



T.C.

AKSARAY ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

WEB KAZIMA VE DUYGU ANALİZİ TEMELLİ ÜRÜN ANALİZ SİSTEMİ

YÜKSEK LİSANS TEZİ

GÖKTUĞ YAVUZ

DANIŞMAN

DR. ÖĞR. ÜYESİ TARIK YILMAZ

AKSARAY 2023

T.C.
AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI

WEB KAZIMA VE DUYGU ANALİZİ TEMELLİ ÜRÜN ANALİZ SİSTEMİ

YÜKSEK LİSANS TEZİ

GÖKTUĞ YAVUZ

DANIŞMAN
DR. ÖĞR. ÜYESİ TARIK YILMAZ

AKSARAY 2023

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren(....) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Göktuğ
Soyadı : YAVUZ
Bölümü : Yönetim Bilişim Sistemleri
İmza :
Teslim tarihi :

TEZİN

Türkçe Adı : Web Kazıma ve Duygu Analizi Temelli Ürün Analiz Sistemi

İngilizce Adı : Web Scraping and Sentiment Analysis Based Product Analysis System

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Göktuğ YAVUZ

İmza:

TC.
AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
JÜRİ ONAY SAYFASI

Göktuğ YAVUZ tarafından hazırlanan “Web Kazıma ve Duygu Analizi Temelli Ürün Analiz Sistemi” başlıklı tez çalışması aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Aksaray Üniversitesi Yönetim Bilişimleri Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: (Dr. Öğr. Üyesi Tarık YILMAZ)

(Yönetim Bilişim Sistemleri ABD, Aksaray Üniversitesi)

İmza

.....

Üye: (Doç. Dr. Aydın KARAKOCA)

(İstatistiksel Bilgi Sistemleri ABD, Necmettin Erbakan Üniversitesi)

.....

Üye: (Doç. Dr. Volkan Soner ÖZSOY)

(Yönetim Bilişim Sistemleri ABD, Aksaray Üniversitesi)

.....

Tez Savunma Tarihi: 22/06/2023

Sosyal Bilimler Enstitüsü Yönetim Kurulu’nun tarih ve
..... sayılı kararı ile onaylanmıştır.

Sosyal Bilimler Enstitüsü Müdürü
Doç. Dr. Sevilay USLU DİVANOĞLU
İmza

TEŞEKKÜR

Bu çalışmanın gerçekleşmesinde yardımlarını esirgemeyen ve yol gösteren danışman hocam Dr. Öğr. Üyesi Tarık YILMAZ' a, her daim yanımda olan nişanlım Maide Nur ER' e ve aileme teşekkür ederim.

AKSARAY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
WEB KAZIMA VE DUYGU ANALİZİ TEMELLİ ÜRÜN ANALİZ
SİSTEMİ

(Yüksek Lisans Tezi)

Göktuğ YAVUZ, Aksaray 2023

ÖZET

Müşteri yorumları, işletmeler için ürün veya hizmet hakkında gerçek zamanlı geri bildirim sağlar ve potansiyel müşterilerin satın alma kararlarını etkiler. İşletmelere olan güven ve inanırlığı geliştirir. Müşterilerin ürün hakkında karar verme sürecine katkıda bulunur ve ürüne yapılan yorumlarla birlikte rekabet avantajı sağlar. Ürünü almak isteyen müşteriler için ise ürün kalitesini ve performansı hakkında ön fikir edinme, ürün hakkında güncel ve doğru bilgilere erişim gibi fırsatları vardır. Bu çalışmada Trendyol üzerinden 15 markaya ait 34007 yorum web kazıma kütüphanesi olan Scrapy ile çekilmiştir. Bu yorumlara sözlük tabanlı duygu analizi kütüphaneleri olan Textblob ve Vader uygulanarak ürünler ve markalar hakkında olumlu, nötr ve olumsuz değerler hesaplanmış, metin özetleme işlemi uygulanarak tüm yorumların anahtar kelimeleri elde edilmiştir. Bu sayede müşterilerin yorum okumasına gerek kalmadan ürünler ve markalar ile ilgili insanların ne düşündüğünün tespiti yapılmıştır. Ayrıca ürünlere verilen yıldız değerleri ile yorumlar analiz edilerek düşük yıldızların yanıltıcı sonuçlar çıkartılabileceği tespit edilmiştir.

Bilim Kodu: 92416

Anahtar Kelimeler: veri, web kazıma, duygu analizi, sözlük tabanlı duygu analizi, metin özetleme

Sayfa Adedi: 77

Danışman: Dr. Öğr. Üyesi Tarık YILMAZ

AKSARAY UNIVERSITY
GRADUATE SCHOOL OF SOCIAL SCIENCES
WEB SCRAPING AND SENTIMENT ANALYSIS BASED PRODUCT
ANALYSIS SYSTEM

(M.A. Thesis)

Göktuğ YAVUZ, Aksaray 2023

ABSTRACT

Customer reviews provide real-time feedback on the product or service and influence potential purchasing decisions of the customers. Develops trust and credibility in businesses. It contributes to the decision-making process of the customers about the product and can provide a competitive advantage with the comments made on the product. For customers who want to buy the product, there are opportunities such as getting a preliminary idea about the quality and performance of the product, and access to up-to-date and accurate information about the product. In this study, 34007 comments from 15 brands on Trendyol were taken with Scrapy, a web scraping library. By applying the dictionary-based sentiment analysis libraries Textblob and Vader to these comments, positive, neutral and negative values were calculated about the products and brands, and the keywords of all comments were obtained by applying the text summarization process. In this way, it has been determined what people think about products and brands without the need for customers to read comments. In addition, by analyzing the star values and comments given to the products, it was determined that the given stars could be misleading.

Science Code: 92416

Key Words : data, web scraping, sentiment analysis, dictionary-based sentiment analysis, text summarization

Page Number: 77

Supervisor: Assist. Prof.Tarik YILMAZ

İÇİNDEKİLER

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU.....	I
ETİK İLKERE UYGUNLUK BEYANI.....	II
JÜRİ ONAY SAYFASI.....	III
TEŞEKKÜR.....	IV
ÖZET	V
ABSTRACT	VI
TABLolar LİSTESİ.....	X
ŞEKİLLER LİSTESİ	XI
SİMGELER VE KISALTMALAR LİSTESİ.....	XIII
1. GİRİŞ.....	1
2. WEB KAZIMA.....	3
2.1. Web Kazıma Literatür Taraması.....	3
2.2. Web Kazıma Nedir?.....	5
2.3. Web Kazıma Çalışma Prensipleri.....	7
2.4. Web Kazıma Teknikleri.....	8
2.4.1. Kopyala Yapıştır.....	9
2.4.2. Metin Yakalama ve Düzenli İfade.....	9
2.4.3. Köprü Metni Aktarım Protokolü (HTTP) Programlama	9
2.4.4. HTML Ayırıştırma.....	9
2.4.5. DOM Ayırıştırma.....	9
2.4.6. Web Kazıma Yazılımı.....	10
2.4.7. Dikey Kümelenme Platformları.....	10
2.4.8. Semantik Açıklama.....	10
2.5. Web Kazıma İçin Kullanılabilecek Python Kütüphaneleri	10
2.5.1. Request Kütüphanesi.....	11
2.5.2. lxml Kütüphanesi.....	11
2.5.3. BeautifulSoup Kütüphanesi.....	12

2.5.4. Selenium Kütüphanesi.....	12
2.5.5. Scrapy Kütüphanesi.....	13
2.6. Web Kazıma Kullanım Alanları.....	13
3. DUYGU ANALİZİ.....	14
3.1. Duygu Analizi Literatür Taraması.....	16
3.2. Duygu Analizi Teknikleri.....	19
3.2.1. Sözlük Tabanlı Duygu Analizi.....	20
3.2.1.1. Sözlük Tabanlı Duygu Analizi Araçları.....	20
3.2.1.1.1. Natural Language Tool Kit.....	20
3.2.1.1.2. Textblob.....	21
3.2.1.1.2.1. Textblob Özellikleri.....	21
3.2.1.1.2.2. Textblob ile Bir Metin Sınıflandırma Sistemi Oluşturma	22
3.2.1.1.3. Vader.....	23
3.2.2. Makine Öğrenmesi.....	24
3.2.2.1. Denetimli Öğrenme.....	24
3.2.2.2. Denetimsiz Öğrenme.....	25
3.2.2.3. Takviyeli Öğrenme.....	26
3.2.2.1.1. Olasılıksal Sınıflandırıcılar.....	26
3.2.2.1.1.1. Naive Bayes Sınıflandırıcısı.....	27
3.2.2.1.1.2. Maximum Entropy Sınıflandırıcısı.....	28
3.2.2.1.1.3. Bayesian Network.....	28
3.2.2.1.2. Lineer Sınıflandırıcılar.....	28
3.2.2.1.2.1. Destek Vektör Makineleri.....	28
3.2.2.1.2.2. Yapay Sinir Ağları.....	29
3.2.2.1.2.3. Karar Ağaçları.....	30
4. METİN ÖZETLEME.....	32
4.1. Yorumlama Tabanlı Metin Özetleme.....	33
4.2. Çıkarımsal Metin Özetleme.....	33
4.3. Denetimli ve Denetimsiz Metin Özetleme.....	34
5. UYGULAMA.....	35
5.1. Materyal ve Yöntem.....	35
5.1.1. Kullanılacak Veri Seti Seçimi.....	35
5.1.2. Web Kazıma Programı.....	35
5.1.3. Veri Tabanı.....	36
5.1.4. Duygu Analizi.....	37

5.1.5. Web Arayüzü.....	37
5.2. Veri Setinin Web Kazıma ile Çekilmesi.....	38
5.3. Veri Tabanı Oluşturma.....	38
5.4. Veri Tabanına Veri Kaydetme.....	38
5.5. Duygu Analizi İşlemleri.....	40
5.5.1. Veri Ön İşleme.....	40
5.5.2. Veri Tabanı İşlemleri.....	41
5.6. Metin Özetleme Adımları.....	43
5.6.1. Veri İşlemleri.....	43
5.6.2. Veri Ön İşleme.....	43
5.6.3. Metin Özetleme.....	43
5.7. Django ile Sonuçların Gösterimi.....	44
6. BULGULAR.....	47
6.1. Alışveriş Yapan İnsanların Markaya ve Ürüne Göre Yaptığı Yorum Arasındaki Duygu Analizi Bulguları.....	47
6.1.1. Markaya Göre Yapılan Çalışmalar.....	47
6.1.2. Ürüne Göre Yapılan Çalışmalar.....	48
6.2. Alışveriş Yapan İnsanların Ürünlere Verdikleri Yıldızlara Göre Metin Özetleme Bulguları.....	49
6.3. Alışveriş Yapan İnsanların Verdiği Yıldız ile Yaptığı Yorum Arasındaki Duygu Analizi Bulguları.....	50
7. SONUÇLAR VE ÖNERİLER.....	52
8. KAYNAKÇA.....	53
AKSARAY ÜNİVERSİTESİ TEZ DEĞERLENDİRME FORMU.....	61

TABLÖLAR LİSTESİ

Tablo 1. Textblob Kütüphanesi ile Örnek Yorumlar.....	22
Tablo 2. Vader Kütüphanesi ile Örnek Yorumlar.....	23
Tablo 3. Textblob ve Vader’e Göre Yorumu En Olumlu Olan Markalar.....	47
Tablo 4. Textblob ve Vader’e Göre Yorumu En Olumsuz Olan Markalar.....	48
Tablo 5. Verilen Yıldızlara Göre Metin Özetleme Sonuçları Tablosu.....	49
Tablo 6. Olumlu Yorum Yapanların Duygu Analizi Sonucu.....	51
Tablo 7. Olumsuz Yorum Yapanların Duygu Analizi Sonucu.....	51

ŞEKİLLER LİSTESİ

Şekil 1. Duygu Analizi Teknikleri.....	19
Şekil 2. Denetimli Öğrenme.....	25
Şekil 3. Denetimsiz Öğrenme	26
Şekil 4. Yapay Sinir Hücresi	29
Şekil 5. Tek Katmanlı Algılayıcı Model.....	30
Şekil 6. Çok Katmanlı Algılayıcı Model	30
Şekil 7. Karar Ağaçları.....	31
Şekil 8. Otomatik Metin Özetleme Türleri.....	32
Şekil 9. Scraping.....	36
Şekil 10. SQLite Arayüzü.....	37
Şekil 11. Django Arayüzü.....	37
Şekil 12. Oluşturulan Veri Tabanı.....	38
Şekil 13. Markalar Tablosu.....	39
Şekil 14. Ürünler Tablosu.....	39
Şekil 15. Ürün Yorumları.....	40
Şekil 16. Emojili Yorumlar.....	40
Şekil 17. Yanlış Yazımı Olan Yorumlar-1.....	41
Şekil 18. Yanlış Yazımı Olan Yorumlar-2.....	41
Şekil 19. Yanlış Yazımı Olan Yorumlar-3.....	41
Şekil 20. Yanlış Yazımı Olan Yorumlar-4.....	41
Şekil 21. Veri Tabanı Şekli.....	42
Şekil 22. Duygu Analizi (Textblob) Sonuçları İşlenmiş Yorumların Şekli.....	42
Şekil 23. Duygu Analizi (Vader) Sonuçları İşlenmiş Yorumların Şekli.....	43
Şekil 24. Navbar Arama Alanı.....	44
Şekil 25. Django ile Arama Yerinde Yapılan Ürün Sonuçları.....	44

Şekil 26. Django Ürünler Sayfası.....	45
Şekil 27. Django ile Duygu Analizi Sonuçlarının Gösterimi.....	45
Şekil 28. Duygu Analizi ve Kullanıcı Yorumlarının Ortalama Değerleri	46
Şekil 29. Alışveriş Yapan İnsanların Verdiği Yıldız ile Yaptığı Yorum Arasındaki Duygu Analizi Sonuçları.....	50



SİMGELER VE KISALTMALAR LİSTESİ

API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag Of Words
CSS	Cascading Style Sheets
DOM	Document Object Model
DVM	Destek Vektör Makineleri
GRU	Gated Recurrent Units
HTML	HyperText Markup Language
HTTP	Hyper-Text Transfer Protocol
LR	Logistic Regresyon
LSTM	Long Short-Term Memory
ML	Makine Öğrenimi Yaklaşımı
NB	Naive Bayes
NLTK	Natural Language Tool Kit
PHP	Hypertext Preprocessor
SVM	Support Vector Machine
TF-IDF	Term Frequency Inverse Document Frequency
UNIX	Uniplexed Information and Computing System
URI	Tekdüzen Kaynak Tanımlayıcısı
URL	Uniform Resource Locator
W3C	The World Wide Web Consortium
XHTML	Extensible HyperText Markup Language
XML	Extensible Markup Language
YSA	Yapay Sinir Ağları

1. GİRİŞ

E-ticaret, internet üzerinden ürün ve hizmetlerin alım satımının yapıldığı elektronik ticaret anlamına gelir. Geleneksel mağaza yerine online platformlar üzerinden ticaret yapılmasını sağlar. E-ticaret, kullanıcılara geniş bir ürün yelpazesi sunar ve alışveriş deneyimini kolaylaştırır. Tüketiciler, istedikleri ürünleri çevrimiçi olarak araştırır, fiyatları karşılaştırır ve güvenli bir şekilde satın alırlar. Ayrıca, satıcılar da çevrimiçi mağaza açarak ürünlerini daha geniş bir kitleye ulaştırarak iş hacimlerini artırma imkânı bulurlar.

Covid-19 salgını, e-ticaret sektöründe büyük bir etki yaratmıştır. Salgının başlamasıyla birlikte birçok ülkede, sosyal mesafe önlemleri ve kısıtlamaları nedeniyle geleneksel mağaza alışverişlerinde azalma yaşamıştır. Bu durum, tüketicilerin çevrimiçi alışverişe yönelmesine ve e-ticaretin önemini artırmasına neden olmuştur.

Covid-19 salgını, 2019 Aralık ayında başlamıştır. Covid-19 sonrası yayılan salgın sebebiyle birçok ülkede karantina uygulamaları başlamıştır. İnsanlar kalabalık ortamlardan izole edilerek daha çok evlerinde yaşamsal faaliyetlerini yapmak zorunda kalmışlardır. (Savrul ve Akın, 2021)

Bu dönemde okullarda uzaktan eğitim, iş yerlerinde uzaktan çalışma gibi yeni normaller hayata girmiştir. Bu sebeple insanlar mobil bağlantı, cep telefonu ve bilgisayar gibi dijital araçlara yönelmiştir. Salgın boyunca internetin kullanımı dikkat çekici bir şekilde artmış, bu süreçte çok daha fazla insan çevrimiçi iş yaparak daha fazla zaman geçirmiştir.

Bu dönemde e-ticaret sitelerinden alışveriş yapan insan sayısı ve e-ticaret sitelerinden satış yapan işletme sayısında da son derece artış olmuştur. Pandemi döneminde e-ticaret faaliyetinde bulunan işletme sayısında %275 oranında artış görülmüştür. E-ticaret hacmi pandemiden önceki sene 136 milyar lira iken, pandemi yılında 226.2 milyar liraya çıkmıştır. 2019 yılında internet sitelerinden verilen sipariş sayısı 1.36 milyar adet iken, 2020 yılında 2.29 milyar adete ulaşmıştır. (Savrul ve Akın, 2021)

E-ticaret sitesinden alışveriş yapan kişiler aldıkları ürünlerle ilgili yorumlarını da paylaşmaktadır. Bu tür uygulamaların kullanılması ile birlikte çok büyük veriler oluşmakta

ve bu veriler büyümektedir. Ham verilerin genişlemesi ile birlikte bu verilerden anlam çıkarılması gibi yeni problemler ortaya çıkmıştır. (Toçoğlu ve Alpkoçak, 2018)

Bu çalışmada bir e-ticaret sitesinden alışveriş yaparken diğer insanların tecrübesinden yararlanılıp ürünler hakkında duygu analizi ve metin özetleme gibi işlemler yapılarak müşterilerin yorumları okumadan zaman tasarrufu sağlayarak hızlı bir biçimde karar vermesine yardımcı olmak hedeflenmiştir.



2. WEB KAZIMA

Günümüzde teknolojinin gelişmesi ve hızlanmasıyla birlikte web kazıma (web scraping) ile ilgili çalışmalar da son derece hızlı bir şekilde artmıştır. Verilerin çok fazla artmasıyla birlikte anlamlı veri çıkartmak çok önemlidir. Bu yüzden web kazıma yöntemleri de son yıllarda oldukça popüler hale gelmiştir. (Mutlu, 2021)

2.1. Web Kazıma Literatür Taraması

Vargiu ve Urru (2013), web kazımadan yararlanarak bir web sayfası için en alakalı reklamları bulmaya çalışan bir web reklam sistemi çalışması gerçekleştirmiştir.

Julian ve Natalia (2015), bilgisayar parçalarını web kazıma kullanarak fiyat karşılaştırılmasını yapmıştır. Bilgisayarların sahip olduğu birçok farklı işlev gibi, işlevlerine göre görevlerini yerine getirebilmeleri için her bilgisayar için farklı özelliklere de ihtiyaçları vardır. Bu nedenle bu uygulama, kullanıcılarına ihtiyaçlarına uygun hesaplamaları bir araya getirme konusunda bir çözüm önermek amacıyla oluşturulmuştur. Bu uygulama aynı zamanda beş bilgisayar mağazasından alınan veri kaynaklarına dayanan bir fiyat karşılaştırma özelliğine sahiptir, böylece kullanıcılar bilgisayar parçaları satın alma ve bilgisayarı daha kolay birleştirme maliyetlerinden tasarruf edebilirler. Bu karşılaştırma özelliği, temel olarak ürünleri yalnızca en düşük fiyatla satın almak istemeyen, aynı zamanda mümkün olan en iyi kaliteyi de bekleyen temel bir tüketici ilkesine dayanmaktadır.

Amalia, Afifa ve Herriyance (2018), Endonezya’da bulunan tropikal hastalıklar için çalışma yapmıştır. Tropikal hastalıklara karşı Endonezya halkının hastalıklar ve ilaçlar hakkında bilgi almak için arama motorlarına güvendiği tespit edilmiştir. Fakat arama motorları, alakasız bazı sonuçlar çıkartmaktadır. Bu soruna çözüm olarak web kazıma teknolojileri kullanılarak ontoloji ile terimler arasındaki ilişki oluşturulmuştur ve istenilen sonuca ulaşılmıştır.

Sundaramoorthy, Durga ve Nagadarshini (2017), NewsOne adında ulusal ve uluslararası kaynaklardan gelen en son haber bilgilerini bir araya getiren ve bunları kısa ve öz biçimde sunan özet bir platform geliştirmiştir. Bu uygulamanın sloganı haberlere hızlı bir şekilde erişmektir. NewsOne, içeriği çeşitli haber sitelerinden çıkarmak için web kazıma kullanır.

Junjoewong, Sangnapachai ve Sunetnanta (2018), insanların promosyon veya kampanyalar için farklı uygulamalardan araştırma yapmak zorunda kaldığını tespit etmiştir. Yapılan ProCircle adlı uygulama sayesinde bilgilendirme haberlerini tek bir sayfadan takip edilmesini sağlamışlardır.

Ertam (2018), bir Türk haber ajansı sitesinde yer alan kategorik haber başlıkları ve özetleri web kazıma yöntemleri kullanılarak toplanmış ve test verileri “one hot encoding” yöntemi ile vektör öğrenme yöntemleri ve derinlemesine öğrenme yöntemleri kullanılarak sınıflandırılmıştır. %90 sınıflandırma doğruluğu elde edilmiştir.

Verma, Mittal ve Dawn (2019), yalan haberlerin tespitini yapabilmek için derin öğrenme ve doğal dil işleme yardımıyla sahte bilgileri ve haberleri tespit eden bir model geliştirdi. Karışıklık matrisi ve sınıflandırma raporları, LSTM modeli kullanılarak yüzde 94'e varan doğruluk puanına ulaşarak yalan haberlerin tespitine yardımcı olmuştur.

Chaudhari, Aparna, Tekkur, Pavan ve Karki (2020), web kazıma işlemini sağlıklı yaşam için kullanmıştır. Yapılan çalışmada bir yemekle ilgili tüm tarifleri web kazıma ile aynı formda ortaya çıkarmıştır. Bu sayede insanlar yemek yaparken farklı tariflere de aynı sayfada bakabilmiştir ve yemekte kullanılan ürünlerin yan etkileri anlatılmıştır ve sağlıklı diyet yemeklerinin tariflerini veri tabanına kaydederek akıllı şef adında bir mobil uygulama geliştirilmiştir.

Krotov, Johnson ve Silva (2020) web kazımanın yasallığı ve etiği çalışmasında web kazıma işlemi yapan kişilere yönelik nelerin yasal nelerin etik olduğunu gösteren bir çalışma yapmıştır.

Bozkuş (2021), web kazıma metotlarını kullanarak bir siber tehdit istihbaratı aracı geliştirmeyi hedeflemiştir. Yapılan çalışma sonucunda geliştirilen uygulama ile açık kaynakla veriler toplanarak API kullanıp verilen kısıtlamalara maruz kalmadan web kazıma ile siber tehdit istihbaratı sağlanabileceği görülmüştür.

Dobashi (2021), modül kullanılarak yapılan öğrenme analizinde, öğretmenlerin işlem yüklerini azaltmak ve ders esnasında analiz sonuçlarını daha hızlı elde edebilmesi için web

kazıma ile gerçek zamanlı işleme önermiştir. Geliştirilen web kazıma fonksiyonu ile modül ders notlarının otomatik olarak indirilmesi sağlanmıştır. Ayrıca, daha önce geliştirilen ön işleme sistemiyle entegre olarak öğretmenler, programı ders sırasında başlatarak öğrencinin materyalinin tıklama akışını gerçek zamanlı olarak bir zaman serisi kesitinde görebilirler. Ayrıca, web kazıma teknolojisinin kullanılmasının, öğrenme günlüklerini çeşitli açılardan gerçek zamanlı olarak analiz etme ve otomatikleştirme olasılığını da arttırmıştır.

Aydın (2022), Türk dizileri üzerinde makine öğrenmesi algoritmaları kullanarak duygu analizi yapmayı amaçlamaktadır. Öncelikle Ekşi Sözlük adındaki web sayfasından diziler hakkındaki yorumlar web kazıma teknikleri ile toplamıştır ve bu verilere ön işleme uygulanmıştır. Daha sonra destek vektör makineleri, lojistik regresyon (LR), K- en yakın komşu algoritması makine öğrenme teknikleri ile kelime çantası modeli, terim frekansı ve ters belge frekansı, Word2vec vektör modelleri ile duyarlılık analizi yapılmıştır. TF-IDF vektör modeli üzerinde eğitilen SVM sınıflandırıcısının 0.631 ile en yüksek tahmin makro ortalamalı F-skorunu vermiştir.

Camargo-Henríquez ve Bernal (2022), Instagram'dan veri çıkartmak için web kazıma tabanlı bir yaklaşım ortaya koymuştur. Bu çalışmayla birlikte insanlar Instagram'dan veri kaynağı olarak yararlanabilirler.

Chiapponi vd. (2022), e-ticaret siteleri ve web kazıyıcılar arasındaki savaştan bahsetmiştir. E-ticaret web sitesi operatörlerinin kendilerini nasıl savunmaya çalıştıklarından ve web kazıyıcılardan bahsedilmiştir.

Raj N, Raj S, Vijayalakshmi (2023), web kazıma tekniklerinin inceler ve bu tekniklerin performansını analiz edip ve sonuçların istatistiksel önemini ortaya çıkartmaktadır.

L.Reddy, B.Reddy, Anil, Mohanty ve Basit (2023), Destek vektör regresyonu ve karar ağacı kullanarak e-ticaret sitesinden alınan verilere göre dizüstü bilgisayar alma maliyetini tahmin etmeyi amaçlamıştır.

2.2. Web Kazıma Nedir?

Günümüzde internet kullanıcısı ve veri sayısı her geçen gün çok hızlı bir şekilde artmaktadır. Hemen hemen herkesin internete katılması ve onu kullanmaya başlaması ile birlikte ağ güçlendirmek için birçok yeni teknik ortaya çıkmıştır. Güçlenen ağ yapıları ve yeni teknolojilerin ortaya çıkması sonucunda maliyetler azalmış ve çok sayıda kullanıcı bir araya

gelmiştir ve çoğu insan internetin olanaklarını kullanmaya başlamıştır. (Gode, Saurkar ve Pathare, 2018)

İnternetin günlük kullanımında çok fazla veri mevcuttur. Ne türde olursa olsun veriler herhangi bir çalışmanın en önemli parçasıdır. (Sirisuriya, 2015)

Önceden çoğu kullanıcı bu verileri alabilmek için geleneksel olarak kopyala yapıştır tekniğini kullanmaktaydı. Fakat verilerin çoğalmasıyla birlikte bu teknik yetersiz kalmıştır. İnternette veri alışverişi, paylaşma ve depolamanın bir sonucu olarak, bu aşırı veri yükünün nasıl ele alınacağı ve kullanıcının en az çabayla en iyi bilgiye nasıl ulaşacağı veya erişeceği gibi yeni bir sorun ortaya çıkmaktadır. Bu sorunları çözmek için, web kazıma kavramı ortaya çıkmıştır. Web kazıma, web üzerinde mevcut yapılandırılmamış veriler temelinde yapılandırılmış veriler oluşturmak için kullanılan bir tekniktir. (Gode vb., 2018)

İnternette otomatik veri toplama aslında internet kadar eski bir kavramdır. Web kazıma yeni bir terim olmasa da geçmiş yıllarda daha çok ekran kazıma, veri madenciliği olarak biliniyordu. Web kazıma, web sitelerinden veri çekme veya kazıma işlemidir. Web hasadı olarak da bilinir. Çeşitli yazılım ve metotlar ile otomatik işlemler kullanılarak istenilen bilgileri bir web kaynağından çıkarma olarak da tanımlanabilir. Otomatik bilgi toplama anlamına da gelir. Web kazıma, çevrimiçi olarak mevcut olan büyük miktardaki veriye hızlı ve kolay erişim sağlar. Bazen de bot olarak tanımlanabilir. Botlar ilgili sitelere göz atıp ve işe yarayacak verileri almak için kullanılır. Bu verilerin veri analitiği anlamında fazlaca kullanımları vardır. (Divya, Rastogi, Yadav ve Perwej, 2022)

Web kazıma, birden çok web sitesinden verileri tek bir tabloya veya veri tabanına çıkarır, böylece verileri analiz etmeyi ve veri görselleştirmeyi kolaylaştıran bir tekniktir. (Sirisuriya, 2015)

Web kazıma, web sitelerinden bir şey ifade etmeyen verileri çıkarmak ve bu verileri anlamlı hale dönüştürmek için kullanılan önemli bir tekniktir. Web kazıma, web verisi çıkarma, web veri kazıma, web toplama veya ekran kazıma olarak da tanımlanır. Web kazıma, bir veri madenciliği şeklidir. Web kazıma işleminin temel ve önemli amacı, farklı ve yapılandırılmamış web sitelerinden gelen bilgileri çıkarmak ve bunları elektronik tablolar, veri tabanı veya virgülle ayrılmış değerler (CSV) dosyası gibi anlaşılır bir yapıya dönüştürmektir. Ürün fiyatlandırması, ürün yorumları, stok fiyatlandırması, farklı raporlar, piyasa fiyatları ve ürün detayları gibi veriler web kazıma yoluyla toplanabilir. (Sirisuriya, 2015)

Kazıma, web sayfalarından bilgi kırmak için kullanılan bir tekniktir. Web sayfaları, HTML ve daha yakın zamanda XML tabanlı XHTML ile yazılmış belgelerdir. Web kazıma işlemi manuel olarak kopyala yapıştır işlemi gibi gözükür. Web kazımanın farkı bu işin bir sanal bilgisayar aracısı tarafından organize ve otomatik bir şekilde yapılmasıdır. (Saurkar vd., 2018)

İnternette bilgi erişmenin tek yolu sadece web tarayıcısından aramak değildir. Tarayıcılar JavaScript'i yürütmek, görüntüleri görüntülemek ve nesneleri insanlar tarafından daha okunabilir bir biçimde düzenlemek için kullanışlı olmasına rağmen web kazıyıcıları, büyük miktarda veri toplama işlemi konusunda çok iyidir. Bir monitörün dar penceresinden her seferinde bir sayfayı görüntülemek yerine web kazıma ile her seferinde aynı anda binlerce hatta milyonlarca sayfaya yayılan veri tabanlarını görüntülenebilir. Ayrıca web kazıyıcıları arama motorlarının getiremeyeceği sonuçları getirebilir. (Mitchell, 2015)

Normal şartlarda web kazıma pek gerekli olmaz. Çünkü bazı web siteleri verilerini yapılandırmış bir biçimde paylaşmak için bir API sağlar. Fakat bu API'ler bazı verileri kısıtlar ya da bazı verilere ne sıklıkta erişilebileceğine engeller koyar. Yani aslında ulaşılabilecek veriler için API'lere çok fazla güvenilmez. Ek olarak, bir web sitesi geliştiricisinin ana önceliği, her zaman arka uç API'si üzerinden ön uç ara yüzünü korumak olacaktır. Kısacası, istenen çevrimiçi verilere erişmek için API'lere güvenilemez ve bu nedenle web kazıma tekniklerinin kullanılması gerekir. (Mitchell, 2015)

2.3. Web Kazıma Çalışma Prensipleri

Web kazıma bir bilgiyi kopyalayıp başka bir yere yapıştırma işlemi gibi görülür. Fakat web kazıma, veriyi organize olarak otomatik ve hızlı bir şekilde yapar. Bu işlemi gerçekleştirirken aslında bir insanın web sitesiyle etkileşim kurarken yaptığı işlemlerin aynısını yapar. Birçok web sayfasında gezinerek bağlantıları izler ve formlar gönderebilir. Bir sonraki adımda ise ayrıştırıcının, önceki adımda alınan verilere dayalı olarak istenen bilgileri almak için kullanıcı tarafından belirlenen yolları izler. Bu yollar CSS seçicileri ve XPATH'ler tarafından belirlenir. Ayrıştırıcıyı web sayfasındaki öğeye yönlendirmek için bazı yollar kullanılır. Ulaşılan bu verileri düzenli olarak bir veri tabanına kaydedebilir ya da istenilen şekilde kullanılabilir. (Saurkar vd., 2018)

Web kazıma için öncelikle bir program ya da yazılım gerekir. Web kazıma işleminde bir URL'ye ihtiyaç vardır. Web kazıma kodu çalıştığında, belirtilen URL'ye bir GET isteği

gönderilir. Gönderilen isteğe cevap olarak, sunucu verileri gönderir ve HTML veya XML sayfasının okunmasına izin verir. Kod daha sonra HTML veya XML sayfasını parçalar. İstenilen verileri bulur ve parçalar. Ardından temizlenen veriler, veri tabanlarında veya istenilen dosya formatında saklanabilir. (Chandra ve Varanasi, 2015)

Bir web kaynağının kazınması, sayfalar tarafından her zaman izin vermez. Bazı şirketler kendilerine karşı bot kullanımına kısıtlama getirir. Kazıma yaparken belirli kurallara uymak gerekir. Web kazıma işleminin yapılması ve yapılmaması gerekenler; (Chandra ve Varanasi, 2015)

- **Şartlar ve koşullar:** Kazıma başlamadan önce akla gelmesi gereken ilk şey şartlar ve koşullardır. Web sitesinin şartlar ve koşullar sayfasını ziyaret edilip sitelerinden kazımayı yasaklayıp yasaklamadıklarını öğrenmektir. Eğer yasak varsa veri çekmemek gerekir. (Chandra ve Varanasi, 2015)

- **Sunucuyu çok sayıda istek göndermek:** Her web sayfası yalnızca belirli miktarda iş yüküne hizmet edebilen bir sunucu üzerinde işlem yapar. Sunucuyu belirli bir zaman diliminde çok sayıda istekle bombardımana tutmak sunucuyu zor duruma sokabilir, bu da ciddi arızalara neden olabilir. Sunucuyu aynı anda çok sayıda istekle bombardıman etmek yerine istekler arasında biraz zaman olması gerekir. Bazı siteler çok istek sayısına bir kısıtlama getirir. Daha fazla devam ederse istek gönderenin IP adresini yasaklayabilir. (Chandra ve Varanasi, 2015)

- **Web kaynağını zaman zaman incelemek:** Bir web sitesi her zaman aynı kalmaz.

Kullanılabilirliği ve kullanıcıların gereksinimlerine göre zaman zaman değişme eğilimindedir. Web sitesinde herhangi bir değişiklik yapılmışsa, kazıma kodu başarısız olabilir. Sitede yapılan değişiklikleri izlemeyi, sıyırıcı komut dosyasını değiştirmeyi ve buna göre kazımayı ihmal etmemek gerekir. (Chandra ve Varanasi, 2015)

2.4. Web Kazıma Teknikleri

Programlama dili ve uygulamalar ile birlikte birçok farklı teknik ve işlem ile web kazıma işlemi gerçekleştirilebilmektedir. Web kazıma tekniği, web madenciliği teknolojisinin bir alt dalıdır. Web kazıma teknikleri kullanılarak, birden fazla web sitesinde bulunan bilgiler tek bir veri tabanına veya elektronik tablolara çıkartılır. Böylece veriler daha fazla kullanım için kolayca görselleştirilip analiz edilebilir. Kopyala-yapıştır, metin yakalama ve düzenli

ifade eşleştirme, HTTP programlama, HTML ayrıştırma, DOM ayrıştırma, web kazıma yazılımı, dikey toplama platformları, semantik açıklama tanıma web kazıma için kullanılan yaygın tekniklerden bazılarıdır. (Sirisuriya, 2015)

2.4.1. Kopyala Yapıştır

İnsanın elle kopyala yapıştır tekniği ile yaptığı basit ve uygulanabilir web kazıma teknolojisidir. Ancak bu, uygulayıcının hatalara neden olma eğilimi ve kullanıcının çok sayıda veri kümesini analiz etmesi ve depolaması gerektiğinde pek de işe yarayacak bir teknik değildir. (Khder, 2021)

2.4.2. Metin Yakalama ve Düzenli İfade

Web sayfalarından bilgi çıkarmak için basit ve güçlü bir yaklaşımdır. Bu teknik, UNIX komutuna veya programlama dilinin düzenli ifade eşleştirme olanaklarına dayanır. (Khder, 2021)

2.4.3. Köprü Metni Aktarım Protokolü (HTTP) Programlama

HTTP programlama ile kullanıcı statik ve dinamik web sayfalarından bilgi alır. Soket programlama kullanılarak uzak web sunucusuna HTTP istekleri gönderilerek veriler alınır. (Khder, 2021)

2.4.4. HTML Ayrıştırma

XQuery ve Hyper Text Query Language (HTQL) gibi yarı yapılandırılmış veri sorgulama dilleri, HTML sayfalarını parçalamak, sayfa içeriğini almak ve dönüştürmek için kullanılır. (Khder, 2021)

2.4.5. DOM Ayrıştırma

Internet Explorer veya Mozilla tarayıcı denetimi gibi bir web tarayıcısını gömerek, programlar, istemci tarafı komut dosyaları tarafından oluşturulan dinamik içeriği alır. Bu

tarayıcı denetimleri ayrıca, hangi programların sayfaların bölümlerini alabileceğine bağlı olarak web sayfalarını bir DOM ağacına ayrıştırır. (Khder, 2021)

2.4.6. Web Kazıma Yazılımı

Günümüzde web kazıma çözümlerini özelleştirmek için kullanılabilen birçok yazılım aracı mevcuttur. Bu yazılım, bir sayfanın veri yapısını otomatik olarak tanımaya çalışır veya web-kazıma kodunu manuel olarak yazma gerekliliğini ortadan kaldıran bir kayıt ara yüzü veya içeriği çıkarmak ve dönüştürmek için kullanılabilecek bazı komut dosyası oluşturma işlevleri ve depolayabilen veri tabanı arayüzleri sağlar. (Khder, 2021)

2.4.7. Dikey Kümelenme Platformları

Dikey özel hasat platformları geliştiren birkaç şirket vardır. Bu platformlar, doğrudan insan katılımı olmadan ve belirli bir hedef siteyle ilgili hiçbir çalışma olmadan belirli dikeyler için çok sayıda "bot" oluşturur ve izler. Hazırlık, tüm dikey için bilgi tabanının oluşturulmasını içerir ve ardından platform, botları otomatik olarak oluşturur. Platformun sağlamlığı, bilgilerin kalitesiyle ölçülür. Genellikle alan sayısı ve ölçeklenebilirliğini (yüzlerce veya binlerce siteye ne kadar hızlı ölçeklenebildiğini) alır. Bu ölçeklenebilirlik çoğunlukla, yaygın toplayıcıların içerik toplamak için karmaşık veya çok emek yoğun bulduğu sitelerin uzun kuyruklarını hedeflemek için kullanılır. (Sirisuriya, 2015)

2.4.8. Semantik Açıklama

Sökülen sayfalar, belirli veri parçacıklarının yerini belirlemek için kullanılabilen meta verileri veya anlamsal işaretlemeleri ve ek açıklamaları kapsar. Ek açıklamalar, Microformat'ın yaptığı gibi sayfalara gömülüyse, bu teknik, DOM ayrıştırmanın özel bir durumu olarak görülür. Başka bir durumda, semantik bir katmanda düzenlenen ek açıklamalar web sayfalarından ayrı olarak depolanır ve yönetilir, böylece kazıyıcılar sayfaları kazımadan önce bu katmandan veri şemasını ve talimatları alır. (Sirisuriya, 2015)

2.5. Web Kazıma İçin Kullanılabilecek Python Kütüphaneleri

Web kazıma için çok sayıda kullanılabilecek Python kütüphanesi mevcuttur. Bunlar;

2.5.1. Request Kütüphanesi

Web tarayıcısı bir web sunucusuyla iletişim kurmaya çalıştığında, bu, istek-yanıt protokolü olarak işlev gören HTTP kullanılarak yapılır. Bu iletişim sürecinde web sunucusuna bir istek gönderir ve karşılığında bir yanıt beklenir. Request ile web istekleri kolaylıkla yönetilebilir. HTTP istek yöntemlerinden bazıları: (Chandra ve Varanasi, 2015)

GET: Bu, verilen URI'yi kullanarak verilen sunucudan bilgi getirir.

HEAD: Bunun işlevi GET'e benzer, ancak fark, yalnızca durum satırını ve başlık bölümünü iletmesidir.

POST: İşlemek istenen sunucuya veri gönderir.

PUT: Yeni bir URL oluşturmak istendiğinde, hedef kaynağın tüm mevcut temsillerini oluşturur veya üzerine yazar.

DELETE: Tüm kaynakları kaldırır.

OPTIONS: Bir istek/yanıt döngüsü için iletişim seçeneklerini belirtir.

Request URI: Tekdüzen Kaynak Tanımlayıcısı (URI), kaynağın adını tanıma yeteneğine sahiptir. URL'lerin açılmasına yardımcı olan yaygın olarak kullanılan bazı Python modülleri vardır.

- `httplib2`: Bu, kapsamlı bir HTTP istemci kitaplığıdır. Diğer HTTP kitaplıklarının dışında kalan birçok özelliği destekler. Önbelleğe alma, canlı tutma, sıkıştırma, yönlendirmeler ve birçok kimlik doğrulama türü gibi özellikleri destekler.

- `urllib2`: Bu, karmaşık bir dünyada HTTP URL'lerini getirmek için yaygın olarak kullanılan bir modüldür. Temel ve özet kimlik doğrulama, yeniden yönlendirmeler, tanımlama bilgileri vb. gibi URL eylemlerine yardımcı olan işlevleri ve sınıfları tanımlar.

Requests: Python'da yazılmış, üretkenlikle sonuçlanacak birçok yetenekle donatılmış, Apache2 lisanslı bir HTTP kitaplığıdır. (Chandra ve Varanasi, 2015)

2.5.2. lxml Kütüphanesi

HTML ve XML ayrıştırma yapmayı sağlayan yüksek performanslı bir Python kütüphanesidir. Tüm Python API'leri ile uyumludur. Dinamik ve karmaşık sitelerdeki verilerin kazanmasında son derece başarılıdır. Fakat kötü tasarlanmış HTML ile iyi çalışmaz.

Web kazıma işleminde Request ve lxml kombinasyonu çok yaygındır. Ayrıca, XPath ve CSS seçicileri kullanarak Html'den veri ayıklamaya olanak sağlar. Diğer ayrıştırıcıların çoğundan hızlıdır. (Richter, 2020)

2.5.3. BeautifulSoup

BeautifulSoup kütüphanesi, adını Alice Harikalar Diyarındaki aynı adı taşıyan bir Lewis Carroll şiirinden almıştır. Web kazıma için kullanılan en yaygın kütüphanedir. HTML ve XML belgelerini ayrıştırmak için kullanılan bir kitaplıktır. Verileri kolayca çıkarmaya yardımcı olmak için ayrıştırma ağaçları oluşturur. BeautifulSoup kütüphanesi varsayılan bir Python kütüphanesi olmadığından, yüklenmesi gerekir. BeautifulSoup, web sayfalarının DOM ağacındaki bilgileri ayıklamak, analiz etmek ve düzenlemek için kullanılan HTML/XML analiz motorunun temeline dayanan bir Python veri tabanıdır. Bir dizi özlü DOM ziyaretçi arabirimi sağlayarak geliştiricilerin hızla bir sistem prototipi oluşturmaya ve deney verilerini almasına yardımcı olur. Ek olarak, yüksek platformlar arası esnekliğe sahiptir. (Zheng, He ve Peng, 2014)

2.5.4. Selenium Kütüphanesi

Selenium, tarayıcıları otomatikleştirmek için bir dizi araçtır. Büyük ölçüde uygulamaları test etmek için kullanılır, ancak kullanımları yalnızca test etmekle sınırlı değildir. Ayrıca, bir tarayıcı penceresinde ekran kazıma ve tekrarlayan görevleri otomatikleştirme için de kullanılır. Selenium, Firefox, Internet Explorer, Google Chrome, Safari ve Opera dahil olmak üzere tüm büyük tarayıcılarda otomasyonu destekler. Selenium WebDriver W3C standartlarının bir parçasıdır ve büyük tarayıcılar tarafından desteklenmektedir. (Gundecha, 2014)

Selenium IDE: Selenium IDE, Selenium komut dosyalarını Firefox ile kaydetmek ve oynatmak için kullanılan bir Firefox eklentisidir. Firefox kullanarak kullanıcı eylemlerini kaydetmek için bir grafik kullanıcı arayüzü sağlar. Selenium'u öğrenmeye ve kullanmaya başlamak için kullanışlı bir araçtır, ancak yalnızca Firefox ile kullanılabilir ve diğer tarayıcılar desteklenmez. Kaydedilen komut dosyalarını, Firefox dışındaki tarayıcılarda komut dosyalarını çalıştırmayı destekleyen Selenium WebDriver tarafından desteklenen çeşitli programlama dillerine dönüştürebilir. (Gundecha, 2014)

Selenium WebDriver: Bu, programlama dillerini kullanarak gelişmiş Selenium betikleri geliştirmek için bir programlama arayüzüdür. Ayrıca birden çok işletim sisteminde Selenium tarafından desteklenen birden çok tarayıcıda testler çalıştırılabilir. Bu, Selenium'u gerçek bir tarayıcılar arası test aracı yapar. Selenium WebDriver, Java, C#, Python, Ruby, PHP ve JavaScript dahil olmak üzere çeşitli dillerde istemci kitaplıkları sunar ve daha çok test komut dosyaları yazmaya yöneliktir. (Gundechea, 2014)

Selenium bağımsız sunucusu: Selenium Grid olarak da bilinir ve WebDriver ile oluşturulan Selenium komut dosyalarının uzaktan ve dağıtılmış olarak yürütülmesine izin verir. Ayrıca, iPhone ve iPad için Android veya Apple iOS gibi mobil platformlardaki testler de dahil olmak üzere testleri paralel olarak çalıştırmak için bağımsız sunucunun ızgara özelliği kullanılabilir. (Gundechea, 2014)

2.5.5. Scrapy Kütüphanesi

Girilen kurallara göre siteyi parçalara ayırıp, sitenin içindeki bilgileri almayı sağlayan açık kaynaklı bir kütüphanedir. Scrapy Linux, Mac ve Windows üzerinde çalışır. Kolayca genişletilebilir. Scrapy bozuk HTML'i anlar. Kafa karıştırıcı kodlamaları verimli şekilde anlayabilir. Scrapy kullanabilmek için Xpath ya da CSS ifadesi kullanarak HTML kaynaklarından veri çıkarmak gerekir. BeautifulSoup' un kullandığı lxml ayrıştırıcısına göre XPath daha yüksek seviyeli bir ara birimdir. Scrapy'nin canlı bir topluluğu vardır. Dakikalar içinde sorulan sorular yanıt bulur. (Kouzis-Loukas, 2016)

2.6. Web Kazıma Kullanım Alanları

Web kazıma ile toplanan bilgiler veri tabanı ya da bir dosya sistemine kaydedilebilir. Bu bilgiler birçok alanda kullanılabilir. Web kazıma araştırma çalışmaları; web sitesinde değişiklik tespiti, satış, ürün-fiyat karşılaştırma, pazar analizi, reklam analizi, finans, web indeksleme, hava durumu, marka izleme, iş ilanları, emlak ilanları vb. gibi birçok farklı alanlar için kullanılır. Bunların dışında ise istenilen bir araştırma konusunda da web kazıma yapılabilir. Çoğu zaman ise rakipler hakkında daha fazla bilgi edinmek için kullanılır. (Sirisuriya, 2015)

3. DUYGU ANALİZİ

İnternet, herkesin açık bir şekilde yorumlarını, görüşlerini ve düşüncelerini rahatça paylaştıkları bir iletişim aracıdır. İnternetle birlikte sosyal medyanın ve e-ticaretin yaygınlaşmasıyla insanlar aldıkları ürünlerden ziyaret ettikleri yerlere kadar tüm konular hakkında görüşlerini dijital ortamlarda paylaşmaya başladılar. İnsanlarda bir ürün sipariş ederken ya da bir yeri ziyaret etmeden önce diğer insanların yorumuna bakarak o ürün hakkında karar vermeye başladı. Örneğin internetten bir sipariş vermeden önce çoğu insan bu ürün hakkındaki yorumlara bakarak karar verir. Bir ürün hakkında insanların büyük çoğunluğu olumsuz yorum yaparsa bu insanların o ürün hakkındaki düşüncelerini etkiler. Duygu analizi (sentiment analysis) de kişilerin herhangi bir konu hakkındaki görüşlerinin sınıflandırılmasına yardımcı olmakta ve o konu hakkındaki genel görüşü tespit etmede kullanılır. Aynı zamanda bu görüşler özellikle firmalar için de ürünlerini ve hizmetlerini değerlendirmede büyük bir fırsattır. Firmalar ürünlerini ya da hizmetlerini yapılan yorumlara göre düzenleyip aldıkları olumlu yorumlar sayesinde daha fazla satış, izlenme vs. sayesinde kazançlarını artırma fırsatı bulurlar. (Demir, Chawai ve Doğan, 2019)

İnternetin yaygınlaşması ve sosyal medya, e-ticaret vs. gibi uygulamaların kullanım sayısının artması ile birlikte yapılan yorumların önemi son derece arttı. İnsanların düşüncelerini bu denli fazla bir şekilde aktarması ile birlikte yorumlar analiz edilmeye başlandı ve metin madenciliği, duygu analizi, fikir madenciliği gibi kavramlar ortaya çıktı. Duygu analizi, duyguların metinlerde nasıl anlatıldığını ve bu anlatımların incelenerek olumlu ya da olumsuz yönlerinin olup olmadığının belirlenmeye çalışıldığı bir analiz yöntemidir. Sentiment kelimesi Fransızca kökenli bir kelime olup anlamı ise kişinin duyguları anlamına gelmektedir. Duygu analizi yapılırken metnin içerisinde duygu belirten kelimelerin doğru bir şekilde tespit edilmesi gerekir. Bu tespitin yapılabilmesi için önceden belirlenmiş kelime listeleri kullanılabilir ya da öznetelik seçimi yöntemi kullanılarak metnin içerisindeki her bir kelimenin hangi duyguyu içerdiği tespit edilebilir. Metin madenciliğinden farklı olarak sadece özneye odaklanmadan duygusal sinyalleri yakalamaya çalışır. Duygu analizinin temel amacı verilen bir metnin hangi grupta yer aldığını belge veya

cümle düzeyinde sınıflamaktır. Konu hakkında yapılan görüşlerin pozitif mi negatif mi nötr mü olduğunu belirlemeyi amaçlar. Fakat başka birçok türü de mevcuttur. Gelişmiş düzeydeki duygu analizi metinlerindeki duygu durumlarını mutlu, üzgün şaşkın vb. şeklinde de sınıflamalar yaptırılabilir. Kısaca duygu analizi asıl olarak, insani duyguları makineye öğretmeyi amaçlar. Bir konu hakkında yapılan yorumun ya da düşüncenin hangi kategoride yer alacağını bir insan gibi tahmin ettirip otomatik olarak yapabilmesini sağlar. (Ahmetoğlu ve Daş, 2020)

Metinlerin ifade ettiği duygusal anlamları tespit edebilmek için birçok adım içermektedir. Bu adımlar fikir çıkarımı, fikir madenciliği, duygu analizi/madenciliği, özellik analizi gibi pek çok dal içerir. Duygu analizi yapılırken doğal dil işleme, veri madenciliği, makine öğrenmesi, istatistiki yaklaşımlar gibi yöntemlerin iş birliğinden bahsedilir. Duygu analizi yapabilmek için ön işleme, öznitelik seçimi, sözlük oluşturma, özellik azaltma, kelime ağırlıkları ile matris oluşturma, cümlelerin öğelerine ayrılması şeklinde verilebilir. Duygu analizi doküman seviyesinde, cümle seviyesinde, kelime grupları seviyesinde, kelime seviyesinde yapılabilmektedir. Bu yaklaşımları sözlük tabanlı ya da makine öğrenmesi tabanlı gerçekleştirilmektedir. (Yengi ve Omurca, 2016)

Doküman seviyesinde yapılan analizlerde tüm metin içerisinde bir fikir çıkartılması hedeflenir. Tüm metin hakkında sonuç olarak pozitif veya negatif olarak bir fikre varılmaktadır. Doküman seviyesi uygulamada kolaylık sağlar. Genel bir duygu polaritesi ortaya çıkarır. Doküman seviyesindeki dezavantaj ise ele alınan nesne hakkında tüzel kişiliklerin fikrine veya duygusuna ayrı ayrı ulaşamaz olmasıdır. Dolayısıyla burada genel bir duygu polaritesi sonucuna varılır. (Özyurt ve Akcayol, 2018)

Cümle seviyesinde ise, her bir cümle için ne tür bir duygu taşıdığı tespit edilir. Pozitif veya negatif bir fikir beyan ederken, nötr bir fikir beyan etmez. Bu seviye analizinde genel olarak öznel değeri olan cümleler ile nesnel değer taşıyan cümleler ayrılmaya çalışılır. Bu ayrım öznitelik sınıflandırma teknikleri ile yapılmaktadır. Bu analiz seviyesinde ima veya ironi içeren öznel cümlelerde çok net sonuçlar çıkaramamaktadır. (Özyurt ve Akcayol, 2018)

Varlık-görüş seviyesinde ise metindeki cümleler yerine komple metnin ortaya çıkardığı fikir ile ilgilenir. Metnin bir duygu ya da fikir içerip içermediği ile ilgilenir. Hedefin tanımlı olmadığı bir metin kısıtlıdır. Metnin vermek istediği mesajı yakalamaya çalışmak ve fark etmek, duygu analizi problemini daha iyi anlamamızı sağlamaktadır. Bu analiz düzeyinin amacı varlıkları ve yönlerini iyi anlamaktır. (Özyurt ve Akcayol, 2018)

Günümüzde, internetin hayatın her alanına giderek yaygınlaşmış olması, insanlara, fikirlerini, duygularını ve görüşlerini paylaşabildikleri sosyal medya, forumlar, bloglar, e-ticaret siteleri gibi sanal ortamlar sunmaktadır. Bu sanal ortamlarda insanların fikirlerinin ifade edildiği milyonlarca veri bulunmaktadır. Bu veriler, başta bu ürün ve hizmetlerin üreticisi ve satıcısı firmalar olmak üzere birçok kişinin ve araştırmacının ilgisini çekmektedir.

Duygu analizi, insanların sanal ortamlarda ürünler, hizmetler, organizasyonlar, olaylar, siyasi düşünceler gibi konular hakkında görüşlerini ifade ettikleri metinler içinde saklı olan duygu, fikir ve düşünceleri ortaya çıkarmayı amaçlamaktadır. 2000’li yılların başından itibaren duygu analizi, doğal dil işlemenin bir alt dalı olarak oldukça aktif çalışma alanı haline gelmiştir.

3.1. Duygu Analizi Literatür Taraması

Duygu analizi ifadesi ilk olarak 2003 yılında Tetsuya ve Jeonghee tarafından kullanılmıştır. Bu ifadeler ilk olarak 2003 yılında kullanılsa da ilk çalışmalar 2000’li yıllarda Vasileios ve Jancye, 2001 yılında Tong ve arkadaşları, 2002 yılında Turney ve gene 2002 yılında Pang’ın çalışmaları ilk çalışma olarak gösterilebilir. Pang ve arkadaşları tarafından sinema filmlerine ait yapılan yorumlar, pozitif ve negatif olarak belirlemek için yapılmıştır. Yapılan çalışmada kelime torbası modeli, sınıflandırıcı olarak ise NB ve DVM’yi kullanarak istenilen sonuç elde edilmiştir. (Özyurt ve Akcayol, 2018)

Pak ve Paroubek (2010), twitterda paylaşılan tweetlerin tarafsız, olumlu veya olumsuz olarak sınıflandırmak için teknikler önermiştir.

Sevindi (2013), Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarını karşılaştırmıştır. Sonuçlara göre makine öğrenmesi yaklaşımları daha başarılı sonuçlar elde etmiştir. Bunun nedeni olarak öğrenme adımı ile daha iyi temsil edebilen modeller oluşturması olarak düşünülmektedir.

Liu (2013), Twitter’da paylaşılan verilerden insanların duygularını önermede fikir madenciliğine ilişkin çeşitli zorluklar ve sorunlar olduğunu analiz etmiştir. Ayrıca, farklı dillerin fikir madenciliğini zorlu bir görev haline getirdiğini aktarmış, e-posta konuşmasının duyarlılığını anlamak için Textblob’un nasıl kullanılacağını göstermiştir. (Gujjar ve Kumar, 2021)

Mishra, Rajnish ve Kumar (2016), örnek olay üzerinde Twitter verilerinin duyarlılık analizini incelemiştir. Hindistan başbakanının kampanyası hakkında yorumlar üzerinde sözlük tabanlı duygu analizi çalışması yapılmıştır. Yapılan incelemeye göre %50 olumlu, %20 olumsuz ve geri kalan %30 yorumun tarafsız olduğu sonucunu elde etmiştir.

Ding, Li, Zhao ve Cheng (2017), Çinli seyahat rezervasyon platformu üzerinden alınan yorumlar üzerinde, Çin duygu sözlüğüne dayanan, yorumlara göre turisti otomatik puanlayan bir duygu analiz sistemi tasarlamıştır. Toplanan verilerden, sitenin sıralama sonuçlarına göre Wuhan'daki ilk on turistik mekânın incelemesini seçmiş ve hesaplama yöntemleri ile ilk on turistik mekânı puanlanmış ve çalışma sonunda sıralama sonuçlarının seyahat rezervasyon platformu sıralamasıyla oldukça tutarlı olduğu sonucuna varılmıştır.

Ashna ve Sunny (2017), Malayalam dili için sözlük tabanlı duyarlılık analiz sistemi çalışması yapmıştır. Bu çalışmada makine öğrenmesi yöntemleri kullanılamamıştır. Çünkü yeterli bir eğitim seti olmadığı için iyi bir performans gösteremez. Bu yüzden sözlük tabanlı yaklaşımlar kullanılarak duygu analizi yapabilmek için bir duygu sözlüğü oluşturulmuştur. Malayalam duygu sözlüğünü geliştirmek için sözlük tabanlı bir yaklaşım önermiştir. Önerilen sistem cümle düzeyinde sınıflandırma için %87.5 ve belge düzeyinde sınıflandırma için %90 doğruluk vermiştir.

Amanet (2017), sosyal medyadaki yazılı Türkçe metinlerde duygu analizi yapabilmek için bir yöntem önermiştir. Yapılan çalışma ile belirlenen duygu kategorileri kullanılarak bir sosyal medya sitesinin metin verileri üzerinde duygu analizi yapılmıştır. Kullanıcıların paylaştıkları metinleri hangi duygu durumuna ve hangi konuya bağlı olarak yazdığı tespit edilmiştir.

Toçoğlu ve Alpkoçak (2018), Türkçe metinlerde duygu analizi çalışmaları yapmak için yeni bir veri seti ve yeni bir sözlük ortaya koymaktadır. Yapılmış olan deneyler, önerilmiş olan sözlükteki anahtar kelime tespiti tekniğinin kullanımının Türkçe metinlerden duygu çıkarımı için tatmin edici sonuçlar verdiğini göstermiştir.

Pek (2019), makine öğrenmesi kullanarak duygu analizi yapmış, diğer yaygın uygulamaların yanı sıra özellik seçimi konusunda iyileştirilmeler yapılarak başarı oranı artırılmaya çalışılmıştır. Naive Bayes sınıflandırıcı eğitim setinin dışında herhangi bir veri setinde sınanarak %80 oranında başarı elde ederken, derin öğrenme yöntemleri içinde bulunan konvolüsyonel sinir ağlarında %97 başarı elde edilmiştir.

Yıldırım, Salman ve Ayvaz (2019), Türkçe için yeni bir duygu kütüphanesi geliştirmeyi hedeflemiştir. Çalışma kapsamında, var olan kütüphanelere emojiler ve puanlama altyapısı eklenerek genişletilmiştir. Oluşturulan bu duygu kütüphanesi iki bağımsız konu üzerinde uygulanmıştır. Çalışma sonucunda, Twitter kullanıcılarının kış aylarında daha olumsuz paylaşımlar yaparken, bahar ve yaz aylarında daha olumlu paylaşımlar yaptıkları tespit edilmiştir. Diğer çalışmada ise televizyonda yayınlanan Survivor için insanların Twitter'da yaptıkları yorumların olumsuz yönde olduğu gözlemlenmiştir.

Ayan, Kuyumcu ve Cıylan (2019), islamofobik tweetlerin sınıflandırılması için yapay zekâ tekniklerinden gözetimli makine öğrenmesi yöntemini kullanmıştır. Testlerde %95'in üzerinde doğruluğa ulaşılmıştır.

Özkan (2020), bir akıllı telefon hakkındaki Türkçe incelemelerin duygu sınıflarının belirlenmesi için doğal dil işleme kullanılarak; performans, fiyat ve kamera hedefleri bazında hedef-tabanlı duygu analizini gerçekleştirmeyi amaçlamıştır. Sonuçlar, incelenen hedefler için en yüksek kesinlik, duyarlılık ve F1 ölçümü değerlerinin sırasıyla %93, %94 ve %93 olduğunu göstermiştir.

Polat (2021), bir e-ticaret sitesinde bulunan farklı kategorilerdeki ürünler için Türkçe yorumları içeren verileri kullanmıştır. Her bir yorumun duygu durumunu tek tek belirlemeyi hedeflemiştir. Deneysel sonuçlar, Türkçe metinlerin duygu analizinde GRU yöntemi ile oluşturulan modellerin diğer modellerden daha başarılı sonuç verdiğini göstermiştir.

Eskin (2022), Twitter paylaşımlarını duygusal analiz ederek, sosyal medyanın BTC fiyatı üzerinde etkisini göstermek ve geleceğe yönelik fiyat tahmini yapmayı hedeflemiştir. Var model dahilinde Granger nedensellik testi ile bu üç zaman serisi (duygusal polarite puanı, BTC, ETH fiyatları) arasında nedensellik bulunmuştur ve model 7814.08 (\$)'lık ortalama mutlak hatayla tahmin yapabilmektedir.

Koç (2022), Türkiye gibi döviz kurlarının karışık ve dalgalı olduğu ülkelerde döviz kurlarını doğru tahmin edebilmeyi amaçlamıştır. Çalışma sonucunda, Twitter duygu analizi ile döviz fiyatları arasında ilişki tespit etmiştir. Olumsuz tweetler ile döviz fiyatları arasında pozitif bir ilişki olduğunu, olumlu tweetler arasında ise negatif bir ilişki olduğunu ortaya koymuştur.

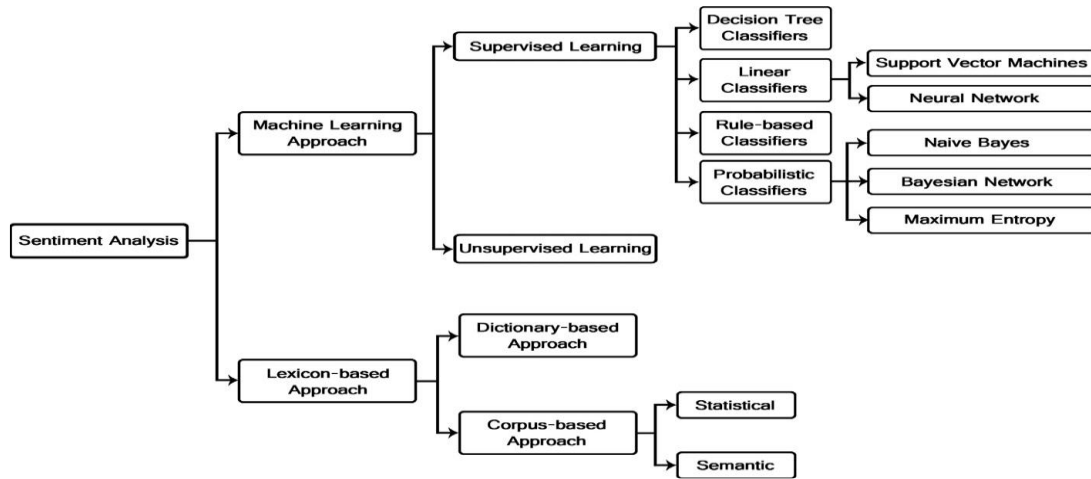
Akdeniz ve Babaoğlu (2022), makine öğrenmesi yöntemlerini kullanarak uzaktan eğitim konulu Türkçe tweetlerin duygu analizini yapmıştır. Türkçe metinlerin manuel etikete sahip olduğu yapıda en iyi TF-IDF – LR ikilisi ile 0.79'luk bir sınıflandırma başarısı elde edilmiştir. Manuel yöntemle etiketlenen tarafsız olarak işaretlenmiş metinler veri

setinden çıkarıldığında başarı oranının arttığı ve BoW – LR ikilisinin 0.84'lük oranla en iyi sonucu verdiği görülmüştür. Dil çeviri işlemi ile hazır modeller tarafından etiketlenerek oluşturulan modellerde Türkçe metinler için istenilen seviyede bir başarı elde edilememiştir.

İlgün (2022), kullanıcıların sosyal medya platformlarında paylaştığı gönderileri duygulara göre sınıflandırılması üzerine çalışma yapmıştır. Çalışmada dönüştürücü ve derin öğrenme modellerinin birlikte kullanımının dönüştürücü kullanılarak ya da kullanılmadan oluşturulan makine öğrenmesi modellerine ve literatürde bulunan geleneksel yöntemle oluşturulan modellere göre başarıyı arttırdığı görülmüştür.

Mutlu (2022), hedef odaklı duygu analizi verilen metinden hedef kelimeye yönelik olan duyguyu çıkarmayı hedeflemiştir. Geleneksel duygu analizi modellerinin performansının hedef odaklı duygu analizi görevinde düştüğünü gözlemlenmiştir. En iyi sonuç veren hedef belirteçli ve maksimum havuz katmanlı modeller geleneksel BERT tabanlı duygu analizi modellerinden %13 daha iyi F1 skoru verdiğini gözlemlenmiştir.

3.2. Duygu Analizi Teknikleri



Şekil 1. Duygu Analizi Teknikleri (Medhat, Hassan ve Korashy, 2014)

DA çalışmalarında kullanılan teknikler genel olarak 2 gruba ayrılır. Bunlar: Makine Öğrenmesi Tabanlı Yöntemler ve Sözlük Tabanlı Yöntemlerdir. Makine Öğrenimi Yaklaşımı (ML), makine öğrenmesi algoritmalarını uygular ve dilsel özellikleri kullanır. Sözlük Tabanlı Yaklaşım, bilinen ve önceden derlenmiş duygu terimlerinin bir koleksiyonu olan bir duygu sözlüğüne dayanır. Duygu polaritesini bulmak için istatistiksel veya anlamsal yöntemler kullanan sözlük tabanlı yaklaşım ve derlem tabanlı yaklaşım olarak ikiye ayrılır.

ML yaklaşımını kullanan metin sınıflandırma yöntemleri, denetimli ve denetimsiz öğrenme olarak ikiye ayrılır. Her iki yaklaşımın ortak kullanıldığı birçok çalışma mevcuttur. Makine öğrenmesi ve sözlük tabanlı yöntemlerin bir arada kullanıldığı kombine çalışmalara multi-modal yöntemler denilmektedir. (Medhat vd., 2014)

3.2.1. Sözlük Tabanlı Duygu Analizi

Sözlük tabanlı yöntemlerde, ilgili dil için daha önceden oluşturulmuş derlemeler kullanılmaktadır. Duygu anlamı taşıyan kelimeleri bir sözlük aracılığıyla seçmek için kullanılan en yaygın yöntemlerden biridir. Sözlük tabanlı yöntem ile elde edilen metinler fiil köklerine ayrılır ve sözlük ile karşılaştırılır ve her kelimenin frekansı sözlük sayesinde belirlenerek kelimenin ifade ettiği duygu düzeyi belirlenir. Bu yaklaşımda kelimelerin zıt ve eş anlamlı özelliklerinden yararlanılarak ilk anda verilen az kelime ile metnin özü yakalanmakta ve daha sonra diğer kelimelerle anlam güçlendirilmektedir. Örneğin iğrenç, mutsuz, kötü gibi kelimeler negatif duygu içerirken güzel, mutlu, harika gibi kelimeler pozitif duygu içerir. Bu şekilde anlamları önceden belirlenmiş olan kümelerin oluşturduğu sözlüğe duygu sözlüğü denir. Daha sonra yapılan çalışmalar ile birlikte zıt anlamlı kelimeleri tespit eden bir yöntem geliştirilerek sözlüğün kapsamı genişletilmiştir. Bu yöntemin eksik kalan yönleri ise eş anlamlı kelimeler veya söz öbekleridir. Aynı zamanda duygu kelimelerinin her zaman duygu için kullanılmaması da bir başka problemidir. Örneğin “Samsung telefon iyi midir?” sorusunda geçen iyi kelimesi Samsung için kullanılmamıştır. Alaylı ve imalı sözlerde başka bir problem konusudur. Örneğin; kötü bir ev için “aman ne güzel ev” sözü olumsuz bir anlam kast etmektedir. Bunlar sistemin sıkıntı olduğu konulardır. (Yengi ve Omurca, 2016)

3.2.1.1. Sözlük Tabanlı Duygu Analizi Araçları

Python’da kullanılabilecek bazı duygu analizi araçları NLTK, Textblob ve Vader olarak adlandırılır.

3.2.1.1.1. Natural Language Tool Kit

NLTK, insan dili verileriyle çalışmak için Python için geliştirilmiş açık kaynaklı bir kütüphanedir. NLTK, "Python kullanarak hesaplamalı dilbilimi öğretmek ve çalışmak için

harika bir araç" ve "doğal dille oynamak için harika bir kitaplık" olarak anılır. Python ile dil işleme için pratik bir giriş yapılmasını sağlar. (Bird, Klein ve Loper, 2009)

NLTK ilk olarak 2001 yılında Pensilvanya Üniversitesi'nde oluşturuldu. Daha sonra ise çok fazla geliştirildi. NLTK 4 hedef doğrultusunda tasarlanmıştır.

Basitlik: Kullanıcılara açıklamalı dil verilerinin işlenmesiyle pratik bir doğal dil işleme bilgisi verir.

Tutarlılık: Tutarlı arabirimler, veri yapıları ve kolayca tahmin edilebilen yöntem adları ile tek tip bir çerçeve sağlar.

Genişletilebilirlik: Alternatif uygulamalar ve aynı göreve rakip yaklaşımlar dahil olmak üzere yeni yazılım modüllerinin kolayca yerleştirilebileceği bir yapı sağlar.

Modülerlik: Araç setinin geri kalanını anlamaya gerek kalmadan bağımsız olarak kullanılabilen bileşenler sağlar. (Bird vd., 2009)

3.2.1.1.2. TextBlob

Textblob, NLTK kütüphanesi kullanan, metin verilerini işlemek için geliştirilmiş, Python'da versiyon 2 ve 3 de desteği olan doğal dil işleme kütüphanesidir. Bu kütüphane isim tamlaması çıkarma, duygu analizi, sınıflandırma, metni kelimelere ve cümlelere bölme, ayrıştırma ve çeviri gibi yaygın dil işleme görevlerinde kullanılmak için basit bir API desteği sağlar. (Yılmaz ve Yumuşak, 2021)

3.2.1.1.2.1. TextBlob Özellikleri

a. Part of Speech tagging:

Bir yapının yapısal elemanlarının yani filler, isimler, sıfatlar ve zarflar gibi belirlenmesi sürecidir. (Hazarika, Konwar, Deb ve Bora, 2020)

b. Noun Phrase Extraction:

İsimleri içeren belirli bir bağlamdan tümceler çıkarılmasını içerir. (Hazarika vb., 2020)

c. Duygu Analizi:

Analiz işleminde polarite ve öznellik olarak iki değer döndürür. Polarite (-1, +1) arasında değer alır. Değer -1'e doğru gittikçe yorumun negatifliği artar, 0 ise nötr, değer 1'e doğru

gittikçe yorumun pozitifliği artar. Öznellik ise (0,1) arasında değer alır. Değer 1'e ne kadar yakınsa yorumun o kadar öznel yani kişinin kendi duygularını belli eden bir yorum olduğu, sıfıra yakınsa ise yorumun çok fazla duygu ifade etmeyen bir yorum olduğu ortaya çıkar. (Loria, 2020)

Tablo 1'de bazı yorumların polartiy ve subjectivity değerleri gösterilmiştir.

Tablo 1. Textblob Kütüphanesi ile Örnek Yorumlar

Cümle	Polarity (Duygu polaritesi)	Subjectivity(Öznellik)
Bu ürünü seviyorum	0,5	0,6
Az maliyet için inanılmaz ürün	0,15	0,65
Kullanışsız ürün sakın almayın	-0,5	0,2

d. Tokenization:

Bir cümlemin sözcüklere ya da ifadelere bölünmesidir. Bir paragrafı farklı ifade türlerine bölmek veya bir ifadeyi farklı kelime türlerine bölmek anlamında da söylenir. (Hazarika vb., 2020)

e. Lemmatization:

Sözlükte verilen bir kelimenin kökenine kadar izini sürmeyi ifade eder. (Hazarika vb., 2020)

f. WordNet Entegrasyonu

Textblob kütüphanesiyle wordnet sentezlerine erişilebilir. (Loria, 2020)

Wordnet, ingilizce sözlük veri tabanıdır. Bu veri tabanı içinde eş anlamlı sözcükleri de bulundurulur ve sözcüklerin anlamına yer verir. (Aktaş, İnce, Çakır ve Kutlu, 2016)

3.2.1.1.2.2. Textblob ile Bir Metin Sınıflandırma Sistemi Oluşturma

İlk olarak bazı eğitim ve test verileri oluşturulur. Analiz için yeni bir metin verildiğinde yeni metni pozitif ya da negatif olarak kutuplandırmak için Naive Bayes sınıflandırıcısı kullanılır. (Loria, 2020)

3.2.1.1.3. Vader

Vader (Valence Aware Lexicon and Sentiment Reasoner), sosyal medyaya uygun olarak ayarlanmış, sözlük ve kural tabanlı bir duygu analizi aracıdır. Hutto ve Gilbert tarafından sosyal medyada kullanılan dilin sembollerini ve stilini analiz ederek geliştirilmiştir. Metindeki duyguların pozitif, nötr ve negatif kutupları tespit edebilir. linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW) ve General Inquirer (GI) üç sözlüğün özellikleri incelenip seçilmesiyle oluşturulmuştur. (Al-Shabi, 2020)

Vader, sözlük tabanlı bir duygu analizi aracıdır. Sosyal medya araçları, film incelemeleri ve ürün yorumlarıyla uğraşırken oldukça başarılıdır. Bunun nedeni, Vader'in yalnızca olumluluk ve olumsuzluk puanını değil, aynı zamanda bir duygunun ne kadar olumlu veya olumsuz olduğunu da söylemesidir. Herhangi bir eğitim verisine ihtiyaç duymaz, ancak genelleştirilebilir.

Vader, duygu sınıflandırması için emojileri de kullanabilir. Çevrimiçi olarak kullanılabilir kadar hızlıdır. Hız-performans değiş tokuşundan ciddi şekilde zarar görmez. (Bonta, Kumaresh ve Janardhan, 2019)

Vader, metindeki kelimelerden birinin kendi sözlüğünde olup olmadığını görmek için bir metin parçasını analiz eder. Polarity_scores() özelliğini kullanarak polarite indekslerini bulur. Bu, cümle için negatif, nötr, pozitif ve bileşik metrik değerlerini döndürür. Bileşik puan, -1 ile +1 arasında normalize edilmiş tüm sözlük derecelendirmelerinin toplamını hesaplayan bir ölçüdür; burada -1, en olumsuz ve +1 en pozitif gösterir. Cümleleri pozitif, nötr veya negatif olarak sınıflandırmak için standartlaştırılmış eşikler belirlemek faydalıdır. (Bonta vb, 2019)

Tablo 2. Vader Kütüphanesi ile Örnek Yorumlar

Cümle	Pozitif(pos)	Bileşik Puan(com)	Nötr(neu)	Negatif(neg)
Kitap güzeldi	0,492	0,4404	0,508	0,0
Vader eğlenceli ya da komik değil	0,0	-0,7424	0,354	0,646
Bugün sadece biraz eğlenceli! Ama idare edeceğim	0,317	0,5249	0,556	0,127

Bileşik puan(compound) sözlükteki her kelimenin eşdeğer değerleri toplanarak hesaplanmış ve -1 (en olumsuz) ile +1 (en olumlu) arasında olacak şekilde kurallara göre ayarlanmıştır. (Al-Shabi, 2020)

Olumlu Duygu: bileşik puan $\geq 0,05$

Nötr Duyarlılık: bileşik puan $> -0,05$ ve $< 0,05$

Olumsuz Duygu: bileşik puan $\leq -0,05$

Vader, duyguları öncelikle noktalama, büyük harf, bağlaçlar, gibi belirli kilit noktalara dayalı olarak analiz eder. Hem duygu kutupluluğunu hem de -4 ila +4 arasında değişen duygu yoğunluğunu gösteren doğrulanmış değerlik puanlarına sahip 7.500'den fazla sözcüksel özellik vardır. (Bonta vb., 2019)

3.2.2. Makine Öğrenmesi

Makine öğrenmesi, elde olan veriden öğrenip öğrendikleri ile tahminlerde bulunabilen algoritmalara denir. Yüklenen verilerden sonuç çıkarıp bu sonuçlardan tahminde bulunan, verilen bilgi dışında da bilgiler çıkartan ve kendisini geliştirebilen algoritma kümesidir. Makine öğrenmesi yapay zekanın alt bir birimidir. (Akbulut, 2022)

Makine Öğrenmeye Dayalı Sınıflandırma Yöntemleri

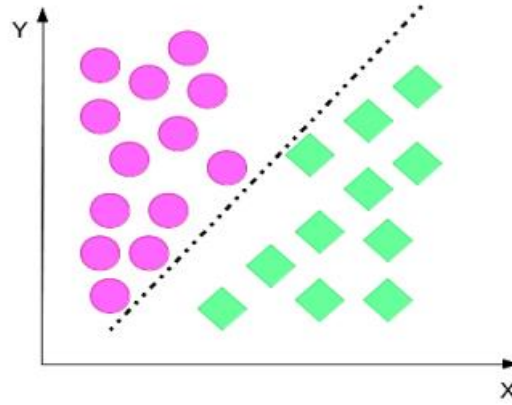
Makine öğrenmeye dayalı sınıflandırma yöntemleri 3'e ayrılır. Bunlar denetimli, denetimsiz ve takviyeli öğrenmelerdir. (Er, 2019)

3.2.2.1. Denetimli Öğrenme

Denetimli öğrenmede daha önceden belirlenmiş eğitim verileri kullanılır. Denetimli bir öğrenme algoritması, bilinen bir eğitim seti kullanır ve verilere bilinen yanıtları alır, ardından yeni verilere yanıt için kendisi tahminler oluşturur. (Pek, 2019)

Denetimli öğrenmenin en büyük sınırlılığı eğitim verilerinin miktarı ve bu verilerin kalitesine bağlı olmasıdır. Bu verilerin miktarının az ya da kalitesiz olması durumunda denetimli öğrenme ile istenilen sonuç alınamayabilir. Denetimli öğrenme verilen eğitim

verilerine dayanan bir model oluşturur. Buna göre kolaylıkla modelleyeceği bir sınıflandırma oluşturur. (Hasanlı, 2019)



Şekil 2. Denetimli Öğrenme (Avcı ve Yıldırım, 2021)

Denetimli öğrenmenin asıl amacı, girdiler ve girdilerden elde edilen sonuçları algoritmaya aktararak algoritmanın bu verilere göre bir sonuç oluşturmasıdır. Sonucun, girdi (x) ve çıktısı (y) daha önceden belirlenir. Asıl amaç, girdiler kullanılarak çıktıyı bulmaktır. (Aykul, 2019)

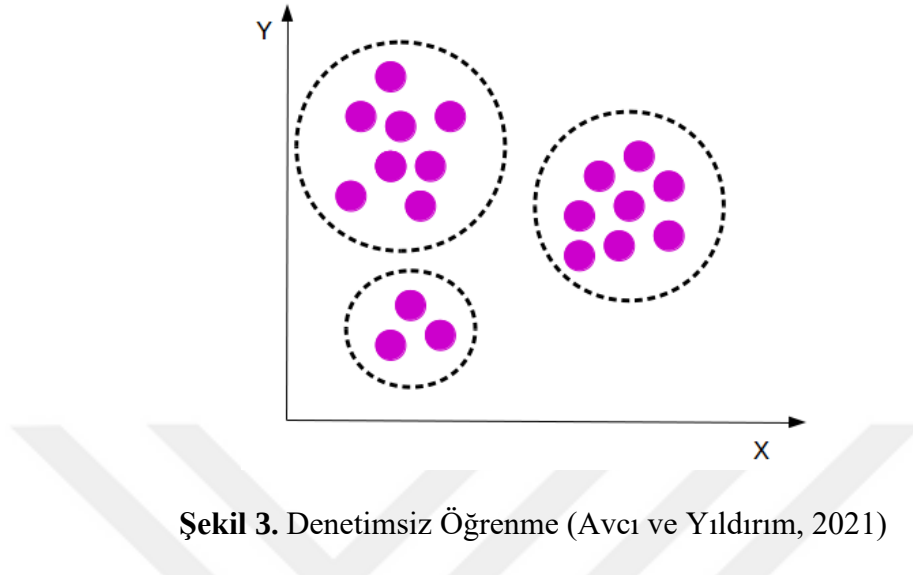
$$y=g(x|\theta) \quad (1)$$

Denklem (1) de g oluşturulan modeli, θ parametreleri, y ise sonucu ifade etmektedir. Girdi ve çıktı arasındaki bağlantının tümevarım yaklaşımıyla bulunma süreci öğrenme olarak isimlendirilir. Öğrenme evresinde ortaya çıkan model ise daha sonraki sonuçların tahmin edilmesi için kullanılır. Denetimli öğrenme alanında en önemli olay algoritmaların bir test kümesi ile çalışmasıdır. Test veri kümesi ile çalışan algoritmalar arasından en iyi çalışan seçilir. Seçim işleminde en çok kullanılan yol doğru sayısının toplam sayıya bölünmesiyle elde edilen tahmin doğruluğu metodudur. (Aykul, 2019)

3.2.2.2. Denetimsiz Öğrenme

Denetimsiz öğrenmede daha önceden belirlenmiş bir eğitim seti yoktur. Bu öğrenmede modele girdi verilmekte ve sonuç istenmektedir. Modele eğitim verisi verilmez, sadece test verisi verilir. Elimizdeki veriden sistem anlamlı bir şey çıkarmaya çalışır. Denetimsiz öğrenmede sistem öğretilmez, verilerden öğrenir. Veriler benzer özelliklerine ve birbirlerine

yakınlık durumuna göre gruplanır. En yaygın kullanıldığı yöntem kümelemedir. Kümeleme yönteminde yakın niteliklere sahip veriler aynı gruba gelecek şekilde kümelenir. (Pek, 2019)



Şekil 3. Denetimsiz Öğrenme (Avcı ve Yıldırım, 2021)

3.2.2.3. Takviyeli Öğrenme

Takviyeli öğrenme, en üst düzeyde kazanç sağlamak için yapılması gereken uygun adımları bulma ile ilgilenir. Buradaki öğrenme algoritması denetimli öğrenmeye göre farklı olarak deneme yanılma süreci ile keşfeder. Gerçekleştirilen her algoritmik adımda ödül veya ceza alınır. İstenilen sonuç üreten adımlar ödüllendirilirken, istenmeyen süreç üreten her adım da cezalandırılır. (Er, 2019)

Takviyeli öğrenme, sayısal bir durum sinyalini en yukarıda çıkarmak için ne tür eylemler yapmak gerektiğini gösteren bir öğrenme şeklidir. Öğrenen duruma, makine öğrenmesinin çoğu yöntemindeki gibi neler yapması gerektiğini söylemeden deneyerek bulmasını sağlar. (Aykul, 2019)

3.2.2.1.1. Olasılıksal sınıflandırıcılar

Olasılıksal sınıflandırıcılar, karışım modellerini kullanır. Olasılıksal sınıflandırıcılar;

3.2.2.1.1.1. Naive Bayes Sınıflandırıcısı

Olasılık teorisinde, Y olayına koşullu bir X olayı için olasılık değeri, X olayına koşullu bir Y olayı için olasılık değerleri farklıdır. Bu iki durum için birbirine ters koşulluluk ilişkisi vardır. Bu ilişkiye Bayes Teoremi denir. (Yumak, 2011)

Bayes teoremi, istatistik ve olasılık teorisi alanında kullanılan bir teoremdir. Bu teorem, bir olayın olasılığını hesaplamak için kullanılan bir formülasyon sunar. Naive Bayes sınıflandırıcısı da bu teoremi kullanır.

Naive Bayes sınıflandırıcısı birçok farklı alanda yaygın kullanıma sahip denetimli öğrenme algoritmasıdır. Belirli bir özellik kümesinin belirli bir sınıfa ait olma olasılığını tahmin etmede kullanılır. Olasılıksal hesaplamalar içinden en yüksek olasılığı seçmeyi hedefler. (Şahinaslan, Dalyan ve Şahinaslan, 2021)

Naive Bayes'in çalışma mantığı, bir eleman için her koşulun olasılığı hesaplanır ve olasılığı en yüksek olana göre sınıflandırılır. Temel işlevi sınıflandırma yapmaktır. (Düven, 2021)

Bayes teoremi, koşullu olasılıkların birbirine olan ilişkisini ifade eder. Koşullu olasılık formülü;

$$P(X|Y) = \frac{P(Y|X) \times P(X)}{P(Y)}$$

P(X): X olayının olasılığı

P(Y): Y olayının olasılığı

P(X|Y): Y olayı bilindiğinde X olasılığının gerçekleşme olasılığı

P(Y|X): X olayı bilindiğinde Y olasılığının gerçekleşme olasılığı

Naive Bayes sınıflandırıcısına göre P(X|Y)' yi maksimum yapan olasılıklar hesaplanır.

Naive Bayes, modelin öğrenilmesi sırasında, her sonucun meydana gelme olasılığını hesaplar. Bulunan bu değere öncelikli olasılık denir. Örneğin; hava durumuna göre tenis oynanma durumu “oynanır” ya da “oynanamaz” gibi gruplandırılmak istenmektedir. Verilen 10 durumdan 7'si oynanır olarak meydana geldiyse tenis oynanması için önsel olasılık 0.7'dir. Bu durum “tenis oynamak için başka hiçbir bilgi yoksa 0.7 olasılıkla oynanır grubundadır” olarak yorumlanır. (Yumak, 2011)

3.2.2.1.1.2. Maximum Entropy Sınıflandırıcısı

Maxent Sınıflandırıcı koşullu üstel modeller sınıfına ait olan olasılıksal bir sınıflandırıcıdır. Naive Bayes aksine Max Entropy, özelliklerin birbirinden bağımsız olduğunu varsaymaz. MaxEnt, Maksimum Entropi Prensipli temel alır ve eğitim verilerine uyan tüm modellerden en büyük entropiye sahip olanı seçer. Max Entropy sınıflandırıcı, dil algılama, konu sınıflandırma, duygu analizi ve daha fazlası gibi çok çeşitli metin sınıflandırma problemlerini çözmek için kullanılabilir. (Medhat vd., 2014)

3.2.2.1.1.3. Bayesian Network

Bayesian Network, değişken ve değişkenlerin ilişkileri için eksiksiz bir model olarak düşünülür. Bu sebeple tüm değişkenler için eksiksiz ortak olasılık dağılımı bir model için belirtilmiştir. Metin madenciliğinde Bayesian Network'ün hesaplama karmaşıklığı çok masraflıdır ve bu yüzden çok tercih edilmeyen bir yöntemdir. (Can ve Alataş, 2017)

3.2.2.1.2. Lineer Sınıflandırıcılar

Makine öğrenimi alanında, istatistiksel sınıflandırmanın amacı, bir nesnenin özelliklerini kullanarak hangi sınıfa ya da gruba ait olduğunu belirlemeye çalışmaktır. Lineer sınıflandırıcı sınıflandırmayı, özelliklerin doğrusal bir kombinasyonunun değerine göre sınıflandırma yaparak bulur. Nesnenin özellikleri aynı zamanda özellik değerleri olarak da bilinir ve tipik olarak makineye özellik vektörü adı verilen bir vektörde sunulur. Bu tür sınıflandırıcılar, basit problemler için iyi çalışır. Eğitim ve kullanım için daha az zaman alır. (Yuan, Ho ve Lin, 2012)

İki lineer sınıflandırıcı vardır. Bunlar destek vektör makineleri ve sinir ağıdır.

3.2.2.1.2.1. Destek Vektör Makineleri

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından temelleri atılmıştır. Destek Vektör Makineleri (DVM) istatistiksel öğrenme teorisine dayalı bir denetimli öğrenme algoritmasıdır. Bu algoritmanın temelleri 1960'lı yıllara dayansa da daha sonradan 1995'li yıllarda Vladimir Vapnik, Bernhard Boser ve Isabella Guyon tarafından geliştirilmiştir. El yazısı tanıma, yüz algılama ve ses analizi gibi alanlarda sık kullanılır.

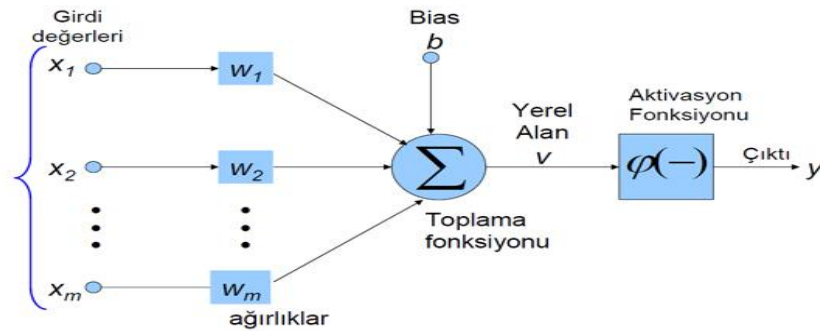
Destek Vektör Makineleri, özellikle sınıflandırma ve regresyon problemlerinde kullanılan bir makine öğrenme algoritmasıdır. DVM, veri noktalarını bir uzayda temsil eden vektörler olarak düşünür ve bu vektörlerin birbirlerinden ayrılabilmesi için bir hiper düzlem bulmaya çalışır. (Kör, 2019)

Destek Vektör Makineleri istatistiksel öğrenme teorisine dayalı bir kontrollü sınıflandırma algoritmasıdır. (Çölkesen ve Kavzoğlu, 2010)

Destek Vektör Makineleri, özellikle orta büyüklükteki veri setleri üzerinde iyi performans gösterir ve geniş bir makine öğrenme uygulama yelpazesine sahiptir. Ancak, büyük veri setleriyle çalışırken hesaplama süresi artabilir. SVM ayrıca, hiper parametre ayarının önemli olması nedeniyle deneme yanılma gerektirebilir.

3.2.2.1.2.2. Yapay Sinir Ağları

YSA sinir hücrelerinden oluşur. Bir YSA hücresinin yapılarında ise girdiler, ağırlıklar, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılarıdır.



Şekil 4. Yapay Sinir Hücresi (Karabulut, 2023)

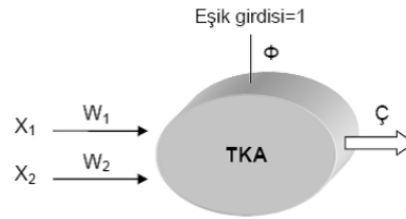
Yapay sinir hücresine çevreden gelen bilgiler girdi değerlerini oluşturur. Girdilerin sinir üzerindeki etkisi ağırlıklardır. Toplam fonksiyonu, ile net girdi hesaplanır. Hesaplanan girdi değerleri ile ağırlıkların çarpımı ile aktivasyon fonksiyonu elde edilir. (Karabulut, 2023)

YSA, insan beyninden esinlenerek geliştirilmiş bir teknolojidir. YSA, sinir sisteminin çalışma şeklini taklit eder. Örneklerden olaylar arasındaki ilişkileri öğrenerek sonraki görmediği örnekler hakkında öğrendiklerinden yola çıkarak karar veren bir sistemdir. Tek başına kullanılmaz. Birden fazla yapay sinir birbiri ile bağlantılı bir şekilde çalışır. Biyolojik nöron hücrelerinin ve bu hücrelerin birbiri arasında kurduğu sinaptik bağın dijital olarak modellenmesidir. (Pek, 2019)

Yapay sinir ağı katmanları giriş, ara ve çıkış olmak üzere 3 katmandan oluşur. Ağa girdi katmanı ile iletilen bilgiler ara katmanda işlenir, bilgiler işlendikten sonra çıktı katmanına ulaşır. Doğru çıktıların elde edilebilmesi için ağırlığın doğru değerlere sahip olması çok önemlidir. (Karabulut, 2023)

a. Tek Katmanlı Algılayıcılar

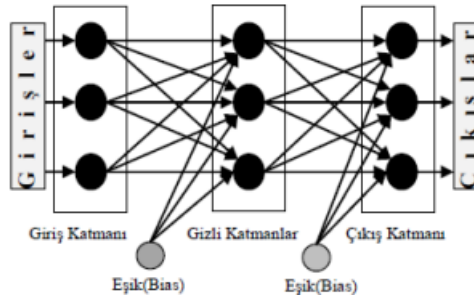
Sadece girdi ve çıktılarından oluşur. Çıktı fonksiyonu doğrusaldır. 1 veya -1 değerini alır.



Şekil 5. Tek Katmanlı Algılayıcı Model (Öztürk ve Şahin, 2018)

b. Çok Katmanlı Algılayıcılar

Tek katmanlı algılayıcıların doğrusal olmayan problemlerin çözümünde yetersiz kalmasından dolayı geliştirilmiştir. Tek katmanlı algılayıcılardan farklı olarak bir veya birden fazla ara katmanın olmasıdır. (Arı ve Berberler, 2017)



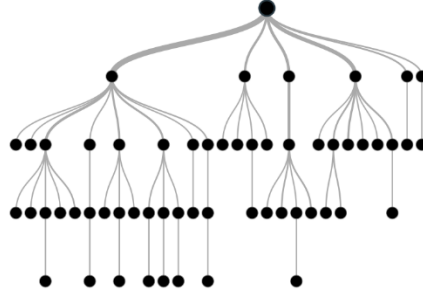
Şekil 6. Çok katmanlı Algılayıcı Modeli (Arı ve Berberler, 2017)

Çok katmanlı algılayıcılarda ileri ve geri yayılım yani geçişler bulunur. İleri yayılım sırasında ağı çıktısı değeri ve hata değeri hesaplanırken, geri yayılım sırasında hata değerinin en minimuma indirilmesi için katmanlar arası bağlantı ağırlık değerleri güncellenir. (Arı ve Berberler, 2017)

3.2.2.1.2.3. Karar Ağaçları

Örneklemeleri sınıflandırmak için ağaç yapısını şeklini kullandığı için bu ismi almıştır. Verileri sınıflandırmak için kullanılır. Başta bir tane kök düğüm belirlenir. Daha sonra kök

düğümünden itibaren aşağı yaprak düğümlere doğru sıralanarak devam eder. Yaprak düğümleri nitelikleri test ederek örnekleri ayırır. Yaprak düğümlerden ayrılan dallar da düğümlerden elde edilen test sonucuna göre ayırım işleminden sonra örnekleri temsil eder. CART, ID3, C4.5, J48 gibi çok fazla algoritması vardır. (Kara, 2019)



Şekil 7. Karar Ağaçları

Id3 Algoritması:

Temel algoritması ID3'tür. ID3 makine öğrenmesinde yeni olan kişilerin rahatlıkla kavrayabileceği ve karar ağaçlarının temel şeklini öğrenebileceği bir algoritmadır. ID3 algoritması, entropiden faydalanarak sınıflandırma problemlerinin çözülmesine imkân tanır. Entropiye dayalı bir algoritmadır. Ayrık veriler üzerinde çalışır. Eksik veriler üzerinde iyi değildir. Veri tabanında fazla değişken varsa fazla hesaplama yapmadan uygun bir karar ağacı oluşturmak isteniyorsa kullanılabilir. (Şatır, Azboy, Aydın, Arslan ve Hacıfendioğlu, 2016)

4. METİN ÖZETLEME

Her gün çok büyük miktarda veri ortaya çıkmaktadır. Bir kişi herhangi bir metnin ya da verinin özetini elde etmek için metni tamamen okuması gerekir. Bu olay tek bir makalede veya tek bir yazı için kolay durabilir fakat binlerce metin olduğunda bir kişi bunu yapamaz. Bu nedenle bu kadar fazla metni daha verimli hale getirecek bir araca ihtiyaç vardır. Bu noktada otomatik özetler ortaya çıkmıştır. (Alipour, 2023)

Metin özetleme, bir kaynaktan ya da aynı konunun anlatıldığı birden fazla kaynaktan kullanıcı için gerekli olan bilgilerin elde edilmesidir. Bu işlem eğer bir bilgisayar tarafından yazılım ile gerçekleştiriliyor ise otomatik metin düzenleme adını alır. (Bal, 2022)

Metin özetleme ile ilgili ilk çalışma 1959 yılında Luhn adlı bir bilim adamı tarafından yapılmıştır. Luhn çalışmasında, metin özetlemesi için kelimelerin cümle içerisindeki frekans sayılarından yararlanmıştır. En çok frekansa sahip kelimelerin o yazı hakkındaki en önemli görüşü verdiğini düşünmüştür. 1969 yılında Edmundson, 1989 yılında da Salton bu konuyla ilgili farklı çalışmalarda bulunmuşlardır. Edmundson yaptığı çalışmasında kelime frekanslarına ek olarak, ipucu veren bazı kelimeleri dikkate almıştır. (Uzundere, Dedja, Diri ve Amasyalı, 2008)

Otomatik metin özetleme yaklaşımları çeşitleri açısından kategorize edilmiştir. Özetlenecek doküman sayısı, özet türüne göre ve özet içeriğine göre farklı alanlara ayrılmaktadır. (Erdağı, 2023)



Şekil 8. Otomatik Metin Özetleme Türleri (Erdağı, 2023)

Metin özetlemeler, girdi olarak alınan doküman sayısına göre sınıflandırılabilir. Bunlar tek belgeli özetleme ve çok belgeli özetlemelerdir. Tek belgeli özetleme sisteminde girdi olarak sadece bir bilgi verilip özet alınır. Çok belgeli özetlemelerde ise birden fazla doküman sisteme girdi olarak verilip özet alınmaktadır. Çok belgeli özetlemelerde sağlıklı sonuçlar elde edebilmek için girdilerin aynı konu üzerinde olması gerekir. (Uçkan ve Karcı, 2020)

Özetler içeriğine bağlı olarak alana özgü özet, sorgu bazlı özet ve genel özet olmak üzere üçe ayrılır. Alana özgü özetlerde belirli bir alan üzerinde yazılmış bir doküman özetlenir. Alana özgü özetlerde verilen alana özgü kelimeler ve kavramları içerir. Genel amaçlı özetler ise belgenin tüm içeriği hakkında genel fikir elde etmek için kullanılır. Sorgu bazlı özetler ise kullanıcı tarafından üretilen ya da iletilen sorgular doğrultusunda çıkarılan bir özet yöntemidir. (Erdağı, 2023)

4.1. Yorumlama Tabanlı Metin Özetleme

Özetsel metin özetleme yönteminde metine bağlı kalınarak orijinal metinde bulunmayan kelimeler kullanarak yeni bir özet meydana getirilir. Farklı kelimeler kullanılması gerektiğinden dilbilimsel açıdan zorlukları vardır. Nitekim insanlar da bir metni okurken neden sonuç ilişkisini özümser, başka kişiye anlatırken farklı kelimeler ve cümleler ile bunu aktarır. (Erdağı, 2023)

4.2. Çıkarımsal Metin Özetleme

Bir metnin içerisinden doğal dile uygun cümleler alındığı için ek çalışmaya gerek duyulmaz. Bu kolaylığı sağladığı için çok fazla tercih edilir. Çıkarımsal metin özetlemenin avantajları olduğu gibi dezavantajları da vardır. Genellikle uzun cümleler daha fazla sayıda bilgi verdiği için daha fazla tercih edilir. Fakat burada gereksiz kelimelerin tekrarlanması biz dezavantaja dönüşebilir. (Erdağı, 2023)

Çıkarımsal metin özetlemede kelimeler, kelime öbekleri ve cümleler alınarak işlem yapılır, okunabilirliği yüksektir. İstatiksel ve sezgisel yöntemler kullanılır. İki türü bulunur. Bunlar kelime frekans değerine göre özetleme ve Textrank algoritmasına göre özetlemedir. (Erhandı, 2020)

Tez kapsamında yapılan çalışmada çıkarımsal metin özetleme yöntemi olan kelime frekans değerine göre, çok belgeli ve genel özet yöntemleri kullanılarak metin özetleme çalışması yürütülmüştür.

4.3. Denetimli ve Denetimsiz Metin Özetleme

Metin özetleme sistemi, denetimli ve denetimsiz metin özetleme olarak iki kısma ayrılır. Denetimli metin özetlemelerde büyük miktarda eğitim verisine ihtiyaç duyulmaktadır. Bunun sebebi eğer küçük boyutlu bir eğitim seti verilirse öğrenme az olacağı için doğru sonuçlar üretmek zorlaşacaktır. Denetimsiz özetlemelerde ise herhangi bir ön bilgi ya da eğitim verisi gerekmez. Bu tür sistemler genellikle yeni veri setleri için uygundur. (Uçkan ve Karcı, 2020)

5. UYGULAMA

5.1. Materyal ve Yöntem

Bu çalışmada ürünlere yapılan yorumlar analiz edilerek müşterilerin zaman tasarrufu ve maddi kayıplar yaşamasını engelleyecek bir proje oluşturulmaya çalışılmıştır. Çalışma yapılırken veri setinin seçimi, web kazıma programı, duygu analizi programı, metin özetleme ve verilerin hangi web yazılımıyla gösterileceği son derece önemlidir.

Son yıllarda çok popüler olan e-ticaret sitesi Trendyol üzerindeki yorumların toplanmasına karar verilmiştir. Trendyol üzerinde çok fazla sayıda ürün ve bu ürünlerde çok fazla sayıda ürün yorumu, ürün yıldızlı değerlendirmesi bulunmaktadır.

Web kazıma işlemi için Python oldukça popüler bir dildir. Python'un web kazıma kütüphanesi Scrapy kullanılarak web siteleri derinlemesine taranabilir. Verilerin kaydedilmesi için kullanımı basit ve Python ile uyumlu olan SQLite veri tabanı seçilmiştir.

Sözlük tabanlı duygu analizi kütüphanelerinden Vader ve Textblob seçilmiştir. Metin özetleme için frekans değerine göre özetleme yöntemleri kullanılmıştır.

Duygu analizi sonuçları, ürünlerin listesi, ürün değerlendirmeleri Python web arayüzü olan Django ile gösterilmiştir.

5.1.1. Kullanılacak Veri Seti Seçimi

Veri seçimi oldukça önemlidir. İnsanlar internet üzerinden bir ürün sipariş ederken yorumlara çok fazla dikkat eder. Bu yüzden çalışmada ürün sayısının ve ürün yorumunun çok fazla olduğu e-ticaret sitesi Trendyol seçilmiştir.

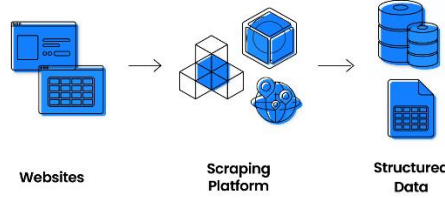
5.1.2. Web Kazıma Programı

Python, Guido Van Rossum tarafından geliştirilmiş bir programlama dilidir. Van Russom ABC programla dilinin geliştirilmesinde yer alırken ABC' nin iyi yönlerinden yola çıkarak,

okunması ve yazması mevcut güçlü dillere göre çok daha kolay olan Python üzerinde çalışmaya başladı. Python, ismini Rossum 'un çok sevdiği MonthyPython isimli komedi grubundan alır. Python nesne yönelimli, yorumsal, modüler ve etkileşimli, yüksek seviyeli bir dildir. Yüksek seviye bir dil olduğu için öğrenilmesi kolaylaşmaktadır. Python' un kodlaması kolaydır; Python kodunun okunabilirliği İngilizce bir ifadenin okunmasına çok benzer olması sebebiyle yeni başlayanların öğrenmesi kolaydır, kodlama yaparken noktalı virgül gerekmez. Yapılmak istenen işler hızlı bir şekilde daha az kod satırıyla yapılabilir. Gelişiminde bu özellik hep ön planda tutulmuştur. Değişken tanımlarken veri türü tanımlamasına gerek kalmaz. (Akça, 2017)

Desteklendiği platformlar oldukça fazladır. Birçok işletim sistemiyle uyumludur. Python oldukça genel bir programlama dilidir. Kullanım alanlarının çok fazlası olmasından dolayı çok fazla tercih edilir. Python, birçok alanda kullanılmaktadır. (Saralıoğlu, 2015)

Bu çalışmada web kazıma için Python kütüphanesi olan Scrapy kütüphanesi seçilmiştir. Scrapy diğer mevcut kütüphanelere göre daha hızlı sonuç verir. Scrapy, diğer kütüphanelere göre çok daha az bellek ve CPU kullanır.



Şekil 9. Scraping

5.1.3. Veri Tabanı

SQLite hafif ve yerel bir veri tabanıdır, C diliyle yazılmıştır. Python ile kullanımı oldukça kolaydır. Python, SQLite ile etkileşim kurmak için özel bir modül içerir. Python ve SQLite'nin entegrasyonu, veri tabanı işlemlerini kolaylaştırır ve Python programlarıyla veri tabanı işlemleri hızlı bir şekilde gerçekleşir. Bu çalışmada web kazımadan gelen veriler SQLite ile veri tabanına kaydedilmiştir.

İsim	Tip	Şema
▼ Tablolar (15)		
> analiz_brand		CREATE TABLE "analiz_brand" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "product_brand" varchar(200) NOT NULL, "product_image"
> analiz_comment		CREATE TABLE "analiz_comment" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "comment_content" text NOT NULL, "comment_evatula"
> analiz_commenten		CREATE TABLE "analiz_commenten" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "comment_content" text NOT NULL, "comment_evat"
> analiz_product		CREATE TABLE "analiz_product" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "product_name" varchar(200) NOT NULL, "product_brand"
> auth_group		CREATE TABLE "auth_group" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "name" varchar(150) NOT NULL UNIQUE)
> auth_group_permissions		CREATE TABLE "auth_group_permissions" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "group_id" integer NOT NULL REFERENCES "au"
> auth_permission		CREATE TABLE "auth_permission" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "content_type_id" integer NOT NULL REFERENCES "dja"
> auth_user		CREATE TABLE "auth_user" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "password" varchar(128) NOT NULL, "last_login" datetime NU
> auth_user_groups		CREATE TABLE "auth_user_groups" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "user_id" integer NOT NULL REFERENCES "auth_user"
> auth_user_user_permissions		CREATE TABLE "auth_user_user_permissions" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "user_id" integer NOT NULL REFERENCES "
> django_admin_log		CREATE TABLE "django_admin_log" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "object_id" text NULL, "object_repr" varchar(200) NO
> django_content_type		CREATE TABLE "django_content_type" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "app_label" varchar(100) NOT NULL, "model" varc
> django_migrations		CREATE TABLE "django_migrations" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "app" varchar(255) NOT NULL, "name" varchar(255) I
> django_session		CREATE TABLE "django_session" ("session_key" varchar(40) NOT NULL PRIMARY KEY, "session_data" text NOT NULL, "expire_date" datetime NOT
> sqlite_sequence		CREATE TABLE sqlite_sequence(name,seq)

Şekil 10. SQLite Arayüzü

5.1.4. Duygu Analizi

Sözlük tabanlı duygu analizi kütüphaneleri Vader ve Textblob kullanılmıştır.

5.1.5. Web Arayüzü

Django, Python programlama dili kullanılarak web uygulamaları geliştirmek için kullanılan yüksek seviye bir web çerçevesidir. Model-View-Controller (MVC) mimari desenini takip eder. Django, güvenlik, ölçeklenebilirlik, hızlı geliştirme ve kapsamlı topluluk desteği gibi avantajlarıyla öne çıkar.



Şekil 11. Django Arayüzü

Django projesi, büyük ölçekli web uygulamalarının geliştirilmesinde kullanılır ve birçok başarılı web sitesi ve uygulama tarafından tercih edilmektedir. Django, açık kaynaklı bir projedir ve geniş bir topluluk tarafından desteklenir. Dokümantasyonu kapsamlıdır ve çeşitli kaynaklar ve örnekler bulunmaktadır.

5.2. Veri Setinin Web Kazıma ile Çekilmesi

Web kazıma web üzerinden istenilen verilerin otomatik olarak çekilme işlemidir. Trendyol üzerinden Avva, Adidas, US Polo Assn, Puma, Hummel, De Facto, Mavi, Colin's, Derimod, Koton, LC Waikiki, Loft, Nike, Suwen ve Tudors olmak üzere 15 markaya ait toplam 4109 ürünle alakalı 34007 yorum web kazıma ile çekilmiştir.

5.3. Veri Tabanı Oluşturma

SQLite veri tabanında ürün markalarının bulunduğu analiz_brand, ürün yorumlarının bulunduğu analiz_comment, ürün yorumları ileride İngilizceye çevrileceği için analiz_commenten ve ürünlerin bulunduğu analiz_product tabloları oluşturulmuştur. Oluşturulan 4 tablonun dışında Django' nun otomatik olarak oluşturduğu tablolarda şekil 12'de gösterilmiştir. Ürün markaları tablosunda ürünün markası ve resmi alanı oluşturulmuştur. Ürün tablosunda, ürünün adı ve ürünün linki, ürün yorumları tablosunda ise ürüne yapılan yorum, verilen yıldız, textblob ve vader alanları oluşturulmuştur.

Tablolar (15)	
> analiz_brand	CREATE TABLE "analiz_brand" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "product_brand" varchar(200) NOT NULL, "product_image"
> analiz_comment	CREATE TABLE "analiz_comment" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "comment_content" text NOT NULL, "comment_evatule"
> analiz_commenten	CREATE TABLE "analiz_commenten" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "comment_content" text NOT NULL, "comment_evatule"
> analiz_product	CREATE TABLE "analiz_product" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "product_name" varchar(200) NOT NULL, "product_brand"
> auth_group	CREATE TABLE "auth_group" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "name" varchar(150) NOT NULL UNIQUE)
> auth_group_permissions	CREATE TABLE "auth_group_permissions" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "group_id" integer NOT NULL REFERENCES "au"
> auth_permission	CREATE TABLE "auth_permission" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "content_type_id" integer NOT NULL REFERENCES "djc"
> auth_user	CREATE TABLE "auth_user" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "password" varchar(128) NOT NULL, "last_login" datetime NL
> auth_user_groups	CREATE TABLE "auth_user_groups" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "user_id" integer NOT NULL REFERENCES "auth_user"
> auth_user_user_permissions	CREATE TABLE "auth_user_user_permissions" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "user_id" integer NOT NULL REFERENCES
> django_admin_log	CREATE TABLE "django_admin_log" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "object_id" text NULL, "object_repr" varchar(200) NO
> django_content_type	CREATE TABLE "django_content_type" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "app_label" varchar(100) NOT NULL, "model" vard
> django_migrations	CREATE TABLE "django_migrations" ("id" integer NOT NULL PRIMARY KEY AUTOINCREMENT, "app" varchar(255) NOT NULL, "name" varchar(255)
> django_session	CREATE TABLE "django_session" ("session_key" varchar(40) NOT NULL PRIMARY KEY, "session_data" text NOT NULL, "expire_date" datetime NOT
> sqlite_sequence	CREATE TABLE "sqlite_sequence"(name,seq)

Şekil 12. Oluşturulan Veri Tabanı

5.4. Veri Tabanına Veri Kaydetme

Öncelikle marka isimleri yani analiz_brand tablosuna seçilen 15 marka, isimleri ve resimlerin linki ile birlikte veri tabanına kaydedilmiştir.

	id	product_brand	product_image
	Filtre	Filtre	Filtre
1	1	Avva	https://www.avva.com.tr/Uploads/...
2	2	US Polo Assn	https://aydinli-polo.akinoncdn.com/...
3	3	Puma	https://...
4	9	Defacto	https://www.mallofantalya.com.tr/...
5	10	Mavi	https://yt3.ggpht.com/...
6	11	Hummel	https://www.logoeeps.net/wp-conten...
7	12	Colin's	https://www.colins.com.tr/Themes/...
8	13	Derimod	https://www.marmarapark.com/...
9	14	Koton	https://upload.wikimedia.org/...
10	15	LC Waikiki	https://margioutlet.com/wp-content/...
11	16	Loft	https://anninc.scene7.com/is/image/...
12	17	Nike	https://c.static-nike.com/a/images/...
13	18	Suwen	https://cdn-suwen.mncdn.com/...
14	19	Tudors	https://www.tudors.com/Uploads/...
15	20	adidas	https://1000logos.net/wp-content/...

Şekil 13. Markalar Tablosu

Daha sonra bu markaya ait ürünler yani analiz_product; ürün ismi, ürün linki ve marka_id'sine göre çekilerek veri tabanına kaydedilmiştir.

id ▼1	product_name	product_link	brand_id
Filtre	Filtre	Filtre	Filtre
1	Erkek %100 Pamuk Koyu Vizon Polo ...	https://www.trendyol.com/avva/...	1
2	Erkek %100 Pamuk Beyaz Polo Yaka ...	https://www.trendyol.com/avva/...	1
3	Erkek %100 Pamuk Açık Turuncu Pol...	https://www.trendyol.com/avva/...	1
4	Erkek Lacivert Polo Yaka %100 Pamu...	https://www.trendyol.com/avva/erke...	1
5	Erkek %100 Pamuk Koyu Mavi Polo ...	https://www.trendyol.com/avva/...	1
6	Erkek %100 Pamuk Siyah Polo Yaka ...	https://www.trendyol.com/avva/...	1
7	Beyaz Erkek T-Shirt ...	https://www.trendyol.com/us-polo-...	2
8	Erkek %100 Pamuk Açık Mavi Polo ...	https://www.trendyol.com/avva/...	1
9	Siyah Erkek T-Shirt ...	https://www.trendyol.com/us-polo-...	2
10	Lacivert Erkek T-Shirt ...	https://www.trendyol.com/us-polo-...	2

Şekil 14. Ürünler Tablosu

Toplam 4109 ürün veri tabanına kaydedilmiştir. Ürünleri alan kişilerin yorumları yani analiz_comment alanına; ürün yorumları ve ürüne verilen yıldız değerleri veri tabanına kaydedilmiştir.

	id	comment_content	comment_evaluation	product_id
	Filtre	Filtre	Filtre	Filtre
1	44557	Ürün güzel ve kaliteli.Sorun yok ...	5	4109
2	44556	Yakası çok büyük ve kumaşı maviye ...	1	4109
3	44555	kaliteli ama biraz kalın geldi bana	1	4109
4	44554	rengi tam resimdeki gibi değil kırmızı ...	4	4109
5	44553	Kumaşı rengi güzel	5	4109
6	44552	çok güzel bir kalıp, mavi kalitesi çok ...	5	4109
7	44551	ürünü aldıktan 20 gün sonra fiyat yarı...	1	4109
8	44550	Kumaşına bayıldık fiyatıda çok iyi	5	4109
9	44549	Ürün gerçekten kaliteli ve güzel.	5	4109
10	44548	Süper tek kelime ile SÜPER	5	4109

Şekil 15. Ürün Yorumları

Toplam 4109 ürünle ilgili 34007 yorum web kazıma işlemi ile çekilerek SQLite veri tabanına sırasıyla kaydedilmiştir.

5.5. Duygu Analizi İşlemleri

Duygu analizi yapılabilmesi için öncelikle verilerin ön işleminden geçmesi gerekir.

5.5.1. Veri Ön İşleme

Yorumların üzerinde analiz yapabilmek için cümlelerin bir takım ön işlemlerden geçmesi gerekir. Duygu analizi işlemleri veri toplama ve veri temizleme işlemleri ile başlamaktadır. Toplanan verilerden bazılarında hata, eksiklik olabilir. Bunlara gürültü denir. Başarılı sınıflandırma yapabilmek için bunların doğru bir şekilde yapılandırılması gerekir.

Veri ön işlemenin ilk adımı olarak veri tabanında bulunan bütün yorumlar büyük, küçük harflerin kullanımıyla ilgili uyumsuzluğu kaldırmak için tüm kelimeler küçük harfe dönüştürülmüştür. Daha sonra sade bir metin elde etmek için noktalama işaretleri, emojiler ve etkisiz kelimeler (bağlaç vb.) kaldırılmıştır.

3156	👍 👍 👍 👍 👍 👍 👍 👍 👍 👍	5	127
18039	👍 👍 👍 👍 👍 👍	5	886

Şekil 16. Emojili Yorumlar

Daha sonra cümlelerdeki yazım yanlışları incelenmiştir.

Iade ettifim tahirt elime ulasti saka mi?

Sintetik, Kötü lkalite.....

Şekil 17. Yanlış Yazımı Olan Yorumlar-1

Birçok yorumdaki yazım yanlışları tek tek düzeltilmiştir.

çokkkk beğendik tavsiye ediyorum

Şekil 18. Yanlış Yazımı Olan Yorumlar-2

Bazı yorumlarda bulunan kelimelerdeki uzatma harfler düzenlemiştir.

MÜ KEM MEL ÜRÜN TAM BEDENİNİZ...

Şekil 19. Yanlış Yazımı Olan Yorumlar-3

Bazı kelimelerdeki boşluklu kullanım tarzındaki yanlışlar düzeltilmiştir.

cok sık bir urun cok yakisti gonul .

Şekil 20. Yanlış Yazımı Olan Yorumlar-4

Türkçe karakterlerdeki yanlış kullanımlar düzeltilmiştir. Tüm bu düzenlemeler yapıldıktan sonra veri tabanındaki yorumlar düzenlenerek kaydedilmiştir.

5.5.2. Veri Tabanı İşlemleri

Bu çalışmada duygu analizi yöntemlerinden sözlük tabanlı yaklaşımlar kullanılmıştır. Veri ön işlemlerinden geçmiş ürün yorumları Python programlama dilinin sözlük tabanlı yaklaşım mantığı ile çalışan TextBlob kütüphanesi kullanılarak tüm yorumlar analiz edilmiştir. Textblob Türkçe desteği olmadığı için yorumlar İngilizceye çevrilip veri tabanına kaydedilmiştir ve Şekil 21’de gösterilmiştir.

id ▼1	comment_content	comment_evaluation	product_id
Filtre	Filtre	Filtre	Filtre
1	I bought it for my dad it's very nice	5	1
2	the texture is very nice, the quality is...	5	1
3	The color and quality are very nice.	5	1
4	i like it very nice	5	1
5	nice Tshirt	5	1
6	thanks	5	1
7	the product is very nice	5	1
8	I was expecting something better ...	3	1
9	this product is good quality	5	1
10	very nice product	5	1

Şekil 21. Veri Tabanı Şekli

Daha sonra Python'da Textblob ile duygu analizi yapıp sonuçlar veri tabanına kaydedilmiştir ve Şekil 22'de gösterilmiştir.

id ▼1	comment_content	comment_evaluation	product_id	comment_textblob
Filtre	Filtre	Filtre	Filtre	Filtre
1	I bought it for my dad it's very nice	5	1	0.78
2	the texture is very nice, the quality is...	5	1	0.845
3	The color and quality are very nice.	5	1	0.78
4	i like it very nice	5	1	0.78
5	nice Tshirt	5	1	0.6
6	thanks	5	1	0.2
7	the product is very nice	5	1	0.78
8	I was expecting something better ...	3	1	0.5
9	this product is good quality	5	1	0.7
10	very nice product	5	1	0.78

Şekil 22. Duygu Analizi (Textblob) Sonuçları İşlenmiş Yorumların Şekli

Aynı şekilde Vader kütüphanesi kullanılarak yorumlar tek tek değerlendirilip veri tabanına kaydedilmiştir ve Şekil 23'de gösterilmiştir.

	comment_content	product_id	comment_textblob	comment_vader
	Filtre	Filtre	Filtre	Filtre
1	I bought it for my dad it's very nice	1	0.78	{'neg': 0.0, 'neu': 0.721, 'pos': 0.279, 'compound': 0.4754}
2	the texture is very nice, the quality is...	1	0.845	{'neg': 0.0, 'neu': 0.56, 'pos': 0.44, 'compound': 0.742}
3	The color and quality are very nice.	1	0.78	{'neg': 0.0, 'neu': 0.66, 'pos': 0.34, 'compound': 0.4754}
4	i like it very nice	1	0.78	{'neg': 0.0, 'neu': 0.349, 'pos': 0.651, 'compound': 0.6801}
5	nice Tshirt	1	0.6	{'neg': 0.0, 'neu': 0.263, 'pos': 0.737, 'compound': 0.4215}
6	thanks	1	0.2	{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.4404}
7	the product is very nice	1	0.78	{'neg': 0.0, 'neu': 0.564, 'pos': 0.436, 'compound': 0.4754}
8	I was expecting something better ...	1	0.5	{'neg': 0.0, 'neu': 0.633, 'pos': 0.367, 'compound': 0.4404}
9	this product is good quality	1	0.7	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
0	very nice product	1	0.78	{'neg': 0.0, 'neu': 0.393, 'pos': 0.607, 'compound': 0.4754}

Şekil 23. Duygu Analizi (Vader) Sonuçları İşlenmiş Yorumların Şekli

5.6. Metin Özetleme Adımları

Metin özetleme işlemleri ile ürün almak isteyen kişilere ürün hakkında genel olarak insanların ne düşündüğü sonucu kısaca özetleyerek ortaya çıkartmak istenmiştir.

5.6.1. Veri işlemleri

Veri tabanında bulunan ürün yorumları kullanılmıştır.

5.6.2. Veri Ön İşleme

Algoritmaların düzgün çalışabilmesi için verilerin belli bir düzende olması gerekir. Bu yüzden veri ön işleme son derece önemli bir adımdır. Temel olarak veri ön işleme tokenization (bölme), etkisiz elemanları atma, kelimenin kökünü bulma adımlarından oluşur. İlk olarak veri tabanından gelen ürün yorumları Python'da tokenization işlemi ile paragraflar cümlelere bölünmüştür. Noktalama işaretleri de temizlenmiştir. İngilizce etkisiz elemanlar da yani stopwords kelimeler etkisiz hale getirilmiştir.

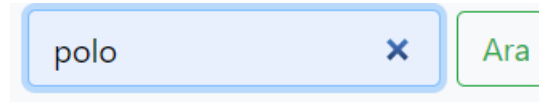
5.6.3. Metin Özetleme

Çıkarımsal metin özetleme yöntemlerinden kelime frekans değerine göre özetleme işlemi yapılmıştır. Bu yöntemde veriler temizlendikten sonra en çok tekrar eden yani en ağırlıklı kelime bulunur. Daha sonra diğer kelimelerin ağırlık sıklığını bulmak için tüm sözcüklerin geçiş sayısını en çok geçen sözcüğün sayısına bölerek kelimelerin ağırlığı bulunur. Bu

kelimelerin ağırlığı cümle puanlaması yapmak için bulunur. Daha sonra her cümlenin cümle puanı hesaplanır. Cümle puanı en yüksek olan cümleler karşımıza metin özetleme olarak çıkartılmıştır.

5.7. Django ile Sonuçların Gösterimi

Django sayfasının sağ üstüne navbar eklenerek ürün arama bölümü oluşturulmuştur.



Şekil 24. Navbar Arama Alanı

Arama bölümünde aratılmak istenen ürün ismi ya da ürün adında geçen herhangi bir şey ile sonuç döndürülmüştür. Şekil 25’de ürün adında Polo geçen her şey web sayfasında gösterilmiştir.

Sıra	Ürün Adı
1	Erkek %100 Pamuk Koyu Vizon Polo Yaka Düz T-shirt E001004
2	Erkek %100 Pamuk Beyaz Polo Yaka Düz T-shirt E001004
3	Erkek %100 Pamuk Açık Turuncu Polo Yaka Düz T-Shirt E001004
4	Erkek Lacivert Polo Yaka %100 Pamuk Düz T-shirt E001004

Şekil 25. Django ile Arama Yerinde Yapılan Ürün Sonuçları

Python web uygulaması Django ile ürünlerin bulunduğu bir sayfa tasarlanmıştır. Bu sayfada ürün adları, markaları, ürün linkleri ve ürünün değerlendirmesini sağlayan bir buton eklenmiş ve Şekil 26’da gösterilmiştir.

Ürünler

Sıra	Ürün Adı	Marka	Link	Analiz
4001	Kadın Yeşil Çizgili Cepli Pamuklu Gömlek 1YAK68399OW	KOTON	https://www.trendyol.com/koton/kadin-yesil-cizgili-cepli-pamuklu-gomlek-p-103609911	Ürün Değerlendirmesi
4002	Kadın Siyah Firfir Detaylı Bluz 9KAK68434PW	KOTON	https://www.trendyol.com/koton/kadin-siyah-firfir-detayli-bluz-9kak68434pw-p-4226509	Ürün Değerlendirmesi
4003	Carmen Pantolon 1YAK47402DD	KOTON	https://www.trendyol.com/koton/carmen-pantolon-p-98341173	Ürün Değerlendirmesi
4004	Crop Bluz İnce Askılı Çizgili 1YAK63729EW	KOTON	https://www.trendyol.com/koton/crop-bluz-ince-askili-cizgili-p-105557783	Ürün Değerlendirmesi
4005	Regular Fit Polo Yaka Uzun Kollu Gabardin Gömlek W7830AZ22SP	DeFacto	https://www.trendyol.com/defacto/regular-fit-polo-yaka-uzun-kollu-gabardin-gomlek-p-227175236	Ürün Değerlendirmesi

Şekil 26. Django Ürünler Sayfası

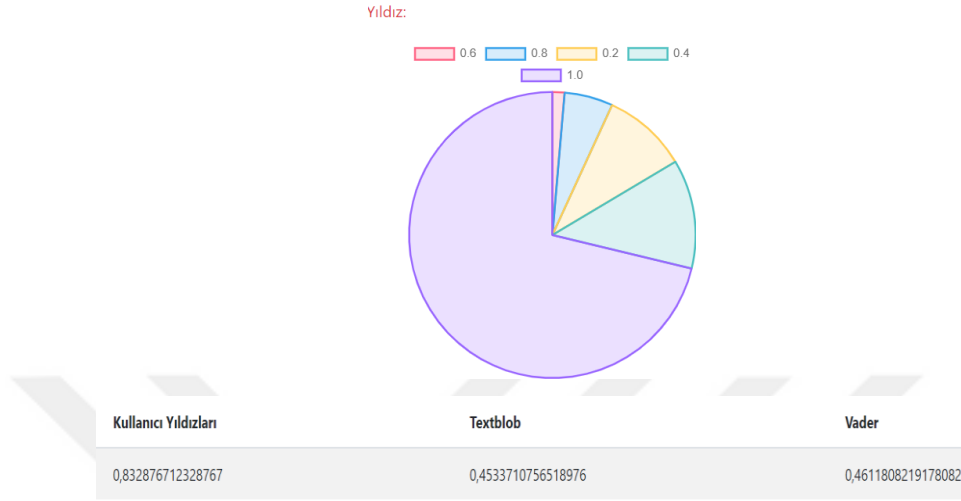
Ürün sırasına tıklanıldığında o ürünle ilgili ürün yorumları, yorumların İngilizceye çevrilmiş hali, ürün yorumlarının duygu analizi kütüphanelerine göre sonuçları Şekil 27’de gösterilmiştir.

Ürün Yorumu-en	Yıldız	Ürün Yorumu-TR	TextBlob	Vader
I bought it for my dad it's very nice	1,0	Babam için almıştım aşırı güzel	0.78	{'neg': 0.0, 'neu': 0.721, 'pos': 0.279, 'compound': 0.4754}
the texture is very nice, the quality is very good	1,0	Dokusu çok güzel kalitesi çok iyi	0.845	{'neg': 0.0, 'neu': 0.56, 'pos': 0.44, 'compound': 0.742}
The color and quality are very nice.	1,0	renk ve kalite çok güzel AVVA sonuçta indirimde almıştım nişanlım severek kullanıyor	0.78	{'neg': 0.0, 'neu': 0.66, 'pos': 0.34, 'compound': 0.4754}
i like it very nice	1,0	Çok beğendim güzel	0.78	{'neg': 0.0, 'neu': 0.349, 'pos': 0.651, 'compound': 0.6801}

Şekil 27. Django ile Duygu Analizi Sonuçlarının Gösterimi

Ayrıca ürünler sayfasında bulunan her ürünün yanında olan ürün değerlendirmesi butonuna tıklayınca ürün analiz sayfası açılmaktadır. Bu sayfada ürüne verilen yıldızların sayısına göre pasta grafiği şeklinde görünümü, müşterilerin verdiği yıldızların, vader ve textblob

kütüphanelerinin de ortalama değeri Şekil 28’de gösterilmiştir. Müşterilerin verdiği 1 yıldız 0.2, 2 yıldız 0.4, 3 yıldız 0.6, 4 yıldız 0.8, 5 yıldız 1.0 olarak gösterilmiştir.



Şekil 28. Duygu Analizi ve Kullanıcı Yorumlarının Ortalama Değerleri

6. BULGULAR

6.1. Alışveriş Yapan İnsanların Markaya ve Ürüne Göre Yaptığı Yorum Arasındaki Duygu Analizi Bulguları

34007 yorum, markalara ve ürünlere göre gruplanmıştır.

6.1.1. Markaya Göre Yapılan Çalışmalar

Gruplanan bu veriler yapılan duygu analizi çalışmasıyla yorumu en olumlu olan markalar, değerleri ile birlikte Tablo 3’de gösterilmiştir.

Tablo 3. Textblob ve Vader’e Göre Yorumu En Olumlu Olan Markalar

Sıra	Marka	Textblob değeri
1	Avva	%82.8
2	Suwen	%80.5
3	US Polo	%79.9
4	Hummel	%78.2
5	Mavi	%77.6

Sıra	Marka	Vader değeri
1	Avva	%82.1
2	Hummel	%79.4
3	Suwen	%78.9
4	US Polo	%78.1
5	Mavi	%77

Markalara göre yapılan çalışmada iki kütüphane de Avva ürünlerine yapılan yorumların en olumlu yorumlar olduğunu tespit etmiştir. Sıralamada bulunan değerler birbirine çok yakındır. İki sözlük tabanlı duygu analizi kütüphanesi de birbirlerine benzer değerler üretmiştir. Ayrıca iki duygu analizi algoritması da ilk 5 içinde aynı markaları bulmuştur. Fakat sıralamalarda farklılıklar gözükmemektedir. Textblob kütüphanesine göre sıralama Avva’dan sonra Suwen, Polo, Hummel ve Mavi şeklindedir. Vader kütüphanesinde ise Hummel, Suwen, Polo ve Mavi şeklinde olmuştur.

Duygu analizi çalışmasına göre yorumları yüzde olarak en olumsuz olan markalar Tablo 4’de gösterilmiştir.

Tablo 4. Textblob ve Vader’e Göre Yorumu En Olumsuz Olan Markalar

Sıra	Marka	Textblob değeri
1	Colin’s	%23.4
2	Koton	%17
3	Loft	%13.8
4	De Facto	%12

Sıra	Marka	Vader değeri
1	Loft	%25
2	Colin’s	%24.8
3	Koton	%19
4	De Facto	%15

Yapılan çalışmada iki ayrı kütüphane aynı dört markayı bulmuştur. Colin’s, Koton ve De Facto markalarındaki değerler birbirine çok yakın çıkmıştır. Buradaki en büyük farklılık Loft markasında olmuştur.

6.1.2. Ürüne Göre Yapılan Çalışmalar

Ürün yorumlarına göre yapılan duygu analizi çalışmasına göre en olumlu ve en olumsuz yorumlar aşağıda gösterilmiştir.

Vader’a göre en olumlu yorumlar;

1.Ürünün başlığında Cindy yazıyor ama Star modeli geliyor açıklamada da yazmışlar ikisinin arasındaki fark Cindy; yüksek bel, Star; süper yüksek bel. 160 boy 53 kilodayım 25/25 beden aldım beden tam oldu boy da tam oldu ama eğer paçalarını kıvrırmayı seviyorsanız boy bedenini büyük seçebilirsiniz. Ben çok beğendim şahsen rengi de güzel belinin süper yüksek bel olması da güzel. (Vader değeri: 0.99)

2. Bayıldım iyi ki almışım benim boyum 163 kilom 48 50 arası biraz sıktı ama yine de süper çok hoş çok sevdim bugüne kadar aldığım en güzel Jean aşırı kaliteli ve hayalimde ki Jean böyle bir jeani o kadar aradım ki ama bulamamıştım 25 27 de alsam olurmuş ama iyi ki almışım alın pişman olmazsınız (Vader değeri: 0.98)

Textblob’a göre en olumlu yorumlar;

1. Özenle paketlenmiş, kumaşı ve rengi çok güzel. (Textblob değeri:1.0)
2. Aşırı mükemmel tavsiye ederim kumaşı her şeyi efsane (Textblob değeri:1.0)

6.2. Alışveriş Yapan İnsanların Ürünlere Verdikleri Yıldızlara Göre Metin Özetleme Bulguları

Alışveriş yapan insanların aldıkları ürüne verdikleri yıldızla göre gruplayıp yorumları metin özetleme ile Tablo 5’de gösterilmiştir.

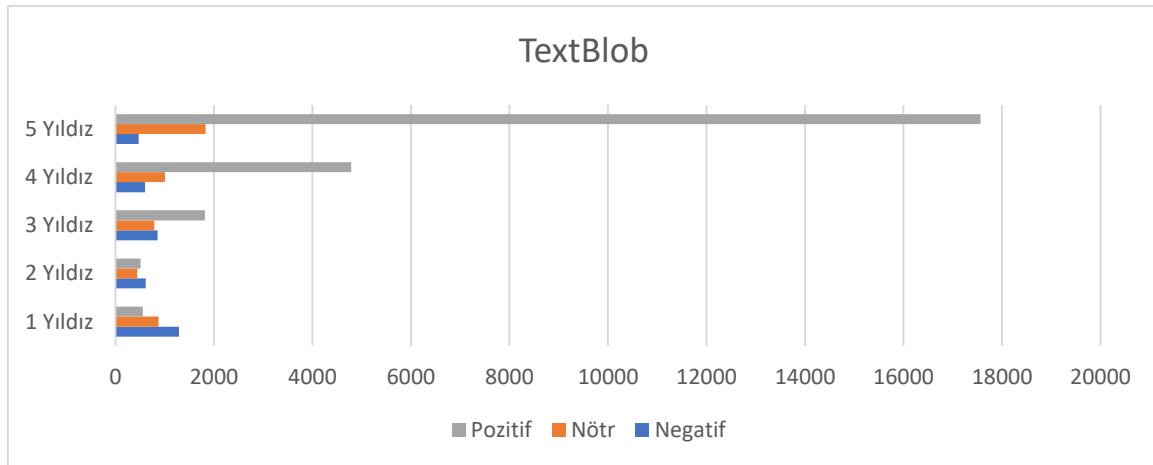
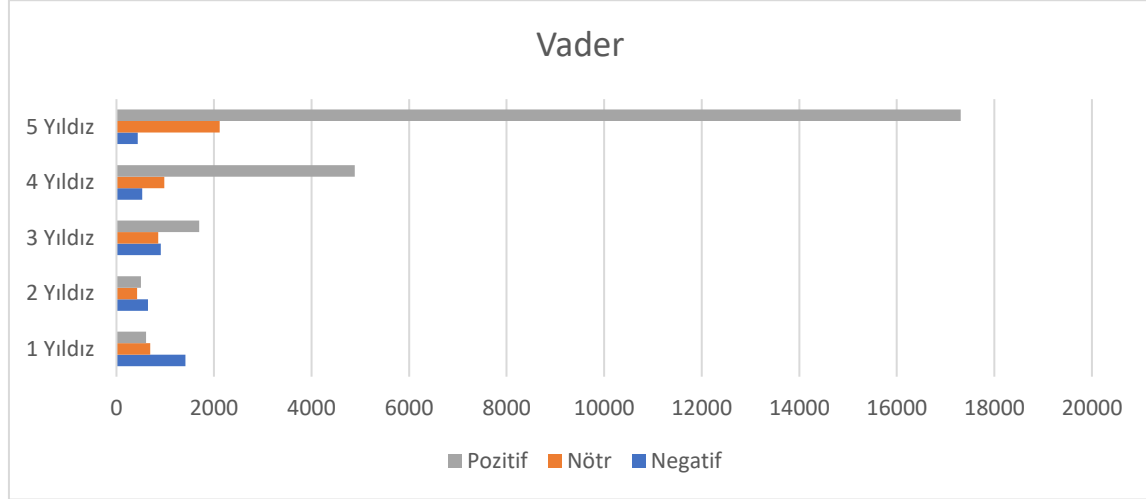
Tablo 5. Verilen Yıldızlara Göre Metin Özetleme Sonuçları Tablosu

Marka	Ürün Adı	1 Yıldız	2 Yıldız	3 Yıldız	4 Yıldız	5 Yıldız
Avva	Erkek Polo Yaka Düz Tişört	Kullanışsız, Geniş kalıp, Çok bol, Küçük alın	Satıcı ilgisiz, Değişim, Küçük alın	Kaliteli, Güzel, Satın alınabilir	Ürün kaliteli, Biraz büyük, Uygun	Kargo hızlı, Avva kalitesi, Harika, Kaliteli
Mavi	Serenay X Icon Broadway Flare Siyah Jean Pantolon	Kalıp dar, Kalıbı hoş değil Fotoğraftaki, gibi değil	Toz topluyor, Büyük alın, Kumaş ince, Kumaş tatmin edici değil	Kumaş ince, Toz topluyor	Kalıbı güzel, Toz topluyor, esnek kumaş, kalite güzel fakat dar	Çok şık, Harika model, Tam bedeninizi alın, Hızlı kargo
Hummel	Arus Erkek Lacivert Kapüşonlu Sweatshirt	Ürün kusurlu, Rengi soluk, Defolu	Kalıp dar, Rengi soluk, Mevsimlik, Dikişi sorunlu	Güzel ürün, Bedeni dar, İade	Baharlık, Hediye, Kalite güzel, Bir beden büyük alınabilir	Çok şık, Harika model, İnce dokulu, Hediye, Kumaşı güzel
US POLO	Gri Erkek T-Shirt	Kalıp küçük, Ürün tüylü, Kalitesiz, Dar kesim	Beden küçük, Kumaş farklı	Kargo geç, Ürün güzel, Yıkınca çekme sorunu	Kargo geç, Dokusu sert, Ürün kaliteli	Kalitesi güzel, Rengi güzel, Tam beden alın
Suwen	Ful Dolgulu Fırfırlı Straplez Üst	Bel kısmı çok dar, Aşırı destekli, Sorulara cevap verilmedi	Çok dar, Küçük	Güzel ama sıkıyor, Büyük alın	Ürün çok kaliteli, Çok tatlı, Arkası kısa	Ürün çok güzel, Büyük alın, Çok beğendim, Çok iyi
Koton	Yüksek Bel, Düğme Detaylı, Pantolon	Ürün farklı, Fotoğraftaki ürünle alakası yok, İade	Kalıbı büyük, Ürün görseldekiyle alakasız, Farklı	Rengi kötü, Başka ürün geldi, Daha koyu, Kumaş normal	Farklı ama güzel, Kalıp geniş, Benzer ürün, Yumuşak	Tam beden, Çok güzel model

Yapılan metin özetleme işlemi ile birlikte verilen yıldızlara göre kısa ve önemli detaylar ortaya çıkartılmıştır. Bu sayede alışveriş yapmak isteyen insanlara yorumların önemli noktaları daha hızlı ve kolay bir şekilde göstererek zaman ve zihin yönünden kolaylık sağlanmaya çalışılmıştır.

6.3. Alışveriş Yapan İnsanların Verdiği Yıldız ile Yaptığı Yorum Arasındaki Duygu Analizi Bulguları

Aşağıdaki tabloda Textblob ve Vader kütüphanelerinde değeri 0'dan küçük olanlar negatif, 0 ve 0.1 arası olanlar nötr, 0.1 ve 0.1' den büyük olanlar ise pozitif yorum olarak değerlendirilip kullanıcıların verdiği yıldızla karşılaştırmıştır ve sonuçlar Şekil 29'da gösterilmiştir.



Şekil 29. Alışveriş Yapan İnsanların Verdiği Yıldız ile Yaptığı Yorum Arasındaki Duygu Analizi Sonuçları

Müşterilerin 4 ve 5 yıldız verdiği ürünler genellikle olumlu yorumlar içeren ürünlerdir. Bu yorumlar sözlük tabanlı duygu kütüphaneleri olan Textblob'a göre %88,4 olumlu, Vader'a göre ise %87,1 olumlu tespit edilmiş ve Tablo 6'da gösterilmiştir.

Tablo 6. Olumlu Yorum Yapanların Duygu Analizi Sonucu

Müşterinin verdiği yıldız (4)	Negatif	Nötr	Pozitif
Textblob	%9.4	%15.7	%74.9
Vader	%8.3	%15.2	%76.4
Müşterinin verdiği yıldız (5)			
Textblob	%2.3	%9.2	%88.4
Vader	%2.2	%10.6	%87.1

Genellikle üç yıldızlı yorumlar nötr olarak değerlendirilebilir. Fakat Textblob kütüphanesine göre sadece %22' si nötr yorum olarak, Vader kütüphanesine göre %24'ü nötr yorum olarak bulunmuştur. 3 yıldız verilen ürünler Vader 'a göre %49 olumlu, Textblob'a göre ise %52 olumlu yorum olduğu tespit edilmiştir. 1 ve 2 yıldız verilen yorumlar genellikle olumsuz yorumların olduğu ürünler olarak düşünülür. Yapılan çalışmaya göre 1 ve 2 yıldız verilen ürünlerin yorumları da duygu analizi kütüphanelerine göre negatif ağırlıklı çıkmıştır. Fakat pozitif ve nötr yorum oranların toplam oranı da yüksek çıkmıştır. Tablo 7'de gösterilmiştir.

Tablo 7. Olumsuz Yorum Yapanların Duygu Analizi Sonucu

Müşterinin verdiği yıldız (1)	Negatif	Nötr	Pozitif
Textblob	%47.4	%32.1	%20.4
Vader	%52	%25.5	%22
Müşterinin verdiği yıldız (2)			
Textblob	%39.3	%28.3	%32.3
Vader	%41	%27	%32

7. SONUÇLAR VE ÖNERİLER

E-ticaret, günümüzde müşteriler için önemli bir alışveriş yöntemi haline gelmiştir. İnternetin ve mobil teknolojilerin yaygınlaşmasıyla birlikte, insanlar artık istedikleri ürünleri çevrimiçi olarak satın alma imkanına sahiptirler. Bu, müşterilere daha geniş bir ürün yelpazesi, kolay erişim ve zaman tasarrufu gibi avantajlar sunmaktadır. E-ticaretin müşteriler açısından bir diğer önemli yanı da duygu analizidir. Duygu analizi, müşterilerin çevrimiçi platformlarda yaptıkları yorumların ve geri bildirimlerin analiz edilmesiyle müşteri duygularının anlaşılmasıdır. Bu, müşterilerin alışveriş deneyimlerini değerlendirmek ve ifade etmek için bir fırsat sunar. Duygu analizi, müşterilerin ürünler ve hizmetler hakkında olumlu, olumsuz veya tarafsız duygularını belirlemek için kullanılan bir tekniktir. Müşteri yorumlarındaki duygusal tonlar, diğer potansiyel müşteriler için önemli bir referans olabilir. Olumlu yorumlar, ürün veya hizmetin kalitesini ve müşteri memnuniyetini yansıtırken, olumsuz yorumlar müşterilerin yaşadığı sorunları ve beklentilerini ifade edebilir.

İnsanlar e-ticaret siteleri üzerinden alışveriş yaparken ürün yorumlarına baktıklarında çok fazla yorum görmektedir. Bu tez çalışmasında sözlük tabanlı araçlar ve kelime frekans özetleme yöntemleri kullanılarak müşteriler için ürün ve marka analiz sistemi geliştirilmiştir. Bu sistemle birlikte müşteriler, yorumları tek tek okumak yerine direk duygu analizi ve metin özetleme çalışmasının sonucuna bakarak almak istedikleri ürün ya da marka için fikir sahibi olacaklardır. Bu sayede müşterilerin zamandan tasarruf etmesi de sağlanmıştır. Aynı zamanda müşterilerin doğru ürüne karar vermesine yardımcı olarak maddi kayıpların yaşanması engellenmeye çalışılmıştır. Ayrıca alışveriş yapan müşterilerin verdiği yıldızlar ile yaptıkları yorum üzerine yapılan çalışmalar sonucunda, yüksek yıldız veren müşterilerin yorumlarının da olumlu olduğu fakat düşük yıldız veren müşterilerin yaptıkları yorumların yanıltıcı sonuç verdiği tespit edilmiştir.

Bu tez çalışması giyim üzerindeki ürün yorumları üzerine yapılmıştır. Diğer alanlar üzerinde de çalışma yapılarak farklı yorumlar yapılabilir. Ayrıca yorumlar İngilizceye çevrilerek çalışma yapılmıştır. Türkçe duygu analizi yapabilen kütüphaneler kullanılarak farklı çalışmalar yapılabilir.

KAYNAKÇA

- Ahmetoğlu, H. ve Daş, R. (2020). Türkçe Otel Yorumlarıyla Eğitilen Kelime Vektörü Modellerinin Duygu Analizi ile İncelenmesi. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 24 (2), 455-463.
- Akbulut, S. (2022). *Makine Öğrenmesi ve Derin Öğrenme Yöntemleri Kullanılarak Bist100 Endeksi ile Emtia, Döviz Fiyatları ve Gelişmekte Olan Ülkelerin Borsa Endeksleri Arasındaki İlişkinin İncelenmesi*. (Yüksek Lisans Tezi), Aksaray Üniversitesi Sosyal Bilimler Enstitüsü, Aksaray.
- Akça, M. (2017). *Polimeraz Zincir Reaksiyonlarında Primer Tasarımı İçin Bir Python Modülü*. (Yüksek Lisans Tezi), Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük.
- Akdeniz, A. ve Babaoğlu İ. (2022). *Makine Öğrenmesi Yöntemleri Kullanılarak Uzaktan Eğitim Konulu Türkçe Tweetlerin Duygu Analizi*. (Yüksek Lisans Tezi), Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Konya.
- Aktaş, Y., İnce, E., Çakır, A. ve Kutlu, A. (2016). *Wordnet ve Bilgisayar Ağ Terimleri Sözlüğünün Oluşturulması*. 15 Mayıs 2023 tarihinde <https://ab.org.tr/ab16/bildiri/21.pdf> sayfasından erişilmiştir.
- Alipour, N. (2023). *Türk Dilinde Derin Öğrenme ile Metin Özetleme*. (Doktora Tezi), Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum.
- Al-Shabi, M. (2020). *Evaluating The Performance Of The Most Important Lexicons Used To Sentiment Analysis And Opinions Mining*. IJCSNS International Journal of Computer Science and Network Security. 15 Nisan tarihinde https://www.researchgate.net/publication/343473213_Evaluating_the_performance_of_the_most_important_Lexicons_used_to_Sentiment_analysis_and_opinions_Mining sitesinden erişilmiştir.
- Amalia, A., Afifa, R. ve Herriyance, H. (2018). "Resource Description Framework

- Generation for Tropical Disease Using Web Scraping," 2018 IEEE International Conference on Communication, Networks and Satellite (Comnetsat), Medan, Indonesia, 2018, pp. 44-48, doi: 10.1109/COMNETSAT.2018.8684030.
<https://ieeexplore.ieee.org/document/8684030>
- Amanet, H. (2017). *Türkçe Sosyal Medya Metinlerinde Duygu Analizi*. (Yüksek Lisans Tezi), Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Ashna, M and Sunny, A. (2017). "Lexicon based sentiment analysis system for malayalam language," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2017, pp. 777-783, doi: 10.1109/ICCMC.2017.8282571.
<https://ieeexplore.ieee.org/document/8282571>
- Ayan, B., Kuyumcu, B. ve Ciylan, B. (2019). Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. *Gazi Üniversitesi Fen Bilimleri Dergisi*, 7 (2), 495-502.
- Aydın, Z. (2022). "Sentiment Analysis about Turkish TV Series with Web Scraping," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2022, pp. 1-4, doi: 10.1109/HORA55278.2022.9800089.
<https://ieeexplore.ieee.org/document/9800089>
- Aycul, F. (2019). *Veri Madenciliği Teknikleri ile Hava Yolu Firmalarının Tweetleri Üzerinden Duygu Analizi*. (Yüksek Lisans Tezi), Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü, Burdur.
- Bal, S. (2022). *Çıkarıcı Türkçe Metin Özetleme Performansını İyileştirmek için Yeni Yöntemler*. (Doktora Tezi), Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir.
- Bird, S., Klein, E. ve Loper, E. (2009). *Natural Language Processing with Python*. 1 Mayıs 2023 tarihinde <https://tjzhifei.github.io/resources/NLTK.pdf> sayfasından erişilmiştir.
- Bonta, V., Kumares, N. ve Janardhan, N. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. *Asian Journal of Computer Science and Technology* ISSN: 2249-0701, 8(2), 1-6.
- Bozkuş, A. (2021). *Veri kazıma yöntemleri ile siber tehdit istihbaratı aracının Geliştirilmesi*. (Yüksek Lisans Tezi), Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elâzığ.
- Camargo-Henríquez, I. and Núñez-Bernal, Y. (2022). "A Web Scraping based approach for

- data research through social media: An Instagram case," 2022 V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC), San Jose, Costa Rica, 2022, pp. 1-4, doi: 10.1109/AmITIC55733.2022.9941290.
<https://ieeexplore.ieee.org/document/9941290>
- Can, U. ve Alataş, B. (2017). Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi. *Int. J. Pure Appl. Sci.* 3(1),75-111.
- Chandra, R. ve Varanasi, B. (2015). *Python Request Essentials*.
- Chaudhari, S., Aparna, R., Tekkur, V., Pavan, G., and Karki, S. (2020), "Ingredient/Recipe Algorithm using Web Mining and Web Scraping for Smart Chef," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2020, pp. 1-4, doi: 10.1109/CONECCT50063.2020.9198450.
<https://ieeexplore.ieee.org/document/9198450>
- Chiapponi, E., Dacier, M., Thonnard, O., Fangar, M., Mattsson, M. and Rigal, V. (2022). "An industrial perspective on web scraping characteristics and open issues," 2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S), Baltimore, MD, USA, 2022, pp. 5-8, doi: 10.1109/DSN-S54099.2022.00012.
<https://ieeexplore.ieee.org/document/9833640>
- Çölkesen, İ. ve Kavzoğlu, T. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*, 144, 73-82.
- Demir, Ö., Chawai, A. ve Doğan, B. (2019). Türkçe Metinlerde Sözlük Tabanlı Yaklaşımlar Duygu Analizi ve Görselleştirme. *Int. Per. of Recent Tech. in App.*, 2: 58-66.
- Ding, Y., Li, B., Zhao Y. and Cheng, C. (2017). "Scoring tourist attractions based on sentiment lexicon," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2017, pp. 1990-1993, doi: 10.1109/IAEAC.2017.8054363.
<https://ieeexplore.ieee.org/document/8054363>
- Dobashi, K. (2021). "Real-Time Web Scraping for Analyzing Moodle Course Material

- Clickstream," 2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET), Sydney, Australia, 2021, pp. 1-10, doi: 10.1109/ITHET50392.2021.9759582.
<https://ieeexplore.ieee.org/document/9759582>
- Durga, R., Sundaramoorthy, K. and Nagadarshini, S. (2017). "NewsOne — An Aggregation System for News Using Web Scraping Method," 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), Melmaurvathur, India, 2017, pp. 136-140, doi: 10.1109/ICTACC.2017.43.
<https://ieeexplore.ieee.org/document/8067594>
- Düven, B. (2021). *Yapay Öğrenme Yöntemleri ile Türkçe Metinlerde Duygu Analizi*. (Yüksek Lisans Tezi), Maltepe Üniversitesi, Lisansüstü Eğitim Enstitüsü, İstanbul.
- Er, M. (2019). *Makine Öğrenmesi ile Türk Müziğinde Duygu Analizi*. (Doktora Tezi), Maltepe Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Erdağı, E. (2023). *Türkçe Metinlerde Çıkarım Tabanlı Otomatik Metin Özetleme*. (Doktora Tezi), Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.
- Erhandı, B. (2020). *Derin Öğrenme ile Metin Özetleme*. (Yüksek Lisans Tezi), Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya.
- Ertam, F. (2018). "Deep learning based text classification with Web Scraping methods," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-4, doi: 10.1109/IDAP.2018.8620790.
<https://ieeexplore.ieee.org/document/8620790>
- Eskin, K. (2022). *Sosyal Robotlar İçin Makine Öğrenmesi Tabanlı Duygu Analizi*. (Yüksek Lisans Tezi), Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Gujjar, P. ve Kumar, P. (2021). *Sentiment Analysis:Textblob For Decision Making*. 3 Mayıs 2023 tarihinde
https://ijsret.com/wp-content/uploads/2021/03/IJSRET_V7_issue2_289.pdf
erişilmiştir.
- Gundecha, U. (2014). *Learning Selenium Testing Tools with Python*.
- Hasanlı, H. (2019). *Büyük Veri Üzerinde Duygu Analizi Yöntemleri ve Azerbaycan Diline Uygulanması*. (Doktora Tezi), Ege Üniversitesi Fen Bilimleri Enstitüsü, İzmir.
- Hazarika, D., Konwar, G., Deb, S. ve Bora, D. (2020). *Sentiment Analysis on Twitter by Using TextBlob for Natural Language Processing*. Proceedings of the International Conference on Research in Management & Technovation pp. 63–67.

- İlgün, H. (2022). *Dönüştürücüler ve derin öğrenme modelleriyle sosyal medya duygu Analizi*. (Yüksek Lisans Tezi), On Dokuz Mayıs Üniversitesi, Lisansüstü Eğitim Enstitüsü, Samsun.
- Julian, L. and Natalia, F. (2015). "The use of web scraping in computer parts and assembly price comparison," 2015 3rd International Conference on New Media (CONMEDIA), Tangerang, Indonesia, 2015, pp. 1-6, doi: 10.1109/CONMEDIA.2015.7449152.
<https://ieeexplore.ieee.org/document/7449152>
- Junjoewong, L., Sangnapachai, S. and Sunetnanta, T. (2018). "ProCircle: A promotion platform using crowdsourcing and web data scraping technique," 2018 Seventh ICT International Student Project Conference (ICT-ISPC), Nakhonpathom, Thailand, 2018, pp. 1-5, doi: 10.1109/ICT-ISPC.2018.8524003.
<https://ieeexplore.ieee.org/abstract/document/8524003>
- Kara, E. (2019). *Metin Türünün Duygu Analizi Sınıflandırmasına Etkisi*. (Yüksek Lisans Tezi), Düzce Üniversitesi Fen Bilimleri Enstitüsü, Düzce.
- Karabulut, S. (2023). *Yapay Sinir Ağları Algoritmaları ile Pvp Nanoliflerin Çaplarının Tahmini*. (Yüksek Lisans Tezi), Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü, Afyon.
- Khder, M. (2021). *Web Scraping Or Web Crawling: State Of Art, Techniques, Approaches And Application*. International Journal of Advances in Soft Computing and its Applications, 13(3),145-168.
- Koç, Y. (2022). *Döviz Kuru Tahmini İçin Türkçe Tweetler ile Twitter Duygu Analizi*. (Yüksek Lisans Tezi), Başkent Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Kouzis-Loukas, D. (2016). *Learning Scrapy, Learn The Art Of Efficient Web Scraping and Crawling With Python*.
- Kör, H., (2019). *Meme Kanserinin Makine Öğrenmesi Yöntemleri ile İkili Sınıflandırılması*. 4th International Symposium on Innovative Approaches in Engineering and Natural Sciences, Çorum, Türkiye.
- Krotov, V., Johnson, L. and Silva, L. (2020). *Tutorial: Legality and Ethics of Web Scraping*. 5 Nisan tarihinde
<https://digitalcommons.murraystate.edu/cgi/viewcontent.cgi?article=1071&context=faculty> sayfasından erişilmiştir.
- Loria, S. (2020). *Textblob Documentation*. 25 Mayıs 2023 tarihinde

- <https://textblob.readthedocs.io/en/dev/> sayfasından erişilmiştir.
- Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment Analysis Algorithms And Applications: A survey, *Ain Shams Engineering Journal*, 1093-1113.
- Mishra, P., Rajnish, R. and Kumar, P. (2016). "Sentiment analysis of Twitter data: Case study on digital India," 2016 International Conference on Information Technology (InCITE) - The Next Generation IT Summit on the Theme - Internet of Things: Connect your Worlds, Noida, India, 2016, pp. 148-153, doi: 10.1109/INCITE.2016.7857607.
<https://ieeexplore.ieee.org/document/7857607>
- Mitchell, R. (2015). *Web Scraping With Python*.
- Mutlu, M. (2021). *Web Kazıma (Scraping) Tekniği ile Www'de Anahtar Kelime Tabanlı Literatür Taraması Uygulamasının Gerçekleştirilmesi*. (Yüksek Lisans Tezi), Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Mutlu, M. (2022). *Türkçe Metinlerde Hedef Odaklı Duygu Analizi*. (Yüksek Lisans Tezi), Boğaziçi Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul.
- Özkan, D. (2020). *Türkçe Hedef-Tabanlı Duygu Analizi*. (Yüksek Lisans Tezi), Atılım Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Öztürk, K. ve Şahin, M. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-i Vekayi Dergisi*, 6(2), 25-36.
- Özyurt, B. ve Akcayol, M. (2018). Fikir Madenciliği ve Duygu Analizi, Yaklaşımlar, Yöntemler Üzerine Bir Araştırma. *S.Ü. Müh. Bilim ve Tekn. Dergisi*, 6(4), 668-693.
- Pak, A. ve Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *In LREc*, 10, 1320-1326.
- Pek, M. (2019). *Makine Öğrenmesi Yöntemleri ile Sosyal Medya Verilerinden Duygu Analizi*. (Yüksek Lisans Tezi), İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
- Polat, B. (2021). *Türkçe Ürün Yorumları Verisi ile Duygu Analizi*. (Yüksek Lisans Tezi), Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- R. R. N. R, N. R. S and V. M., "Web Scrapping Tools and Techniques: A Brief Survey," 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2023, pp. 1-4, doi: 10.1109/ICITIIT57246.2023.10068666.
<https://ieeexplore.ieee.org/document/10068666>

- Rastogi, N., Yadav, K., Divya, K. and Perwej, Y. (2022). Sentimental Analysis on Web Scraping Using Machine Learning Method. *Journal of Information and Computational Science*, 12(8), 24-29.
- Reddy, C., Reddy, K., Anil, G., Mohanty, S. and Basit, A. (2023). "Laptop Price Prediction Using Real Time Data," 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), Jeddah, Saudi Arabia, 2023, pp. 1-5, doi: 10.1109/ICAISC56366.2023.10085473.
<https://ieeexplore.ieee.org/abstract/document/10085473>
- Richter, S. (2020). *Lxml*. 5 Mayıs 2023 tarihinde <https://lxml.de/> sayfasından erişilmiştir.
- Saralıoğlu, E. (2015). *Python Programlama Dili Kullanılarak Uzaktan Algılama Amaçlı Arayüz Tasarımı*. (Yüksek Lisans Tezi), Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Trabzon.
- Saurkar, A., Pathare, K. and Gode, A. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(4), 363-367.
- Savrul, B. ve Akin, F. (2021). Koronavirüs Salgınının Türkiye’de E-Ticaret Sektörüne Etkileri. *Erzincan Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, Özel Sayı, 15 Haziran tarihinde <https://dergipark.org.tr/tr/download/article-file/1840759> sayfasından erişilmiştir.
- Sevindi, B. (2013). *Türkçe Metinlerde Denetimli ve Sözlük Tabanlı Duygu Analizi Yaklaşımlarının Karşılaştırılması*. (Yüksek Lisans Tezi), Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Sirisuriya, S. (2015). A Comparative Study on Web Scraping. *Proceedings of 8th International Research Conference*, Ratmalana, Sri Lanka.
- Şahinaslan, Ö., Dalyan, H. ve Şahinaslan, E. (2021). Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi. *Bilişim Teknolojileri Dergisi*, 15(2), 221-229. 15 Nisan tarihinde <https://dergipark.org.tr/en/download/article-file/1991478> sayfasından erişilmiştir.
- Şatır, E., Azboy, F., Aydın, A., Arslan, H. ve Hacıefendioğlu, Ş. (2016). Veri İndirgeme ve Sınıflandırma Teknikleri ile Glokom Hastalığı Teşhisi. *El-Cezerî Fen ve Mühendislik Dergisi*, 3(3), (485-497). 15 Nisan tarihinde <https://dergipark.org.tr/tr/download/article-file/225051> sayfasından erişilmiştir.

- Toçoğlu, M. ve Alpkoçak, A. (2018). Lexicon-Based Emotion Analysis In Turkish. *Turk J Elec Eng & Comp Sci*, 27(2), 1213 – 1227.
- Uçkan, T. ve Karcı, A. (2020). "Extractive multi-document text summarization based on graph independent sets". *Egyptian Informatics Journal*, 21(3), 145-157.
- Uzundere, E., Dedja, E., Diri, B., & Amasyalı, M. F. (2008). Türkçe Haber Metinleri İçin Otomatik Özetleme. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu (pp.1-4). Isparta, Turkey.
- Vargiu, E. and Urru, M. (2013). Exploiting web scraping in a collaborative filteringbased approach to web advertising. *Artificial Intelligence Research*, 2(1), 45-54.
- Verma, A., Mittal, V. and Dawn, S. (2019). "FIND: Fake Information and News Detections using Deep Learning," 2019 Twelfth International Conference on Contemporary Computing (IC3), Noida, India, 2019, pp. 1-7, doi: 10.1109/IC3.2019.8844892. <https://ieeexplore.ieee.org/document/8844892>
- Yengi, Y. ve Omurca, S. (2016). Büyük Verilerde Duygu Analizi Öneri Sistemleri. *Avrupa Bilim ve Teknoloji Dergisi*, 4(7), 51-57.
- Yıldırım, M. ve Avcı, İ. (2021). Görme Engelli Bireyler İçin Derin Öğrenme Tabanlı Nesne Tanıma Modeli. *Avrupa Bilim ve Teknoloji Dergisi*, Özel Sayı 28, 220-227.
- Yıldırım, S., Salman, Y. ve Ayvaz, S. (2019). Türkçe Duygu Kütüphanesi Geliştirme: Sosyal Medya Verileriyle Duygu Analizi Çalışması. *Avrupa Bilim ve Teknoloji Dergisi*, Sayı 16, 51-60.
- Yılmaz, H.ve Yumuşak, S. (2021). Açık Kaynak Doğal Dil İşleme Kütüphaneleri. *İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 3 (1), 81-85.
- Yuan, G., Ho, C. and Lin, C. (2012). "Recent Advances of Large-Scale Linear Classification". 100(9), 2584-2603. 15 Nisan 2023 tarihinde <http://www.jcomputers.us/vol10/jcp1006-03.pdf> sayfasından erişilmiştir.
- Yumak, B. (2011). *Elektronik Postaların Ayırıştırılmasında Naive Bayesian ve Bulanık Mantık Yöntemlerinin Karşılaştırılması*. (Yüksek Lisans Tezi), Gazi Üniversitesi Bilişim Enstitüsü, Ankara.
- Zheng, C., He, G. and Peng, Z. (2014). *A Study of Web Information Extraction Technology Based on BeautifulSoup*. 10 Mayıs 2023 tarihinde <http://www.jcomputers.us/vol10/jcp1006-03.pdf> sayfasından erişilmiştir.

Aksaray Üniversitesi Tez Değerlendirme Formu		
Öğrencinin Adı Soyadı:		EVET
Kapak		
1	Tez Başlığı tutanaktaki başlıkla aynı mı?	
2	Kapakdaki ay ve yıl savunmaya girilen tarihle tutarlı mı?	
3	Kapak format kılavuzdaki kapak formatına uygun mu?	
4	Kapakta yazılan tüm yazılar doğru olarak verilmiş mi?	
İçindekiler		
5	Sayfa numaraları tam verilmiş mi?	
6	Şekil, Çizelge vb. listeleri verilmiş mi? Sıralaması doğru mu?	
7	Özet, Abstract, Giriş, Sonuçlar vb. bölümler var mı?	
8	Yazım hataları kontrolü yapıldı mı?	
Giriş		
9	Hazırlanan tezin önemini anlatıyor mu?	
10	İkinci ve Üçüncü dereceden başlık içermemeli kuralına uyuldu mu?	
Özet/Abstract		
11	Kılavuza uygun mu?	
12	Ay ve yıl savunmaya girilen tarihle tutarlı mı?	
13	Özet; tek sayfa, tek aralık, tek paragraf kuralına uygun olarak yazıldı mı?	
14	Bilim kodu, sayfa adedi, anahtar kelimeler ve tez danışmanı yazıldı mı?	
Kaynakça		
15	Kaynakların tamamına metin içinde atıf yapıldı mı?	
16	Kaynak formatı Kılavuzdaki kaynak formatına uygun olarak hazırlanmış mı?	
17	Atıf formatı kılavuzdaki atıf formatına uygun mu?	
Atıf Yöntemi APA 6 ☐ CMS ☐ İSNAD ☐		
Genel Değerlendirme		
18	Etik Beyan açıklaması okundu, uyuldu ve imzalandı mı?	
19	Kabul/Onay sayfası kılavuzdaki formata uygun olarak düzenlenmiş mi?	
20	Kabul /Onay sayfasında belirtilen oy birliği/oy çokluğu seçeneklerinden uygun olanı savunmayla tutarlı olacak şekilde belirlenmiş mi?	
21	Sayfa kenar boşluklar ve sayfa numaraları kılavuzdaki formatına uygun mu?	
22	Paragraf boşlukları ve metin satır aralığı kılavuza uygun olacak şekilde düzenlenmiş mi?	
23	Başlık yazımları kılavuzdaki başlık formatlarına uygun mu?	
24	Yazı tipi ve boyutu kılavuzdaki yazı tipi ve boyutu formatına uygun mu?	
25	Şekil, Çizelge vb. açıklama ve numaralandırmaları kılavuzdaki formata uygun olarak yazılmış mı?	

Bu tezin tarafımdan “Tez yazım kuralları” okunarak dikkatlice hazırlanmış olduğunu ve doğabilecek her türlü olumsuzluktan sorumlu olacağımı kabul ederim.

Öğrencinin imzası