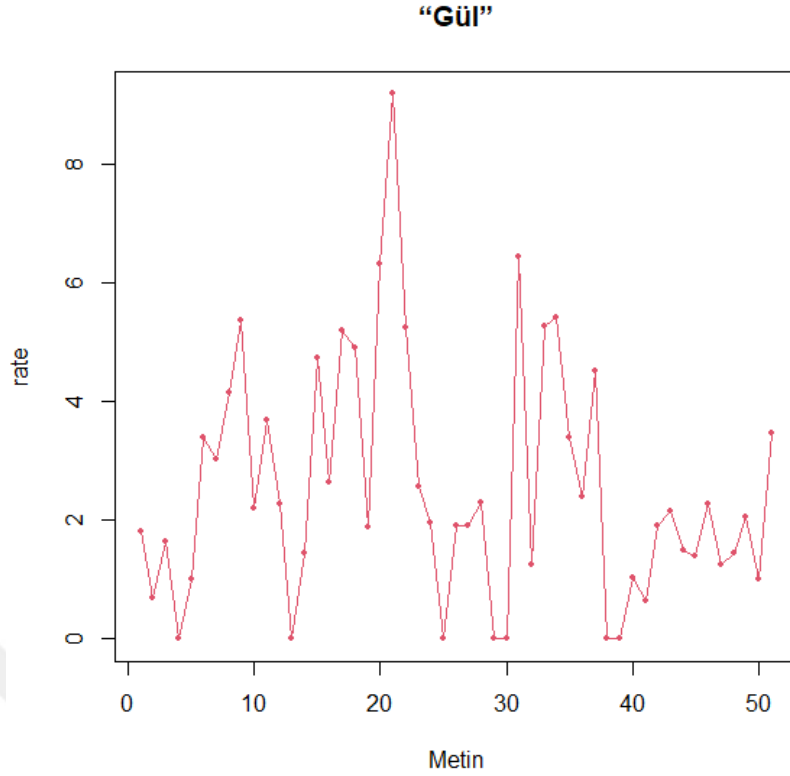
Tablo 4.4. *Yazıldıkları Yıllara Göre Gruplanan Metinlerde Sıkça Kullanılan Duygu Kelimeleri*

Hikâye	Yıl	Metinlerde En Sık Kullanılan Duygu Kelimeler
1, 2	1905	ask, sev, gül, nazik, acı, merhamet, sükun, mahcup, nefret, tehalük
3	1907	bedbaht, mesut, hiddet, minnettar, sevgili, vahşet, asabi, fena, gül, mahcup, memnun, sakın
4	1909	matem, sükun, tehalük, sakın, aci, asabi, ask, ciddi, hicran, mesut, nazik, tereddüt
5	1910	sev, aşk
6-9	1914	gül, hiddet, mesut, küstah, sevgili, ciddi, daril, esef, sakın, sas, sev , fena, korku
10-13	1917	gül, hiddet, sev, aci, fena, sevinç, ümit, ask, mesut, sakın, asabi, alem, istirap
14- 21	1918	gül, sev, hurmet, aci, sakın, sas, daril, azim, ask, heyecan, asabi, fena, hoş, korku, huzur
22-43	1919	gül, sev, fena, aci, korku, sakın sas, ask, heyecan, hiddet, ümit, hoş, dehşet, sükun, elem, sevinç, ciddi
44- 47	1920	gül, hoş, sas, sev, fena, sevgili, ask, cesaret, ümit, aci, fazilet, hiddet, merhamet rezalet, sakın, sefkat
48	1927	dehşet, fena, gül, keyif
49-51	1947	gül, fena, hoş, sev, sükun, itimat, evham, hırçın, intikam

Tablo 4.4 incelendiğinde, grup içerisindeki eleman sayısının en yüksek olduğu 1918 ve 1919 yıllarında, sıkça kullanılan kelimelerin (gül, sev, aşk, fena ve diğerleri) kelime bulutu yardımıyla elde edilen ve metinlerin genelinde frekansı yüksek olan duygu kelimeleri ile uyduğu görülmektedir.

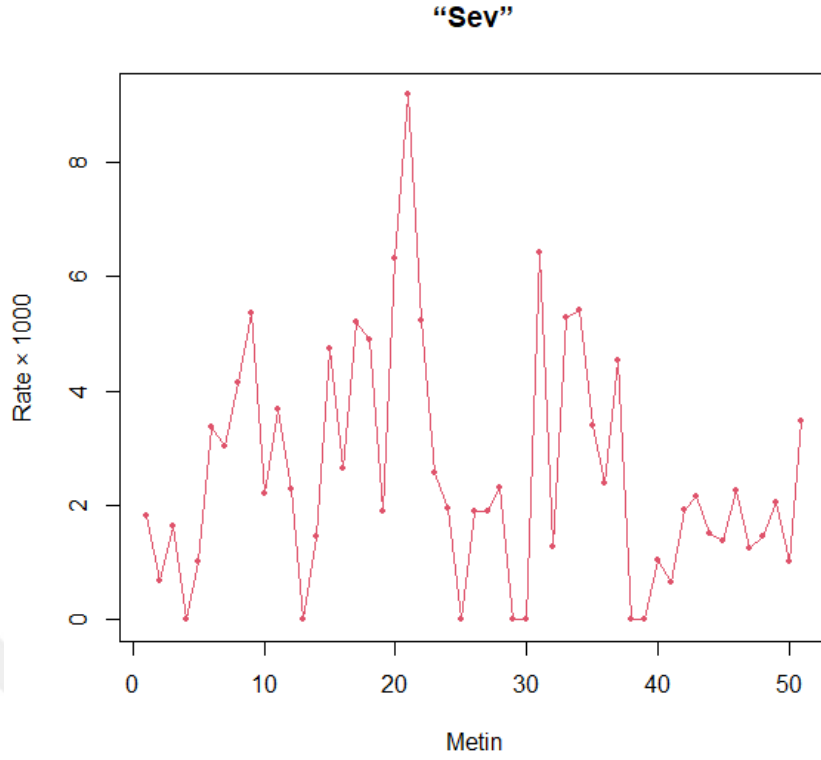
4.3.4. En Sık Kullanılan Duygu Kelimelerinin Metin İçerisindeki Yerlerinin Belirlenmesi

Bu başlık altında, ikinci duygu sözlüğüyle tanımlanmış olan ve metinlerde en sık kullanılan duygu kelimelerinin metinler arasındaki değişimi incelenmiştir.



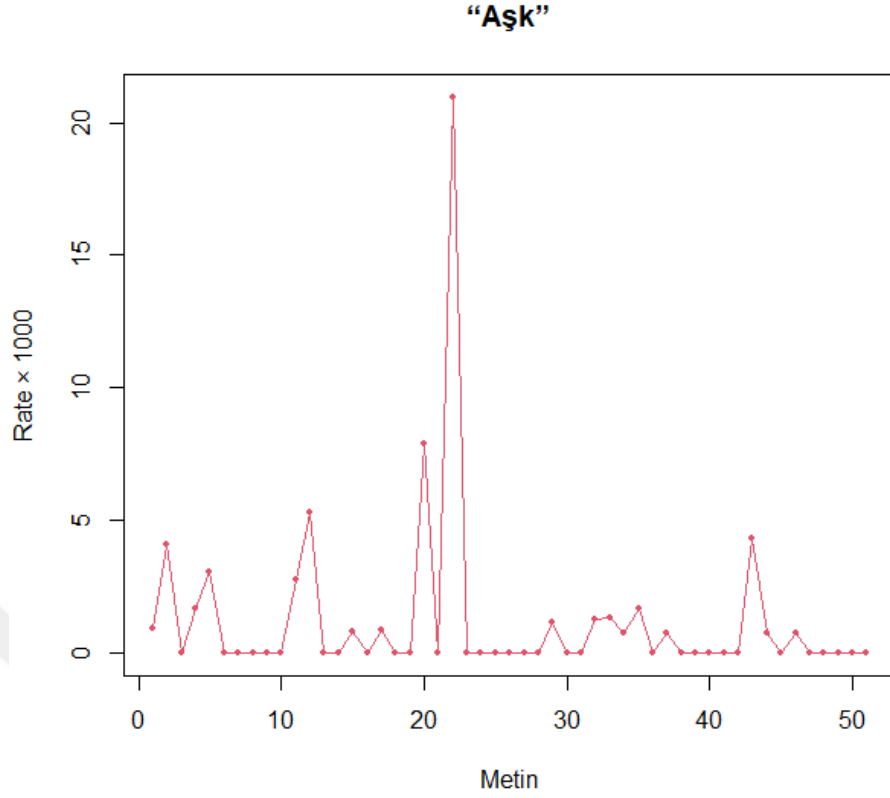
Şekil 4.9. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Gül-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.9’da “gül” kelimesinin 4., 13., 25., 29., 30., 38., ve 39. hikâye dışındaki bütün hikâyelerde kullanıldığı görülmektedir. Ek olarak kelimenin 21. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



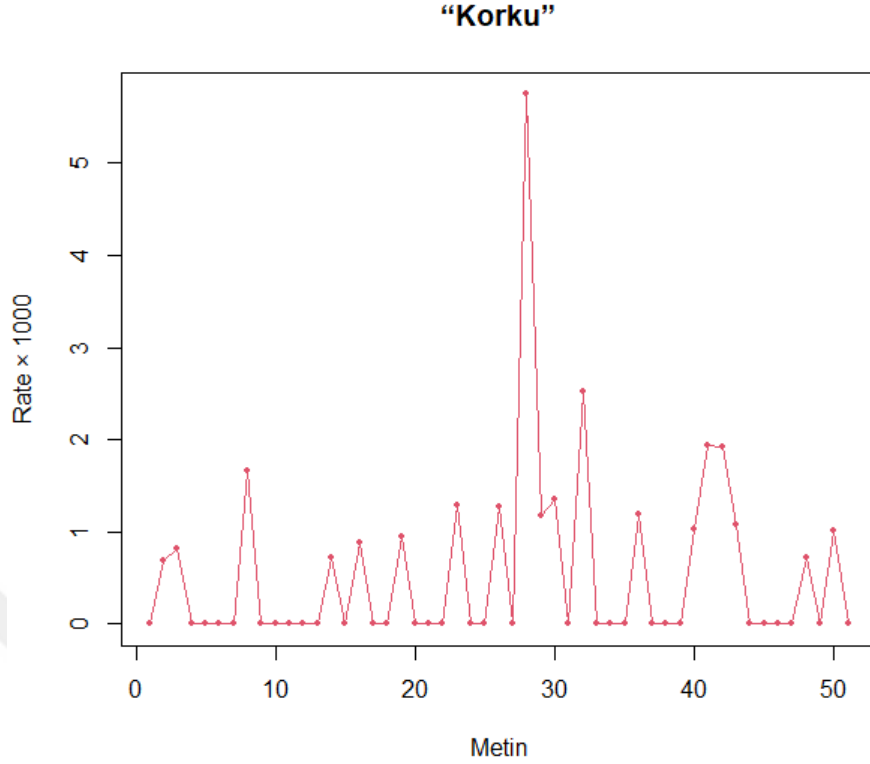
Şekil 4.10. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Sev-” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.10’da “sev” kelimesinin 4., 13., 25., 29., 30., 38., ve 39. hikâye dışındaki bütün hikâyelerde kullanıldığı görülmektedir. Ek olarak kelimenin 21. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



Şekil 4.11. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Aşk” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.11’de “aşk” kelimesinin 51 hikâyenin yarısından fazlasında kullanılmadığı görülmektedir. Ayrıca kelimenin 22. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.



Şekil 4.12. İkinci Duygu Sözlüğü Kullanılarak Yapılan Analiz Sonucuna Göre Metinlerde En Sık Kullanılan Duygu Kelimesi Olan “Korku” Kelimesinin Metinlerin Her Biri İçerisindeki Kullanım Sıklığının Belirlenmesi

Şekil 4.12’de “korku” kelimesinin, 51 hikâyenin yarısından fazlasında kullanılmadığı görülmektedir. Ayrıca kelimenin 28. hikâyede en yüksek frekansa sahip olduğu belirgin bir şekilde ortaya konmuştur.

4.4. Birinci ve İkinci Sözlüğü ile Yapılan Duygu Analizi Sonuçlarının Karşılaştırılması

Birinci ve ikinci duygu sözlüğü kullanılarak 51 hikâye metni üzerinde gerçekleştirilen duygu analizi sonuçları Tablo 4.5’te gösterilmiştir.

Tablo 4.5. Birinci ve İkinci Duygu Sözlüğü ile Yapılan Duygu Analizi Sonuçları

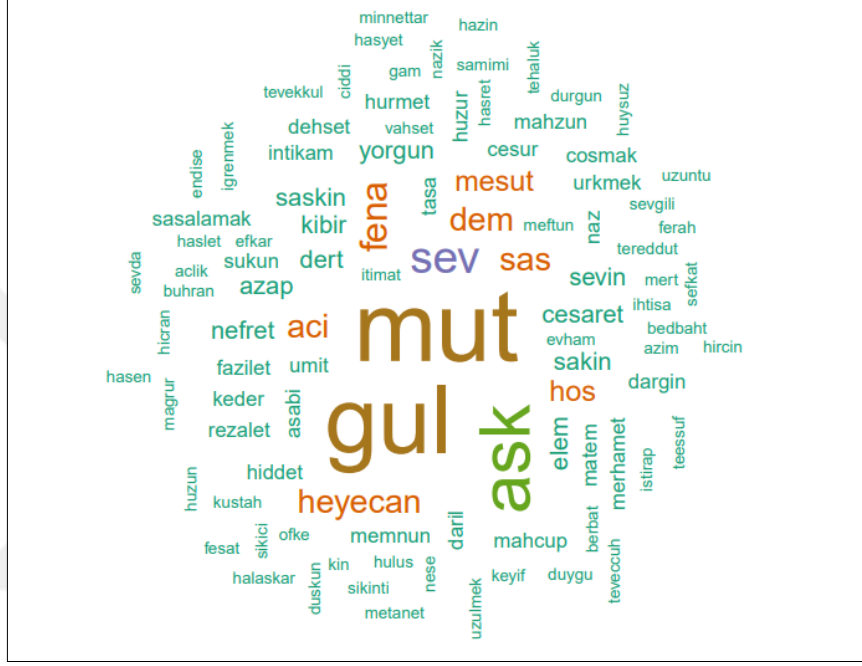
Tespit Edilen Duygu Kelimesi Sayısı									
1.Sözlük				2.Sözlük					
Metinler	Pozitif	Negatif	Toplam	Pozitif		Negatif			Toplam
				Coşku	Keder	Korku	Hayret	Öfke	
1. Hikâye	152	75	227	10	1	0	0	2	13
2. Hikâye	168	122	290	19	7	1	1	7	35
3. Hikâye	153	114	267	24	8	2	3	13	50
4. Hikâye	151	128	279	20	7	2	0	2	31
5. Hikâye	113	55	168	12	1	0	0	3	16
6. Hikâye	124	47	171	4	3	0	0	1	8
7. Hikâye	165	104	269	13	1	1	1	6	22
8. Hikâye	189	95	284	13	3	2	1	5	24
9. Hikâye	118	53	171	8	0	0	2	2	12
10. Hikâye	225	122	347	26	2	0	0	2	30
11. Hikâye	152	72	224	9	3	0	0	7	19
12. Hikâye	192	117	309	17	3	0	0	5	25
13. Hikâye	84	109	193	6	2	0	1	3	12
14. Hikâye	171	142	313	18	0	4	1	2	25
15. Hikâye	188	95	283	21	0	2	1	1	25
16. Hikâye	141	140	281	18	5	2	0	12	37
17. Hikâye	144	76	220	13	0	0	3	2	18
18. Hikâye	69	117	186	8	1	0	2	1	12
19. Hikâye	117	77	194	6	0	2	0	1	9
20. Hikâye	113	53	166	9	3	0	0	3	15
21. Hikâye	93	72	165	11	5	0	0	3	19
22. Hikâye	136	48	184	13	4	0	2	1	20
23. Hikâye	107	77	184	4	2	1	2	0	9
24. Hikâye	60	34	94	5	0	0	1	0	6
25. Hikâye	166	77	243	10	4	0	0	1	15
26. Hikâye	217	147	364	13	1	3	1	6	24
27. Hikâye	240	137	377	22	7	2	0	1	32
28. Hikâye	86	70	156	10	0	5	0	5	20
29. Hikâye	153	81	234	13	0	2	1	2	18
30. Hikâye	193	81	274	3	3	3	2	2	13
31. Hikâye	79	57	136	7	0	1	0	1	9
32. Hikâye	105	49	154	7	1	2	0	3	13

Tespit Edilen Duygu Kelimesi Sayısı									
33. Hikâye	74	53	127	4	0	0	1	0	5
34. Hikâye	184	115	299	17	0	5	0	7	29
35. Hikâye	111	40	151	8	5	0	1	1	15
36. Hikâye	83	71	154	6	3	2	0	3	14
37. Hikâye	162	111	273	21	5	1	3	6	36
38. Hikâye	88	111	199	6	5	1	0	6	18
39. Hikâye	78	89	167	4	1	0	1	2	8
40. Hikâye	106	62	168	5	2	1	1	0	9
41. Hikâye	215	120	335	14	6	7	0	7	34
42. Hikâye	122	123	245	4	2	3	1	3	13
43. Hikâye	134	69	203	15	3	1	1	3	23
44. Hikâye	146	84	230	14	1	0	3	4	22
45. Hikâye	96	94	190	10	1	2	0	1	14
46. Hikâye	180	64	244	8	0	1	1	2	12
47. Hikâye	100	80	180	7	1	0	1	2	11
48. Hikâye	167	57	224	5	1	4	2	4	16
49. Hikâye	128	55	183	10	0	0	0	3	13
50. Hikâye	142	61	203	10	2	2	0	2	16
51. Hikâye	166	78	244	10	2	2	0	6	20
Toplam	7046	4380	11426	570	117	69	41	167	964

Tabloda, metinler içerisinde ilgili sözlüklerde etiketlenmiş duygu kategorilerine göre sınıflandırılarak tespit edilen toplam kelime sayılarına yer verilmiştir. Birinci duygu sözlüğü kullanılarak yapılan sözlük tabanlı duygu analizi sonucuna göre; 51 hikâye metninin tamamında 7046'sı pozitif, 4380'i negatif yönlü olmak üzere toplamda 11,426 duygu kelimesi tespit edilmiştir. İkinci duygu sözlüğü kullanılarak yapılan sözlük tabanlı duygu analizi sonucuna göre ise; 570'i coşku, 117'si keder, 69'u korku, 41'i hayret ve 167'si öfke duygusu yönlü olmak üzere toplamda 964 duygu kelimesi tespit edilmiştir.

4.5. Ömer Seyfettin’e Ait 165 Hikâye Üzerinde Yeniden Uygulanan Sözlük Tabanlı Duygu Analizine İlişkin Kelime Bulutunun Yorumlanması

İkinci duygu sözlüğünde tanımlı duygu kelimelerine göre belirlenen ve Ömer Seyfettin’e ait 165 hikâye metninin tamamında sıkça geçen ilk 100 duygu kelimesi şeklindeki kelime bulutunda gösterilmiştir.



Şekil 4.13. Ömer Seyfettin’e Ait 165 Hikâye Üzerinde Gerçekleştirilen Duygu Analizine İlişkin Kelime Bulutu

Kelime bulutu incelendiğinde en yüksek frekansa sahip kelimenin “mut” kelimesi olduğu görülmektedir. “Mut” kelimesini sırasıyla; “gül”, “aşk” ve “heyecan” kelimeleri takip etmektedir. Mut kelimesi “mutlu, mutluluk, mutlu olmak” gibi kelime ve kelime gruplarının kökünü oluşturmaktadır. Bu duruma örnek olarak heyecan kelimesi verilebilir. Heyecan hissine sahip olan kişi nasıl “heyecanlı” olarak adlandırılıyorsa, mut hissine sahip kişi de “mutlu” olarak adlandırılmaktadır.

BÖLÜM 5

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma

Bir yazarın çokça eseri arasındaki ya da bir eserin kendi içerisindeki duygu durumu değişiminin incelenmesi şeklindeki amaçlar doğrultusunda ortaya konan bu tez çalışması, hikâye ve roman türündeki metinlerin duygu durumları açısından incelenebilirliğini, sözlük duygu tabanlı duygu analizi yönteminin bu amaçları gerçekleştirebilme konusundaki başarısını ve ilgili uygulamanın eğitim alanındaki işlevini, gerekliliğini ve uygulanabilirliğini ortaya koymaktadır. Bu amaçlar doğrultusunda, aynı zamanda bir eğitim materyali olan, Ömer Seyfettin'e ait hikâye türündeki 165 eser; sözlük tabanlı duygu analizi yöntemi ile incelenmiştir. Duygu analizinin en temel bileşenini oluşturan duygu sözlüğü, sözlük boyutu ve ayrıntı düzeyinin analize etkisinin ortaya konulması bakımından iki farklı şekilde düzenlenmiştir. Metin içerisinde duygu bildiren fiil, isim ve sıfat türündeki her kelime, alanında uzman araştırmacı tarafından belirtilen duygu yönüne göre pozitif veya negatif şekilde etiketlenerek birinci duygu sözlüğü oluşturulmuştur. Yine aynı metin içerisinde amacı bir duyguyu nitelemek olan ve belirgin bir duyguyu içeren kelimeler çeşitli duygu kategorilerinde (coşku, korku, öfke, keder, hayret) etiketlenerek ikinci duygu sözlüğü oluşturulmuştur. Devamında iki ayrı duygu sözlüğü kullanılarak, tutarlı bir analiz yapılabilmesi amacıyla analiz öncesinde belirli kriterlere göre seçilen 51 hikâye üzerinde sözlük tabanlı duygu analizi gerçekleştirilmiştir. Yapılan duygu analizinin sonuçları; kelime bulutu ve frekans poligonu şeklindeki metin madenciliği temelli analiz görselleştirme araçları yardımıyla, kullanılan sözlüğe göre ayrı ayrı yansıtıldığı için bulgulara ilişkin sonuçlar da öncelikle kelime bulutları ve frekans poligonları temelinde açıklanmıştır. Bu doğrultuda;

Birinci duygu sözlüğü kullanılarak yapılan duygu analizine ilişkin frekans poligonu sonuçları incelendiğinde; sahip oldukları bölüm sayısının 4 olduğunun görüldüğü 26. ve 27. hikâyenin, aynı zamanda en çok tokene ve token çeşitliğine sahip hikâyeler oldukları yorumu yapılabilmektedir. Benzer şekilde en az token sayısına sahip olan 24. hikâyenin 2 bölüme ayrıldığı görülmektedir. Dolayısıyla henüz ortaya konan bu yorumlar dahilinde bile grafiğin analiz sonuçlarıyla örtüştüğü ve analiz sonuçlarını açık ve anlaşılır şekilde yansıttığı görülmektedir. Grafikte total-toplam kategorisi incelendiğinde; ortalamadan en az iki standart

sapma miktarı kadar saparak duygu içeren kelime sayısı toplamı bakımından diğerlerine göre belirgin bir farklılık ortaya koyan ve 9., 16., 20., 29., 41., 46., 48. hikâyelerde yer almak üzere toplamda 7 hikâye bölümü tespit edilmiştir. Bölümler içerisinde pozitif ve negatif kategorisindeki çizgi noktalarının konumları ve 1000 paydasındaki oranları, bölümün baskın olan duygu kutbu hakkında bilgi vermektedir. Bu doğrultuda duygu yoğunluğu bakımından anlamlı bir farklılık gösteren 9, 16., 20., 29., 41., 46. ve 48. hikâyelerdeki bölümler incelendiğinde tamamının pozitif bir duygu yönüne sahip olduğu görülmektedir. Toplam duygu kelimesi miktarı bakımından, ortalamada veya ortalamanın altında seyir eden hikâye bölümlerinin de ayrıca incelendiğinde genel olarak pozitif duygu kelimesi oranlarının daha yüksek olması neticesinde pozitif bir duygu yönüne sahip oldukları görülmektedir. Ayrıca belirtmek gerekirse ise 3.,13.,16.,18., 28., 38., 39., 41.,42.,45.,47. hikâyelerde, negatif duygu kelimesi oranının pozitif duygu kelimesi oranından daha yüksek olduğu bölümlerin yer aldığı görülmektedir. Bu bölümler arasında ise negatif duygu kelimesi oranının pozitif duygu kelimesi oranından daha belirgin bir şekilde yüksek olduğu bölümlerin 13., 18., ve 28. hikâyelerde yer aldığı görülmektedir. Hikâyelerin kendi içerisinde 500 tokenlik bölümlere ayrılarak analiz edilmesi; birbirleri arasındaki duygu değişimlerinin ve karşılaştırmaların ortaya konulması haricinde, kendi içerisindeki duygu değişimleri hakkında bilgi edinilmesi imkanını sağlamaktadır. Bu doğrultuda; örnek olarak 16. hikâye incelendiğinde 1. ve 2. bölümünün pozitif, 3. bölümünün ise negatif bir duygu yönüne sahip olduğu görülmektedir.

İkinci duygu sözlüğü kullanılarak yapılan duygu analizine ilişkin frekans poligonu sonuçları incelendiğinde; ortalamadan en az iki standart sapma miktarı kadar saparak duygu içeren kelime sayısı toplamı bakımından diğerlerine göre belirgin bir farklılık ortaya koyan ve 3., 4., 12., 16., 20., 22., ve 34. hikâyelerde yer almak üzere toplamda 8 hikâye bölümü tespit edilmiştir. Bölümler içerisinde coşku, hayret, keder, korku ve öfke kategorisindeki çizgi noktalarının konumlarının ve 1000 paydasındaki oranlarının, bölümün baskın olan duygu kutbu hakkında bilgi verdiği yargısı doğrultusunda; duygu yoğunluğu bakımından anlamlı bir farklılık gösteren 3. hikâyenin birinci bölümünün ve 4., 12., 20., 22., 34. hikâyelerdeki bölümlerin coşku; 3.hikâyenin ikinci bölümünün ve 16. hikâyedeki 3. bölümün öfke yönünde duygu kutbuna sahip olduğu görülmektedir. Toplam duygu kelimesi miktarı bakımından, ortalamada veya ortalamanın altında seyir eden hikâye bölümlerinin de ayrıca incelendiğinde genel olarak coşku duygusu içeren kelimelerin oranlarının daha yüksek olması neticesinde coşku ağırlıklı bir duygu yönüne sahip oldukları görülmektedir. Duygu kutupları ayrı ayrı ele alındığında,

bölümler arasında coşku duygusunu içeren kelimelerin oranının en yüksek olduğu bölümün 22. hikâyede yer aldığı; öfke duygusunu içeren kelimelerin oranının en yüksek olduğu bölümün 3. hikâyede yer aldığı; korku duygusunu içeren kelimelerin oranının en yüksek olduğu bölümün 12. ve 34. hikâyelerde yer aldığı; keder duygusunu içeren kelimelerin oranının en yüksek olduğu bölümün 41. hikâyede yer aldığı; hayret duygusunu içeren kelimelerin oranının en yüksek olduğu bölümün 18. ve 24. hikâyelerde yer aldığı görülmektedir. Hikâyelerin kendi içerisinde 500 tokenlik bölümlere ayrılarak analiz edilmesi; birbirleri arasındaki duygu değişimlerinin ve karşılaştırmaların ortaya konulması haricinde, kendi içerisindeki duygu değişimleri hakkında bilgi edinilmesi imkanını sağladığı bilgisi doğrultusunda; birinci duygu sözlüğü ile yapılan analizde de örnek olarak verilen 16. hikâye incelendiğinde 1. ve 3. bölümünün öfke yönünde, 2. bölümünün ise coşku yönünde bir duygu kutbuna sahip olduğu görülmektedir.

Birinci duygu sözlüğü ile yapılan duygu analizi sonucuna ilişkin kelime bulutu incelendiğinde; en yüksek frekansa sahip kelimenin “gör” fiili olduğu tespit edilmiştir. Gör” fiilini sırasıyla; “büyük”, “başla” ve “bul” kelimeleri takip etmiştir. Birinci duygu sözlüğünde tanımlanan bu kelimelerin pozitif kelimeler olduğu bilgisi eşliğinde genel olarak metinlerde sıkça geçen duygu kelimelerinin pozitif olduğu yorumu yapılabilmektedir.

İkinci duygu sözlüğü ile yapılan duygu analizi sonucuna ilişkin kelime bulutu incelendiğinde; en yüksek frekansa sahip kelimenin “gül” fiili olduğu tespit edilmiştir. Gül” fiilini sırasıyla; “ask”, “sev” ve “acı” kelimeleri takip etmektedir. İkinci duygu sözlüğünde tanımlanmış olması dahilinde metinde sıkça geçen kelimelerden olan gül, sev ve aşk kelimelerinin coşku yönünde bir duygu kutbuna sahip olduğu; acı kelimesinin ise keder yönünde bir duygu kutbuna sahip olduğu yorumu yapılabilmektedir.

Her iki duygu sözlüğü temelinde, kelime bulutu yardımıyla, 51 hikâye metninin tamamında en sık geçen duygu kelimeleri oldukları belirlenen kelimeler; metinlerin her biri içerisindeki kullanım sıklıkları bakımından ayrıca incelenmiştir. Yapılan bu incelemeler neticesinde; sırasıyla “aşk” ve “korku” kelimelerinin hikâyelerin yarısından fazlasında yer almamalarına rağmen, en sık kullanılan duygu kelimeleri içerisinde dahil olmalarının nedeninin, 22. ve 28. hikâyede genelin üstündeki yüksek frekanslar ile aykırı bir değer oluşturmalarının olduğu tespit edilmiştir.

Tüm bu analize ek olarak Ömer Seyfettin'e ait 51 hikâye metninin tamamının yazım yılı dikkate alınarak, metinlerde sıkça kullanılan duygu kelimeleri ile ilgili metnin yazıldığı tarihte yazarın yaşadıklarının, psikolojik özelliklerinin ve yazarın yaşadığı tarihteki olaylar arasındaki ilişki incelenmiştir. Bu doğrultuda en sık kullanılan her bir duygu kelimesinin metinler içerisindeki sıklıkları hesaplanmıştır. Eserlerin büyük çoğunluğunun 1917 ve 1920 yıllarında yazıldığı bilgisi eşliğinde; her iki duygu sözlüğüne göre sıkça kullanıldığı belirlenen (birinci sözlük: gör, bul, başla, büyük; ikinci sözlük: gül, aşk, sevgi) duygu kelimelerinin tamamının, bu yıllar arasında yazılan eserlerin tamamında da sıkça kullanılan duygu kelimeleri olduğu tespit edilmiştir.

Birinci ve ikinci duygu sözlüğü ile gerçekleştirilen sözlük tabanlı duygu analizi sonuçlarının, oluşturulan frekans poligonları ve kelime bulutları yardımıyla incelenmesinin ardından; ilgili sözlüklerin kelime boyutlarının ve sözlüklerdeki kelime türlerinin aynı metinler üzerindeki duygu analiz sonuçlarına etkisi karşılaştırmalı olarak incelenmiştir. Bu doğrultuda, birinci duygu sözlüğü ile gerçekleştirilen sözlük tabanlı duygu analizi sonuçlarına göre; 38. ve 42. hikâyeler haricindeki hikâyelerde genel olarak pozitif kutuplu duygu kelimesi sıklığının daha yüksek olduğu tespit edilmiştir. Pozitif ve negatif yönlü duygu kutbuna ek olarak; coşku, keder, korku, hayret ve öfke duygusu şeklinde, ayrıntılı duygu etiketi bulunan ikinci duygu sözlüğü ile gerçekleştirilmiş sözlük tabanlı duygu analizi sonuçları iki açıdan değerlendirilmiştir. Beş ana duygu ele alınarak yapılan incelemeye göre; 51 hikâye metninin tamamında coşku duygusunu içeren duygu kelimesi sayısı daha fazladır. Beş ana duygu, 51 hikâye metninin genelinde toplam kullanım sıklığı bakımından incelendiğinde, sırasıyla en sık kullanılan duygu kelimesinin coşku, öfke, keder, korku ve hayret şeklinde olduğu görülmektedir. İkinci duygu sözlüğü, polarite bakımından değerlendirildiğinde ise negatif (öfke, keder, korku ve hayret) duygu yönüne sahip duygu kelimesi sayısı daha yüksek olan ve dolayısıyla negatif duygu yönüne sahip olduğu görülen 9 hikâye (3., 11., 16., 30., 36., 38., 41., 42. ve 48.hikâye) tespit edilmiştir. Birinci sözlük ile negatif duygu yönüne sahip olduğu belirlenen 38. ve 42. hikâye, ikinci duygu sözlüğü kullanılarak yapılan duygu analizi sonucuna göre yeniden değerlendirildiğinde; 38. ve 42. hikâyenin, ikinci duygu sözlüğü kullanılarak gerçekleştirilen duygu analizi sonuçlarına göre de negatif duygu yönüne sahip olduğu görülmektedir (38.hikâye: 11 negatif, 6 pozitif kelime; 42.hikâye: 9 negatif, 4 pozitif kelime). Her iki duygu sözlüğü birlikte değerlendirildiğinde; birinci duygu sözlüğü kullanılarak yapılan duygu analizi ile 51 hikâye metninin genelinde tespit edilen toplam duygu kelimesi sayısı, ikinci

duygu sözlüğü kullanılarak tespit edilen toplam duygu kelimesi sayısından kayda değer şekilde daha fazladır. Bu sonuç, ilgili duygu sözlüklerinin boyutları ile örtüşmektedir.

Daha dengeli ve tutarlı bir karşılaştırma yapılabilmesi amacıyla özellikle seçilmiş (eşit veya yakın sayfa sayısı) 51 hikâye metni üzerinde gerçekleştirilen sözlük tabanlı duygu analizi, analize daha önce dahil edilmemiş diğer hikâyelerinde de eklenmesiyle toplamda 165 hikâye üzerinde yeniden uygulanmış ve sonuçlar kelime bulutu ile görselleştirilmiştir. Kelime bulutu incelendiğinde en yüksek frekansa sahip kelimenin “mut” kelimesi olduğu tespit edilmiştir. “Mut” kelimesini sırasıyla; “gül”, “aşk” ve “heyecan” kelimeleri takip etmektedir. Mut kelimesi “mutlu, mutluluk, mutlu olmak” gibi kelime ve kelime gruplarının kökünü oluşturmaktadır. Bu duruma örnek olarak heyecan kelimesi verilebilmektedir. Heyecan hissine sahip olan kişi nasıl ki “heyecanlı” olarak adlandırılıyorsa, mut hissine sahip kişi de “mutlu” olarak adlandırılmaktadır.

Metinlerde sıkça kullanılan duygu kelimeleri yardımıyla; bir yazarın kronolojik sıradaki her bir eserinin duygu yönünün ve dolayısıyla yazarın genel duygu durumu değişiminin tespit edilmesi amacıyla yapılan bu sözlük tabanlı duygu analizi neticesinde; Ömer Seyfettin’ in hikâyelerinin yoğun şekilde duygu kelimesi içerdiği tespit edilmiştir.

Hikâyeler hem çocuklar hem yetişkinler için kalıcı, eğitici ve eğlendirici bir niteliğe sahip metin kaynaklarıdır. Bir olayı, durumu çeşitli duygusal düzeylerde ileten en güçlü iletişim kanallarıdır. Bu doğrultuda kuşaklar boyunca farklı disiplinlerdeki çeşitli eleştirmen ve araştırmacılar tarafından ayrıntılı olarak incelenmiştir. Örneğin, Sallabaş (2012), çalışmasında Türkçe öğretimi açısından Ömer Seyfettin’in hikâyelerinin değer aktarımında ne derecede başarılı olduğunu incelemiştir. Gerçel ve Sarıçan (2011), aynı hikâyeleri dil, üslup ve çocuk eğitime uygunluğu açısından incelemiştir. Aydemir (2005) ise Ömer Seyfettin’in hikâyelerinin modern Türk hikâyeciliğine etkisini inceleyen bir çalışma oraya koymuştur. Ancak, özellikle yurt içinde yapılan çalışmalar dikkate alındığında, hikâyelerdeki duygusal içeriği tespit eden ve duyguların hikâye boyunca nasıl bir değişim gösterdiğini analiz eden çok az sayıda çalışma bulunmaktadır.

Tezde ulaşılan sonuçlar kullanılan materyal bakımından ele alındığında; edebiyat ve çocuk eğitimi gibi alanlar açısından değerli ve önemli bir eğitim-öğretim kaynağı olduğu bilinen ve Ömer Seyfettin tarafından kaleme alınan hikâyelerin, duygu içeren kelime sayısı bakımından zengin olduğu tespit edilmiştir. Keleş ve Çepni (2006)’ nin de ortaya koyduğu

üzere, duygu beyin ve öğrenme aktivitesi ile ilişki temel bir bileşenler. Bu doğrultuda duygusal açıdan zengin içeriğe sahip eğitim materyalleri, öğrenmenin etkinliğini ve kalıcılığı artıran önemli materyallerdir. Dolayısıyla yapılan sözlük tabanlı duygu analizi neticesinde, Ömer Seyfettin'in hikâyelerinin, içeriğindeki duygu kelimelerinin yoğunluğu ve duygu yönleri dikkate alınarak, eğitim ve öğretim açısından faydalı bir materyal olduğu sonucuna varılmıştır.

Tezde ulaşılan sonuçlar kullanılan yazılım bakımında ele alındığında ise; bir istatistiksel hesaplama ve görselleştirme programı olan R'nin analiz uygulama, analiz değerlendirme ve analiz sonucunu görselleştirme işlemleri bakımında ne kadar güçlü olduğu ve sanılanın aksine kolayca kullanılabilir olduğu yargısına varılmıştır. Ayrıca R'nin alan yazında alışlagelmiş, klasik işlemlere (içerik-kavram analizi vb.), uygulama sürecinde kolay, hızlı ve güvenilir bir bakış açısı kazandırılması açısından önemli bir araç olduğu ortaya konmuştur.

5.2. Öneriler

Veri Madenciliğinin bir alt dalı olan metin madenciliğine ilişkin uygulamaların, eğitimde kullanımının önemi ve gerekliliği ile ilgili olarak tezin kuramsal çerçeve ve ilgili araştırmalar kısmında da belirtildiği üzere yurt dışında çokça çalışma bulunmaktadır. Ancak yurt içindeki çalışmalarda eğitim ve öğretim alanının içinde bulunan taraflar açısından (öğrenci, öğretici, öğretim kurumu, öğretim platformu) faydaları kanıtlanmış olan bu uygulamalara gereken ölçüde yer verilmemiştir. Bu durum, metinlerdeki duygu eğilimlerinin keşfedilmesi amacını taşıyan duygu analizi uygulamaları açısından ele alındığında ise; yurt içinde, duygu analizi uygulamasının çoğunlukla ticaret ve ekonomi gibi alanlarda kullanımına ilişkin çalışmaların çoğunluğu oluşturduğu görülmektedir. Dolayısıyla ilgili uygulamanın ve uygulamanın gerçekleştirildiği yazılımın, eğitimde kolaylıkla kullanılabilir olduğunun ortaya konulması amacını taşıyan bu tez çalışmasında da belirtildiği üzere, öncelikle eğitim alanında duygu analizi ile ilgili olarak daha fazla çalışma yapılması gerektiği tavsiye edilmektedir.

Duygu analizi uygulamasının içeriğine ilişkin öneride bulunmak gerekirse;

- ❖ Türk diline özgü özellikler nedeniyle morfolojik analiz adımı ortaya çıkan zorlukların giderilmesi amacıyla, temelde R yazılımında kullanılmak üzere R programlama diliyle yazılmış bir morfolojik çözümleme aracı geliştirilmedi.
- ❖ Sözlük tabanlı duygu analizi ile ilgili çalışmaları kolaylaştırmak ve araştırma sürecini kısaltmak bakımından alan bağımlılığı olmayan, içerisinde Osmanlı Türkçesine özgü kelimelerinde bulunduğu genel bir duygu sözlüğü oluşturulmalıdır.

Duygu analizi ile ilgili gelecek araştırmalara yönelik öneride bulunmak gerekirse;

- ❖ Doğrudan çocuk eğitiminde kullanılan roman ve hikâyeler çocuğa yansıtılan duygu kutbu bakımından sözlük tabanlı duygu analizi ile incelenmelidir.
- ❖ Türkçe ders kitaplarında çeşitli temalar içerisinde yer alan metinler, öğrenci seviyesine uygunluğu ve yansıttığı duygu durumu bakımından incelenmelidir.
- ❖ Eğitim alanında kullanılan metin türündeki materyallerin, ilgili eğitim-öğretim kazanımını belirten kavramlar bakımından değerlendirildiği ya da doğrudan kazanıma uygunluğunun incelendiği çeşitli çalışmalar bulunmaktadır. Aslında bir çeşit kavram çıkarma işlevine sahip olan duygu analizindeki temel mantık, bu uygulamalarla birleştirilerek daha hızlı, kolay ve güvenilir kavram-içerik analizi çalışmaları yapılabilir.
- ❖ Eğitimde Ölçme ve Değerlendirme alanı hedef alınarak yapılan alan yazın incelemesi neticesinde; ilgili alan yazında ortaya konulan lisansüstü tezlerde genel olarak nitel araştırmalarda sıkça kullanılan doküman incelemesi gibi yöntemlerin göz ardı edildiği görülmektedir. Eğitimde önemi herkesçe bilinen ve çeşitli çalışmalarla gerekliliği ispatlanan metin türündeki eğitim materyallerinin nitelik ve eğitime uygunluk gibi kavramlar açısından Ölçme ve Değerlendirme araştırmaları dahilinde incelenmesi gerekmektedir. Bu doğrultuda nitel araştırma yöntemlerine ilişkin çeşitli uygulamalarda daha nicel ölçümler ya da analizler gerçekleştirmek adına, duygu analizi gibi güncel, nesnel ve istatistiksel açıdan doğrulanabilir analiz yöntemlerinin uygulanabileceği önerilmektedir.
- ❖ Gelişen teknolojinin beraberinde getirdiği yeni yöntem, teknik ve uygulamalar; başarı testleri gibi sonuç odaklı, geleneksel ölçme ve değerlendirme araçlarına süreç odaklı yeni

alternatifler oluşturmaktadır. Örneğin, öğretimi gerçekleştirilen bir konu hakkında, öğrenciden çeşitli zaman aralıklarında elde edilen; deneme, kompozisyon gibi metin türündeki veriler üzerinde metin madenciliği yöntemleri uygulanabilir. Otomatik olarak gerçekleştirilen bu uygulamalar ile ortaya koyduğu metinler üzerinden öğrencinin konuya hâkim olup olmama durumu hakkında kolaylıkla yorum yapılabilir. Bu yeni yöntem, teknik ve uygulamaların ölçme ve değerlendirme açısından kullanımına ilişkin daha kapsamlı bir örnek vermek gerekirse; öğrencilerin, öğretim sürecinin başında ortaya koydukları performans göstergeleri üzerinde birtakım veri ve metin madenciliği yöntemleri uygulanarak, öğretim sürecinin sonundaki performansları tahmin edilebilir. Dolayısıyla bunun gibi analiz ve tahmin uygulamaları ile öğrencilere doğru zamanda müdahale edilmesi ve eğitim desteği verilmesi gibi birçok konuda eğitim ve öğretim sürecine yardımcı olacak potansiyel bilgilerin edilebileceği önerilmektedir. Ayrıca bu bilgiler doğrultusunda eğitimcilerin, çeşitli eğitim politikaları geliştirebileceği öngörülmektedir.

KAYNAKÇA

- Abdoulaye, I. B. C. (2019). *Türkçe metinlerde sözlük tabanlı yaklaşımla duygu analizi ve görselleştirme* (Yayımlanmamış Yüksek Lisans Tezi). Marmara Üniversitesi, İstanbul.
- Acosta, O. C., Behar, P. A. ve Reategui, E. B. (2014). Content recommendation in an inquiry-based learning environment. *Paper persented at Frontiers in education conference*, IEEE, 1–6.
- Aggarwal, C. C. ve Zhai, C. (2012). *Mining text data*. Springer Science ve Business Media.
- Aghaee, N. (2015). Finding potential problems in the thesis process in higher education: Analysis of e-mails to develop a support system. *Education and Information Technologies*, 20, 21–36.
- Akba, F. (2014). *Duygu analizinde öznitelik seçme metriklerinin değerlendirilmesi: Türkçe film eleştirileri* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Akbudak, İ. (2019). *Konuşma bilgisi ve makine öğrenmesi kullanılarak duygu analizi*. (Yayımlanmamış Yüksek Lisans Tezi). Pamukkale Üniversitesi, Denizli.
- Akın, B. K ve Şimşek U. T. G (2018). Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi. *Bilişim Teknolojileri Dergisi*, 11(3), 245-253. Doi: 10.17671/gazibtd.342419
- Akısoğlu, B. (2019). *Predicting economic growth using machine learning techniques and sentiment analysis*. (Yayımlanmamış Yüksek Lisans Tezi). TED Üniversitesi, Ankara.
- Aktaş, Ş. (2009). Edebi Metin ve Özellikleri. *Türkiyat Araştırmaları Enstitüsü Dergisi*, 39, 187-200.
- Albayrak, A. (2018). *Duygu analizinde farklı vektör temsil yöntemleri ve sınıflayıcıların karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Sivas Cumhuriyet Üniversitesi, Sivas.
- Albayrak, N., B. (2011). *Opinion and sentiment analysis using natural language processing techniques*. (Yayımlanmamış Yüksek Lisans Tezi). Fatih Üniversitesi, İstanbul.

- Alm, C. O. ve Sproat, R. (2005). Emotional sequencing and development in fairy tales. *Proceedings of the first international conference on affective computing and intelligent interaction*. 668-674.
- Alm, C., O. ve Sproat, R. (2005). Emotional sequencing and development in fairy tales. *In Affective Computing and Intelligent Interaction*. Springer, 668–674.
- Amanet, H. (2017). *Türkçe sosyal medya metinlerinde duygu analizi*. (Yayımlanmamış Yüksek Lisans Tezi). Karadeniz Teknik Üniversitesi, Trabzon.
- Araki, J., Rajagopal, D., Sankaranarayanan, S., Holm, S., Yamakawa, Y. ve Mitamura, T. (2016). Generating questions and multiple-choice answers using semantic analysis of texts. *The 26th international conference on computational linguistics technical papers*, 1125–1136, Association for Computational Linguistics.
- Arguedas, M., Xhafa, F., Casillas, L., Daradoumis, T., Peña, A. ve Caballé, S. (2018). A model for providing emotion awareness and feedback using fuzzy logic in online learning. *Soft Computing*, 22(3), 963–977.
- Aung K., Z. ve Myo N., N. (2017). Sentiment analysis of students' comment using lexicon based approach, *16th International Conference on Computer and Information Science (ICIS)*, 2017, 149-154.
- Aydemir, C. (2005). Modern Türk Hikâyeciliğinde Ömer Seyfettin Etkisi, *Balıkesir Üniversitesi Sosyal Bilimler Dergisi*, 8(13), 54-62.
- Azevedo, B. F. T., Behar, P. A. ve Reategui, E. B. (2011). Automatic analysis of messages in discussion forums. *In 14th international conference on interactive collaborative learning (ICL)*, 76–81.
- Baccianella, S., Esuli, A. ve Sebastiani. F., (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *In LREC*, (10), 2200–2204.
- Baker, R. S. ve Inventado, P. S. (2014). Educational data mining and learning analytics. *In Learning analytics*, Springer, 61–75.

- Baker, R. S. ve Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.
- Baker, R. S. ve Yacef, K. (2010). *Data mining for education*. In McGaw B, Peterson P, Baker E, eds. International Encyclopedia of Education. 3rd ed. Vol. 7. Oxford, UK: Elsevier, 112–118.
- Bala, M. ve Ojha D. B. (2012). Study of applications of data mining techniques in education. *International J Res Sci Technol*, 1: 1–10.
- Baltacı, A. (2019). Nitel Araştırma Süreci: Nitel Bir Araştırma Nasıl Yapılır?. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(2), 368-388.
- Balyan, R., McCarthy, K. S. ve McNamara, D. S. (2017). Combining machine learning and natural language processing to assess literary text comprehension. Grantee Submission.
- Barbosa, L. ve Feng, J. (2010). Robust sentiment detection on Twitter from biased and noisy data. *In Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10)*, 36–44.
- Berry, M. W. ve Castellanos, M. (2004). Survey of text mining. *Computing Reviews*, 45, 548.
- Bhonde, R., Bhagwat, B., Ingulkar, S. ve Pande, A. (2015). Sentiment Analysis Based on Dictionary Approach. *Int. J. Emerg. Eng. Res. Technol.*, 3(11). 51–55.
- Bing Liu. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.
- Blitzer, J., Dredze, M. ve Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *In ACL*, (7), 440–447.
- Boiy, E. ve Moens, F. (2008). A machine learning approach to sentiment analysis in multilingual Web texts, *Inform. Retrieval* 12 (5), 526–558.
- Boynukalın, Z. (2012). *Emotion analysis of Turkish texts by using machine learning methods* (Yayımlanmamış Yüksek Lisans Tezi). ODTÜ, Ankara.

- Bozkaplan, Ş. A., Kubilay, İ. A. ve Karademirlidağ. Y. (2016). Mukaddime ve Piyale Önsözü'nden Yola Çıkararak Türkçede Duygu Sözlüğü Oluşturulması Üzerine Metin Tabanlı Bir Deneme. III. Uluslararası Sözlükbilimi Sempozyumu
- Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K. ve Reimann, P. (2011). Collaborative writing support tools on the cloud. *Transactions on Learning Technologies, IEEE*, 4, 88–97.
- Caspi, A., Gorsky, P. ve Chajut, E. (2003). The influence of group size on nonmandatory asynchronous instructional discussion groups. *The Internet and Higher Education*, 6, 227–240.
- Chakrabarti, S. (2002). *Mining the Web: Analysis of Hypertext and Semi-Structured Data*, Morgan Kaufmann, San Francisco.
- Chakraborty K., Bhattacharyya S., Bag R. ve Hassanien A., A. (2019). Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Techniques. *Social Network Analytics: Computational Research Methods and Techniques*. Academic Press. 127-147.
- Chakraborty K., Bhattacharyya S., Bag R. ve Hassanien, A. E. (2018). *Comparative Sentiment Analysis on a Set of Movie Reviews Using Deep Learning Approach*. In: Hassanien A., Tolba M., Elhoseny M., Mostafa M. (Ed.) The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018). AMLTA 2018. Springer, Cham. doi.org/10.1007/978-3-319-74690-6_31
- Chen, N.-S., Wei, C.-W. ve Chen, H.-J. (2008). Mining e-learning domain concept map from academic articles. *Computers & Education*, 50, 1009–1021.
- Chinkina, M., Oswal, A. ve Meurers, D. (2018). Automatic input enrichment for selecting reading material: An online study with english teachers. *In Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, 35–44.
- Chowdhury, G. G. (2003). Natural language processing. *In Annual review of information science and technology* (Vol. 37, pp. 51–89).

- Cobo, G., García-Solórzano, D., Santamaría, E., Morán, J. A., Melenchón, J. ve Monzo, C. (2010). Modeling students' activity in online discussion forums: A strategy based on time series and agglomerative hierarchical clustering. *In Educational data mining 2011*.
- Coutinho, E., Moreira, L., Paillard, G. ve de Lima, E. T. (2016). Análise do sentimento de mensagens de chats em uma turma de graduação de um curso de educação à distância. *In Anais dos Workshops do Congresso Brasileiro de Informática na Educação*.
- Crawley, M. J. (2007). *The R Book*. Wiley. Chichester.
- Croft, B., Metzler, D. ve Strohman, T. (2009). *Search Engines: Information Retrieval in Practice*. Addison-Wesley. Boston, MA.
- Crossley, S. A., Varner, L. K., Roscoe, R. D. ve McNamara, D. S. (2013). Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. *In International conference on artificial intelligence in education*, Springer, 269–278.
- Crossley, S., Allen, L. K., Snow, E. L. ve McNamara, D. S. (2015). Pssst... textual features... there is more to automatic essay scoring than just you! *In Proceedings of the fifth international conference on learning analytics and knowledge*, ACM, 203–207.
- Cunningham-Nelson, S., Baktashmotlagh, M. ve Boles, W. (2019). Visualizing student opinion through text analysis. *IEEE Transactions on Education*, 62(4), 305–311.
- Cutrone, L. A. ve Chang, M. (2010). Automarking: Automatic assessment of open questions. *In IEEE 10th international conference on advanced learning technologies (ICALT)*, IEEE, 143–147.
- Çalışkan, S. (2020). *Text mining analysis of translation, social communication and literary writing for turkish*. (Yayımlanmamış Yüksek Lisans Tezi). Bilkent Üniversitesi, Ankara.
- Çetin, F. S. (2017). *Türkçe hedef tabanlı duygu analizi için alt görevlerin incelenmesi – hedef terim, hedef kategori ve duygu sınıfı belirleme* (Yayımlanmamış Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, İstanbul.

- da Costa Pinho, I., Epstein, D., Reategui, E. B., Correa, Y. ve Polonia, E. (2013). The use of text mining to build a pedagogical agent capable of mediating synchronous online discussions in the context of foreign language learning. *In 2013 IEEE Frontiers in education conference*, San Jose, CA: IEEE, 393–399.
- Das, A. ve Bandyopadhyay, S. (2010). SentiWordNet for indian languages. *Asian Federation for Natural Language Processing*. China, 56–63.
- Dascalu, M., Trausan-Matu, S., Dessus, P. ve McNamara, D. S. (2015). *Discourse cohesion: A signature of collaboration*. In Proceedings of the fifth international conference on learning analytics and knowledge, ACM, 350–354.
- de Micheaux, P. L., Drouilhet, R. ve Liquet B. (2013). *The R Software: Fundamentals of Programming and Statistical Analysis*. Statistics and Computing. Springer Science+Business Media, New York.
- De Wever, B., Schellens, T., Valcke, M. ve Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46, 6–28.
- Değer, N. S. (2017). *Sosyal medya mesajlarında veri madenciliği ile duygu analizi* (Doktora Tezi). İstanbul Üniversitesi, İstanbul.
- Dehkhargani, R., Saygın, Y., Yanıkoğlu, B. ve Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), 667-685.
- Dehkharghani, R. (2015) *Sentiment Analysis in Turkish: Resources and Techniques* (Doktora Tezi). Sabancı Üniversitesi, İstanbul.
- Demirci, S. (2014). *Emotion analysis on turkish tweets*. (Yayımlanmamış Yüksek Lisans Tezi). ODTÜ, Ankara.
- Denecke, K. ve Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artif. Intell. Med.* 17–27.

- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5.
- Distante, D., Fernandez, A., Cerulo, L. ve Visaggio, A. (2014). Enhancing online discussion forums with topic-driven content search and assisted posting. *In International joint conference on knowledge discovery, knowledge engineering, and knowledge management*, Springer.161–180.
- Dzikovska, M., Steinhäuser, N., Farrow, E., Moore, J. ve Campbell, G. (2014). Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education*, 24, 284–332.
- Ekman, P. (1992). *An argument for basic emotions*. 6, 3–4 (1992), 169–200
- Elia, G., Solazzo, G., Lorenzo, G. ve Passiante, G. (2019). Assessing learners' satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, 92, 589–599.
- Elkins, K. ve Chun J. (2019). Can Sentiment Analysis Reveal Structure in a Plotless Novel?. Cornell University,
- Elouazizi, N., Birol, G., Jandciu, E., berg, G., Welsh, A., Han, A. ve Campbell, A. (2017). *Automated analysis of aspects of written argumentation*. In Proceedings of the seventh international learning analytics & knowledge conference. ACM, 606–607.
- Enginün, İ. (1985). “Ömer Seyfettin'in Hikâyeleri”, *Doğumunun Yüzüncü Yılında Ömer Seyfettin*, Atatürk Kültür Merkezi Yayınları: Ankara.
- Ergüllü, B. (2018). *Twitter Sentiment analysis*. (Yayımlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, İstanbul.
- Escudeiro, N., Escudeiro, P. ve Cruz, A. (2011). Semi-automatic grading of students' answers written in free text. *Electronic Journal of e-Learning*, 9, 15–22.

- Esuli, A. ve Sebastiani F. (2006) Determining term subjectivity and term orientation for opinion mining, in: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL06)*, 193–200.
- Evirgen, E. (2016). *Sentiment analysis of turkish tweets*. (Yayımlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, İstanbul.
- Feldman, R. ve Sanger J. (2007). *Text mining handbook*. New York, Cambridge University Press
- Ferguson, R. ve Shum, S. B. (2011). Learning analytics to identify exploratory dialogue within synchronous text chat. In *Proceedings of the 1st international conference on learning analytics and knowledge*. ACM, 99–103.
- Fırat H. ve Coşkun M., V. (2018). Mustafa Ruhi Şirin'in çocuk şiirlerinde duygu değeri doğrultusunda tanımlanan kavramlar. *The Journal of Academic Social Sciences*.
- Figueira, I. D. B. B. R. (2008b). A repository with semantic organization for educational content. In *Eighth IEEE international conference on advanced learning technologies*, 2008. ICALT0'8 .114–116). IEEE.
- Filatova, E. (2012). Irony and sarcasm: corpus generation and analysis using crowdsourcing, in: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, 392–398.
- Flor, M. ve Riordan, B. (2018). A semantic role-based approach to open-domain automatic question generation. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, 254–263.
- Fox, E. (2008). *Emotion Science Cognitive and Neuroscientific Approaches to Understanding Human Emotions*. Palgrave Macmillan.
- Ganesan, K., Zhai, C. ve Han, J. (2010). Opinosis: a graph-based approach to abstractive summarization of highly redundant opinions, in: *Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10)*, 340–348.

- Gao, B., Berendt, B. ve Vanschoren, J., (2015). Who is more positive in private? Analyzing sentiment differences across privacy levels and demographic factors in facebook chats and posts. *In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 2015. ACM, 605–610.
- Garrido, A. L., Pera, M. S. ve Ilarri, S. (2014). Sole-r: A semantic and linguistic approach for book recommendations. *In IEEE 14th international conference on advanced learning technologies (ICALT)*. IEEE, 524–528).
- Geçgel, H. ve Sarıçan, E. (2011). Ömer Seyfettin’in Hikâyelerinde Çocuk ve Eğitim Teması. *Uşak Üniversitesi Sosyal Bilimler Dergisi*. 4 (2), 164-175.
- Gezici, G. ve Yanıkoğlu, B. (2018). *Sentiment Analysis in Turkish*.
- Gibson, A., Aitken, A., Sándor, A., Buckingham Shum, S., Tsingos-Lucas, C. ve Knight, S. (2017). Reflective writing analytics for actionable feedback. *In Proceedings of the seventh international learning analytics & knowledge conference*. ACM, 153–162.
- Giraz, H. E. (2020). *Emotion analysis on Turkish text*. (Yayımlanmamış Yüksek Lisans Tezi). İzmir Ekonomi Üniversitesi, İzmir.
- Gunnarsson, B. L. ve Alterman, R. (2012). Predicting failure: A case study in co-blogging. *In Proceedings of the 2nd international conference on learning analytics and knowledge*, 263–266, Vancouver.
- Guo, L. ve Wan, X. (2012). Exploiting syntactic and semantic relationships between terms for opinion retrieval, *J. Am. Soc. Inform. Sci. Technol.* 63 (11), 2269–2282.
- Gümüş, C. (2017). *Kelime kullanım oranları ve kullanıcı istatistikleri kullanılarak Türkçe twitter verisi üzerinde duygu analizi* (Yayımlanmamış Yüksek Lisans Tezi). Doğu Üniversitesi, İstanbul.
- Hacızade, N. (2012). *Bilişsel Dilbilim Açısından Duyguların Dili*. Konya: Çizgi Kitabevi.
- Hasanlı, H. (2019). *Büyük veri üzerinde duygu analizi yöntemleri ve Azerbaycan diline uygulanması* (Doktora Tezi) Ege Üniversitesi, İzmir.

- He, Y., Lin, C. ve Alani H. (2011). Automatically extracting polarity-bearing topics for cross-domain sentiment classification, in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT '11)*,123–131.
- Hirschberg, J. ve Manning, C. D. (2015). *Advances in natural language processing*. Science, 349, 261–266.
- Hixson, T. (2020). Reactions vs. reality: Using sentiment analysis to measure university students' responses to learning ArcGIS. *Journal of Map & Geography Libraries*, 1–14.
- Hsu, C.-K., Hwang, G.-J. ve Chang, C.-K. (2010). Development of a reading material recommendation system based on a knowledge engineering approach. *Computers & Education*, 55, 76–83.
- Hu, M. ve Liu, B., (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 168–177.
- Huang, C. Q., Han, Z. M., Li, M. X., Jong, M. S. ve Tsai, C. C. (2019). Investigating students' interaction patterns and dynamic learning sentiments in online discussions. *Computers & Education*, 140, 103589.
- Hughes, S., Hastings, P., Magliano, J., Goldman, S. ve Lawless, K. (2012). *Automated approaches for detecting integration in student essays*. In *International conference on intelligent tutoring systems*,Springer, 274–279.
- Ihaka R. ve Gentleman R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.
- Jacobs, A., M. (2019). Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics. *Language Representation and Learning in Cognitive and Artificial Intelligence Systems*, Vol,6.
- Jindal, N. ve Liu, B. (2007). Review spam detection, in: *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*, ACM Press,1189–1190.

- Jo, Y. ve Oh, A. (2011). Aspect and sentiment unification model for online review analysis. *In Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. ACM, 815–824.
- Joksimovic, S., Gasevic, D., Kovanovic, V., Adesope, O. ve Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, 22, 1–10.
- Jorge-Botana, G., Luzón, J. M., Gómez-Veiga, I. ve Martín-Cordero, J. I. (2015). Automated lsa assessment of summaries in distance education: Some variables to be considered. *Journal of Educational Computing Research*, 52, 341–364.
- Kama, B. (2016). *Feature based sentiment analysis on informal turkish texts* (Yayımlanmamış Yüksek Lisans Tezi). ODTÜ, Ankara.
- Kang, D. ve Park, Y., (2014). Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. *Expert Syst. Appl.* 41, 4 (2014), 1041–1050.
- Kang, M., Chaudhuri, S., Kumar, R., Wang, Y.-C., Rosé, E., Rosé, C. ve Cui, Y. (2008). *Supporting the guide on the side*. *In Intelligent tutoring systems*. Springer, 793–795.
- Kaplan, B. A. (2016). *Twitter üzerindeki Türkçe mesajlarda veri madenciliğiyle duygu analizi*. (Yayımlanmamış Yüksek Lisans Tezi). Beykent Üniversitesi, İstanbul.
- Karamanlı, E. (2019). *Makine öğrenmesi algoritmaları kullanarak, metin madenciliği ve duygu analizi ile müşteri deneyiminin geliştirilmesi* (Yayımlanmamış Yüksek Lisans Tezi) İstanbul Üniversitesi, İstanbul.
- Karaöz, B. A. (2018). *Büyük veri ve işletme analitiği: Sosyal medya ve duygu analizi ile bir öngörü modeli* (Yüksek Lisans Tezi). İstanbul Üniversitesi, İstanbul.
- Katz, P., Singleton, M. ve Wicentowski, R. (2007). Swat-mp: The semeval-2007 systems for task 5 and task 14. *In Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 308–313.

- Kaya M., Fidan G. ve Toroslu I. H. (2012). Sentiment analysis of turkish political news. *International Conferences on Web Intelligence (WI 2012)*, Macau, China, IEEE Computer Society, 174-180.
- Kechaou, Z., Ammar, M. B. ve Alimi, A. M. (2011). Improving e-learning with sentiment analysis of users' opinions. *In Global engineering education conference (EDUCON)*, IEEE, 1032–1038,
- Keleş, E. ve Çepni, S. (2006). Beyin ve öğrenme. *Journal of Turkish science education*. 3(2).
- Khribi, M. K., Jemni, M. ve Nasraoui, O. (2007). Toward a hybrid recommender system for e-learning personalization based on web usage mining techniques and information retrieval. *In E-learn: World conference on E-learning in corporate, government, healthcare ve higher education*. Association for the Advancement of Computing in Education (AACE). 6136–6145).
- Khribi, M. K., Jemni, M. ve Nasraoui, O. (2008). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. *In 8th IEEE international conference on advanced learning technologies*. IEEE, 241–245.
- Kızılkaya, Y. M. (2018). *Duygu analizi ve sosyal medya alanında uygulama* (Doktora Tezi). Uludağ Üniversitesi, Bursa.
- Kim, J. ve Shaw, E. (2014). Scaffolding student online discussions using past discussions: Pedabot studies. *Artificial Intelligence Review*, 41, 97–112.
- Kiritchenko, S., Zhu, X., Cherry, C. ve Mohammad, S., A. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 437–442.
- Koedinger K, Cunningham K, Skogsholm A. ve Leber, B. (2008). An open repository and analysis tools for finegrained, longitudinal learner data. *In: First International Conference on Educational Data Mining*. Montreal, Canada;, 157–166.

- Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtins contribution to understanding learning in settings of collaboration. *In Proceedings of the 1999 conference on computer support for collaborative learning*. Nashville.
- Kotu V. ve Deshpande B. (2019). *Data Science, Concepts and Practice*. Cambridge: Morgan Kaufmann Publishers
- Kumar A. ve Paul A. (2016) Mastering Text Mining with R: Master text-taming techniques and build effective text-processing applications with R. Birmingham: Packt Publishing
- Kumar A. ve Sebastian T., M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *Intelligent Systems and Applications*. 10, 1-14.
- Latif, M. M. A. (2008). A state-of-the-art review of the real-time computer-aided study of the writing process. *International Journal of English Studies*, 8, 29–50.
- Lee, L. ve Pang, B. (2008) Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* Vol. 2(1), 1–135
- Lee, S., W., Song, Y., I., Lee, J., T., Han, K., S. ve Rim, H., C. (2011). A new generative opinion retrieval model integrating multiple ranking factors, *J. Intell. Inform. Syst.* 38 (2), 487–505.
- Lei, J., Rao, Y., Li, Q., Quan, X. ve Wenyin, L. (2014). Towards building a social emotion detection system for online news. *Fut. Gen. Comput. Syst.* 37, 438–448.
- Leiyue, Y. ve Jianying, X. (2012). Chat analysis to understand students using text mining. *In Advances in control and communication*. Springer, 235–243.
- Leong, C. K., Lee, Y. H. ve Mak, W. K. (2012). Mining sentiments in sms texts for teaching evaluation. *Expert Systems with Applications*, 39, 2584–2589.
- Lewkow, N., Feild, J., Zimmerman, N., Riedesel, M., Essa, A. ve Boulanger, D. (2016). A scalable learning analytics platform for automated writing feedback. *In Proceedings of the third (2016) ACM conference on learning@ scale* .109–112, ACM

- Li, L. (2018). Sentiment-enhanced learning model for online language learning system. *Electronic Commerce Research*, 18 (1), 23–64.
- Lin, C. ve He, Y. (2009). Joint sentiment/topic model for sentiment analysis. *In Proceedings of the 18th ACM Conference on Information and Knowledge Management*. ACM, 375–384.
- Lin, F.-R., Hsieh, L.-S. ve Chuang, F.-T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education*, 52, 481–495.
- Lin, H.-C. K., Chen, N.-S., Sun, R.-T. ve Tsai, I.-H. (2014). Usability of affective interfaces for a digital arts tutoring system. *Behaviour ve Information Technology*, 33(2), 105–116.
- Lin, Q., Zhu, Y., Zhang, S., Shi, P., Guo, Q. ve Niu, Z. (2019). Lexical based automated teaching evaluation via students' short reviews. *Computer Applications in Engineering Education*, 27(1), 194–205
- Litman, D. (2016). Natural language processing for enhancing teaching and learning. *In Proceedings of the thirtieth AAAI conference on artificial intelligence* .4170–4176). AAAI Press.
- Liu B. (2010). *Sentiment Analysis and Subjectivity*. Handbook of Natural Language Processing,
- Liu B., Hu, M. ve Cheng, J. (2005). Opinion observer: Analyzing and comparing opinions on the web. *In Proceedings of the 14th International Conference on World Wide Web*. ACM, 342–351
- Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers
- Liu, M., Calvo, R. A. ve Pardo, A. (2013). Tracer: A tool to measure and visualize student engagement in writing activities. *In 2013 IEEE 13th international conference on advanced learning technologies (ICALT)* .421–425).
- Liu, M., Pardo, A. ve Liu, L. (2017). Using learning analytics to support engagement in collaborative writing. *International Journal of Distance Education Technologies (IJDET)*, 15, 79–98.

- Liu, S., Peng, X., Cheng, H. N. H., Liu, Z., Sun, J. ve Yang, C. (2019). Unfolding sentimental and behavioral tendencies of learners' concerned topics from course reviews in a MOOC. *Journal of Educational Computing Research*, 57(3), 670–696.
- Lovheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Med. Hypoth.* 78, 2 (2012), 341–348
- Lui, A. K.-F., Li, S. C. ve Choy, S. O. (2007). An evaluation of automatic text categorization in online discussion analysis. In *7th IEEE international conference on advanced learning technologies*. ICALT 2007. 205–209. Niigata, Japan.
- Luque, C., Luna, J. M., Luque, M. ve Ventura, S. (2019). An advanced review on text mining in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1302.
- Lyons, M., Aksayli, N. D. ve Brewer, G. (2018). Mental distress and language use: Linguistic analysis of discussion forum posts. *Computers in Human Behavior*, 87, 207–211.
- Maks, I. Ve Vossen, P. (2012). A lexicon model for deep sentiment analysis and opinion mining applications, *Decis. Support Syst*, 53 (4) ,680–688.
- Manning C, Raghavan P. ve Schütze H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
- Manning C. ve Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Manning, C. D., Raghavan, P. ve Schütze, H. (2008). *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.
- Mansur, A. B. F. ve Yusof, N. (2013). Social learning network analysis model to identify learning patterns using ontology clustering techniques and meaningful learning. *Computers & Education*, 63, 73–86.
- Matthew, J. (2014). *Text Analysis with R for Students of Literature*. Springer, Switzerland.

- Mazidi, K. ve Tarau, P. (2016). Automatic question generation. *In International conference on intelligent tutoring systems*. 23–33.
- Mazzolini, M. ve Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, 49, 193–213.
- Medhat, W., Hassan, A. ve Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113
- Medler, D. A., Arnoldussen, A., Binder, J. R. ve Seidenberg, M. S. (2005). *The Wisconsin Perceptual Attribute Ratings Database*.
- Melville, P., Gryc, W. ve Lawrence, R., (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. *In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1275–1284.
- Miner G., Delen D., Elder J., Fast A., Thomas H. T. ve Nisbet R. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. USA, Academic Press
- Mohammad S., M. (2011). *Sentiment Analysis of Mail and Books*. Institute for Information Technology, National Research Council, Canada.
- Mohammad, S., A. (2012). # Emotional tweets. In Proceedings of the 1st Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the Main Conference and the Shared Task. Volume 2: Proceedings of the 6th International Workshop on Semantic Evaluation. *Association for Computational Linguistics*, 246–255.
- Mohammad, S., A. ve. Turney, P., (2010). *Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon*. In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text. *Association for Computational Linguistics*, 26–34.
- Mohammad, S., M., Kiritchenko, S. ve Zhu, X. (2013). NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. *In Proceedings of the 7th International Workshop on Semantic Evaluation Exercises (SemEval'13)*.

- Montoyo, A., Martínez-Barco, P. Ve Balahur, A. (2012). Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments, *Decis. Support Syst.* 53 (4) 675–679.
- Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., Estevez-Ayres, I. ve Kloos, C. D. (2019). A learning analytics methodology for understanding social interactions in MOOCs. *IEEE Transactions on Learning Technologies*, 12(4), 442–455.
- Morgan, D. L. (1996). *Focus groups as qualitative research* (C. 16). New York: Sage publications
- Mostow, J. ve Beck, J. (2006). *Some useful tactics to modify, map and mine data from intelligent tutors.*
- Mukherjee, A, Liu, B., Glance, N. (2012) Spotting fake reviewer groups in consumer reviews, *in: Proceedings of the 21st International Conference on World Wide Web (WWW '12)*, ACM Press, 191–200.
- Mukherjee, A. ve Liu, B. (2012). Aspect extraction through semi-supervised modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*. Association for Computational Linguistics, 339-348.
- Mulki, H., Haddad, H. ve Babaoğlu, İ. (2018). Preprocessing impact on Turkish sentiment analysis. *2018 26th Signal Processing and Communications Applications Conference (SIU)*, İzmir.
- Nagashree, N. ve Pujari, N. V. (2016). A tutor assisting novel electronic framework for qualitative analysis of a question bank. *Computers in Human Behavior*, 65, 9–13.
- Neviarouskaya, A., Prendinger, H. ve Ishizuka, M., (2007). Textual affect sensing for sociable and expressive online communication. *In Affective Computing and Intelligent Interaction*. Springer, 218–229.
- Newman, H. ve Joyner, D. (2018). Sentiment analysis of student evaluations of teaching. *In International conference on artificial intelligence in education*. 246–250.

- Nimala, K. ve Jebakumar, R. (2019). Sentiment topic emotion model on students feedback for educational benefits and practices. *Behaviour & Information Technology*, 1–9.
- Nisbet, R., Elder, J. ve Yale K. (2018). *Handbook of Statistical Analysis & Data Mining Applications*. Academic Press.
- Nisbet, R., J. Elder ve G. Miner. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier, Burlington, MA.
- Nishikawa, H., Hasegawa, T., Matsuo, Y. ve Kikui, G. (2010). Opinion summarization with integer linear programming formulation for sentence extraction and ordering, in: *Proceedings of the 23rd International Conference on Computational Linguistics (COLING '10)*, 910–918.
- Nitin, G. I., Swapna, G. ve Shankararaman, V. (2015). Analyzing educational comments for topics and sentiments: A text analytics approach. In *Frontiers in education conference (FIE)*, IEEE, 1–9.
- Nunes, B. P., Mera, A., Kawase, R., Fetahu, B., Casanova, M. A. ve de Campos, G. H. B. (2014). A topic extraction process for online forums. In *2014 IEEE 14th international conference on advanced learning technologies (ICALT)*. IEEE. 541–543.
- Oramas B. R., Zatarain C. R., Barrón E. M. L. ve Hernández P. Y. (2019). Opinion mining and emotion recognition in an intelligent learning environment. *Computer Applications in Engineering Education*, 27(1), 90–101.
- O'Rourke, S. T., Calvo, R. A. ve McNamara, D. S. (2011). Visualizing topic flow in students essays. *Educational Technology & Society*, 14, 4–15.
- Ortigosa, A., Martín, J. M. ve Carro, R. M. (2014). Sentiment analysis in facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527–541.
- Osmanoglu,U.O., Atak,O.N., Caglar,K., Kayhan, H. ve Can, T.C. (2020). Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach. *Journal of Educational Technology & Online Learning*, 3(1), 31-48

- Önal, M. (2017). *Sözlüksel benzeşim yöntemi ile android tabanlı duygu analizi uygulaması* (Yayımlanmamış Yüksek Lisans Tezi). Fırat Üniversitesi, Elazığ.
- Özdeş, M. (2017). *Büyük veri araçlarını kullanarak duygu analizi gerçekleştirimi* (Yayımlanmamış Yüksek Lisans Tezi). Pamukkale Üniversitesi, Denizli.
- Özkul, M. (2019). *Metin madenciliği yöntemi ile akıllı ulaşım sistemleri algısının duygu analizi ile ölçülmesi* (Yayımlanmamış Yüksek Lisans Tezi). Bandırma Onyedli Eylül Üniversitesi, Çanakkale.
- Pak, M. Y., (2015). *Metinlerde duygu analizi ve sınıflandırma yöntemleri* (Yayımlanmamış Yüksek Lisans Tezi). Anadolu Üniversitesi, Eskişehir.
- Pang, B ve Lee L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*,1–135.
- Pang, B. ve Lee, L. (2004). A sentimental education, in: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04), 271–278.
- Pang, B. ve Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 271.
- Pang, B. ve Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics*, 115–124.
- Pang, B. ve Lee, L. ve Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10. Association for Computational Linguistics, 79–86.
- Park, K.-M., Park, H., Kim, H., G. ve Ko, H. (2012). Review summarization based on linguistic knowledge, in: H. Yu, G. Yu, W. Hsu, Y.-S. Moon, R. Unland, J. Yoo (Ed.), *Proceedings*

of the 17th International Conference on Database Systems for Advanced Applications (DASFAA'12), Lecture Notes in Computer Science, vol. 7240, Springer Berlin Heidelberg, Berlin Heidelberg, 105–114.

Patel F. ve Soni, N. (2012). Text mining: A Brief survey. *International Journal of Advanced Computer Research*. 2 (4).

Patel, D. (2018). Sentiment analysis of harry potter book series using r. *Journal of Applied Science and Computations*,5 (11).

Pathak, M., A. (2013). *Beginning Data Science with R*. Springer International Publishing, Switzerland.

Pennebaker, J., Booth, R. J. ve Francis, M., (2007). *Linguistic inquiry and word count: LIWC*. Austin, TX: liwc.net

Plutchik, R. ve Kellerman, H. (1986). *Emotion: Theory, Research and Experience*. Academic Press, New York.

Pong-inwong, C. ve Songpan, W. (2019). Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining. *International Journal of Machine Learning and Cybernetics*, 10(8), 2177–2186.

Raaijmakers, S. ve Kraaij, W. (2008) A shallow approach to subjectivity classification, *in: Proceedings of the Second International Conference on Weblogs and Social Media (ICWSM '08)*, 216–217.

Rahimi, Z., Litman, D. J., Correnti, R., Matsumura, L. C., Wang, E. ve Kisa, Z. (2014). Automatic scoring of an analytical response-to-text assessment. *In International conference on intelligent tutoring systems* .601–610). Springer.

Rajput, Q., Haider, S. ve Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. *Applied computational intelligence and soft computing*, pages 12.

Ramachandran, L. ve Gehringer, E. F. (2011). Automated assessment of review quality using latent semantic analysis. *In 2011 11th IEEE international conference on advanced learning Technologies*.136–138.

- Rani, S. ve Kumar, P. (2017). A sentiment analysis system to improve teaching and learning. *Computer*, 50(5), 36–43.
- Ravi, S. ve Kim, J. (2007). Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications*, 158, 357.
- Reagan, A. J. (2017). Towards a science of human stories: using sentiment analysis and emotional arcs to understand the building blocks of complex social systems (Graduate College Dissertations and Theses). University of Vermont.
- Reimann, P., Frerejean, J. ve Thompson, K. (2009). Using process mining to identify models of group decision making in chat data. *In Proceedings of the 9th international conference on computer supported collaborative learning-volume 1. International Society of the Learning Sciences*. 98–107.
- Reyes, A. ve Rosso, P. (2012). Making objective decisions from subjective data: detecting irony in customer reviews, *Decis. Support Syst.* 53 (4), 754–760.
- Romero C, Ventura S, Pechenizky M. ve Baker, R. (2010). *Handbook of Educational Data Mining. Data Mining and Knowledge Discovery Series*. Boca Raton, FL: Chapman and Hall/CRC Press; 2010.
- Romero, C. ve Ventura, S. (2007) Educational data mining: a survey from 1995 to 2005. *J Expert Syst Appl*, 1:135–146.
- Romero, C. ve Ventura, S. (2013). Data Mining in Education. *WIREs Data Mining Knowl Discov* 2013, 3: 12–27
- Romero, C. ve Ventura, S. (2017). Educational data science in massive open online courses. *Data Mining and Knowledge Discovery*.
- Rus, V., Banjade, R., Lintean, M., Niraula, N. ve Stefanescu, D. (2013). Semilar: A semantic similarity toolkit for assessing students natural language inputs. *In Educational data mining 2013*. Memphis, Tennessee: University of Memphis.

- Ruseti, S., Dascalu, M., Johnson, A. M., Balyan, R., Kopp, K. J., McNamara, D. S. ve Trausan-Matu, S. (2018). Predicting question quality using recurrent neural networks. *In International conference on artificial intelligence in education*, Springer, 491–502.
- Rushdi-Saleh, M, Martín-Valdivia, M., Montejo-Ráez, A. ve Ureña López, L. (2011). Experiments with SVM to classify opinions in different domains, *Expert Syst. Appl.* 38 (12) 14799–14804.
- Sachan, M. ve Xing, E. P. (2016). Easy questions first? A case study on curriculum learning for question answering. *In Proceedings of the annual meeting of the association for computational linguistics*.
- Sağlam, F. (2019). *Otomatik duygu sözlüğü geliştirilmesi ve haberlerin duygu analizi*. (Doktora lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Saha, S., Dhamecha, T. I., Marvaniya, S., Sindhgatta, R. ve Sengupta, B. (2018). Sentence level or token level features for automatic short answer grading?: Use both. *In International conference on artificial intelligence in education*, Springer, 503–517.
- Sallabaş, M. E. (2012). Ömer Seyfettin hikâyelerinin Türkçe öğretiminde değer aktarımı bakımından incelenmesi. *Mustafa kemal üniversitesi sosyal bilimler enstitüsü dergisi*, 59-68.
- Santos, C. L., Rita, P. ve Guerreiro, J. (2018). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. *International Journal of Educational Management*, 32(3), 431–447.
- Sarvabhotla, K., Pingali, P. ve Varma, V. (2011). Sentiment classification: a lexical similarity based approach for extracting subjectivity in documents, *Inform. Retrieval*, 14 (3). 337–353.
- Savoy, J. (2012). Authorship attribution based on specific vocabulary, *ACM Trans. Inform. Syst.* 30 (2), 1–30.
- Scherer, K. ve Wallbott, H. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *J. Pers. Soc. Psychol.* 66, 2 (1994), 310.

- Scheuer, O. ve McLaren, B. (2008). *Helping teachers handle the flood of data in online student discussions*. In *Intelligent tutoring system*. Springer, 323–332.
- Seki, Y., Kando, N. ve Aono, M. (2009). Multilingual opinion holder identification using author and authority viewpoints, *Inform. Process. Manage.* 45 (2). 189–199.
- Sevindi, B. I. (2013). *Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarının karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Gazi Üniversitesi, Ankara.
- Shahraki, A., G., (2015). *Emotion Detection from Text*. Master's thesis. University of Alberta.
- Shaver, P., Schwartz, J., Kirson, D. ve O'connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *J. Pers. Soc. Psychol.* 52, 6 (1987), 1061
- Sheeha, M. (2019). *Real-time social media sentiment analysis using big data technologies*. (Yayımlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, İstanbul.
- Shehu, H. A. (2019). *Kutupsallık sözlüğü ve yapay zeka yardımı ile Türkçe twitter verileri üzerinde duygu analizi*. (Yayımlanmamış Yüksek Lisans Tezi). Pamukkale Üniversitesi, Denizli.
- Shi, Y., Larson, M. ve Jonker, C. M. (2015). Recurrent neural network language model adaptation with curriculum learning. *Computer Speech & Language*, 33, 136–154.
- Siemens, G. ve Baker, R. S. J. d. (2012). Learning analytics and educational data mining: towards communication and collaboration. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. Vancouver, British Columbia, Canada; 1–3.
- Sims, S. ve Bokhove, C. (2020) Demonstrating the potential of text mining for analyzing school inspection reports: a sentiment analysis of 17,000 Ofsted documents, *International Journal of Research & Method in Education*.
- Simsek, D., Sándor, A., Shum, S. B., Ferguson, R., De Liddo, A. ve Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. In *Proceedings of the fifth international conference on learning analytics and knowledge*. ACM. 355–359.

- Sinclair, A., Lopez, A., Lucas, C. ve Gasevic, D. (2018). Does ability affect alignment in second language tutorial dialogue? *In Proceedings of the 19th annual SIGdial meeting on discourse and dialogue* .41–50.
- Sindhu, I., Muhammad Daudpota, S., Badar, K., Bakhtyar, M., Baber, J. ve Nurunnabi, M. (2019). Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access*, 7, 108729–108741
- Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A. ve McNamara, D. S. (2015). Youv'e got style: Detecting writing flexibility across time. *In Proceedings of the fifth international conference on learning analytics and knowledge* .194–202). ACM.
- Stanley, H., H. (1978). *Information retrieval: Computational and theoretical aspects*. Academic Press. New York, San Francisco.
- Stone, P., Dunphy, D., Smith, M. ve D. M. Ogilvie. (1968). The general inquirer: A computer approach to content analysis. *J. Region. Sci.* 8, 1 (1968), 113–116
- Strapparava C. ve Valitutti, A. (2004). WordNet affect: An affective extension of wordnet. *In LREC*, 1083–1086.
- Strapparava, C. ve Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*. Association for Computational Linguistics, 70–74.
- Sung, Y.-T., Liao, C.-N., Chang, T.-H., Chen, C.-L. ve Chang, K.-E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The lsa-based technique. *Computers & Education*, 95, 1–18.
- Suwal, S. ve Singh, V. (2018). Assessing students' sentiments towards the use of a Building Information Modelling (BIM) learning platform in a construction project management course. *European Journal of Engineering Education*, 43(4), 492–506.
- Tata, S. ve Di Eugenio, B. (2010). Generating fine-grained reviews of songs from album reviews, in: *Proceedings of the 48th Annual Meeting of the Association for*

Computational Linguistics, Association for Computational Linguistics, Uppsala, Sweden, 1376–1385.

Tian, F., Gao, P., Li, L., Zhang, W., Liang, H., Qian, Y. ve Zhao, R. (2014). Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems. *Knowledge-Based Systems*, 55, 148–164.

Tobarra, L., Robles-Gómez, A., Ros, S., Hernández, R. ve Caminero, A. C. (2014). Analyzing the students' behavior and relevant topics in virtual learning communities. *Computers in Human Behavior*, 31, 659–669.

Toçoğlu, M. A. (2018). *Lexicon-based emotion analysis in Turkish*. (Yayımlanmamış Yüksek Lisans Tezi). Dokuz Eylül Üniversitesi, İzmir.

Trausan-Matu, S., Dascalu, M. ve Rebedea, T. (2012). A system for the automatic analysis of computer-supported collaborative learning chats. *In Advanced learning technologies (ICALT)*, 95–99.

Trivedi, S., Pardos, Z. A. ve Heffernan, N. T. (2011). Clustering students to generate an ensemble to improve standard test score predictions. *In International conference on artificial intelligence in education*. Springer. 377–384).

Troisi, O., Grimaldi, M., Loia, F. ve Maione, G. (2018). Big data and sentiment analysis to highlight decision behaviours: A case study for student population. *Behaviour & Information Technology*, 37(10–11), 1111–1128.

Tseng, C.-W., Chou, J.-J. ve Tsai, Y.-C. (2018). Text mining analysis of teaching evaluation questionnaires for the selection of outstanding teaching faculty members. *IEEE Access*, 6, 72870–72879.

Turna, M. (2020). Makale ve Günlüklerinden Hareketle Ömer Seyfettin'in Görüşleri Üzerine Bir Değerlendirme. *Dil ve Edebiyat Araştırmaları*. (22), 127-157.

Türk Dil Kurumu (2019). *Güncel Türkçe Sözlük*. Ankara: TDK Yayınları.

Türkmenoğlu, C. (2015). *Türkçe Metinlerde Duygu Analizi* (Yayımlanmamış Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, İstanbul.

- Uçan, A. (2014). *Otomatik duygu sözlüğü çevirimi ve duygu analizinde kullanımı* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Valjataga, T., Põldoja, H. ve Laanpere, M. (2011). Open online courses: Responding to design challenges. Stanford University.
- Varişoğlu, B. (2017). Dil Kullanımının Bir Yansıması Olan Duyguları Dile Getirme Becerisiyle Duygusal Zekâ Arasındaki İlişki: Türkçe Öğretmeni Adayları Örnekleme. *Turkish Studies.International Periodical for languages, Letarure and History of turkish or Turkic*. 12(14), 513-528.
- Villalón, J. J. ve Calvo, R. A. (2011). Concept maps as cognitive visualizations of writing assignments. *Educational Technology & Society*, 14, 16–27.
- Vu, X. S. ve Park, S. B. (2014). Construction of vietnamese SentiWordNet by using vietnamese dictionary. *In Proceedings of the 40th Conference of the Korea Information Processing Society*. 745–748.
- Vural, A. G., Cambazoğlu, B. B., Senkul, P. Ve Tokgöz, Z. O. (2013). A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. *In Computer and Information Sciences III* ,437-445. Springer, London.
- Wang, D., Zhu, S. ve Li, T. (2009). SumView: a Web-based engine for summarizing product reviews and customer opinions, *Expert Syst. Appl.* 40 (1), 27–33.
- Wang, Z., Pan, X., Miller, K. F. ve Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Computers & Education*, 78, 115–123.
- Warschauer, M. ve Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research*, 10,157–180.
- Weiss S., Indurkha N. ve Zhang T. (2019). *Fundamentals of Predictive Text Mining*. New York: Springer
- Whitelock, D., Twiner, A., Richardson, J. T., Field, D. ve Pulman, S. (2015). Openessayist: A supply and demand learning analytics tool for drafting academic essays. *In Proceedings*

of the fifth international conference on learning analytics and knowledge, ACM, 208–212.

Wiebe, J., Wilson, T., Bruce, R., Bell, M. ve Martin, M. (2004). *Learning subjective language. Computational Linguistics*. 30(3):277–308

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, J., Cardie, C., Riloff, E. ve Patwardhan, S. (2005a). OpinionFinder: A system for subjectivity analysis. In *Proceedings of hlt/emnlp on Interactive Demonstrations. Association for Computational Linguistics*, 34–35.

Wilson, T., Wiebe, J. ve Hoffmann, P. (2005b). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, 347–354.

Wilson, T., Wiebe, J. ve Hwa, R. (2004). Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of AAAI*, 761–769.

Witten, I. H. (2004). *Text mining*. Boca Raton, FL: CRC Press.

Woods, B., Adamson, D., Miel, S. ve Mayfield, E. (2017). Formative essay feedback using predictive scoring models. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 2071–2080.

Wulf, J., Blohm, I., Leimeister, J. M. ve Brenner, W. (2014). Massive open online courses. *Business ve Information Systems Engineering (BISE)*, 6, 111–114.

Xie, S., Wang, G., Lin, S. ve Yu, P. S. (2012). Review spam detection via temporal pattern discovery, in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, 823–831.

Xu, S., Chen, J. ve Qin, L. (2018). Cluf: A neural model for second language acquisition modeling. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications* .374–380)

- Yadollahi, A., Shahraki A. G., ve. Zaiane O. (2017). Görüşten Duygu Madenciliğine Metin Duyarlılığı Analizinin Mevcut Durumu. *ACM Comput. Surv.* 50, 2.
- Yelmen, İ. (2016). *Doğal dil işleme yöntemleriyle Türkçe sosyal medya verileri üzerinde duygu analizi* (Yayımlanmamış Yüksek Lisans Tezi). İstanbul Aydın Üniversitesi, İstanbul.
- Yıldırım., S. (2018). *Twitter verileriyle duygu analizi ve Türkçe duygu kütüphanesi* (Yayımlanmamış Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, İstanbul.
- Yoo, J. ve Kim, J. (2014). Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, 24, 8–32.
- Younis, E. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *International Journal of Computer Applications*. 112(5), Egypt
- Yu, L. C., Lee, C. W., Pan, H. I., Chou, C. Y., Chao, P. Y., Chen, Z. H., Tseng, S. F., Chan, C. L. ve Lai, K. R. (2018). Improving early prediction of academic failure using sentiment analysis on self-evaluated comments. *Journal of Computer Assisted Learning*, 34(4), 358–365.
- Yu, L.C., Wu, J.L., Chann, P.C. ve Chu, H.S. (2013). Using a contextual entropy model to expand emotionwords and their intensity for the sentiment classification of stockmarket news. *Knowl-Based Syst.* 41:89–97.
- Yurt, E. A. (2015). *Türkçe metinlerde duygu analizi* (Yayımlanmamış Yüksek Lisans Tezi). Maltepe Üniversitesi, İstanbul.
- Yurtalan, G., Koyuncu M. ve Turhan, Ç. (2019). A polarity calculation approach for lexiconbased Turkish sentiment analysis. *Turkish Journal of Electrical Engineering & Computer Sciences*.
- Yurtöz, Ç. Ü. (2019). *Measuring the sentiment effects using emoticon features for a general turkish corpus* (Yayımlanmamış Yüksek Lisans Tezi). Galatasaray Üniversitesi, İstanbul.

Zhu, X., Kiritchenko, S. ve Mohammad, S., A. (2014). Sentiment analysis of short informal texts. *J. Artif. Intell. Res.* 50 (2014), 723–762.



EKLER

Ek-1: Duygu Analizine İlişkin R Kodları

```
belge<-file(description = "hikâye.txt", blocking = TRUE, encoding = getOption("encoding"),
raw = FALSE,method = getOption("url.method", "default"))
satir <- readLines(belge, encoding = "UTF-8")
belge_basi<- grep("omer seyfettin", satir) + 1
belge_sonu <- grep("son", satir) - 1
satir_kumesi <- satir
anabaslik <- grep("baslangic", satir_kumesi )
metin_bolum <- grep("^[:space:]*[[:digit:]]+\\. ", satir_kumesi )
metin_bolum <- metin_bolum[metin_bolum > anabaslik]
start <- c(1, metin_bolum + 1)
end <- c(metin_bolum - 1, length(satir_kumesi ))
metin<- mapply(function(s, e) paste(satir_kumesi[s:e], collapse = "\n"), start, end)
metin<- trimws(metin)
metin <- metin[-1]
baslik<- sub("^[:space:]*[[:digit:]]+\\. ", "", satir_kumesi [metin_bolum])
baslik <- trimws(baslik)
h_korpus <- corpus_frame(baslik, metin)
rownames(h_korpus) <- sprintf("ch%02d", seq_along(metin_bolum))
text_filter(h_korpus)
text_tokens(h_korpus)
text_filter(h_korpus) <- text_filter(drop_punct = TRUE)
term_stats(h_korpus)
sturk<-stopwords("tr",source="stopwords-iso")
term_stats(h_korpus, subset = !term %in% sturk)
x <- term_matrix(h_korpus)
dim(x)
bt_matris <- term_matrix(h_korpus, transpose = TRUE)
library(foreign)
duygusozluk<-read.csv("duygusozluk.csv", header = TRUE, sep = ";")
```

```

colnames(duygusozluk) <- c("kelime", "duygu")
term_stats(h_korpus, subset = term %in% duygusozluk$kelime)
terim_skor<- with(duygusozluk, unclass(table(kelime, duygu)))
head(terim_skor)
kategori<- rowSums(terim_skor> 0)
hikaye_bolum <- text_split(h_korpus, "tokens", 500)
(t_tokensayisi <- text_ntoken(hikaye_bolum))
x <- term_matrix(hikaye_bolum, select = rownames(terim_skor))
metin_skor <- x %*% terim_skor
birim<- 1000
oran <- list(poz = metin_skor[, "pozitif"] / t_tokensayisi * birim,
            neg = metin_skor[, "negatif"] / t_tokensayisi * birim)
oran$toplamlam <- rate$pos + rate$neg
sh<- lapply(oran, function(r) sqrt(r * (birim - r) / t_tokensayisi))
i <- seq_len(nrow(hikaye_bolum))
par(mar = c(4, 4, 11, 9) + 0.1, las = 1)
xlim <- range(i - 0.5, i + 0.5)
ylim <- range(0, oran$toplamlam + sh$toplamlam, oran$toplamlam - sh$total)
plot(xlim, ylim, type = "n", xlab = "Segment", ylab = "Rate \u00d7 1000", axes = FALSE,
     xaxs = "i")
usr <- par("usr")
axis(1, at = i[i %% 5 == 0], labels = FALSE)
axis(1, at = i[i %% 10 == 0], labels = TRUE)
axis(2)
abline(v = tapply(i, hikaye_bolum$parent, min) - 0.5, col = "gray")
labels <- h_korpus$baslik
at <- tapply(i, hikaye_bolum$parent, mean)
text(at, usr[4] + 0.01 * diff(usr[3:4]),
     labels = labels, adj = 0, srt = 45, cex = 0.8, xpd = TRUE)
box()
col <- c(total = "#000000", pos = "#FC8D62", neg = "#8DA0CB")
legend(usr[2] + 0.015 * diff(usr[1:2]), usr[3] + 0.8 * diff(usr[3:4]),

```

```

legend = c("toplaml", "pozitif", "negatif"),
title = expression(bold("duygu")),
fill = col[c("total", "poz", "neg")],
cex = 0.8, xpd = TRUE)
abline(h = mean(oran$toplaml), lty = 2, col = col[["toplaml"]])
for (t in c("neg", "poz", "toplaml")) {
  r <- rate[[t]]
  s <- se[[t]]
  cl <- col[[t]]
  lines (i, r, col = cl)
  points(i, r, col = cl, pch = 16, cex = 0.5)
  if (t == "total") {
    int <- abs((r - mean(r)) / sd(r)) > 2
    segments(i[int], (r - s)[int], i[int], (r + s)[int], col = cl)
    segments((i - .2)[int], (r - s)[int], (i + .2)[int], (r - s)[int], col = cl)
    segments((i - .2)[int], (r + s)[int], (i + .2)[int], (r + s)[int], col = cl)
  }
}

```

Ek-2: Etkisiz İfadeler Listesi

acaba, acep, açıkça, açıkçası, adamakıllı, adeta, ait, altı, altmış, ama, amma, anca, ancak, arada, artık, aslında, aynen, ayrıca, az, bana, bari, başka, başkası, bazen, bazı, belki, ben, benden, beni, benim, beri, beriki, beş, bilcümle, bile, bin, binaen, binaenaleyh, bir, biraz, birazdan, birbiri, birçoğu, birçok, birden, birdenbire, biri, birice, birileri, birisi, birkaç, birkaçı, bir kez, birlikte, bir şey, bir şeyi, bitevi, biteviye, bittabi, biz, bizatihi, bizce, bizcileyin, bizden, bize, bizi, bizim, bizimki, bizzat, boşuna, böyle, böylece, böylecene, böylelikle, böylemesine, böylesine, bu, buna, bunda, bundan, bunlar, bunları, bunların, bunu, bunun, buracıkta, burada, buradan, burası, büsbütün, bütün, çabuk, çabukça, çeşitli, çoğu, çoğun, çoğunca, çoğunlukla, çok, çokça, çokları, çoklarınca, çokluk, çoklukla, cuk, cümlesi, çünkü, da, daha, dahi, dahil, dahilen, daima, dair, dayanarak, de, defa, değil, değin, dek, demin, demincek, deminden, denli, derakap, derhal, derken, diğer, diğeri, diye, doğru, doksan, dokuz, dolayı, dolayısıyla, dört, edecek, eden, eder, ederek, edilecek, ediliyor, edilmesi, ediyor, eğer, elbet, elbette, elli, emme, en, enikonu, epey, epeyce, epeyi, esasen, esnasında, etmesi, etraflı, etraflıca, etti, ettiği, ettiğini, evleviyetle, evvel, evvela, evvelce, evvelden, evvelemirde, evveli, fakat, filanca, gah, gayet, gayetle, gayri, gayrı, geçende, geçenlerde, gelgelelim, gene, gerçi, gerek, gibi, gibilerden, gibisinden, gine , gırla, göre, hakeza, halbuki, halen, halihazırda, haliyle, handiyse, hangi, hangisi, hani, hariç, hasebiyle, hasılı, hatta, hele, hem, henüz, hep, hepsi, her, herhangi, herkes, herkesin, hiç, hiçbir, hiçbir, hoş, hulasaten, için, iken, iki, ila, ile, ilen, ilgili, ilk, illa, illaki, imdi, indinde, inen, iş, ise, işte, ister, itibaren, itibariyle, itibarıyla, iyi, iyice, iyicene, kaç, kadar, kaffesi, kah, kala, kanımca, karşın, katrilyon, kaynak, kelli, kendi, kendilerine, kendini, kendisi, kendisine, kendisini, kere, keşke, kez, keza, kezalık, ki, kim, kimden, kime, kimi, kimisi, kimse, kimsecik, kimsecikler, kırk, kısaca, külliye, lakin, leh, lütfen, maada, madem, mademki, mamafih, mebni, meğer, meğerki, meğerse, milyar, milyon, mı, mu, mü, naşı, nasıl, nasılsa, nazaran, ne, neden, nedeniyle, nedenle, nedense, nerde, nerden, nerdeyse, nere, nerede, nereden, neredeyse, neresi, nereye, nitekim, neye, neyi, neyse, nice, niçin, nihayet, nihayetinde, nitekim, niye, o, öbür, öbürkü, öbürü, olan, olarak, oldu, olduğu, olduğunu, oldukça, olduklarını, olmadı, olmadığı, olmak, olması, olmayan, olmaz, olsa, olsun, olup, olur, olursa, oluyor, on, ona, onca, önce, önceden, önceleri, öncelikle, onculayın, onda, ondan, onlar, onlardan, onları, onların, onu, onun, oracık, oracıkta, orada, oradan, oranca,

oranla, oraya, öteki, ötekisi, otuz, öyle, öylece, öylelikle, öylemesine, oysa, oysaki, öz, pek, pekala, pekçe, peki, peyderpey, rağmen, sadece, sahi, sahiden, sana, sanki, şayet, sekiz, seksen, sen, senden, seni, senin, şey, şeyden, şeyi, şeyler, şimdi, siz, sizden, sizi, sizin, sonra, sonradan, sonraları, sonunda, şöyle, şu, şuna, şuncacık, şunda, şundan, şunlar, şunları, şunu, şunun, şura, şuracık, şuracıkta, şurası, tabii, tam, tamam, tamamen, tamamıyla, tarafından, tek, trilyon, tüm, üç, üzere, var, vardı, vasıtasıyla, ve, velev, velhasıl, velhasılıkelam, veya, veyahut, ya, yahut, yakinen, yakında, yakından, yakınlarda, yalnız, yalnızca, yani, yapacak, yapılan, yapılması, yapıyor, yapmak, yaptı, yaptığı, yaptığını, yaptıkları, yedi, yeniden, yenilerde, yerine, yetmiş, yine, yirmi, yok, yoksa, yoluyla, yüz, yüzünden, zarfında, zaten, zati, zira

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Leyla Karakuş

Eğitim Durumu

Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Fakültesi, Türkçe Öğretmenliği Bölümü.

Yüksek Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Programı.

Bildiği Yabancı Diller : İngilizce.

Tarih : 25/06/2021

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- ☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.
 - ☐ Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun 3 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

25.06.2021

Leyla Karakuş

İNTİHAL RAPORU

Turnitin Orijinallik Raporu

İşleme kondu: 07-Haz-2021 12:04 +03
NUMARA: 1602030124
Kelime Sayısı: 34837
Gönderildi: 1

Benzerlik Endeksi	Kaynağa göre Benzerlik	
	İnternet Sources:	%8
%9	Yayımlar:	%1
	Öğrenci Ödevleri:	%2

EĞİTİMDE METİN MADENCİLİĞİ: TÜRKÇE METİNLERDE SÖZLÜK TABANLI DUYGU ANALİZİ
Leyla Karakuş tarafından

1% match (09-Haz-2020 tarihli internet)
<https://cran.r-project.org/web/packages/corpus/vignettes/corpus.html>

1% match (19-May-2021 tarihli internet)
<https://gist.github.com/hdk5/b4774da59e4716c929fc4db8f89948c7>

1% match (28-Tem-2018 tarihli internet)
<http://hothikayeler.blogspot.com/2011/09/scack-hikayeler-03.html>

< 1% match (01-Haz-2021 tarihli öğrenci ödevleri)
[Submitted to Akdeniz University on 2021-06-01](#)

< 1% match (07-Ağu-2018 tarihli öğrenci ödevleri)
[Submitted to Akdeniz University on 2018-08-07](#)

< 1% match (01-Ağu-2018 tarihli öğrenci ödevleri)
[Submitted to Akdeniz University on 2018-08-01](#)

< 1% match (10-Haz-2016 tarihli öğrenci ödevleri)
[Submitted to Akdeniz University on 2016-06-10](#)

< 1% match (02-Nis-2021 tarihli internet)
<http://adikerisim.akdeniz.edu.tr/bitstream/handle/123456789/2724/T03878.pdf?isAllowed=y&sequence=1>

< 1% match (24-Oca-2021 tarihli internet)
<http://adikerisim.akdeniz.edu.tr/bitstream/handle/123456789/2205/T03330.pdf?s=>

< 1% match (26-Nis-2021 tarihli internet)
<http://adikerisim.akdeniz.edu.tr/bitstream/handle/123456789/2929/T04089.pdf?isAllowed=y&sequence=1>

< 1% match (26-Nis-2021 tarihli internet)
<http://adikerisim.akdeniz.edu.tr/bitstream/handle/123456789/2936/T04096.pdf?isAllowed=y&sequence=1>

< 1% match (26-Nis-2021 tarihli internet)
<http://adikerisim.akdeniz.edu.tr/bitstream/handle/123456789/2937/T04097.pdf?isAllowed=y&sequence=1>