



**MARMARA ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**TÜRKÇE METİNLERDE SÖZLÜK TABANLI YAKLAŞIMLA DUYGU**  
**ANALİZİ VE GÖRSELLEŞTİRME**

ISSA BABAN CHAWAI ABDOULAYE

**YÜKSEK LİSANS TEZİ**

Bilgisayar Mühendisliği Anabilim Dalı  
Bilgisayar Mühendisliği Programı

**DANIŞMAN**

Dr. Öğr. Üyesi Önder DEMİR

**EŞ-DANIŞMAN**

Dr. Öğr. Üyesi Buket DOĞAN

İSTANBUL, 2019



**MARMARA ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**TÜRKÇE METİNLERDE SÖZLÜK TABANLI YAKLAŞIMLA DUYGU**  
**ANALİZİ VE GÖRSELLEŞTİRME**

---

ISSA BABAN CHAWAI ABDOULAYE

(523616901)

**YÜKSEK LİSANS TEZİ**

Bilgisayar Mühendisliği Anabilim Dalı  
Bilgisayar Mühendisliği Programı

**DANIŞMAN**

Dr. Öğr. Üyesi Önder DEMİR

**EŞ-DANIŞMAN**

Dr. Öğr. Üyesi Buket DOĞAN

İSTANBUL, 2019

---

**MARMARA ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

Marmara Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Öğrencisi ISSA BABAN CHAWAI ABDOULAYE'nin “ **Türkçe Metinlerde Sözlük Tabanlı Yaklaşımla Duygu Analizi Ve Görselleştirme**” başlıklı tez çalışması, 26 Ağustos 2019 tarihinde savunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

**Jüri Üyeleri**

Dr. Öğr. Üyesi Önder DEMİR (Danışman)  
Marmara Üniversitesi Bilgisayar Mühendisliği Bölümü .....

Doç. Dr. Ali BULDU (Üye)  
Marmara Üniversitesi Bilgisayar Mühendisliği Bölümü .....

Dr. Öğr. Üyesi Mustafa Cem KASAPBAŞI (Üye)  
İstanbul Ticaret Üniversitesi Bilgisayar Mühendisliği Bölümü .....

**ONAY**

Marmara Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun **26.08.2019** tarih ve **2019/18-02** sayılı kararı ile ISSA BABAN CHAWAI ABDOULAYE'nin Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programında Yüksek Lisans derecesi alması onanmıştır.

**Fen Bilimleri Enstitüsü Müdürü**  
**Prof. Dr. Bülent EKİCİ**



## ÖNSÖZ

İlk teşekkürlerim doğal olarak Tez çalışmam boyunca benden bilgisini, desteğini ve sabrını esirgemeyen hocalarım Dr. Öğr. Üyesi **Önder Demir**'e ve Dr.Öğr.Üyesi **Buket Doğan**'a gider ve onlara teşekkürü bir borç bilirim.

Bu tezin danışmanlığı kabul ettikleri için teşekkürlerimi sunarım. Onlar benim için danışmandan daha fazladır, onlar benim için mentorlarımdır ve onlarda izlenecek örnekler bulduğumu iddia edebilirim.

Ticaret üniversitesinde bilgisayar mühendisliği bölümünde öğretim üyesi olan Dr. Öğr. Üyesi **Mustafa Cem KASAPBAŞI**'ya bana jüri üyesi olma onurunu verdiği için de teşekkürlerimi sunarım.

Sonunda, ailemin tüm üyelerine daha doğrusu babama (**GAMBO ABDOULAYE**) ve anneme (**SALAMADOU GARBA**) bütün bu seneler boyunca güvenleri, destekleri, sevgileri, cesaretlendirmelerinden dolayı teşekkür ederim.

## İÇİNDEKİLER

ÖZET.....	v
ABSTRACT .....	vi
KISALTMALAR.....	viii
ŞEKİL LİSTESİ.....	ix
TABLO LİSTESİ.....	x

1. GİRİŞ.....	1
1.1 Metin Madenciliği.....	1
1.2 Duygu Analizi.....	2
1.3 Tez Çalışmasının Amacı.....	2
1.4 Metin madenciliği ile ilgili Alanlar.....	3
1.5 Literatür Araştırması.....	5
2. MATERYAL VE YÖNTEM: .....	9
2.1 istatistiksel yaklaşımlarla doküman korpusu işleme .....	9
2.1.1 Duygu analizi için metin sınıflandırma: .....	10
2.1.2 Belge korpusun temsili: .....	11
2.1.2.1 Dilsel birimi: .....	11
2.1.2.2 Metin Ön işleme Aşaması .....	12
2.1.3 Sınıflandırma teknikleri.....	13
2.1.3.1 Bayes sınıflandırıcı.....	13
2.1.3.2 polarite ve duygu yoğunluğuna göre sınıflandırma.....	15
2.2 Duygu analizi: .....	16
2.2.1 Duygu notasyonunun karmaşıklığı.....	16
2.2.2 Öznel cümle ile nesnellik cümle arasındaki karşılaştırma.....	16
2.2.3 Duygu analizine farklı yaklaşımlar.....	17
2.2.3.1 N-gramların sınıflandırmadaki rolü.....	17
2.2.3.2 Olumsuzluk işleme.....	17
2.2.4 Tez çalışmasında kullanılan sözlükler.....	17

2.3. Geliştirilen duygu analizi yazılım sisteminin mimarisi .....	22
2.3.1 Duygu analizi aşamaları .....	23
2.3.1.1 Ön işleme aşaması .....	23
2.3.1.2 Ön işleme örneği .....	23
2.3.1.2.1. Verilerin temizlenmesi ve bölünmesi .....	25
2.3.1.2.2. Morfolojik analiz .....	26
2.3.1.2.3 Boyut indirgeme .....	29
2.3.1.3 Duygu analizi aşaması .....	30
2.3.1.4 Üçüncü aşama: duygu görselleştirme, Bilginin çıkarılması .....	36
2.3.1.4.1 Frekans analizi .....	36
2.3.1.4.2 Kelimeler arasındaki birliktelik (cooccurrences) .....	36
2.3.1.4.3 Merkezleştirilme - kelimenin Önemi .....	38
2.3.1.4.4. Kelime Bulutları .....	39
2.3.1.4.5 Histogram .....	40
2.3.1.4.6 Metinden çıkartılabilecek diğer veriler .....	41
2.3.2 Duygu derecelendirme ve sınıflandırma modülü .....	45
2.3.2.1 Duygu derecelendirme .....	45
2.3.2.2 Metin sınıflandırma .....	47
2.3.3 Kullanılan veriseti .....	48
2.3.4 Uygulamanın çalışma prensibi .....	49
2.3.5 Uygulamanın Mimari Yapısı .....	51
3. BULGULAR VE TARTIŞMALAR .....	53
4. SONUÇLAR: .....	57

## **KAYNAKLAR**

## ÖZET

# TÜRKÇE METİNLERDE SÖZLÜK TABANLI YAKLAŞIMLA DUYGU ANALİZİ VE GÖRSELLEŞTİRME

Metin madenciliği (TM) yaklaşımları, yapılandırılmamış metinden otomatik olarak faydalı bilgilerin çıkarılmasını kolaylaştırmak için önem kazanmaktadır. Bu amaçla kullanılan metin madenciliği yöntemleri ile büyük miktardaki metin verileri, kısa zamanda ve yüksek performans ile analiz edilebilmektedir.

Bu çalışmanın temel amacı, kelime sıklığı, bilgi çıkarma, sınıflandırma, makine öğrenmesi, veri işleme ve çıkarma gibi araçlarla Türkçe metinlerinin işlenmesi ve görselleştirilmesidir.

Metin madenciliği alanlarından biri olan duygu analizi veya fikir madenciliği, görüşler, tutumlar ve duygular gibi öznel bilgilerin algılanmasını otomatikleştirmek için kullanılmaktadır. Bu tezde sunulan çalışmada, tweet, roman, müşteri yorumları gibi Türkçe metinlerin analizine ihtiyaç duyulduğu durumlarda kullanılacak yeni bir araç önerilmektedir. Geliştirilen uygulama esas olarak duygu analizi yaparken "kelime düzeyinde" çalışmaktadır. Buna ek olarak, duygu analizinde kullanılan sözlük sıfatlar, isimler, fiiller, zarflar, ön ekler, son ekler tğrgnde kelimeler içermektedir. Kullanılan birinci sözlük 5.000 kelime, ikinci sözlük 25.000 kelime ve üçüncü için de yaklaşık 25.000 kelime içermektedir.

Tez üç bölümden oluşmaktadır.:

Bölüm 1'de Tezin genel girişinde sırasıyla metin madenciliğine ve duygu analizine özgü tanımlar verilmiştir, ardında tez çalışmanın amacı ve önemi hakkında daha fazla bilgi paylaşmıştır. Son olarak metin madenciliği ile ilgili alanlardan ve çalışmalardan bahsedilmiştir.

Bölüm 2'de ise kullanılan materyal ve yöntemlerden bahsedilmiştir. Bu bölümde belge korpusunun nasıl temsil edildiği, kullanılan sınıflandırma teknikleri, duygu analizi yöntemleri, tez çalışmasında kullanılan sözlükler ve genel sistem mimarisi açıklanmaktadır.

Bölüm 3'te ise geliştirilen duygu analizi sistemine ait sonuçlar ve önerileri içermektedir

**Anahtar Kelimeler:** Metin madenciliği, duygu analizi, metin sınıflandırması, makine öğrenmesi.

## **ABSTRACT**

### **SENTIMENT ANALYSIS AND VISUALIZATION BY DICTIONARY BASED APPROACH IN TURKISH TEXTS**

Text mining (TM) approaches are important in order to facilitate the automatic extraction of useful information from unstructured text. In this purpose, with the utilization of Text Mining methods, large amounts of text data can be analyzed in a short time and with high performance. The main purpose of this study is to process and visualize Turkish texts with the help of tools such as word frequency, information extraction, classification, machine learning, data processing and extraction.

Sentiment Analysis or Opinion Mining is one of the text mining fields which is used to automate the perception of subjective information such as opinions, attitudes and emotions. The study presented in this thesis is a new tool proposed that can be used when the analysis of Turkish texts such as tweets, novels, customer comments is needed.

While searching or making sentiment analysis, the developed application works mainly at the word level. In addition, the dictionary used in emotion analysis includes different words like adjectives, nouns, verbs, adverbs, prefixes, suffixes. The first dictionary used is AFINN with 5,000 words, the second dictionary BING contains 25,000 words and the third dictionary NRC contains approximately 25,000 words.

The thesis consists of three parts:

In Chapter 1, in the general introduction of the thesis, the specific definitions of Text Mining and Sentiment Analysis are given respectively, after which the thesis shares more information about the purpose and importance of the study. Finally, the fields and studies related to text mining are mentioned.

In Chapter 2, the materials and methods used are mentioned. This section describes how the document corpus is represented, the classification techniques used, the sentiment analysis methods, the dictionaries used in the thesis study and the architecture of the general system.

Chapter 3 contains the results and recommendations of the developed Sentiment Analysis system.

**Keywords:** Text Mining, Sentiment Analysis, Text Classification, Machine Learning.



## KISALTMALAR

MA:	Makine Öğrenme
MM:	Metin Madenciliği
BA:	Bayes Algoritması
DA:	Duygu Analizi
MS:	Metin Sınıflandırması
ST:	Sözlüksel Tablosu
SO:	Sentiment Orientation
TM:	Text mining

## ŞEKİL LİSTESİ

- Şekil 1.1 Metin madenciliği ile ilgili Alanlar
- Şekil 2.2 "Afinn" sözlüğü örneği
- Şekil 2.3 "öznellik" sözlüğü örneği
- Şekil 2.4 "Duygular" sözlüğü örneği
- Şekil 2.5 "SentiTurkNet" sözlüğü örneği
- Şekil 2.6 Önişleme aşamasının adımları
- Şekil 2.7 Korpustaki kelimelerinin dağılımı
- Şekil 2.8 Örnek bir DTM tablosu
- Şekil 2.9 DuyguSınıflandır() fonksiyonu
- Şekil 2.10 duyguSınıflandır() Fonksiyonun ekran çıktısı
- Şekil 2.11 polariteSınıflandır() Fonksiyonu
- Şekil 2.12 polariteSınıflandır() Fonksiyonun ekran çıktısı
- Şekil 2.13 DuyguSınıflandır() fonksiyonu
- Şekil 2.14 Ortak oluşum grafi
- Şekil 2.15 Merkeziyet ölçülerinin grafiksel gösterimi
- Şekil 2.16 Kelime sıklıkları
- Şekil 2.17 Kelime bulutları
- Şekil 2.18 Fonksiyon çıktısı
- Şekil 2.19 Histogram
- Şekil 2.20 kitap kelimesi geçen dökümanlar.
- Şekil 2.21 Korpus hakkında sözcüksel rapor
- Şekil 2.22 143 numaralı belgede kelime inceleme
- Şekil 2.23 « kitap » kelimesiyle eş zamanlı gelen kelimeler.
- Şekil 2.24 Sözlüksel Tablosu (TLE)
- Şekil 2.25 « kitap » köküyle olan kelimeler
- Şekil 2.26 Samsung Q9 ile ilgili yorumlar

## TABLO LİSTESİ

<b>Tablo2.1</b>	Bayes Teoremine ait örnek tablo
<b>Tablo2.2</b>	Afinn sözlüğü örneği
<b>Tablo2.3</b>	Bing sözlüğü örneği.
<b>Tablo2.4</b>	NRC sözlüğü örneği.
<b>Tablo2.5</b>	Korpus oluşturma
<b>Tablo2.6</b>	Noktalama işaretlerini kaldırılması
<b>Tablo2.7</b>	Örnek durağan kelime listesi
<b>Tablo2.8</b>	StopList'deki kelimelerden arındırılmış korpus
<b>Tablo2.9</b>	Örnek metinde yer alan Stoplist kelimeleri
<b>Tablo2.10</b>	İleri morfolojik analiz öncesi Korpus
<b>Tablo2.11</b>	İleri morfolojik analiz sonrası Korpus
<b>Tablo2.12</b>	Örnek Frekans Tablosu
<b>Tablo2.13</b>	Measurement Metrics Table
<b>Tablo2.14</b>	Ci kategori için sınıflandırma sonucunun olasılığı :
<b>Tablo2.15</b>	Özet Tablosu.
<b>Tablo2.16</b>	DATA1 sonuçları
<b>Tablo2.17</b>	DATA2 sonuçları
<b>Tablo2.18</b>	DATA3 sonuçları

# 1. GİRİŞ

## 1.1 Metin Madenciliği

Günümüzde, internet hem kişisel hem de profesyonel bilgi alışverişi için vazgeçilmez bir araçtır. 2019 yılında metin yazmak veya okumak için internet platformunu kullanan 4,39 milyar aktif çevrimiçi kullanıcı bulunmaktadır. bu rakam Ocak 2018'e göre 366 milyon (%9) artış göstermektedir [1].

İnternet; kullanıcılara, bilgi aramada ve erişiminde büyük ölçüde yardımcı olmaktadır. İnternet ortamı; bilgi almak, otel rezervasyonu yapmak, ürün satın almadan önce diğer kullanıcıların incelemelerini veya yorumları kontrol etmek, bir sinema filmi ile ilgili yorumları okumak gibi amaçlarla ürünler ile ilgili bilgi sahip olmak için kullanılan ilk araçtır.

İnternet üzerinde dolaşan verilerin hızlı artışıyla metin madenciliği kullanımı da artmıştır. İnternet kullanıcılarının ürettikleri verinin %80'i yapılandırılmamış verilerden oluştuğunu bilinmektedir[2]. Yapılandırılmamış veriler, yasal belgeler, tıbbi kayıtlar, web sayfaları, görüntüler, gibi veri türlerinden oluşur. Metinsel veriler kitaplar, gazeteler, makaleler, bloglar, sosyal ağlar gibi birçok farklı biçimde bulunabilmektedir. Metin madenciliği, bu yapısal olmayan (unstructured) veri kaynaklarından verilerin analizini yaparak, yapılandırılmış (structured) veri elde etmeyi amaçlamaktadır.

Metin madenciliği, doğal dil halindeki metinlerden anlamlı bilgileri toplamaya çalışan bir alandır. Belirli amaçlar için yararlı olan bilgilerin elde edilmesi için metnin analiz edilmesi süreci olarak tanımlanabilir[3]. Elde edilen bilgi ile, analiz edilen metin kaynaklarında açık olarak görülmeyen ilişkiler, hipotezler ve eğilimler tespit edilir. Metin madenciliği, karmaşık metinlerin daha kolay bir şekilde anlaşılmasını sağlar.

2019 yılında 82,4 milyon nüfusa sahip Türkiye'nin nüfusunün %72'sini oluşturan 59,36 milyon internet kullanıcısı vardır[4]. Türkiye'nin yalnızca bir aylık e-ticaret kullanıcı etkinliklerine göre; internet kullanıcılarının %52'si satın almak üzere çevrimiçi(online) ürün veya hizmet aramakta ve bir perakende e-ticaret mağazasını ziyaret etmektedir.

Bu kullanıcıların %43'ü de aradığı ürünle ilgili satın almayı gerçekleştirmektedir [5].

## 1.2 Duygu Analizi

İnternet tüm dünyada, halka açık yorumların, görüşlerin ve duyguların paylaşımı için en

yaygın kullanılan iletişim platformudur. Birçok internet kullanıcısı sahip olduğu ürünler, ziyaret ettiği yerler, okuduğu kitaplar veya günlük olaylar hakkında görüşlerini, düşüncelerini ve duygularını internet siteleri veya sosyal ağlar aracılığıyla paylaşmaktadır.

Başkalarının görüşlerini bilmek, karar verme sürecinin her zaman önemli bir bilgi parçası olmuştur. Bir çok kişi herhangi bir karar vermeden önce, benzer durumdaki diğer kişilerin görüşlerini araştırır. Örneğin, bir satın alma işlemi yapmadan önce, diğer tüketicilerin görüşlerine danışılması, sinemada bir filmi izlemeden önce diğer kişilerin görüşlerine bakılması çok yaygın bir durumdur. İnternet sayesinde uzman olunmayan alanlarda, uzman kişilerin görüşlerinden ve deneyimlerinden yararlanılarak doğru tercihler yapılabilir. İnternet üzerindeki kullanıcı veya uzman görüşlerinin ise metin şeklinde yer aldığı görülmektedir. Metin madenciliğinde sunulan yöntemleri ve duygu analizinde kullanılan teknikleri aracıyla metnin ana duygusu tespit edilir. *Duygu analizi* metin madenciliğinde de önemli bir yere sahiptir. Metin içerisindeki anahtar kelimelere bakılarak metnin konusu belirlendikten sonra metnin “duygu haritası” çıkarılır. Bu işlem metnin okuyucuya aktarmak istediği duygunun daha kolay anlaşılmasını sağlar.

Duygu analizi, insanların sanal ortamlarda ürünler, hizmetler, organizasyonlar, olaylar, siyasi düşünceler, bazı konulardaki tutumları, duruşları gibi konular hakkında görüşlerini ifade ettikleri metinler içinde saklı olan duygu, fikir ve düşünceleri ortaya çıkarmayı, genel duyarlılık kutuplarını belirlemeyi amaçlamaktadır [6].

Metin madenciliğinin bir alt bölümü olan ve fikir madenciliği olarak da bilinen duygu analizi, doğal dil işleme, metin analizi ve hesaplamalı dilbilimciliği gibi bilimlerle yakın ilişki içindedir. Duygu analizi bu bilim dallarındaki yaklaşımları da kullanarak, kaynak materyallerdeki öznel bilgiyi tanımlamakta ve ortaya çıkarmaktadır [7,8]. Duygu analizi, pazarlamadan müşteri hizmetlerine kadar çeşitli uygulamalarda yaygın olarak uygulanmaktadır.

Duygu analizi, genellikle duygu polaritesi ve duygu skoru olarak adlandırılan iki kavrama bağlıdır. Duygu polaritesi veya kategorisi pozitif veya negatif bir ikili değerdir [9].

### **1.3 Tez Çalışmasının Amacı**

Bu tez çalışmasının amacı metin madenciliği yöntemleriyle “Duygu Analizi”

işlemini gerçekleştirmektedir. Gerçekleştirilen çalışmada, Türkçe metin verileri ile metin madenciliği uygulamaları geliştirilerek, bu alanda yer alan sınırlı çalışmalara katkıda bulunmak amaçlanmıştır. Bu amaçla İngilizce ve Türkçe sözcükler içeren Afinn, Bing, Nrc ve Sentitürk sözlükleri bir araya getirilip, değerlendirilerek duygu analizindeki kutupsallıkların belirlenmesi hedeflenmiştir. Böyle bir sistemin görevi, kendisine sunulan metni oluşturan sözcükleri analiz ederek, o metnin içerdiği en baskın duyguya göre veya polariteye göre sınıflandırma gerçekleştirmektedir.

Duyguların notasyon sistemine ilişkin olarak, çeşitli yaklaşımlar bulunmaktadır. Bu çalışmada notasyon sistemi 3 farklı sözlük üzerine odaklanarak yapılmaktadır. Birinci sözlükte her kelime için farklı sayısal değer (-2, -1.5, -1, 1, 1.5, 2) ile duygu yoğunluğu hesaplanmaktadır. İkinci sözlük ise eleştiri görüşünün kutupsallığı üzerine işlem yapmaktadır: negatif pozitif, zayıf pozitif, güçlü pozitif, zayıf negatif ve güçlü negatif. En son sözlük 6 farklı duygu sınıfı: üzüntü, mutluluk, sürpriz, korku, öfke şeklinde metinleri sınıflandırmaktadır. Bu sayede tüm yorumların bu yorumların bu kategorilere göre bir dağılımı oluşur [10].

Bu tez çalışmasında aşağıda belirtilen işlemler gerçekleştirilmiştir:

- Sözlüklerin çevrilerek Türkçe sözlüklerin oluşturulması,
- Metnin ön işlenmesi yapılarak analiz için uygun hale getirilmesi,
- Metnin sözcüklere ayrılması,
- Sözcüklerin duygu puanının sözlüklere göre belirlenmesi,
- her duygunun kategorisine göre toplam duygu puanının hesaplanması,
- en yüksek duygu puanına sahip olduğu kategoriye göre sınıflandırılması
- görselleştirme ile daha kolay yorumlanabilir hale getirilmesi sağlanmıştır.

Farklı metinlerin sınıflandırılması gerçekleştirmek için tüm bu yaklaşımların sonuçlarını bir araya getirecek bir sistem oluşturulmuştur.

#### **1.4 Metin Madenciliği Uygulama Alanları**

Metin madenciliği; Şekil 1.1’de görülen dilbilimi, bilgisayar bilimi, veri madenciliği, makine öğrenmesi teknikleri, web madenciliği olarak belirtilen beş alan ile yakın ilişki içerisinde.



**Şekil 1.1** : Metin madenciliği ile ilgili Alanlar[11].

Dilbilim, dilbilimin bilimsel sorularını cevaplamak için bilgisayar yöntemlerinin geliştirilmesi ile ilgilenen alandır. Dilbilim ile ilgili temel sorular; dil bilgisi, dil becerileri, dilin üretilmesinde ve anlaşılmasında dil bilgisinin nasıl edinildiği ve konuşlandırıldığıdır. Dilbilimciler, insanların dili nasıl kullandıklarını merak eder ve matematiksel modeller kullanarak modelleştirirler [12].

Bilgisayar biliminde amaç, metnin analiz edilmesi için gereken mühendislik problemlerine ilişkin en uygun çözümleri oluşturmaktır. Buradaki rasyonel, bilimsel bir teoriyi düşünmek ya da dilbilimde olduğu gibi X ve Y dillerinin tarihsel olarak ilişkili olduğunu kanıtlamak değildir. NLP (Natural Language Processing) gibi yöntemleri kullanarak büyük miktarda mevcut bilgiyi metin biçiminde keşfetmektir [13].

İstatistik alanı , metin madenciliğinin merkezinde yer almaktadır.

Veri madenciliği ve makine öğrenme tekniklerinde amaç az bilginin mevcut olduğu bir alanı keşfetmek veya gelecekteki gözlemleri doğru bir şekilde tahmin etmektir.

Veri madenciliği, veri analiz ve karar verme sürecini ifade eder.

Makine öğrenmesi, gelecekteki verilerin gözlemlenenlerden nasıl tahmin edileceğini bilmeyi amaçlamaktadır[14].

Web Madenciliği, internet kullanıcılarının davranış analizini içermektedir. Web

madenciliği üç çeşit analiz içerir. İlk analiz, sitelerin yapısını web sitelerinin arasındaki bağlantı sayısına göre araştırmaktır (bağlantılar). İkinci analiz, kullanıcıların tarihçelerini anlamak için web sitelerinin kullanılmasıdır[15]. Böylece, bağlantıların sayısı ve web sitesinde nasıl gezinildiği incelenir. Üçüncü analiz ise web sitesi temalarında yer alan kelimelerin tanımlanmasına karşılık gelen içerik analizidir. İçerik, veri kaynağına bağlı olarak çeşitli şekillerde gelir (sosyal ağlar, forumlar, bloglar, siteler).

Metin madenciliği işlemlerinde metin sınıflandırma önemli bir alandır. Bu işlemin yapılabilmesi için yukarıda belirtilen beş alan içerisinde makine öğrenme ve bilgi arama öncelikli alanlar olacaktır. Duygu analizi yaparken bilgi arama (Information Retrieval), metin sınıflandırma (Text Categorization), makine öğrenmesi (Machine Learning) veya metin arama etki alanlarımız olacaktır.

### **1.5 Literatür Araştırması**

Dehkharghani'nin gerçekleştirdiği çalışmada, SENTITURKNET isimli bir Türkçe kutup sözlük oluşturmuştur. SENTITURKNET sözlüğü English WordNet, SentiWordNet ve SenticNet sözlüklerinden yararlanarak oluşturulmuştur. Kelimenin İngilizcedeki karşılığının doğal anlamını bulmak amacıyla English WordNet Gloss, kelimenin olumlu, olumsuz, tarafsız gibi kutupsal anlamı için SentiWordNet ve duygusal anlamı için senticNet sözlükleri kullanılmıştır. Kelimelere olumlu, olumsuz ve tarafsızlık olarak toplamları 1 eden puanlar verilmiştir. Örneğin good kelimesinin olumlu puanı P:0.75, Olumsuz puanı N:0, Tarafsız puanı O:0.25'dir. Bu kelime için olumlu bir anlam yansıttığı söylenebilir. Duygu analizi gerçekleştirebilmek için kelime düzeyinde inceleme yeterli değildir. Bu amaçla cümle, doküman bazlı inceleme ve sınıflandırma yapılmıştır. Çalışmada lojistik regresyon, yapay sinir ağları, destek vektör makineleri algoritmaları kullanılmıştır. Geliştirilen uygulama WEKA yazılımı üzerinde çalıştırılarak Twitter üzerinden elde edilen veri seti üzerine uygulanmıştır [16].

Çelik ise çalışmasında farklı alanlara ait yeni veri grupları(kitap, film, ve alışveriş sitesi yorumları) oluşturulmuş ve oluşturduğu veri setiyle Türkçe duygu analizi yapmıştır. Weka doğal dil işleme kütüphanesi kullanılan çalışmada destek vektör makinesi, naive bayes sınıflandırıcı, lojistik regresyon, karar ağaçları gibi farklı makine öğrenmesi algoritmaları ile testler yapılmıştır. Çalışmada, naive bayes sınıflandırıcısı ile % 85 doğruluk elde edilmiştir [17].

Salman ve Al-Bndi'nin çalışmasında bir mikroblog sitesi olan twitter'dan elde edilen veriler üzerinde duygu analizi yapılmıştır. Bir öznitelik çıkarıcı ile kelimelerin frekansı bulunmuştur. Sonra öznitelikler kullanılarak tweet cümleleri sınıflandırılmıştır. Şirkete, kişiye, ürüne, filme, yerlere, etkinliğe gibi farklı farklı alana ait twitter data seti ele alınmıştır: 1000, 5000, 10000, 25000, 5000, 131 042 tweet cümleleri ve sırasıyla unigram için %70,19; %74,37; %74,93; %76,32; %76,32; %77,15 , bigram için ise %69,08; %75,20; %75,14; %74,65; %76,60; %76,88 doğruluk oranı elde edilmiştir. Kelime sayısının azlığı sınıflandırma işlemini zorlaştırdığından duygu belirten , 😞 , 😊 gibi simgelerde değerlendirilmiştir [18].

Yeliz'in çalışmasının amacı ise kullanıcı puanlama temelli öneri sistemlerinin, kullanıcı puanları yerine duygu analizinden elde edilen değerler ile gerçekleşmesidir. Bu çalışmada büyük veri analiz yöntemleri ile öneri sistemlerinin kişiselleştirilmesinin önemi ortaya konmuş ve bu sistemlerin temel yöntemlerine sunduğu katkı raporlanarak sonuçları analiz edilmiştir. Analiz sonuçlarına göre, verinin hacim olarak büyümesi ve kişiselleştirme adımları başarımlarını arttırmıştır. Aynı zamanda verinin hacminin büyümesinin hata değerine olan olumsuz etkisi de görülmektedir. Çalışmada sözlük tabanlı ve doğal dil işleme tabanlı duygu analizi kullanılmıştır. Geliştirilen yöntem Amazon ve Movie Lens verisetleri üzerine uygulanmıştır. Ayrıca Mahaout makine öğrenmesi ve doğal dil işleme kütüphaneleri kullanılmıştır. Her cümlelerin içerdiği duygu tespit edilip toplanarak tüm metnin duygusu analiz edilmektedir. Duygu değeri çok kötü (-2), kötü (-1), nötr (0), iyi (1), çok iyi (2) olarak beş farklı şekilde temsil edilmektedir: [19].

Ahmed'in gerçekleştirdiği tez çalışmasında, duyguları sınıflandırmak için sıklıkla kullanılan makine öğrenme (ML) yöntemlerini araştırarak, Arapça belgeler üzerinde duygu analizinin performans sonuçlarının iyileştirilmesi amaçlanmıştır. Aynı zamanda ML ile modelleme sürecinde girdilerin temsilinde yeni vektör biçimleri oluşturulmuştur. İkinci olarak maksimum, ortalama ve çıkarma gibi toplama işlevlerine dayalı farklı formüller kullanarak terim ve belge düzeylerinde sözcük tabanlı yaklaşımın uygulanması amaçlanmıştır. Arapça'da sözcüğe dayalı duygu analizinin terim ve belge seviyesi ile birlikte kullanılabileceğini göstermek için kullanılmıştır. Destek Vektör Makinesi (SVM), Karar Ağacı (D-Ağacı) ve Yapay Sinir Ağı (ANN) gibi farklı ML yöntemleri hem tek başlarına ve hem de kurallarla birlikte uygulanmıştır. Bu çalışmada elde edilen en başarılı sonuçlar yapay sinir ağı ile sınıflandırma gerçekleştirildiğinde elde edilmiştir. Genel olarak makine

öğrenmesi yaklaşımı sonuçları sözlük kullanımı yaklaşımına göre daha iyi sonuçlar vermiştir. Çalışmada Rushdi-Saleh tarafından oluşturulan OCA veri seti kullanılmıştır. Bu veri seti yarısı olumlu, yarısı olumsuz 500 film yorumundan oluşmaktadır [20].

Cumali tarafınadn gerçekleştirilen çalışmada ise, hem İngilizce hem de Türkçe için yapılan çalışmalarda kullanılan Makine Öğrenmesi (MÖ) ve sözlük tabanlı metotlara yeni özellikler eklenerek iki farklı veri kümesi üzerinde değerlendirme yapılmıştır. Daha önce İngilizce ve Türkçe metinler için çalışılmış metotlardan MÖ ve sözlük tabanlı DA olmak üzere iki ayrı metot Türkçe metinler için uygulanmıştır. Bu metotlar kısa ve uzun metinler olmak üzere iki farklı Türkçe veri kümesine uygulanıp başarımları ölçülmüştür. Kelimeler asıl anlamlarını köklerinde barındırdığından gereksiz ekler atılıp, asıl anlamı içeren kelime köklerine ulaşılmıştır. Varlık/yokluk (-lı,-li,-sız,-süz) eklerini ve olumsuzluk bildiren ekleri (-me,-ma) kaldırılmamış, özel bir işaret koyarak muhafaza edilmiştir. Sözlük tabanlı yöntem için her kelimesi taşıdığı duygu yönelimine göre puanlandırılmış bir sözlük kullanılarak bir metnin duygu yönelimini bulmaya çalışılmıştır. Kullanılan sözlüğü oluşturmak için Thellwal ve diğ. [21] çalışmalarında kullandıkları Sentistrength sözlüğünü Türkçeye çevrilmiştir. Veri seti Twitter yorumları veri kümesi ve film yorumlarından oluşturulmaktadır. Twitter veri kümesine uygulandığında sözlük tabanlı metod ile %75,2, makine öğrenmesi tabanlı metod ile ise, Karar Destek Makineleri (KDM) sınıflandırıcısı kullanılarak, %85 başarı elde edilmiştir. Film yorumları veri kümesine uygulandığında ise sözlük tabanlı metod ile %79,5, MÖ tabanlı metod ile KDM sınıflandırıcısı kullanılarak %89 başarı elde edilmiştir [22].

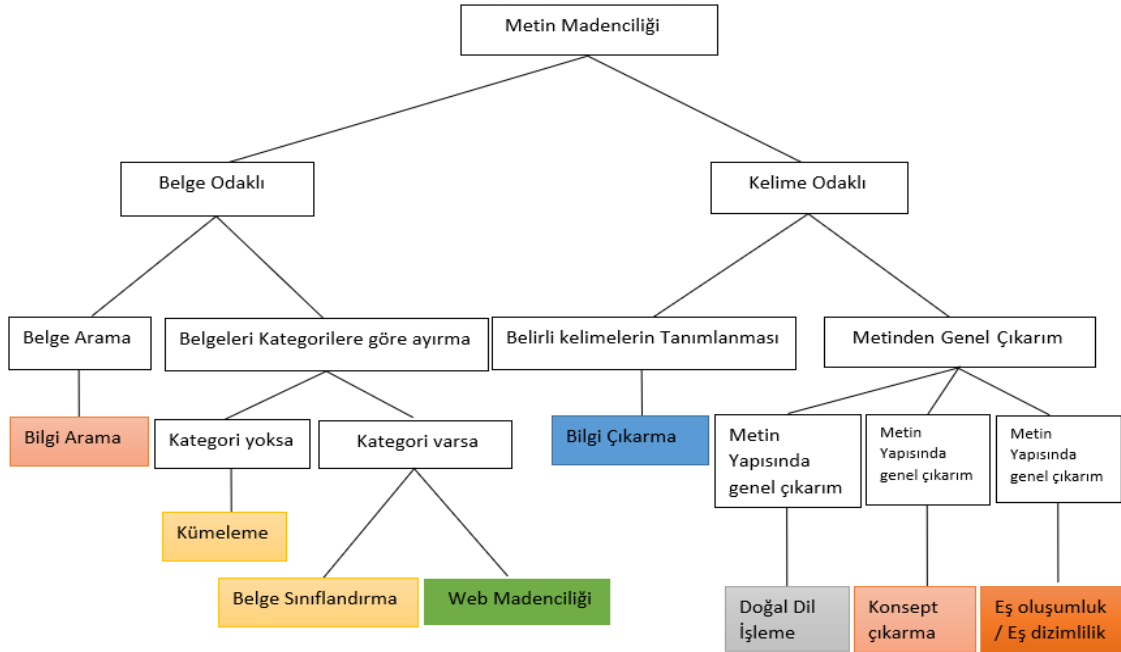


## 2. MATERYAL VE YÖNTEM

Tez çalışmasında kullanılan materyal ve yöntemler bu bölümde açıklanmıştır.

### 2.1 İstatistiksel Yaklaşımlarla Doküman Korpusu İşleme

Metin madenciliği alanındaki çalışmalar Şekil 2.1’de görüldüğü gibi gruplandırılabilir. Belge odaklı metin madenciliği çalışmaları, belge arama ve belgeleri kategorilere ayırma amacıyla kullanılırken, kelime odaklı metin madenciliği çalışmaları, belirli kelimelerin ortaya çıkarılması veya metinden genel çıkarım işlemlerini gerçekleştirebilmektedir.



Şekil 2.1. Metin madenciliği alanının amacını temsil eden karar ağacı [23].

#### Uygulama alanlarının amacı:

Şekil 2.1’de, Miner ve arkadaşları tarafından metin madenciliği alanındaki farklı amaçları ve analizleri göstermek için önerdikleri model görülmektedir.

Şekil 2.1’de görüldüğü gibi iki adet temel metin madenciliği analiz türü bulunmaktadır. Birinci analiz belgelere odaklanırken, ikincisi ise kelimelere odaklanmaktadır. Belge odaklı metin madenciliğinde belgenin bir kategoride olup olmadığını kontrol edilir, varsa o zaman belge üzerine bir belge sınıflandırma veya web madenciliği işlemleri yapılabilir. Yoksa, bir kümeleme işlemi yapılır.

Kelime odaklı bir metin madenciliği ise belirli bir kelime üzerine bilgiye sahip olmak istenirse bir bilgi çıkarma işlemi yapılır. Metnin genel bir çıkarımı yapmak istenirse 3 tane farklı yol vardır: birincisi metnin yapısına göre bilgisayar destekli bir metindeki kelimelerin analizine tahsis edilen bir bilim olan doğal dil işleme, ikincisi metin anlamından genel çıkarım için bir konsept çıkarma, üçüncüsü ise metin bağlantılardan bir eş oluşumluk veya eş dizimlilik(co-occurrence) işlemleri uygulanır.

### **2.1.1 Duygu analizi için metin sınıflandırma**

Sözlük temelli belge indeksleme, belge filtreleme, otomatik meta veri oluşturma, kelimelerin anlamlarından belirsizliği kaldırma, Web kaynaklarının hiyerarşik kataloglarına göre ayırmak ve genel olarak, belgelerin düzenlenmesi veya belgelerin seçici olarak işlenmesi ve uyarlanması gerektiren tüm işlemler için Metin sınıflandırma (Text Categorization). işlemi kullanılmaktadır.

Metin sınıflandırma, metin madenciliğinde çok önemli bir alandır, çünkü bu sayede bir metnin olumlu veya olumsuz olması durumunda nasıl karar verileceği konusunda bir kural oluşturulabilir.

Metin sınıflandırma birçok farklı alanda kullanılmaktadır:

Çok sayıda reklamı olan gazeteler, bir reklam için en uygun kategoriyi seçebilecekleri otomatik metin sınıflandırma yapan bir sistemden yararlanabilir.

Gazete makalelerinin uygun bölümler altında (örneğin, politika, olaylar, yaşam tarzları, vb.) otomatik olarak sınıflandırılması veya konferans bildirileri oturumlarında otomatik gruplama işlemi için metin sınıflandırma kullanılabilir [24].

Benzer şekilde, bir e-posta filtreleme sistemi, spam'i filtreleyebilir ve mesajları kullanıcı için tematik kategorilere ayırabilir.

Metin sınıflandırma teknikleri aynı zamanda kelimelerin anlamlarının belirsizliğini azaltmaya da yardımcı olur. (İng : Word Sense Disambiguation - WSD). WSD, belirsiz kelimelerin anlamlarını içeren bir metindeki araştırma faaliyetidir. Tek bir kelimenin birkaç anlamı olabilir. Bu nedenle, WSD sisteminin görevi, anlamlarından hangisinin olduğuna karar vermektir. Doğal dil işleme ve sözcüklerin anlamı ile belge indeksleme dahil birçok uygulama için WSD çok önemlidir.

Metin oluşturulan koşullarında hangisine uygun ise, o kategoride olacak şekilde

sınıflandırılmaktadır.

**if** <Koşul> **then** <category>

1990 yılı öncesinde belge sınıflandırma işlemi için yukarıdaki gibi bir koşul içeren ifade sonucunda, bir belge <category> kategorisine sınıflandırılarak işlem gerçekleştirilmektedir. Benzer şekilde birden fazla koşul bulunabilir ve hangi koşul sağlanıyorsa, yani sınırlamadan en az birisiyle uyduğu durumda belge o kategoriye ait olarak sınıflandırılır. 1990'ların başından ise, Makine öğrenmesi ile metin sınıflandırma yaklaşımı, popülerlik kazanmış ve en baskın kullanılan yaklaşım haline gelmiştir [25]. Bu yaklaşımda, bir öğrenme sürecinde, bir uzman bir belge kümesinin özellikleri gözlemleyerek, bir kategori  $C_i$  için bir sınıflandırıcı oluşturur. Bu yaklaşımda, bir öğrenme sürecinde, bir uzman elle sınıflandırarak bir belge kümesinin özelliklerini gözlemleyip bir kategori  $C_i$  için bir sınıflandırıcı oluşturulur. Örneğin, oluşturulan kurallardan bir metnin negatif olarak sınıflandırılması için, sözlükte negatif olarak yer alan kelimelerin sınıflandırmak istenilen metinde var olması gerekmektedir. Bu işlemi sağlayacak bir dizi kuralların da tanımlanması gerekmektedir. Endüktif işlem yöntemiyle, yeni belgenin  $C_i$  kategorisinde sınıflandırılması için gereken özellikleri belirlenir.

### **2.1.2 Belge korpusun temsili**

Metin madenciliğinde korpus veya külliyat kelimesi, işlenecek metinde yapısal değişiklik yapılarak, düzenli hale getirilmesi sonucunda ortaya çıkan yeni metin biçimini ifade etmektedir. Bu yeni metin biçimi veri analizi için kullanılacaktır.

#### **2.1.2.1 Dilsel birimi**

Doğal dilde metinler bir sınıflandırıcı veya sınıflandırma algoritmaları ile doğrudan yorumlanamaz. Bir belgenin anlamı, belgesel korpus analizinden kaynaklanan az ya da çok ayrıntılı özelliklere sahip belirli dilsel birimler kümesi tarafından belirlenmektedir. Anlamı temsil eden ilk dilsel birimler, kelimelerin yardımcı önermeleri, yani lemmasıdır. Bu dilbilimsel birimlerin tanımlanması için, metnin kelimelerinin dilsel ön işleminin yapılması gerekir. Dilsel birim, kelime veya cümle ile temsil edilebilir. İlk durumda, dilsel birim belgede görüldüğü gibi bir kelimedir. Her kelimedeki, boşluk, noktalama işaretleri gibi ayırıcılar dikkate alınarak metinden çıkarılır. Belge grubunu karakterize eden kelimelerin sayısı çok fazla olabilir. Bu nedenle, bu kelimelerin bir alt kümesini

tutmak gereklidir. Bu filtreleme, korpustaki kelimelerin ortaya çıkma sıklığına dayanır. Diğer yaklaşımlar kelimeleri değil kelime gruplarını kullanır. Cümleler, anlamı tanımlayan dilsel birim olarak kabul edilir. Bu tür bir dilbilimsel birim, bir kelimedenden daha çok anlam ifade eder. [26].

Bu yaklaşım sayesinde, kelimeler arasındaki sıralı ilişkisi temsil edilebilir. Bu yöntemin olumsuz yönü, kelime gruplarının görülme sıklığına dair güvenilir istatistikler sunmamasıdır. Çünkü kelimeler arasındaki çok sayıda kombinasyon, çok sayıda düşük frekans değeri üretir ve bu durum analiz için faydalı bir bilgi içermez.

### **2.1.2.2 Metin ön işleme aşaması**

Bir kelimenin benzer anlama sahip bir çok biçimi olabilir. Örneğin sevmek, sevdim, seviyorum gibi kelimeler aslında aynı anlama sahiptir.

Eğer kelimeleri dilsel birlik olarak kullanırsak, birçok kelimenin ortak anlamlara sahip olduğunu veya basitçe başka bir konjugasyon şekli oluşturduğunu fark ederiz. Bu kelimelere farklı bir anlam vermek, anlamsal bir ilişki olmadan gereksiz olacaktır. Bu nedenle metinlerde ekler atılarak, her bir kelimenin kökünü bulan, kök ayırma işlemi (stemming) adı verilen bir işlem yapılmaktadır. Kök ayırma işlemi metnin morfolojik analizini yapan bir işlemdir [27]. Bu işleme kelimelerin morfolojik incelemesi ile ve ekler sözlüğüne dayanarak, kelimenin kökünü çıkarmayı mümkün kılmaktır.

Bu noktada tez çalışmasında karşılaşılan en büyük zorluk geliştirme ortamında kullanılan kök ayırma kütüphaneleri elverişli olmamasıdır ve Rstudio ortamında Türkçe karakterleriyle bozukluk sürekli yaşanmasıdır .

Kökleşmeden veya kök bulmasından daha karmaşık bir analiz, bir sözlüğe dayanan lemmatizasyondur. Lemmatizasyonun amacı, her kelimeyle sözlükteki bir giriş ile ilişkilendirmektir. Aynı anlamdaki birçok kelime farklı kelimelerden gelebildiğinden, morfolojik analiz yetersizdir.

Kelimenin sözlükteki “doğru” hâlinin tespit edilen Lemmatizasyon, belirsizlikleri gidermek için sözdizimsel analizin eklenmesini gerektirir, bu yüzden morfosintaktik bir analiz yapılır [28].

Önceki işlemlerden herhangi birini gerçekleştirilmeden önce, belgenin anlamına aktif olarak katılmayan tüm sözcükleri kaldırmak için bir "stoplist" kullanmak yaygın bir uygulamadır. İşlemek istenilen metinden, zamirler gibi önemli olmayan kelimeleri yok edilmesi bu aşamada gerçekleştirilmektedir. Durma listesinin(stop-list) olumsuz yönü,

sınıflandırma için faydalı olabilecek kelimelerin de atılması riskini de içermesidir. Korpus oluştururken kelimelerin var olma veya olmama durumlarının belirlenmesi gerekir. Bir belgede herhangi bir kelimenin varlığını 1 değeri ile, yok olma durumunu ise 0 değeri ile göstermek mümkündür.

Korpusu oluşturulurken birimlerin(kelimelerin) ağırlığının belirlenmesi gerekir. Bu değer sıklıkla 0 ila 1 arasında değişmektedir. Özel bir durum olarak ağırlık iki değerli olarak, (1, belgede kelimenin varlığını ve 0 olmadığını şeklinde de gösterilebilir.

Bu işleminin sonunda, cümlenin bölümlenmesi, lemmatizasyonu ve belgeye anlam katmayan tüm kelimelerin silinmesinden sonra dilsel ön işlem yapılan bir vektör temsili elde edilmektedir.

### **2.1.3 Sınıflandırma teknikleri:**

Denetimli öğrenme olarak da adlandırılan sınıflandırma, verileri özelliklerine veya niteliklerine göre önceden tanımlanmış sınıfa atama işlemidir. Sınıflandırma yöntemlerinin amacı, nesnelerin belirli tanımlayıcı özelliklere göre ait oldukları sınıfların saptanmasıdır. Sınıflandırma yöntemini kullanan sistemlere örnek olarak hastanın klinik tanılarına göre tıbbi bir teşhis koymak gibi karar verme için destek sağlayan bir sistem verilebilir. Bir müşterinin kişisel verilerine göre banka kredisi talebine cevap verilmesi veya sensörler tarafından alınan sinyallere dayanarak bir uyarı sürecini tetiklenmesi gibi sistemler de sınıflandırma uygulamaları içerisinde yer alan sistemlerdir.

Sınıflandırma prosedürü bir dizi örnek kelimededen ve bir sınıflandırma metodundan oluşmaktadır. Sözlük tabanlı sınıflandırmada kelimeler gruplara göre; pozitif, negatif, mutlu veya üzüntülü gibi gruplara ayrılabilir. Bu örnek veri setinden ve sınıflandırma yönteminden yola çıkılarak metnin sınıfı belirlenir. Gerçekleştirile tez çalışmasında bayes sınıflandırıcı kullanılmıştır.

#### **2.1.3.1 Bayes sınıflandırıcı:**

Naive Bayes sınıflandırıcı, makine öğrenmesinde yaygın olarak kullanılan denetimli bir öğrenme algoritmasıdır. Metin sınıflandırma sorunları için özellikle yararlıdır[29].

Bayes teoremine ait formül Denklem 2.1’de görülmektedir.

$$P(A/B) = \frac{P(B|A) * P(A)}{P(B)} \quad (2.1)$$

“Bayes sınıflandırıcı”, Bayes teoremine ve koşullu olasılıklara dayanmaktadır. *Koşullu olasılık bir olayın gerçekleştiği bilindiğinde başka bir olayın ortaya çıkma olasılığını*

ifade etmektedir.

**Örnek :**

Bir lise sınıfı ve .  $A$  ve  $B$  aşağıdaki iki olay olduğu varsayıldığında,

$A$  Olayı: öğrenci bir kızdır durumuna,

$B$  Olayı : öğrenci Almanca öğrenir durumuna karşılık gelmektedir.

Almanca öğrenen bir kızın rastgele seçilme olasılığı nedir? gibi sorunun cevabı Bayes teoremi ile hesaplanabilir.

$P$  bir olayın olasılığını göstermektedir.

$P(\text{öğrenci bir kızdır ve öğrenci Almanca öğrenir}) = P(\text{öğrenci bir kızdır}) * P(\text{öğrenci Almanca öğrenir} | \text{öğrenci bir kızdır})$

Bu hesaplama buna eşdeğerdir:

$P(\text{öğrenci bir kızdır ve öğrenci Almanca öğrenir}) = P(\text{öğrenci Almanca öğrenir}) * P(\text{öğrenci bir kızdır} | \text{öğrenci Almanca öğrenir})$

**Tablo2.1** Bayes Teoremine ait örnek tablo

	Kız ( $A$ )	Erkek ( $\neg A$ )	Toplam
Almanca ( $B$ )	10	7	17
Diğer Diller ( $\neg B$ )	4	9	13
Toplam	14	16	30

$P(A/B)$  terimi,  $B$  olayının zaten gerçekleştiği bilindiğinde ,  $A$  olayının gerçekleşme olasılığını ifade etmektedir.  $A$  terimi kanıt ,  $B$  terimi Sonuç olarak adlandırılır.

**Örneğin,** Almanca konuşan bir öğrencinin kız olduğunu bilerek rastgele çıkartma olasılığı hesaplanmak istendiğinde Bayes teoremine göre Denklem ....'da görülen işlemler yapılmalıdır.

$$P(\text{Almanca} | \text{kız}) = P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B) * P(B)}{P(A)}$$

$P(A)$  Sınıf popülasyonundan bir kız rastgele alma olasılığıdır.  $P(A)$ 'ya önceki olasılığı (prior probability) ifade etmektedir. Kardinal ifadesi kümedeki öğelerin sayısını ifade etmektedir Örneğin, **cardinal( $\Omega$ )** örnekteki tüm öğrencileri temsil etmektedir.  $\Omega$  uzay kümesi temsil ediyor.

$$P(A) = \frac{\text{cardinal}(A)}{\text{cardinal}(\Omega)} = \frac{14}{30} = 0.4666$$

$$P(B \cap A) = \frac{\text{cardinal}(B \cap A)}{\text{cardinal}(\Omega)} = \frac{10}{30} = 0,3333$$

Bu sonucu verir :

$$P(B|A) = \frac{\frac{10}{30}}{\frac{14}{30}} = \frac{0,3333}{0,4666} = 0,7143$$

### 2.1.3.2 Polarite ve Duygu Yoğunluğuna Göre Sınıflandırma

Duygu analizinin ilk çalışmaları duygusal kutupsallık (sentimental polarity) veya duygu polaritesi olarak geçmekte olup verilen metni olumlu (positive) veya olumsuz (negative) olarak sınıflandırmayı amaçlamaktadır [30]. Duygu yoğunluğu ise bir metin ne kadar negatif (güçlü, zayıf) veya ne kadar pozitif (güçlü, zayıf) anlamına gelmektedir.

Duygu polaritesinin belirlenmesi için bir belgenin veya metnin önceden belirlenen hangi sınıfa ait olduğuna dair kuralların oluşturulması gerekmektedir. İnternette yer alan metin biçimindeki film eleştirilerinin olumlu ve olumsuz olarak iki sınıfa ayrılmasını sağlayan polarite tespiti bu konuya örnek olarak verilebilir ve bu alanda birçok çalışmanın yer aldığı görülmektedir.

Kelimelerin kutupsallığını veya polaritelerini göstermek için anlamsal yönelim adı verilen bir kavram kullanılmaktadır. Anlamsal yönelim pozitif ve negatif olarak iki boyutta değişmektedir. Kelimelerin anlamsal yönelimini hesaplamak için birkaç yöntem vardır. Genellikle ilişkilerinin anlamsal yönelim olan SO (semantic orientation), olumlu veya negatif kelimelerin toplam değeri olarak eksi, negatif kelimelerin toplam değeri olarak hesaplanır:

$$SO = \sum_{pkelime \in Pkelimeler} A(kelime; pkelime) - \sum_{nkelime \in Nkelimeler} A(kelime; nkelime) \quad (2.2) \quad [31]$$

- **A(kelime, pkelime)** incelenmiş kelimelerde pozitif olan kelimelerin adetini ifade etmektedir.

- **A(kelime, nkelime)** incelenmiş kelimelerde negatif olan kelimelerin adetini ifade etmektedir.

- **pkelime**  $\in$  **Pkelimeler** pozitif kelimeler kümesinde olan her pozitif kelime ifade edilir.

- **nkelime**  $\in$  **Nkelimeler** negatif kelimeler kümesinde olan her negatif kelime ifade edilir

Toplam pozitif bir değer ise kelime pozitif olarak, toplam negatif bir değer ise kelime negatif olarak kabul edilir. Toplamın mutlak değeri, polaritenin yoğunluğunu gösterir.

Duygular ve öznellik içeriğe çok duyarlıdır ve etki alanına bağımlıdır. Alan bağımlılığı, kısmen kelime değişimlerinin bir sonucudur. Örneğin aynı ifade, farklı alanlarda farklı duygular gösterebilir. Üstelik internette herkes kendi kelime dağarcığını kullanmaktadır. Bu da aynı etki alanında olsa bile duygu analizi sürecini zorlaştırmaktadır. Kritik cümleler için doğru ağırlık veya puanı vermek de çok zordur. **Örneğin** birçok olumlu kelimedede sonra bir film yorumunda "... ona rağmen filmin sonundan önce sinemadan çıktım." ifadesi yer aldığında yorumun anlamı değişmektedir.

Bu örnekler, çeşitli kullanıcılar tarafından yazılan bir metinde ideal bir duygu notasyonu örneğine ulaşmanın hala çok zor olduğunu göstermektedir. Çünkü kullanıcılar tarafından yazılan metinler herhangi bir kurala uyum göstermemekte ve olası tüm durumları öngörmek mümkün olmamaktadır.

## 2.2 Duygu Analizi:

Bu bölümde duygu analiz ile ilgili bilgilendirme yapılacaktır

### 2.2.1 Duygu notasyonunun karmaşıklığı:

Duygu gösteriminin karmaşıklığını göstermek için aşağıdaki sinema eleştiri cümleleri ele alınabilir:

1. **Harika bir yaşam (Wonderful Life).**
2. **Tüm zamanların en sevdiğim filmi !**
3. **Bundan nefret eden, gerçek yaşamda sadece iki kişi gördüm.**

Görüldüğü gibi, eleştiri karşıt kutuplara sahip üç cümleden oluşmaktadır. İlk cümle, "**Harika bir Yaşam**" ifadesi filmin adıdır. İkinci cümle "Tüm zamanların en sevdiğim filmi" ifadesi olumlu bir ifadedir. fakat üçünü cümle : "**Bundan nefret eden, gerçek yaşamda sadece 2 kişi gördüm**" ifadesi istatistiksel bir çalışma yapıldığında olumsuz olarak değerlendirilirken gerçekte olumlu bir durumu ifade etmektedir.

Duygular kelimelerle genellikle çok karmaşık bir şekilde ifade edilebilmektedir. Duygu analizi gerçekleştirmek için cümleler , kelimeler şeklinde ayrı ayrı düşünüldüğünde bazen belgenin tanımlaması zorlaşmaktadır.

### 2.2.2 Öznel cümle ile nesnellik cümle arasındaki karşılaştırma:

Bir metnin nesnelliği, kullanılan ifadelere ve kelimelere bağlıdır. Genel olarak nesnel cümleler, tarafsızlık, nötrlük, ilgisizlik veya kişiliksizlik kelimeleri ile ifade edilir.

Nesnel bireyin, yargı veya karar anında, bir tür evrensellik elde etmek için kendine özgü fikirleri, inançları veya kişisel tercihleri terk etmesi beklenir.

Öznellik cümleleri ise , yazarın kişisel fikirlerinin ifade etmesine izin verdiği için nesnellığe karşıdır. Bir kişiye veya belirli bir bakış açısına bağlı olan şey öznedir. Bir ifade bireyin tutkularını, önyargılarını ve kişisel seçimlerini yansıtıyorsa, yargı öznedir ve önyargıyla eş anlamlıdır.

### **2.2.3 Duygu analizine farklı yaklaşımlar**

Duygu analizi yaparken farklı yaklaşımları veya yöntemlerin aracıyla yapılır.

#### **2.2.3.1 N-gramların sınıflandırmadaki rolü**

Kelimelerin bir metin içerisindeki konumu metnin duygu düzeyini veya öznelliğini değiştirebilir. Pang ve arkadaşları [32] tarafından gerçekleştirilen çalışmada sinema eleştirilerine ait metin verisinde, unigram yöntemi kullanarak gerçekleştirilen sınıflandırmanın, iki gram yöntemi kullanarak elde edilen sonuçlardan daha iyi olduğu görülmektedir. Ancak, farklı verilerde ve durumlarda iki-gram ve trigram yöntemlerinin daha iyi polarite sınıflandırma araçları olduğu da görülebilmektedir.

Öznellik tespitinde yapılan çalışmalarda sıfatların varlığı ile cümle öznelliği arasında güçlü bir ilişki olduğu görülmektedir [33]. Bu nedenle bu tez çalışmasında kullanılan sözlük içerisinde birçok sıfat yer almaktadır.

#### **2.2.3.2 Olumsuzluk işleme**

Türkçe dilinde olumsuzluk ifade etmek için çok farklı kelimeler ve ekler kullanılabilir. Bu durumda da olumsuzluk tespitini zorlaştıran bir etkidir. Örneğin, “güzel değil” ifadesindeki olumsuzluğu ulaşabilmek için bi-gram yöntemi kullanılabilir. “gel-mi-yorum” kelimesinde ise, kelimenin köküne kadar inmek gerekmektedir. Gerçekleştirilen uygulamada R programlama ortamında Türkçe dil desteği sınırlı olduğu ve mevcut kütüphane yüksek verimli olmadığı için zorluklar yaşanmıştır. Bu zorlukları gidermek için algoritma tasarlanırken me, ma, mi mü, gibi negatif ekleri üzerine dayanarak yapılmalıdır.

#### **2.2.4 Tez çalışmasında kullanılan Sözlükler:**

Tez çalışması kapsamında duygu analizi için oluşturulan sözlükler farklı sözlüklerin bir araya getirilmesi ile ortaya çıkmıştır.

R sınıflarda kullandığımız sözlüklerin orijinalı İngilizce dilindedir ve farklı

arařtırmalarda kullanılmıřtır. Bu szlgn Trke diline evrimi tarafımızca gerekleřtirilmiřtir.

- “znellik” olan birinci szlk <https://mpqa.cs.pitt.edu/> web sitesinde aık kaynak olarak yayınlanmıřtır ve [https://mpqa.cs.pitt.edu/lexicons/subj\\_sense\\_annotations/](https://mpqa.cs.pitt.edu/lexicons/subj_sense_annotations/) sitesinden indirilebilir. University of pittsburgh’deki “School of Computing and Information” blm tarafından oluřturulup zerine farklı arařtırma yapılmıřtır. Bu szlkte toplam 24 774 kelime yer almaktadır.
- “duygular” olan ikinci szlk de [https://mpqa.cs.pitt.edu/lexicons/subj\\_sense\\_annotations/](https://mpqa.cs.pitt.edu/lexicons/subj_sense_annotations/) web sitesinde yayınlanmıřtır. Bu szlkte toplam 2 477 kelimeler vardır.

Bu iki tane szlk, **znellik duygu notasyonu (Subjectivity Sense Annotations)**[34], **rn tartıřma verileri (Product Debate Data)** [35], **Siyasi Tartıřma Verileri (Political Debate Data)** [36], **Grř Bildirici Sistemi (OpinionFinder System)** [37] gibi farklı farklı alıřmalarda kullanılmıřtır.

- nc szlk olan afinn.txt ise sentitrk adlı Trke szlkten alınmıřtır. Bu szlk, **Dehkharghani ve arkadaşları** [38] tarafından oluřturulup *Sentiment Analysis in Turkish: Resources and Techniques* konulu Doktora tezinde kullanılmıřtır. **Wordnet Gloss, SentiwordNet ve SenticNet** olarak 3 ayrı szlkten yararlanarak oluřturulmuřtur. Afinn szlgnde 24 682 kelime vardır.
- Drdnc szlk olan SentiTurkNet szlg, Dehkharghani ve arkadaşları tarafınan oluřturulmuřtur . İlk kapsamlı Trk kutupluluk szlk kaynađı olan Trke WordNet’ten yararlanarak,  kutupluluk puanı atanmıř SentiTurkNet, pozitiflik, olumsuzluk ve nesnellik/ntr (tarafsızlık) seviyelerini gsterir.

İngilizce'deki duygu analizi alıřmalarında , sıka kullanılan  tr szlk bulunmaktadır, bu szlklerin  de unigramlara, yani tek kelimelere dayanmaktadır. Btn bu szlkler, duygular veri kmesinde tablolanmıřtır ve “Tidytext” ktphanesinde `get_sentiments()` aracıyla elde edilir.

- **AFINN**: -5 ile 5 arasında deđiřen puanlara sahip kelimeleri iermektedir. Negatif deđerler olumsuz duyguları, pozitif deđerler ise pozitif duygular temsil eder.

**Kod:**

```
get_sentiments("afinn")
```

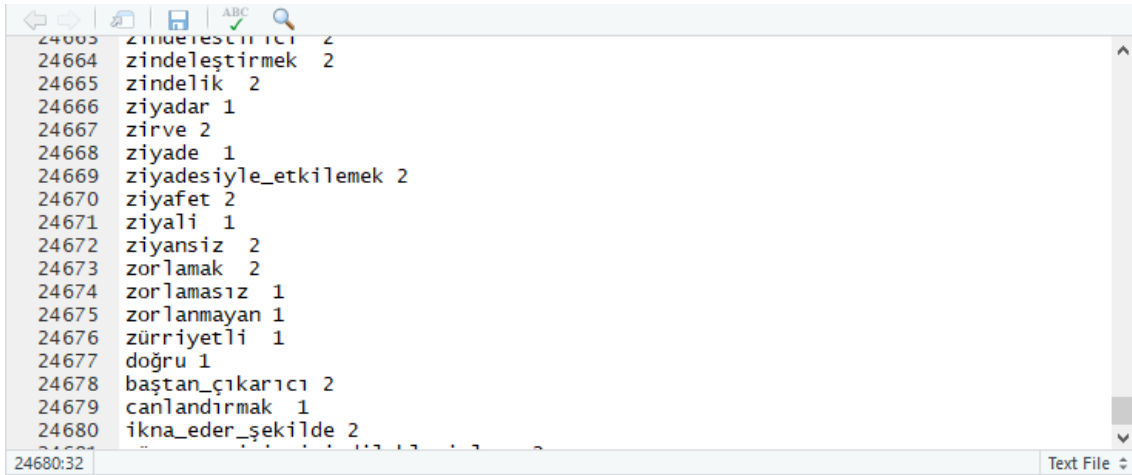
### Cıktı:

**Tablo2.2** Afın sözlüğü örneđi

##	Word	Score
##	<chr>	<int>
## 1	abandon	-2
## 2	abandoned	-2
## 3	abandons	-2
## 4	abducted	-2
## 5	abduction	-2
## 6	abductions	-2
## 7	abhor	-3
## 8	abhorred	-3
## 9	abhorrent	-3
## 10	abhors	-3
## # ...	with 2,466 more rows	

**Afın sözlüğü** 2009 ve 2011 yılları arasında Finn Årup Nielsen tarafından -5 (negatif) ve +5 (pozitif) arasında bir tamsayı ile ve elle derecelendirilen İngilizce terimler listesidir.

"AFİNN" ile aynı özelliklere sahip sözlüğü "Afinn" sözlüğüdür ve bu şekildedir:



**Şekil 2.2** "Afinn" sözlüğü örneđi

- **BİNG:** Bing Liu and collaborators tarafından oluşturulmuştur, kelimeleri ikili bir biçimde pozitif ve negatif kategorilere sınıflandırır.

**Kod:**

```
get_sentiments("bing")
```

**çıktı:**

**Tablo2.3** Bing sözlüğü örneği.

##	Word	Sentiment
##	<chr>	<chr>
## 1	2-faced	Negative
## 2	2-faces	Negative
## 3	a+	Positive
## 4	abnormal	Negative
## 5	abolish	Negative
## 6	abominable	Negative
## 7	abominably	Negative
## 8	abominate	Negative
## 9	abomination	Negative
## 10	abort	Negative
## #	... with 6,778 more rows	

"bing" ile aynı özelliklere sahip sözlüğümüz "öznellik" sözlüğüdür .

Bu [https://mpqa.cs.pitt.edu/lexicons/subj\\_sense\\_annotations/](https://mpqa.cs.pitt.edu/lexicons/subj_sense_annotations/) siteden yayınlanmış bir çalışmada olan subjectivity sözlüğünün tercümesi yaparak elde edilmiştir.

Sözlük bu şekilde görülmektedir:

```
17 nefret_et, strongsubj, negative
18 nefret_etmek, strongsubj, negative
19 iğren, strongsubj, negative
20 iğrenmek, strongsubj, negative
21 tiksinti, strongsubj, negative
22 iğrenç, strongsubj, negative
23 nefret_uyandıran, strongsubj, negative
24 ahlak_dışı, strongsubj, negative
25 hoşlanmama, strongsubj, negative
26 hoşlanmamak, strongsubj, negative
27 aşırı_derecede_nefret, strongsubj, negative
28 aşırı_derecede_nefret_etme, strongsubj, negative
29 aşırı_derecede_nefret_etmek, strongsubj, negative
30 kurallara_uy, strongsubj, positive
31 kurallara_uymak, strongsubj, positive
32 itaat, strongsubj, positive
33 itaat_et, strongsubj, positive
34 itaat_etmek, strongsubj, positive
```

Şekil 2.3 "öznellik" sözlüğü örneği

- **NRC**: Saif Mohammad and Peter Turney. tarafından oluşturulmuştur, kelimeleri pozitif, negatif, öfke, beklenti, tiksinti, korku, sevinç, üzüntü, şaşkınlık ve güven kategorilerine sınıflandırır.

**Kod:**

```
get_sentiments("nrc")
```

**Çıktı:**

Tablo2.4 NRC sözlüğü örneği.

##	Word	sentiment
## 1	abacus	Trust
## 2	abandon	Fear
## 3	abandon	Negative
## 4	abandon	Sadness
## 5	abandoned	Anger
## 6	abandoned	Fear
## 7	abandoned	Negative
## 8	abandoned	Sadness
## 9	abandonment	Anger
## 10	abandonment	Fear

"NRC" ile aynı özelliklere sahip sözlüğü "**Duygular**" sözlüğüdür ve bu şekildedir:

17	acımasızca,korku
18	acımasızlık,korku
19	acınacak,üzüntü
20	acınaklı,üzüntü
21	acı_ver,üzüntü
22	acı_vermek,üzüntü
23	aç_göz,öfke
24	aç_gözlü,öfke
25	afallamış,sürpriz
26	afallama,sürpriz
27	afallamak,sürpriz
28	afallaşma,sürpriz
29	afallaşmak,sürpriz
30	afallaştırma,sürpriz
31	afallaştırmak,sürpriz
32	afallatma,sürpriz
33	afallatmak,sürpriz
34	afallatan,sürpriz

Şekil 2.4 "Duygular" sözlüğü örneği.

- **SentiTurkNet**: 3 tane farklı sözlükten yararlanarak oluşturulmuştur.

İlki: ingilizce “**wordNet**” içinde yaklaşık 117 000 cümle vardır ve bu sözlük bir küme olarak eşanlı terimlerini gruplandırır.

İkincisi: “**sentiWordNet**”, ingilizcedeki duyarlık analizi görevlerini desteklemek amacıyla oluşturulmuştur. Neg, pos, nötr olarak her kelimenin 3 farklı kutuplu puanı verir.

Üçüncüsü: “**senticNet**” her kelimenin hoşluğuna, dikkatine, uyarlığına , kabiliyetine ve genel polarite gücüne göre sayısal değer verir.

1	synonyms	Turkish Gloss	Polarity	Label	POS tag	neg value	obj value	pos value	Eng Synor	English G	SWNpos	SWNneg
2	ayva reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	o	n		0,06	0,872	0,068	jam#1	preserve	0	0
3	gül reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	o	n		0,06	0,872	0,068	jam#1	preserve	0	0
4	incir reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	o	n		0,06	0,872	0,068	jam#1	preserve	0	0
5	ıştah	Yemek yeme isteği.	o	n		0,06	0,872	0,068				
6	ıştahlı	İştahı olan, boğazlı	p	a		0,06	0,462	0,478				
7	ıştahsız	Yemek yeme isteği olmayan, boğazsız	n	a		0,48	0,452	0,068				
8	ıştahsızlık	İştahsız olma durumu	n	n		0,512	0,42	0,068				
9	kayısı reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	o	n		0,06	0,872	0,068	jam#1	preserve	0	0
10	kayısı suyu	meyve suyu	o	n		0,06	0,872	0,068	fruit_juic	drink pro	0	0
11	kısayol	bilgisayarda bir dosyadan diğer dosyaya ulaşmak	o	n		0,06	0,872	0,068				
12	mikrobiyolojik	mikroplar la uğraşan bilim	o	a		0,06	0,872	0,068				
13	minibüs	10, 12 kişilik küçük otobüs	o	n		0,06	0,872	0,068				
14	müsteşarlık	Müsteşarın görevi veya makamı	o	n		0,06	0,872	0,068				

Şekil 2.5 “SentiTurkNet” sözlüğü örneği

### 2.3. Geliştirilen Duygu Analizi Yazılım Sisteminin Mimarisi

Tez çalışması kapsamında geliştirilen duygu analizi sisteminin ana görevleri, duygu analizin sayesinde, istenilen metni seçilen sözlüğe göre sınıflandırabilmektir. Rstudio'da

mevcut olan kütüphaneler (**RTextTools**, **plyr**, **e1071**, **wordcloud**, **RColorBrewer**, **languageR**, **Snowball**, **graphics**,... ) ve *Naive Bayes* gibi algoritmalar kullanarak metinlerin duygu özellikleri vurgulanarak, her metnin bir grafik gösterimi elde edilebilmektedir.

### 2.3.1 Duygu analizi aşamaları

Geliştirilen duygu analizi yazılım sistemi üç aşamalı çalışarak analiz işlemini gerçekleştirmektedir.

#### 2.3.1.1 Ön işleme aşaması

Önişleme aşamasının dikkatli biçimde gerçekleştirilmesi sistemin başarısı için çok önemlidir. Çünkü eğer bu aşama doğru bir biçimde tamamlanırsa, takip eden diğer kısımlar da başarılı bir şekilde tamamlanabilecektir. Bu aşamada metnin temizlenmesi ve özelliklerinin ortaya çıkarılması için kullanıma hazır hale getirmesi sağlanmaktadır.

**Şekil 2.6**'te önişleme aşamasının işlem adımları görülmektedir. Metin üzerine yapılacak değişiklikler şu şekildedir:

- ÜÇ HARFTEN DAHA KISA KELİMELERİ KALDIRILIR,
- Büyük harfleri küçük harflere çevirilir (tolower=true),
- Rakamlar kaldırılır,
- Noktalama işaretleri kaldırılır,
- Kelimeler arasında ekstra bırakılmış boşluklar kaldırılır,
- Seyrek terimleri kaldırılır,
- Durdurma kelimeleri kaldırılır,
- Stemming/lemming yapılır,
- Belge terim matrisi (document term matrix) dönüştürülür,

Bu bölümün sonunda dönüştürülmüş bir metin elde edilmekte ve (TİDY TEXT) veya işlenmiş metin adı verilmektedir. Bu aşamanın amacı metinleri almak ve bir sonraki katman tarafından işlenebilecekleri şekilde biçimlendirmektir.

Düzenli veri ilkelerini kullanarak birçok metin incelemesi görevi daha kolay, etkili ve yaygın kullanılan araçlarla tutarlı hale getirilir. Önişleme aşaması sonrasında elde edilen düzenli metin ile, metnin özelliklerinin kolayca manipüle edilmesi, özetlenmesi,

görselleştirilmesi ve doğal dil işleme kullanarak etkin iş akışlarına entegre edilmesine olanak mümkün olmaktadır.

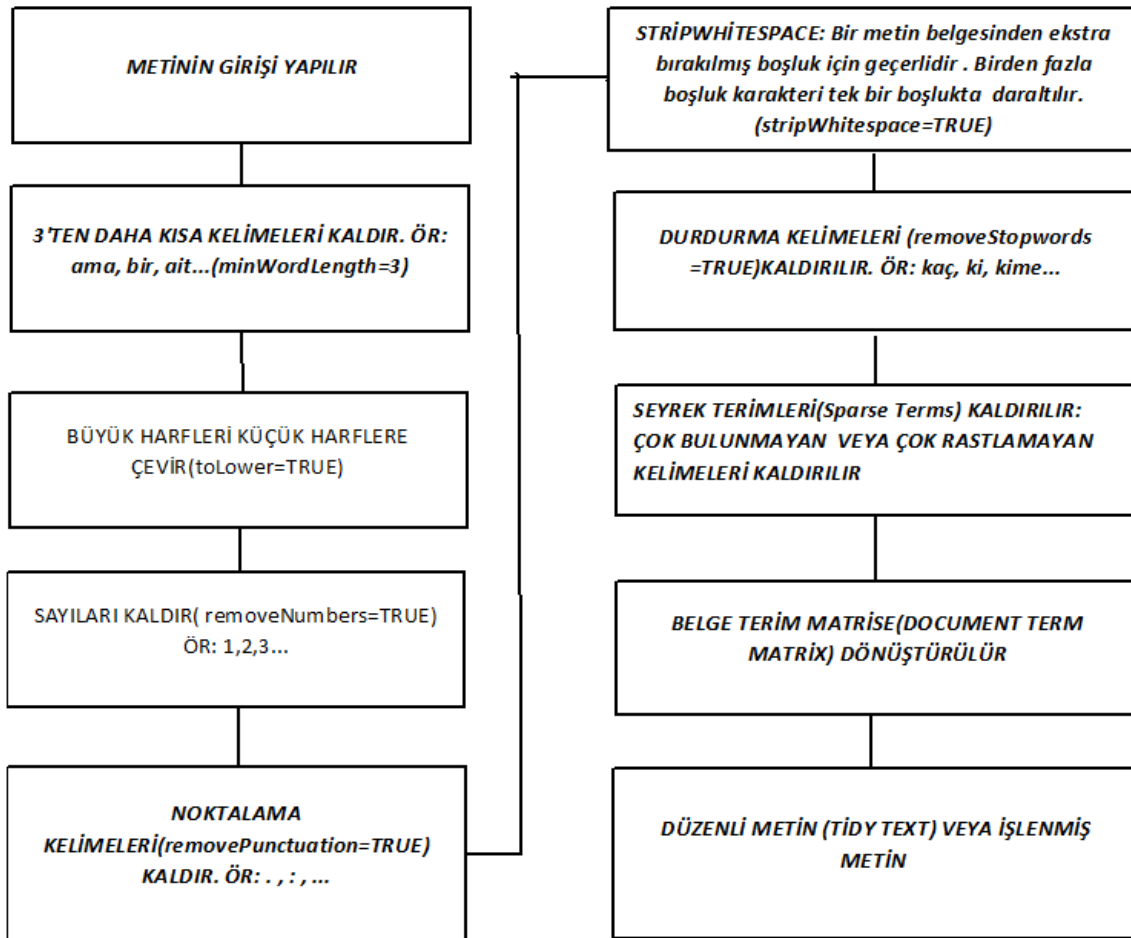
Düzenli metin verileri belirli bir yapıya sahiptir:

Her değişken bir sütun

Her gözlem bir satır

Her gözlemsel birim, bir tablo

Düzenli metin formatını, one-token-per-row ile bir tablo olarak tanımlanır. Bir token, analiz için kullanılacak sözcük (terim) gibi anlamlı bir metin birimidir. Tokenization(dizgi parçalama) ise metni token halinde bölme işlemidir. Düzenli metin incelemesi için, her bir satırda saklanan token genellikle tek bir kelimedir. Ancak n-gram, cümle veya paragraf da olabilir.



Şekil 2.6 Ön işleme aşamasının adımları

### 2.3.1.2 Ön işleme örneği

Bu bölümde metin örneği olarak ön işleme adımları uygulanacaktır.

#### 2.3.1.2.1. Verilerin temizlenmesi ve bölünmesi

Ön işleme adımı verileri temizler ve yapılandırır. Asıl amaç, anlamlı bilgileri çıkarmak için metinlerin içeriğini işlemektir. İkinci amaç, metinlerinde görünen her kelime için yapılandırılmış bir temsiline dönüştürmektir. Anahtar kelimeler ile belgeler arasındaki ilişkiler, kelimelerin görünme sıklığı ile incelenir.

Hazırlık aşaması dört ana bileşene sahiptir: **morfolojik analiz**, **sözdizimsel analiz**, **sözcüksel analiz** ve **boyut küçültme veya indirgeni**.

Morfolojik analizi, kelimeler dizgeciklere (Tokenization) ayırarak, ardından "kök" veya "lematizasyon" ile normalleştirilmesini sınırlandırır. Bu adımda, cümleleri birbirinden ve kelimelerden ayırarak metni otomatik olarak keser. Bu analizin her bir aşaması aşağıdaki örnek metin ile açıklanmaktadır:

**"1- Bir veba bu sesiymiş giderken bıraktığın... Kim bilir bende**

**2- hangi yaralarım öptün bana ait sandığın? Bu ayrılık ne**

**3- iyileştirdi ne öldürdü... Beni ağlattı» seni güldürdü. Şimdi**

**4- benim anı diye sakladıklarım, senin çoktan unuttukların...**

**5- Bak, sabah olsa bile içim hâlâ gece. Bilmiyorum bu acı kaçınıcı.**

**6- Senin olmadan sana solmak yok mu? İşte o çok acı.**

**7- Gözümdeki damlalar, içimde kuduran nehirlerin zerrecikleridir.**

**8- Gidişin öğretti bana; bazı aşklar aşka ihanettir!**

**9- Ben ağlarım içten ve sessiz; sen, sana suskun sözlerimi gözlerimden" [39]**

**Tablo2.5** Korpus oluşturma

Sıra No	Metin
1	Bir veba bu sesiymiş giderken bıraktığın... Kim bilir bende
2	hangi yaralarım öptün bana ait sandığın? Bu ayrılık ne
3	iyileştirdi ne öldürdü... Beni ağlattı» seni güldürdü. Şimdi
4	benim anı diye sakladıklarım, senin çoktan unuttukların...

5	Bak, sabah olsa bile içim hâlâ gece. Bilmiyorum bu acı kaçınıcı.
6	Senin olmadan sana solmak yok mu? İşte o çok acı.
7	Gözümdeki damlalar, içimde kuduran nehirlerin zerrecikleridir.
8	Gidişin öğretti bana; bazı aşklar aşka ihanettir!
9	Ben ağlarım içten ve sessiz; sen, sana suskun sözlerimi gözlerimden

**Tablo2.5'**te görülen dokuz cümle noktalama işaretleri , büyük harf ve rakam içeren bir korpus örneğidir. Ön işleme adımları, noktalama işaretleri ve küçük harfli kelimeler olmadan bu korpusu basitleştirerek kelimeleri analiz için uygun hale getirecektir.

### 2.3.1.2.2. Morfolojik analiz

İlk adımda morfolojik analiz noktalama işaretlerini kaldırılması ve metni küçük harfe dönüştürülmesini sağlar. "-", "'", "&" gibi noktalama işaretleri kaldırılmıştır. Morfolojik analiz sonucu Tablo 2.6'de görülmektedir.

**Tablo2.6** Noktalama işaretlerini kaldırılması

Sıra No	Metin
1	bir veba bu sesiymiş giderken bıraktığın kim bilir bende
2	hangi yaralarım öptün bana ait sandığın bu ayrılık ne
3	iyileştirdi ne öldürdü beni ağlattı seni güldürdü şimdi
4	benim anı diye sakladıklarım senin çoktan unuttukların
5	bak sabah olsa bile içim hâlâ gece bilmiyorum bu acı kaçınıcı
6	senin olmadan sana solmak yok mu işte o çok acı
7	gözümdeki damlalar içimde kuduran nehirlerin zerrecikleridir
8	gidişin öğretti bana bazı aşklar aşka ihanettir
9	ben ağlarım içten ve sessiz sen sana suskun sözlerimi gözlerimden

Bu örnekte noktalama işaretlerinin, sayılar ve vurguların kaybolduğu, büyük harflerin küçük harflere dönüştürüldüğü görülebilmektedir.

Bir sonraki adımda ise , durağan(stoplist) listesinde yer alan kelimelerin kaldırılması işlemi gerçekleştirilmektedir.

Durağan kelime listesinde toplam 188 kelime yer almaktadır. **Tablo 2.7'**da bu listeden örnekler görülmektedir.

**Tablo2.7** Örnek durağan kelime listesi

Acaba	artık	bana	Belki	beş	birçok	birkaç,	biz
altı	asla	bazı	Ben	bile	birçokları	birkaçı	bize
Ama	aslında	bazıları	Beni	bir	biri	birşey	bizi
ancak	az	bazısı	Benim	birçoğu	birisi	birşeyi	Bizim

Bilgilendirici olmayan kelimeler listesini ("sen", "biz", "ama") gibi otomatik olarak silebilmek için R Studio'da snowball veya Rtemis gibi bazen kütüphanelerde bir "durma kelimeler" listesi önceden tanımlanmıştır.

Durma listesindeki kelimeler, zarfların veya sıfatların aksine metinlerin anlamı veya eleştirilerde açıklanan görüş üzerinde zayıf bir etkiye, değer yargısına sahiptir.

**Tablo2.8** StopList'deki kelimelerden arındırılmış korpus

Sıra No	Metin
1	veba sesiymiş giderken bıraktığın bilir
2	yaralarım öptün ait sandığın ayrılık
3	iyileştirdi öldürdü ağlattı güldürdü
4	anı diye sakladıklarım çoktan unuttukların
5	bak sabah olsa içim hâlâ gece bilmiyorum acı kaçınıcı
6	olmadan solmak yok işte o çok acı
7	Gözümdeki damlalar içimde kuduran nehirlerin zerrecikleridir
8	gidişin öğretti aşklar aşka ihanettir
9	ağlarım içten sessiz suskun sözlerimi gözlerimden

**Tablo 2.9** örnek metinden kaldırmış kelimeleri altı çizgili olarak göstermektedir.

**Tablo2.9** Örnek metinde yer alan Stoplist kelimeleri

Sıra No	Metin
1	<u>bir</u> veba <u>bu</u> sesiymiş giderken bıraktığın <u>kim</u> bilir <u>bende</u>
2	<u>hangi</u> yaralarım öptün <u>bana</u> ait sandığın <u>bu</u> ayrılık <u>ne</u>
3	iyileştirdi <u>ne</u> öldürdü <u>beni</u> ağlattı <u>seni</u> güldürdü <u>şimdi</u>
4	<u>benim</u> anı diye sakladıklarım <u>senin</u> çoktan unuttukların
5	bak sabah olsa <u>bile</u> içim hâlâ gece bilmiyorum <u>bu</u> acı kaçınıcı
6	<u>senin</u> olmadan <u>sana</u> solmak yok <u>mu</u> işte o çok acı
7	gözümdeki damlalar içimde kuduran nehirlerin zerrecikleridir
8	gidişin öğretti <u>bana</u> <u>bazı</u> aşklar aşka ihanettir
9	<u>ben</u> ağlarım içten <u>ve</u> sessiz <u>sen</u> <u>sana</u> suskun sözlerimi gözlerimden

Daha sonra, bir kelimenin formunu türevlendirme, ön ekler ve son ekler olmadan kökenine indirgemek için bir algoritma uygulanır , örneğin "alışkanlık" kelimesi alış kelimesine indirir. Aynı zamanda şimdiki / geçmiş zaman ve tekil / çoğul gibi ekler de kaldırılır. R Studio içerisinde bu işlemin otomatik yapılmasını sağlayan bir kütüphane yer almadığı için, bu işemi gerçekleştirecek bir yazılım hazırlanmıştır.

**Tablo2.10** İleri morfolojik analiz öncesi Korpus

Sıra No	Metin
1	<b>v</b> eba <b>s</b> esiymiş <b>g</b> iderken <b>b</b> ıraktıkğın <b>b</b> ilir
2	<b>y</b> aralarım <b>ö</b> ptün <b>a</b> it <b>s</b> andıkğın <b>a</b> yrılık
3	<b>i</b> yileştirdi <b>ö</b> ldürdü <b>a</b> ğlattı <b>a</b> ğlattı <b>g</b> üldürdü
4	<b>a</b> nı <b>d</b> iye <b>s</b> akladıklarım <b>ç</b> oktan <b>u</b> nuttukların
5	<b>b</b> ak <b>s</b> abah olsa <b>i</b> çim <b>h</b> âlâ <b>g</b> ece <b>g</b> eç <b>b</b> ilmiyorum <b>a</b> cı_ <b>a</b> ç <b>k</b> açınıcı
6	<b>o</b> lmadan <b>s</b> olmak <b>y</b> ok <b>i</b> şte o <b>ç</b> ok <b>a</b> cı_ <b>a</b> ç
7	<b>g</b> özümdeki <b>d</b> amlalar <b>i</b> çimde <b>k</b> uduran <b>n</b> ehirlerin <b>z</b> errecikleridir
8	<b>g</b> idişin <b>ö</b> ğretti <b>a</b> şklar <b>a</b> şka <b>i</b> hanettir
9	<b>a</b> ğlarım/ <b>i</b> çten <b>s</b> essiz <b>s</b> uskun <b>s</b> özlerimi <b>g</b> özlerimden

Sözdizimsel analiz her cümlenin farklı bölümleri arasındaki bağlantı yapısını belirlemek için kullanılır. Daha gelişmiş bir formda hem kelimeyi çevreleyen bağlamı, hem de ek gramer bilgilerini kullanan *lemmatization* olarak da bilinir.

Kelimededen son ekler ve farklı değişkenler silinir. Örneğin "gitmek" fiili farklı biçimlerde bulunabilmektedir: "gidiyorum", "gittim", "gittiler" gibi. Analiz listedeki her kelime tek bir sembole indirgenir. Bu adım, bazı hataları kontrol edilmesini sağlar. Ancak, kalitesi bir yazılımdan diğerine ve bir dilden diğerine bağlıdır.

**Tablo2.11** İleri morfolojik analiz sonrası Korpus

1	veba ses gider bıraktık_ bırak** bil*
2	yara öp ait sandık ayrılık
3	iyileştir öldür ağlat ağlattı*** güldür
4	anı di sakladık çok unuttuk_unut****
5	bak sabah ol iç hâlâ gece geç bilmiyor acı_ aç kaçınıcı
6	olma solmak yok işte çok acı_ aç
7	göz damla içi kudura nehir zerrecik
8	gidiş öğret aşk aşka ihanet
9	ağla* iç sessiz susk söz göz

\* Bunu yakalanmadı

\*\* bırak yerine bıraktık bulundu

\*\*\* ađlat yerine ađlattı bulundu

\*\*\*\*unut yerine unuttuk bulundu...

### 2.3.1.2.3 Boyut indirgeme

Veri hazırlamanın son bileşeni, her belgede kullanılan terimlerin sıklığının temsil etmektir. Bu aşamada öncelikle belge terim matrisi (DTM) gibi verilerin yapılandırılmış bir gösteriminin oluşturulması sağlanır. Her satır bir dokümanı temsil eder ve her sütun görünen terimleri gösterir. Şekil 2.7’da örnek bir kopus için kelimelerin dağılımı görülmektedir.

	Global	occ.	Global %
bir	236	4.6283585	
ve	114	2.2357325	
çok	111	2.1768974	
kitap	101	1.9807805	
bu	76	1.4904883	
ama	57	1.1178662	
da	56	1.0982546	
kadar	45	0.8825260	
için	44	0.8629143	
de	42	0.8236909	
kitabı	42	0.8236909	
gibi	38	0.7452442	
ben	27	0.5295156	

Şekil 2.7 Korpustaki kelimelerinin dağılımı

Bu kelimelerin korpustaki frekansları analiz etmek için, veri hazırlama aşamasının sonunda korunan her kelimenin bir olasılık durum tablosu oluşturulur. Bu tablonun içeriği, kelimenin tekrar değeridir. Bu tablo, Belge Terim Matrisi (DTM) olarak adlandırılır. DTM tablosu, bir kelimenin bir metinde kaç kez kullanıldığını göstermektedir. Kelimelerin büyük çoğunluğu sadece birkaç metinde veya belgede görünür. Sonuç olarak, bir DTM kullanılmayan kelimelerin çok sayıda sıfır değerine sahiptir.

Docs	ama	bir	bu	çok	da	de	için	kadar	kitap	ve
x100	0	4	2	1	0	0	0	2	5	0
x116	0	5	0	1	1	0	1	0	1	3
x123	0	4	1	1	1	5	0	0	1	2
x26	3	8	1	4	3	3	1	0	1	4
x32	2	5	3	0	1	0	0	0	0	0
x40	0	2	1	0	0	0	0	1	1	2
x66	0	4	0	0	0	1	0	2	2	1
x71	0	4	2	1	1	0	0	0	0	2
x87	0	4	2	6	0	0	1	0	1	5
x92	0	5	0	1	2	1	0	0	0	1

**Sekil 2.8:** Örnek bir DTM tablosu

Yukardaki tabloya göre metnimizdeki 100. Döküman'da veya cümlede 4 tane "bir" kelimesi, 2 tane "bu kelimesi", 2 tane "kadar" kelimesi ve 5 tane "kitap" kelimesi mevcuttur.

Hazırlık aşamaları, incelenen metnin türüne bakılmaksızın benzer bir şekilde gerçekleştirilir. Öte yandan, tıp alanı gibi belirli bir alanda yer alan terminolojik kelimeler kullanılıyorsa, araçlar ve sözlük bu alana göre uyarlanmalıdır. Tüm araçlar tüm diller için uyumlu değildir, genel bir çözüm üretmemektedir. Bu nedenle farklı dillere ait analiz çalışmalarında veri yapılandırması çalışmaları gerektirmektedir.

### 2.3.1.3 Duygu analizi aşaması

Bu bölüm metin sınıflandırma modülüdür. Burada hangi sözlüğün kullanılacağına ve Bayes algoritmasını nasıl hesap yapacağına burada karar verilmektedir. Metin analizi yapmak için metinde kelimelerin oluşumunu ve dağılımını kullanılır. Genel olarak duygu analizinde kullanılan bir yaklaşım sunulacaktır. Bilinen algoritmalar arasında araştırmalara göre en iyi sonuç veren Bayes sınıflandırıcı seçilmiştir. Bu çalışmada Bayes sınıflandırıcı skorları belirlemek için kullanılmıştır ve bu bölüm farklı farklı yaklaşımlarla üç tane R sınıfına ayrılmıştır.

#### **Duygu\_sınıflandır.R**

Bu sınıfta "Duygular.csv" sözlüğünü kullanılmıştır. Metnin sınıflandırması öfke, nefret, korku, mutluluk, üzüntü, sürpriz olarak adlandırılan altı duyguya göre yapılmaktadır.

**duyguSınıflandır()** fonksiyonun aşamaları aşağıdaki gibidir:

**1-** Düzenli metin oluşturmak için metnimiz `matris_olustur()` fonksiyonuna gönderilir ve çıktı olarak matris yapısında olan bir metin gövdesi elde edilir.



```
Console Terminal x
~/Metinmadenciligi/
[1] "KELİME: ađırlık KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: bađlı KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: becerikli KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bekle KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bıkmak KAT: nefret SKOR: 4.95582705760126"
[1] "KELİME: bravo KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bu KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: bunalımlı KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: büyüle KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: çekememezlik KAT: öfke SKOR: 5.97888576490112"
[1] "KELİME: çekicilik KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: iç KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: in KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: in KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: yazık KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: zor KAT: üzüntü SKOR: 6.38687931936265"
ÖFKE NEFRET KORKU MUTLULUK ÜZÜNTÜ
[1,] "7.81480342948936" "7.81480342948936" "7.81480342948936" "48.4050372901726" "39.749200
0263026"
SÜRPRİZ EN_UYGUN_DUYGU
[1,] "3.31499375915909" "mutluluk"
> |
```

Şekil 2.10 `duyguSınıflandır()` Fonksiyonun ekran çıktısı

### Polarite\_sınıflandır.R

Bu sınıfta "öznellik.csv" sözlüğünü kullanılır, burada sınıflandırma kutupluluđa (Olumlu veya Olumsuz) ve kategoriye (zayıf veya Güçlü) göre yapılır.

Aşağıdaki şekil fonksiyonun farklı aşamalarını daha iyi anlamak ve özümsemek için yardımcı olacaktır.

**polariteSınıflandır()** Fonksiyonun aşamaları aşağıdaki fonksiyonda görölmektedir:

1- Düzenli metin oluşturmak için metnimiz **matris\_olustur()** fonksiyonuna gönderilir ve çıktı olarak matris yapısında olan bir metin gövdesi elde edilir.

2- öznellik.csv sözlüğün özellikleri çıkarttıktan sonra, oluşturduğumuz matristeki her kelime kelime teker teker `duygu.csv` sözlükteki kelimelerle karşılaştırılır.

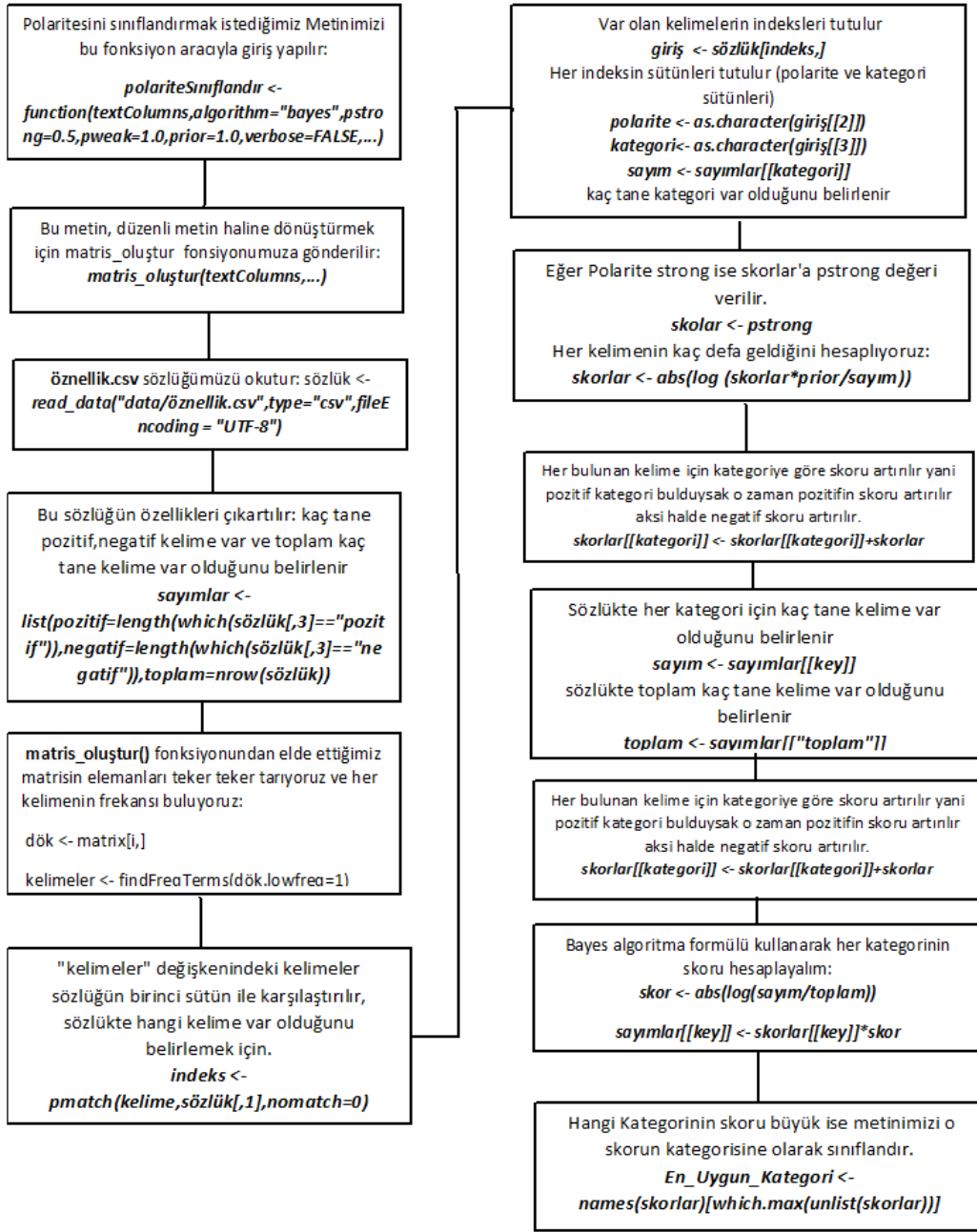
3- Hem öznellik.csv ve hem de matriste bulunan kelimelerin **indeksini**, **polaritesini** ve **kategorisini** tutulur.

4- Eğer polaritesi **güçlü** (strong) ise skoruna güçlü değeri verilir aksi halde **zayıf**(weak) değeri verilir.

5- Her kelime kaç defa geldiđini belirlenir.

6- Her kelime için hem kategorisine hem de polaritesine göre skoru hesaplanır

7- Hangi kelimenin kategori skoru büyükse, metni o skorun kategorisine göre sınıflandırılır.



Şekil 2.11 polariteSınıflandır() Fonksiyonu

Konsoldan edilen sonuç aşağıda olduğu gibidir:

```
Console Terminal x
~/Metinmadenciligi/
[1] "WORD: ağrı CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: ancak CAT: negative POL: weaksubj SCORE: 9.58877680769643"
[1] "WORD: bağı CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: baş CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: başlatmak CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: becerikli CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: bekle CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: birlik CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: bunalımlı CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: çalışmas CAT: negative POL: weaksubj SCORE: 9.58877680769643"
[1] "WORD: çek CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: çekememezlik CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: çekicilik CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: çöz CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: değil CAT: negative POL: weaksubj SCORE: 9.58877680769643"
[1] "WORD: dur CAT: positive POL: strongsubj SCORE: 9.86806760730487"
[1] "WORD: fiyatlı CAT: negative POL: weaksubj SCORE: 9.58877680769643"
[1] "WORD: karakter CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: motor CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: sınırlı CAT: negative POL: weaksubj SCORE: 9.58877680769643"
[1] "WORD: sorun CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: var CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: vermek CAT: positive POL: weaksubj SCORE: 9.17492042674492"
[1] "WORD: yazık CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: yok CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "WORD: zor CAT: negative POL: strongsubj SCORE: 10.2819239882564"
[1] "POS: 0.924955939796596 NEG: 0.51109955884509 RATIO: 1.80973730810232"

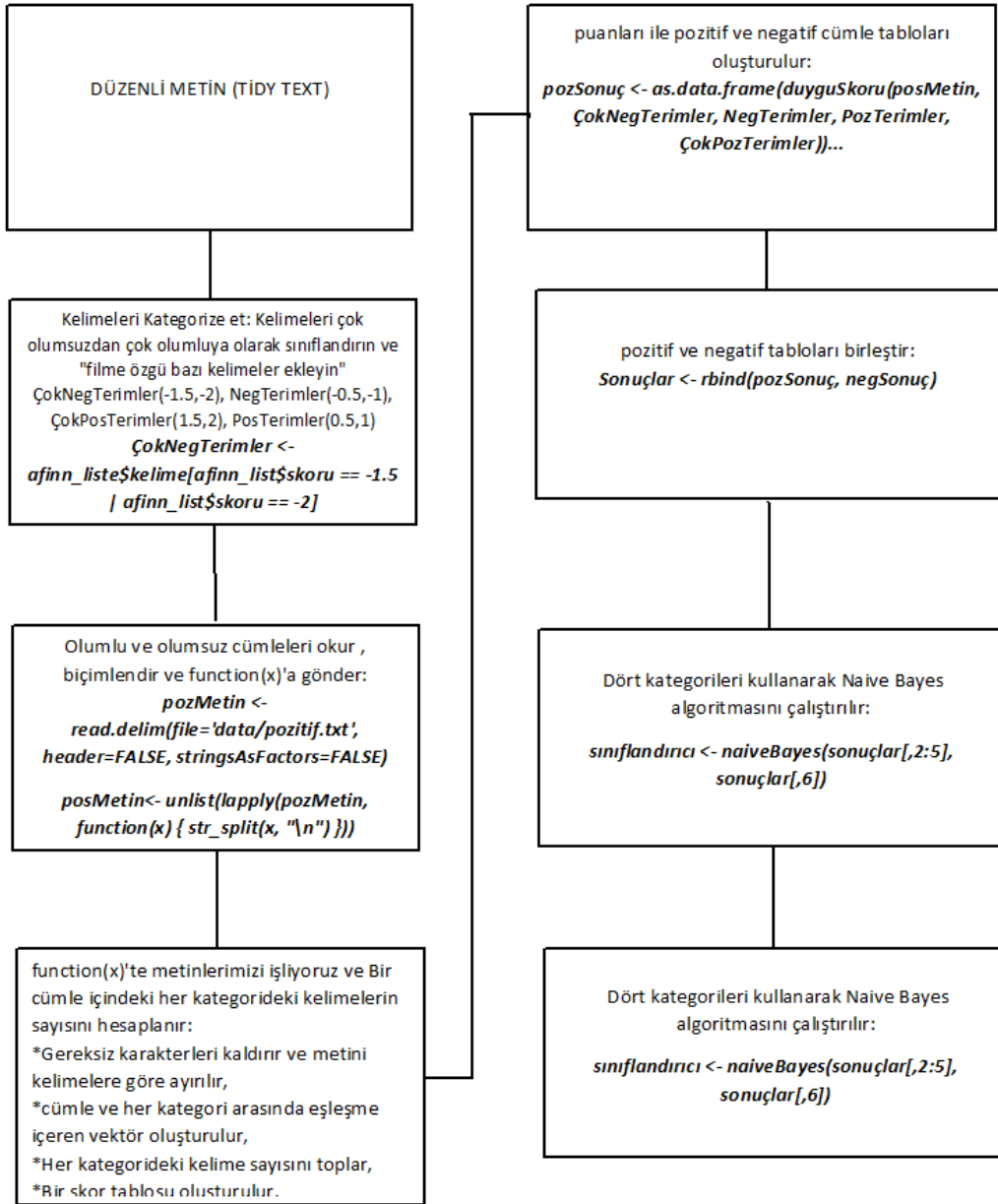
[1,] POS NEG POS/NEG BEST_FIT
> | "0.924955939796596" "0.51109955884509" "1.80973730810232" "positive"
```

Şekil 2.12 polariteSınıflandır() Fonksiyonun ekran çıktısı

### Duygu işleme.R

Bu yöntemde "afinn" sözlüğünü kullanılmaktadır. Bu fonksiyonda, Bayes fonksiyonu önceki yöntemden farklı olarak Rstudio'da bulunan Kütüphane'nin bayes fonksiyonu ile hesaplanmıştır. Bu fonksiyonda kullanılan sözlük ile her kelime -2 ile +2 arasında 0.5 değer aralığı ile (-2, -1,5, -1, 0, 0,5, 1, 1,5, 2) değerlerinden uygun olan değer ile skorlanmıştır. Bu fonksiyon metinde yer alan her kelimenin skor değerine göre Bayes sınıflandırıcı ile uygun sınıfı belirler. Bu fonksiyonun işleyişi aşağıdaki gibidir:

- 1- Sözlüğün 1. sütunu kelime, 2. sütunu skor bilgisini verir.
- 2- Kelimeler çok olumsuzdan az olumsuza kadar [-2, +2] aralığında kategorize edilir.
- 3- Cümleler okutulur ve işlenmek üzere fonksiyona gönderilir.
- 4- Fonksiyon içerisinde metindeki gereksiz karakterler kaldırılır, kelimelere göre ayırır, cümleler ve her kategori arasında eşleme içeren vektör oluşturulur. Her kategorideki kelime sayısı toplanarak bir skor tablosu oluşturulur.
- 5- Skor tablosundaki puanlara göre ile pozitif ve negatif cümlelerin tabloları oluşturulur.
- 6- Oluşturulan tabloları birleştirildikten sonra Naives Bayes algoritması çalıştırılır.
- 7- Sonuçları değerlendirilir.



Şekil 2.13 DuyguSınıflandır() fonksiyonu

Fonksiyonun çıktısı şekil 2.13'deki gibidir..

```

          actual
predicted positive negative
positive    27     10
negative     8     25

95 percent confidence interval:
 0.6243856 0.8399301
sample estimates:
probability of success
      0.7428571

```

şekil 2.14 DuyguSınıflandır() Fonksiyonunun ekran çıktısı

### 2.3.1.4 Üçüncü aşama: duygu görselleştirmesi, bilginin çıkarılması

Cümle analizinden çıkan sayısal verilerin grafiklere dönüştürülmesi daha iyi anlaşılmasını sağlar. Bu grafiklerin elde edilmesi için kelimelerin frekansları, birlikte kullanılma durumları, kelimenin metindeki önemini belirten merkezilik değerleri bulunur

#### 2.3.1.4.1 Frekans analizi

Veri hazırlama aşamasından sonra tutulan kelimeler DTM tablosuna dahil edilir. Daha sonra, bu kelimelerin her birinin sıklığı hesaplanır ve azalan düzende sıralanırlar. Tablo 2.12’de örnek bir kelime frekans tablosu görülmektedir.

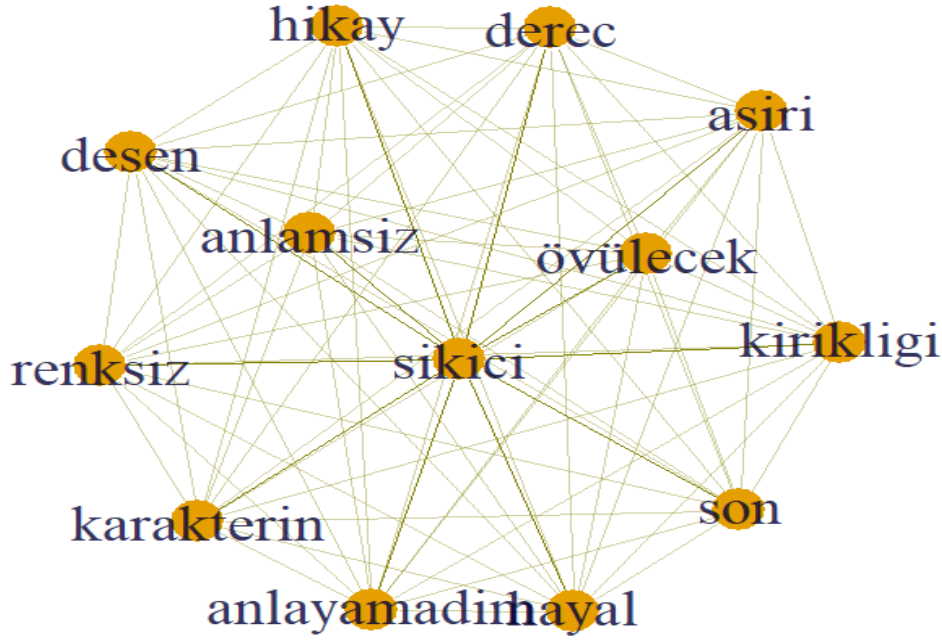
**Tablo2.12** Örnek Frekans Tablosu

	Kelime	Frekans
1	otel	239
2	kitap	109
3	gün	88
4	degil	87
5	yemek	80
6	oteld	70
7	ayni	65
8	kötü	64
9	güzel	60
10	oda	60

#### 2.3.1.4.2 Kelimeler arasındaki birliktelik (cooccurrences)

Metindeki kelimeler arasındaki bağlantıları incelemek için kullanılır. Sık sık birlikte kullanılan kelimeler arasındaki birliktelik değeri yüksektir. Bu ölçümün yapılması için ilişkilendirme kuralları, korelasyon testleri, eş oluşumlar ve benzerlik endeksleri

kullanılır. İki terim metindeki cümleler içerisinde hep birlikte kullanılıyorsa korelasyon değeri 1'e yakın olacaktır. Eğer iki kelime nadiren bir arada kullanılıyorsa korelasyon 0'a yakın olacaktır. Şekil 2.14'te örnek bir metinden elde edilen birliktelikleri gösteren bir grafik görülmektedir. Grafın düğümleri arasındaki bağlantı sayısı, o iki kelimenin birlikte kullanım sıklığını göstermektedir.



Şekil 2.14 Ortak oluşum grafi.

Şekil 2.14'teki grafa birkaç kelime arasında güçlü bir eş oluşum olduğu görülmektedir:

« son » ve « derece » bir bi-gram oluşturup « son derece » verir

« Hayal » ve « kirikliği » bir bi-gram oluşturup « hayal kirikliği » verir

« Hikaye » ve « sıkıcı » bir bi-gram oluşturup « Hikaye sıkıcı » verir

« son » , « derece » ve « sıkıcı » bir bi-gram oluşturup « Son derece sıkıcı » verir

« anlamsız » ve « hikaye » bir bi-gram oluşturup « anlamsız hikaye » verir

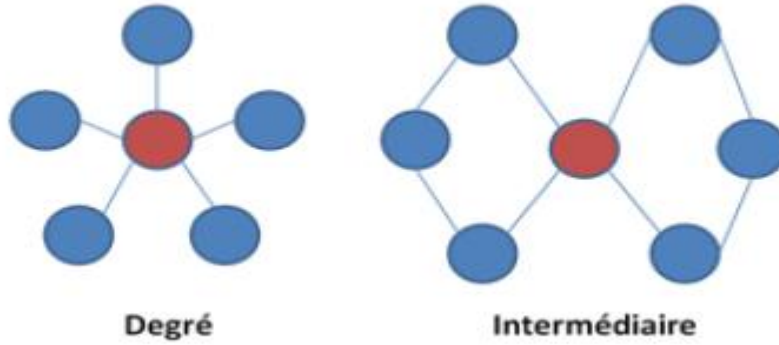
« aşırı » ve « renksiz » bir bi-gram oluşturup « aşırı renksiz » verir

Bu kelimelerin kombinasyonuyla, metnin negatif bir polaritesine veya kategorisine sahip olduğu doğrudan anlaşılmaktadır.

### 2.3.1.4.3 Merkezileştirilme - kelimenin önemi

Kelimelerin birlikte bulunma bağlantılarını incelenerek kelimelerin önemi de belirlenebilir. Kelimenin metinde bulunma sayısının yanı sıra başka kelimeleri bağlamak içinde kullanılma sayısı önemlidir. Bağlantı sayısı ile ölçülen derece merkezietir (degree centrality). Bir kelimenin ne kadar merkezi olduğu, ona ne kadar çok kelime bağladığına bağlıdır.

Eğer bir kelimededen başka bir kelimeye geçilmesi için üçüncü bir kelimededen geçmek gerekiyorsa, o zaman bu kelime diğer kelimeler arasında bir etkileşim sağlamaktadır. Böyle kelimelere aracılık merkezi (betweenness centrality) denilir. **Şekil 14**'te merkezi ve aracı merkezi kelime grafik örnekleri görülmektedir. Merkezietiyi incelemek için iki endeks kullanılıyor ve **Şekil 14**'de gösterilmektedir.



**Şekil 2.15** Merkeziet ölçülerinin grafiksel gösterimi.

**Şekil 2.14** incelendiğinde en merkezi kelimelerin "anlamsız", "sıkıcı", "övülecek" olduğunu söylenebilir. Çünkü bu noktalardan grafin bütün noktalara ulaşılabilir.

Kelimeler grafikte komşu olmadığında grubun bu kelimeler arasındaki yol üzerinde yer alan diğer kelimelere bağlı olduğu söylenebilir. Örneğin: son -> derece -> hayal -> kırıklığı, aşırı -> derece -> anlamsız.

**Merkez derecesi**, bir kelimededen başlayan doğrudan bağlantıların sayısıdır ve kavşak kavramına karşılık gelir. **Aracılık merkezi** ise, metindeki bilgi alışverişi üzerinde en fazla kontrolü olan kelimeler ile temsil etmektedir.

#### 2.3.1.4.4. Kelime bulutları

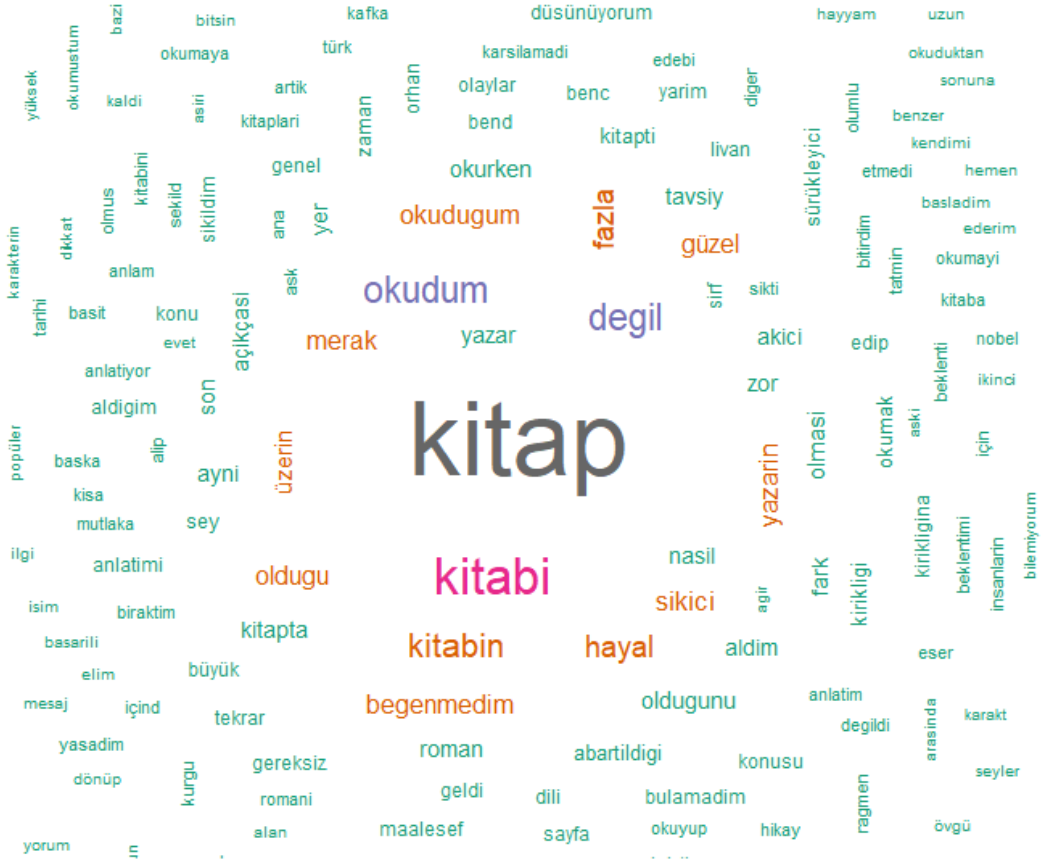
Metin verilerini görselleştirmenin en kolay ve en yaygın yolu kelime bulutlarıdır. Amaç, her bir kelimeyi görüntülemektir. Bu gösterimi kelimenin metin içerisindeki sıklığı o kelime için kullanılan fontun büyüklüğünü etkilemektedir. En sık görülen kelime, en büyük yazı tipiyle gösterilir. Bu kelimeler, metinde ifade edilen ana düşünceyi veya konuyu yansıtmaktadır.

Şekil 2.16'da, şekil 2.15'te sıklıkları verilen kelimelerden oluşan kelime bulutu görülmektedir.

	<u>word</u>	<u>freg</u>	V3
<u>kitap</u>	kitap	100	2.82
<u>kitabı</u>	kitabı	43	1.21
<u>değil</u>	değil	26	0.73
<u>okudum</u>	okudum	26	0.73
<u>kitabın</u>	kitabın	23	0.65
<u>hayal</u>	hayal	20	0.56
<u>fazla</u>	fazla	17	0.48
<u>okuduğum</u>	okuduğum	16	0.45
<u>sıkıcı</u>	sıkıcı	16	0.45
<u>begenmedim</u>	begenmedim	15	0.42
<u>güzel</u>	güzel	15	0.42
<u>merak</u>	merak	14	0.40
<u>olduğu</u>	olduğu	14	0.40
<u>yazarın</u>	yazarın	14	0.40
<u>üzerin</u>	üzerin	13	0.37
<u>okurken</u>	okurken	12	0.34
<u>olduğunu</u>	olduğunu	12	0.34
<u>olması</u>	olması	12	0.34
<u>açıkçası</u>	açıkçası	11	0.31
<u>akıcı</u>	akıcı	11	0.31
<u>kitapta</u>	kitapta	11	0.31
<u>tavsiy</u>	tavsiy	11	0.31
<u>yazar</u>	yazar	11	0.31
<u>yer</u>	yer	11	0.31
<u>aldım</u>	aldım	10	0.28



Şekil 2.16 Kelime sıklıkları



Şekil 2.17 Kelime bulutu

Yukarıdaki kelime bulutu incelendiğinde "kitap", "kitabi" "okudum" ve "degil" frekans tablosundaki mevcut verileri doğrulayan 4 kelime olduğu görülmektedir. Diğer önemli kelimeler ise "begenmedim", "hayal", "sıkıcı", "merak" ve "güzel" olup kitabın karakteristiğine odaklandığını göstermektedir. Buradan da kitapla ilgili taki yorumların çoğunun olumsuz olduğunu anlaşılabilir.

### 2.3.1.4.5 Histogram

Analiz edilen bir metnin önışleme sonrası elde edilen sözcük kökleri ve elde edilen sayısal değerler görülmektedir. Bu skor değerlerinin toplanması sonucu her duygunun skoru elde edilir. Bu skorlar üzerinden çizilen histogram grafiği ile metnin duygu dağılımı daha iyi anlaşılabilir.

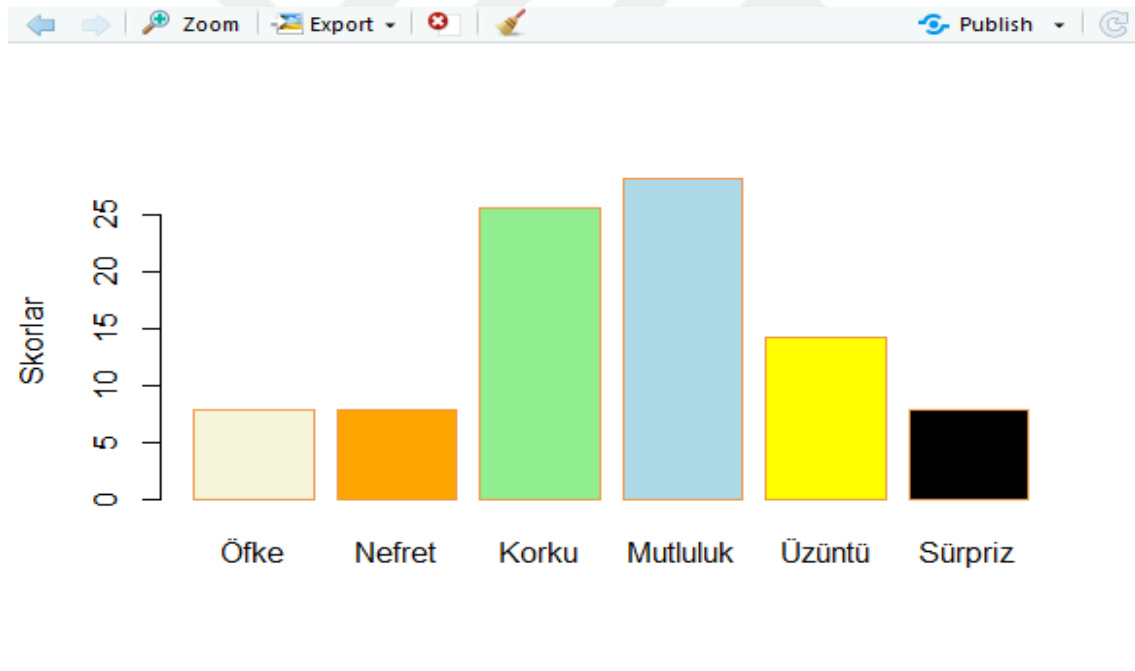
```

[1] "KELİME: be KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: bil KAT: sürpriz SKOR: 4.49980967033027"
[1] "KELİME: bu KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: çok KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: çok KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: di KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: geç KAT: mutluluk SKOR: 6.76503897678054"
[1] "KELİME: göz KAT: öfke SKOR: 5.97888576490112"
[1] "KELİME: iç KAT: üzüntü SKOR: 6.38687931936265"
[1] "KELİME: mu KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: se KAT: nefret SKOR: 4.95582705760126"
[1] "KELİME: se KAT: korku SKOR: 5.92958914338989"
[1] "KELİME: Se KAT: mutluluk SKOR: 6.76503897678054"
ÖFKE             NEFRET             KORKU             MUTLULUK
[1,] "7.81480342948936" "7.81480342948936" "25.603570859659" "28.109920359831"
ÜZÜNTÜ             SÜRPRİZ             EN_UYGUN_DUYGU
[1,] "14.201682748852" "7.81480342948936" "mutluluk"

```

**Şekil 2.18** Fonksiyon çıktısı.

Bu metin biraz üzüntülü, biraz öfkeli cümleler yer almasına rağmen metnin en yoğun duygusunun “mutluluk” olduğu görülmektedir. **Şekil 2.18**'de metnin duygu histogramı görülmektedir.



**Şekil 2.19** Histogram

### 2.3.1.4.6 Metinden çıkartılabilecek diğer veriler

Bir metin içerisindeki kelimelerle ilgili sayısal veriler elde edilebilir. Ayrıca belirli bir kelimenin diğer kelimelerle olan ilişkilerini gösteren sayısal veriler elde edilebilir.

rmeye çalışmış. Toplum için anlamı olan ne varsa kitabında toplamış. Zaman kaybetmiş tufleri okumaktan kaçınılmış.

**X139** bana göre orhan pamuk'un en aceleyle gelmiş, ne anlattığı belli olmayan, en zorlama kitabı bu **kitap**. kara **kitap**'ı veya benim adım kırmızı'yı okumuşsanız çok iyi anlarsınız ne demek istediğimi...

**X142** orhan pamuk söz konusu olduğunda beklenti de yüksek oluyor haliyle. ancak kendi adıma, tam anlamıyla hayal kırıklığına uğradığım bir **kitap** oldu. orhan pamuk bu kez ruhunu katmamış eserine.

**X143** ahlaksızlıkta sınır tanımayan bir **kitap**, bu tür kitapların türk aile yapısını ve türk insanının değerleri ni sarsıntıya uğratmak için kasıtlı olarak yazıldığını düşünüyorum.

**X146** neredeyse tüm orhan pamuk kitaplarını okudum, bence en güzeli kara kitaptır. masumiyet müzesi kitabını da beğenmiştim. ama bu **kitap** beklentimi karşılamadı. sanki **kitap** özeti gibi kısa ve özensiz buldum.

Şekil 2.20 kitap kelimesi geçen dökümanlar.

	X22	X23	X24
Number of terms	24.00000	39.00000	29.00000
Number of unique terms	22.00000	38.00000	27.00000
Percent of unique terms	91.66667	97.43590	93.10345
Number of hapax legomena	21.00000	37.00000	25.00000
Percent of hapax legomena	87.50000	94.87179	86.20690
Number of words	24.00000	39.00000	29.00000

	Corpus mean	Corpus total
Number of terms	34.68707	5099.00000
Number of unique terms	31.53741	2119.00000
Percent of unique terms	91.95955	41.55717
Number of hapax legomena	29.14966	1461.00000
Percent of hapax legomena	85.39151	28.65268
Number of words	34.68707	5099.00000

Şekil 2.21 Korpus hakkında sözcüksel rapor

Şekil 2.21'deki veriler incelendiğinde, Bu metnin 2119 farklı kelimedenden oluştuğu görülmektedir. X23 ile temsil edilen belgede (38), X22'den daha fazla benzersiz veya farklı kelime vardır.

Şekil .22'de örnek bir belgenin kelime inceleme sonucu görülmektedir.

\$X143	% Term/Level	% Level/Term	Global %	Level %	Global occ.
türk	9.090909	28.571429	0.13728182	2	7
kitapların	4.545455	50.000000	0.03922338	1	2
aile	4.545455	33.333333	0.05883507	1	3
düşünüyorum	4.545455	14.285714	0.13728182	1	7
olarak	4.545455	5.555556	0.35301039	1	18
	t value	Prob.			
türk	3.375525	0.0004			
kitapların	2.381918	0.0086			
aile	2.229496	0.0129			
düşünüyorum	1.883280	0.0298			
olarak	1.439537	0.0750			

Şekil 2.22 143 numaralı belgede kelime inceleme

"Türk" kelimesi, 143 numaralı belgede belirtilen kelimelerin tümünün %9.09'unu temsil eder.

Bu metindeki "Kitap" kelimesinin detaylı incelemesi yapılabilir.

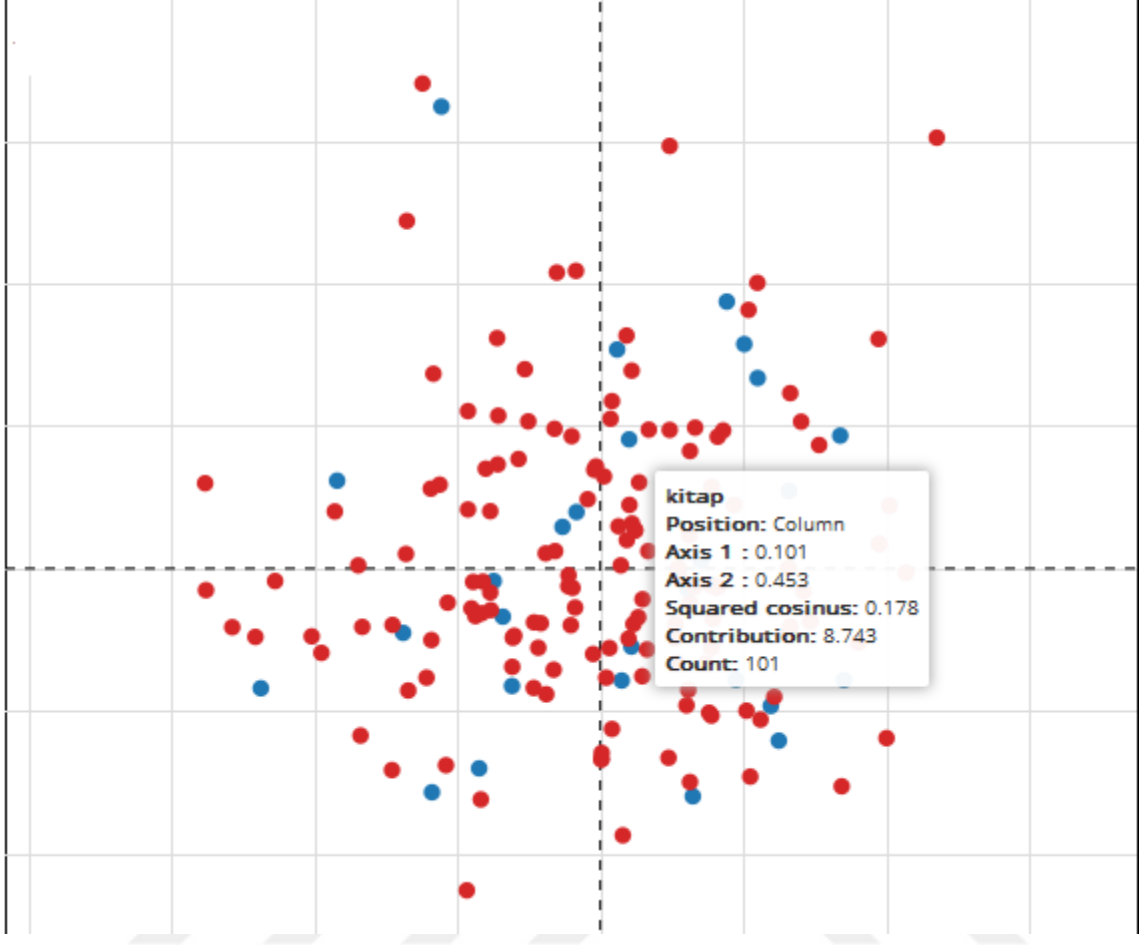
	% Term/Cooc.	% Cooc./Term	Global %	Cooc.	Global
kitap	3.83738602	100.00000	1.98078055	101	101
bir	5.43313070	60.59322	4.62835850	143	236
bitiremedim	0.18996960	100.00000	0.09805844	5	5
kesinlikle	0.18996960	100.00000	0.09805844	5	5
rağmen	0.18996960	100.00000	0.09805844	5	5
bence	0.26595745	87.50000	0.15689351	7	8
maalesef	0.26595745	87.50000	0.15689351	7	8
sayfa	0.26595745	87.50000	0.15689351	7	8
ya	0.26595745	87.50000	0.15689351	7	8
başarılı	0.15197568	100.00000	0.07844675	4	4
dolu	0.15197568	100.00000	0.07844675	4	4
evet	0.15197568	100.00000	0.07844675	4	4

**Şekil 2.23** « kitap » kelimesiyle eş zamanlı gelen kelimeler.

"KİTAP" kelimesini içeren yorumların % 0,18'ı "bitiremedim" kelimesi ile birlikte kullanılmıştır. Aynı zamanda da "bitiremedim" kelimesi içeren yorumların %100'ü "KİTAP" kelimesiyle birlikte kullanılmıştır.

#### **Sözlük tablo şemasından kelime eşleşme Analizi :**

Ayrıca, metindeki her bir kelimeyi, görünme sayısını, metne katkısını belirtilmesini sağlayan bir nokta bulutu grafiği de çıkarılabilmektedir.



**Şekil 2.24** Sözlüksel Tablosu (ST)

Şekil 2.24'te mavi ile işaretlenen "kitap" kelimesi, metnimize 8743 katkı payı ile 101 kez yer almaktadır.

kitap kökü olan kelimeleri de çıkarabilir:

Dimension: Axis 1 (1.27%)  
Hide: None

Active levels  
Show 10 entries  
Search: kitap

Level	Position	Coord	Contrib	Cos2	Count
kitaplarda	Column	0.360	0.01	0.002	2
kitap	Column	-0.027	0.00	0.000	101
kitaplardan	Column	-0.128	0.00	0.001	4
kitapları	Column	-0.101	0.00	0.000	6
kitaptı	Column	-0.090	0.00	0.000	8
kitaplar	Column	-0.122	0.00	0.000	3
kitaplarından	Column	-0.128	0.00	0.000	2
kitaplarıyla	Column	0.203	0.00	0.000	1
kitapta	Column	-0.055	0.00	0.000	11
kitaptan	Column	-0.102	0.00	0.000	2

Showing 1 to 10 of 18 entries (filtered from 2,265 total entries) Previous 1 2 Next

## Şekil 2.25 « kitap » köküyle olan kelimeler

Şekil 2.25'te « kitap » köküyle olan kelimelerin sayısı(2265 toplam) ve hangi biçimde olduklarına dair bilgileri görüntüleyen ekran yer almaktadır.

## 2.3.2 Duygu derecelendirme ve sınıflandırma modülü

### 2.3.2.1 Duygu derecelendirme

Duygu derecelendirmesi ve tespiti modülü geliştirilen sistemin en önemli bölümüdür. Amaç, analiz edilen metinlerde ifade edilen duyguları değerlendirmek ve sınıflandırılmasıdır. Metindeki duyguyu belirlemek amacıyla bir metne not ya da polarite atamak için üç farklı yöntem geliştirilmiştir. Duygu değerlendirme modülü, oluşturulan farklı sözlükleri ve ayrıca metinlerin sınıflandırılmasında farklı yaklaşımları kullanmaktadır. Bir metindeki duygu değerlendirmeleri cümle cümle yapılır. Sürecin sonunda ilişkili cümleler skorlarla birlikte alınarak her cümle için performans ölçümü yapılır. Tez çalışmasında duyguları sınıflandırmak için sözlüklere özgü farklı yöntemler kullanılmıştır.

Birinci sözlük kullanılarak duygu skoru hesaplama yöntemi şu şekildedir.

```
sayım <- sayımlar[[anahtar]]
```

```
toplam <- sayımlar[["toplam"]]
```

```
skor <- abs(log(sayım/toplam))
```

```
skorlar[[anahtar]] <- skorlar[[anahtar]]+skor
```

Burada örneğin « öfke » duygu skoru isteniyorsa:

```
sayım <- sayımlar[[öfke]] #Sözlükte öfke duygusu için kaç tane
kelime var?
toplam <- sayımlar[["toplam"]] #Sözlükte toplam kaç tane kelime var?
skor <- abs(log(sayım/toplam)) # Bu sözlükte öfke duygusu çıkma olasılığı
nedir?
skorlar[[öfke]] <- skorlar[[öfke]] + skor #Metinde öfke duygusu içeren kelime ile
karşılaştığında öfke duygusunun skorunu arttır.
```

**İkinci sözlük kullanılarak duygu skoru hesaplama yöntemi şu şekildedir.**

```
sayım <- counts[[Anahtar]] # Pozitif kelimelerin ve negatif kelimelerin sayısı çıkarılır
#örnek count[pozitif]=7345
toplam <- counts[["toplam"]] # Sözlükte toplam kaç tane kelime var?
skor <- abs(log(sayım/toplam)) #Bu sözlükte pozitif/negatif duygu çıkma
olasılığı nedir?
skor[[Anahtar]] <- skor[[Anahtar]]+skor #Metinde öfke duygusu içeren kelime ile
karşılaştığında öfke duygusunun skorunu arttır.
```

**Bu yöntemde polariteden dolayı skorlarda değişim olabilir.**

**örnek:**

```
if (polarite == "GüçlüÖznel") skor <- pstrong #pstrong = 0.5, pweak = 1
```

```
skor <- abs(log(skor*prior/count))
```

```
skorlar[[anahtar]] <- skorlar[[anahtar]]+skor
```

**Değerler çok küçük olduğu için logaritma kullanılmıştır.**

**İkinci sözlük kullanılarak duygu skoru hesaplama yöntemi şu şekildedir.**

#Her kategorideki kelime skorları veya puanları topluyoruz

```
çPozEşleme <- sum(!is.na(çPozEşleme))
```

```
pozEşleme <- sum(!is.na(PozEşleme))
```

```
çNegEşleme <- sum(!is.na(çNegEşleme))
```

```
negEşleme <- sum(!is.na(NegEşleme))
```

Bu skorlar toplandıktan sonra bayes fonksiyonuna gönderilir.

### 2.3.2.2 Metin sınıflandırma:

Metin sınıflandırma işleminin kodu **Duygu\_sınıflandır.R** dosyasında yer almaktadır.

Burada sınıflandırma unlist fonksiyon ile yapılmaktadır :

**unlist(skorlar)**

Bu fonksiyon parametreleri(skorlar) ile birlikte kullanılmaktadır.

Bu fonksiyon en belirgin duyguyu belirlemek için kullanılır ve metnin sınıflandırılması en yüksek duygu skoru alınarak karar verilir (**which.max(unlist(skorlar))**). Bu sadece metni en yüksek duygu skoruna göre sınıflandırılmasını sağlar.

Polariteye göre sınıflandırma işleminin kodu **Polarite\_sınıflandır.R** dosyasında yer almaktadır. Burada bir önceki sınıfta olduğu gibi aynı şekilde **unlist(skorlar)** fonksiyonu kullanılarak yapılır.

Metnin en yüksek puana sahip polariteye göre sınıflandırılmasını sağlar:

**which.max(unlist(skorlar))**

**Tablo2.12** Ölçüm Metrikleri Tablosu

	Gerçek Pozitif	Gözlem Negatif
Pozitif	x	y
Negatif	z	w

Doğruluk ve performans değerlendirmesini (F-Ölçümü) tanımlamak için aşağıdaki işlemler göz önüne alınır.

Bir metin doğru şekilde olumlu olarak sınıflandırıldığında doğru pozitif (x) kabul

edilsin.

Yanlış pozitif (y), pozitif olarak sınıflandırılmış bir negatif metindir, Yanlış Olumsuz (z), olumlu bir metindir ancak olumsuz olarak sınıflandırılmıştır ve son olarak doğru negatif (w) doğru olarak sınıflandırılmış negatif metindir. [40,41].

gerçek/doğru pozitif (TP) oranı veya geri çağırma(Recall):  $R = x / (x + z)$ , yanlış negatif (FP) oran veya hassasiyet(precision):  $P = x / (x + y)$ , doğruluk(Accuracy):  $A = (x + w) / (x + y + z + w)$  ve F ölçümü (performans):  $F = 2 * (P * R) / (P + R)$ . Birçok durumda Fmeasure'ü kullanacağız, çünkü bu bir testin doğruluğunun bir ölçüsüdür ve hem hassasiyete hem de hatırlamaya/Geri çağırmaya dayanır. [42]

### 2.3.3 Kullanılan veriseti

İnternet, film eleştirmenleri, kitaplar ve daha fazlası için önemli bir bilgi kaynağıdır. Tüm filmler için tanımlayıcı sayfalar bulabiliriz ve sayısız sinematografi sitelerinin veya kitaplarının gelişimi sayesinde, internet kullanıcılarının topluluğu tüm eserlere bakış açıları ifade edebilirler. Bu nedenle, birçok site, üyelerine film veya kitapları yorumlaması ve notlanması için izin vermektedir. .

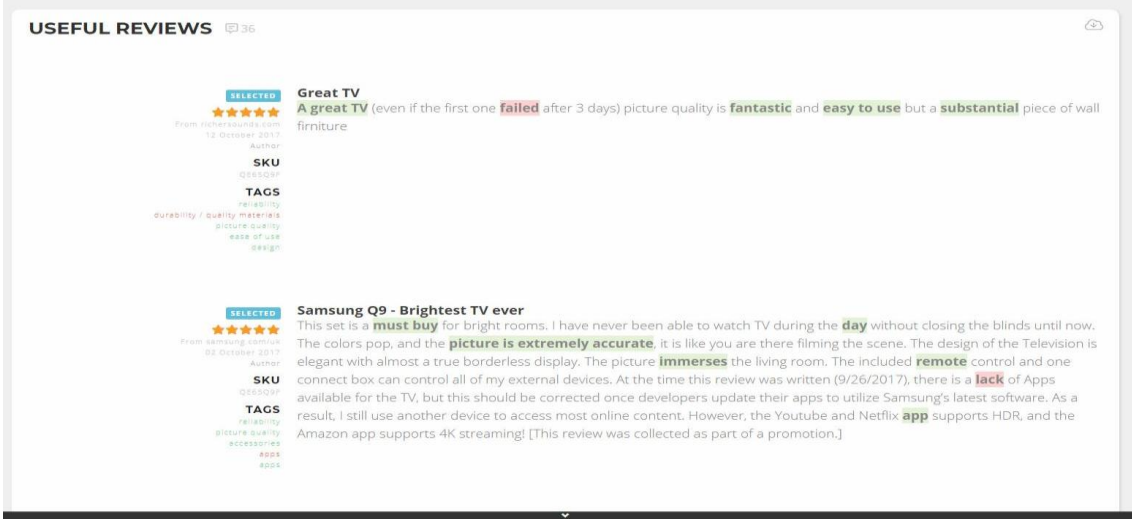
Tez çalışmasında üç farklı yorum sınıfına ait 35'er, toplamda 105 film yorumu içeren veriseti uygulama için kullanılmıştır [44]. Film yorumlarında negatif, pozitif ve tarafsız olmak üzere üç sınıf bulunmaktadır.

Bu uygulamada negatif ve pozitif yorumlar kullanılmıştır. Metinlerdeki ortalama kelime sayısı 49'dur

**“Tüm Müşteri Yorumları (91309 adet)”** dosyasında bir çevrimiçi kitap satış sitesinde (kitapyurdu.com) yer alan 47 kitaba ait 91309 müşteri yorumu listelenmektedir[45].

### Humur Sentiment Dataset

Duygu analizi üzerine bir sistem kurmak istersek, öncelikle görüşlerini ve öznel cümleleri gösteren belgelerde tüketicilerden ve kullanıcılardan fikir toplamak zorunda kalacaktır. Bazen, **kitapyurdu.com**, **Imdb.com**, **Epinions.com**, **Amazon.com** gibi kullanıcı görüşlerinin iyi yapılandırıldığı büyük siteler söz konusu olduğunda, bu nispeten kolaydır.[46]



Şekil 2.26 Samsung Q9 ile ilgili yorumlar

İlgimizi çeken belgeler toplandıktan sonra, bu belgeler tarafından ifade edilen tüm görüş ve duyguları tanımlama sorunu ile karşı karşıyayız.

### 2.3.4 Uygulamanın çalışma prensibi:

Öncelikle aşağıdaki kod çalıştırılarak kütüphane yüklenmesi gereklidir.

**library(stopwords)** #durdurucu kelimeler

**library(stringr)** #Basit, Tutarlı Sarmalayıcılar Ortak Dize İşlemleri

**library(plyr)** #plyr, verileri parçalamayı, üzerine bir şeyler yapmayı ve bir araya getirmeyi kolaylaştıran bir R paketidir.

Bu ortak bir veri işleme adımdır. Önemli olarak, plyr, giriş ve çıkış veri formatlarını sözdizimsel olarak tutarlı bir fonksiyonlar setinden kontrol etmeyi kolaylaştırır.

**library(graphics)** # Grafik dilbilgisi, grafiğin bileşenlerini net bir şekilde tanımlamamızı sağlayan bir araçtır. Böyle bir dilbilgisi, adlandırılmış grafiklerin (örneğin "saçılma grafiği") ötesine geçilmesine ve istatistiksel grafiklerin temelini oluşturan derin yapı hakkında bilgi edinilmesine izin verir.

**library(RTextTools)**

**library(tm)**

**library(igraph)**

```
library(ggplot) .
```

```
library("languageR", lib.loc="~/R/win-library/3.4")
```

```
library("SnowballC", lib.loc="~/R/win-library/3.4")
```

```
devtools::load_all()
```

Yok olan kütüphaneler ise bu kodla indirilmektedir:

```
install.packages("RTextTools")
```

Ardından yüklenen kütüphaneleri çağırmak için aşağıdaki komutu kullanılmaktadır:

```
library(RTextTools)
```

Bu kütüphaneler sırasıyla yüklenmelidir

Konsol üzerine değerlendirmek veya sınıflandırmak istenilen metin yüklenir:

```
dökümanlar <- c("hayal kırıklığı bu kadar övülecek ne var anlayamadım. karakterin kendisi aşırı renksiz")
```

Eğer birden fazla döküman yüklenmek istiniyorsa konsoldan aşağıdaki komut gönderilmelidir.

```
dökümanlar <- c("hayal kırıklığı bu kadar övülecek ne var anlayamadım. karakterin kendisi aşırı renksiz sıkıcı.",
```

```
    "hikaye desen son derece sıkıcı, anlamsız. sonuna kadar gideyim belki değer diye düşündüm ama o da olmadı.",
```

```
    "bir an önce bitsin dediğim bir kitaptı.")
```

Her dokümanı diğerinden ayırmak için, tırnaklar içerisinde yazıldıktan sonra, dokümanın sonunda bir virgül konulur, böylece kod bir sonraki gelen metnin başka bir doküman olduğunun farkına varır.

Metnin ardından aşağıda açıklanan fonksiyonların çağırılması gerekir. Bu fonksiyonlar metnin nasıl sınıflandırılmak istendiğine bağlıdır:

- Eğer 6 duyguya göre sınıflandırmak isteniyorsa aşağıda belirtilen fonksiyonu çağırılır:

**duygu\_sınıflandır(dokümanlar,algorithm="bayes",verbose=TRUE)**

Bu fonksiyonun parametreleri şu şekildedir:

**Dökümanlar:** sınıflandırmak istenilen dokümanlar

**algorithm="bayes"** kullanılmak istenilen algoritma

**verbose=TRUE:** Yapılandırmada ayrıntılı olarak True olarak ayarlamak, hizmetin daha fazla çıktı üretmesine neden olur (hem UYARI hem de info log düzeylerini gösterir), normalde yalnızca UYARI veya daha yüksek (örneğin, ERROR) görürsünüz.

- Eğer pozitif veya negatif olarak sınıflandırmak isteniyorsa:

**polarite\_sınıflandır(dokümanlar,algorithm="bayes",verbose=TRUE)** fonksiyonu çağırılır.

- **Duygu\_işleme.R** sınıfı(class) için satır satır derlenir.

### 2.3.5 Uygulamanın Mimari Yapısı

Farklı sözlüklere göre metinlerin sınıflandırılması geliştirilen yazılımı ile mümkün olmaktadır. Uygulamamızın doğruluk oranını artırmamıza, metnin tüm özelliklerini vurgulamamıza izin vermektedir. Bu çalışma, 3 adet farklı sözlük kullanarak Türkçe metin madenciliğinde gerçekleştirilen az sayıda uygulamalardandır.

Duygu analizi esasen istatistiksel bir işlem olduğundan, sözlük daha büyük olduğunda, uygulamanın başarı oranı artmaktadır.



### 3. BULGULAR VE TARTIŞMALAR

Bu bölümde deneysel çalışmalar sonucunda elde edilen bulgular sunulmuştur.

Tablo3.1’de önerilen tekniğin adımları görülmektedir.

**Tablo 3.1.** Özet Tablosu.

No.	Adım
1.	Metin girişi.
2.	<ul style="list-style-type: none"><li>o Metin analizi süreci .</li><li>o Kelimelere bölme</li><li>o Metindeki her kelime için</li><li>o Durdurma kelimesi listesinde yer alıyor ise kaldır</li><li>o Noktalama listesinde yer alan işaretleri kaldır.</li><li>o Tüm kelimeleri küçük harflere çevir</li></ul>
3	<ul style="list-style-type: none"><li>o Analiz süreci elde edilen önışlemeden geçmiş cümle yapısı oluşturulur.</li><li>o Kelimelerin kökü bulunur</li><li>o Her kelimelerin sözlükte belirtilen değeri bulunur (pozitif veya negatif). Sözlükte yer almayan kelimeler kaldırılır</li></ul>
4	<ul style="list-style-type: none"><li>o Olumsuzluk eklerinin kontrolü .</li><li>o kelimeler, sıfatlar, zarf... üzerine kontrol edilir.</li></ul>
5.	<ul style="list-style-type: none"><li>o Her cümle incelemesinde, cümleyi oluşturan her kelime için</li><li>o SA(duygu analizi) puanı 0 yap.</li><li>o Her kelimeyi polaritesini değerlendir,</li><li>o Kelimelerin polaritesine göre cümlenin ağırlıklı duygu polaritesini belirle.</li><li>o Cümlenin 6 duyguya göre sınıfını ve kategorisini belirle</li></ul>

Tez çalışmasında dört farklı veriseti kullanılmıştır.:

**İlk Veri setimiz : (Data1), 35 olumlu cümle ve 35 olumsuz cümle içermektedir.**

**İkinci Veri setimiz : (Data2), 700 olumlu cümle ve 700 olumsuz cümle içermektedir.**

**Üçüncü veri seti: (Data3), yaklaşık 10000 olumlu cümle ve 10000 olumsuz cümle içermektedir.**

PolariteSınıflandır.R fonksiyonu yöntem 1 ve Duyguİşleme.R fonksiyonu ise yöntem 2 olarak adlandırılmıştır.

Geliştirilen sistemin üzerine 3 farklı data seti test edilmiştir:

İlk Test edilen veriseti Data1 olarak adlandırılmıştır. Bu veriseti sinema filmi yorumları içermektedir. 35 negatif cümle ve 35 pozitif cümle içermektedir. Bu veriseti üzerinde

yapılan deneylerde yöntem 1 kullanılarak 35 negatif yorumun 27'si negatif polariteli olduğu sistem tarafından tespit edilmiştir. Sekiz adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 31 tanesi negatif 4 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür.

Pozitif yorum cümleleri üzerinde yapılan deneylerde yöntem 1 ve yöntem 2 35 cümlenin 27 tanesini pozitif olarak nitelendirmiştir. Sekiz cümle ise negatif olarak nitelendirilmiştir. Yöntem 1'in toplam doğruluk oranı %77,14 iken yöntem 2'nin toplam doğruluk oranı %82,85 olmuştur. Deneylerle ilgili veriler **Tablo 2.15**'te sunulmuştur.

**Tablo2.15** DATA1 sonuçları

	<b>DATA1</b>	<b>Yöntem 1</b>	<b>Doğruluk Oranı</b>	<b>Yöntem 2</b>	<b>Doğruluk Oranı</b>
<b>35 Cümle</b>	<b>Negatif</b>	<b>27 Negatif ve 8 Pozitif</b>	<b>%77,14</b>	<b>31 Negatif ve 4 pozitif</b>	<b>%88,57</b>
<b>35 Cümle</b>	<b>Pozitif</b>	<b>27 Pozitif ve 8 Negatif</b>	<b>%77,14</b>	<b>27 Pozitif ve 8 Negatif</b>	<b>%77,14</b>

İkinci verisetinde 700 adet negatif, 700 adet pozitif cümle yer almaktadır. Bu veriseti üzerinde yapılan deneylerde yöntem 1 kullanılarak 700 negatif yorumun 491'si negatif polariteli olduğu sistem tarafından tespit edilmiştir. 210 adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 525 tanesi negatif 175 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür.

Pozitif yorum cümleleri üzerinde yapılan deneylerde yöntem 1 700 cümlenin 528'ini pozitif, yöntem 2 ise 524 tanesini pozitif olarak değerlendirmiştir. Yöntem 1'in toplam doğruluk oranı %72,78,14 iken yöntem 2'nin toplam doğruluk oranı %74,92 olmuştur. Deneylerle ilgili veriler tablo 2.16'te sunulmuştur.

**Tablo2.16** DATA2 sonuçları

DATA2	Yöntem 1	Doğruluk Oranı	Yöntem 2	Doğruluk Oranı
700 Negatif Cümle	491 Negatif ve 210 Pozitif	%70,14	525 Negatif ve 175 pozitif	%75
700 Pozitif Cümle	528 Pozitif ve 172 Negatif	%75,42	524 Pozitif ve 176 Negatif	%74,85

· Üçüncü veri setimiz **Data3** olarak adlandırılacak , **Humir** data setinden alındı [47] , 10 329 negatif cümle ve 10 833 pozitif cümleleri içermektedir. Bu Data seti daha iyi değerlendirebilmek için 20 tane pozitif data seti ve 20 tane negatif data seti olarak bölünmüştür:

**Tablo2.17** DATA3 sonuçları

DATA3	Yöntem1	Doğruluk Oranı	Yöntem2	Doğruluk Oranı
10 833 Pozitif Cümle	Positive: 8 273 Negative: 2 560	%76,36	Positive: 8 628 Negative: 2 205	%79,64
10 329 Negatif cümle	NEGATIVE: 7 423 POSITIVE: 2 906	%71,86	NEGATIVE: 7 773 POSITIVE: 2 556	%75,25

Bu veriseti üzerinde yapılan deneylerde yöntem 1 kullanılarak 10 329 negatif yorumun 7 423'si negatif polariteli olduğu sistem tarafından tespit edilmiştir. 2 906 adet negatif yorum ise pozitif polariteli olarak nitelendirilmiştir. Bu negatif cümlelerin yöntem 2 ile değerlendirilme işlemi sonrasında ise 7 773 tanesi negatif 2 556 tanesi pozitif olarak nitelendirilmiş ve doğruluk oranının arttığı görülmüştür.

pozitif yorum cümleleri üzerinde yapılan deneylerde yöntem 1 10 833 cümlelerin 8 273 tanesini pozitif olarak nitelendirmiştir. 2 560 cümle ise negatif olarak nitelendirilmiştir. yöntem 2 10 833 cümlelerin 8 628 tanesini pozitif olarak nitelendirmiştir. 2 205 cümle ise

negatif olarak nitelendirilmiştir

Yöntem 1'in toplam doğruluk oranı %74.11 iken yöntem 2'nin toplam doğruluk oranı %77,445 olmuştur. Deneyle ilgili veriler tablo 2.17'de sunulmuştur.



## 4. SONUÇLAR

Bu tez çalışmasında metin madenciliği ile bir duygu analizi sistemi geliştirilmiştir. Çeşitli veri setleri üzerinde yapılan testlerle sistemin performansı ölçülmüştür. Geliştirilen sistem ayrıca incelediği metindeki duygu türlerini görselleştirmektedir. Sistem duyguları sınıflandırmak için üç farklı sözlük kullanmaktadır. Geliştirilen sistemin inceleyeceği metin uzunluğu ile ilgili bir kısıtı bulunmamaktadır. Fakat analiz edilecek metin veya kelime sayısı arttıkça, işleme süresinin arttığı görülmüştür. İşlem süresini azaltmak için metinler üzerinde bir önışleme prosedürü uygulanmış ve bir duygu taşımayan kelimeler ve noktalama işaretlerinin listeden çıkarılması sağlanmıştır. Ancak önışleme duygu analizinin tüm zorluklarını giderememiştir.

Hatalı yazılan kelimeler, kinayeli yorumlar, art niyetli yazılmış yorumlar, bazı deyimlerin ve kelimelerin birden fazla anlamının olması, karşılaştırma içeren cümlelerin bir taraf için olumlu duygu taşıırken diğer taraf için olumsuz duygu taşıması, yorum yazarının kültür farklılığı gibi durumlar duygu analizini zorlaştırmaktadır. Yapılan test çalışmalarında sistemin hatalı karar verdiği cümleler incelendiğinde bu gibi durumlarla karşılaşılmıştır.

Örneğin “Hastanızın iyileşmesini umuyorum.” cümlesinde iyileşmek fiili olumlu anlam taşıırken umuyorum ifadesi bunun henüz gerçekleşmediğini belirtmektedir. Geliştirilen sistem bunu tespit etmekte başarılı olamamıştır. Bazı cümlelerdeki anlamı çıkarabilmek için sadece kelimelerin sözlük anlamları yetmez. Örneğin “Bu kadının sese Sezen Aksu’ya benziyor.” cümlesinde gerçek anlamı çıkarabilmek için Sezen Aksu’nun sesi güzel biri olduğu bilinmelidir. Duygu anlamının tespitinin zor olduğu bir türde karşılaştırma cümleleridir. “İzmir, Antalya’dan daha güzel bir şehir.” ifadesi İzmir için olumlu anlam taşıırken Antalya için olumsuz olarak nitelendirilebilir. Sistemin bu gibi durumları da değerlendirebilme imkanı yoktur.

Üç farklı veri setiyle yapılan deneysel çalışmalarda elde edilen sonuçları aşağıda olduğu gibidir:

- **1. Data:** 1. yöntem için **35** Negatif cümlede **27**’si negatif, doğruluk oranı **%77,14**; 2. yöntem ise **31**’i negatif olarak bulunmuştur, doğruluk oranı ise **%88,57**’dir.
  1. yöntem için **35** pozitif cümlede **27**’si pozitif, doğruluk oranı **%77,14**;
  2. yöntem ise **27**’si pozitif olarak bulunmuştur, doğruluk oranı ise **%77,14**’tür.

- **2. Data:** 1. yöntem için **700** Negatif cümlede **491**'i negatif, doğruluk oranı **%70,14**; 2. yöntem ise **525**'i negatif olarak bulunmuştur, doğruluk oranı ise **%75**'tir.

1. yöntem için **700** pozitif cümlede **528**'i pozitif, doğruluk oranı **%75,42**; 2. yöntem ise **524**'ü pozitif olarak bulunmuştur, doğruluk oranı ise **%74.85**'tir.

- **3. Data:** 1. yöntem için **10 833** pozitif cümlede **8 273**'ü pozitif, doğruluk oranı **%76,36**; 2. yöntem ise **8 628**'i pozitif olarak bulunmuştur, doğruluk oranı ise **%79,64**'tür.

1. yöntem için **10 329** negatif cümlede **7 423**'i negatif, doğruluk oranı **%75,42**;  
2. yöntem ise **7 773**'ü negatif olarak bulunmuştur, doğruluk oranı ise **%75,25**'tir.

Bu alandaki gelecek çalışmalarda negatif anlam içeren cümlelerin analiz üzerinde daha fazla çalışılmalıdır. Türkçe'de olumsuzluk anlamı veren bir çok kelime vardır. Bu kelimeler tek başına olumlu görülen bir cümleyi olumsuz hale getirebilmektedirler. Bu da pek çok duygu analizi sisteminin başarısını etkilemektedir. Fiillerin belirttiği duygular üzerine daha çok çalışma yapılmalıdır. Sıfat ve zarflar üzerine yoğunluklu yapılan çalışmalar çok yüksek başarıya erişememektedir. Türkçe için geliştirilen sözlük tabanlı çalışmaların başarısının artması için mecaz anlamları da içeren sözlüklerin üretilmesi de gereklidir.

## KAYNAKLAR

- [1] Simon Kemp, report author , Digital around the world in 2019, the essential line data you need to understand global mobile, internet, and social media use P7. <https://wearesocial.com/global-digital-report-2019>.
- [2] Dr. Cebrail Taşkın, MBA Chief Technology Officer (CTO) & Chief Digital Officer, YENİ DEĞER: BÜYÜK VERİ, 2017 [https://www.linkedin.com/pulse/yeni-de%C4%9Fer-b%C3%BCy%C3%BCk-veri-dr-cebrail-ta%C5%9Fk%C4%B1n-mba?articleId=6328625124677341184#comments-6328625124677341184&trk=public\\_profile\\_post](https://www.linkedin.com/pulse/yeni-de%C4%9Fer-b%C3%BCy%C3%BCk-veri-dr-cebrail-ta%C5%9Fk%C4%B1n-mba?articleId=6328625124677341184#comments-6328625124677341184&trk=public_profile_post)
- [3] Text mining, Ian H. Witten Computer Science, University of Waikato, Hamilton, New Zealand.
- [4] Simon Kemp, report author, Digital around the world in 2019, the essential line data you need to understand global mobile, internet, and social media use, we are social 2019 report for Turkey <https://wearesocial.com/global-digital-report-2019>
- [5] [www.marketingturkiye.com.tr](http://www.marketingturkiye.com.tr), haberler, 2017 rakamlar, Türkiye’de e-ticaret istatistikleri <https://www.marketingturkiye.com.tr/haberler/rakamlarla-2017-turkiye-e-ticaret-istatistikleri/>
- [6] Akcayol, M & Özyurt, Barış. (2018). FİKİR MADENCİLİĞİ VE DUYGU ANALİZİ, YAKLAŞIMLAR, YÖNTEMLER ÜZERİNE BİR ARAŞTIRMA. Selcuk University Journal of Engineering ,Science and Technology. 6. 668-693. 10.15317/Scitech.2018.160.
- [7] Cothrel J, Williams RL. On-line communities: helping them form and grow. J Knowl Manag. 1999 Mar; 3(1): 54-60
- [8] Hsiung RC. Suggested principles of Professional ethics for the online provision of mental health services. Stud Health Technol inform. 2001;84(Pt 2): 1296-300
- [9] Liu, B., Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
- [10] Akgül, E. S., Ertano, C., & Banu, D. (2016). Twitter verileri ile duygu analizi.

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22(2), 106-110

- [11] Abbé, Adeline. Analyse de données textuelles d'un forum médical pour évaluer le ressenti exprimé par les internautes au sujet des antidépresseurs et des anxyolitiques. Diss. Paris Saclay, 2016.
- [12] Stubbs M. Text and corpus analysis: Computer-assisted studies of language and culture. Blackwell Oxford; 1996.
- [13] Roche E, Schabes Y. Finite-state Language Processing. 1997. (MIT Press).
- [14] <https://www.quora.com/What-is-sentiment-analysis-and-real-life-examples>, Wonderflow BV Panamalaan 6h
- [15] O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2): 183-8.
- [16] Sarasohn-Kahn J. The wisdom of patients: Health care meets online social media. 2008;
- [17] Kontos E, Blake KD, Chou W-YS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. Journal of medical Internet research. 2014;16(7):e172.
- [18] Al-Bnd, A., & Salman, M. (2015). Sentiment analysis and opinion mining via microblogging in social media like: twitter.
- [19] Yengi, Y., & Omurca, S. İ. (2016). Distributed Recommender Systems with Sentiment Analysis Büyük Veride Tavsiye Sistemlerini Duygu Analizi ile Desteklemek. Eur. J. Sci. Technol, 4(7), 51-57.
- [20] NASSAR, A. (2017). MULTILEVEL SENTIMENT ANALYSIS IN ARABIC. Hacettepe University/ Graduate School of Science and Engineering/ Department of Computer Engineering.
- [21] Peng, L., Cui, G., Zhuang, M., and Li, C., "What do seller manipulations of online product reviews mean to consumers? " (HKIBS Working Paper Series 070-1314). Hong Kong: Hong Kong Institute of Business Studies, Lingnan University, 2014.
- [22] Türkmenoğlu, C. (2015). Türkçe Metinlerde Duygu Analizi (Doctoral dissertation, Fen Bilimleri Enstitüsü). İSTANBUL TEKNİK ÜNİVERSİTESİ/ FEN

BİLİMLERİ ENSTİTÜSÜ/ Bilgisayar Mühendisliği Anabilim Dalı.

[23] Yan L, Tan Y. Feel Blue so Go Online: An Empirical Study of Online Supports among Patients. SSRN Electron J [Internet]. 2010 [cited 2016 Apr 20]; Available from:

<http://www.ssrn.com/abstract=1697849>

[24] Machine Learning in Automated Text Categorization, FABRIZIO SEBASTIANI, Consiglio Nazionale delle Ricerche, Italy Received December 1999; revised February 2001; accepted July 2001, ACM Computing Surveys, Vol. 34, No. 1, March 2002.

[25] Compilation for critically constrained knowledge bases robert C. Schrag and Daniel Research Laboratories University of Texas at Austin august 30, 1996. Mitchell

[26] Aligning Grammatical Theories and Language Processing Models, Shevaun Lewis · Colin Phillips, 19 November 2014, Johns Hopkins University

[27] stemming algorithm in Snowball. the definition of the Porter stemmer, Program, Vol 14 no. 3 pp 130-137, July 1980 Porter, Martin F. "Snowball: A language for stemming algorithms." (2001).

[28] PACHUNKE, T., O. MERTINEIT, K. WOTHKE, and R. SCHMIDT, (1994), "The Linguistic Knowledge in a Morphological Segmentation Procedure of German", Computer Speech and Language, vol. 8, pp. 233–245.

[29] Tantuğ, A . "Metin Sınıflandırma". Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi 5 (2016): <<http://dergipark.org.tr/tbbmd/issue/22245/238796>>

[30] SEKER, Sadi Evren. "Duygu Analizi (Sentimental Analysis)." Yönetim Bilişim Sistemleri Ansiklopedisi, Cilt 3, Sayı 3, Eylül 2016: 21.

[31] Grzegorz Dzielkowski. Analyse des sentiments : système autonome d'exploration des opinions exprimées dans les critiques cinématographiques. Automatique / Robotique. École Nationale Supérieure des Mines de Paris, 2008. Français. <NNT : 2008ENMP1637>. <tel-00408754>

[32] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10.

Association for Computational Linguistics, 2002.

[33] Hatzivassiloglou & Wiebe (2000) Hatzivassiloglou, Vasileios, and Janyce M. Wiebe. "Effects of adjective orientation and gradability on sentence subjectivity." Proceedings of the 18th conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 2000.

[34] Cem Akkaya, Janyce Wiebe, Alexander Conrad and Rada Mihalcea (2011). Improving the Impact of Subjectivity Word Sense Disambiguation on Contextual Opinion Analysis. Conference on Computational Natural Language Learning (CoNLL 2011).

[35] Swapna Somasundaran and Janyce Wiebe (2009). Recognizing Stances in Online Debates. In Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, August 2-7, 2009, Singapore (ACL 2009)

[36] Swapna Somasundaran and Janyce Wiebe (2010). Recognizing Stances in Ideological On-line Debates In Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pages 116-124, Los Angeles, CA. Association for Computational Linguistics, 2010 (NAACL-HLT, 2010).

[37] Saifee, V., & Jay, T., "Applications and Challenges for Sentiment Analysis: A Survey", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue2, 2013

[38] Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. Language Resources and Evaluation, 50(3), 667-685.

[49] Kahraman Tazeoğlu, Bukre kitabı, Bazı aşklar aşka ihanettir, Veba busesi bölümü, Sayfa 177. Destek yayını.

[40] OpinionFinder, Jan Wiebe's Natural Language Processing (NLP) Group at the University of Pittsburgh, Copyright (c) 2006 University of Pittsburgh

[41] Seyed, M.M, Mustafa, E., Dincturk, Salman, H., Gregor, V. B., Guy-Vincent, J.,

and Iosif, V.O., "A brief history of web crawlers", CASCON'13 proceedings of the 2013 Conference of the center for advanced studies on collaborative research, pages 40-54, USA, 2013.

[42] Saifee, V., & Jay, T., "Applications and Challenges for Sentiment Analysis: A Survey", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue2, 2013.

[43] "TÜRKÇE METİNLERİN SINIFLANDIRILMASINDA METİN TEMSİL YÖNTEMLERİNİN PERFORMANS KARŞILAŞTIRILMASI", M.Fatih AMASYALI, Sümevra BALCI, Esra Nur VARLI, Emrah METE ,EMO Bilimsel Dergi, Cilt 2, Sayı 4, 95-104, 2012.

[44] Apoorv, A., Boyi, X., Ilia, V., Owen, R., and Rebecca, P., "sentiment Analysis of Twitter Data", proceeding LSM '11 Proceedings of the Workshop on Languages in Social Media, Pages 30-38, USA, 2011.

[45] "Türkçe Sıfat Sözlüğü ile Duygu Analizi", İslam Mayda, Mehmet Korkmaz, Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (ASYU), 4-6 Ekim 2018, Adana.

[46] Naderalvojud, B., Ucan, A., & Sezer, E. A. (2018), "HUMIR at IEST-2018: Lexicon-Sensitive and Left-Right Context-Sensitive BiLSTM for Implicit Emotion Recognition", In Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 182-188) <http://humirapps.cs.hacettepe.edu.tr/tsad.aspx>

[47] Rao, D., & Ravichandran, D., "Semi-supervised polarity lexicon induction". In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. EACL '09, Stroudsburg, PA, USA, Association for Computational Linguistics (2009) 675–682.

## ÖZGEÇMİŞ

**Adres** : MERKEZ MAHALESİ CENGİZ SK. NO: 14-/9 AVCILAR,  
İSTANBUL

**Cep telefonu** : 05079632372

**E-posta** : [issa132@hotmail.com](mailto:issa132@hotmail.com)

### **Kariyer hedefi :**

Çalıştığım şirketin ilerlemesini sağlamak. İşlemi severek iyi bir şekilde yapmak, kendimi de geliştirmek. Okulda öğrendiğim şeyler pratik yaparak en iyi yazımcılardan biri olmak. İşimi ciddi olarak çalışacağım şirkette hep beraber üste çıkarmak, başarıyı imza atmak.

**AD** : ISSA BABAN CHAWAI **SOYAD** : ABDOULAYE

**Eğitim durumu** : lisans  
23/03/89 **Doğum tarihi** :

**Medeni durum** : bekar **Doğum yeri** :  
YAOUNDE

**Toplam tecrübe** : yok **Ehliyet** : yok

**Çalışma durumu** : yok **Ask. durumu** : yok

**Uyruk** : KAMERUN

**Eğitim durumu** :  
Marmara üniversitesi , yüksek lisans bilgisayar  
mühendisliği (2016-2018)

**GANO:** 3.80

**Üniversite** : İSTANBUL TİCARET ÜNİVERSİTESİ, lisans  
bilgisayara  
Mühendisliği 2011-2016)

**GANO:** 3,64

**Lise** : İNTERNATİONAL COMPLEXE "LA GAÏETE"

**Yabancı diller** :

✓ **Türkçe:** **Okuma:** Çok iyi  
**Yazma:** Çok iyi  
**Konuşma:** İyi

✓ **İngilizce:** **Okuma:** İyi  
**Yazma:** Çok iyi  
**Konuşma:** İyi

- ✓ **Fransızca: Okuma:** Çok iyi  
**Yazma:** Çok iyi  
**Konuşma:** Çok iyi

**Mesleki bilgiler :**

**Programlama dilleri** : C#, C, C + +, R Programming  
JAVA,HTML,CSS,JAVASCRIPT,PHP,JQUERY, .NET,  
XML,XAML,WPF,ASP.NET,ANDROİD,ADO.NET,LİNQ

**Yazılımlar** : OpenGL,ACCESS,SQL,MYSQL,MİCROSOFT,  
MATLAB,PAİNT,ADOBE PHOTOSHOP,  
OFFİCE(EXCEL,WORD...), R Studio, WEKA.

**Sertifika bilgileri** :  
- Data Visualization by University of Illinois at Urbana-  
champaign.  
- İletişim Sihirbazı.  
- Eğitim ve sosyo-kültürel semineleri.  
- Tüketilen dünya'nin yenilenebilir sektörü enerji.

**Katıldığım seminerleri** :  
- Network ve haberleşme(Business mind, network  
marketing, bilgi teknolojileri, elektronik haberleşme, network  
aktiviteleri).

**Katıldığım konferanslar** :  
- TASAM(TÜRK ASYA STRATEJİK  
ARAŞTIRMALAR MERKEZİ), 16-17 aralık 2010.  
- GENÇ TÜRKİYE ZİRVESİ 10-11-12 NİSAN 2015.

**Makaleler:**  
- MAMOGRAFİ GÖRÜNTÜLERİNDEN KİTLE  
TESPİTİ AMACIYLA ÖZNİTELİK  
ÇIKARIMI(UBICTUS 2017), academia.edu üzerine  
yayınladı.