

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SÖZLÜKSEL BENZEŞİM YÖNTEMİ İLE ANDROID TABANLI
DUYGU ANALİZİ UYGULAMASI

YÜKSEK LİSANS TEZİ

Muharrem ÖNAL

Anabilim Dalı: Mekatronik Mühendisliği

Program: Bilgisayar Sistemleri

Danışman: Yrd. Doç. Dr. Cafer BAL

MART-2017

T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SÖZLÜKSEL BENZEŞİM YÖNTEMİ İLE ANDROID TABANLI DUYGU
ANALİZİ UYGULAMASI

YÜKSEK LİSANS TEZİ

Muharrem ÖNAL
(132134107)

Tezin Enstitüye Verildiği Tarih: 14 Nisan 2017

Tezin Savunulduğu Tarih: 17 Mart 2017

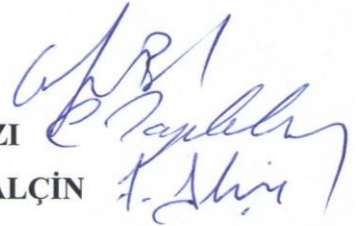
Anabilim Dalı: Mekatronik Mühendisliği

Programı: Bilgisayar Sistemleri

Danışman: Yrd. Doç. Dr. Cafer BAL

Doç. Dr. Erkan TANYILDIZI

Yrd. Doç. Dr. Ömer Faruk ALÇİN



MART-2017

ÖNSÖZ

Tez çalışmalarım süresince yardımlarından dolayı danışmanım Sayın Yrd. Doç. Dr. Cafer BAL başta olmak üzere, Doç. Dr. Erkan TANYILDIZI'na, çalışma arkadaşlarıma ve hayatımın her döneminde bana destek olan aileme teşekkürlerimi sunarım.

Muharrem ÖNAL

ELAZIĞ-2017



İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	II
İÇİNDEKİLER.....	III
ÖZET	V
SUMMARY	VI
ŞEKİLLER LİSTESİ	VII
TABLolar LİSTESİ	IX
KISALTMALAR LİSTESİ	X
SEMBOLLER LİSTESİ	XII
1. GİRİŞ.....	1
1.1. Problem Tanımı	2
1.2. Çalışma Konusu ve Kapsamı	3
1.3. Amaç	4
1.4. Motivasyon ve Özgün Değer	4
2. LİTERATÜR ÖZETİ.....	5
2.1. Duygu Analizi ile İlgili Çalışmalar	5
2.2. Android Uygulamalar ve Mobil Duygu Analizi ile İlgili Çalışmalar.....	15
3. DUYGU SÖZLÜĞÜ	18
3.1. Türkçe Duygu Sözlüğü Oluşturulması	18
3.2. Sözlüksel Benzeşim Yöntemi.....	18
3.3. WordNet Sözcük Veri Tabanı	20
3.3.1. Online WordNet	22
3.3.2. WordNet Benzerlik Ölçümleri	23
3.4. Sözlük Veri Tabanı.....	25
3.5. Veri Setleri	28
3.6. Türkçe Duygu Sözlüğü ve Veri Setleri ile Duygu Analizi.....	31
3.6.1. Başarı Ölçüm Yöntemleri.....	33
3.6.2. Duygu Analizi Testi Başarı Ölçümleri.....	35
4. MOBİL UYGULAMALAR	39
4.1. Mobil İşletim Sistemleri	39
4.1.1. Android Mobil İşletim Sistemi	39
4.1.2. iOS Mobil İşletim Sistemi	40

4.1.3.	Windows Phone İşletim Sistemi	41
4.2.	Mobil Uygulama Geliştirme Yaşam Döngüsü	41
5.	MOBİL DUYGU ANALİZİ UYGULAMA TASARIMI	46
5.1.	Mobil Duygu Analizi	46
5.2.	Mobil Duygu Analizi Uygulaması Yazılım Geliştirme Yaşam Döngüsü	46
5.3.	Mobil Duygu Analizi Veri Tabanı Özellikleri	49
5.4.	Duygu Analizi Algoritması Yapısı	50
5.5.	Mobil Duygu Analizi Uygulama Arayüzü	52
5.6.	Mobil Uygulamanın Yayınlanması	60
6.	BULGULAR.....	61
7.	SONUÇLAR ve TARTIŞMA.....	62
8.	ÖNERİLER.....	63
KAYNAKLAR.....		64
EKLER		69
ÖZGEÇMİŞ		75

ÖZET

Bu çalışmada, Android tabanlı mobil cihazlarda sözlüksel benzeşim yöntemiyle duygu analizi yapan bir mobil uygulama gerçekleştirilmiştir. Duygu analizi sözlüksel benzeşim yönteminde kullanılan duygu sözlüğü, kapsamayı üst anlam (hypernym), kapsanmayı alt anlam (hyponym) ve aynı seviyede olmayı eş düzeylik (coordinate) olarak ifade eden ve bunlara benzer ilişkiler içeren WordNet sözcük veri tabanı ve Resnik WordNet benzerlik ölçümü yöntemi kullanılarak geliştirilmiştir. Google, Yandex gibi sözlük uygulamalarıyla İngilizce sözcükler Türkçeye çevrilmiştir. Olumlu ve olumsuz olarak etiketlenmiş film ve otel yorumları ile veri setleri oluşturulmuştur. Duygu sözlüğü ve veri setleri ile duygu analizi yapılmıştır. Duygu analizinde veri setleri, normal ve etkisiz kelimesiz formlarıyla kullanılmıştır. Duygu analizi sonuçları doğruluk ve F-Ölçütü başarımları ölçütleriyle değerlendirilmiştir. Duygu analizinde yaklaşık %56 oranında başarı sağlanmıştır. Duygu sözlüğünün oluşturulma, duygu analizinde kullanılma sürecinde Java platform ve kütüphaneleri kullanılmıştır. Türkçe duygu sözlüğü ile Android Studio platformunda duygu analizi Android mobil uygulaması geliştirilmiştir.

Anahtar Kelimeler: Duygu Analizi, Sözlüksel Benzeşim, Duygu Sözlüğü, Mobil Uygulama

SUMMARY

Android Based Sentiment Analysis Application with Lexical Similarity Method

In this study, a mobile application which does sentiment analysis with lexical similarity method was performed in Android based mobile devices. Sentiment dictionary used in the sentiment analysis with lexical similarity method was developed using the WordNet lexical database method which expresses the scope as upper meaning (hypernym), inclusion as sub meaning (hyponym), being at the same level as coordination (coordinate) and the Resnik WordNet similarity measure method. English words to Turkish words were translated by using like Google, Yandex dictionary applications. Positive and negative labeled film and hotel comments used to make datasets. Sentiment analysis was made with sentiment dictionary and datasets. Datasets as normal and without stop words forms were used in sentiment analysis. Sentiment analysis results were evaluated by accuracy and F-Measure performance measures. Sentiment analysis success rate is about %56. Java platform and libraries were used in the process of generating sentiment dictionary and using it in sentiment analysis. Android mobil application developed in the Android Studio platform with Turkish sentiment dictionary.

Key Words: Sentiment Analysis, Lexical Similarity, Sentiment Dictionary, Mobile Application

ŞEKİLLER LİSTESİ

Sayfa No

Şekil 3.1. Duygu analizi yöntemleri	19
Şekil 3.2. WordNet veri tabanının fiil ve isim sözcük türleri arasındaki hiyerarşik yapısı.	20
Şekil 3.3. WordNet veri tabının sıfat sözcük türleri arasındaki küme yapısı	21
Şekil 3.4. WordNet online kullanım arayüzü	22
Şekil 3.5. Sözcük çeviri algoritması	26
Şekil 3.6. Duygu sözlüğü veri tabanı	28
Şekil 3.7. Film yorumları pozitif veri seti	29
Şekil 3.8. Film yorumları negatif veri seti	30
Şekil 3.9. Otel yorumları veri seti	31
Şekil 3.10. Pozitif ve negatif veri seti cümlelerinin polaritelerinin belirlenme ekran sonucu	32
Şekil 3.11. Otel yorumları veri seti cümlelerinin polaritelerinin belirlenme ekran sonucu	33
Şekil 3.12. Film yorumları doğruluk sonuçları karşılaştırılması	36
Şekil 3.13. Otel yorumları duygu analizi sonuçları karşılaştırılması	37
Şekil 4.1. Android mimarisi	40
Şekil 4.2. iOS mimarisi	40
Şekil 4.3. Windows phone mimarisi	41
Şekil 4.4. Yazılım geliştirme yaşam döngüsü kademeleri	42
Şekil 5.1. Duygu sözlüğü veri tabanı tablo ve sütunlar başlıkları	49
Şekil 5.2. Mobil duygu analizi uygulaması veri tabanı	50
Şekil 5.3. Duygu analizi algoritması akış şeması	51
Şekil 5.4. Mobil duygu analizi arayüz tasarımına ait görüntü	52
Şekil 5.5. Mobil duygu analizi uygulamasına ait arayüz görüntüsü	53
Şekil 5.6. Mobil duygu analizi uygulamasında mobil klavye ile analiz metni girişi arayüz görüntüsü	54
Şekil 5.7. Mobil duygu analizinin metnin panoya yapıştırılmasıyla uygulanması arayüz görüntüsü	55
Şekil 5.8. Mobil duygu analizi uygulamasında sesli ifadenin metne çevrilim arayüz görüntüsü	56
Şekil 5.9. Mobil duygu analizi uygulamasında metnin txt dosyası olarak kaydedilme arayüz görüntüsü	57

Şekil 5.10. Mobil duygu analizi uygulamasında kayıtlı txt dosyalarının açılma arayüz görüntüsü	58
Şekil 5.11. Analiz edilmiş metnin e-posta olarak gönderilme arayüz görüntüsü.....	59
Şekil 5.12. Analiz edilmiş metnin kısa mesaj olarak gönderilme arayüz görüntüsü.....	60



TABLÖLAR LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1. Veri tabanı sözcüklerinin tür dağılımı	25
Tablo 3.2. Sözlük çeviri yüzdeleri	26
Tablo 3.3. Resnik Java kütüphanesi ile elde edilen sözcük değerlerinin yön dağılımı	27
Tablo 3.4. Film yorumları dağılımı.....	29
Tablo 3.5. Otel yorumları dağılımı	30
Tablo 3.6. Film yorumlarının duygu puanı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları	35
Tablo 3.7. Film yorumlarının duygu yönü baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları	35
Tablo 3.8. Film yorumlarının duygu puanı ve kelime varlığı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları.....	36
Tablo 3.9. Film yorumlarının duygu yönü ve kelime varlığı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları.....	36

KISALTMALAR LİSTESİ

1-NN	: 1-Nearest Neighbours (En Yakın 1 Komşu)
API	: Application Programming Interface (Uygulama Programlama Arayüzü)
A-TF	: Augmented Term Frequency
AVD	: Android Virtual Device (Android Sanal Aygıtı)
BOW	: Bag of Words (Kelimeler Çantası)
B-TF	: Boolean Term Frequency
DAL	: Dictionary of Affect in Language (Dil Etkileşim Sözlüğü)
DD	: Duygu Değeri
DVM	: Destek Vektör Makineleri
DY	: Duygu Yönü
EVA	: Evaluative
HDFS	: Hadoop Distributed File System
HTML	: Hypertext Markup Language (Hiper Metin İşaretleme Dili)
IC	: Information Content (Bilgi İçeriği)
IG	: Information Gain (Bilgi Kazanımı)
ISEAR	: International Survey on Emotion Antecedents and Reactions
JDK	: Java Development Kit (Java Geliştirme Kiti)
JSON	: JavaScript Object Notation
KNN	: K-Nearest Neighbors (En Yakın K Komşu)
LA-TF	: Log-Average Term Frequency
L-TF	: Logarithmic Term Frequency
MAE	: Mean Absolute Error (Mutlak Hata Ortalaması)
ME	: Maksimum Entropi
MNB	: Multinomial Naive Bayes
NB	: Naive Bayes
NN	: Neural Networks (Sinir Ağları)
N-TF:	: Normalized Term Frequency
PCA	: Principal Component Analysis (Temel Bileşen Analizi)
PDA	: Personal Digital Assistant (Kişisel Dijital Yardımcı)
PMI	: Pointwise Mutual Informaiton (Noktasal Karşılıklı Bilgi)
POS	: Part of Speech, Position of Speech

RBEM	: Rule-Based Emmission Model
RF	: Random Forest (Rastgele Orman Karar Ağaçları)
RMSE	: Root Mean Square Error (Hata Karelerinin Ortalama Karekökü)
SARE	: Sentiment Analysis Research Environment
SDK	: Software Development Kit (Yazılım Geliştirme Kiti)
SMO	: Sequential Minimal Optimization (Sıralı Minimal Optimizasyon)
SQL	: Structured Query Language (Yapılandırılmış Sorgu Dili)
SSTF	: Structured and Statistical Text Features
SVM	: Support Vector Machine (Destek Vektör Makinesi)
TF	: Term Frequency (Terim Frekansı)
TF-IDF	: Term Frequency – Inverse Document Frequency (Terim Frekansı – Ters Metin Frekansı)
UML	: Unified Modelling Language (Tümleşik Modelleme Dili)
URL	: Uniform Resource Locator (Birörnek Kaynak Konumlayıcı, Tekdüzen Kaynak Bulucu)
WEKA	: Waikato Environment for Knowledge Analysis
WLLR	: Weighted Log Likelihood Ratio
XML	: Extensible Markup Language (Genişletilebilir İşaretleme Dili)

SEMBOLLER LİSTESİ

B	: Birleştirme
C, k	: Katsayılar
c₁, c₂	: Benzerlik değişkenleri
d	: Yön değişim sayısı
D	: Derinlik değişkeni, duyarlık
dist_{JC}	: Jiang-Conrath yaklaşımı uzaklığı
FN	: False Negative (Yanlış Negatif Değeri)
FP	: False Positive (Yanlış Pozitif Değeri)
IS-A	: Kelimeleri birbirine bağlayan link
i	: Terim sırası
K	: Anma
len	: İki eş anlamlı grup arasındaki en kısa yol
lso	: En yakın ortak kavram
n	: Toplam terim sayısı
P_K	: Kelime puanı
rel_{HS}	: Hirst-St-Onge yaklaşımı benzerliği
s	: Kelime skoru
sim_L	: Lin yaklaşımı benzerliği
sim_{LC}	: Leacock-Chodorow yaklaşımı benzerliği
TN	: True Negative (Doğru Negatif Değeri)
TP	: True Positive (Doğru Pozitif Değeri)
Y_K	: Kelime yönü

1. GİRİŞ

İnsan hayatında sürekli çevresiyle iletişim halindedir. Bu iletişim, insanların birbirlerine duygu ve düşüncelerini sözlü olarak aktararak gerçekleşebildiği gibi, görsel olarak mimikler veya beden dili ile dolaylı olarak ta gerçekleşebilmektedir. Bu tür iletişim kişilerin, çoğunlukla birbirleriyle eşzamanlı olarak etkileşiminden dolayı, birbirlerinin aktarmak istedikleri duygu ve düşünceleri rahatlıkla anlayabildikleri yöntemlerdir. Ancak bir metnin anlatmak istediğini ve ifade ettiği duyguyu anlamak büyük oranda mümkün gibi görünse de metnin uzunluğuyla beraber metindeki ima, mecaz gibi faktörler metindeki duygu ve görüşleri tam anlamıyla anlamayı güçleştirebilir. Günümüzde sosyal ağların son derece yaygınlaşması ile iletişimde duygu ve düşüncelerin aktarımını küçümsenemeyecek boyutta yazılı hale getirmiştir.

İnsanlar iletişim aracılığıyla, muhatabının düşüncesini, hislerini anlamaya ve ona uygun şekilde davranmaya çalışır. Daha geniş kapsamda ise belirli bir konuya ilişkin fikir geliştirme, öneri değerlendirme veya toplumsal bir görüş belirleme gibi büyük boyutta özellikle metindeki duygu ve fikirleri daha detaylı olarak belirleme çıkarımları için bir kişi veya bir grup insanın önemli bir zaman harcaması gerekir.

Günümüzde internetin birçok olanak sağlaması, sosyal ağların yanı sıra, ticari faaliyetler içinde ürün ve hizmetlerle ilgili yorumların son derece önemli hale gelmesine zemin oluşturmuştur. İnternette, e-ticaret, sosyal medya gibi sanal ortamlarda birçok insanın görüşlerini belirtebildiği yorumlar bulunmaktadır. Bunun gibi metinlerin sanal ortamda yaygınlaşmasıyla insanların duygu ve görüşlerini metinler üzerinden ortaya çıkarmanın önemi artmıştır. İnternet ortamındaki metinlerle duygu analizi yapılması birçok alanda faydalı amaçlar için kullanım alanı sağlayabilir. Satıcıların pazarlama politikalarını geliştirmek için müşteri memnuniyeti belirleyebilir. Anket ve araştırmalar daha hızlı ve kolay değerlendirilebilmesi ile örneklem uzayının genişletilmesi ve yapılan çalışmanın doğruluğu ve etkinliği arttırılabilir. Siyaset alanında vatandaşların görüşleri doğrultusunda yatırım planlamasından, kişisel ihtiyaçlara kadar büyük boyutlu bilgiler değerlendirilebilir.

İnternet gibi sayısal ortamdaki büyük boyutlu verilerin duygu analizinin manuel olarak yapılmasının zorluğunun yanı sıra, duygu analizini gerçekleştirmeye çalışan yöntem ve algoritmalar bilgisayarlar üzerinde çalışmaktadır. Dolayısıyla duygu analizini gerçekleştirmek için yazılı dokümanların bilgisayar ortamında olması belirli bir avantaj kazandırmaktadır.

Duygu analizi genel olarak metindeki duygunun pozitif, negatif ve nötr olarak tespit edilmesi şeklinde tanımlanabilir. Daha detaylı analizi için farklı duygu tespitleri yapılabilmektedir. Literatürde duygu analizi için, daha çok makine öğrenmesi ve sözlüksel benzeşim yöntemleri kullanılmaktadır. Makine öğrenmesi yönteminin başarımı, öğrenme verilerinin oluşturulması, verilerden temsil oranı yüksek öznitelikler çıkartılması gibi yoğun işlemleri gerektirir. Sözlüksel benzeşim yönteminde sözcüklerin duygu değeri tanımlanmış bir sözlükten faydalanılarak metindeki duygunun olumluluk, olumsuzluk ve yönsüzlük durumu tespit edilmektedir. Ayrıca sözlüksel benzeşim yöntemi, konu kapsamından bağımsız ve kullanımı kolaydır.

Türkçe duygu analizi gerçekleştirilen çalışmalar bulunmasına karşın, bununla ilgili bir yazılım bulunmamaktadır. Ayrıca mobil cihazlara yönelik herhangi bir uygulamada mevcut değildir. Günümüzde mobil uygulamaların geliştirilmesi ve kullanımı oldukça yaygındır. Mobil uygulamalar, özellikle açık kaynak kodlu mobil işletim sistemleri için rahat bir şekilde geliştirilebilen, birçok yararlı işlevi olan, hayatı kolaylaştıran, kolay ulaşılabilen mobil yazılımlardır. Mobil uygulamaların bir kısmı mobil işletim sistemler ile beraber hazır gelirken bir kısmı uygulama marketlerinden indirilir. Akıllı telefon ve tablet bilgisayarların yaygınlaşmasıyla beraber çoğu masaüstü bilgisayar işlevleri mobil cihazlarda da yerine getirilebilmektedir. Mobil cihazlar ihtiyacı karşıladığı durumlarda, mekâna bağımlı olmaması ve kullanım kolaylığıyla masaüstü ve dizüstü bilgisayarlara tercih edilebilmektedir. Mobil uygulama sektörünün popülaritesi son birkaç yılda önemli artışlar göstermiştir. Günümüzde mobil uygulama geliştirme ve kullanım oranı artışı devam etmektedir.

Türkçe duygu analizi uygulamasının sözlüksel benzeşim yöntemiyle, kullanım avantajları göz önüne alındığında mobil uygulama olarak geliştirilmesi ile metinlerin duygu analizi yapılabilir.

1.1. Problem Tanımı

Duygu analizi, ifade edilmiş bilgilerin duygusal yönünün olumlu, olumsuz ve nötr olmak üzere belirlenip kategorize edilmesidir.

Makine öğrenmesi yöntemiyle duygu analizi gerçekleştirmek için, eğitim veri seti yüklenip sonraki veriler bu öğrenmelere göre sınıflanır. Sözlüksel benzeşim yöntemiyle

genel bir sözlük oluşturulup konu kapsamından ve veri setinden bağımsız olarak duygu analizi yapılabilir.

Makine öğrenmesi ile duygu analizini uygulamak için genel kullanıma açık, özelleştirilmiş bir arayüz mevcut olmasa da makine öğrenmesi yöntemlerini başarıyla uygulayan WEKA (Waikato Environment for Knowledge Analysis) yazılımı [1] mevcuttur. Eğitim ve test kümelerinden öznitelikler seçilip öznitelik vektör yapısı oluşturulur. Veriler bir dosyaya kaydedilip WEKA yazılımına yüklenir. Eğitim kümesi ile oluşturulan bağıntı ile test kümesi sınıflandırılır. Sözlüksel benzeşim yöntemi ile duygu analizi için duygu sözlüğüne ihtiyaç vardır. Türkçe duygu sözlüğü ve Türkçe metinler için duygu analizi yapan bilgisayar veya taşınabilir cihaz uygulaması tespit edilememiştir.

1.2. Çalışma Konusu ve Kapsamı

Duygu analizi için sözlüksel benzeşim, istatistiksel yöntemler ve bu iki yöntemin birleşimi gibi çeşitli yöntemler bulunmaktadır. Sözlüksel benzeşim yönteminde pozitif ve negatif olarak puanlanmış sözcükler içeren bir duygu sözlüğü ile sınıflama yapılmaktadır. İstatistiksel yöntem ise KNN [K-Nearest Neighbors (En Yakın K Komşu)], NB (Naive Bayes), SVM [Support Vector Machine (Destek Vektör Makinesi)] gibi makine öğrenmesi yöntemleriyle yapılan duygu analizidir.

Bazı diller için SentiWordNet [2] gibi duygu sözlükleri mevcuttur. Bu sözlüklerle ve sözlüksel benzeşim yöntemiyle duygu analizi yapılabilmektedir. Sözcüklerin anlamsal olarak birbirine bağlı olduğu WordNet [3] gibi bir sözcük veri tabanı ile kelimelerin birbirine benzerliği değer olarak elde edilebilir. Kelimelerin pozitif ve negatif kelimelerine benzerlik değerleri ile duygu puanı bulunarak duygu sözlüğü oluşturulabilir. Duygu sözlüğü Türkçeye çevrilip Türkçe metinler için duygu analizi yapılabilir. Android Studio [4] gibi bir platformda mobil uygulama geliştirilerek mobil cihazlarda duygu analizi uygulanabilir.

Bu çalışma kapsamında Android mobil uygulamaların kullanım potansiyeli göz önüne alınarak Türkçe duygu sözlüğü ile mobil olarak duygu analizi yapan Android uygulama geliştirilmesi düşünülmüştür.

1.3. Amaç

Bu çalışmada, Türkçe metinlerin duygu analizinin sözlüksel benzeşim yöntemiyle yapılmasına imkân tanıyacak Türkçe duygu sözlüğü veri tabanının oluşturulması ve yazılım geliştirme yaşam döngüsü çerçevesinde bu veri tabanı kullanılarak Android Studio platformunda arayüzlerin tasarlanıp programlamalar yapılarak mobil duygu analizi uygulaması gerçekleştirilmesi hedeflenmiştir.

1.4. Motivasyon ve Özgün Değer

Mobil duygu analizi uygulaması, mobil cihazlarda Türkçe metinler için duygu analizi yapılmasını sağlayabilir. Duygu analizi sonuçları kısa mesaj veya e-mail yoluyla kişilere yönlendirilerek ve dosya olarak kaydedilip duygu analizinden etkili bir şekilde yararlanılabilir. Metin dosyaları cihaza kaydedilip analiz edilebilir. Sadece kısa metinler değil sayfalarca metinden oluşan ifadeler mobil olarak analiz edilebilir.

Araştırmalarda ulaşılamayan Türkçe duygu sözlüğü oluşturulabilir. Mobil uygulamalarda tespit edilemeyen Türkçe duygu analizi uygulaması geliştirilebilir.

2. LİTERATÜR ÖZETİ

2.1. Duygu Analizi ile İlgili Çalışmalar

Sadi Evren [5] çalışmasında literatürdeki duygu analizi kavramlarını, bu alanda kullanılan terimleri, öznitelik çıkarma yöntemlerini, karşılaşılan problemler ve çözümlerini incelemiştir.

Kelimelerin negatif veya pozitif olarak ayrılması ve yorumlardaki kelimelerin çoğunluk duygu yönünün yorumun duygu yönünü sınıflaması ilk ve en basit yöntemdir. Fikir madenciliği teknikleri, özellik tabanlı, frekans tabanlı ve model tabanlı teknikler olmak üzere üç farklı açıdan incelenebilir. Özellik tabanlı tekniklerde duygusal kutupsallığın hangi bakış açısına göre olumlu veya olumsuz olduğunu incelenir. Frekans tabanlı tekniklerde isim, sıfat, zarf, fiil kelimelerin sıklıkları (frekansları) üzerinden fikir madenciliği yapılır. Bu kelimelerin, fikirlerin dayanağı olduğu kabul edilmektedir. Örnek olarak 2011 yılında yapılan bir araştırmada [6] fikirlerin %60-%70'nin metindeki isimlere dayandığı bulunmuştur. Model tabanlı tekniklerde, daha önceki metinlerden oluşturulan modelleri etiketlenmemiş metinlere uygulayarak fikir elde edilmektedir.

Duygu analizinde karşılaşılan güncel problemler, kapalı duygu, alaycılık, alan bağımlılığı, ters kabuller, pragmatik, dünya bilgisi, öznellik yakalanması, varlık tanımlama, ters ifadeler olarak sıralanabilir.

En çok kullanılan öznitelik vektör yapısı yöntemleri TF [Term Frequency (Terim Frekansı)], TF-IDF'de [Term Frequency – Inverse Document Frequency (Terim Frekansı – Ters Metin Frekansı)] metinde geçen kelimeler çıkarılıp dokümanda geçtiği miktara göre hesaplamalar yapılır.

Standart ölçeklemede, ölçekleme etkisi özniteliklerin dağılımına uygulanır. Normalleştirmede veri setleri ortak bir yapıya indirgenir. PCA [Principal Component Analysis (Temel Bileşen Analizi)] veri üzerindeki bileşenlerin çıkartıldığı filtrelemelerdir.

Önder ve arkadaşlarının [7] Türkçe twitter mesajlarının duygu analizi için oluşturduğu sistem üç aşamadan oluşmaktadır. Birinci aşamada eğitim ve test verisi ön işlemeden geçirilmiştir. İkinci aşamada ön işlemeden geçirilen mesajlardan BOW [Bag of Words (Kelimeler Çantası)] ve N-gram model olmak üzere iki farklı yöntemle öznitelikler çıkarılmıştır. Daha sonra çıkarılan öznitelikler ağırlıklandırılmış ve elde edilen veri arff

dosya formatına dönüştürülmüştür. Son aşamada veriyi temsil eden en iyi öznitelikler seçilerek sınıflandırma yapılmıştır.

Twitter API kullanılarak geliştirilen Java tabanlı uygulama ile mesajların içerdiği his simgelerine göre (“:-)”, “:- (“ ...) eşit sayıda pozitif ve negatif toplam 20000 mesaj ile veri seti oluşturulmuştur. Ön işlemede pozitif ve negatif grupta bulunan his simgelerinin beraber bulunduğu çelişkili mesajlar ve terim sayısı ikiden az olan mesajlar elenmiştir. Mesajların içeriği küçük harfe dönüştürülmüştür. His simgeleri, URL [Uniform Resource Locator (Birörnek Kaynak Konumlayıcı, Tekdüzen Kaynak Bulucu)], hashtag kullanıcı adları, etkisiz kelimeler, üçten az karakterden oluşan kelimeler ve harf dışındaki karakterler silinmiştir. Terimlerin dokümandaki önem derecesini belirlemek amacıyla oluşturulan terim vektöründe TF, TF-IDF, Boolean olmak üzere üç farklı ağırlıklandırma yöntemi kullanılmıştır. WEKA kütüphanesi kullanılarak geliştirilen Java tabanlı uygulama ile makine öğrenmesi yöntemleriyle yapılan sınıflamada %66,06 oranında başarı elde edilmiştir.

Hatice ve Saliha Sıla'nın [8] çalışmasında makine öğrenmesi ile duygu analizinde veri seti performansını incelemek amacıyla gıda sektörünün çeşitli firma ürünlerine ait tweetlerden dengeli ve dengesiz olmak üzere iki veri seti oluşturulmuştur. Tweetlerdeki Türkçe ç,ğ,ı,ö,ş,ü karakterleri c,g,i,o,s,u karakterlerine dönüştürülmüştür.

Öznitelik vektör yapısı metotlarından TF kullanılmıştır. Deneyler 10 katmanlı çapraz geçerleme stratejisi ile WEKA 3.6 versiyon yazılımı kullanılarak yapılmıştır. NB, RF [Random Forest (Rastgele Orman Karar Ağaçları)], SMO [Sequential Minimal Optimization (Sıralı Minimal Optimizasyon)], Decision Tree, 1-NN [1-Nearest Neighbours (En Yakın 1 Komşu)] yöntemleri kullanılmıştır. Dengeli ve dengesiz veri setinde en yüksek doğruluk başarımları ölçütü oranları (%72,33-%66,40) SMO yöntemi ile elde edilmiştir. Dengeli veri setinin başarısının daha yüksek olduğu gözlenmiştir.

Ferhat Özgür [9] Eşle/İndirge tekniğinin DVM (Destek Vektör Makineleri) ile uygulanabilirliğini araştırmıştır. Eşle/İndirge yöntemi ile yüksek boyutlu veri seti alt kümelere parçalanarak problemin daha küçük parçalar şeklinde çözülmesi sağlanmıştır. Eşle fonksiyonu bir veri alanındaki veri çiftini birleştirerek başka bir alana veri çift listesi olarak göndermektedir. İndirge fonksiyonu ise eşle fonksiyonu çıktısına uygulanıp yeni indirgenmiş liste oluşturmaktadır. Eşle/İndirge yöntemi uygulanarak oluşturulmuş yeni liste başka bir Eşle/İndirge yöntemine tabi tutularak daha fazla sayıda alt çıktı matrisine dönüştürülebilmektedir.

Veri setinde Türkiye’de bulunan 108 adet devlet üniversitesi ve 66 adet vakıf üniversitesi twitter hesaplarındaki mesajlardan oluşmaktadır. Mesajlar Twitter Stream API v1.1 kullanılarak geliştirilen Java tabanlı uygulama ile elde edilmektedir. Veri setindeki etkisiz kelimeler elenmiş, metin vektör uzayı şekline dönüştürülmüş, vektör uzayında öznitelik seçimi yapılmıştır. İki ve üç sınıflı sınıflandırma için her sınıf grubundaki tweetler birbirine yakın sayıda seçilmiştir. Sınıflandırmalar olumlu, olumsuz mesajlar birlikte ve ayrı şekilde değerlendirilerek doğruluk oranları belirlenmiştir. Her üniversitenin mesajları ayrı değerlendirilmiştir. İki sınıflı sınıflama için olumlu ve olumsuz mesajlar ayrı değerlendirilerek en yüksek %95’in üzerinde doğruluk başarı oranına ulaşılmıştır. Olumlu ve olumsuz mesajlar birlikte değerlendirildiğinde bütün metin için %50 civarında olumluluk, %49 civarında olumsuzluk oranı elde edilmiştir. Üçlü sınıflama için ayrı sınıflamada en yüksek doğruluk oranı olumlu mesajlar için %60’ın üzerinde, olumsuz ve nötr mesajlar için %90 üzerinde bulunmuştur. Olumlu, olumsuz, nötr mesajların birlikte sınıflanmasında en iyi doğruluk dağılım oranı %10 civarında olumsuzluk, %30 civarında olumluluk, %60 civarında nötr oranı olarak görülmektedir.

Zeynep [10] öznitelik belirleme ve vektör yapısında WLLR (Weighted Log Likelihood Ratio) skorlaması, TF, TF-IDF, BOW, N-gram yöntemlerini kullanmıştır. Veri setleri olarak ISEAR (International Survey on Emotion Antecedents and Reactions) veri seti Türkçeye çevrilmiş ve Türkçe peri masalları kullanılmıştır. ISEAR veri setinde toplam yedi duygu ifadeleri içermektedir bunların dördü olan mutluluk, hüzn, öfke ve korku içeren ifadeler çalışma kapsamında Türkçeye çevrilmiştir. Diğer veri seti yirmi beş çocuk peri masallarından oluşmaktadır. Ön işleme aşamasında noktalama işaretleri, morfolojik analizler ve kelimelerin yazım denetimi uygulanmıştır. Morfolojik analiz, Türkçe metinler için sık kullanılan açık kaynak kodlu Zemberek doğal dil işleme kütüphanesi ile yapılmıştır. Deneylerde ISEAR veri setinde en yüksek doğruluk başarı oranı (%81.34) Complete NB sınıflandırıcı, 2400 öznitelik ve TF öznitelik vektör yapısı metoduyla elde edilmiştir. Türkçe peri masalları en yüksek doğruluk başarı oranına (%74.50) NB sınıflandırıcı, 800 öznitelik, TF yöntemiyle ulaşılmıştır.

Burak İbrahim [11] duygu analizinde makine öğrenmesi ve sözlüksel tabanlı yaklaşımların karşılaştırmasını yapmıştır. Veri seti için bir film web sitesinden kullanıcı yorumları indirilmiştir. En az iki farklı dokümanda geçmeyen öznitelikler öznitelik kümesinden çıkarılmıştır. Öznitelik vektörünün boyutunun gereksiz olarak artması engellenmiştir. Bazı yorumlardaki ç,ğ,ı,ö,ş,ü gibi Türkçe karakterlerin kullanılmaması

yanlışığı Turkish Deasciifier Java kütüphanesi ile kısmen hatayla düzeltilmiştir. Özel isimler, yanlış yazılan, anlamı olmayan kelimeler Zemberek kütüphanesi yardımıyla veri setinden çıkarılmıştır. Aynı kütüphane kelime köklerini ayırmak için kullanılmıştır. Öznitelik vektör yapısı için TF-IDF, A-TF (Augmented Term Frequency), B-TF (Boolean Term Frequency), LA-TF (Log-Average Term Frequency), L-TF (Logaritmik Term Frequency), N-TF (Normalized Term Frequency), TF yöntemleri kullanılmıştır. Yapılan deneylerde en yüksek F-Ölçütü değerine (0,82), kelimeler köklerine ayrılmadan, etkisiz kelimeler elenerek, TF-IDF yöntemi kullanılarak SVM sınıflandırıcısı ile ulaşılmıştır.

Hatime Dilek [12] twitter mesajlarının duygu analizi ile pazar değerlendirilmesini yapmıştır. Twitter verilerinin transferi için Duygu Analizi yazılımı kullanılmıştır. Yazılımda veri kaynağı ile eğitim yapılmaktadır. Bu işlem için en az 2000 kelime gerekmektedir. Daha sonra yeni gelen ifadeler otomatik olarak sınıflandırılmaktadır. Avea, Turkcell ve Vodafone sözcüklerinin geçtiği tweetler veri tabanına çekilmiş Excel dosyası olarak kaydedilmiştir. Aynı ifadede iki operatörün ismi geçtiği tespit edilmişse ifade silinmiştir. Operatörlerin kendi resmi sayfalarından yayınlanmış oldukları tweetler, uygulamadan yapılan konum paylaşımları silinmiştir. Metne, harf olmayan karakterlerin çıkarılması, tamamının küçük harfe çevrilmesi gibi ön işleme işlemleri uygulanmıştır. Öznitelik vektör yapısında TF-IDF yöntemi kullanılmıştır. Twitter kullanımı ile operatör abone sayısı ve geliri paralellik gösterirken bu uyum, yapılan duygu analizi ile elde edilen pozitif veri dağılım oranı ile operatör sayısı ve geliri için söylenememektedir.

Aysun ve arkadaşları [13] duygu analizinin SVM yönteminin performansını etkileyen parametre değişimlerinin sınıflandırma performansı üzerinde etkilerini incelemişlerdir. Olumlu, olumsuz, nötr metin içeren iki veri seti ve dört farklı ruh hali içeren üçüncü veri seti ile yapılan deneyde parametre seçimine bağlı olarak en yüksek ve en düşük başarı değerleri 0,633-0,429, 0,751-0,305, 0,752-0,342 olarak değişmektedir. Farklı parametrelerin SVM başarısı üzerinde etkili olduğu görülmüştür.

Eyüp Sercan ve arkadaşları [14] Duygusal Twitter programı ile verileri analiz etmişlerdir. Duygusal Twitter programında işleyişin tamamı neredeyse otomatik olarak yapılmakta ve tüm işlemler bir çatı altında toplanmaktadır. Eğitim veri setindeki tweetler Duygusal Twitter arayüzü kullanılarak elle etiketlenmiştir. Veri ön işlemede gereksiz karakter, kelimeler, linkler temizlenip JSON (JavaScript Object Notation) formatı için Türkçe karakter dönüşümleri yapılmıştır. Sözlük yöntemi ile yapılan duygu analizinde olumlu kelime, olumsuz kelime, kelime çiftleri ve jargon sözlüğü olmak üzere dört farklı

sözlük kullanılmıştır. Sözlük ve karakter tabanlı N-gram yöntemleriyle sırasıyla 0,72 ve 0,70 F-Ölçütü değerlerine ulaşılmıştır.

Fırat [15] duygu analizinde Ki-Kare (Chi-Square) ve IG [Information Gain (Bilgi Kazanımı)] yöntemleri ve değişik sayılarda özniteliklerle başarı oranlarını karşılaştırarak incelemiştir. Veri seti bir web sitesinde bulunan 0,5 ile 5,0 aralığında puanlanmış film yorumlarından oluşmaktadır. Olumlu, olumsuz ve nötr kategorileri olmak üzere her birinde eşit sayıda toplam 80.100 adet yorum kullanılmıştır. Veri ön işlemede Zemberek kütüphanesi kullanılarak kelimeler köklerine ayrılmış, anlamsız kelimeler silinmiştir. SQL Server 2012 Enterprise veri tabanına kaydedilmiştir. Veri tabanının Reviews tablosunda, film yorumlarının bağlantı adresi, içeriği, puanı, hangi kategoriye ait olduğu, köklerden oluşan cümle halini barındıran kısımların tutulduğu tablodur. Words tablosunda, veri ön işlemeden geçmiş 80.100 adet film yorumunun kelimeleri tutulur. Filtered tablosunda depolanan yorumların orjinalleri ve veri ön işlemeden geçmiş köklerinin tutulduğu tablodur. Distinctwords tablosu yorumlardaki kullanılmış olan kelime köklerinin bulunduğu tablodur. Deneyde SVM sınıflandırıcısı kullanılmıştır. En yüksek F-Ölçütü değeri 375 öznitelik ile olumlu kategoride 0,83 olumsuz kategoride 0,84 olarak bulunmuştur. Aynı veri seti 375 ve 1000 öznitelik için NB sınıflandırıcısı kullanılarak sınıflandırılmıştır. SVM sınıflandırıcısının başarısının daha yüksek olduğu gözlenmiştir.

Hakan [16] var olan Türkçe duygu analizi çalışmalarının başarısını yükseltmek amacıyla yeni veri setlerinde ve Türkçeye özel veri hazırlama algoritmaları ile duygu analizi yapmıştır. Book, film, shopping, all adlarında veri setleri oluşturulmuştur. Zemberek ile hece kontrolü ve karakter düzeltimi yapılmıştır. Veri setindeki kelimeler küçük harflere dönüştürülmüştür. Bir ya da iki karakterli kelimeler elenmiştir. Olumsuzlaştıran değil kelimesi, kendinden önceki kelime ile birleştirilmiştir. Arff dosyası içinde özniteliklerde 0 negatif 1 pozitif sınıfı belirtmek üzere hazırlanmıştır. Deneylerde en yüksek doğruluk başarı oranı (%85,80) kitap yorumlarında NB sınıflandırıcı ile yakalanmıştır.

Mahmut ve M. Fatih [17] eşit sayıda toplam 12000 tweetten oluşan iki veri seti üzerinde gözetimli ve gözetimsiz yöntemlerin karşılaştırılması amacıyla duygu analizi yapmıştır. Gözetimli yöntemlerin daha başarılı olduğu deneylerde en yüksek başarı oranına (%65,5) SMO algoritmasıyla elde edilmiştir.

Sadi Evren ve Khaled [18] çalışmalarında beş farklı Türkçe web sitesinden alınan 1,6 milyondan fazla etiketli yorum üzerinde TF-IDF yöntemi ve SVM, KNN, RF

sınıflandırıcıları ile duygu analizi yapmışlardır. KNN sınıflandırıcı ile yakalanan %57,69 doğruluk başarı oranı bütün özellikler kullanılarak %61,84'e çıkarılmıştır.

Yeliz [19] Java dili ağırlıklı geliştirdiği duygu analizi sisteminde, makine öğrenmesi algoritmaları için sisteme Mahout kütüphanesini eklemiştir. Doğal dil işleme CoreNLP kütüphanesi yanında kelimelerin pozitif ve negatif değerlerinin bulunduğu duygu sözlüğü sisteme eklenmiştir. Doğruluk ölçümlerinde, MAE [Mean Absolute Error (Mutlak Hata Ortalaması)] ve RMSE [Root Mean Square Error (Hata Karelerinin Ortalama Karekökü)] yöntemleri kullanılmıştır. Veri seti olarak Amazon ve MovieLens'in kullanıldığı deneyde en yüksek RMSE başarı değeri (0,92) işbirlikçi filtreleme yöntemi ile elde edilmiştir.

Rahim [20] Türkçe duygu analizi yapmak için kapsamlı bir sistem tasarlayıp geliştirmiştir. Veri ön işlemede, olumsuzluk getiren kelime veya ekleri belirlemek için olumsuzlaştırma, polaritenin zayıf veya güçlü olduğunu belirtmek için kelime yoğunlaşması, koşul cümlelerindeki görünen polaritenin değişimini kontrol etmek için koşul cümleleri, cevap çıkarmak yerine duygu ifade eden retorik cümleler için retorik sorular, deyimlerin gerçek polaritesini belirlemek için deyimsel kullanımlar aşamaları gerçekleştirilmiştir. SentiWordNet ve SentiTurkNet'in kullanıldığı sınıflandırma deneylerinde en yüksek doğruluk oranı (%91.11) SMO, NN [Neural Networks (Sinir Ağları)], Logistic sınıflandırıcılarının kombine kullanılmasıyla yakalanmıştır.

Önder [21] literatürdeki duygu analizi başarısını arttırmak için farklı ön işleme, etiketleme, sınıflandırma ve benzerlik yöntemleri etkileri incelemiştir. Duygu analizi sistemi ön işleme, veri işleme ve sınıflandırma aşamalarından oluşmaktadır. Ön işleme, kelimelere ayırma, etkisiz kelimelerin çıkarılması, kök bulma işlemlerinden oluşmaktadır. Veri işlemede ön işlenmiş metinden öznitelik çıkarılıp ağırlığı belirlenmekte ve veri arff dosyasında saklanmaktadır. Son aşamada verilerle sınıflandırma yapılmaktadır. Öznitelikler BOW, N-gram ve SSTF (Structured and Statistical Text Features) modelleri kullanılarak çıkarılmıştır. Öznitelik vektör yapısında TF-IDF yöntemi kullanılmıştır. Boyut indirgeme aşamasında Ki-Kare, Karşılıklı Bilgi, Bilgi Kazanımı, Korelasyon Tabanlı Öznitelik Seçme, öznitelik aşamasında Saklı Anlam Analizi yöntemleri kullanılmıştır. En yüksek doğruluk başarı oranı (%98,61) tüm öznitelikler kullanılarak SVM sınıflandırıcısı ile yakalanmıştır.

Erkut [22] Türkçe tweetlerin R programlama diliyle işlenebilmesini amaçlamıştır. 3579 kelimelik -5 ile +5 arasında puanlanmış bir duygu sözlüğü oluşturulmuştur. Veri

setindeki tweetler duygu puanı ve tekrar sayıları ile skorlanmıştır. SVM, RF, Boosting, ME (Maksimum Entropi) yöntemlerinin kullanıldığı sınıflandırmada en yüksek doğruluk başarı oranı (%54) RF yöntemiyle elde edilmiştir.

Batuhan [23] forumlardaki yorum ve incelemelerin toplanmasıyla oluşturulan veri setine daha önceden İngilizce için geliştirilmiş işlemleri uygulamıştır. Duygu kelimelerini vurgulayan, duygu yükselticiler ve duygu yönünü değiştiren duygu değiştirenleri kullanılmıştır. PMI [Pointwise Mutual Information (Noktasal Karşılıklı Bilgi)] istatistiksel ölçüm kullanılmıştır. Veri seti MySQL veri tabanında tutulmaktadır. Deneylerde %60 duyarlılık, %85 anma başarı oranına ulaşılmıştır.

Tugay [24] sosyal medya verileri üzerinde duygu analizi ve nitelik çıkarımı üzerine çalışma yapmıştır. Twitter verileri Excel dosyasında saklanıp insan eliyle etiketlenmiştir. Daha sonra Cassandra veri tabanına kaydedilmiştir. Kelimeler Apache Lucence kütüphanesi ile N-gramlara ayrılmıştır. Nitelik çıkarımı için terim varlığı ve frekans analizi yapılmıştır. Yapılan deneyde en yüksek doğruluk oranı (%89,84) Logistic Regresyon yöntemiyle yakalanmıştır.

Yeliz ve Sevinç İlhan'ın [25] çalışmalarının amacı kullanıcı puanlı tavsiyelerin, kullanıcı puan yerine duygu analizi ile bulunan değerler ile büyük verilerde uygulanmasıdır. Büyük miktarda veri depolamak için HDFS (Hadoop Distributed File System) kullanılmıştır. Makine öğrenmesi yöntemi kullanımı için Mahout kütüphanesi sisteme eklenmiştir. Harita indirgeme yöntemleri ile tüm eğitim veri seti okunmuş ve öğrenme modeli oluşturulmuştur. Doğal dil işleme için CoreNLP kütüphanesinden faydalanılmıştır. Sözlük tabanlı duygu analizi modülü için kelimelerin skorlandığı duygu sözlüğü oluşturulmuştur. Yapılan deneylerde en yüksek RMSE değerine (1,3418) Amazon Electronics puanlamalarında ulaşılmıştır.

Mustafa [26] duygu analizini twitter verileri ile SMO ve MNB (Multinomial Naive Bayes) yöntemleriyle uygulamıştır. Twitter mesajları Twitter API ile elde edilmiştir. Mesajlar olumlu ve olumsuz yüz ifadelerine (“:-”), “:~”, “:-(”, “:(“ vb.) göre pozitif veya negatif olarak etiketlenmiştir. Veri ön işlemede iki zıt duygu ifadelerini birlikte içeren mesajlar elenmiştir. Gönderi adresi, hashtag kullanıcı adları, etkisiz kelimeler ve URL’ler, sıralı olarak ikiden fazla tekrar eden harflerin ilk çiftten sonraki harfleri silinmiştir. Mesajların etiketlenmesinden sonra yüz ifadeleri kaldırılmıştır. 1-4-gram yöntemleri, MNB ve SMO algoritmalarının kullanıldığı deneylerde en yüksek doğruluk başarı oranına (%77.15) Bigram yöntemi ve MNB algoritması ile ulaşılmıştır.

Muhammet Yasin [27] Türkçe ve İngilizce için makine öğrenmesi yöntemleriyle duygu analizi yapmıştır. Ön işlemede olumsuzluk eki alan kelimeler neg_kelime kökü şeklinde etiketlenmiş, yüz ifadeleri işlenmiştir. Öznitelik çıkarma işlemi için kelime çantası, N-gram yöntemi kullanılmıştır. Özniteliklerden ayırt ediciliği yüksek olan kelimelerin seçilmesi ve öznitelik sayısının düşürülmesi için bilgi kazanımı tekniği kullanılmıştır. Bilgi kazanımı tekniği öznitelik teriminin doğru sınıflandırmaya katkısını ölçer. Öznitelik vektörü oluşturmada Binary ve TF-IDF yöntemleri kullanılmıştır. Türkçe veri seti için en yüksek doğruluk başarı oranı (%80,86) Binary, 5-gram yöntemleri, SVM sınıflandırıcısıyla, İngilizce veri seti için en yüksek başarı oranı (%84,55) 2-gram, TF-IDF yöntemleri ve SVM sınıflandırıcısıyla elde edilmiştir.

Süleyman Fatih [28] sosyal medya metinlerine makine öğrenmesi metotlarıyla duygu analizi uygulamıştır. İki ayrı web sitenin ürün yorumlarından pozitif ve negatif olmak üzere veri setleri oluşturulmuştur. Tr-tagger morfolojik analizde kullanılmıştır. Negatif yapan kelimeler olarak “değil” ve “ne...ne” biçimindeki kalıplar ele alınmıştır. Binary, frekans öznitelik metotları ve BayesNet, VotedPerceptron ve J48 karar ağacı makine öğrenmesi algoritması kullanılmıştır. Yapılan deneylerde en iyi doğruluk oranına (%94,9) Binary metodu ve BayesNet sınıflandırıcı ile ulaşılmıştır.

Mus'ab Habib [29] duygu analizi problemlerini çözmek için kamuya açık bir sistem olan duygu analizi araştırma ortamı SARE'yi (Sentiment Analysis Research Environment) sunmaktadır. Uygulama Persistence, Data Acces, Logic, Web Application katmanlarından oluşmaktadır. Corpus Reduction modülüyle belirli bir veya daha fazla POS (Part of Speech) türüne göre dokümanı ortak köklere optimize ederek indirger. Aspect Expression Extraction modülü verilerin bir bölümünün SVM sınıflandırıcı için eğitim kümesi olarak kullanılmasını sağlar. Aspect Lexicon Builder modülü ile anahtar kelimeler içeren bir görünüş sözlüğü elde edilir. Aspect Based Sentiment Analysis modülü tüm kelimelerin duygu skorlarını hesaplamayı içerir. SARE programıyla yapılan deneyde en yüksek %55 F-Ölçütü oranı elde edilmiştir.

Sinem [30] çalışmasında kısa Twitter mesajlarının işlenmesini kolaylaştırmak için yeni bir Türkçe tweet veri seti oluşturmuştur. Kızgınlık, korku, sevinç, iğrenme, üzüntü şaşırma kelime ve eş anlamlılarını içeren tweetler Twitter Streaming API ile toplanmıştır. Veri ön işlemede link ve kullanıcı adları gibi duygu içermeyen ifadeler ve etkisiz kelimeler silinmiştir. Yanlış yazılmış kelimelerin düzeltilmesi için bazı metotlar uygulamıştır. Öznitelik belirleme ve vektör yapısında N-gram, POS etiketi TF, TF-IDF, IG, WordNet

Affect Lexicon yöntemleri kullanılmıştır. Olumsuzluk taşıma, özel morphem analizleri yapılmıştır. En yüksek sınıflandırma doğruluk oranı (%69,92) 800 öznitelik, TF, TF-IDF yöntemleriyle ve SVM sınıflandırıcısı ile elde edilmiştir.

Hu ve Liu [31] çalışmalarında kullanıcı yorumlarını sınıflamışlardır. Olumlu ve olumsuz kelimelerden oluşan kelime listesi elle hazırlanıp WordNet gibi bir sözlükle kelimelerin eş ve zıt anlamları aranarak genişletilmiştir. Çalışma üç aşamada gerçekleştirilmiştir. Birinci aşamada, müşteri yorumlarında, doğal dil işleme teknikleri uygulanmıştır. İkinci aşamada görüş cümlesinin negatif veya pozitif olup olmadığı belirlenmiştir. Alt görevler olarak, bazı sıfat sözcüklerini doğal dil işleme yöntemleriyle belirleme, bu kelimeler için anlamsal yönelimi belirleme ve cümledeki görüş yönelimlerini etkili bir algoritma ile bulma işlemleri yapılmıştır. Üçüncü aşamada sonuçlar özetlenmiştir. Dijital Camera1, Dijital Camera2, Cellular Phone, Mp3 Player, DVD Player veri setlerinde duygu analizi yapılmıştır. %84 başarı oranı gösterilmektedir.

Kamps ve arkadaşları [32] mesafe tabanlı algoritma tekniğini kullanarak kelimelerin duygu yönünü, kelime listesine olan uzaklıklarını bulmuşlardır. WordNet uzaklık ölçümleri, Rada Et Al (1989), Hirst and St-Onge (1998), Leacock and Chodorow (1998), Resnik (1995) Lin (1998) yöntemleriyle yapılmaktadır. Mesafe tabanlı algoritmada EVA (Evaluative) fonksiyonu her sözcük için -1 ile +1 arasında bir uzaklık değeri vermektedir.

Steinberger ve arkadaşları [33] bazı dillerde duygu sözlükleri oluşturmuşlardır. Sözlük oluşturma yaklaşımı, öznel terimlerin, yarı otomatik olarak, çok dilli terim Ontopopulis algoritması kullanılarak İngilizce ve İspanyolca olarak toplanması ve diğer dillere yansıtılmasıyla meydana gelmiştir. Sonraki aşamada diğer terimler bulunmaktadır. Yüksek puanlanmış bazı terimler ile diğer içerik kalıpları doldurulmaktadır. Sözlükler elle kontrol edilmiştir. Çekçe, Fransızca, Almanca, İtalyanca, Rusça dillerine ortalama %90 doğruluk oranı ile çeviri gerçekleştirilmiştir.

Esuli ve Sebastiani [34] makine öğrenmesi yöntemleri ile WordNet sözlüğündeki kelimelerin pozitif, negatif, nötr puanları hesaplamışlardır. Bu çalışma doğrultusunda, her kelimenin üç değeri toplamı 1 olacak şekilde duygu değeri gösterimi yapılan SentiWordNet online duygu sözlüğü yayınlanmıştır. Üç farklı yaklaşım ile deneyler yapılmıştır. Birinci yaklaşımda ilk sınıflandırıcı, terimi öznel ve nesnel olmasına göre sınıflandırmıştır. İkinci sınıflandırıcı ise öznel olarak sınıflandırılmış terimi pozitif veya negatif olarak sınıflandırmıştır. İkinci yaklaşım da iki sınıflandırma içerir. Birinci

sınıflandırma pozitif, pozitif değil sınıflandırması iken, ikinci sınıflandırma negatif, negatif değil sınıflandırmasıdır. Pozitif, negatif ile pozitif değil, negatif değil olarak sınıflandırılmış terimler objektif olarak tutulmuştur. Üçüncü yaklaşımda, pozitif, negatif ve objektif kategorileri eşit statüde, üçlü sınıflandırıcı öğrenme ile her terim üç kategori içinde birbirine göre sınıflandırılmıştır.

Akbaş [35] çalışmasında sözlüksel benzeşim ve makine öğrenmesinin birleşik olarak kullandığı bir yöntemle duygu analizi yapmıştır. Sözlük, metni duygu yönünden etkilemeyen kelimelerin elenmesinde kullanılmıştır. Metindeki kelimelerin öznitelikleri belirlendikten sonra öznitelik vektörleri oluşturulup, makine öğrenmesi algoritmalarıyla eğitim ve test verileri duygu analizinde kullanılmıştır. Sentistrength'in Türkçe konfigürasyonu olarak 1-5 arasında ilişki derecelendirme ölçüsüne sahip, pozitif ve negatif sözcükler içeren sözlük Türkçe kelimeler için düzenlenmiştir. Duygu analizinde %85 oranında başarı yakalanmıştır.

Alaettin [36] duygu analizini, sözlüksel benzeşim ve makine öğrenmesi ile uygulamış, sonuçları karşılaştırmıştır. Sözlüksel benzeşim yöntemi için WordNet kullanılarak geliştirilmiş SentiWordNet duygu sözlüğü Türkçeye çevrilmiştir. Terim duygu puanı, pozitif puandan negatif puan çıkarılarak elde edilmiştir. Tekrar eden eş sesli kelimelerin duygu puanları SentiWordNet sitesindeki örnek kodda bulunan ağırlıklandırma formülü ile birleştirilmiştir. Tureng.com, Zargan.com, tr.bab.la sözlükleri ve Google Translate API servisi kullanılarak İngilizceden Türkçeye yapılan çeviride 39,965 terimin %97,5'i çevrilmiştir. Veri seti, Html Agility Pack kütüphanesi kullanılarak elde edilen film ve otel yorumlarının Microsoft SQL Server 2012 programı kullanılarak veri tabanına kaydedilmesiyle oluşturulmuştur. Ön işleme de harf dışındaki karakterler silinmiş, Zemberek kütüphanesi kelime denetle fonksiyonu ile yazım yanlışı denetlenmiş, öner fonksiyonunun kelime önerisi ile düzeltilmeye çalışılmıştır. Sözlüksel benzeşim yöntemi duygu analizi cümle polaritesi göz önüne alınarak yapılmıştır. Cümle polaritesi, cümledeki kelime duygu puanlarının toplamının negatif veya pozitif olmasına göre belirlenmektedir. Sözlüksel benzeşim yöntemiyle yapılan deneyde en yüksek doğruluk değeri (0,80) otel yorumlarında kelimelerin gövde hali ile yapılan deneyde elde edilmiştir. Aynı veri seti makine öğrenmesi yöntemlerinde duygu analizine tabi tutulmuştur. 65000 yorumun tamamı test ve eğitim kümesi olarak ikiye ayrılmıştır. En yüksek doğruluk değerine (0,84) film yorumlarında, Ki-Kare yöntemi, 375 öznitelik ve SVM sınıflandırıcı ile ulaşılmıştır.

Cumali [37] makine öğrenmesi ve sözlük tabanlı metotları yeni özelliklerle veri setlerine uygulamıştır. Twitter mesajları ve bir film web sitesinde bulunan metinlerle veri seti oluşturmuştur. Veri ön işlemede Zemberek kütüphanesi ile Türkçe olmayan karakterler ve basit yazım hataları düzeltilmiştir. Olumsuzluk meydana getiren varlık/yokluk ekleri ve birleşik kelimelerin kullanılması ele alınmıştır. Web site ve resim linkleri silinmiş ve gürültü oluşturan karakter tekrarları normalize edilmiştir. Olumsuzluk kelimeleri duygu barındıran kelimelerden sonra gelmişse duygu puanı -1 ile çarpılarak hesaba katılmıştır. Duygu sözlüğündeki her eleman -5 ile +5 arasında duygu değerine sahiptir. His simgelerinin duygu değerlerinin bulunduğu his simgeleri sözlüğü oluşturulmuştur. Sözlük tabanlı duygu analizi en yüksek F-Ölçütü oranı (%79,5) film yorumları veri setinde tüm veri ön işleme modülleri kullanılarak ulaşılmıştır. Makine öğrenmesi metodunda ise en yüksek başarı oranı (%89,5) NB sınıflandırıcı ve TF-IDF vektör yapısı yöntemiyle yakalanmıştır.

2.2. Android Uygulamalar ve Mobil Duygu Analizi ile İlgili Çalışmalar

Nisarg ve Rahila [38] çalışmalarında Android mobil cihaz işletim sistemi özelliklerini incelemiş, Android, Symbian ve Windows Mobile işletim sistemleri karşılaştırmalarını yapmışlardır. Android işletim sistemi Google tarafından 2007 yılında akıllı telefon, PDA [Personal Digital Assistant (Kişisel Dijital Yardımcı)] ve netbooklar için kullanıma sunulmuştur. Android uygulamalar Java tabanlıdır ve bu faktör sanal makine kullanımını gerektirir. Android Dalvik olarak adlandırılan kendi sanal makinesini kullanmaktadır. Android programlamada Android SDK, JDK [Java Development Kit (Java Geliştirme Kiti)] ve Eclipse Java platformu kullanılmaktadır. Taşınabilirlik, güvenilirlik, bağlanabilirlik, açık platform, çekirdek boyutu, standartlar, özel özellikler konularından Symbian, Windows Mobile, Android karşılaştırmalarında en iyi işletim sistemi olarak Android öne çıkmaktadır.

Mustafa ve arkadaşları [39] Android işletim sisteminin gelişimi hakkında araştırmalar yapmıştır. İlk Android işletim sistemi 2007 yılında Android 1.0 sürümü ile kullanıma sunulmuştur. Google Android kaynak kodlarını 2008 yılında sunmuştur. Cupcake Android 1.5 sürümü ile dokunmatik ekran, yapay ekran klavyesi gibi yenilikler ortaya konmuştur. 1.5 sürümünden sonra her sürüme kod ad verilmiştir. 1.6 Donut sürümü ile yeni ekran çözünürlüğü standartları ve text to speech özelliği getirilmiştir. Text to

speech özelliği ile konuşarak arama yapılabilir. Eclair 2.0 sürümü ile birlikte Android market uygulamaları App Store uygulamalarından daha yaygın hale gelmiştir. Froyo 2.2 sürümü ile 2.1 sürümü iyileştirilmiş, bunun yanında bluetooth 2.1, HTML 5 yenilikleri getirilmiştir. Gingerbread 2.3 sürüm ile internet arama özelliği getirilmiş, Android market ve klavye düzeni yenilenmiştir.

Mevlüt ve Emre [40] mobil sipariş sistemi Android uygulaması geliştirmişlerdir. Google tarafından geliştirilen JDK, AVD [Android Virtual Device (Android Sanal Aygıtı)] içeren Android Studio uygulaması programlamada kullanılmıştır. Web uygulamalarında kullanılan, birden çok teknolojinin kullanıldığı Ajax programlama yaklaşımı kullanılmıştır. JSON veri değişim formatı Ajax ile kullanılmıştır. Programda web sunucularda en çok tercih edilen MySQL veri tabanı kullanılmıştır.

Margaret [41] çalışmasında Android işletim sisteminin akıllı telefonlarda kullanımına ve Android işletim sistemi yazılım platformu yapısına yer vermiştir. Android işletim sistemi modifiye edilmiş Linux çekirdeği üzerine kurulmuştur. Sistem bileşenleri Java, C, C++ ve XML [Extensible Markup Language (Genişletilebilir İşaretleme Dili)] ile yazılmıştır. Android mimarisi iPhone uygulamalarından daha güvenlidir. Android güvenlik kontrolünü kullanıcıya bırakmıştır. Erişime izin verildiğinde yazılımlar tarafından bu izin kullanılabilir. Android marketler kötü amaçlı yazılımları sistemden kaldırmaktadır. Bu yazılımlarla ilgili ciddi hasar yaşanmamıştır.

Amit Kumar [42] Android yapısını bir geliştiricinin bakış açısından incelemiştir. Android temel e-mail, sms, takvim, harita programları ve diğer özellikler içerir. Bütün programlar Java programlama dili kullanılarak yazılmıştır. C/C++ kütüphanelerini içerir ve Android sistemde değişik bileşenlerle kullanılır. Android uygulamalar Java uygulamaları gibi Eclipse gibi bir platformda geliştirilebilirler. Android programları Android Emülatörleri ile test edilebilir.

Tor Morten ve arkadaşları [43] çalışmalarında Android, Windows Phone, iOS ve Firefox OS işletim sistemlerini karşılaştırmışlardır. iOS ve Android işletim sistemlerinin gelişmişlik düzeyleri Windows Phone ve Firefox OS işletim sistemlerinden ileridedir. Geriye dönük uyumluluk ve güncellemeleri her sürümün alamaması iOS ve Android işletim sistemlerinin dezavantajıdır. iOS, Objective-C, Android, Java, Firefox OS, Web, Windows Phone .Net C# programlama platformları ile uygulama geliştirmeye açıktır. iOS, Android, Windows Phone ve Firefox OS işletim sistemleri uygulama biçimleri sırasıyla .IPA, .APK, Web Application, .XAP'tır. Bu işletim sistemlerinin Firefox OS hariç üçünde

yerel SQL [Structured Query Language (Yapılandırılmış Sorgu Dili)] veri tabanı desteklenmektedir.

Chambers ve arkadaşları [44] mobil cihazlarda, kişisel mesajlar ve sosyal medya içeriğinin çabuk ve özel mobil duygu analizini uygulamışlardır. Android için mobil duygu analizi SentiCorr duygu analizi motoru prensipleri ile yerel olarak mobil duygu analizi gerçekleştirilmiştir. SentiCorr ile cümle seviyesinde POS (Position of Speech) kullanılarak yapılan mobil duygu analizinde başarı oranı %90'dır.



3. DUYGU SÖZLÜĞÜ

3.1. Türkçe Duygu Sözlüğü Oluşturulması

Duygu analizinin sözlüksel benzeşim yöntemiyle yapılabilmesi için duygu içeren kelimelerin duygu puanlarıyla veri tabanı gibi bir yapıda tutulması gerekmektedir. Bu oluşuma duygu sözlüğü denmektedir. Duygu analizi, duygu sözlüğü kelimelerinin metinde bulunma durumuyla ilgili olarak gerçekleştirilir.

Türkçe duygu sözlüğü oluşturulurken WordNet sözcük veri tabanından yararlanılmıştır. WordNet kelimelerin ilişkisel olarak birbirine bağlı olduğu büyük bir sözcük veri tabanıdır. WordNet'in sözcükleri veri tabanından çekilmiş, farklı sözlük uygulamaları yardımıyla Türkçeye çevrilmiştir. Kelime duygu puanları, WordNet'in ilişkisel yapısı ve bu yapıyla ilgili benzerlik ölçümlerinden yararlanılarak elde edilmiştir. Kelimelerin pozitif ve negatif puanları bu kelimelere olan benzerlik değerleriyle bulunmuştur. Türkçe duygu sözlüğünün duygu analizinde başarısının ölçülmesi için veri setine ihtiyaç duyulmuştur. Film ve otel tanıtım sitelerindeki [45,46] yorumlar Jsoup Java kütüphanesi [47] yardımıyla siteden metin dosyasına aktarılmıştır. Genel ön işleme işlemleriyle beraber duygu analizi ile duygu sözlüğünün başarısı ölçülmüştür.

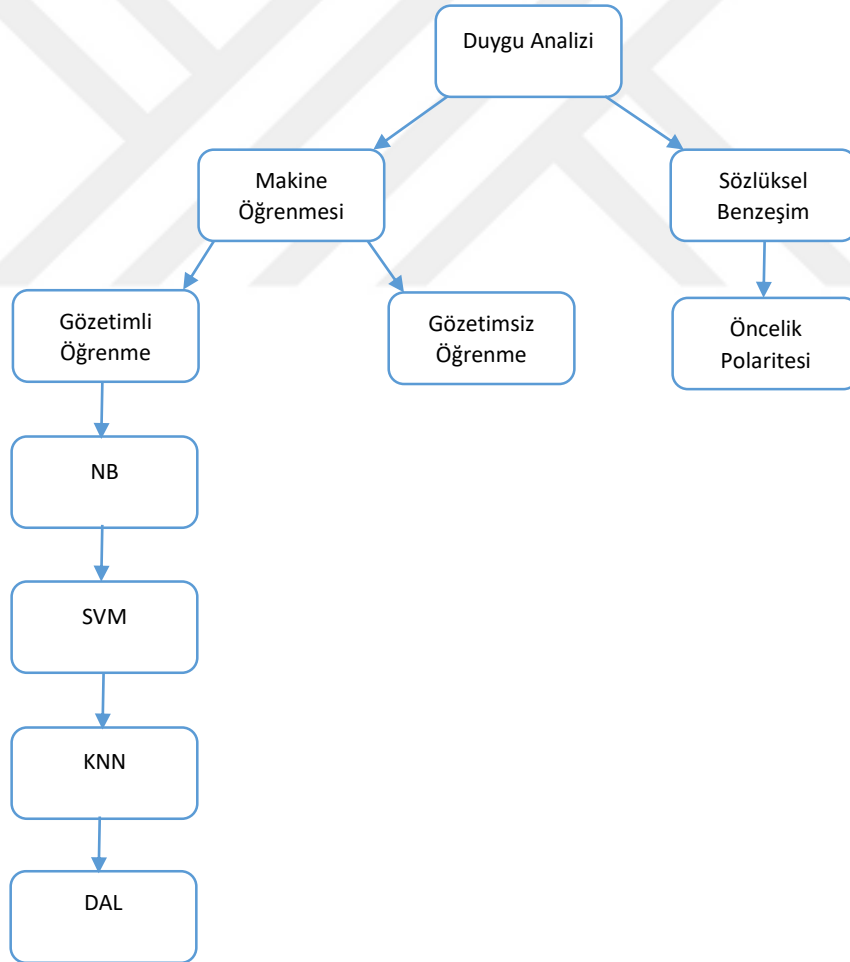
3.2. Sözlüksel Benzeşim Yöntemi

Duygu analizi iki ana yöntemle gerçekleştirilir. Şekil 3.1'de duygu analizi yöntemleri görülmektedir. Makine öğrenmesi yöntemlerinde gözetimsiz öğrenmede eğitim kümesi test kümesi ile eşleştirilirken, gözetimli öğrenmede istatistiksel yöntemler için eğitim kümesinden üretilen bir model söz konusudur. Eğitim kümesinde genellikle BOW, N-gram, TF, TF-IDF yöntemleri kullanılarak öznitelik seçimi yapılır ve öznitelik vektör yapısı oluşturulur. Veriler bir dosyaya kaydedilerek test kümesini sınıflayacak bir model oluşturmak üzere makine öğrenmesi sistemine yüklenir.

En çok kullanılan makine öğrenmesi yöntemleri NB, SVM, DAL [Dictionary of Affect in Language (Dil Etkileşim Sözlüğü)], KNN'dir. NB yönteminde, olasılık hesaplama yöntemleriyle sınıflandırma yapılır. KNN yönteminde bulunmak istenen yeni öge sınıfı için, alınan verilerin çıkarılan özelliklerinden k gibi bir sayı kadar eleman yakınlığı aranır. SVM yönteminde, iki boyutlu, üç boyutlu veya çok boyutlu uzayda veriyi

ayırma mekanizmaları ile verileri sınıflama yöntemidir. Bu yöntemde belirli kısıtlar göz önüne alınarak oluşturulan sınıflama denklemi aracılığıyla yeni öge bir bölgeye yerleştirilip sınıflandırılır. DAL yönteminde duygu içeren kelimelere 1 ile 5 puan aralığındaki gibi bir puanlama yapılır. Test verisinde cümledeki kelimelerin tamamı sistem içinde bulunuyorsa kelimelerin puanlarının aritmetik ortalamasının alınması ile cümlelerin duygu tespiti ve duygu analizi yapılır.

Sözlüksel benzeşim yönteminde, bir dile ait kelimeler, veri tabanında duygu değeriyle beraber kaydedilir. Duygu analizinde, bu kelimelerin tutulduğu yapıdaki değerler ile cümle veya kelime bazında metindeki duygu yönü tespit edilir. Duygu analizi için yapılan deneyler başarımlar ölçütleri yardımıyla karşılaştırılmaktadır. Doğruluk, F-Ölçütü gibi başarımlar ölçütleriyle sağlanan başarı gözlenir.



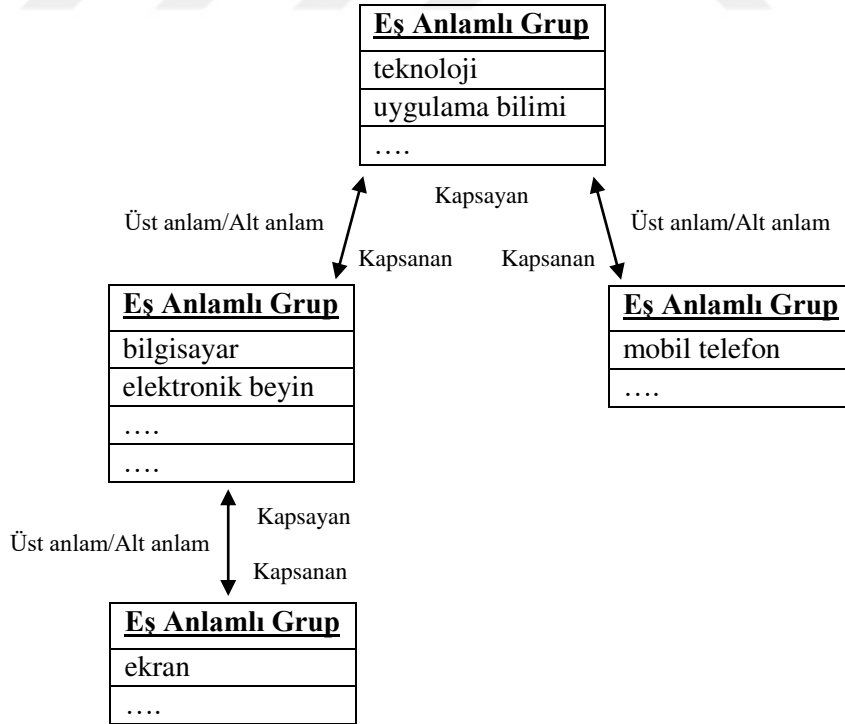
Şekil 3.1. Duygu analizi yöntemleri

3.3. WordNet Sözcük Veri Tabanı

WordNet, Princeton Üniversitesinde psikolog ve dilbilimciler tarafından geliştirilen bir sözcük veri tabanıdır. Bu veri tabanının oluşturulmasındaki temel düşünce, sözcükleri alfabetik olmaktan ziyade kavramsal olarak ilişkilendirmektir.

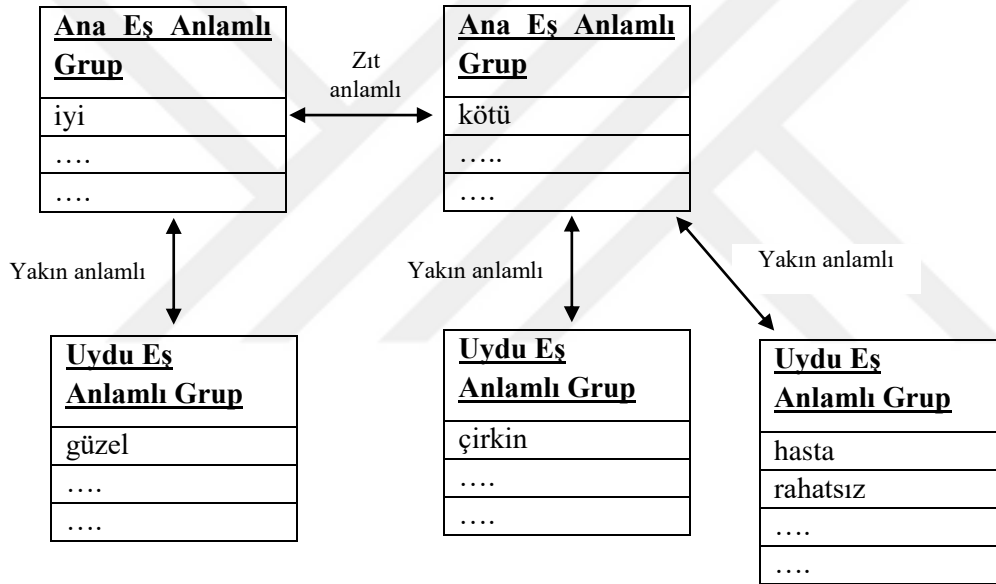
WordNet, veri tabanı, sözcükler, sözcükleri işlemek için kodlar, arama rutinleri ve bilgilerin görüntülenmesi için bir arayüzden oluşur. Sözcüklerin isim, sıfat, zarf, fiil türüne göre etiketlendiği ve normal sözlük işlevlerine göre sıradan sayılmayacak üst anlamlılık (hypernym), alt anlamlılık (hyponym), eş düzeylik (coordinate), parça bütün (meronym holonym) gibi birçok anlamsal ilişkileri ihtiva eden büyük bir kelime veri tabanı içerir. 117.000 kelimenin bütünü anlamsal olarak diğer kelimelerle ilişkilendirilmiştir.

WordNet, synsets adı verilen eş anlamlı gruplar halinde sözcüklerin aralarındaki sözcüksel anlam ve ilişkileri açıklar. İsimler ve fiiller, eş anlamlı grupların arasındaki üst anlam ve alt anlam ilişkilerine dayanılarak hiyerarşik olarak düzenlenmiştir. Şekil 3.2’de görüldüğü gibi kapsayan ile kapsanan isimler arasında alt anlam ilişkisinde kapsamın daraldığı, üst anlam ilişkisinde ise kapsamın genişlediği bilgisi edinilebilir [3].



Şekil 3.2. WordNet veri tabanının fiil ve isim sözcük türleri arasındaki hiyerarşik yapısı

Sıfatlar, ana eş anlamlı grup (head synsetler) ve uydu eş anlamlı grup (satellite synsetler) içeren kümeler şeklinde düzenlenmiştir. Her küme, zıt grup çifti veya üçlüsü tarafından meydana gelmektedir. Zıt grup çifti veya üçlüsü ana eş anlamlı gruplarla belirtilir. Çoğu ana eş anlamlı grup, bir veya daha fazla uydu eş anlamlı gruba sahiptir. Bu uydu eş anlamlı gruplar, ana eş anlamlı grup tarafından gösterilen kavrama benzer anlamda kavramları gösterir. Şekil 3.3'te WordNet sıfat kümeleri yapısı görülmektedir. İki sıfat eş anlamlı grubu ana eş anlamlı grup olarak zıt anlam (antonym) ilişkisindeyken, uydu eş anlamlı grupları ise bağlı olduğu ana eş anlamlı grupla benzer, yakın anlamlı kelimeler içermekte ve bağlı olmadığı ana eş anlamlı grupla dolaylı bir zıtlık (indirect antonym) ilişkisi içerisindeydir [3].

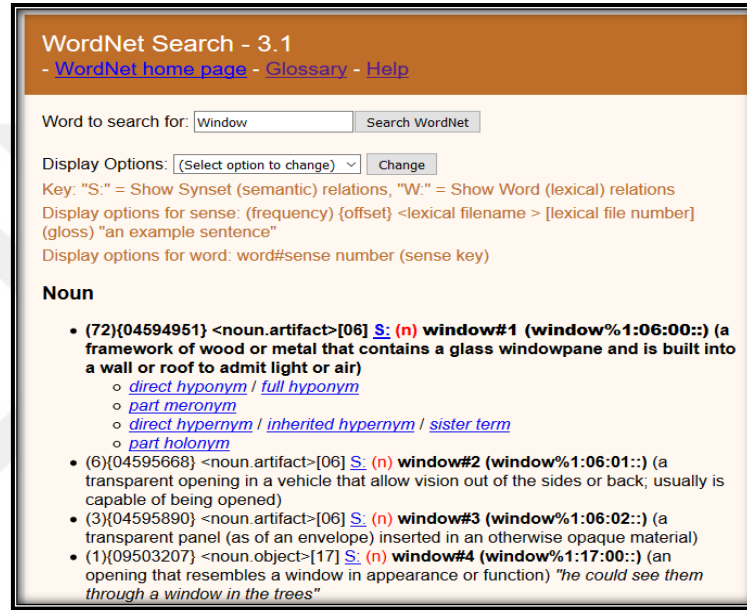


Şekil 3.3. WordNet veri tabının sıfat sözcük türleri arasındaki küme yapısı

Pertainym zıt anlamları olmayan ilişkisel bir sıfattır. Pertainyms oluşumlarında eş anlamlı gruplarda bir kelime veya kelime dizilerinin ilgili isimle olan ilişkisinden dolayı sözlüksel bir işaretçi ile gösterimi yapılır. Ortaçlar ve eylem soylu sözcükler de türetildiği fiillere sözcüksel işaret ile ulaşır. Zarflar ise genellikle sıfatlardan derlenmiştir. Bazen zıt anlam içeren kelimeler ihtiva eder. Bunlardan dolayı zarf eş anlamlı grup genellikle türettiği sıfata doğru sözlüksel bir işaretçi içerir. WordNet yapısı için tüm tanımlamalar Ek Tablo 2’de verilmiştir [3].

3.3.1. Online WordNet

WordNet web sitesinde, isteğe bağlı, örnek cümleleri gizleme, açıklamaları gizleme, veri tabanındaki konumları gösterme gibi görüntüleme tercihleri ile kelime araması online olarak yapılabilmektedir. Şekil 3.4'te WordNet online arama arayüzü görülmektedir. Aranacak kelime ve opsiyonel görüntüleme tercih seçenekleri belirlenip kelimeyle ilgili ilişki ve örnek cümleler görüntülenebilmektedir.



Şekil 3.4. WordNet online kullanım arayüzü [3].

WordNet online kullanım arayüzünde çeşitli kısaltmalar kullanılmaktadır. Bunlardan S, bir eş anlamlı grup için ihtiva ettiği anlam ilişkilerini görüntülerken W kelimenin sözlüksel ilişkilerini gösterir. Aramada belirtilen yapı şu şekildedir: (Tekrar sayısı) {Kelimelerin tüm kelimelerin yazımlarındaki sıra numarası veya id'si} <Sözlükteki noun, verb, adverb, adjective dosya> [Sözlükteki dosya numarası] (Sözcük türü) Kelime#Anlam numarası (Index.sense dosyasındaki anlam numaraları) (Kelime açıklaması) “Varsa örnek cümle”

3.3.2. WordNet Benzerlik Ölçümleri

WordNet benzerlik ölçümleri, iki sözcük arasındaki benzerliği WordNet'in ilişkisel ağından yararlanarak sayısal değer olarak veren yaklaşımlardır. Hirst-St-Onge, Leacock-Chodorow, Resnik, Jiang-Conrath ve Lin olmak üzere beş benzerlik ölçüm yaklaşımı tanımlanmıştır.

Hirst-St-Onge Yaklaşımı

Hirst ve St-Onge'nin WordNet için öne sürdüğü anlamsal benzerlik ölçüm yaklaşımıdır. WordNet eş anlamlı grupları, çok uzun olmayan bir yol ile bağlanmışsa ve bu yol içinde çok fazla yön değiştirme olmuyorsa, bu iki eş anlamlı gruptaki kelimeler anlamsal olarak birbirine yakındır (Eşitlik 3.1) [48].

$$rel_{HS}(c_1, c_2) = C - path\ length - k \times d \quad (3.1)$$

Eşitlikte “ c_1 ” ve “ c_2 ” benzerliği aranan kelime eş anlamlı gruplarını, “ d ” yön değişim sayısını, “ C ” ve “ k ” ise sabit sayıları belirtmek üzere ölçüm değeri hesaplanmaktadır. Sabit sayılar önceden tanımlanmamıştır. Hesaplama varsayımlarına bağlı olarak değişen değerler olarak belirlenir. Hirst ve St-Onge yaklaşımında yol belirlenemiyorsa ölçüm değeri 0'dır. Bu eş anlamlı grupların benzerlik ilişkisi yoktur.

Leacock-Chodorow Yaklaşımı

Bu ölçümde, iki eş anlamlı grup arasında en kısa yol ölçümünün yapılmasına ($len(c_1, c_2)$) güvenilmektedir (Eşitlik 3.2). İki kavram arasındaki en kısa yol en az ara kavram içeren yoldur. Bununla birlikte tüm yol uzunluğundaki D derinlik değişkeni ve IS-A bağlantısı ile hesaplama yapılır. Derinlik yaprak düğümden kök düğüme olan hiyerarşide en uzun yol olarak tanımlanır. IS-A hiyerarşik bağlantısı bir kavramın diğer kavrama ne kadar yakın olduğunu belirtir. Örneğin otomobil ve tekne ortak üst kavram olarak aracı paylaştığı için otomobil bir ağaçtan çok bir teknedir veya tekneye benzerdir [48,54].

$$sim_{LC}(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2D} \quad (3.2)$$

Eşitlikte “ c_1 ” ve “ c_2 ” benzerliği aranan eş anlamlı grupları, 2D derinlik değişkeninin değerinin iki katını belirtmek üzere, değer oranlarının negatif logaritması alınarak benzerlik değeri bulunur.

Resnik Yaklaşımı

Resnik ilişki tabanlı ölçümünde IC [Information Content (Bilgi İçeriği)] formülasyonunu temel almıştır (Eşitlik 3.3.a). IC hiyerarşideki kavramlara atanan değerdir. Bir kavramda IC büyük bir metinde bulunma frekansı ile hesaplanır. Maksimum olasılık hesabı ile olasılığı bulunur [48].

Resnik ile WordNet’teki iki kavram arasında benzerlik en düşük ortak üst kavram bilgisi ($lso(c_1, c_2)$) ve içerik bilgisi beraber kullanılarak bulunur. Eşitlik 3.3.b’de Resnik benzerlik ölçüm formülü görülmektedir.

$$IC(Concept) = -\log(p(Concept)) \quad (3.3.a)$$

$$sim_R(c_1, c_2) = -\log p(lso(c_1, c_2)) \quad (3.3.b)$$

Eşitlikte “ c_1 ” ve “ c_2 ” benzerliği aranan kelimelerin eş anlamlı grupları belirtmektedir. Benzerlik değeri en düşük ortak üst kavram bilgisinin IC bilgisiyle hesaplanır.

Jiang-Conrath Yaklaşımı

IC kavramı kullanılır. Fakat bu verilen ana eş anlamlı gruptan çocuk eş anlamlı gruba rastlamanın koşullu olasılığı biçimindedir. Bu yaklaşım benzerlik değerini tersten veren uzaklık ölçüm yaklaşımıdır Eşitlik 3.4’te uzaklık ölçüm formülü görülmektedir [48].

$$dist_{JC}(c_1, c_2) = 2 \log(p(lso(c_1, c_2))) - (\log(p(c_1)) + \log(p(c_2))) \quad (3.4)$$

Eşitlikte “ c_1 ” ve “ c_2 ” uzaklığı aranan kelimelerin eş anlamlı gruplarını belirtmek üzere, $p(lso(c_1, c_2))$ en yakın ortak üst kavram, $p(c_1)$ ve $p(c_2)$ en özel alt çocuk maksimum olasılık hesap değerini göstermektedir.

Lin Yaklaşımı

Lin eş anlamlı gruplar arasındaki anlamsal benzerliğin ölçümünü kendi benzerlik teoreminden [49] türetmiştir. İki kavram benzerliği, iki kavramın ortak noktasını belirtmek için gerekli olan bilginin, onları türetmek için gerekli olan bilgi miktarına oranı ile ölçülür. İki kavramın ortak noktası en düşük ortak eş anlamlı grup ve aralarındaki IC ile bulunur. Eşitlik 3.5'te benzerlik formülü görülmektedir.

$$sim_L(c_1, c_2) = \frac{2 \times \log p(lso(c_1, c_2))}{\log p(c_1) + \log p(c_2)} \quad (3.5)$$

Eşitlikte “c₁” ve “c₂” benzerliği aranan kelimelerin eş anlamlı gruplarını belirtmek üzere, p(lso(c₁, c₂)) en yakın üst kavram ve p(c₁) ve p(c₂) en özel alt çocuk maksimum olasılık hesap değerleri kullanılarak kelimeler arasındaki benzerlik değeri bulunur.

3.4. Sözlük Veri Tabanı

Sözcük seçimi ve sözcüklerin duygu yön ve değerlerinin belirlenmesi için RiWordNet [50], Resnik Java kütüphanelerinden yararlanılmıştır. WordNet veri tabanında bulunan isim, sıfat, zarf, fiil sözcüklerin RiWordNet kütüphanesinin getRandomWords fonksiyonu ile elde edilmesiyle sözcük seçimi gerçekleştirilmiştir. Tablo 3.1'de sözcük tür dağılımı görülmektedir. WordNet sözcük veri tabanından 44502 isim, 13969 sıfat, 2212 zarf, 7900 fiil türünde olmak üzere 68583 İngilizce sözcük elde edilmiştir.

Tablo 3.1. Veri tabanı sözcüklerinin tür dağılımı

Sözcük Türü	Sözcük Sayısı
İsim	44502
Sıfat	13969
Zarf	2212
Fiil	7900
Toplam	68583

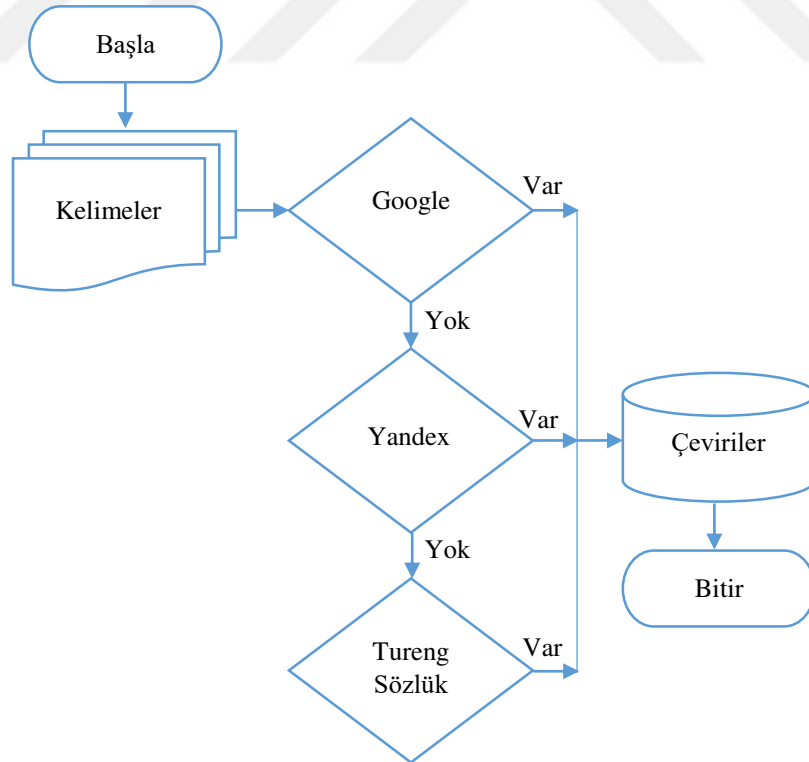
WordNet veri tabanından alınan 68583 sözcüğün Türkçe diline çevrilmesi için Google [51], Yandex [52] ve Tureng sözlük [53] uygulamalarından yararlanılmıştır.

Uygulamaların ayrı olarak kullanıldığı çeviride, çeviri oranı düşük olmuştur. Tablo 3.2’de uygulamaların çeviri oranları görülmektedir. Yandex ve Tureng sözlük uygulamaları çeviri oranı %50’nin altında kalmıştır. Google uygulama %68 çeviri oranı, uygulamalar bağlanmış şekilde kullanılarak %78 oranına arttırılmıştır.

Tablo 3.2. Sözlük çeviri yüzdeleri

Sözlükler	Çevrilen Terim	Çeviri Yüzdesi
Google	46636	%68
Yandex	33605	%49
Tureng	29490	%43
Bağlı	53494	%78

Çeviri uygulamaları sırasıyla Google, Yandex, Tureng olarak bağlanmış, ilk sözlük uygulamasında bulunmayan sözcük bir diğer sözlük uygulamasında aranarak çeviri işlemi gerçekleştirilmiştir. Çeviriler veri tabanına kaydedilmiştir. Sözlük çeviri algoritması Şekil 3.5’te görülmektedir.



Şekil 3.5. Sözcük çeviri algoritması

İsim ve fiil sözcüklerinin pozitif (positive) ve negatif (negative) kelimelere olan benzerlik değerleri Resnik Java kütüphanesi ile elde edilip, pozitif değerden negatif değer çıkarılarak kelimenin duygu değeri bulunmuştur (Eşitlik 3.6). Bazı kelimelerin birden çok benzerlik değeri belirtilmiştir. Bu değerler, SentiWordNet'in web sitesinde bulunan örnek kodda belirtilen birleştirme formülüyle kelime önceliğine göre ağırlıklandırılıp birleştirilmiştir. Eşitlik 3.7.'de görüldüğü gibi tekrar eden kelimenin, ilk kelimedenden son kelimeye doğru sırasıyla ağırlıkları azaltılarak birleştirilmektedir. Sıfat ve zarf türündeki sözcükler için pozitif ve negatif benzerlik değerleri döndürülememiştir. Bunun nedeni sıfat ve zarf eş anlamlı gruplarının hiyerarşik olarak düzenlenmiş olmamalarıdır [54].

$$Skor_i = Pozitif_i - Negatif_i \quad (3.6)$$

$$Skor_B = \frac{\sum_{i=1}^n \frac{1}{i+1} s_i}{\sum_{i=1}^n \frac{1}{i}} \quad (3.7)$$

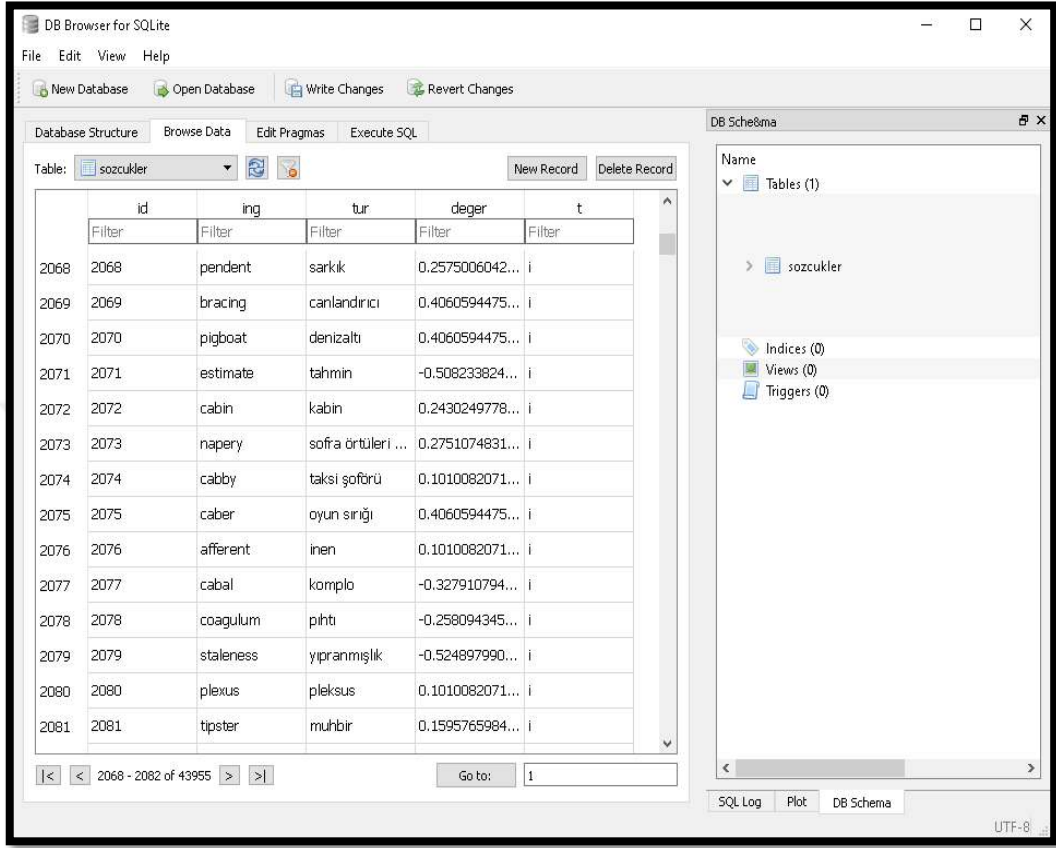
Tablo 3.3'te kelimelerin duygu değeri yön dağılımı görülmektedir. Fiil türü sözcüklerin duygu değerinin tamamı negatif yönlü değerlerden oluşmaktadır. WordNet'te positive sözcüğünün isim, sıfat, negative sözcüğünün isim, sıfat ve fiil türleri bulunduğundan fiil sözcüklerinin positive kelimesine olan anlamsal benzerlik değeri bulunamamıştır. Fiil türündeki sözcükler sözlük veri tabanında kullanılamamıştır. İsim türündeki sözcüklerde pozitif yönlü duygu değerleri negatif yönlü değerlerden daha fazla bulunmuştur. İsim türündeki kelimelerin 25473 pozitif, 17918 negatif yönlü duygu değeri elde edilmiştir.

Tablo 3.3. Resnik Java kütüphanesi ile elde edilen sözcük değerlerinin yön dağılımı

Sözcük Türü	Pozitif Yön	Negatif Yön	Nötr
İsim	25473	17918	1111
Sıfat	-	-	13969
Zarf	-	-	2212
Fiil	-	564	7336
Toplam	25473	18482	24628

WordNet'ten alınan İngilizce sözcükler ve Türkçeye çevrimlerinin duygu değerleriyle beraber tutulduğu duygu sözlüğü veri tabanı Şekil 3.6'da görülmektedir

(Duygu sözlüğünde ilk ve son elli terimden oluşan örnek kelimeler listesi Ek Tablo 1’de belirtilmiştir.).



	id	ing	tur	deger	t
2068	2068	pendent	sarkık	0.2575006042...	i
2069	2069	bracing	canlandırıcı	0.4060594475...	i
2070	2070	pigboat	denizaltı	0.4060594475...	i
2071	2071	estimate	tahmin	-0.508233824...	i
2072	2072	cabin	kabin	0.2430249778...	i
2073	2073	napery	sofra örtüleri ...	0.2751074831...	i
2074	2074	cabby	taksi şoförü	0.1010082071...	i
2075	2075	caber	oyun sınığı	0.4060594475...	i
2076	2076	afferent	inen	0.1010082071...	i
2077	2077	cabal	komplot	-0.327910794...	i
2078	2078	coagulum	pıhtı	-0.258094345...	i
2079	2079	staleness	yıpranmışlık	-0.524897990...	i
2080	2080	plexus	pleksus	0.1010082071...	i
2081	2081	tipster	muhbir	0.1595765984...	i

Şekil 3.6. Duygu sözlüğü veri tabanı

3.5. Veri Setleri

Sözlüğün test edilmesi için pozitif ve negatif olarak etiketlenmiş metinler gereklidir. Bu şekilde sunulan bir veriye ulaşılamamıştır. Veri seti oluşturmada film web sitesinde [45] bulunan izleyici yorumları ve otel rezervasyon sitesindeki [46] yorumlar kullanılmıştır.

Film Yorumları Veri Seti

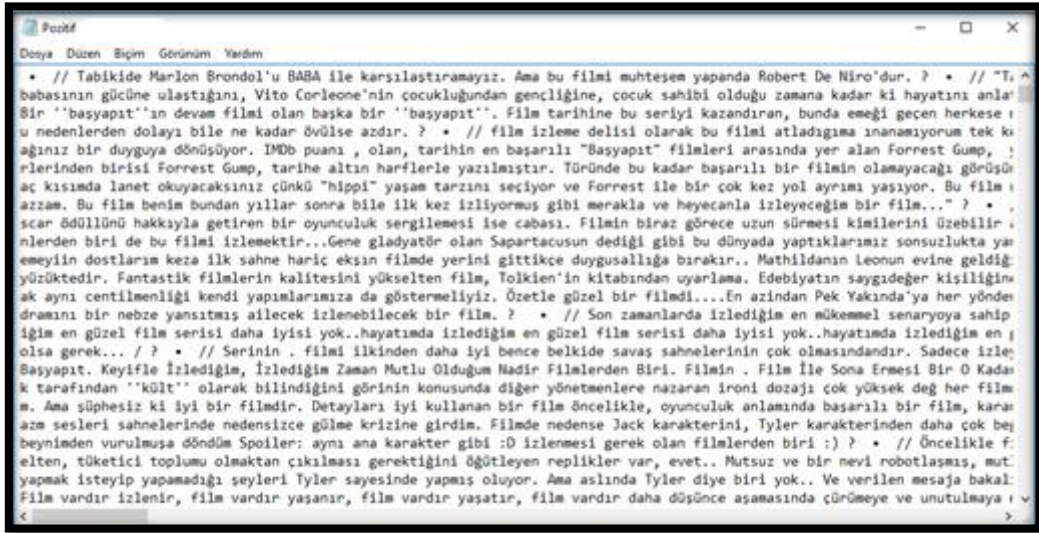
Film web sitesindeki yorumlar 0,5 ile 5 değerleri arasında puanlanmakta, tekrar eden bir yapı olarak tutulmaktadır. Jsoup Java kütüphanesi ile tüm yorumlar ayrılıp metin

dosyasında tutulmuştur. 4,0 ile 5,0 puanları arasındaki yorumların pozitifliği ve 0,5 ile 2,0 puanları arasındaki yorumların ise negatifliği kontrol edilmeyip etiketlemelerin doğru olduğu kabul edilmiştir. Yorumların 7966'sı pozitif, 3911'i negatif etiketlidir. Tablo 3.4'te film yorumlarının puanlamalara göre dağılımı görülmektedir.

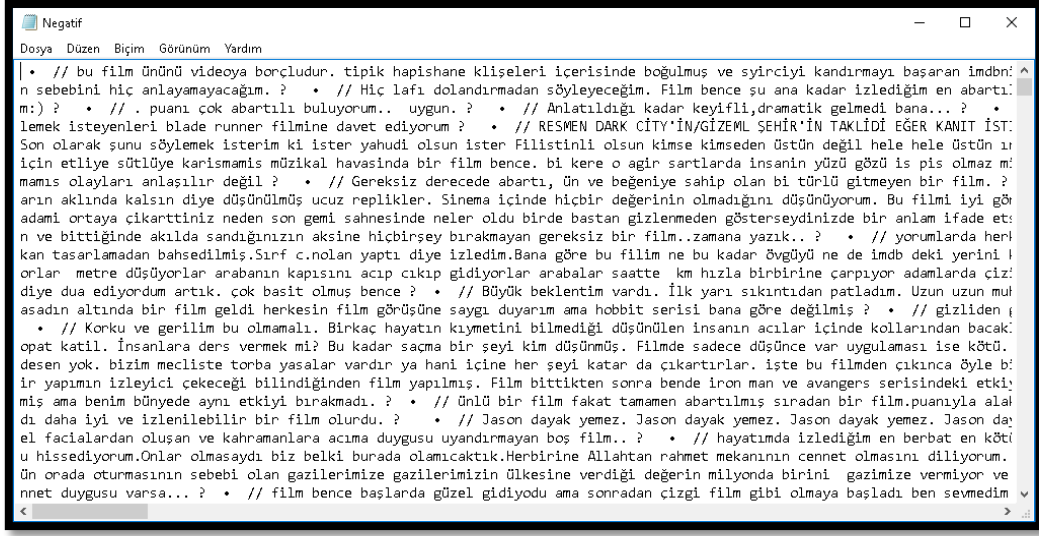
Tablo 3.4. Film yorumları dağılımı

Film Yorumları	Puanlama	Yorum Sayısı
	5,0	7966
	4,5	
	4,0	
	3,5	12425
	3,0	
	2,5	
	2,0	3911
	1,5	
	1,0	
	0,5	

Sonraki aşamada 3911 pozitif ve negatif yorum Şekil 3.7 ve 3.8'de görüldüğü gibi metin dosyalarında tutulmuştur.



Şekil 3.7. Film yorumları pozitif veri seti



Şekil 3.8. Film yorumları negatif veri seti

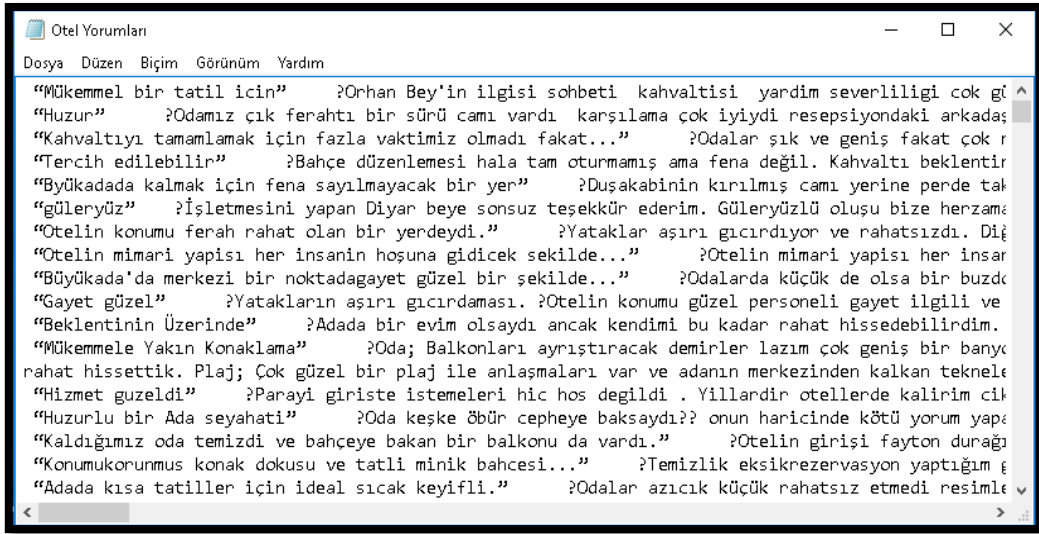
Otel Yorumları Veri Seti

Otel rezervasyon web sitesinde bulunan otel yorumları 1 ile 10 değerleri arasında ile kullanıcılar tarafından puanlanmıştır. Yorumların puanlamaları kontrol edilmeyip doğru kabul edilmiştir. Otel yorumlarının, 1 ile 5 puan aralığındaki yorumları negatif, 5 ile 10 puan aralığındaki yorumları pozitif olarak kabul edilmiştir. Rastgele seçilen 500 yorumun 362'si pozitif 138'i negatif etiketlidir. Tablo 3.5'te otel yorumlarının kullanıcı puanlamalarına göre dağılımı görülmektedir.

Tablo 3.5. Otel yorumları dağılımı

Otel Yorumları	Puanlama	Yorum Sayısı
	10-5,1	362
	5-1	138

Pozitif ve negatif tüm otel yorumları Şekil 3.9'da görüldüğü gibi metin dosyasında tutulmuştur.



Şekil 3.9. Otel yorumları veri seti

3.6. Türkçe Duygu Sözlüğü ve Veri Setleri ile Duygu Analizi

Sözlüksel benzeşim yöntemiyle duygu analizi, ön işleme aşamasından geçirilerek duygu analizine hazır hale getirilen metnin cümle polariteleri belirlenerek, sayıca fazla olan cümle polaritesi yönü metnin duygu yönünü belirtmek üzere yapılmaktadır.

Türkçe duygu sözlüğü ile hangi doğruluk oranında duygu analizi yapılabildiğini, duygu yönü oranlarını ne kadar doğru tahmin edildiğini belirlemek amacıyla duygu yönü belirli film veri setleri ve duygu yüzde oranları belirli otel veri seti ile duygu analizi yapılmıştır.

Film ve otel yorumları veri setleri için metni noktalama işaretlerinden arındırma, metnin içeriğini küçük harfe dönüştürme gibi veri ön işleme işlemleri duygu analiziyle birlikte analizden önce yapılmıştır. Yanlış yazılan kelimeler Zemberek doğal dil işleme kütüphanesi ile düzeltilmeye çalışılmış, farklı öneriler sunabildiğinden düzeltme uygulanmamıştır.

Veri setlerindeki ifadelerin cümle polariteleri, kelime, duygu değeri, duygu yönü, frekansı ve varlığı gibi kriterler baz alınarak farklı şekillerde bulunmuş sonuçlar karşılaştırılmıştır. Duygu değeri baz alınarak yapılan duygu analizinde cümledeki kelimelerin duygu değeri toplanıp toplam değerın duygu yönüne göre cümle pozitif veya negatif olarak etiketlenmiştir (Eşitlik 3.8.a). Duygu yönü baz alınarak yapılan değerlendirmede ise cümlede sayısal olarak fazla olan kelime yönü cümledeki duygu

yönünü göstermektedir (Eşitlik 3.8.b). Kelime frekansı kriterinde sözcük tekrarları göz ardı edilmiş, kelime varlığı kriterinde ise tekrarlar elenmiştir.

Karşılaştırma yapmak üzere film veri seti, normal olarak ve etkisiz kelimeler (stop words) elenerek (Stop words tüm kelimeleri Ek 1’de belirtilmiştir.) iki ayrı formda değerlendirilmiştir.

Film veri setlerinde cümle etiketlerine göre doğru ve yanlış tespit değerleri başarı ölçüm yöntemleri ile hesaplanmıştır. Şekil 3.10’da film yorumları pozitif ve negatif veri setindeki kelime duygu yönü ve varlığı baz alınarak bulunan 14949 cümle polaritelerinin belirlendiği sonuç ekranı görülmektedir.

$$Cümle Polaritesi_{DD} = \begin{cases} Pozitif, \sum_{i=1}^n P_{K_i} > 0 \\ Negatif, \sum_{i=1}^n P_{K_i} < 0 \end{cases} \quad (3.8.a)$$

$$Cümle Polaritesi_{DY} = \begin{cases} Pozitif, \sum_{i=1}^n Y_{K_i} > 0 \\ Negatif, \sum_{i=1}^n Y_{K_i} < 0 \end{cases} \quad (3.8.b)$$

```

14943. Cümle---> 0.0,Objective
14944. Cümle---> 1.0,Pozitif
14945. Cümle---> 1.0,Pozitif
14946. Cümle---> 1.0,Pozitif
14947. Cümle---> -2.0,Negatif
14948. Cümle---> -1.0,Negatif
14949. Cümle---> -1.0,Negatif
14949 Cümle---> Pozitif Cümle:5926Negatif Cümle:54540bjektif Cümle:3569

14943. Cümle---> 2.0,Pozitif
14944. Cümle---> 1.0,Pozitif
14945. Cümle---> -1.0,Negatif
14946. Cümle---> 1.0,Pozitif
14947. Cümle---> -1.0,Negatif
14948. Cümle---> -1.0,Negatif
14949. Cümle---> 0.0,Objective
14949 Cümle---> Pozitif Cümle:5643Negatif Cümle:58060bjektif Cümle:3500

```

Şekil 3.10. Pozitif ve negatif veri setleri cümlelerinin polaritelerinin belirlenme ekran sonucu

Otel yorumları veri seti için pozitif ve negatif yorumlar birlikte değerlendirilmiştir. Veri setinin gerçek pozitiflik ve negatiflik yüzdeleri ile duygu analizinden sonraki oranlar karşılaştırılmıştır. Şekil 3.11’de otel yorumlarının kelime duygu yönü ve varlığı baz alınarak cümle polariteleri belirleme sonuç ekranı görülmektedir.

```

2175. Cümle---> -2.0,Negatif
2176. Cümle---> 0.0,Objective
2177. Cümle---> 1.0,Pozitif
2178. Cümle---> 1.0,Pozitif
2179. Cümle---> 5.0,Pozitif
2180. Cümle---> 4.0,Pozitif
2181. Cümle---> 1.0,Pozitif
2181 Cümle---> Pozitif Cümle:1257Negatif Cümle:4450objektif Cümle:479

```

Şekil 3.11. Otel yorumları veri seti cümlelerinin polaritelerinin belirlenme ekran sonucu

3.6.1. Başarı Ölçüm Yöntemleri

Duygu analizi için yapılan deneyler başarımlar ölçütleri yardımıyla karşılaştırılmaktadır. Başarım ölçütlerinin değeri arttıkça yöntem daha doğru ve kapsamlı sonuçlar verir.

Doğruluk–Hata Oranı (Accuracy-Error Rate)

Doğruluk ve hata oranı başarımlar ölçütlerinde kullanılan popüler ve basit bir yöntemdir. Duygu analizinde bütün örneklerin doğru sınıflandırılmasında ne kadar başarılı olduğunu belirler. Doğru tahmin edilmiş örnek sayısının ($TP+TN$), toplam örnek sayısına ($TP+TN+FP+FN$) oranıdır (Eşitlik 3.9.a). Hata oranı ise yanlış tahmin edilmiş örnek sayısının ($FP+FN$), toplam örnek sayısına oranıdır (Eşitlik 3.9.b). Doğruluk oranı ile hata oranı toplam değeri 1’dir.

$$Doğruluk = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.9.a)$$

$$Hata = \frac{FP+FN}{TP+TN+FP+FN} \quad (3.9.b)$$

TP doğru pozitif tahminlerini, TN doğru negatif tahminlerini, FP yanlış pozitif tahminlerini ve FN yanlış negatif tahminlerini belirtmektedir. Diğer başarı ölçüm yöntemlerinde de terimler aynı tahminleri belirtmek üzere kullanılmaktadır.

Duyarlık (Precision)

Duyarlık, polaritenin ne kadar doğru tahmin edildiğinin ölçümüdür. Duygu analizinde örneklerin ne kadarının düzgün olarak sınıflandırıldığını gösterir. Pozitif örnekler için, doğru pozitif tahmin sayısının (TP) tüm pozitif tahmin sayısına ($TP+FP$) oranı, negatif örnekler için, doğru negatif tahmin sayısının (TN) tüm negatif tahmin sayısına ($TN+FN$) oranıdır (Eşitlik 3.10).

$$Duyarlık = \frac{TP}{TP+FP} \quad (3.10)$$

Anma (Recall)

Anma, kategoriye ait elemanların ne kadarının tahmin edildiğinin ölçümüdür. Pozitif örnekler için, doğru tahmin edilmiş pozitif örnek sayısının (TP), doğru pozitif ve yanlış negatif tahmin örnek sayısına ($TP+FN$) oranı, negatif örnekler için, doğru tahmin edilmiş negatif örnek sayısının (TN) doğru negatif ve yanlış pozitif tahmin örnek sayısına ($TN+FP$) oranıdır (Eşitlik 3.11).

$$Anma = \frac{TP}{TP+FN} \quad (3.11)$$

F-Ölçütü (F-Measure)

Anma ve duyarlık ölçütleri tek başına anlamlı karşılaştırma sonuçları elde etmek için yeterli değildir. Her iki ölçüt beraber değerlendirilir. F-Ölçütü, anma (K) ve duyarlığın (D) harmonik ortalamasıdır (Eşitlik 3.12).

$$F = \frac{2DK}{D+K} \quad (3.12)$$

3.6.2. Duygu Analizi Testi Başarı Ölçümleri

Türkçe Duygu Sözlüğü ve Film Veri Setiyle Yapılan Duygu Analizi Başarı Ölçümleri

Film veri seti, NetBeans Java platformunda kodlar derlenerek cümle polariteleri, kelime duygu puanı, duygu yönü, tekrarlarının çıkarıldığı kelime varlığı, kelime frekansı baz alınarak bulunup farklı formlarda yapılan duygu analizi sonuçları F-Ölçütü ve doğruluk başarı ölçütü yöntemleriyle değerlendirilmiştir. F-Ölçütü değeri, başarı ölçüm yöntemleri başlığında belirtildiği gibi duyarlılık yani duygu analizi için pozitiflik, negatiflik oranının doğru ayırt edilme ve pozitif, negatif kategorideki elemanlarının tahmin miktarını gösteren anma ölçümlerinin beraber kullanımı ile bulunan değerdir. Duygu analizi değerlendirilmesinde F-Ölçütü değeri arttıkça duygu yönü tespitinin başarısının arttığı anlaşılmaktadır. Doğruluk değeri polarite ayırımına bakılmaksızın doğru tahmin miktarını gösterir. Doğruluk oranı arttıkça daha doğru analiz sonuçlarının alındığı bilgisine ulaşılabilir.

Cümle polaritesinin duygu puanı bazında bulunarak yapılan duygu analizinde pozitif ve negatif veri setleri için F-Ölçütü ve doğruluk başarı oranları Tablo 3.6’da görülmektedir.

Tablo 3.6. Film yorumlarının duygu puanı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları

	Pozitif F-Ölçütü	Negatif F-Ölçütü	Doğruluk
Normal	0,469	0,557	0,512
Etkisiz Kelimesiz	0,443	0,568	0,514

Veri setleri, kelime duygu yönü baz alınarak analiz edildiğinde, doğruluk başarı oranının %4 oranında arttığı gözlenmiştir. Tablo 3.7’de değerlendirme değerleri görülmektedir.

Tablo 3.7. Film yorumlarının duygu yönü baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları

	Pozitif F-Ölçütü	Negatif F-Ölçütü	Doğruluk
Normal	0,560	0,429	0,558
Etkisiz Kelimesiz	0,541	0,477	0,511

Veri setleri, tekrarlar çıkarılıp kelime frekansı yerine varlığı ve puanı baz alınarak duygu analizine tabi tutulmuştur. Tablo 3.8’de F-Ölçütü ve doğruluk değerleri görülmektedir. Bu şekilde yapılan analizde, anlamlı doğruluk başarı oranı artışı gözlenmemiştir.

Tablo 3.8. Film yorumlarının duygu puanı ve kelime varlığı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları

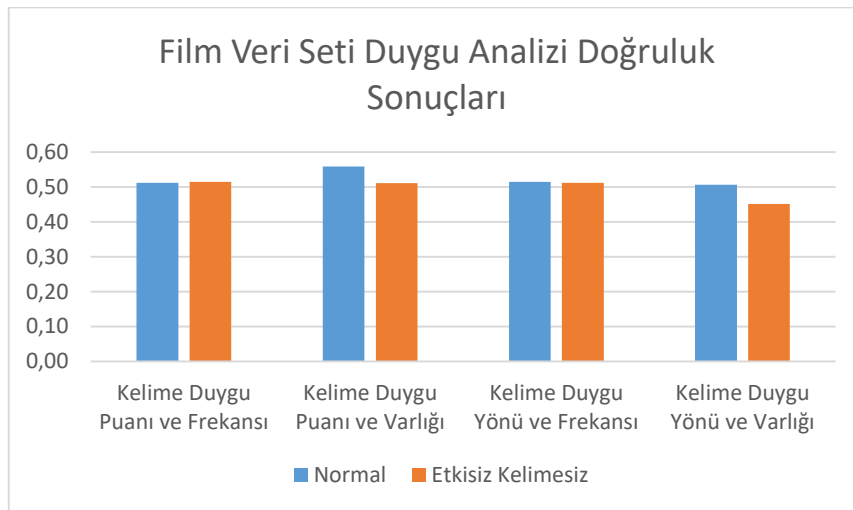
	Pozitif F-Ölçütü	Negatif F-Ölçütü	Doğruluk
Normal	0,440	0,571	0,514
Etkisiz Kelimesiz	0,408	0,586	0,512

Duygu yönü ve kelime varlığı baz alınarak yapılan duygu analizinde doğruluk değerlerinde artış olmazken, pozitif ve negatif F-Ölçütü 0,5 değerinin üzerindedir. Tablo 3.9’da F-Ölçütü ve doğruluk değerleri görülmektedir.

Tablo 3.9. Film yorumlarının duygu yönü ve kelime varlığı baz alınarak yapılan duygu analizinin F-Ölçütü ve doğruluk yöntemleri ile değerlendirilme sonuçları

	Pozitif F-Ölçütü	Negatif F-Ölçütü	Doğruluk
Normal	0,516	0,511	0,506
Etkisiz Kelimesiz	0,488	0,445	0,451

Film veri setinde normal olarak ve etkisiz kelimeler elenerek yapılan duygu analizi doğruluk sonuçları karşılaştırması Şekil 3.12’de görülmektedir.

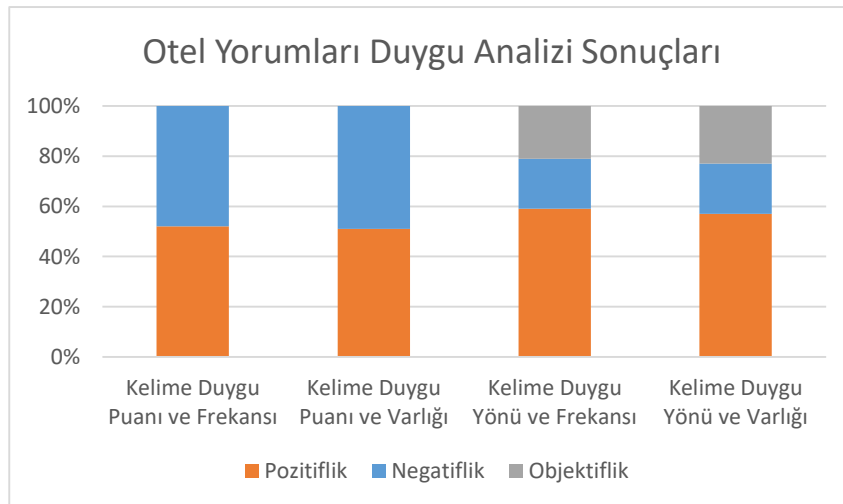


Şekil 3.12. Film yorumları doğruluk sonuçları karşılaştırması

Etkisiz kelimeler tarafsız, metne katkısı bulunmayan ve duygu analizinde etkisi olmayan kelimeler olarak tanımlanmaktadır. Karşılaştırmalarda görüldüğü gibi etkisiz kelimelerin elendiği formda yapılan duygu analizi sonuçları, etkisiz kelimelerin elenmediği formdaki analiz sonuçlarından farklılıklar göstermektedir. Türkçe duygu sözlüğü incelendiğinde bir (-0.115551066862812), bize (0.159576598417647), rağmen (-0.401111008467646), tamam (0.148917762430234) gibi etkisiz kelimelerin kutupsal duygu puanının bulunduğu tespit edilmiştir. Aynı kelimeler İngilizce olarak SentiWordNet online duygu sözlüğünde aranmıştır. a (-0.02410468319559229), despite (-0.5), ok (0,3135) gibi sözcüklerin bazılarının duygu değerlerinin bulunduğu tespit edilmiştir. Duygu analizinde veri ön işleme aşamasında uygulanan etkisiz kelimelerin elenme işleminin genel etkisiz kelime listesi ile yapılmasının duygu analizi sonuçlarını etkilediği görülmektedir.

Türkçe Duygu Sözlüğü ve Otel Veri Setiyle Yapılan Duygu Analizi Başarı Ölçümleri

Otel yorumları 2181 cümleden oluşmaktadır. Yorumların %62'si olumlu, kalan %38'i olumsuz kategoridedir. Duygu puanı ve duygu yönü için ayrı ayrı kelime frekansı, varlığıyla bulunan cümle polariteleriyle yapılan duygu analizinin sonuçları Şekil 3.13'te görülmektedir.



Şekil 3.13. Otel yorumları duygu analizi sonuçları karşılaştırılması

Duygu analizi sonucunda veri seti pozitiflik ve negatiflik oranına en yakın sonuçların, cümlelerdeki kelimelerin duygu yönü, frekansı ve varlığının seçildiği analizlerde alındığı görülmektedir.



4. MOBİL UYGULAMALAR

4.1. Mobil İşletim Sistemleri

Günümüzde akıllı telefon ve tablet bilgisayarlar mobil cihazlar olarak adlandırılmaktadır. Mobil cihazların yazılım işlevini bir mobil işletim sistemi yerine getirmekte ve bu işletim sistemiyle beraber bazı temel mobil uygulamalar bulunmaktadır. Ayrıca işletim sistemiyle ilgili kurumsal veya kişisel olarak geliştirilen mobil uygulamalar uygulama marketlerinden indirilebilmektedir.

Kullanımı yaygın olan üç tür mobil işletim sistemi mevcuttur. Bunlar mobil cihaz pazar payına göre sırasıyla Android, iOS, Windows Phone mobil işletim sistemleridir. Android ve iOS işletim sistemlerinin uygulama marketlerinde bir milyondan fazla uygulama bulunmaktadır. Bu iki işletim sistemi mobil cihaz pazar payının yaklaşık %95'ini oluşturmaktadır.

4.1.1. Android Mobil İşletim Sistemi

İlk olarak 2003'te geliştirilmeye başlanan Android sistemler gelişerek günümüzde açık kaynak kodlu, yüz binlerce uygulamaya sahip uygulama marketi ile mobil cihazların pazar payının yaklaşık %80'i oranını yakalamış en popüler mobil işletim sistemidir. Andy Rubin'in de aralarında bulunduğu bir grup California'da Android Inc. Adında bir şirketin açık kaynak kodlu işletim sistemlerini cep telefonlarına uyarlamayı başarmışlardır. Google Inc. iki yıllık Android Inc.'yi satın alıp kurucularını Google şirketi bünyesine katıp 2007 yılında Android işletim sistemi geliştirdiklerini açıklamıştır. Bu Android 1.0 mobil işletim sistemi Linux 2.6 çekirdeğini kullanan ilk android işletim sistemidir. 2008 yılında Google Inc. tarafından kaynak kodlar sunulmuştur. Önemli versiyonlar 2.1, 2.2, 2.3, 3.0, 3.1, 3.2, 4.0, 4.0.3, 4.1.2, 4.2.2, 4.3.1, 4.4.2, 4.4.W.2, 5.0.1, 5.1.1, 6.0, 7.0 şeklinde artarak günümüzdeki popülerliğine ulaşmıştır. Android mimarisi uygulamalar, uygulama ana çatısı, kütüphaneler gibi ana yapılarından oluşmaktadır. Şekil 4.1'de Android mimarisi ana yapıları görülmektedir [39].

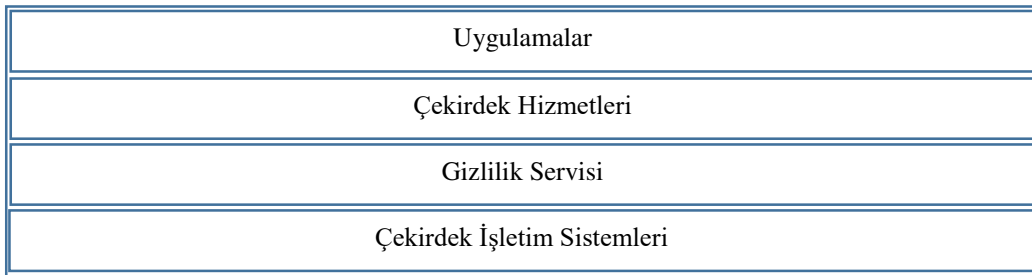


Şekil 4.1. Android mimarisi [43].

Android uygulamaları Netbeans ve Eclipse Java platformlarında Android SDK ile geliştirilebilir. Google Inc. son yıllarda Android uygulamalar için Java platform desteğini çekmiş bulunup Android Studio android mobil uygulama geliştirme platformunu kullanıma sunmuştur. Android Studio 2.2 Preview 7 son sürüm olarak yayınlanmıştır.

4.1.2. iOS Mobil İşletim Sistemi

iOS, iPhone gibi Apple mobil cihazların işletim sistemidir. iOS uygulamaları, Objective-C programlama dili ve Cocoa Touch kütüphanesini kullanarak yazılmıştır. Objective-C, C dilinin bir uzantısıdır. C# ve Java'nın benzer sentaksına alternatif sunmaktadır. iOS uygulamaları geliştirmek için bilgisayar Mac OS işletim sistemine ihtiyaç vardır. iOS uygulamaları XCode ile yazılır. Uygulama geliştiricisi, güçlü bir editör, iyi bir analiz ekipmanı, iOS simülatörü ve SDK [Software Development Kit (Yazılım Geliştirme Kiti)] içerir. iOS mimarisi; uygulamalar, çekirdek hizmetleri, gizlilik servisi gibi ana yapılarından oluşmaktadır. Şekil 4.2'de iOS mimarisi ana yapıları görülmektedir [43].



Şekil 4.2. iOS mimarisi [43].

4.1.3. Windows Phone İşletim Sistemi

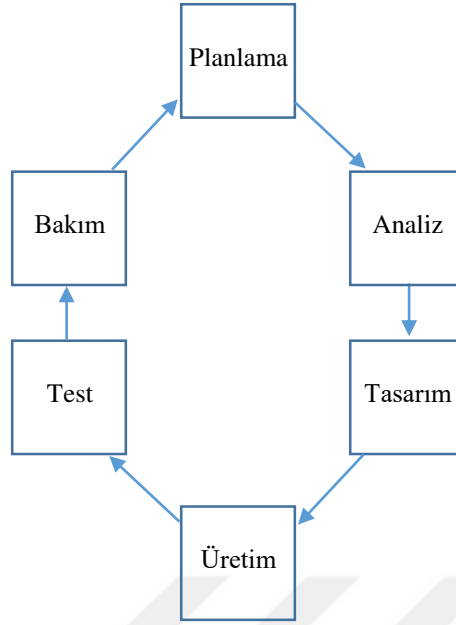
Windows Phone 7 .NET platformları ile yazılmıştır. Windows Phone 7 Silverlight ve XNA platformlarını destekler. Silverlight ile kullanıcı arayüzleri geliştirilir. XNA Microsoft'un oyun platformudur, 2D ile 3D grafik tasarımını destekler ve Visual Studio'da tasarlanmıştır. Visual Basic ve C# programlama dilleri ile Windows Phone için program yazılabilir. Windows phone mimarisi ana görev, çekirdek uygulama gibi ana yapılarından oluşmaktadır. Şekil 4.3'te Windows Phone mimarisi ana yapıları görülmektedir [43].



Şekil 4.3. Windows phone mimarisi [43].

4.2. Mobil Uygulama Geliştirme Yaşam Döngüsü

Bir yazılımın proje planlamasından teslimata kadar gerçekleştirilen kademelerden oluşan döngüye yazılım geliştirme yaşam döngüsü denir. Diğer yazılımlar gibi mobil uygulamalar da kod yazımıyla beraber çeşitli aşamalardan geçerler. Bu aşamalar planlama, analiz, tasarım, üretim, test ve bakım aşamalarıdır. Bu aşamalar gerçekleştirilirken güncel değişimler ve genişlemelerle geriye dönüşü olabilen bir döngü şeklinde ilerlenmektedir. Bu döngü Şekil 4.4'te görülmektedir. Aşamalar gerçekleştirildikten sonra hata düzeltimi, yeni istekler, bakım gibi nedenlerden dolayı döngüye tekrar girmek gerekebilir.



Şekil 4.4. Yazılım geliştirme yaşam döngüsü kademeleri

Planlama

Yazılım yaşam döngüsünün başlangıcıdır. Temel ihtiyaçlar belirlenir, sistem getirileri ve maliyet hesaplanır, proje planlaması oluşturulur.

Analiz

En önemli kademelerden biridir. Analiz sürecinde uygulamaların tüm işlevleri detaylarıyla belirlenir. Sistemin gerek duyduğu işlevler, gereksinimler netleşir. Gerçekleştirilmek istenen uygulamanın tüm detayları ortaya çıkarılır. Bu aşamada müşteri, yazılım mühendisi, sistem analisti, iş analisti, ürün yöneticisi gibi görevler üstlenmiş kişilerin oluşturduğu grup tarafından yürütülebilir.

Tasarım

Analiz kademesinde ortaya çıkan uygulama detayları, işlemlerin adım adım yapılacağı bileşenlere ayrılır. Uygulama planı oluşturulur. Bunun yanında uygulama bilgileri, sistem tasarım bilgileri, tasarım detayları, veri model, arayüz tasarımı, UML

diyagramlarını içeren tasarım dokümanı hazırlanmalıdır. Tasarım dokümanı, uygulama geliştirilmesini ve sonrasında uygulamayı anlatan, her geliştirici tarafından anlaşılabilir teknik bir harita görevini görecektir. Yüksek düzeyde tasarım ve detaylı tasarım olmak üzere iki tür tasarım yapılır. Yüksek düzeyde tasarımda modüllerin yapısı ve etkileşimleri tasarlanır. Detaylı tasarımda ise yüksek düzeyde tasarım gözden geçirilip tekrar ele alınarak tasarım tamamlanır. Tasarım kademesinde programlama problemlerinin çözümlerini daha basit olarak ele almak için soyutlama ve program çalışma sırasındaki yapısı ile ilgili dinamik, statik olmak üzere modelleme teknikleri gibi teknikler kullanılır.

Üretim

Planlama, analiz, tasarım kademelerinden sonra tasarım aşamasında oluşturulan planlama doğrultusunda yapılacak işlemlerin geliştirilme aşamasıdır. Programlama aşaması da denilebilir. Kodlamalar bu kademedeyi yapar. Alfa testleri bu aşamada gerçekleştirilmelidir.

Test

Uygulama geliştirilmesi tamamlandıktan sonra müşteri veya ilgili kişilere sunulmadan önce test ekibi tarafından beta testlerinin gerçekleştirilmesi aşamasıdır. Bu aşamada hatalar varsa giderilir, uygulama kullanıma alınır.

Bakım

Proje kullanıma sunulduktan sonra uygulamanın geliştirilmesi, iyileştirilmesi, oluşabilecek hataların giderilmesi gibi süreçler bu kademedeyi gerçekleştirilir.

Siyah Kutu Testi

Siyah kutu (black box) testi, fonksiyonellik testi, closed box testi, opaque testi gibi isimlerle de adlandırılır. Bu yöntemde kod ve tasarımla değil, gereksinim ve fonksiyonellik ile ilgilenilir. Siyah kutu testinde işleve uygun veri seçimi ve bu seçim doğrultusunda programın normal veya anormal davranış gösterimi gözlemlenir. Bu test

stratejisinde sistem gereksinimleri ve bu gereksinimlere nasıl cevap verileceği bilinmelidir. Siyah kutu testleri kullanıcı gereksinimi duymayan testler, kullanıcı gereksinimi duyan testler olmak üzere iki gruba ayrılabilir. Kullanıcı gereksinimi duymayan fonksiyonellik testi, araştırma testi, ad-hoc testi gibi siyah kutu testleri bulunmaktadır.

- Fonksiyonellik testinde uygulamaya ait fonksiyonlar test edilir. Sorun olmadığı sürece teste devam edilir.
- Araştırma testi, test edilecek olan uygulama hakkında test öncesinde fikir elde edilmesidir.
- Ad-Hoc testinde bütün testlerin sürelerinin belirlenmesini amaçlanır.
- Yenilenme testinde uygulamanın herhangi bir hata karşısında ne kadar sürede eski durumuna döneceğini test edilir.
- Seviye testi işlemlerle sistemin uç noktasına kadar gidilip bu uç noktanın belirlenmesini kapsar.
- Kullanılabilirlik testi arayüzün, birçok kullanıcı tarafından kullanılacak olmasından dolayı önemli olduğu uygulamalarda yapılması gereken test çeşididir.
- Stres testinde çok sayıda veri girişi, sorgu, büyük sayısal değerler işlemleri, aynı işlemin aynı anda yüksek tekrarlarla yapılması ile sistemin dayanıklılığı test edilir.
- Yükleme testinde kullanılacak sistemin nerede ve hangi şartlarda çökebileceğini belirlemekle, sistem performansı ölçülür.
- Duman testi genel olarak uygulanan mantık testidir. Bu testten geçen uygulama için daha sonra büyük çaplı testler uygulanmalıdır.

Kullanıcı gereksinimi duyan testler kabul testi, alfa testi, beta testinden oluşmaktadır.

- Kabul testi, kullanıcıların sistemin beklentileri karşılayıp karşılamadığına yönelik yaptığı testlerdir.
- Alfa testi, kullanıcıların uygulama geliştirme merkezine çağrılıp yapacağı işlemlere göre geliştiricilerin, düzeltilmesi veya değiştirilmesi gereken kısımları not ettiği testlerdir.
- Beta testi, kullanıcılar için uygulamanın beta versiyonunun yayınlanıp gelecek geri bildirimler doğrultusunda düzeltmelerin yapıldığı testlerdir.

Beyaz Kutu Testi

Beyaz kutu (white box) testinde uygulama kodu temel alınır. Cam testi, açık kutu testi, temiz kutu testi olarak da adlandırılır. Testi yapan kişi sorunlu kısmı bulmak amacıyla kodları inceler. Kod optimizasyonu ile fazla kodlar temizlenir, bütüne bakınca ortaya çıkan sorunlar belirlenir. Kodun en iyi şekilde test edildiği veri tespit edilir. Bu avantajları yanında belirli özelliklere sahip test yapıldığında maliyetin artması ve kodların tek tek incelenip hata bulma zahmetinin çok fazla olması dezavantajlarındandır. Birim testi, statik ve dinamik analiz testi, açıklama kapsamı testi, güvenlik testi, değişim testi beyaz kutu testleridir.

- Birim testinde yazılım geliştiriciler tarafından küçük kod blokları test edilir.
- Statik ve dinamik analiz testi için statik analizde kod sıralı bir şekilde incelenip hatalar araştırılır. Dinamik analizde ise kodun çalışması ve çıktı analiz edilir.
- Açıklama kapsamı testinde kod içerisindeki açıklamaların doğruluğu incelenerek test edilir.
- Güvenlik testinde izinsiz erişim gibi konularda sistem açıklarının tespit edilip bu açıkların kapatılmasını amaçlanır.
- Değişim testi, sistemdeki bir sorunun giderildikten sonra yapılan bir genel testtir.

5. MOBİL DUYGU ANALİZİ UYGULAMA TASARIMI

5.1. Mobil Duygu Analizi

Mobil duygu analizi, mobil cihazlarda duygu analizinin etkili olarak uygulanması olarak tanımlanabilir. Android işletim sistemi için mobil duygu analizi SentiCorr duygu analizi motoru prensipleri ile yerel olarak mobil duygu analizi gerçekleştirilmiştir. SentiCorr cümle düzeyinde POS kullanarak, cümledeki kelime türlerini tanımlayarak öznellik algılama aşaması ve daha sonra POS etiketlerini görüş sözlüğü tanımlamak için kullanır. Eğer cümle objektif ise POS etiketleri, olumlu ya da olumsuz ifadeleri işaret eden cümle kalıplarını aramak için polarite algılama aşamasında kullanılır. Öznellik (subjectivity) algılaması için AdaBoost ve polarite algılaması için RBEM (Rule-Based Emmission Model) kullanılmıştır. Android telefonlarının işlemcisinden dolayı orijinal POS etiketleyici kullanılamamaktadır. Javada yazılmış kütüphane yardımıyla POS etiketleyici kullanılabilir. SentiCorr cümle düzeyinde POS kullanarak yapılan mobil duygu analizinde %90 oranında başarı yakalanmıştır [44].

Mobil duygu analizi, istatistiksel ve sözlüksel benzeşim yöntemleriyle de gerçekleştirilebilir. İstatistiksel yöntemler ile yapılacak olan mobil duygu analizinde eğitim kümesine ihtiyaç vardır, konu kapsamına bağımlı olarak öncelikle eğitim veri seti oluşturulmalıdır. Otomatik bir öznitelik seçme algoritması ile istatistiksel yöntemler uygulanarak mobil duygu analizi gerçekleştirilebilir. Eğitim veri seti verileri arttıkça duygu analizi daha hassas olarak yapılabilir. Dezavantajı, her konu kapsamı için ayrı eğitim kümesi ve daha fazla işlem gerektirmesidir. Sözlüksel benzeşim yöntemiyle bir duygu sözlüğü veri tabanı ile duygu puanı veya duygu puanının kutupsal yönü baz alınarak mobil duygu analizi konu kapsamından bağımsız, fazla işlem gerektirmeyen bir şekilde uygulanabilir.

5.2. Mobil Duygu Analizi Uygulaması Yazılım Geliştirme Yaşam Döngüsü

Planlama

Planlama kademesinde, uygulama oluşturulurken yapılacak işlemlerin genel yapısı ortaya çıkarılmıştır.

- Duygu analizi işlemlerinin yapılması için duygu sözlüğü veri tabanına ihtiyaç vardır.
- Duygu analizi yapılacak metin girişinin nasıl yapılacağı belirlenmelidir.
- Duygu analizi sonucunun ne üreteceği belirlenmelidir.

Uygulama için yapılacak işlemler ana hatlarıyla belirlendikten sonra uygulamanın geliştirilmesi için gerekli programlama ve veri tabanı yazılımları belirlenmiştir.

- Android mobil uygulama, Google Şirketinin android uygulama geliştirmek için resmi olarak önerdiği Android Studio mobil programlama platformunun 2.1.2 versiyonu ile geliştirilmelidir.
- Uygulamanın gerektirdiği minimum Android işletim sistemi versiyonu, günümüzde hemen hemen tüm Android mobil cihazları kapsayacak şekilde 2.2 seçilmelidir.
- Duygu analizi için duygu sözlüğü veri tabanı, mobil uygulamalar için ideal olan SQLite veri tabanı yazılımında [55] oluşturulmalıdır.

Analiz

Uygulamanın, planlama aşamasındaki ortaya çıkarılan ana hatlarını ile uygulama detaylarıyla belirlenmiştir.

- Uygulama üç bileşenden oluşacaktır: Duygu analizinin yapılacağı analiz ekranı bileşeni, duygu analizi algoritması, menü bileşeni.
- Metin girişi üç şekilde yapılabilecektir:
 - Android klavye ile metin yazılarak
 - Başka bir uygulamadan kopyalanmış metin yapıştırılarak
 - Konuşulup ses kaydı ile alınan ifadeler metine dönüştürülerek
- Duygu sözlüğü veri tabanı, duygu içeren kelimeler ve bu kelimelerin duygu puanlarıyla oluşturulacaktır.
- Duygu analizi, duygu sözlüğü veri tabanı yardımıyla kutupsal değer içeren kelimelerin bulunmasıyla tespit edilen cümle duygu yönü ile metnin duygu yönü ve cümle veya kelime bazında kutupsal yön ağırlığıyla duygu oranı tespiti şeklinde gerçekleştirilmelidir.

- Duygu analizi sonucu, pozitif, negatif, duygu içermemeyi ya da eşit derecede negatiflik ile pozitiflik içermeyi ifade eden objektif duygu yönleri yanında pozitiflik ve negatiflik duygu oranı yüzdelik olarak ekrana yansıtılmalıdır.
- Menü bileşeni, dosya kaydı, dosyanın açılması ile e-posta ve kısa mesaj gönderme seçeneklerini içerecektir.

Tasarım

Analiz aşamasında detaylandırılan işlemlerin, bir zamanlama ile aşamalar halinde yapılması için uygulama planı oluşturulmuştur.

- Duygu sözlüğü veri tabanı oluşturulmalıdır.
- Duygu analizi algoritması geliştirilmelidir.
- Menü, metin giriş kutusu, ses kaydı butonu, analiz etme butonu, sonuç kutusu ve yeni analiz butonu bileşenleriyle uygulama arayüzü oluşturulmalıdır.
- Fonksiyon ve arayüz testleri yapılmalıdır.

Üretim

Tasarım aşamasında basamaklara ayrılan işlemlerin, belirtilen zaman zarfı aralığında arayüz düzenlemesi, kodlaması ile arayüz ve fonksiyon testleri yapılmalıdır.

Test

Uygulamanın bir beta sürümü yayınlanması yerine Android uygulama marketi olan Google Play Store'a yüklenmesi ve geri bildirim yorumlarına göre varsa hataların düzeltilmesi amaçlanmıştır.

Bakım

Uygulama Google Play Store'a yüklendikten sonra yapılan yorumlar dikkate alınarak geliştirme, iyileştirme ve oluşabilecek yeni hataların giderilmesi amaçlanmıştır.

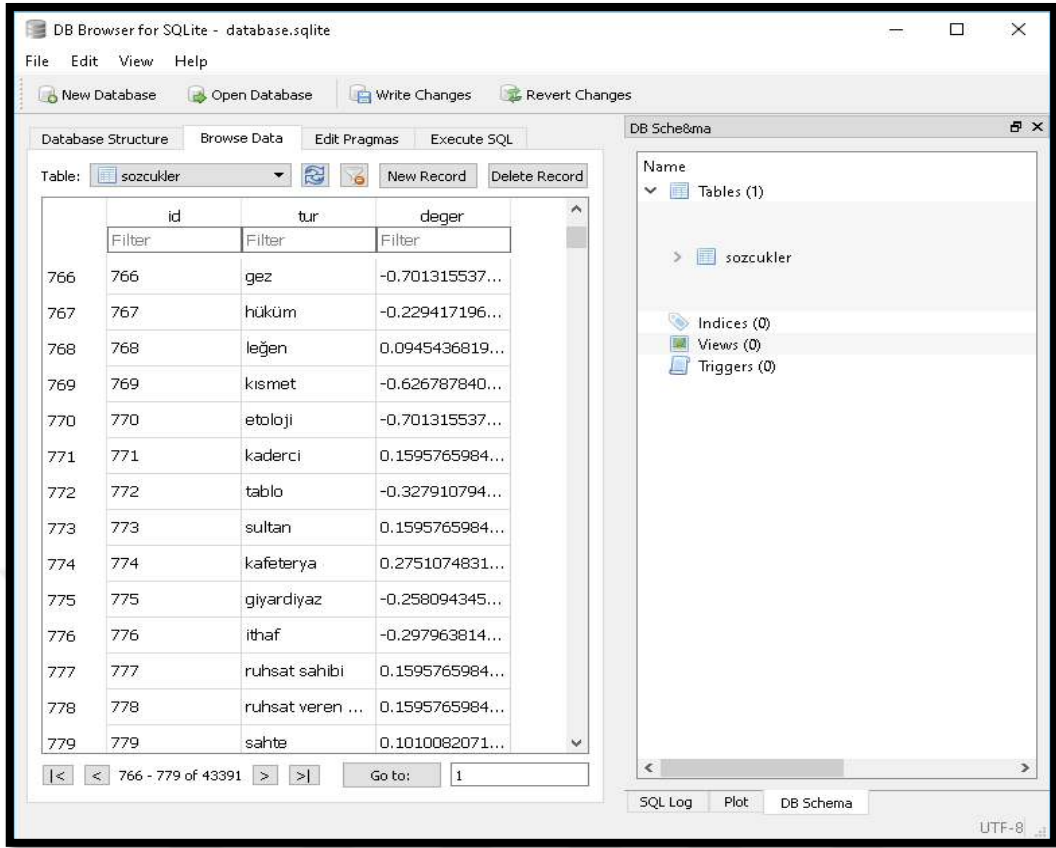
5.3. Mobil Duygu Analizi Veri Tabanı Özellikleri

Bölüm 3’te oluşturulma detayları belirtilen, test sonucunda doğru analiz yapabilen duygu sözlüğü, mobil uygulamalar için tercih edilen SQLite veri tabanı yazılımında tutulmuştur. Veri tabanında kayıtlar, “sozcukler” tablosu altında bulunan kelimenin tekil numarası “id”, kelimenin İngilizce karşılığı “ing”, kelimenin Türkçe karşılığı “tur”, kelimenin duygu puanı “deger” ve kelimenin türü “t” başlıklarını barındıran veri alanlarına kaydedilmiştir. Tablonun “id” veri alanı türü, primary key özelliği ile beraber ondalık olmayan sayı tipi integer, “değer” veri alanı türü, duygu puanını belirtebilecek hassas sayı tiplerini türü real, “ing”, “tur”, “t” veri alanları türü metin kaydı alabilecek text tipi olarak belirlenmiştir. Şekil 5.1’de duygu sözlüğü veri tabanının veri alanı adları ve türleri görülmektedir.



Şekil 5.1. Duygu sözlüğü veri tabanı tablo ve sütun başlıkları

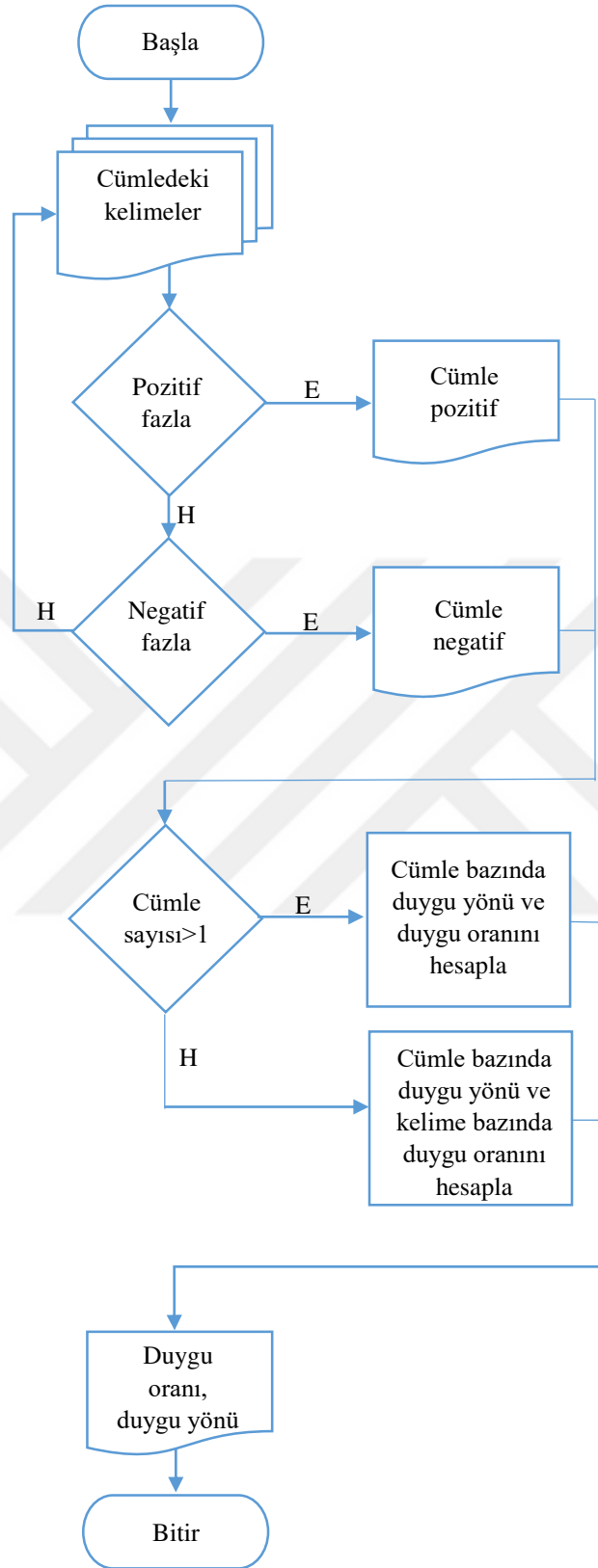
Mobil duygu analizi uygulamasının bu versiyonunda kullanılmayan “ing” ve “t” veri alanları işlem performansını arttırmak amacıyla veri tabanından çıkarılmıştır. Şekil 5.2’de veri tabanı içeriği görülmektedir. Veri tabanında “id”, “tur”, “değer” alanlarında kelime tekil numarası ile Türkçe sözcükler duygu değerleriyle beraber tutulmaktadır.



Şekil 5.2. Mobil duygu analizi uygulaması veri tabanı

5.4. Duygu Analizi Algoritma Yapısı

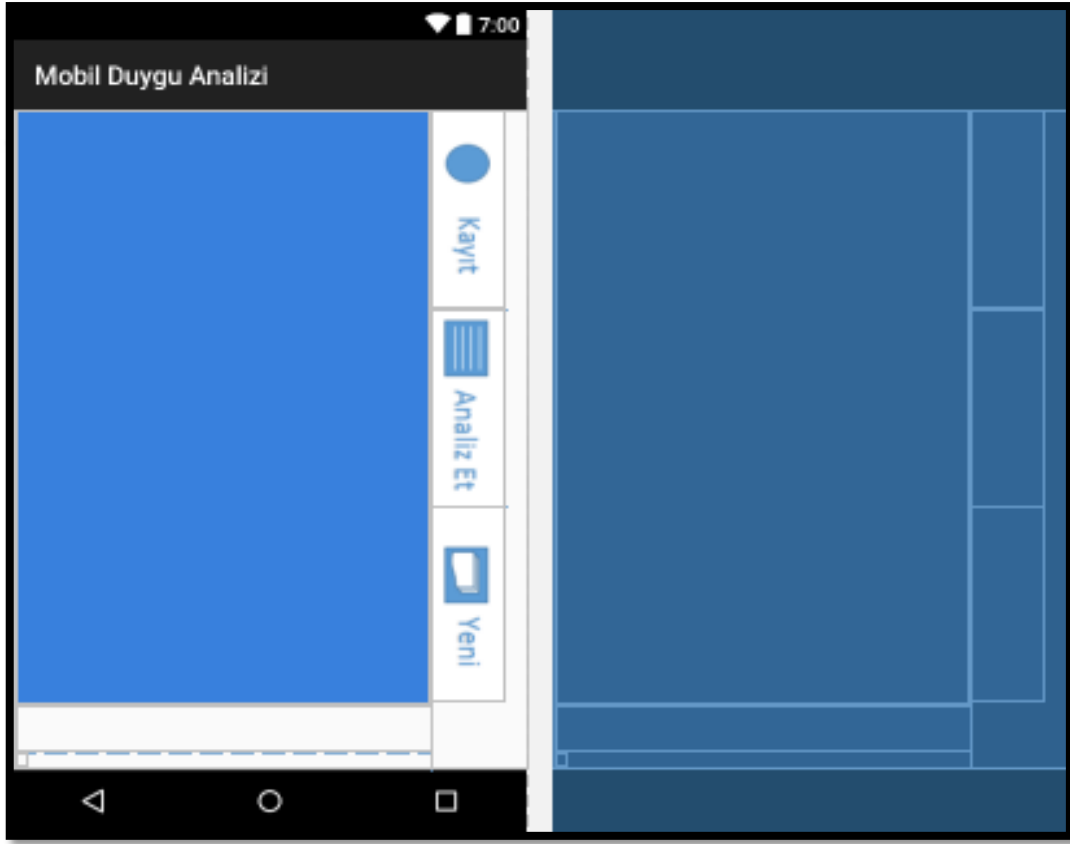
Resnik yöntemiyle oluşturulan duygu sözlüğü veri tabanı ile cümlede kelimelerin varlık ve duygu yönü baz alınarak veri setlerinde uygulanan duygu analizi en iyi başarı oranını göstermiştir. Mobil uygulama duygu analizinde de cümle polaritesini belirlemede kelime varlığı ve duygu yönü baz alınmıştır. Metin duygu yönü, sayıca fazla olan cümle duygu yönü olarak belirlenmiştir. Metnin bulunan duygu yönü yanında, metin bir cümleden oluşuyorsa kelime bazında, birden fazla cümleden oluşuyorsa cümle bazında duygu oranı hesaplanmıştır. Hesaplama basitçe duygu yönüyle paralel cümleler/kelimelerin toplam cümle/kelimeye oranı alınmıştır. Şekil 5.3'te duygu analizi algoritması görülmektedir.



Şekil 5.3. Duygu analizi algoritması akış şeması

5.5. Mobil Duygu Analizi Uygulama Arayüzü

Mobil duygu analizi uygulamasının arayüzü yapı olarak menü ve analiz ekranından oluşmaktadır. Arayüz için ImageButton, textView ve editText öğeleri kullanılmıştır. ImageButton öğesi ile görsel butonlar oluşturulabilmekte, textView ile sonuçlar görüntülenebilmekte ve editText alanı ile metin girişi alınabilmektedir. Şekil 5.4'te uygulamanın geliştirilme aşamasındaki arayüz tasarımı görülmektedir.

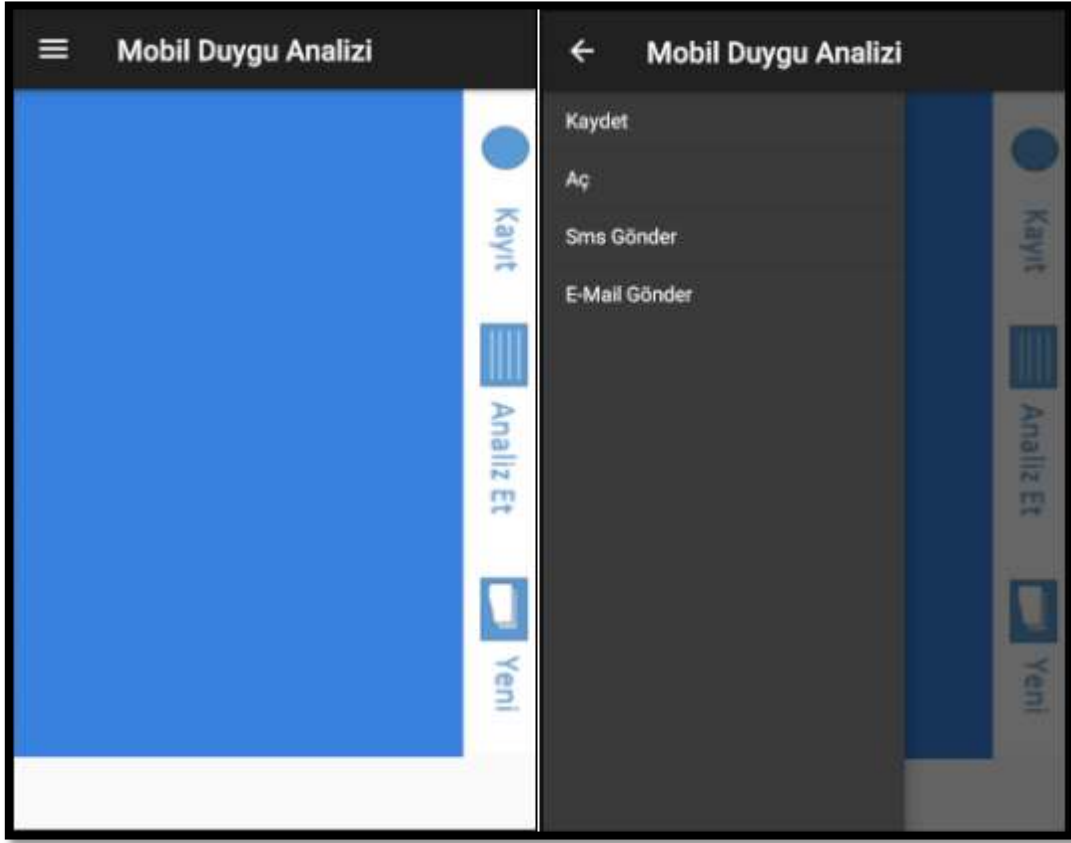


Şekil 5.4. Mobil duygu analizi arayüz tasarımına ait görüntü

Uygulama ile ilgili analiz ekranı ve menü arayüz görüntüleri Şekil 5.5'te görülmektedir. Analiz ekranında Kayıt butonu veya başka bir yolla girilen metnin Analiz Et butonu ile duygu analizi yapılmakta, sonuç ekrana yansıtılmakta ve yeni butonuyla metin ve analiz sonuçları silinerek yeni duygu analizi için uygulama bekletilmektedir.

Menü yapısında dosya işlemleri Aç, Kaydet seçenekleriyle sağlanmaktadır. Kaydet seçeneği seçilerek dosya otomatik olarak kaydedilmekte ve Aç seçeneği seçilerek mevcut

dosyaların metinlerine erişilebilmektedir. Sms Gönder, E-mail Gönder seçenekleri seçilerek metin ve duygu analizi sonuçları bu yollarla başka kişilere iletilebilmektedir.



Şekil 5.5. Mobil duygu analizi uygulamasına ait arayüz görüntüsü

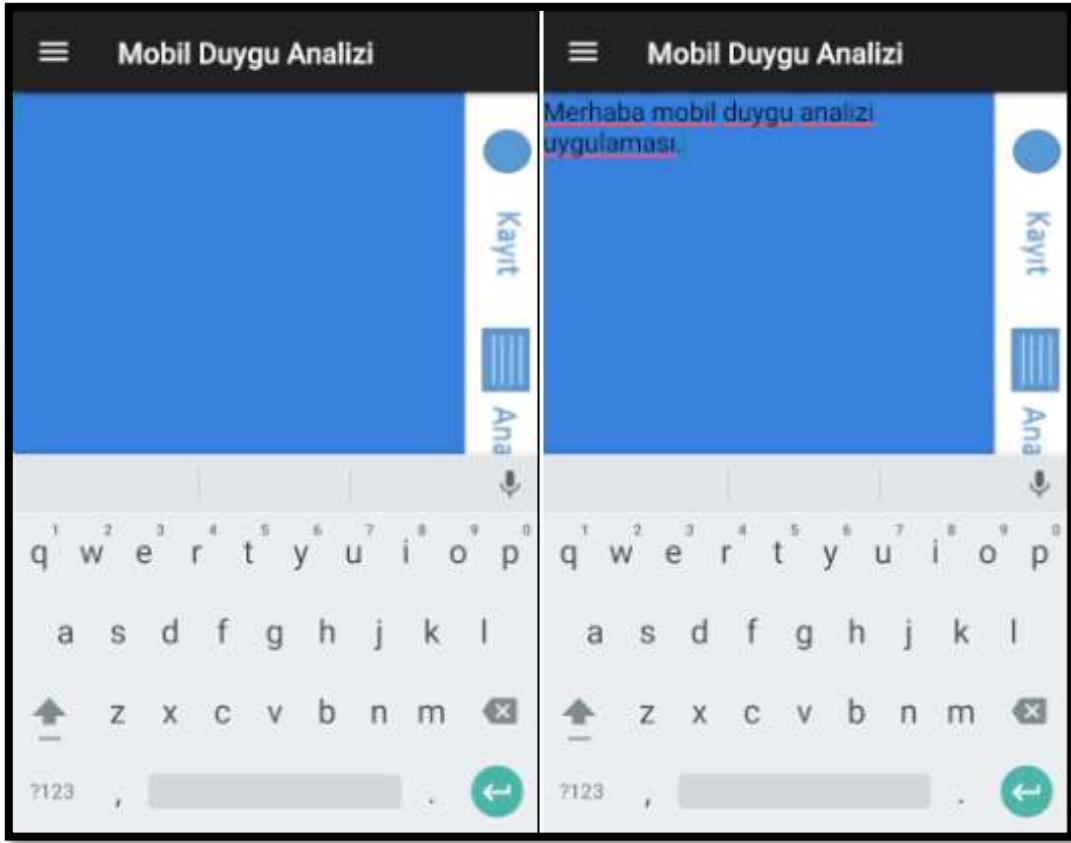
Mobil duygu analizi uygulamasında analiz edilecek metin,

- Mobil klavye ile yazılarak
- Kopyalanmış metin yapıştırılarak
- Ses kaydı ile alınan ifade metne çevrilerek üç farklı şekilde uygulamaya girilebilir.

Mobil Klavye ile Yazılarak Metin Girişi

Duygu analizi yapılacak metin, metin kutusu alanına mobil klavye ile girilebilmektedir. Ayrıca daha sonra kayıt butonuyla konuşmaların metne çevrilip eklenmesiyle karma olarak metin girişi yapılabilir. Şekil 5.6’da mobil klavye ile

metin giriş görüntüsü görülmektedir. Metin giriş alanına dokunulmasıyla imleçle beraber mobil klavye otomatik olarak ekrana yansımaktadır. Kısa mesaj ve E-postanın bu yolla yazılıp metnin içerdiği duygunun farkında olarak gönderilmesi avantaj sağlayabilmektedir.

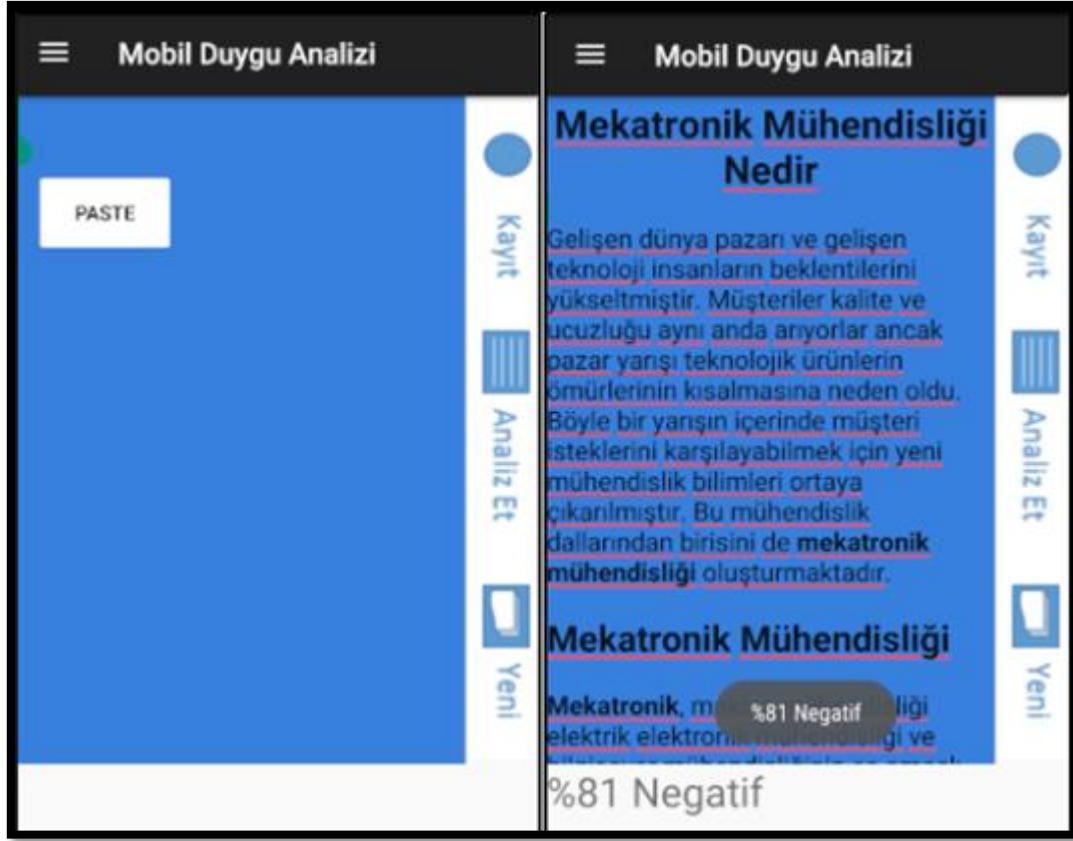


Şekil 5.6. Mobil duygu analizi uygulamasında mobil klavye ile analiz metni girişi arayüz görüntüsü

Kopyalanmış Metnin Yapıştırılmasıyla Metin Girişi

Kısa mesaj, e-posta, web tarayıcı vb. yollarla elde edilen metinler kopyalanıp metin kutusu alanına yapıştırılıp duygu analizine tabi tutulabilir. Metin giriş alanına dokunulup basılı tutulduğunda beliren yapıştır seçeneği ile metin eklenmektedir. Analiz et butonuyla duygu analizi yapılmaktadır. Şekil 5.7’de kısa sayılmayacak uzunlukta makalenin duygu analizi görüntüsü görülmektedir. Sayfalarca kopyalanmış metnin hızlı bir şekilde duygu analizinin yapıldığı gözlenmiştir.

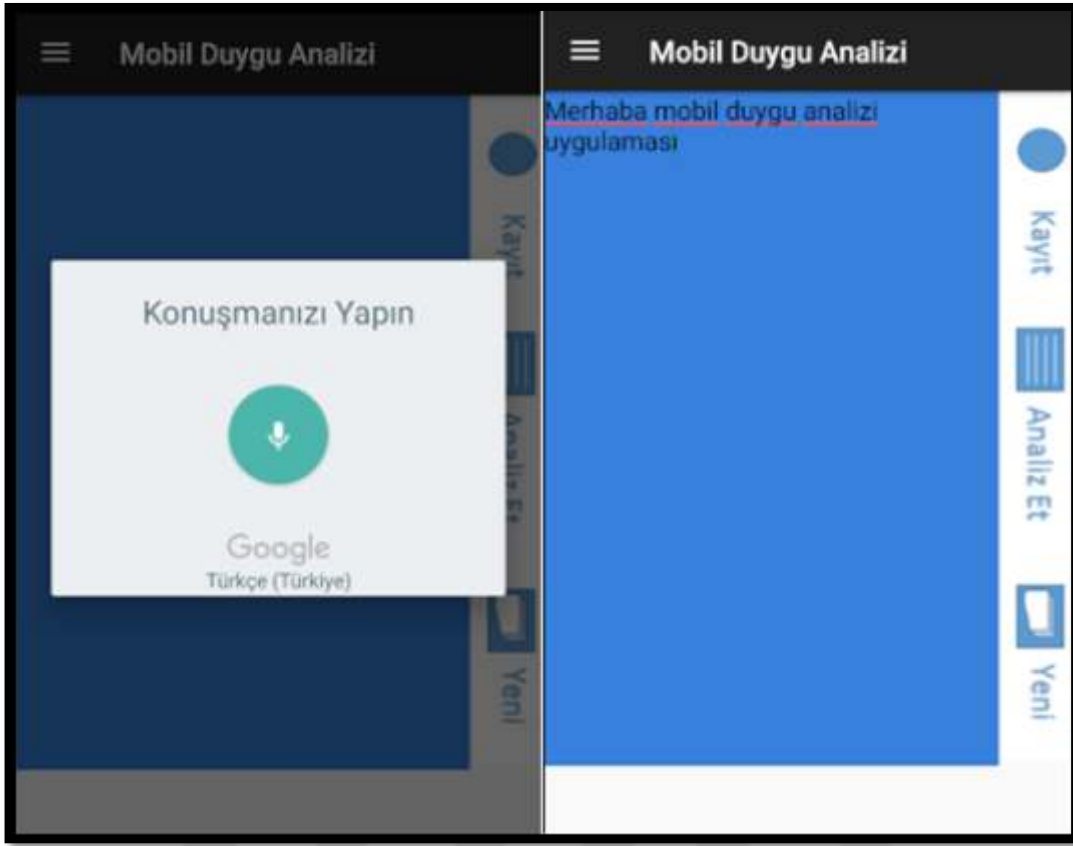
İnternetteki mevcut yorumlar, bu şekilde alınıp duygu analizine tabi tutulabilmekte, konu hakkında fikir sahibi olunabilmektedir. Ayrıca gelen kısa mesaj ve E-posta metinlerinin içerdiği duyguları tespit etmek iletişimde yarar sağlayabilmektedir.



Şekil 5.7. Mobil duygu analizinin metnin panoya yapıştırılmasıyla uygulanması arayüz görüntüsü

Ses Kaydı ile Alınan İfadenin Metne Çevrilmesiyle Metin Girişi

Mobil duygu analizi uygulamasında kayıt butonu ile Türkçe dili için ses verisi metne dönüştürülmektedir. Kayıt butonu ile çeviri aracı çalıştırılmakta ve konuşma sonlandırılana kadar ses verisi metne dönüştürülmek üzere kayıt edilmektedir. Şekil 5.8’de ilgili arayüz görüntüsü görülmektedir. Ses verisinin metne çeviriminde oluşan bazı kelime hataları manuel olarak mobil klavye ile giderilebilmektedir. Ses kayıt aracı daha sonra tekrar çalıştırılarak metne eklemeler yapılabilir.



Şekil 5.8. Mobil duygu analizi uygulamasında sesli ifadenin metne çevrilim arayüz görüntüsü

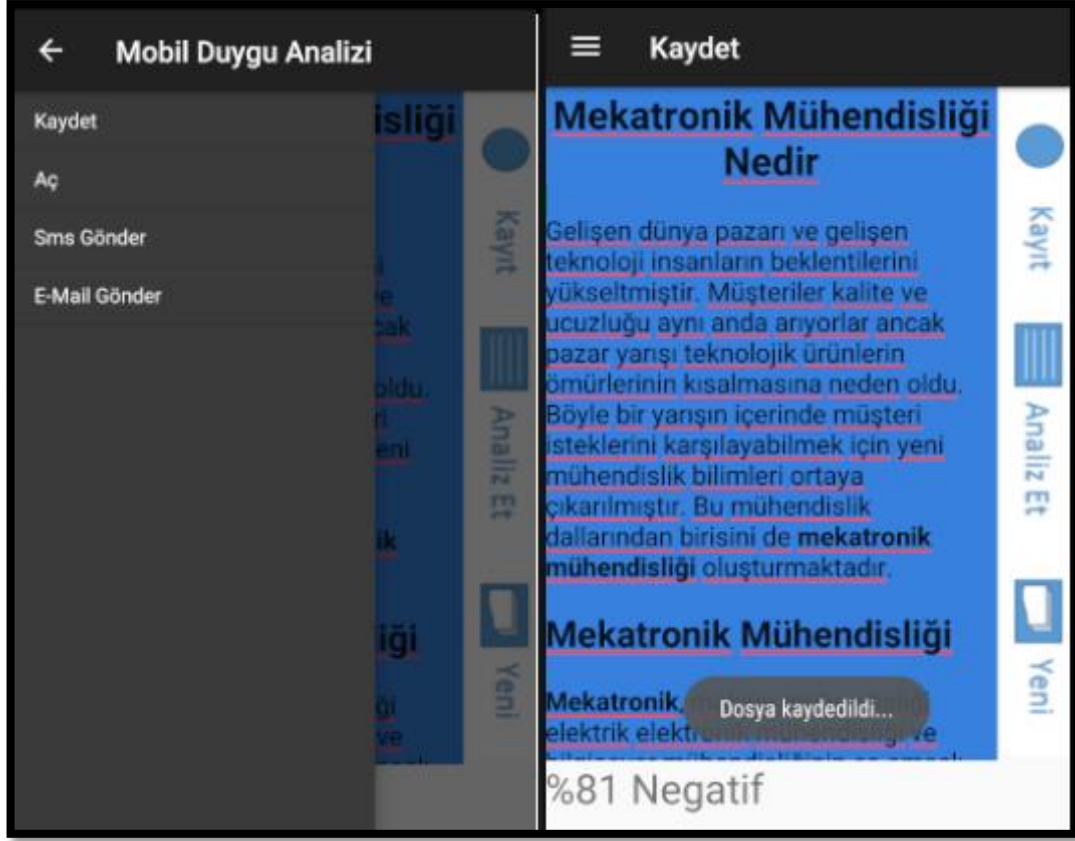
Mobil duygu analizi uygulamasında,

- a) Metinler analiz sonuçlarıyla birlikte txt uzantılı dosya olarak kaydedilebilmekte
- b) Kaydedilmiş dosyalar açılabilme, sdcard/mobilduyguanalizi/dosyalar/ klasörüne haricen yüklenen txt uzantılı dosyaları açılıp analiz edilebilmekte
- c) Analiz edilen metin, kısa mesaj veya e-posta olarak gönderilebilmektedir.

Metinlerin Kaydedilmesi

Duygu analizi yapılmış veya sonradan duygu analizi yapılmak üzere metinler dosya şeklinde kaydedilebilmektedir. Bu menü seçeneği ile ilgili uygulama arayüz görüntüsü Şekil 5.9'da görülmektedir. Kaydet seçeneği sadece butona dokunularak kullanılmaktadır. Beş kelimedenden fazla boyut uzunluğuna sahip olan metinler, menünün Kaydet seçeneği ile dosya ismi olarak metnin ilk beş kelime ve analiz sonucuyla sdcard/mobilduyguanalizi/dosyalar/ klasörüne txt dosyası olarak kaydedilmektedir.

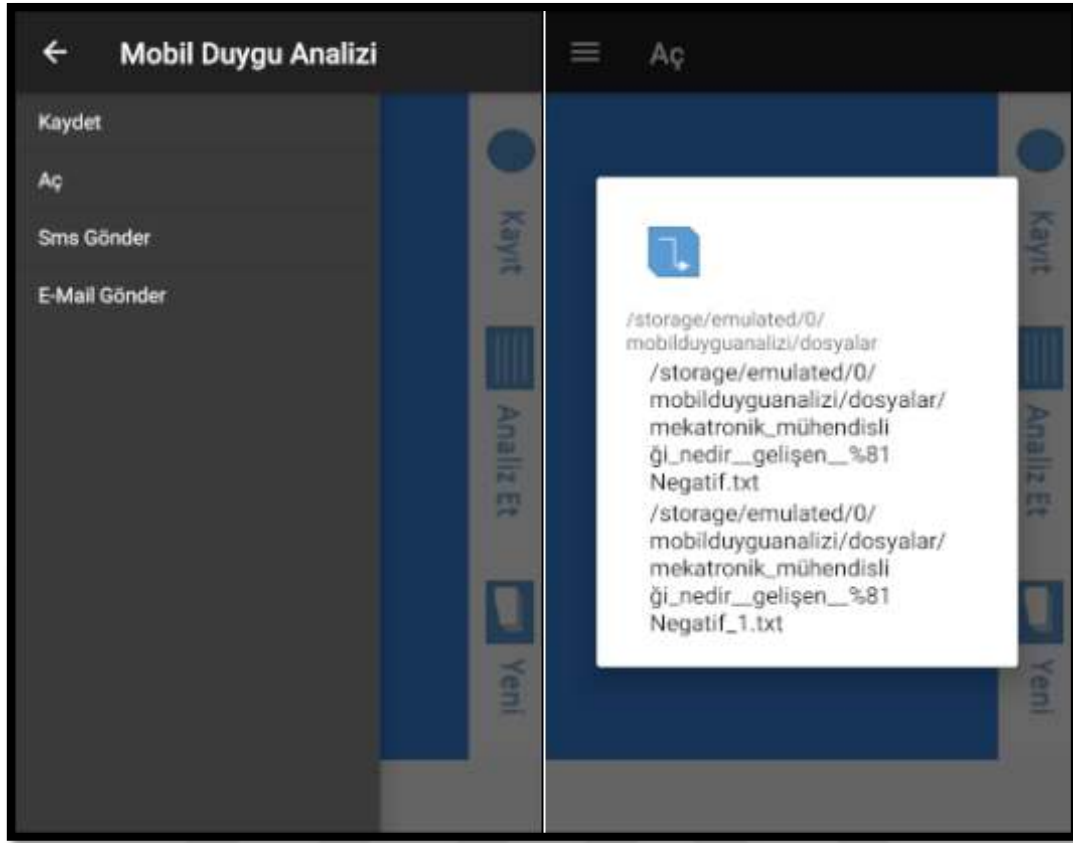
Otomatik olarak verilen dosya isminin tekrarlama sorunu mevcut dosya ismine 1'den başlayarak artan sayılar eklenmesi ile giderilmektedir.



Şekil 5.9. Mobil duygu analizi uygulamasında metnin txt dosyası olarak kaydedilme arayüz görüntüsü

Metinlerin Açılması

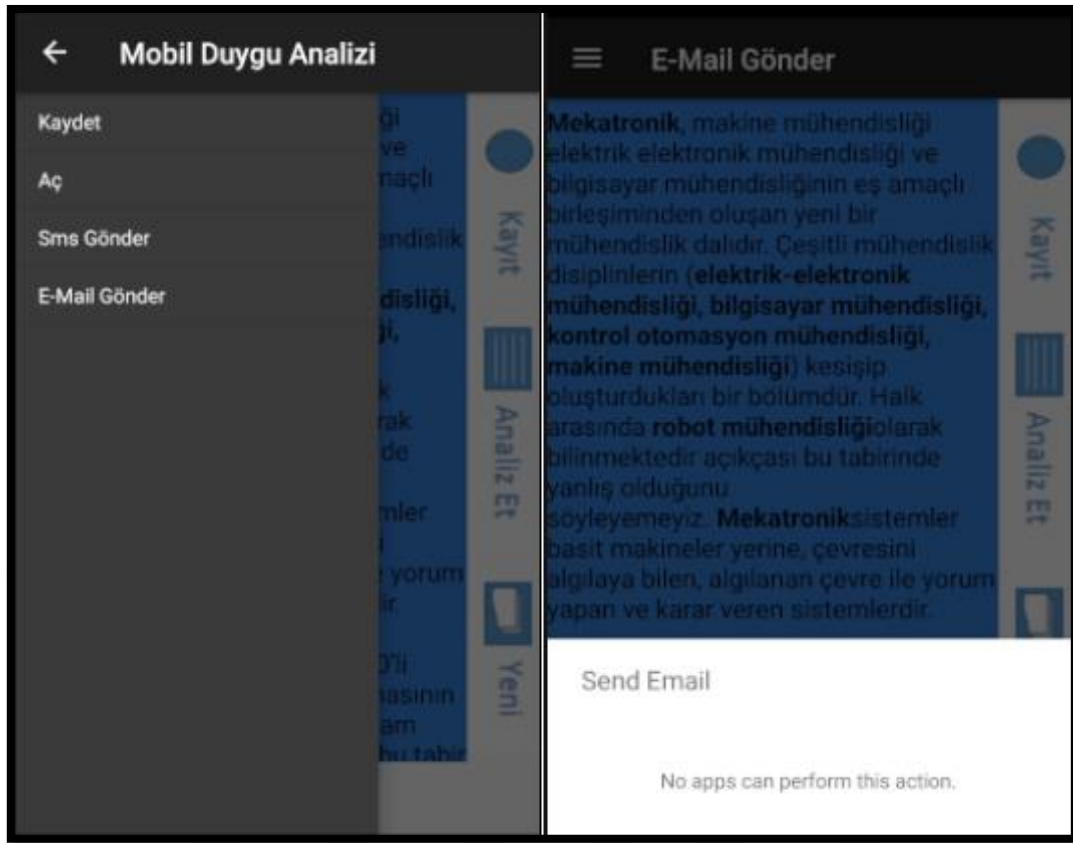
Mobil duygu analizi uygulamasında, kaydedilmiş veya sdcard/mobilduyguanalizi/dosyalar/ klasörüne analiz için yüklenen txt dosyaları menünün Aç seçeneğiyle açılıp kullanılabilir. Şekil 5.10'da Aç seçeneği arayüz ekranı görülmektedir. Menü Aç seçeneği seçildiğinde daha önceden kayıt edilmiş dosyalar listelenmektedir. Bu listeden açılmak istenen dosyanın seçilmesi ile dosyadaki metne ve bulunuyorsa duygu analizi sonucuna ulaşılabilmektedir.



Şekil 5.10. Mobil duygu analizi uygulamasında kayıtlı txt dosyalarının açılma arayüz görüntüsü

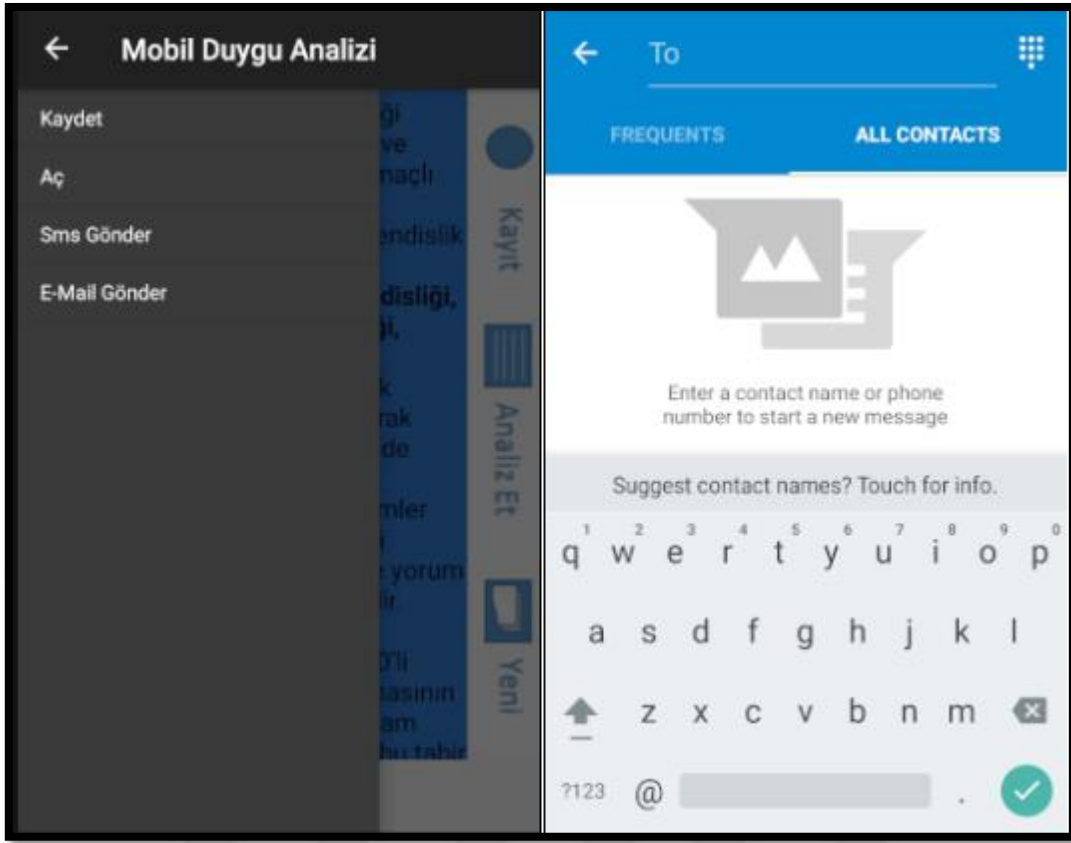
Metinlerin Gönderilmesi

Analiz edilmiş metin duygu analizi sonucuyla beraber menünün E-mail Gönder seçeneği ile iletilebilmektedir. E-mail gönderimi uygulama arayüz görüntüsü Şekil 5.11’de görülmektedir. Menü E-mail Gönder seçeneği seçildiğinde metin ve duygu analizi sonucu yüklü E-Mail programlarının birine yönlendirilmektedir. Önemli iletişim metinleri duygu analizi sonuçlarıyla beraber başka kişilere aktarılabilceği gibi istenirse duygu analizi sonucu silinerek sadece metin iletimi de yapılabilir.



Şekil 5.11. Analiz edilmiş metnin e-posta olarak gönderilme arayüz görüntüsü

Metinler ve analiz sonuçları, E-mail seçeneği dışında menünün Sms Gönder seçeneği ile kısa mesaj olarak iletelebilmektedir. Sms gönderimi uygulama arayüz görüntüsü Şekil 5.12’de görülmektedir. Metin ve duygu analizi yapılmışsa sonucuyla beraber sms gönderim arayüzüne iletilmektedir. Duygu analizi sonucu ile iletimi dışında, sadece gönderilecek kısa mesajdaki duygunun farkında olunması amacıyla duygu analizi yapıp sonuç metinden çıkarılabilmektedir.



Şekil 5.12. Analiz edilmiş metnin kısa mesaj olarak gönderilme arayüz görüntüsü

5.6. Mobil Uygulamanın Yayınlanması

Mobil uygulama yazılım yaşam döngüsü kademeleri tamamlandıktan sonra son kullanıcıların erişimine açılması yayınlama olarak adlandırılmaktadır.

Android uygulamalar Google Play Store uygulama marketinde yayınlanmaktadır. Android App Developer Zone olarak adlandırılan sisteme üye olunduktan sonra giriş yapıp Google kullanım sözleşmesine uygun olarak uygulamaya ait olacak bir sertifika ile uygulama imzalanmakta, Google Play Store'a yüklenip yayınlanmaktadır.

Root cihazlar olarak adlandırılan erişim yasaklı alanlara korsan şekilde erişebilen cihazlarda yüklü uygulamaların apk uzantılı dosyasına erişilebilmektedir. Dosya çözülüp kodlara erişim sağlanabilmektedir. Bu durumun önüne geçmek için Android Studio platformunda apk dosyasının çözülmesi zorlaştıran ProGuard kütüphanesi [56] kullanılmaktadır.

6. BULGULAR

Resnik Java kütüphanesi yardımıyla oluşturulan Türkçe duygu sözlüğü, film ve otel yorumlarını içeren veri setleri ile test edilmiş, başarımları ölçütleriyle değerlendirilmiştir. Film yorumları duygu analizinde kelime duygu yönü ve varlığı baz alındığında pozitif, negatif F-Ölçütü değerleri oranı %50'nin üzerinde bulunmuştur. %56 civarında doğruluk ve %59 civarında F-Ölçütü değeri oranı en yüksek başarı oranı olarak gözlenmiştir. Otel yorumları ile yapılan duygu analizinde en iyi sonuçlara kelimelerin duygu yönü, frekansı, varlığı baz alındığında ulaşılmıştır.

Türkçe metinler için duygu analizinin mobil cihazlarda yapılabilmesi için Türkçe duygu sözlüğü ile mobil duygu analizi uygulaması geliştirilmiştir. Mobil uygulama ile sesli ifadeler metne çevrilip duygu analizi yapılabilmiştir. Mobil uygulamada sayfalarca veri girilerek yapılan duygu analizi ile orta çapta verilerin mobil duygu analizi uygulamalarında analiz edilebileceği ve e-posta veya kısa mesaj yoluyla analiz sonuçlarının aktarılabilceği görülmüştür.

7. SONUÇLAR ve TARTIŞMA

WordNet sözcük veri tabanının bağlantıları ve benzerlik ölçümleri ile kelimelerin duygu puanları bulunmuştur. Sıfat, zarf, fiil kelime türlerinden kutupsal duygu değeri elde edilememiş ve duygu sözlüğü kapsamı daralmıştır. Resnik Java kütüphanesi kullanılarak hazırlanan sözlükte isim türündeki kelimelerin duygu değerleri eksiksiz hesaplanabilmiştir.

Türkçe sözlük veri tabanına ulaşılammıştır. Duygu sözlüğü kelime seçimi WordNet veri tabanından yapılmıştır. Kelime seçiminin genel bir Türkçe sözlüğünden yapılması Türkçe metinler için daha doğru duygu analizi yapılmasına olanak sağlayabilir.

Google, Yandex, Tureng gibi popüler çeviri araçlarıyla kelimelerin tamamının çevirimi gerçekleştirilememiş, kelimelerin Türkçe çevirilerin artırımı için birkaç sözlüğün bağlı olarak kullanılması gerekmiştir.

Duygu analizi veri ön işleme aşamasında genel etkisiz kelimeler listesi kullanılmasının duygu analizi sonuçlarını etkilediği görülmüştür. Duygu analizi için genel etkisiz kelime listesinin özelleştirilmesi gerekmektedir.

Türkçe metinler için sözlüksel benzeşim yöntemi ile duygu analizi mobil cihazlarda uygulanabilmiştir. SQLite veri tabanında oluşturulan duygu sözlüğü ile mobil duygu analizi uygulaması gerçekleştirilmiştir. Sesli ifadelerin metine çevrilmesiyle duygu analizi yapılabilmiştir. SentiCorr duygu analizi motoru prensipleri ile oluşturulan mobil duygu analizi uygulamasının öğrenme verisi bağımlılığı ile öğrenme verisi depolanması gerekliliği dezavantajı ve istatistiksel yöntemlerle yapılacak olan duygu analizinin, öğrenme verisi gerekliliği ile konu kapsam bağımlılığı göz önüne alındığında mobil duygu analizi için sözlüksel benzeşim yöntemi, konu kapsamından bağımsız olması, eğitim kümesine ihtiyaç bulunmaması gibi avantajlarından dolayı tercih edilebilir.

8. ÖNERİLER

Mobil duygu analizi uygulaması ile pozitif, negatif ve objektif olmak üzere metnin duygu yönü sınıflandırılmıştır. Sonraki çalışmalarda hüzün, korku, öfke, sevinç gibi temel duyguları da belirleyebilen mobil duygu analizi uygulaması geliştirilmesi düşünülmektedir.

Kurumsal yönetimlerde, örneğin ticari sektörler için görüşlerin analiz edilip yönetim politikalarının etkileşimli olarak otomatik yönlendirilebileceği bir mobil duygu analizi uygulaması geliştirilmesi planlanmaktadır.



KAYNAKLAR

- [1] <http://www.cs.waikato.ac.nz/ml/weka/> Weka 3: Data Mining Software in Java. 15 Mayıs 2016.
- [2] <http://sentiwordnet.isti.cnr.it/> SentiWordNet. 15 Mayıs 2016.
- [3] <https://wordnet.princeton.edu/WordNet> A Lexical Database for English. 18 Mayıs 2016.
- [4] <http://developer.android.com/studio/index.html> Android Studio the Official IDE for Android. 20 Mayıs 2016.
- [5] **Seker, S. E.**, 2016. Duygu analizi (Sentimental analysis), *YBS Ansiklopedi*, **3**, 21-36.
- [6] **Liu, B.**, 2011. Opinion mining and sentiment analysis, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data.: Springer*, **11**, 459-526.
- [7] **Çoban, Ö., Özyer, B. and Özyer, G.T.**, 2015. Sentiment analysis for Turkish twitter feeds. *23rd Signal Processing and Communications Applications Conference (SIU)*. IEEE, İnönü University, Malatya, Turkey, 16-19 May, 2388-2391.
- [8] **Nizam, H., Akın S.S.**, 2014. Makine öğrenmesi ile duygu analizinde veri seti performansı, Ders notları.
- [9] **Çatak, F.Ö.**, 2015. Eşle/İndirge yöntemi kullanılarak destek vektör makinesi algoritması ile yüksek boyutlu sosyal medya mesajlarının kutupsal değerinin ölçülmesi, Tübitak Bilgem Siber Güvenlik Enstitüsü, 1-12.
- [10] **Boynukalın, Z.**, 2012. Emotion analysis of Turkish texts by using machine learning methods, *MS Thesis*, M.E.T.U. The Graduate School of Natural and Applied Sciences, Ankara.
- [11] **Sevindi, B.İ.**, 2013. Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarının karşılaştırılması, *Yüksek Lisans Tezi*, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- [12] **Beyhan, H.D.**, 2014. Sosyal medya üzerinden metin madenciliği ve duygu analizi ile pazar değerlendirme, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- [13] **Güran, A., Uysal, M., Doğrusöz, Ö.**, 2014. Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi, *DEÜ Mühendislik Fakültesi Mühendislik Bilimleri Dergisi*, **48**, 86-93.

- [14] **Akgül, E.S., Ertano, C., Diri B.,** 2016. Twitter verileri ile duygu analizi, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, **2**, 106-110.
- [15] **Akba, F.,** 2014. Duygu analizinde öznelite seçme metriklerinin değeriendirilmesi: Türkçe film eleştirileri, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- [16] **Çelik, H.,** 2013. Sentiment analysis for Turkish language, *MS Thesis*, İstanbul Kültür University Institute of Science, İstanbul.
- [17] **Çetin, M., Amasyalı, M.F.,** 2013. Eğiticili ve Geleneksel Terim Ağırlıklandırma Yöntemleriyle Duygu Analizi, *21. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU)*. IEEE, KKTC, 24-26 Nisan.
- [18] **Seker, S. E., Al-Naami, K.,** 2013. Sentimental analysis on Turkish blogs via ensemble Classifier, *Proceedings of the 2013 International Conference on Data Mining (DMIN)*, Dokuz Eylül University, İzmir, Turkey 2013, 10-16.
- [19] **Yengi, Y.,** 2016. Büyük veride duygu analizine dayalı öneri sistemleri, *Yüksek Lisans Tezi*, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.
- [20] **Dehkharghani, R.,** 2015. Sentiment analysis in Turkish: Resources and techniques, *PhD Thesis*, Sabancı University Graduate School of Engineering and Natural Sciences, İstanbul.
- [21] **Çoban, Ö.,** 2016. Metin sınıflandırma teknikleri ile Türkçe twitter duygu analizi, *Yüksek Lisans Tezi*, Atatürk Üniversitesi Fen Bilimleri Enstitüsü, Erzurum.
- [22] **Evirgen, E.,** 2016. Sentiment analysis of Turkish tweets, *MS Thesis*, Bahçeşehir University Graduate School of Natural and Applied Sciences, İstanbul.
- [23] **Kama, B.,** 2016. Feature based sentiment analysis on informal Turkish texts, *MS Thesis*, M.E.T.U. The Graduate School of Natural and Applied Sciences, Ankara.
- [24] **Özgirgin, T.,** 2016. Sosyal medyada duygu analizi ve nitelik çıkarımı, *Yüksek Lisans Tezi*, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- [25] **Yengi, Y., Omurca, S.İ.,** 2016. Distributed recommender systems with sentiment analysis, Büyük Veride Tavsiye Sistemlerini Duygu Analizi ile Desteklemek, *Avrupa Bilim ve Teknoloji Dergisi*, **7**, 51-57, Kocaeli.
- [26] **Al-Bnd Abd Salman, M.,** 2015. Sentiment analysis and opinion mining via microblogging in social media like: Twitter, *MS Thesis*, Çankaya University The Graduate School of Natural and Applied Sciences, Ankara.

- [27] **Pak, M.Y.**, 2015. Metinlerde duygu analizi ve sınıflandırma için yeni yöntemler, *Yüksek Lisans Tezi*, Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir.
- [28] **Giriş, S.F.**, 2014. Sentimedia: Opinion mining and sentiment analysis on social media, *MS Thesis*, Fatih University The Graduate School of Sciences and Engineering, İstanbul.
- [29] **Husaini M.H.**, 2013. Sare: a sentiment analysis research environment, *MS Thesis*, Sabancı University Graduate School of Engineering and Natural Sciences, İstanbul.
- [30] **Demirci, S.**, 2014. Emotion analysis on Turkish tweets, *MS Thesis*, M.E.T.U. Graduate School of Natural and Applied Sciences, Ankara.
- [31] **M. Hu and L. Bing**, 2004. Mining and summarizing customer reviews, *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, San Jose, California, USA, July 25-29, 755-760.
- [32] **Kamps, J., Marx M.J., Mokken R.J. and De Rijke M.**, 2004. Using wordnet to measure semantic orientation of adjectives. *Proceedings of LREC 2004, the 4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, May 26-28, 1115-1118.
- [33] **Steinberger, J., Ebrahim, M., Ehrmann M., Hurriyetoglu A., Kabadjov M., Lenkova P., Steinberger R., Tanev H., Vazquez S. and Zavarella V.**, 2012. Creating sentiment dictionaries via triangulation, *Decision Support Systems*, **4**, 689-694.
- [34] **A. Esuli and F. Sebastiani**, 2006. Determining term subjectivity and term orientation for opinion mining, *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics*, (EACI, 06), Trento, İtalya, 3-7 April, 193-200.
- [35] **Akbaş, E.**, 2012. Aspect based opinion mining on Turkish tweets, *MS Thesis*, Bilkent University Graduate School of Engineering and Science, Ankara.
- [36] **Uçan, A.**, 2014. Otomatik Duygu Sözlüğü Çevirimi ve Duygu Analizinde Kullanımı, *Yüksek Lisans Tezi*, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- [37] **Türkmenoğlu, C.**, 2015. Türkçe metinlerde duygu analizi, *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

- [38] **Gandhewar, N., Sheikh, R.,** 2010. Google Android: An emerging software platform for mobile devices, *International Journal on Computer Science and Engineering*, **1**, 12-17.
- [39] **Tufan, M., Baykara, M., Güler, Z., & Avcı, E.,** 2011. Açık Kaynak Mobil İşletim Sistemi: Android İşletim Sistemi, *Elektrik-Elektronik ve Bilgisayar Sempozyumu*, Fırat Üniversitesi, Elazığ, Türkiye, 5-7 Ekim, 349-354.
- [40] **İnan, M., & Biçek, E.** 2016. Restoran Yönetim Süreçlerine Mobil Yaklaşım: Bir Android Uygulama, Akademik Bilişim Konferansı, Adnan Menderes Üniversitesi, Aydın, Türkiye, 30 Ocak-2 Şubat.
- [41] **Butler, M.,** 2011. Android: Changing the mobile landscape. *IEEE Pervasive Computing*, **10**, 4-7.
- [42] **Saha, A.K.,** 2008. A developer's first look at android, *Overview Developers (Linux for you)*, **Vol. Jan.**, 47-50.
- [43] **Gronli, T. M., Hansen, J., Ghinea, G., & Younas, M.,** 2014. Mobile application platform heterogeneity: Android vs Windows Phone vs iOS vs Firefox OS, *IEEE 28th International Conference on Advanced Information Networking and Applications*, Victoria, BC, Canada, 13-16 May, 635-641.
- [44] **Chambers, L., Tromp, E., Pechenizkiy, M., & Gaber, M.,** 2012. Mobile sentiment analysis, *Proc. of the 16th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems, Frontiers in Artificial Intelligence and Applications*, IOS Press, San Sebastian, Spain, September 10-12, 470-479.
- [45] <http://www.beyazperde.com/filmler/tum-filmleri/kullanici-puani> En İyi Filmler. 22 Ağustos 2016.
- [46] <https://www.booking.com/reviews.tr.html> Gerçek gezginlerden seyahat noktası değerlendirmesi. 24 Ağustos 2016.
- [47] <http://jsoup.org/> jsoup: Java HTML Parser. 25 Ağustos 2016.
- [48] **Budanitsky, A., Graeme H.,** 2001. Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures, *Workshop on WordNet and Other Lexical Resources*, **2**, 2-2.
- [49] **Lin, D.** 1997. Using syntactic dependency as local context to resolve word sense ambiguity, *Proceedings of the 35th Annual Meeting of the Association for*

Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain, July 07-12, 64-71.

- [50] <http://rednoise.org/rita/reference/RiWordNet.html> RiWordNet (Java-only). 27 Ağustos 2016.
- [51] <http://cloud.google.com/translate/docs/> Google Translate API. 29 Ağustos 2016.
- [52] <http://tech.yandex.com/translate/> Yandex Translate API. 3 Eylül 2016.
- [53] <http://tureng.com/tr/turkce-ingilizce> Tureng Online Çeviri Sözlüğü. 5 Eylül 2016.
- [54] **Pedersen, T., Patwardhan, S., & Michelizzi, J.,** 2004. WordNet:: Similarity: measuring the relatedness of concepts, *Demonstration papers at HLT-NAACL*, Boston, MA, USA, May 02-07, 38-41.
- [55] <http://sqlite.org/> SQLite. 7 Eylül 2016.
- [56] <http://proguard.sourceforge.net/> Proguard. 10 Eylül 2016.

EKLER

Ek 1. Stop words listesi

STOP WORDS

a	birkaçı	değil	henüz	kendine
acaba	birşey	demek	hep	kendini
altı	birşeyi	diğer	hepsi	ki
ama	biz	diğeri	hepsine	kim
ancak	bize	diğerleri	hepsini	kime
artık	bizi	diye	her	kimi
asla	bizim	dokuz	her biri	kimin
aslında	böyle	dolayı	herkes	kimisi
az	böylece	dört	herkese	l
b	bu	e	herkesi	m
bana	buna	elbette	hiç	madem
bazen	bunda	en	hiç kimse	mı
bazı	bundan	f	hiçbiri	mi
bazıları	bunu	fakat	hiçbirine	mu
bazısı	bunun	filan	hiçbirini	mü
belki	burada	felan	ı	n
ben	bütün	filan	i	nasıl
beni	c	g	için	ne
benim	ç	gene	içinde	ne kadar
beş	çoğu	gibi	iki	ne zaman
bile	çoğuna	ğ	ile	neden
bir	çoğunu	h	ise	nedir
birçoğu	çok	hâlâ	işte	nerde
birçok	çünkü	hangi	j	nerede
birçokları	d	hangisi	k	nereden
biri	da	hani	kaç	nereye
birisi	daha	hatta	kadar	nesi
birkaç	de	hem	kendi	neyse

niçin	ö	siz	şu	v
niye	öbürü	sizden	şuna	var
o	ön	size	şunda	ve
on	önce	sizi	şundan	veya
ona	ötürü	sizin	şunlar	veyahut
ondan	öyle	son	şunu	y
onlar	p	sonra	şunun	ya
onlara	r	ş	t	ya da
onlardan	rağmen	şayet	tabi	yani
onların	s	şey	tamam	yedi
onların	sana	şeyden	tüm	yerine
onu	sekiz	şeye	tümü	yine
onun	sen	şeyi	u	yoksa
orada	senden	şeyler	ü	z
oysa	seni	şimdi	üç	
oysaki	senin	şöyle	üzere	

Ek Tablo 1. Türkçe duygu sözlüğündeki örnek terimler

İngilizce terim	Türkçe terim	Değer	Tür
"somberness"	"koyuluk"	"-0.211168101234138"	"i"
"hollowware"	"kase"	"0.275107483177388"	"i"
"flit"	"taşınma"	"-0.916192444663022"	"i"
"copyhold"	"kayıtlara uyarak araziyi elinde tutma"	"-0.701315537061614"	"i"
"flip"	"çevirmek"	"-0.552099612724328"	"i"
"a"	"bir"	"-0.115551066862812"	"i"
"garland"	"çelenk"	"0.0855971862631534"	"i"
"micropyle"	"mikropil"	"-0.0124862988842425"	"i"
"i"	"ben"	"-0.054921292607109"	"i"
"o"	"o"	"0.0485844767782824"	"i"
"t"	"ton"	"0.0687382369306366"	"i"
"wiliness"	"hinlik"	"-0.701315537061614"	"i"
"attempt"	"girişim"	"-0.916192444663022"	"i"
"quartern"	"ölçeğin dörtte biri"	"-0.258094345952835"	"i"
"quarters"	"kışla"	"0.275107483177388"	"i"
"better"	"daha iyi"	"0.169857412292924"	"i"
"flop"	"düşme"	"0.0374651312064825"	"i"
"flow"	"akış"	"-0.44952529166095"	"i"
"flavivirus"	"flavivirüs"	"0.159576598417647"	"i"
"luridness"	"dehşetlilik"	"-0.211168101234138"	"i"
"floc"	"topak"	"0.10100820715967"	"i"
"floe"	"yüzen buz kütlesi"	"0.159576598417647"	"i"
"weldment"	"kaynaklanmasına"	"0.406059447528715"	"i"
"greyback"	"leş kargası"	"0.140967082567193"	"i"
"eurasia"	"avasya"	"0.159576598417647"	"i"
"longitude"	"boylam"	"-0.0124862988842425"	"i"
"calendula"	"nergis"	"0.159576598417647"	"i"
"flux"	"akı"	"-0.238932328141551"	"i"
"condyle"	"kondil"	"0.10100820715967"	"i"
"rhinorrhea"	"rinore"	"-0.701315537061614"	"i"
"flub"	"gaf"	"-1.0307165002459"	"i"
"flue"	"baca"	"0.243024977850579"	"i"
"sketch"	"kroki"	"-0.067459476551102"	"i"
"bettor"	"bahisçi"	"0.159576598417647"	"i"
"tracery"	"yaprak şeklinde oyma"	"0.275107483177388"	"i"
"cloudiness"	"bulutluluk"	"-0.15907801556458"	"i"
"chessman"	"satranç taşı"	"0.705047051431444"	"i"
"spitfire"	"püsküren volkan"	"0.159576598417647"	"i"
"deferment"	"tecil"	"-1.0307165002459"	"i"
"abkhasian"	"abaza"	"-0.258094345952835"	"i"
"poignancy"	"dokunaklılık"	"-0.22941719640252"	"i"
"fluorosis"	"yuvaları"	"-0.258094345952835"	"i"
"poseur"	"numaracı"	"0.159576598417647"	"i"
"tubercle"	"tüberkül"	"-0.159483049695104"	"i"
"threads"	"ipler"	"0.275107483177388"	"i"
"exemplar"	"örnek"	"-0.701315537061614"	"i"
"shylock"	"cimri"	"0.0683320913195467"	"i"
"ablepharia"	"ablefaron"	"-0.258094345952835"	"i"
"studbook"	"atların şecere defteri"	"-0.258094345952835"	"i"
"poignance"	"acıyı"	"-0.258094345952835"	"i"
"teakwood"	"hint ağacı odunu"	"0.10100820715967"	"i"
"hematocele"	"hematosel"	"-0.701315537061614"	"i"
"flea"	"pire"	"0.159576598417647"	"i"
"flex"	"esnek"	"-1.0307165002459"	"i"
"iconolatry"	"ikonlara tapma"	"-1.0307165002459"	"i"

"argument"	"tartışma"	"-0.562424950108376"	"i"
"prototype"	"prototip"	"-0.701315537061614"	"i"
"ratepayer"	"vergi yükümlüsü"	"0.159576598417647"	"i"
"flaw"	"kusur"	"-0.22941719640252"	"i"
"flat"	"düz"	"0.069710097553968"	"i"
"flap"	"kapak"	"0.111654899426978"	"i"
"flan"	"turta"	"0.10100820715967"	"i"
"flak"	"uçaksavar"	"0.0176412181445987"	"i"
"subjection"	"boyun eğme"	"-0.401111008467646"	"i"
"brazil"	"brezilya"	"0.134458289911372"	"i"
"russet"	"rustik"	"0.275107483177388"	"i"
"catboat"	"tek direkli küçük yelkenli"	"0.406059447528715"	"i"
"gigabyte"	"gigabayt"	"-0.258094345952835"	"i"
"cheerio"	"şerefe"	"-0.258094345952835"	"i"
"protrusion"	"çıkıntı"	"-0.401111008467646"	"i"
"flag"	"bayrak"	"0.136432929103418"	"i"
"flab"	"şişmanlık"	"0.10100820715967"	"i"
"hobgoblin"	"cin"	"-0.623391588499212"	"i"
"flax"	"keten"	"0.0861098335603023"	"i"
"distiller"	"damıtıcı"	"0.159576598417647"	"i"
"anorchism"	"anorkidi"	"-0.258094345952835"	"i"
"discretion"	"takdir"	"-0.28544576882151"	"i"
"cayman"	"timsah"	"0.159576598417647"	"i"
"foundering"	"çöküyor"	"-0.854150312242486"	"i"
"colophony"	"kolofan"	"0.10100820715967"	"i"
"blackbody"	"kara cisim"	"0.263086363747828"	"i"
"thiotepa"	"tiyotepa"	"0.10100820715967"	"i"
"clapper"	"şakşakçı"	"0.087071415851945"	"i"
"gymnasium"	"spor salonu"	"-0.253322704546842"	"i"
"archaism"	"artık kullanılmayan deyim"	"-0.258094345952835"	"i"
"joiner"	"marangoz"	"0.149363696118918"	"i"
"haggis"	"sakatat yahnisi"	"0.10100820715967"	"i"
"hexane"	"hekzan"	"0.10100820715967"	"i"
"footwork"	"ayak hakimiyeti"	"-0.916192444663022"	"i"
"intersex"	"interseks"	"0.159576598417647"	"i"
"haggle"	"pazarlık etmek"	"-0.258094345952835"	"i"
"wishbone"	"lades kemiği"	"0.10100820715967"	"i"
"catamaran"	"katamaran"	"0.406059447528715"	"i"
"sealant"	"dolgu macunu"	"0.10100820715967"	"i"
"thruster"	"agresif kimse"	"0.113870165846924"	"i"
"valeriana"	"valeryan"	"-0.258094345952835"	"i"
"danube"	"tuna"	"0.10100820715967"	"i"
"sawfish"	"testere balığı"	"0.159576598417647"	"i"
"emigration"	"göç"	"-1.0307165002459"	"i"
"flesh"	"et"	"0.0537889942427643"	"i"

Ek Tablo 2. WordNet tanımlamaları [3].

Terim	Tanımlama
Adjective Cluster	Antonymous çifti veya üçlüsü etrafında organize olan synset grubudur. Bir sıfat kümesi antonymous kavramları temsil eden iki veya daha fazla head synset içerir. Her head synsetin bir veya daha fazla satellite synseti vardır.
Attribute	Sıfatların özelliğini belirten isimdir. Ağır ve hafif sıfatları için isim olan ağırlık kelimesi bir attribute kelimesidir.
Base Form	Çekimlerinin eklendiği kelime ya da kelime grubu için temel formdur.
Basic Synset	Yapı olarak synsetle aynıdır. 5WN’de kullanılır.
Collocation	WordNet’te collocation, boşluk veya tire ile bağlanmış iki ya da daha fazla kelime dizisidir.
Coordinate	Aynı hypernym’e sahip isimler veya fiillerdir.
Cross-Cluster Pointer	Bir sıfat kümesinden diğerine uzanan anlamsal bir işaretçidir.
Derivationally Related Forms	Aynı kök formu olan, anlamsal olarak ilişkili farklı sözdizimsel kategorili terimlerdir.
Direct Antonyms	Sıfat kümelerinde head synsetler arasında görülen doğrudan zıtlıktır.
Domain	Synsetin kategori, bölge veya kullanım işaretçileriyle bağlandığı sınıflamadır.
Domain Term	Yerel sınıfa ait bir synset. Category_Term, Region_Term veya Usage_Term olarak ta belirtilir.
Entailment	Bir X fiilinin Y’yi gerektirmesi, X’in Y’siz üretilmemesi veya üretilmemiş olmasıdır.
Exception List	Kelimeler için morfolojik dönüşümlerin düzenli olmayıp, algoritmik bir şekilde işlenememesidir.
Group	Benzer anlamlı fiillerin manuel olarak gruplanmasıdır.
Gloss	Her synset bir gloss içermektedir. Gloss bir tanım ve opsiyonel olarak örnek cümle içerir.
Head Synset	En az bir tane direct antonym’e sahip kelime içeren sıfat synsetidir.
Holonym	Parça bütün ilişkisinin bütün kısmını ifade eder.
Hypernym	Genel sınıf terimi. X Y’nin hypernymi ise Y X’in bir türüdür.
Hyponym	Belirli bir sınıfın bir üyesini belirtmek için kullanılır.
Indirect Antonym	Satellite synsette direct antonymi olmayan fakat head synsetin direct antonymi yoluyla indirect antonym’e sahip olan sıfattır.
Instance	Hyponym’in belirli bir formudur.
Lemma	WordNet veri tabanı index dosyalarında bulunan küçük harfli ASCII metin kelimeleridir. Genellikle kelime ya da kelime dizisinin temel formudur.
Lexical Pointer	Bir sözcük işaretçisidir. Synsetlerdeki kelimeler arasında ilişki olduğunu gösterir.
Lexicographer File	Sözlük bilimciler tarafından düzenlenmiş WordNet synsetleri için ham data içeren dosyalardır.
Lexicographer Id	Bir anlamı belirten lexicographer dosya içindeki tamsayıdır.
Monosemous	Bir sözdizimsel kategoride yalnızca bir anlama sahip olmayı ifade eder.
Meronym	Parça bütün ilişkisinde parça bölümünü belirtir.
Part Of Speech	Kelimenin isim, sıfat, zarf ve fiil olmak üzere hangi türde olduğunu belirtir.
Participial Adjective	Bir fiilden türetilmiş sıfattır (ortaç).
Pertainym	İlişkisel bir sıfattır. Zıt anlamları yoktur. Bir pertainym bir ismi veya başka bir pertainymi işaret edebilir.
Polysemous	Bir sözdizimsel kategoride birden fazla anlama sahip olmayı ifade eder.
Polysemy Count	WordNet’te bir sözdizimsel kategoride bir kelimenin toplam anlam sayısıdır.
Postnominal	Bir postnominal sıfat sadece ismi takip ederek oluşur.
Predicative	Sadece yüklem pozisyonları içinde kullanılabilen bir sıfattır.
Prenominal	Sadece isimden önce oluşabilecek bir sıfattır. Yüklem olarak kullanılamaz.
Satellite Synset	Bir sıfat kümesinde head synseti tarafından gösterilen ve benzer kavramları içeren bir synsettir.
Semantic Concordance	Bir anlamsal etiket yoluyla, sözlük içerisinde uygun anlamlarla birbirine bağlı bir metin içerisindeki kelimelerin kombine edilmesiyle oluşan bir metinsel yapı ve sözlüktür.

Semantic Tag	Bir metinsel dosyadaki bir kelimedeki WordNet veri tabanı içindeki bu kelimenin belirli anlamına giden bir işaretçidir. Bir semantic tag bir sense key tarafından temsil edilir.
Semantic Pointer	Synsetler arasındaki ilişkiyi gösteren bir anlamsal işaretçidir.
Sense	WordNet'te bir kelimenin anlamıdır. Bir kelimenin her anlamı farklı synsetlerde bulunur.
Sense Key	WordNet veri tabanı içinde bir anlamı bulmak için gerekli bilgidir. Bir sense key, bir lemma field ve kodları, lexicographer id, lexicographer file numarası ve bir satellite head synsetiyle ilgili bilgiyi birleştirir.
Subordinate	Hyponym.
Synset	Bir synonym set; gerçek anlamını yitirmeden bazı bağlamlarda kendi aralarında değiştirilebilen bir dizi kelimedir.
Troponym	Başka bir fiilin belirli bir şekilde detaylandırılmasını belirten bir fiildir.
Unique Beginner	Superordinate olmayan bir isim synsetidir.



ÖZGEÇMİŞ

Muharrem ÖNAL

onal.muharrem@gmail.com

Fırat Üniversitesi Teknoloji Fakültesi

Mekatronik Mühendisliği Bölümü

ELAZIĞ

16 Mayıs 1986 tarihinde Diyarbakır ilinde doğmuştur. Nevzat Ayaz Anadolu Lisesinde ortaöğrenimini bitirdikten sonra eğitim hayatına Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Öğretmenliği Bölümü ile devam etmiştir. Lisans mezuniyeti sonrasında Sağlık Hizmetleri Meslek Yüksekokulu Tıbbi Görüntüleme Teknikleri Bölümü, Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü ve Fen Bilimleri Enstitüsü Mekatronik Mühendisliği Anabilim Dalı Bilgisayar Sistemleri Programında öğrenimlerine başlamıştır. Ön lisans, lisans bölümlerini tamamlamış ve lisansüstü öğrenimine devam etmektedir.