

T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

İŞ ZEKASI VE DUYGU ANALİZİ: SAKARYA MERKEZLİ  
SOSYAL MEDYA VERİLERİNİN DOĞAL DİL İŞLEME  
YAKLAŞIMLARIYLA İNCELENMESİ

YÜKSEK LİSANS TEZİ

Furkan SARAÇOĞLU

Bilişim Sistemleri Mühendisliği Anabilim Dalı

ŞUBAT 2024



T.C.  
SAKARYA ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

İŞ ZEKASI VE DUYGU ANALİZİ: SAKARYA MERKEZLİ  
SOSYAL MEDYA VERİLERİNİN DOĞAL DİL İŞLEME  
YAKLAŞIMLARIYLA İNCELENMESİ

YÜKSEK LİSANS TEZİ

Furkan SARAÇOĞLU

Bilişim Sistemleri Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. İsmail Hakkı CEDİMOĞLU

ŞUBAT 2024



Furkan SARAÇOĞLU tarafından hazırlanan “İş Zekası ve Duygu Analizi: Sakarya Merkezli Sosyal Medya Verilerinin Doğal Dil İşleme Yaklaşımlarıyla İncelenmesi” adlı tez çalışması 09.02.2024 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilişim Sistemleri Mühendisliği Anabilim Dalı, Yüksek Lisans tezi olarak kabul edilmiştir.

### Tez Jürisi

**Jüri Başkanı :**            **Doç. Dr. Halit ÖZTEKİN**            .....

Sakarya Uygulamalı Bilimler Üniversitesi

**Jüri Üyesi :**                **Prof. Dr. İsmail Hakkı CEDİMOĞLU**            .....

Sakarya Üniversitesi

**Jüri Üyesi :**                **Doç Dr. İhsan Hakan SELVİ**                        .....

Sakarya Üniversitesi



## **ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ**

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “İş Zekası ve Duygu Analizi: Sakarya Merkezli Sosyal Medya Verilerinin Doğal Dil İşleme Yaklaşımlarıyla İncelenmesi” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(09/02/2024).

Furkan SARAÇOĞLU





*Değerli aileme ve dostlarıma*



## **TEŐEKKÜR**

Tez sürecimde beni bilgeliđi, önerileri ve sabrıyla yönlendiren saygıdeđer danıőmanım Prof. Dr. İsmail Hakkı CEDİMOĐLU'na teőekkürlerimi ve saygılarımı sunarım.

Eđitim yolculuđumda hiçbir zaman yanımdan ayrılmayan kıymetli annem, Ayőe ve babam Mehmet SARAĐOĐLU'na; bu yolculuktaki büyük emekleri için abim Özgür SARAĐOĐLU'na, sürekli destekleri ve cesaretlendirmeleriyle yanımda olan bütün dostlarıma içtenlikle teőekkür ederim.

Furkan SARAĐOĐLU



## İÇİNDEKİLER

### Sayfa

<b>ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ</b> .....	<b>v</b>
<b>TEŞEKKÜR</b> .....	<b>ix</b>
<b>İÇİNDEKİLER</b> .....	<b>xi</b>
<b>KISALTMALAR</b> .....	<b>xiii</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>xv</b>
<b>ÖZET</b> .....	<b>xvii</b>
<b>SUMMARY</b> .....	<b>xix</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. LİTERATÜR ARAŞTIRMASI</b> .....	<b>3</b>
<b>3. TEORİK ÇERÇEVE</b> .....	<b>9</b>
3.1. Doğal Dil İşleme .....	9
3.1.1. Metin Madenciliği ve Ön İşleme Aşamaları .....	9
3.1.2. Kategori Tahmini ve Duygu Analizi.....	12
3.2. İş Zekası .....	12
3.2.1. Veri Ambarı Mimarisi.....	13
3.2.2. Veri Ambarı Mimari Yaklaşımları.....	14
3.2.3. Veri Ambarı Mimari Katmanları .....	16
3.2.4. Veri Ambarı Mantıksal Model Yaklaşımları .....	17
3.2.4.1. Yıldız Şeması Mantıksal Modeli.....	17
3.2.4.2. Kartanesi Şeması.....	18
3.2.5. Ayıklama, Dönüştürme ve Yükleme (ETL).....	19
3.2.6. Veri Görselleştirme .....	20
<b>4. MATERYAL VE YÖNTEM</b> .....	<b>21</b>
4.1. Projede Kullanılan Araçlar ve Yöntemler .....	21
4.1.1. Twitter API .....	21
4.1.2. Python Programlama Dili.....	22
4.1.3. Spyder IDE.....	24
4.1.4. Azure Data Factory .....	24
4.1.5. Azure SQL Server .....	24
4.1.6. Power BI .....	25
4.2. Metin Sınıflandırma ve Makine Öğrenimi Algoritmasının Karşılaştırılması .	25
4.2.1. Rastgele Orman Sınıflandırıcı Algoritması.....	25
4.2.2. K - En Yakın Komşu Algoritması.....	26
4.2.3. Çok Terimli Naïve Bayes Algoritması.....	27
4.2.4. Lojistik Regresyon Algoritması .....	27
4.2.5. Destek Vektör Makineleri Algoritması.....	28
4.2.6. FastText Algoritması.....	29
4.2.7. Türkçe BERT Algoritması .....	29
<b>5. SAKARYA’NIN DUYGU ATLASI PROJESİ</b> .....	<b>31</b>
5.1. Sakarya’nın Duygu Atlası Proje Mimarisi .....	31
5.1.1. Twitter Verisilerinin Elde Edilmesi .....	31

5.1.2. Veri Ön İşleme .....	33
5.1.3. Modelin Çalıştırılması.....	33
5.1.4. ETL İşlemi.....	34
5.1.5. Veri Modelleme ve Görselleştirme .....	36
<b>6. BULGULAR .....</b>	<b>39</b>
<b>7. SONUÇLAR .....</b>	<b>43</b>
<b>KAYNAKLAR.....</b>	<b>45</b>
<b>ÖZGEÇMİŞ.....</b>	<b>51</b>

## KISALTMALAR

<b>API</b>	: Application Programming Interface
<b>BERT</b>	: Bidirectional Encoder Representations from Transformers
<b>CCG</b>	: Combinatory Categorical Grammar
<b>DA</b>	: Duygu Analizi
<b>DDİ</b>	: Doğal Dil İşleme
<b>ETL</b>	: Extract, Transform, Load
<b>IDE</b>	: Integrated Development Environment
<b>İZ</b>	: İş Zekası
<b>LR</b>	: Lojistik Regresyon
<b>MÖ</b>	: Makine Öğrenimi
<b>NB</b>	: Naive Bayes
<b>NLG</b>	: Natural Language Generation
<b>NLP</b>	: Natural Language Processing
<b>NLU</b>	: Natural Language Understanding
<b>ODS</b>	: Operational Data Source
<b>OLTP</b>	: Online Transactional Processing
<b>SQL</b>	: Structured Query Language
<b>SVM</b>	: Support Vector Machine





## ŞEKİL LİSTESİ

### Sayfa

Şekil 3.1. Metin Ön İşleme Adımları.....	11
Şekil 3.2. Kimball Veri Ambarı Mimarisi, Kimball (2002)'den uyarlanmıştır. ....	15
Şekil 3.3. Inmon Veri Ambarı Mimarisi, Inmon (2002)'den uyarlanmıştır. ....	16
Şekil 3.4. Veri Ambarı Mimari Katmanları, ElGamal (2016)'dan uyarlanmıştır.....	17
Şekil 3.5. Yıldız Şeması Mantıksal Modeli .....	18
Şekil 3.6. Kartanesi Şeması Mantıksal Modeli.....	19
Şekil 4.1. Rastgele Orman Sınıflandırıcı Algoritma Sonuçları.....	26
Şekil 4.2. K - En Yakın Komşu Algoritma Sonuçları.....	26
Şekil 4.3. Çok Terimli Naive Bayes Algoritma Sonuçları .....	27
Şekil 4.4. Lojistik Regresyon Algoritma Sonuçları .....	28
Şekil 4.5. Destek Vektör Makineleri Algoritma Sonuçları.....	28
Şekil 4.6. FastText Algoritma Sonuçları .....	29
Şekil 4.7. Türkçe BERT Algoritma Sonuçları .....	30
Şekil 5.1. Sakarya'nın Duygu Atlası Proje Mimarisi .....	31
Şekil 5.2. İşlenmemiş Tweet Metinleri .....	32
Şekil 5.3. Ön İşleme Yapılmış Tweet Metinleri .....	33
Şekil 5.4. Tahminleme Yapılmış Metinler.....	34
Şekil 5.5. ETL İşlemi Sonrası Metinler .....	35
Şekil 5.6. İşlem Tablosu Veri Aktarımı .....	35
Şekil 5.7. Boyut Tabloları Veri Aktarımı .....	36
Şekil 5.8. Dim ve Fact Tabloları Veri Akışları.....	36
Şekil 5.9. Proje Veri Modeli .....	37
Şekil 6.1. Kategori Analizi Sayfası.....	39
Şekil 6.2. Duygu Analizi Sayfası.....	40
Şekil 6.3. Lokasyon Detay Sayfası .....	41



# İŞ ZEKASI VE DUYGU ANALİZİ: SAKARYA MERKEZLİ SOSYAL MEDYA VERİLERİNİN DOĞAL DİL İŞLEME YAKLAŞIMLARIYLA İNCELENMESİ

## ÖZET

Bu çalışma, sosyal medya kullanımının gün geçtikçe arttığı ve yaygınlaştığı bir döneme odaklanmaktadır. Özellikle Twitter gibi dinamik bir platformdan elde edilen sürekli güncellenen ve hızla artan metinsel veriler üzerinde yoğunlaşarak, bu verilerin analizini gerçekleştirmeyi amaçlamaktadır. Bu metin verileri, doğal dil işleme yöntemleri ve makine öğrenmesi algoritmaları kullanılarak madencilik yapılarak incelenmekte ve bu sayede sosyal medya ortamını, toplumu ve çevreyi oluşturan bireylerin düşünce, his ve duygularını anlaşılır bir şekilde ortaya koymayı hedeflemektedir. Temel olarak bu çalışma, metinsel verilerin derinlemesine incelenmesi ve analiz edilmesi yoluyla, sosyal medya kullanıcılarının zihinsel dünyalarını daha anlamlı bir şekilde ortaya çıkarmayı amaçlamaktadır.

Twitter gibi duygu ve düşüncelerin sıkça paylaşıldığı bir platform üzerinde gerçekleştirilen bu çalışmada, Tweet metinleri üzerinde gerçekleştirilen duygu analizi ve kategori tahminleri sayesinde elde edilen veriler, iş zekası mimarisi kullanılarak anlamlandırılmış ve çeşitli grafiklerle görselleştirilmiştir. Bu yaklaşım, geleneksel anket çalışmalarının yerine geçerek daha düşük maliyetle, hızlı erişim ve sonuçlar sunabilmeyi amaçlamaktadır. Twitter üzerinden elde edilen bu verilerin analizi, duygu ve düşüncelerin toplumda nasıl yayıldığını anlamak ve daha etkili bir veri toplama yöntemi sunmak amacıyla gerçekleştirilmiştir.

Bu araştırma, çeşitli yazılımlar ve yaklaşımların bir araya getirildiği bir çerçeve içinde gerçekleşmiş olup, uzun vadeli çıkarımlara ulaşmayı hedeflemektedir. İş zekası mimarisi ve araçlarıyla otomatikleştirilen çalışmanın adımları, belirli sonuçlara ve başarımlara ulaşmak için özenle düzenlenmiştir. Performans, optimum düzeyde tutularak veri kaybı ve yanlışlık oranı en düşük seviyeye çekilmiştir. Veri koruma ve yedekleme süreçleri için güvenilir veri tabanı sistemleri tercih edilmiştir. Bu şekilde, araştırmanın güvenilirliği artırılmış ve elde edilen verilerin güvenliği sağlanmıştır.

Temizlenmiş ve bilgiye dönüştürülmüş veriler, çeşitli veri görselleştirme uygulamaları ve çözümleri kullanılarak anlamlı bir görsel forma getirilmiş ve bu önemli bilgiler, etkili görsel raporlar şeklinde sunulmaktadır. Bu raporlar, karmaşık veri setlerini anlamak ve anahtar bulguları hızlıca görmek isteyenlere verileri anlaşılabilir kılmak amacıyla tasarlanmıştır.

Araştırmanın temel amacı, sadece geçici sonuçlara değil, aynı zamanda sürdürülebilir bir şekilde uzun vadeli çıkarımlara ulaşmaktır. Bu nedenle, verilerin anlamlandırılması ve görselleştirilmesi, araştırmanın gelecekteki analizlere ve karar süreçlerine değer katmasını sağlamak üzere özenle planlanmıştır. Bu görsel sunumlar, araştırma sürecinin etkili bir iletişim aracı olup, bulguların geniş bir izleyici kitlesine kolayca aktarılmasını amaçlamaktadır.



# **BUSINESS INTELLIGENCE AND SENTIMENT ANALYSIS: EXAMINING SAKARYA-CENTRIC SOCIAL MEDIA DATA THROUGH NATURAL LANGUAGE PROCESSING APPROACHES**

## **SUMMARY**

This study focuses on a period in which the use of social media is increasing and becoming more widespread. Specifically, it aims to analyze the continuously updated and rapidly increasing textual data obtained from dynamic platforms like Twitter. These text data are examined through natural language processing methods and machine learning algorithms, aiming to understand the thoughts, feelings, and emotions of individuals shaping the social media environment, society, and the community. Fundamentally, the study aims to uncover the mental worlds of social media users more meaningfully by in-depth examination and analysis of textual data.

In this context, the application of techniques like sentiment analysis and social media analytics functions as a powerful toolkit that allows for both a deeper understanding and future prediction of user interactions and behaviors on different social media platforms. The ultimate goal of this investigation is to make sure that the knowledge extracted from social media data through the use of these advanced methods reaches a degree of depth and coverage necessary for well-informed decision-making and academic research.

Focused mostly on social media platforms where emotions and thoughts are frequently expressed, such as Twitter, this study involved careful data pretreatment efforts to enable sentiment analysis and category predictions in tweet messages. During this preparation stage, a methodical technique was used to thoroughly clean the text data. This included filtering out irrelevant phrases, eliminating uncertainties, and removing superfluous letters. The cleaned and refined dataset was then carefully selected to enhance the applicability and effectiveness of machine learning algorithms in the later stages of data analysis and prediction.

Using techniques like the Turkish BERT algorithm, which is well-known for having better performance metrics than other algorithms, the cleaned and documented dataset was processed once again. Subsequent data transformation and modeling processes were included into the business intelligence architecture framework in an elegant manner, resulting in data that was carefully prepared for visualization. These deliberate actions serve to improve the effectiveness of business intelligence-driven analysis in addition to enabling a more thorough understanding of insights obtained from Twitter data.

The data moved smoothly to the Power BI application, where it was carefully transformed into a multitude of visual representations through the establishment of interrelationships. These images were carefully selected and arranged on several pages labeled "Location Analysis," "Category Analysis," and "Sentiment Analysis." Notably, a detailed analysis was carried out under these specific parts, concentrating on the connections and discussion regarding shares that come from the province of Sakarya. In the sentiment analysis section, special attention was paid to identifying the

dominant feelings expressed in shares written by users from Sakarya province as well as a detailed examination of how they were distributed.

In the meantime, the category analysis went into explaining the themes contained in the shares and also looked closely at how they were distributed among categories. The location analysis section further clarified any apparent correlations with particular theme motifs and illuminated the spatial distribution of shares within Sakarya province. Essentially, the visual analyses selected by Power BI provide a comprehensive understanding of the sharing preferences and behavioral patterns displayed by Sakarya province's social media users.

The category analysis results explore the shared content's thematic landscape in Sakarya province and indicate a significant focus on topics related to "Politics" and "Culture." In these theme domains, a tendency toward "Very Positive" and "Neutral" attitudes was observed, highlighting Sakarya users' tendency to adopt a positive or neutral manner, especially when participating in conversations about political and cultural issues. These observations provide light on the general mindset of Sakarya province's social media users, suggesting a tendency to take a favorable or impartial viewpoint while discussing issues of political and cultural importance.

After analyzing sentiment analysis in more detail, several different emotions, such as "Very Positive," "Neutral," and "Negative," were shown to be common in shared content. But when one looked closely at measures like retweet and like counts, one interesting finding emerged: shares that elicited happy feelings received more interaction. This observation implies that users are more sensitive to content with a positive tone, as seen by their active desire to promote and spread such content on social media in Sakarya province. As a result, the results highlight the tendency of Sakarya province social media users to favor content that receives good feedback, highlighting a tendency to actively interact with and spread positive information.

The geography study produced some interesting findings: within Sakarya province, the districts with the highest number of tweet shares were Adapazarı and Serdivan. Even though the number of tweets from each of these districts varied, it was clear that shares from Serdivan received more attention and retweets. This phenomena points to the existence of a social media user base in Serdivan that is more vibrant and interaction-driven. These results highlight the potential impact of regional dynamics—which go beyond simple physical proximity—on the degree of participation among social media users.

With the ultimate goal of accelerating the acquisition of findings, this technique aims to replace traditional survey methodologies with more economical and efficient alternatives. The study's stated goal is to use social media data to better understand the mechanisms behind how ideas and emotions spread throughout society and to provide a more effective means of gathering data. Through the use of social media platforms to enable real-time data gathering and analysis, this method makes it possible to quickly and thoroughly generate insights, which in turn makes it possible to more quickly and thoroughly understand the emotional and intellectual dynamics of society.

Here, the goal is to avoid the time and cost constraints that come with conventional survey approaches while attempting to extract meaningful stories from large datasets. Examining Twitter data provides a broad overview of general trends, emotional responses, and mental processes, providing a deeper and more complex understanding. At the heart of this methodology is the effort to examine social media users' real-time

reactions, which promotes agility and quickness in the data collection process and makes the information-gathering process more dynamic and effective.

Furthermore, by addressing the always changing and expanding data pool on social media sites, this method enables more intelligent monitoring and understanding of the changing emotional and cognitive fabric of society. As a result, these data collecting and analysis procedures take on a more flexible and modern color, providing decision-makers with more accurate and fast information.

In order to obtain long-lasting insights, this study has been conducted inside a comprehensive framework that integrates multiple software tools and approaches. The business intelligence architecture and technologies were used to carefully plan and carry out the automated process's methodical steps. Efforts were undertaken throughout this continuum to reduce error rates and data loss, which in turn guaranteed the accuracy and dependability of the analyzed data and strengthened the credibility of the generated results. In addition, the integration of integrated software tools and procedures has expanded the scope of the research, providing a deeper knowledge by careful examination. This enhancement improves the ability to accomplish long-term goals and make more effective decisions.

The need for dependable database systems in data backup and protection procedures emphasizes how crucial it is to guarantee the security of the collected data. By maintaining the highest standards throughout the data analysis process, this robust technique strives to facilitate the achievement of long-term perspectives and encourage confidence in the obtained insights.

This study presents cleaned and converted data in an insightful visual format using visually striking reports that are easily incorporated into the business intelligence architecture and natural language processing techniques.

The value of these data visualizations is enhanced by utilizing business intelligence architecture, which expedites the analysis process and allows the gained insights to be evaluated from a broader perspective. As a result, the research findings are presented in a way that makes them easier to understand, which increases their effectiveness in guiding decision-making processes.





## 1. GİRİŞ

Dijital çağın hızla büyüyen dünyası, sadece dildeki gelişmeleri değil, aynı zamanda değerli ve geniş veri akışını da içerir. Bu veri akışı günümüzde durdurulamaz bir şekilde büyüyerek önemli bir oluşum halini almıştır. Hepimiz, bu oluşumun bir parçası ve yaratıcısı olarak dijital ortamda etkin bir rol oynamaktayız. Sosyal medya ise insanların kendilerine ait paylaşımları yaptıkları dijital evrenin en belirgin ve değerli unsurlarından biri olarak kabul edilmektedir.

Sosyal medya, toplumu saran bir ağ gibi etkisiyle, son yıllarda gücü ve etkisi sıkça tartışılan bir dijital bilişim ortamıdır. Bu etkili ortam, içinde yaptığımız paylaşımlar sonucunda bizimle ilgili her türlü bilgiyi içerebilir. İnsanlar çeşitli sosyal medya araçlarını farklı amaçlarla ve kullanım şekilleriyle kullanırken, bu çalışma özellikle insanların duygularını, düşüncelerini ve hislerini paylaştığı Twitter platformuna odaklanmaktadır.

Twitter, mikroblog odaklı anlık ileti veya durum paylaşım platformu olarak halkın sesine en kolay, hızlı ve amaca yönelik bir şekilde ulaşma imkanı sunar. Twitter'da paylaşılan metinlere "Tweet" denir ve içerikleri metin, lokasyon, zaman ve başlık etiketleri gibi bilgiler içerebilir. Bu çalışmanın amacı, Twitter üzerinden elde edilmiş olan metinsel verileri amaca uygun bir şekilde iş zekası terminolojisi ile analiz ederek bilgiye dönüştürmek, bu bilgiyi değerlendirmek ve sonuç olarak bir karara ulaşmak üzerinedir.

Gerekli veri setini elde etmek için Twitter API kullanılarak, Sakaya il merkezi bölgesinde atılan tweet'ler 3 aylık bir süre için günlük olarak izlenmiş ve çeşitli iş zekası yöntemleri ile analizler yapılmıştır. İş zekası yöntemleriyle yapılan analizler, veri madenciliği ve keşif süreçlerini içermiş ve bu verilerin anlamını çıkarmak için özel algoritmalar kullanılmıştır. Metin sınıflandırma, tahminleme ve duygu analizi gibi doğal dil işleme yöntemleriyle analizler oluşturulmuş ve tahminsel verilerin anlamlı görsellere dönüşümü sağlanmıştır. Bu çalışmanın amacı, yapılan analiz sonuçlarıyla birlikte elde edilen verilerin, lokasyon ve etkin olduğu çevre için iş zekası perspektifiyle toplumsal bir katkı ve geri bildirim sağlamasıdır.



## 2. LİTERATÜR ARAŞTIRMASI

Veri, iş zekası ve doğal dil işleme konuları, bir araya geldiğinde değerli ve verimli projelerin ve araştırmaların kilit noktası olmuştur. Özellikle bu alanların ve bileşenlerinin bir araya getirilmesi, bilgi çıkarma süreçlerini zenginleştirip daha detaylı analiz imkanı sunarak organizasyonlar için önemli avantajlar sağlamaktadır.

Maynard ve ark. (2007) iş zekası bağlamında bilgi entegrasyonu için doğal dil teknolojisinin tasarımını ve uygulamasını ele almışlardır Temel hedefleri ise, finansal risk yönetimi ve uluslararasılaşma gibi konularda belirli bir şirketin finansal durumunu analiz etmek üzere web dahil bir dizi kaynaktan ilgili bilgileri çıkarmaktır. Bu bilgilerin tanımlanması ve çıkarılması, doğal dil işleme teknolojisi temelinde gerçekleştirilmiştir. Çalışma, iş raporları ve şirket web siteleri gibi çeşitli metin türlerine odaklanarak bilgi çıkarma süreçlerini detaylı bir şekilde ele almıştır.

Bahja (2020) ise yaptığı detaylı çalışmasında doğal dil işleme çalışmalarının farklı sektörlerde ve farklı işletmelerde nasıl bir öneme sahip olduğuna değinmiştir. Bahja, Finans sektöründe verimli çalışan makine öğrenmesi ve doğal dil işleme algoritmalarıyla finansal analistlerin ve iş analistlerinin yatırımcılarıyla doğru zamanda doğru kararlar alabilmesi için yaygın olarak kullanıldığını belirtmiştir. Sağlık sektöründe ise, veri kalitesindeki düşüklük ve yapısal veriye daha az derecede sahip olan bir sektör olmasına rağmen doğal dil işleme yöntemlerini kanser vakalarını tanıma, psikolojik hastalıkları öngörme ve doktor tükenmişliğini azaltarak karar verme süreçlerini iyileştirme gibi çalışmalar olduğundan bahsedilmiştir. Çalışmasında eğitim sektörüne de değinen Bahja, doğal dil işleme çalışmalarının eğitim ve öğretimde de epey önemli rol oynayabileceğinden bahsederek, öğretimde kullanılan kaynakların makine çevirimi ile evrensel bir hale dönüşebileceğini ve internet üzerinden öğrenimin de hızlanabileceğinin altını çizmiştir.

Bal ve ark. (2013) ise insan kaynakları platformunda paylaşılan iş ilanları ve öz geçmiş bilgileriyle ilgili değerli bir çalışma yapmışlardır. Bu çalışma detayında paylaşılan iş ilanları ve özgeçmiş paylaşımlarının doğal dil işleme yöntemlerinden geçirilerek iş ilanlarına uygun adaylar bulunması amaçlanmıştır. Bu kapsamda,

semantik web ve kural tabanlı bilgi çıkarım yöntemleri kullanılmış olup 30.000 özgeçmiş incelenmiş içerisinde bulunan 343,20 kelimeyle birlikte yapılan bilgi çıkarım işlemi %71 doğrulukla sonuçlandırılmıştır.

Çalışmasında hizmet sektöründe bulunan otel ve çevrimiçi yorumlarını inceleyen Ay (2021) makine öğrenmesi ve doğal dil işleme yöntemlerini kullanarak manipülasyona sebep olan otel yorumlarını tespit etmeyi amaçlamıştır. Bu çalışma kapsamında DBSCAN algoritmasını kullanmış, 2.623 otele ait 287.463 adet veriyle yazı kümeleme işlemleri yapmıştır.

Dragoni ve ark. (2018) yaptığı çalışmada yasal metinlerden otomatik kural ve bilgi çıkarımı için doğal dil işlemeyi kullanmışlardır. Sözdizimi tabanlı kural çıkarımı aşamasında Stanford Ayırıştırıcısı ve CCG(Combinatory Categorical Grammar) Ayırıştırıcısı ile cümlelerin sözdizimsel yapısını ve deontik bileşenlerini analiz eden ekip bu iki tekniği karşılaştırmış ve çıkarımda bulunan kavramlara anlam kazandırmak ve tanımlayabilmek için de Wordnet'ten faydalanmışlardır.

Alt dalga değerlendirmesinde, altın standartta bulunan 65 terimden 59 terimin çıkarımını yapıp 49'unda doğru sonuca ulaşılmıştır. Değerlendirme sonucu olarak %90,78 kesinlik oranı, %83,05 duyarlılık oranı ve %86,74 F1-Skor oranına ulaşmışlardır. Deontik işaretlemeyle doğru tanımlanan 49 terimin 47'si uygun olarak işaretlenmiş ve %95'92'lik bir hassasiyet elde edilmiştir.

Üst dalga değerlendirmesinde ise ilk adım olarak, ilk dalga değerlendirmesinden çıkan sonuçlar ve CCG Ayırıştırıcısı tarafından oluşturulan kuralların uyumu incelenmiştir. CCG Ayırıştırıcısı ile çıkarılan terimlerde 51 ilişki bulunmuş olup bunların 35'inin altın standartta karşılığı olduğu görülmüştür. Bu inceleme sonucu olarak CCG ayırıştırıcısının 2 yeni kural daha bulup kesinlik oranını %97.22'ye çıkardığı ancak ilişki sayısının yüksek çıkmasından dolayı da hassasiyet oranı %68.63'e düşmüştür. Bu çıkarımlar sonucunda ekip, CCG ayırıştırıcısının Stanford Ayırıştırıcısı'na göre daha detaylı ve mantıksal bir seviyede çalıştığı sonucuna ulaşmıştır. İki ayırıştırıcının de birlikte kullanıldığı senaryoda, sistem hassasiyeti %66,67'den %80,56'ya çıkmış olup iyileştirme sağlandığı gözlemlenmiştir.

Doğal dil işleme ve makine öğrenimi yöntemleriyle benzer bir çalışma yapan Nay (2018) çalışmasında hukuk metinlerini anlamak ve tahmin etme algoritmalarıyla çıkarımlarda bulunuyor. Hukuk metinlerini anayasal, yasal, dava, idari ve sözleşmeli

olarak beş ana başlıkta toplayan Nay, J. çeşitli DDİ modelleri ve Makine Öğrenimi yöntemleriyle içerik özetleme, içerik çıkarma, metinle ilişkili sonuçları tahmin etme ve soruları cevaplandırma gibi analizlerde bulunuyor.

Doğal Dil İşleme (NLP), çeşitli endüstri ve sektörlerde önemli bir etkiye sahiptir. Dilin karmaşıklığını ve çeşitliliğini anlama yeteneği sayesinde, NLP, sağlık, finans, eğitim, perakende ve daha birçok farklı alanlarda kullanılmaktadır. Sağlık sektöründe, hastane kayıtları ve medikal metinler üzerinde NLP, hastalık teşhisi ve bu hastalıklara yönelik tedavi planlaması gibi kritik görevlerde kullanılabilir. Finans sektöründe, finansal raporlar ve haberler üzerinde NLP, piyasa trendlerini analiz edebilme ve yatırım stratejileri geliştirme konusunda değerli bir role sahiptir. Eğitimde, öğrenci performansı değerlendirmesi ve eğitim materyallerinin kişiselleştirilmesi gibi alanlarda NLP'nin kullanımı artmaktadır.

Yüksel ve Tan (2018) herhangi bir lokasyonda bulunan kafeler, restoranlar, ve alışveriş merkezleri gibi mekanların bulunduğu bu mekanların yorumlandığı ve puanlandığı bir uygulama olan Foursquare sosyal paylaşım uygulamasında bulunan yorumlar üzerinde çalışmışlardır. Kendi geliştirdikleri Sosyal Bilgi Keşfi Algoritması (SBKA) ve Bayes gibi diğer algoritmaları karşılaştıran Yüksel,A. ve Tan, F. ikili sınıflandırmada %81,97 başarı oranı elde ederken üçlü sınıflandırmada ise %84,49 başarı oranı elde etmiştir.

Can ve ark. (2012), genelde yerel seçimlerde yönelim etkisi olduğunu düşündükleri sosyal medya platformlarından biri olan Twitter'da yapılan paylaşımlar üzerinden 2012 ABD seçimindeki başkanlık adayları hakkında yapılan paylaşımları gerçek zamanlı doğal dil işlemede önemli yeri olan duygu analiziyle incelemiş ve elde ettikleri sonuçları veri görselleştirme yaparak takip etmişlerdir.

Bu çalışmayla birlikte insan kaynağıyla sağlanabilen anket yada röportaj gibi geri bildirimlerle alınan analizler yerine daha hızlı ve amaca yönelik bir analiz aracı geliştirip izlenimlerini gerçekleştirmişlerdir.

Küçükkartal (2020) metin madenciliğinin en çok kullanılan başlıklarından olan duygu analizi üzerinde bir çalışma gerçekleştirmiştir. Çalışmasında metin madenciliği ön işleme adımlarıyla İngilizce tweetleri incelemiş, TF-IDF (Terim Sıklığı – Ters Doküman Sıklığı) yöntemi ile ağırlıklandırma işlemi yapmıştır.

Elde ettiđi sonuçları veri görselleştirme araçlarıyla görselleştirerek soyut olan verileri anlamlandırarak somut hale getirmiştir.

Çalışma sonucunda vaccine(aşı) kelimesini aratarak oluşturduđu 10.000 ingilizce tweet içerisindeki metinlerin %42'sinin pozitif yönde, %14'ünün negatif yönde ve %43'ünün ise nötr tweet paylaşımı olduđu sonucuna ulaşmıştır.

Karabulut ve Küçüksille (2018) Twitter platformunda çalışan profesyonel bir izleme analiz aracı geliştirmiş olup doğal dil işlemenin çeşitli yöntemleriyle bu araçlarını geliştirmişlerdir. Çalışma kapsamında sadece paylaşılan metinler üzerinde değil, paylaşımları yapan kullanıcılar üzerinde ve paylaşım içeriğinde bulunan etiketler üzerinde de detaylı çalışan bir analiz aracı geliştirmişlerdir.

Kaur (2016) yaptığı çalışmasında iş zekasının internet bankacılığı üzerindeki önemini ve bankacılık sektöründeki rolünü vurgulamıştır. İş zekası, bankaların müşteri ilişkileri yönetimi ve diğer müşteri odaklı süreçlere bağlanarak müşterilerini daha iyi anlamalarını ve daha karlı müşterileri elde etmelerini sağlayan bir araç olarak kabul ediliyor. İş zekası, bankaların tarihsel analiz, performans bütçelemesi, iş performansı analitiđi, çalışan performansı ölçümü, yönetici panelleri, pazarlama ve satış otomasyonu, ürün inovasyonu, müşteri karlılığı, düzenleyici uyum ve risk yönetimi gibi alanlarda kullanabilecekleri teknolojileri içeriyor.

Bankacılık sektöründe iş zekasının, bankacılık müşterilerine zamanında ve doğru bilgi sunmanın değerli bir varlık olduđu, iş zekası platformunun iyi uygulanmasının sektör liderleri arasında farklılık yarattığı, güvenilir ve tek bir veri kaynağı oluşturmanın işbirliği gerektirdiđi ve iş zekası çözümlerinin bankalara yeni ve değerli fırsatlar sunduđunun altını çizmiştir.

Romero ve ark. (2021) yaptıkları projede iş zekası temelli bir ticari web uygulaması tasarlayarak uygulamaya sahip olan kuruluşların müşterilerinin sosyal ve finansal davranışları hakkında analiz yapmak ve daha fazla kar edebilmelerini sağlamak amacıyla bu uygulamayı geliştirmişlerdir. Bu uygulamayla birlikte kuruluşlara karar verme süreçlerinde daha geniş perspektifler sunmak ve iç yatırım maliyetlerini azaltmak gibi stratejik kararların da olumlu yönde etkileneceđini öngörmüşlerdir.

Çalışmasında iş zekası kapsamında incelenen ve önemli bir konu olan Pazar Sepeti Analizine yer veren Gündüz (2015) birliktelik kuralı tekniklerini işleyerek müşterinin satın alma eğilimlerini analiz edip satın alınan ürün ile birlikte alınması muhtemel olan

ürünlerin tavsiye edilme durumunu analiz etmeyi hedeflemiştir. Microsoft'un birliktelik kuralında bulunan Apriori algoritması ve genel olarak büyük veri kümelerinde kullanılan Fp – Growth algoritmalarını kullanarak değerli bir Pazar Sepeti Analizi çalışması yapmıştır.

Halim ve ark. (2019) bir restoran için pazarlama stratejisi tasarlamak amacıyla iş zekası yöntemleriyle Pazar sepeti analizi yapmayı hedeflemiştir.

Yapılan promosyonlarına beklediği dönüşü alamayan restoran Pazar Sepeti Analizi ile tüketicilerinin ve müşterilerinin alışkanlıklarını ürün, yer, fiyat ve promosyon kavramlarıyla uyumlu bir şekilde anlayıp bir pazarlama stratejisi tasarlamayı hedeflemektedir.

Yapılan çalışma sonucunda tüketim saatleri ve tüketici grupları kümelenmiş ilişkisiz gözükmesine rağmen promosyon olarak kullanıldığında verimli olacak menüler saptanmıştır.

Çetin ve Tanrıöver (2020) personel memnuniyetini ölçebilmek adına iş zekası ve veri madenciliği projesi geliştirerek personellerin birimlerindeki memnuniyet noktalarını keşfederek, Apriori ve K-Means algoritmalarıyla işledikleri verilerin anlamlı bir hale dönüşmesini sağlamışlardır. Verinin anlam yolculuğunu tamamlamak üzere Microsoft Power BI aracını kullanmış ve veri görselleştirme adımıyla projelerini tamamlamışlardır.

Türk (2020) otomobil yan sanayi sektöründe numune üretimi yapan bir firmanın süreçlerini iş zekası uygulamaları yoluyla iyileştirmeyi amaçlamıştır. Çalışmanın amacı, şirketin sipariş ve süreç takibini, raporlamayı, analiz etmeyi ve filtrelemeyi en aza indirmek ve bu süreçlerin verimliliğini artırmaktır. Projenin tamamlanmasının ardından takip ettiği süreçlerde iyileştirmelerin olumlu sonuçlarını gören Türk, özellikle firmanın günlük toplantıları için harcanan sürelerde azalma görmüştür. Bununla birlikte, iş birimleri için makine kullanım, verimlilik oranı, müşteri şikayet sayısı, aylık planlı gerçekleşen teslimat, aylık planlı gerçekleşen teslimat süresi ve Power BI ile görsellenen raporlarda %10 ila %15 oranında iyileştirme sağladı.

Bu literatür taramasında bulunan farklı çalışmalar ve konularla birlikte veri, doğal dil işleme ve iş zekası alanlarında yapılan projelerin, organizasyonların stratejik hedeflerine ulaşmasında nasıl değerli bir rol oynadığının altı çizilmektedir





### **3. TEORİK ÇERÇEVE**

#### **3.1. Doğal Dil İşleme**

Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlama ve işleme yeteneği sağlayan disiplinlerarası bir alanı temsil eder. NLP'nin iki önemli bileşeni, Doğal Dil Anlama (NLU) ve Doğal Dil Üretimi (NLG), dilin farklı yönlerine odaklanarak bilgisayar sistemlerine dil yetenekleri kazandırmayı amaçlar. NLU, metin veya konuşma verilerini anlamlandırma konusunda odaklanır. Bu, dilbilgisel analiz, kelime öbeklerinin anlamını çıkarma, duygu analizi gibi süreçleri içerir. Örneğin, bir NLU sistemi, bir metin belgesindeki anahtar bilgileri çıkarabilir veya bir konuşma kaydındaki duygusal tonu anlayabilir.

Diğer taraftan, NLG, bilgisayarların anladığı bilgileri insan dilinde anlaşılır bir şekilde ifade etme sürecini temsil eder. Yaratıcı bir dil kullanarak metin üretme yeteneği, bu alandaki temel hedeflerden biridir. Örneğin, bir NLG uygulaması, sayısal verileri açıklayıcı bir metin haline getirebilir veya veri tablolarını anlaşılır cümlelere dönüştürebilir (Sing ve Mahmood, 2021).

NLP, iş dünyasında müşteri geri bildirimleri analizi, sosyal medya duygu analizi, otomatik dil çevirisi gibi birçok alanda kullanılmaktadır. NLU ve NLG, bilgisayarların dil ile etkileşimde bulunma kabiliyetini genişleterek, dilin anlamını çıkarma ve üretme yeteneklerini daha derinlemesine kullanılabilir kılmaktadır. Bu sayede, NLP, gün geçtikçe gelişen ve dil teknolojilerindeki ilerlemeleri şekillendiren kapsamlı bir araştırma alanı haline gelmiştir (Abro ve ark, 2023).

##### **3.1.1. Metin madenciliği ve ön işleme aşamaları**

Veri madenciliği, istatistiksel analiz, veritabanı yönetimi ve makine öğrenimi gibi farklı alanlardan gelen bilgi birikimini bir araya getirerek, önceden bilinmeyen ve açıkça ortaya çıkmamış verileri çıkarmak için kullanılan bir veri analiz yöntemidir (Arslantekin, 2003).

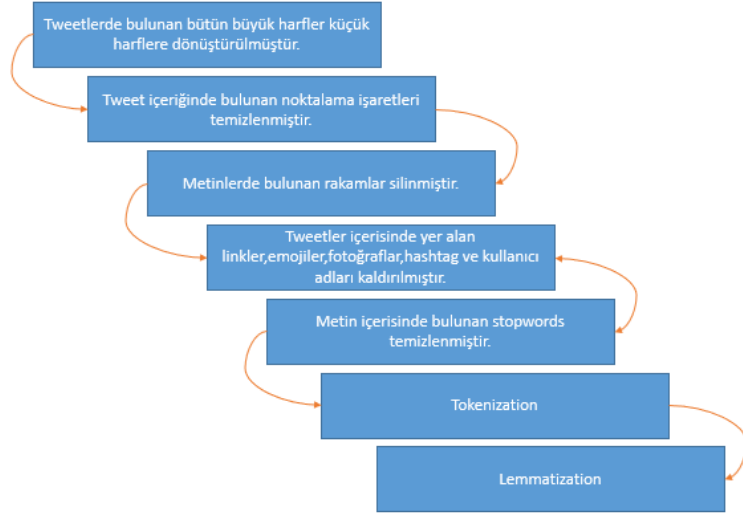
Veri madenciliği, devasa veri yığınları arasında gizli örnek kalıpları ortaya çıkarma, eğilimleri belirleme ve gerekli ilişkileri kurma süreçlerini içermektedir.

Veri madenciliği, veritabanı teknolojileri, istatistik, yapay zeka, örüntü tanıma, yüksek başarılı hesaplama, veri görselleştirme, bilgi erişimi, görüntü işleme ve zamana bağlı veri analizi gibi çeşitli alanlardan elde edilen yöntemleri entegre eden bir uygulama alanını temsil etmektedir (Han ve Kamber, 2006).

Veri madenciliği alanındaki araştırmalar genellikle yapısal verilere, özellikle de ilişkisel veritabanları ve veri ambarlarındaki verilere odaklanmaktadır. Ancak, gerçekte ulaşılabilir ve kullanılabilir verinin önemli bir kısmı, metin veritabanları veya diğer adıyla doküman veritabanlarından gelmektedir. Bu tür veritabanları genellikle e-posta iletileri, araştırma bildirimleri, sayısal kütüphaneler, haberler, kitaplar, makaleler ve internet sayfaları gibi farklı kaynaklardan oluşan geniş bir doküman koleksiyonunu içermektedir. Çevrimiçi ortamdaki bilgi miktarındaki hızlı artış, beraberinde metin veritabanlarının boyutlarının da hızla artmasına ve ilgili teknolojilerinin gelişmesine neden olmaktadır. Tahminlere göre, iş dünyasına ait bilgilerin %85'i metin formatında depolanmaktadır (Hotho ve ark, 2005).

Metin Madenciliği, metinlerin çeşitli ön işleme adımlarından geçirilip, çeşitli yöntemlerin kullanılarak ara sonuçların analiz edilmesini ve sonuçların anlamlı bir hale getirilmek üzere görselleştirilmesini içeren adımlardan oluşmaktadır (Pazarçeviren ve ekibi, 2015). Ek olarak, duygu analizi, metin özetleme, döküman kümeleme, doküman sınıflandırma, metindeki konunun tespiti, intihal tespiti ve spam analizi gibi pek çok farklı alanda kullanılabilir. Son yıllarda daha önemli hale gelmiş olan Duygu Analizi (DA) veya diğer bir adıyla Görüş Madenciliği (GM) uygulamaları da bu alan içerisinde değerlendirilmektedir. GM, insanların bir konuyla ilgili görüşleri, düşünceleri ve duyguları hakkında çıktılar sunabilen bir çalışma alanıdır (Kumar ve Vadlamani, 2015).

Elde edilen tweetler ile metin madenciliği işlemleri yapılırken en önemli kısımlardan birisi metinleri sözdizimsel olarak temiz ve düzenli hale getirmektir. Bu işlemi yaparken bazı aşamaların sırasıyla takip edilmesi gerekmektedir. Bu aşamalar Şekil 3.1'de gösterilmiştir.



**Şekil 3.1.** Metin Ön İşleme Adımları

DDİ alanında, stopwords ya da Türkçe adıyla "durak kelimeleri," genellikle dil analizinde dikkate alınmayan yaygın kelimeleri ifade eder. Bu kelimeler, dilin yapısal yapısını desteklemede temel bir rol oynar, ancak bir metnin anlamını çıkarmak veya özetlemek açısından sınırlı değer taşırlar. Örnek olarak, "ve," "ama," "veya," gibi yaygın bağlaçlar, "bu," "şu," gibi işaretleyiciler ve "oldu," "gitti," gibi yaygın fiiller stopwords olarak kabul edilir.

Durak kelimelerinin çıkarılması, NLP uygulamalarında analiz sürecini kolaylaştıran önemli bir yere sahiptir. Bu kelimelerin filtrelenmesi, metindeki gerçek anlam taşıyan ve öne çıkan kelimeleri belirlemeyi kolaylaştırır.

Durak kelimelerini ayıklamak, kelime dağarcığını daraltır ve modelin daha odaklı, anlamlı sonuçlar üretmesine katkıda bulunur ( Sarica ve Jianxi, 2020).

Tokenizasyon, doğal dil işleme süreçlerinde metni anlamlandırmak için kullanılan temel bir adımdır. Bu aşamada, bir cümle veya belge, kelimelerine veya alt kelimelerine ayrılarak daha küçük ve işlenebilir parçalara dönüştürülür.

Örneğin, "Büyük bir macera yaşadık" cümlesi tokenizasyon sonucunda "Büyük," "bir," "macera," "yaşadık" şeklinde ayrı tokenlara bölünür. Bu, metni analiz etmek, kelimeler arasındaki ilişkileri anlamak ve dilbilgisi yapısını anlamak için temel bir adım oluşturur (Nguyen ve ark, 2021).

Lemmatizasyon ise kelimelerin köklerini elde etme sürecidir. Bu işlem, bir kelimenin sözlükteki temel veya kök halini bulma amacını taşır. Örneğin, "koşuyordum" kelimesi

lemmatizasyon sonucunda "koşmak" kelimesine dönüştürülür. Lemmatizasyon, benzer anlamlı kelimelerin aynı temel forma indirgenmesini sağlayarak metin analizlerinde tutarlılık ve anlam zenginliği sağlar (Uddin ve ark, 2022).

Tokenizasyon ve lemmatizasyon, dil işleme uygulamalarında metni daha anlamlı ve işlevsel bir formata getirme amacını taşırlar. Bu işlemler, dilin karmaşıklığını azaltarak, metni daha etkili bir şekilde analiz etme ve anlam çıkarma süreçlerine katkıda bulunurlar (Kharis ve ark, 2022). Yapılmış olan metin ön işleme adımlarında Python dilinde bulunan, NLTK(Natural Language Toolkit) ve Re (Regular Expression) python kütüphaneleri kullanılmıştır.

### **3.1.2. Kategori tahmini ve duygu analizi**

Kategori tahmini, metin doğal dil işleme metotlarından biri olan metin sınıflandırma metoduyla metnin sınıflandırılıp bu sınıfların algoritmalarla tahmin edilmesi sağlanmıştır.

Duygu analizi, aynı zamanda fikir madenciliği olarak adlandırılır. Metinlerde gizli olan duyguları, fikirleri ve düşünceleri belirlemenin en yaygın yolu, bu metinleri olumlu, olumsuz veya nötr olacak şekilde sınıflandırmaktır (Aravi, 2014).

### **3.2. İş Zekası**

İş Zekası (İZ), organizasyonlara kılavuzluk eden ve değerli içgörüler sunan bir disiplindir, özellikle de veri analizi ve raporlama alanında uzmanlaşmıştır. İZ, büyük miktarda veri setini anlamlı bilgilere dönüştürme amacı güder ve stratejik karar alma süreçlerini destekler. Organizasyonların faaliyetleri sırasında biriktirdikleri veri kaynakları, İZ araçları tarafından incelenir ve çeşitli analitik tekniklerle anlamlandırılır. İZ, belirlenmiş hedeflere ulaşmak, performansı değerlendirmek, trendleri belirlemek ve geleceğe dönük stratejiler oluşturmak gibi amaçlar doğrultusunda değerlidir.

Bu süreçte, İZ araçları, veri madenciliği, raporlama, görselleştirme ve keşif analizi gibi çeşitli teknikleri birleştirerek organizasyonlara bilgi sunar. Veri tabanlı çıktılarla değerli kararlar almak, rakipleri karşısında öne geçmek ve iş süreçlerini optimize etmek, İş Zekası'nın temel amaçları arasında yer alır. Ayrıca, İZ, işletmelerin müşteri davranışlarını analiz etmesi ve anlaması, pazarlama stratejilerini oluşturması ve

finansal performanslarını takip etme konularında da kritik bir rol oynar (Ahmad ve ark. 2020).

İZ'nin önemi gün geçtikçe artmaktadır, çünkü bilgiye dayalı karar alma, dinamik iş dünyasında rekabet avantajı sağlamaktadır. İZ, veri odaklı kültürleri destekler ve işletmelerin karmaşık veri setlerini anlayarak stratejik avantaj elde etmelerine zemin hazırlar. Bu nedenle, İş Zekası, dijital çözümleri içinde barındıran iş dünyasında değerli bir alan haline gelmiş ve organizasyonların sürdürülebilir başarılarını destekleyen önemli bir disiplin haline gelmiştir (Choi ve ark, 2022).

Hannula, Pirttimäki (2003) ve Olexov (2014) tarafından yürütülen geniş kapsamlı bir endüstri anketi, İş Zekası (İZ) benimsemenin algılanan faydalarını doğrulamıştır. İZ benimsemenin en önemli faydaları şunlardır:

- Karar verme için kaliteli bilgi elde etme (95%)
- Beklenen fırsatları ve tehditleri analiz etme yeteneğinin artması (83%)
- Kurumsal/şirket bilgisinin artması (76%)
- Bilgi paylaşımının artması (73 %)
- Daha iyi analiz ve bilgi çekme yeteneği (58%)
- Daha hızlı karar alma (54%)
- Daha doğru ve daha hızlı raporlama (82%)
- Karar alma kalitesinin artması (78%)
- Müşteri hizmetlerinin iyileştirilmesi (58%)

### **3.2.1. Veri ambarı mimarisi**

Veri ambarları, organizasyonların geniş veri setlerini etkili bir şekilde depolaması, yönetebilmesi ve ihtiyaç anında hızlıca erişilebilmesi için tasarlanmış bilişim sistemleridir. Temel amacı, farklı kaynaklardan gelen verileri birleştirip anlamlı bilgiler çıkarmak olan iş zekası ve analitik uygulamalarına hizmet etmektir. İşletme depolarında saklanan ve işlenen verilerin hacmi ve çeşitliliği hızla artmakta ve analitik uygulamalar için veri hazırlama ve verimli erişimde yeni zorluklar ortaya çıkarmaktadır (Rao ve ark, 2019).

Veri ambarının ana bileşenleri, veri entegrasyonu, depolama, yönetim ve sorgulama süreçlerini içerir.

**Veri Entegrasyonu:** Bu bileşen, farklı kaynaklardan gelen verileri toplama ve birleştirme sürecini yönetir. Bu aşamada, çeşitli veri formatları ve yapıları standart bir formata dönüştürülerek homojen bir veri seti oluşturulur.(El Wessimy ve ark, 2013)

**Depolama:** Veri ambarının belki de en kritik bileşeni depolama altyapısıdır. Veri depolama, genellikle çok boyutlu (multidimensional) bir veri modeli kullanır ve bu model, verileri etkili bir şekilde organize etmeyi sağlar. Veri depolama çözümleri, performansı artırmak ve verilere hızlı erişim sağlamak için optimize edilmiş yapılar içerir (Taktak ve ark, 2017).

**Yönetim:** Veri ambarı yönetimi, verilerin güvenliği, bütünlüğü ve erişilebilirliği için gerekli olan süreçleri kapsar. Veri ambarı yönetimi, veri ambarının oluşturulması, bakımı, yedeklenmesi ve güncellenmesi gibi operasyonel görevleri içerir.

**Sorgulama:** Bu bileşen, kullanıcıların veri ambarındaki bilgilere erişmelerini sağlar. SQL sorguları veya iş zekası araçları aracılığıyla, kullanıcılar veri ambarındaki bilgilerle sorgulamalar, analizler ve ihtiyaçlarına yönelik raporlar oluşturabilirler.

Bu bileşenlerin birlikte çalışır hale gelmesiyle, bir veri ambarının etkili bir şekilde çalışması sağlanır ve organizasyonlara büyük veri setlerini anlamak ve stratejik kararlar almada yardımcı olacak güçlü içgörüler sunar.

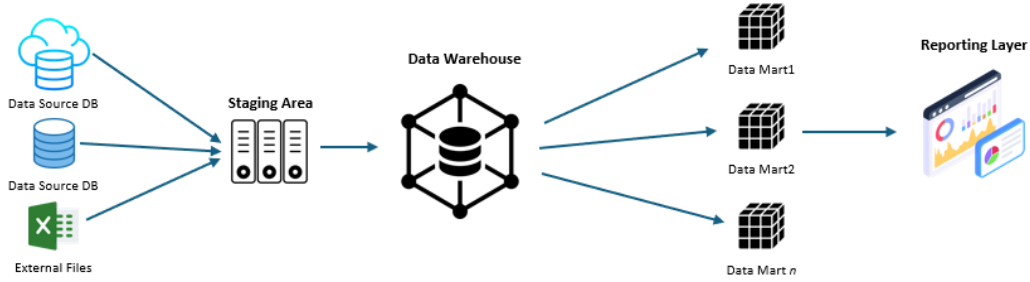
### **3.2.2. Veri ambarı mimari yaklaşımları**

Veri depolama konusunda iki öncü yaklaşım bulunmaktadır; en önemli yaklaşımlar, boyutsal yaklaşım ve normalleştirilmiş yaklaşımdır. Boyutsal yaklaşım, Ralph Kimball'ın önerdiği veri modeli yaklaşımıdır. Bu yaklaşıma göre, veri ambarı, Boyutsal Model/Yıldız şeması kullanılarak modellenmelidir.

Normalleştirilmiş yaklaşım, aynı zamanda 3NF modeli (Üçüncü Normal Form) olarak adlandırılan Bill Inmon'un veri modeli yaklaşımıdır. Bu yaklaşıma göre, veri ambarı, E-R (varlıksal ilişki) modeli/normalleştirilmiş model kullanılarak modellenmelidir. Boyutsal yaklaşımda, işlem verileri genellikle sayısal işlem verileri olan "fact"e yani işlem tablolarına ve fact tablolarına bağlam sağlayan referans bilgileri olan "dimension"a yani boyutsal tablolara bölünür.(Shobirin ve ark, 2017)

Inmon'un modeli, bir organizasyonun bilgi sistemlerini ve veritabanlarını kapsayan muazzam bir yapı önerir; bu devasa yapıya "Corporate Information Factory" veya CIF adı verilir (Inmon ve Imhoff, 2002).

Inmon, organizasyonun genel veritabanı ortamını dört benzersiz seviyeye böler: Operasyonel, Atomik Veri Ambarı, Departman ve Bireysel. Veri ambarını oluşturan son üç seviye, genel veritabanını içerirken, ilk seviye, eski ve diğer işlem sistemlerinden gelen verileri içerir ve bu seviye, organizasyonun günlük işleyişini destekler. Şekil 3.2’de Kimball veri ambarı mimarisi gösterilmiştir.

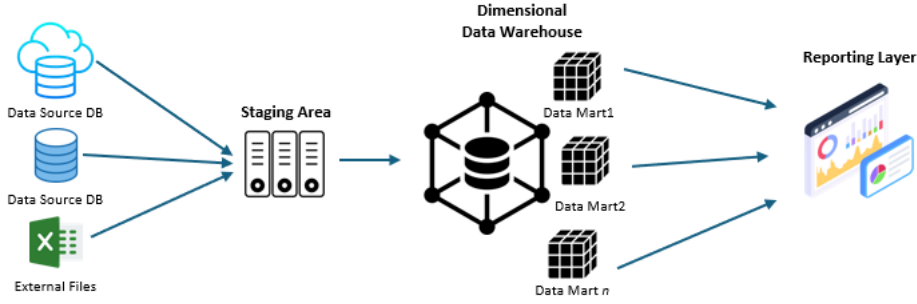


**Şekil 3.2.** Kimball Veri Ambarı Mimarisi, Kimball (2002)’den uyarlanmıştır.

Inmon, atomik veri ambarının inşa edilmesinin başlangıç çabasının, çeşitli bölümler arasında uyumsuz veri riskini ortadan kaldırdığı için değerli olduğunu savunur. Bu, organizasyonun veri düzenini koruyarak çeşitli departman veritabanlarının oluşturulmasına olanak tanır (Inmon, 2002).

Kimball'un modeli, geleneksel ilişkisel veritabanı yaklaşımından birkaç önemli açıdan farklılık gösterir. Bu farklılıklardan biri, Kimball modeli ile oluşturulan veri depolarının, veri ambarına özgü bir veri modelleme yöntemi kullanmasıdır. Bu yöntem, "Boyutsal Veri Modelleme" olarak tanımlanmaktadır (Kimball, 2002).

Başka bir önemli fark, genel mimarinin yüksek derecede etkileşimli olması beklenen birden çok veritabanını içermesidir. Boyutsal modelleme, işlem(fact) ve boyut(dimension) tablolarıyla başlar. İşlem tabloları metrikleri içerirken, boyut tabloları işlem tablolarındaki metriklerin özniteliklerini içerir. Boyut tabloları sıklıkla tekrarlayan gruplar içerir; bu, normleştirme kurallarına aykırıdır. Ancak boyutsal modelleme, veri ambarındaki yüksek performansı ve kullanıcı dostuluğunu sağlamak için normleştirme kurallarını ihlal eder. Son kullanıcılar için modellenmiş bir veri tabanına sorgu yapmak oldukça kolaydır çünkü neredeyse tüm verilerin özetlenmiş yolları zaten boyut tablolarında bulunmaktadır. Bu, kullanım kolaylığı hedefini büyük ölçüde karşılamaktadır. Boyutsal modelleme, veri ambarının benzersiz gereksinimlerinden faydalanarak bir veri modelleme yaklaşımıdır. Şekil 3.3’te Inmon veri Ambarı Mimarisi sunulmuştur.



**Şekil 3.3.** Inmon Veri Ambarı Mimarisi, Inmon (2002)'den uyarlanmıştır.

İşlem tabloları küçük bir satır sayısına ve boyut tablolarını yüksek derecede denormalleştirilmiş şekillerde tutmak, elde edilen veri ambarını son kullanıcı için oldukça erişilebilir hale getirebilir ve makul sorgu yanıt süreleri sağlayabilir (Breslin, 2004).

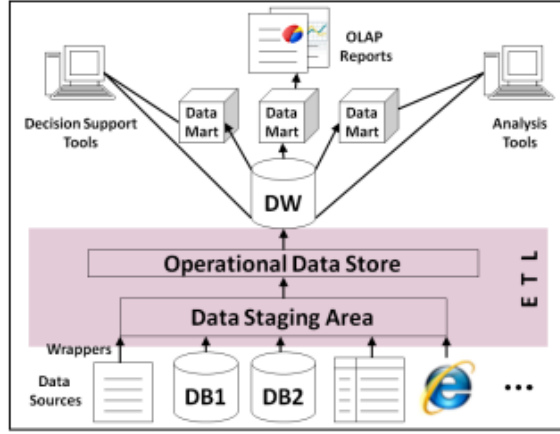
### 3.2.3. Veri ambarı mimari katmanları

Veri ambarlarında mimari önem taşıyan iki kritik katman staging (aşama) ve ODS (Operasyonel Veri Kaynağı) katmanlarıdır.

Staging katmanı, veri ambarına gelen ham verilerin işlenip hazırlanmadan önce geçtiği ilk aşamadır. Burada, kaynak sistemlerden alınan veriler temizlenir, düzenlenir ve standart formata getirilir. Bu aşama, veri ambarına gelen verilerin kalitesini artırmak ve tutarlılığını sağlamak için önemlidir. Ayrıca, çeşitli kaynaklardan gelen verilerin birleştirilmesini kolaylaştırarak, homojen bir veri seti elde etmeye yardımcı olur. Staging katmanı, veri ambarının veri entegrasyonu sürecinde önemli bir ara aşamadır.

ODS katmanı ise, işlenmiş ve hazırlanmış verilerin geçtiği bir diğer önemli aşamadır. Operasyonel Veri Kaynakları, genellikle iş süreçlerini desteklemek amacıyla kullanılan bir katmandır. Burada, güncel ve işlenmiş veriler saklanarak, operasyonel raporlama, sorgulama ve iş zekası ihtiyaçlarına hızlı erişim sağlanır (MadeSukarsa ve ark, 2012). ODS, kullanıcıların gerçek zamanlı veya neredeyse gerçek zamanlı verilere erişimini kolaylaştırarak, iş birimlerinin alacağı kararların hızlanması ve verimin artırılmasına olanak tanır. Şekil 3.4'te bahsedilen mimari görsel olarak gösterilmiştir.





**Şekil 3.4.** Veri Ambarı Mimari Katmanları, ElGamal (2016)'dan uyarlanmıştır.

Staging ve ODS katmanları arasındaki geçişler, veri ambarının mimarisini güçlendiren ve optimize eden önemli noktalardır. Bu geçişler, veri ambarının performansını artırır, kullanıcıların anlık ihtiyaçlarına daha performanslı yanıt vermesini sağlar ve veri bütünlüğünü korur. İki katman arasındaki bu akıcı geçiş, veri ambarının etkili bir şekilde çalışmasını ve organizasyonların veri odaklı kararlar almasını destekler (Yang ve ark, 2019).

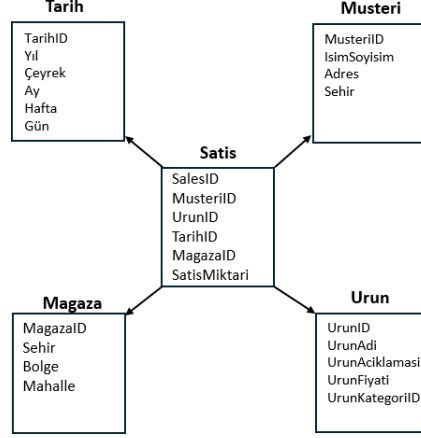
### 3.2.4. Veri ambarı mantıksal model yaklaşımları

Veri ambarı modelleri, büyük miktarda veriyi depolama, yönetme ve analiz etme amacıyla kullanılan karmaşık yapılar içerir. Yıldız şeması ve kartanesi şeması, bu modellerin iki temel tasarım yaklaşımını temsil eder.

#### 3.2.4.1. Yıldız şeması mantıksal modeli

Veri ambarları, operasyonel veritabanlarından alınan ilgili bilgileri kayıtlar halinde toplar ve gelecekteki sorgu işleme, analiz ve değerlendirme için depolar. Bu kayıtlar, bir işlem tablosu ve birden çok boyut tablosu içerir. İşlem tablosu, her boyutlu tablodan gelen anahtarları içererek bir yıldız şeması oluşturur. Bu şema, işlem tablosundaki ilgili kayıtları birleştirerek farklı boyutsal bilgileri destekler. Yıldız şemasında, merkezi bir "İşlem" veya ana veri tablosu, çeşitli boyut tablolarla ilişkilendirilmiştir. Bu boyut tabloları, genellikle veri analizi için önemli olan kategorik bilgileri temsil eder. Yıldız şeması, sorguların performanslı ve verimli bir şekilde sonuç döndürebilmesini sağlar.

Şekil 3.5'te Yıldız şeması mantıksal modelinin ilişkisel şekli gösterilmiştir.



**Şekil 3.5.** Yıldız Şeması Mantıksal Modeli

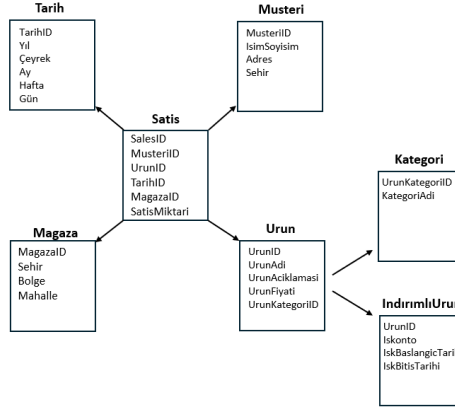
Yıldız şemasında, boyut tablolarındaki tekrarlanan verilerin neden olduğu normalleştirilmemiş tablolara ilgili eksiklikler görülebilir. Bu sorunu çözmek için normalleştirme, tabloyu iki veya daha fazla tabloya ayırarak ve tekrarlanan değerleri yeni tablolardan birine taşıyarak gerçekleştirilebilir.

Örneğin, ürün kategorisi tablosundaki normalleştirme, elektronik ürünleri için birden fazla bölge müdürü veya bölgeyi desteklemesi, ürünün farklı kategori ve departmanlara sahip olması veya takvimin yıl, ay ve hafta gruplandırmasına göre kategorize edilmesi durumunda daha faydalı bir kullanıma sahip olacaktır (Seyed-Abbasi ve Madesi,2015).

### 3.2.4.2. Kartanesi şeması

Yıldız şemasının avantajları olan sadeleştirme ve sorgu optimizasyonunu korurken, yüksek normalleştirilmiş bir yapı olan snowflake şeması, tekrarlanan verileri azaltır ancak daha fazla boyutlu tabloya ve karmaşık sorgu birleştirmelerine yol açacaktır.

Kartanesi şeması, daha karmaşık veri yapılarını temsil eder. Birden çok faktör arasındaki ilişkileri gösteren bu model, geniş veri setlerini daha detaylı anlaşılır bir şekilde analiz etme imkanı sunar. Kartanesi şeması, farklı boyut tabloları arasındaki ilişkileri anlamak ve özellikle büyük ölçekli veri ambarlarındaki karmaşıklığı ele almak için kullanılır (Seyed-Abbasi ve Pusey, 2007). Şekil 3.6'da Kartanesi şeması mantıksal modelinin tablolar arası ilişkisel görünümü gösterilmektedir.



**Şekil 3.6.** Kartanesi Şeması Mantıksal Modeli

Bu modeller, bir organizasyonun iş ihtiyaçlarına ve analitik hedeflerine uygun olarak seçilir ve uyarlanır.

Yıldız şeması ve kartanesi şeması, veri ambarlarının tasarım sürecinde esneklik ve optimize edilmiş performans sağlayarak, veri analizi ve raporlama süreçlerinde daha etkin ve verimli bir yaklaşım sunar. Bu sayede organizasyonlar, topladıkları verilerden daha derinlemesine anlam çıkarabilir ve stratejik kararlar alabilirler.

### 3.2.5. Ayıklama, dönüştürme ve yükleme (ETL)

Günümüz işletmelerinde, çeşitli kaynaklardan elde edilen verilerin veri depolarına aktarılması sıklıkla karşılaşılan bir durumdur. Ancak, bu verilerin çeşitli formatlarda girilmiş olması, sınıflandırma sürecini karmaşık hale getirebilmektedir. Bu sebeple, veri deposunun iş zekâsı unsurlarının merkezi olan bu verileri yüklerken, öncesinde temizlenmeleri, birleştirilmeleri ve sınıflandırılmaları gerekmektedir. Özellikle OLTP(Online Transactional Processing) yapısındaki çok sayıda kullanıcının farklı formatlarda ve kendilerine ait yetkilerle sisteme girdiği veriler, çeşitli yazılımlar aracılığıyla ETL adımlarından tek tek geçirilerek veri deposuna aktarılması gerekmektedir. ETL çalışmalarının temel amacı gereksiz bilgileri elemek ve gerekli bilgileri ise kaliteli bir şekilde ortak kısıtlar altında depolamaktır (Pazarçeviren ve ark, 2015).

Veri kaynağı, dönüştürme ve yükleme araçları, gerekli tüm dönüşümleri, özetlemeleri ve veriyi veri ambarında birleşik bir formata dönüştürmek için kullanılır. Ayrıca Extract, Transform ve Load (ETL) Araçları olarak da bilinen bu araçların işlevleri şunları içerir:

- İşletme veritabanlarından veri ambarına yüklenmesi gerekmeyen gereksiz verilerin ortadan kaldırılması.
- Çeşitli kaynaklardan gelen veriler için ortak isimleri ve tanımlamaları arama ve değiştirme.
- Herhangi eksik veri durumunda, bunları varsayılan değerle doldurma ve özetler ile türetilmiş verileri hesaplama.
- Birden fazla veri kaynağından gelen tekrarlanan verilerin teke indirgenmesi (Hosmane ve Rakshitha, 2021).

### **3.2.6. Veri görselleştirme**

Görsel öğeler, bilgiyi hedef kitleye sunmanın çok etkili bir yöntemidir. Genel kavramlar, süreçler ve birikimler, şekil, grafik, animasyon gibi görsel araçlarla daha iyi aktarılabilir. Bu araçlar, karmaşık verinin kolayca anlaşılmasına katkı sağlar. Veri görselleştirme, genellikle diyagramlar, çizelgeler ve grafikler gibi araçlarla gerçekleşir. Bu grafik öğeler, sayısal değerleri anlamak ve iletmek için verimli bir ortam oluşturur. Bilgilendirme tasarımı, özellikle veriyi net bir şekilde izleyiciye sunma amacını taşır ve bu anlayış, bilginin görselleştirilmesiyle bütünleşir.

Bilginin nasıl iletilmesi gerektiği konusu sadece tasarımcıların değil, aynı zamanda farklı alanlarda çalışan araştırmacıların ve bilim adamlarının da ilgi odağıdır. Veri içeriği, bozulmadan, anlaşılır bir biçimde dışarıya aktarılmalıdır. Verilerin açık ve çarpıcı bir şekilde sunulabilmesi için bilgilendirme tasarımı yöntemlerinin ustalıkla kullanılması gerekmektedir (Gürler ve ark, 2018).

Görselleştirme, istatistiksel ve karmaşık verileri grafiksel arayüzlerle sunarak, insanların daha kolay anlamasını sağlar. İnsanlar genel olarak 3 boyutlu görselleri algılayabilir, ancak daha fazla boyutlu verileri algılamakta zorlanırlar. Görselleştirme, çok boyutlu verileri 2 veya 3 boyuta indirerek, insanların daha rahat kavramasına ve anlamasına imkan verir. (Çelik ve Akdamar, 2018).

## **4. MATERYAL VE YÖNTEM**

### **4.1. Projede Kullanılan Araçlar ve Yöntemler**

Projede kullanılan araçlar aşağıda belirtilmiştir.

1. Twitter API
2. Python Uygulaması
3. Spyder IDE
4. Azure Data Factory
5. Azure SQL Server
6. Power BI

#### **4.1.1. Twitter API**

Twitter API, twitter platformu için sosyal medya analizi, içerik takibi, duygu analizi, trendleri belirleme gibi bir dizi uygulama için güçlü bir veri kaynağı sunar. Bu API sayesinde, milyonlarca tweet arasında arama yapabilir, belirli kelimeleri içeren tweet'leri filtreleyebilir ve belirli coğrafi konumlardan gelen tweet'ler incelenebilir. Gerçekleştirilen uygulamada ise belirli bir coğrafi konumdan gelen tweetler incelenmiştir.

Buna ek olarak, örneğin, iş zekası üzerine bir çalışma yapılıyorsa, Twitter API sayesinde belirli anahtar kelimelerle ilgili tweet'leri analiz edebilir, trendleri takip edebilir ve kullanıcıların duygu durumlarını anlamak için doğal dil işleme teknikleri uygulanabilir.

Ayrıca, Twitter API'nin kullanımını sayesinde farklı sektörlerin yapmış olduğu sosyal medya kampanyalarının etkileşimini ölçebilir, hedef kitlelerinin ne hakkında konuştuğunu anlayabilir ve marka bilinirliğinin artırmak için uygun stratejiler geliştirilebilir. API ayrıca, belirli bir konu veya hashtag etrafında gelişen olayları anlık olarak takip edilebilmesini sağlar.

Twitter API'nin sektörel ve endüstri bazında sağladığı faydalar oldukça geniş kapsamlıdır. Bu sektörler ve faydalar aşağıdaki gibi örneklendirilebilir.

#### **4.1.1.1. Pazarlama ve marka yönetimi**

Pazarlama iş birimi çalışanları, Twitter API'yi kullanarak marka bilinirliğini artırabilir, kampanya etkileşimlerini ölçebilir ve hedef kitleleriyle etkileşimde bulunan içerikleri analiz edebilir. Trend analizi yaparak, sektörel gelişmeleri takip edebilir ve marka stratejilerini güncelleyebilirler (Mattila ve Salman, 2018).

#### **4.1.1.2. Müşteri hizmetleri**

Şirketler, Twitter API'yi müşteri hizmetleri için kullanarak, müşteri geri bildirimlerini anında takip edebilir, olumlu veya olumsuz tepkileri hızlı bir şekilde yanıtlayabilir ve müşteri memnuniyetini artırmak için gerekli önlemleri alabilir.(Shakeel ve ark, 2020)

#### **4.1.1.3. Finans ve yatırım**

Finansal analistler, Twitter API aracılığıyla finansal haberleri ve duyuruları takip edebilir. Finansal piyasalardaki hareketliliği anlamak için tweet analizleri yapabilir ve sosyal medya üzerinden gelen duyarlılık verilerini kullanarak yatırım stratejilerini şekillendirebilirler (Padmanayana ve ark, 2021).

#### **4.1.1.4. Sağlık sektörü**

Sağlık profesyonelleri, Twitter API'yi kullanarak hastalık salgınlarını izleyebilir, halk sağlığı trendlerini analiz edebilir ve toplumda sağlıkla ilgili konularda farkındalık yaratmak için sosyal medya kampanyalarını değerlendirebilir.(Jordan ve ark, 2018)

#### **4.1.1.5. Eğitim**

Eğitim kurumları, Twitter API'yi öğrenci geri bildirimlerini izlemek, eğitim trendlerini takip etmek ve sosyal medya üzerinden öğrenci etkileşimlerini analiz etmek için kullanabilir. Bu, eğitim programlarını iyileştirmek ve öğrencilere daha etkili bir deneyim sunmak için önemli bilgiler sağlayabilir (Grosseck ve Holotescu, 2008).

#### **4.1.2. Python programlama dili**

Python programlama dili, geniş yelpazede kullanım alanına sahiptir. Python'un tercih edilme sebebi, basit ve okunabilir syntax'ı, geniş kütüphane desteği, topluluk katılımı ve çeşitli endüstrilerde kullanılabilirliği sayesinde sağlanıyor. Python'un gücü, zengin kütüphane ve modül ekosistemi ile daha da artar. Veri bilimi, yapay zeka, web geliştirme, otomasyon ve daha birçok alanda kullanılan birçok özel kütüphane bulunmaktadır.

Örneğin, NumPy ve Pandas veri bilimi, TensorFlow ve PyTorch yapay zeka, Django ve Flask web geliştirme alanlarında sıkça kullanılan kütüphanelere örnek olarak verilebilir.

Gerçekleştirilen proje kapsamında ise, veri temizleme, veri işleme, veri analizi, makine öğrenmesi ve duygu analizi için kullanılacak olan programlama dilidir.

#### **4.1.2.1. Bu proje için kullanılan python kütüphaneleri;**

Tweepy, Twitter API ile etkileşimde bulunmayı sağlayan önemli bir kütüphanedir. Twitter API ile Python programlama dili arasında köprü görevi görmektedir. Tweepy, tweetlerin elde edilmesi, tweet paylaşımı ve paylaşım yapan kişilerin bilgilerinin alınması gibi birçok farklı alanda işlemler gerçekleştirebilmektedir.

Pandas, veri bilimi ve analizi alanında kullanışlı bir açık kaynaklı kütüphanedir. Pandas, veri çerçeveleri (dataframes) ile çalışmayı hedeflemiştir ve veri işleme, düzenleme, analiz ve modelleme gibi pek çok veri bilimi işlemini basitleştirir.

Numpy, bilimsel hesaplamalar ve çok boyutlu diziler üzerinde etkili işlemler gerçekleştirmek için kullanılan temel bir kütüphanedir. NumPy, özellikle lineer cebir operasyonları ve matematiksel fonksiyonlar gibi birçok sayısal operasyonu destekler.

Re, Python programlama dilinin en çok kullanılan kütüphanelerinden biri olan "Regular Expressions" için kısaltılmış bir isimdir. Re kütüphanesi, metin içindeki aramalar, eşleştirmeler ve değişiklikler için kullanılır.

Düzenli ifadeler, metin işleme işlemlerinde kullanılan güçlü bir araçtır ve belirli bir desene uyan metinleri bulma veya değiştirme işlemlerini tanımlamak için kullanılır. "re" kütüphanesi, bu tür işlemleri gerçekleştirmek için bir dizi fonksiyon ve metot içerir.

NLTK (Natural Language Toolkit), Python programlama dilinde doğal dil işleme yöntemleri için kullanılan bir kütüphanedir. NLTK, dil işleme, metin madenciliği, belge sınıflandırma, çeviri ve benzeri birçok doğal dil işleme görevini kolaylaştıran birçok kaynak içerir.

Transformers, doğal dil işleme (NLP) alanında öne çıkan büyük ölçekli dil modeli ağlarını ve transfer öğrenme modellerini içeren bir kütüphanedir. Hugging Face tarafından geliştirilen bu kütüphane, BERT, GPT-2, RoBERTa, T5, DistilBERT ve birçok diğer popüler dil modellerini içerir.

### **4.1.3. Spyder IDE**

Spyder, Python programlama dili için özel olarak tasarlanmış açık kaynaklı bir entegre geliştirme ortamıdır (IDE).

Özellikle veri bilimi ve bilimsel hesaplama projeleri için geliştirilen Spyder, IPython entegrasyonu ile interaktif bir çalışma ortamı sunmaktadır. NumPy, SciPy, Pandas gibi veri bilimi kütüphaneleriyle uyumludur ve veri analizi, matematiksel modelleme gibi işlemleri kolaylaştırır. Gelişmiş kod düzenleme arayüzü, renkli vurgular ve otomatik tamamlama gibi özelliklerle yazım sırasında destek sağlar. Veri bilimi projeleri üzerinde çalışanlar için uygun ve etkili bir IDE olan Spyder, çalışmalarında Python temelli uygulamalar geliştirmek isteyen öğrenciler için de güçlü bir araç olarak öne çıkmaktadır.

### **4.1.4. Azure data factory**

Azure Data Factory, Microsoft'un bulut tabanlı veri entegrasyonu hizmetidir, farklı veri depolama sistemleri arasında sorunsuz veri hareketi ve dönüştürme sağlayarak büyük veri işleme ve analitik uygulamalar için bir çözüm aracı sunar. Kullanıcı dostu arayüzü, veri akışlarını görsel olarak tasarlama imkanı sunarken, güçlü işleme motoru veri dönüştürme ve temizleme işlemlerini kolaylaştırır. Otomatikleştirme yeteneği sayesinde zamanlanmış görevleri optimize eder ve büyük veri çözümleriyle uyumlu çalışabilir (Sverrisson ve Hákonarson, 2023).

Azure ekosistemiyle entegre olması, Azure Data Factory'yi diğer Azure hizmetleriyle sorunsuz bir şekilde çalıştırırken, güvenli veri entegrasyonunu da sağlar. Sonuç olarak, Azure Data Factory, veri entegrasyonunu modern ve etkili bir şekilde yönetmek isteyen organizasyonlar için kapsamlı bir çözüm sunar (Jadhav ve ark, 2021).

### **4.1.5. Azure SQL server**

Azure SQL Server, Microsoft'un Azure bulut platformunda barındırılan bir veritabanı hizmetidir. Kullanıcılar, geleneksel SQL Server'ın bulut versiyonunu kullanarak veritabanı yönetimi, yedekleme, güvenlik ve performans ayarlarını kontrol edebilirler. Esnek mimarisi, yüksek performanslı işlemci ve bellek kaynaklarına erişim, otomatik indeksleme gibi özellikleri ile farklı uygulama ihtiyaçlarına uyum sağlar.

Güvenlik odaklı tasarımı, veri şifreleme ve sektörel uyumluluk standartları ile hassas verilerin güvenli depolanmasını sağlar. Azure SQL Server, geniş ekosistemle



entegrasyon, bulut avantajları ve ölçeklendirme imkanları sunarak kullanıcılara etkili bir veritabanı çözümü sunar (Viswanathan ve ark, 2017).

#### **4.1.6. Power BI**

Power BI, Microsoft'un iş zekası ve veri analizi alanında geliştirdiği bir uygulamadır. Power BI, işletmelerin verilerini görsel olarak sunmalarına, yorumlamalarına ve dağıtmalarına imkan veren güçlü bir araçtır. Kullanıcı dostu arayüzü ve çok sayıda entegrasyon seçeneği ile dikkat çeker. Veri kaynaklarından veri çekme, veri modelleri kurma ve farklı grafiksel raporlar oluşturma yeteneği vardır. Power BI, çoklu veri kaynağını birleştirme, gerçek zamanlı güncellemeler ve paylaşımına açık dashboard'lar yaratma konusunda kullanıcılarına esneklik sağlar. (Khilari ve ark, 2022).

Ayrıca, görsel analizler, interaktif raporlar ve etkileşimli grafiklerle veri hikayelerini daha anlamlı hale getirme olanağı sunarak iş kararlarına daha hızlı ve bilgi odaklı yaklaşım sağlar. İş dünyasında veri odaklı kararlar almak isteyen şirketler ve profesyoneller için Power BI, veri analizi süreçlerini kolaylaştıran ve iş zekası çözümlerini güçlendiren etkili bir araçtır (Bhargava ve ark, 2018).

## **4.2. Metin Sınıflandırma ve Makine Öğrenimi Algoritmasının Karşılaştırılması**

Kategori tahmini, doğal dil işleme yöntemlerinden biri olan metin sınıflandırma yöntemidir ve metnin sınıflandırılmasını ve tahmin edilmesini sağlar. Metin sınıflandırma, önceden görülmemiş belgelere otomatik olarak önceden tanımlanmış etiketleri yerleştirme görevidir. Belgelerin dizinlenmesinde, e-posta filtrelemede, web taramasında ve kişisel bilgi ajanlarında kullanılan metin sınıflandırma, makine öğrenimi ve bilgi çıkarma alanlarının etkileştiği aktif ve önemli bir araştırma alanıdır (Scottve Matwin, 1999).

### **4.2.1. Rastgele orman sınıflandırıcı algoritması**

Rastgele Orman Sınıflandırıcı, veri sınıflandırma problemini çözmek için birden fazla karar ağacı kullanan bir algoritmadır. Bu algoritma, farklı veri alt kümelerinden karar ağaçları oluşturarak sınıflandırma performansını iyileştirir. Bu karar ağaçları, karar ormanı adı verilen bir yapıda birleştirilir. Bu algoritma, çeşitli veri türlerine uygulanabilir ve vektör makinelerine göre daha doğru sonuçlar sağlayabilir. Vektör makinelerinden daha kesin sonuçlar üretir. Büyük sayıda değişken ve sınıf etiket içeren kategorik bir dağılıma sahip veri setlerini, kayıp veri içeren veya dengesiz olan veri

setlerini sergileme konusunda iyi çalışır (Aydın, 2018). Şekil 4.1’de görüldüğü üzere, Random Forest makine öğrenimi algoritması eğitim modeli %74 doğruluk oranı başarı sonucuna ulaşmıştır.

	precision	recall	f1-score	support
siyaset	0.77	0.78	0.78	237
dunya	0.43	0.81	0.56	219
ekonomi	0.82	0.59	0.68	236
kultur	0.86	0.75	0.80	230
saglik	0.84	0.85	0.84	231
spor	0.99	0.75	0.85	234
teknoloji	0.85	0.66	0.74	230
accuracy			0.74	1617
macro avg	0.79	0.74	0.75	1617
weighted avg	0.80	0.74	0.75	1617

Şekil 4.1. Rastgele Orman Sınıflandırıcı Algoritma Sonuçları

#### 4.2.2. K - En Yakın Komşu Algoritması

K – En Yakın Komşu, sınıflandırılacak yeni bireyin önceki verilerine karşılık gelir. Bu, bireyler arasındaki k yakınlığını kontrol eden bir algoritmadır (Cover, 1967). Sınıflandırma sırasında test örnekleri, eğitim örnekleriyle karşılaştırılır. Mesafe için Yakınlık Öklidyen mesafesi kullanılır. Komşu örneklerin çoğunluk oylarına dayalı tahminler yapılır. Yüksek k dikkati, değerleri aşırı uyumlu hale getirme eğiliminde olduğu için önemlidir (Hastie ve ark, 2009).

Şekil 4.2.’de görüldüğü üzere, K - En Yakın Komşu algoritmasıyla modelin çalıştırılması sonucu olarak, eğitim modeli %88 doğruluk oranına ulaşılmıştır.

	precision	recall	f1-score	support
siyaset	0.81	0.86	0.84	237
dunya	0.77	0.84	0.80	219
ekonomi	0.83	0.82	0.83	236
kultur	0.92	0.90	0.91	230
saglik	0.93	0.90	0.92	231
spor	0.97	0.96	0.97	234
teknoloji	0.92	0.86	0.89	230
accuracy			0.88	1617
macro avg	0.88	0.88	0.88	1617
weighted avg	0.88	0.88	0.88	1617

Şekil 4.2. K - En Yakın Komşu Algoritma Sonuçları

### 4.2.3. Çok terimli naïve bayes algoritması

Çok terimli NB yöntemi, çok terimli Naive Bayesian sınıflandırıcısını oluşturmak için kullanılabilir. Bu yöntem, pratik, kolay ve verimli bir yöntemdir. Çok terimli Naive Bayesian, birden çok terimde dağılımını belirler. Belge, kelimelerin toplamı olarak görülür ve her kelimenin yeri diğerinden bağımsız olarak oluşmuş sayılır. Naive Bayes'in çok terimli versiyonu olan Multinomial NB, Rennie ve ark. (2003) tarafından öne sürülmüş, analiz edilmiş, geliştirilmiş ve yazılmıştır. Aşağıda bulunan Şekil 4.3'te görüldüğü üzere, Multinomial NB makine öğrenimi algoritması eğitim modeli %89 doğruluk oranına sahiptir.

	precision	recall	f1-score	support
siyaset	0.79	0.91	0.85	237
dunya	0.88	0.84	0.86	219
ekonomi	0.85	0.90	0.87	236
kultur	0.93	0.89	0.91	230
saglik	0.91	0.95	0.93	231
spor	1.00	0.93	0.96	234
teknoloji	0.92	0.82	0.87	230
accuracy			0.89	1617
macro avg	0.90	0.89	0.89	1617
weighted avg	0.90	0.89	0.89	1617

Şekil 4.3. Çok Terimli Naive Bayes Algoritma Sonuçları

### 4.2.4. Lojistik Regresyon Algoritması

LR kategorik bir modeldir ve bir veya daha fazla ön gösterge değişkenine bağlıdır. Bağımlı değişkenlerin sonuçlarını tahmin etmek için kullanılır. Daha önce, temel LR modeline ek olarak çeşitli modeller geliştirilmiştir. Multinomial LR modeli, lojistik regresyonu genelleştiren bir modeldir ve birden fazla farklı sonuç verir. Sonuç olarak, kategorik olarak dağılmış bağımlı değişkenin çeşitli olası sonuçlarını tahmin etmek için kullanılan modeldir. (Feng ve Timmermans, 2010). Şekil 4.4'te gösterildiği üzere, Lojistik Regresyon makine öğrenimi algoritması eğitim modelimiz için %90 doğruluk oranına sahiptir.

	precision	recall	f1-score	support
siyaset	0.88	0.89	0.89	237
dunya	0.82	0.89	0.86	219
ekonomi	0.88	0.88	0.88	236
kultur	0.92	0.94	0.93	230
saglik	0.93	0.92	0.93	231
spor	1.00	0.96	0.98	234
teknoloji	0.90	0.84	0.87	230
accuracy			0.90	1617
macro avg	0.90	0.90	0.90	1617
weighted avg	0.91	0.90	0.90	1617

Şekil 4.4. Lojistik Regresyon Algoritma Sonuçları

#### 4.2.5. Destek Vektör Makineleri Algoritması

Destek Vektör Makineleri (SVM), öğrenme, sınıflandırma, kümeleme, yoğunluk tahmini ve nihayetinde veri üzerinden regresyon kuralları oluşturmak amacıyla kullanılan bir eğitim algoritmasıdır. SVM, iki sınıf ve çoklu sınıf sınıflandırma problemlerini çözmek için uygundur. Veriyi sınıflandırırken SVM'nin amacı, en yakın örnekleri bulmak ve bu örnekler arasındaki ayırıcıyı yüzeye dik olarak maksimize etmektir. Ayırıcı yüzeyin veri setindeki performansı değişebilir ve çok sayıda seçenek vardır. SVM'nin sağladığı avantajlar sayesinde, ayırıcı yüzey her iki sınıfa da aynı uzaklıkta ve maksimum uzaklıkta olabilmektedir (Kazan ve Karakoca, 2019).

Şekil 4.5'te görüldüğü üzere, Destek Vektör Makineleri makine öğrenimi algoritması eğitim modeli %90 doğruluk oranına sahiptir.

	precision	recall	f1-score	support
siyaset	0.894	0.890	0.892	237
dunya	0.844	0.913	0.877	219
ekonomi	0.900	0.919	0.910	236
kultur	0.948	0.952	0.950	230
saglik	0.955	0.926	0.941	231
spor	0.996	0.974	0.985	234
teknoloji	0.918	0.874	0.895	230
accuracy			0.921	1617
macro avg	0.922	0.921	0.921	1617
weighted avg	0.923	0.921	0.922	1617

Şekil 4.5. Destek Vektör Makineleri Algoritma Sonuçları

#### 4.2.6. FastText algoritması

FastText, Facebook tarafından geliştirilmiş açık kaynaklı bir doğal dil işleme kütüphanesidir. Kelime seviyesinde gömülü temsiller oluştururken özellikle alt kelimeleri ve karakterleri dikkate alan değerli bir yapı sunar. Bu sayede, özellikle morfolojik detayları bulunan dillerde verimli çalışmaktadır. Metin sınıflandırma, duygu analizi gibi pek çok doğal dil işleme görevinde verimli bir şekilde kullanılmaktadır (Kuyumcu ve Aksakallı, 2019).

Şekil 4.6'da gösterildiği üzere, FastText algoritması eğitim modeli %87 doğruluk oranına sahip olduğu sonucuna ulaşmıştır.

	precision	recall	f1-score	support
siyaset	0.86	0.82	0.84	158
dunya	0.79	0.79	0.79	131
ekonomi	0.79	0.80	0.80	142
kultur	0.87	0.91	0.89	139
saglik	0.94	0.96	0.95	122
spor	0.96	0.93	0.94	147
teknoloji	0.86	0.85	0.86	141
accuracy			0.87	980
macro avg	0.87	0.87	0.87	980
weighted avg	0.87	0.87	0.87	980

Şekil 4.6. FastText Algoritma Sonuçları

#### 4.2.7. Türkçe BERT algoritması

BERT (Bidirectional Encoder Representations from Transformers), önceden eğitilerek hazır hale getirilen ve açık kaynak kodlu paylaşımına sahip bir DDİ modelini temsil etmektedir. BERT, arama sorgularındaki eksik kelimeleri bulmak için tüm cümleyi inceler dolayısıyla bu özellikle birlikte diğer algoritmalarından farklı bir konumda yer almaktadır. BERT, makine öğrenimi algoritmalarını kullanarak, kelimelerin birbiri arasındaki ilişkiyi analiz etmeye çalışır bundan dolayı da iki yönlü bir dil işleme özelliğine sahiptir. Geleneksel çift yönlü DDİ algoritmalarından farklı olarak, BERT, sağdan sola ve soldan sağa giden daha karmaşık bir maskeli dil modeli kullanılmaktadır (Acikalın ve ark, 2020).

Yapılan çalışmada, Duygu analizi (Yıldırım, 2022a) ve Kategori tahmini (Yıldırım, 2022b) için daha önce eğitilmiş olan transformers modelleri kullanılmıştır. Bu modellerin temelinde Google BERT algoritması bulunmaktadır.

Aşağıdaki Şekil 4.7.'de görüldüğü gibi, Türkçe BERT algoritması %93 doğruluk oranına ulaşmıştır. Bu, çalışma kapsamında yapılan makine öğrenimi algoritmaları karşılaştırma sürecinde en yüksek doğruluk oranını göstermektedir.

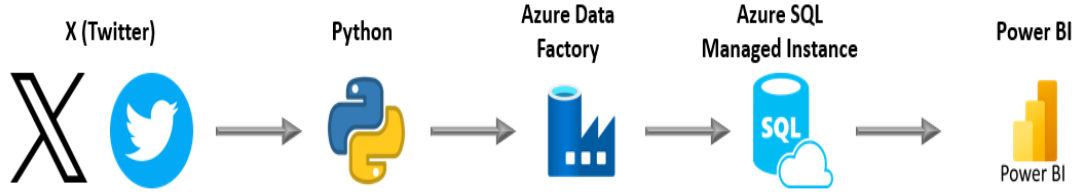
	precision	recall	f1-score	support
siyaset	0.923	0.911	0.917	158
dunya	0.908	0.908	0.908	131
ekonomi	0.919	0.880	0.899	142
kultur	0.949	0.942	0.946	139
saglik	0.950	0.943	0.947	122
spor	0.987	1.000	0.993	147
teknoloji	0.879	0.929	0.903	141
accuracy			0.931	980
macro avg	0.931	0.931	0.931	980
weighted avg	0.931	0.931	0.931	980

**Şekil 4.7** Türkçe BERT Algoritma Sonuçları

Yapılan makine öğrenimi algoritma karşılaştırma sonucuna göre en yüksek başarı oranı Türkçe BERT algoritmasında olduğu için yapılan çalışmada Türkçe BERT algoritması kullanılmasına karar verilmiştir.

## 5. SAKARYA’NIN DUYGU ATLASI PROJESİ

Bu çalışma belirli bir akış ve mimari çerçevesinde oluşturulmuştur. Şekil 5.1’de bulunan görselde aşamaların ilerleyişi ve proje yol haritası gösterilmiştir.



Şekil 5.1. Sakarya'nın Duygu Atlası Proje Mimarisi

### 5.1. Sakarya'nın Duygu Atlası Proje Mimarisi

#### 5.1.1. Twitter verilerinin elde edilmesi

Analiz edilecek veriler Twitter üzerinden günlük olarak elde edilmiştir. Twitter API kullanabilmek için önce Twitter Geliştirici hesabı daha sonrasında ise Twitter Geliştirici platformu üzerinden Twitter Uygulaması oluşturulup gerekli erişim izinleri sağlanmıştır.

Analiz edilecek lokasyon olarak Sakarya ili merkezi, 40.771945624249085, 30.388644933700565 koordinatı alınarak 25 kilometre çapında atılmış olan günlük ortalama 2000 tweet çekilmiş ve csv dosyasına kaydedilmiştir.

Elde edilen tweetler içerisinde, metin , metin uzunluğu , tweetin atıldığı tarih, lokasyon, like ve retweet sayıları , tweetin atıldığı kaynak gibi bilgileri içermektedir.

Sistem 7 günde 1 gün, 7 defa çalıştırılarak ilgili haftaya ait verilerin gün gün veri çekimi sağlanmıştır. Şekil 5.2’de Twitter üzerinden çekilen verilerin işlenmeden önceki hali gösterilmiştir.

metin,uzunluk,tarih,konum,begeni,retweet,kaynak  
 DEME KIZ AÄLARIM,25,2023-03-01 23:51:41,b'serdivan',1,0,Twitter for iPhone  
 Burnum iÄŒten kÄtrÄtkmÄtÄY canÄtmÄt acÄttÄtyor,49,2023-03-01 23:50:31,b'serdivan',1,0,Twitter for iPhone  
 Olum ben niye bu kadar safÄtm ya?,32,2023-03-01 23:50:28,b'Serdivan ',2,0,Twitter for Android  
 GÄnLÄK TÄrk Kahvesi ihtiyacÄtmÄtzÄt bu gÄnde attÄtÄtÄtnÄtz hikayelerden karÄYÄtladÄtk ÄŒok saÄY olun gY'S,92,2023-03-01 23:48:30,b'Sakarya / Kocaeli',0,0,Twitter for Andro  
 Az ÄŒnce evimin sokaÄYÄnda snap attÄt????????,43,2023-03-01 23:46:44,b'serdivan',3,0,Twitter for iPhone  
 TÄm yabancÄt dil derslerimden bu sene de kalÄtsam kalbimde kÄtrÄtlacak baÄYka bir yer kalmamÄtÄY olacak.,98,2023-03-01 23:45:07,b'',4,0,Twitter for Android  
 ErkeÄYin ""e"" si kÄÄŒÄK lÄtfen,40,2023-03-01 23:34:46,b'Adapazar',1,0,Twitter for Android"  
 MUÄLA ÄLKE OLMALI <https://t.co/5zfnd10tC1>,41,2023-03-01 23:33:32,b'Sakarya/Adapazar',1,0,Twitter for Android  
 Benim gÄnLÄm sarhoÄYtur yÄtldÄtzlarÄtn altÄndaa,43,2023-03-01 23:31:26,b'Adapazar',26,0,Twitter for iPhone  
 allah razi olsun bizi dusunmussun,46,2023-03-01 23:30:33,b'Serdivan',1,0,Twitter for iPhone  
 bu gece ÄŒok derin gibi uyusam iyi olurdu,40,2023-03-01 23:23:34,b'adapazar',1,0,Twitter for Android  
 olum biye hesap kapatip kapatip duruyorsun manitalara mi yakalaniyorsun noluyor,92,2023-03-01 23:22:17,b'Serdivan',1,0,Twitter for iPhone  
 Ne eksik ne fazla aynen bu ÄYekil <https://t.co/bw2Ju5YFM6>,56,2023-03-01 23:19:58,b'Sakarya',10,0,Twitter for iPhone  
 yÄtldÄtzlara bakÄtn lan,20,2023-03-01 23:15:08,b'Adapazar ',35,0,Twitter for Android  
 RT GÄdanÄtn sorumluluÄYü biz gÄda mÄhendislerinde olmalÄtdÄtr bunun iÄŒin kamuda GÄdaMuh AlÄtmÄtArtsÄtn,112,2023-03-01 23:15:07,b'',0,93,Twitter for iPhone  
 trileÄŒe.. tam ÄYü an <https://t.co/GX5hFPEM1x>,43,2023-03-01 22:56:28,b'54',7,0,Twitter for Android  
 Araba plakasÄtnÄt bioya yazmak nedirr,35,2023-03-01 23:08:12,b'Akyaz',13,0,Twitter for Android  
 ÄŒiÄY kÄŒfteyle ÄYarap iÄŒerken dinlemelik ÄYarkÄtler <https://t.co/2FCw6i3xFO>,70,2023-03-01 23:02:14,b'Serdivan',4,0,Twitter for iPhone  
 GÄzel mimarÄtm benim,19,2023-03-01 23:03:32,b'Sakaryaspor',6,0,Twitter for iPhone  
 Sana olan tavrÄtmÄtn mimarÄtsÄtn,28,2023-03-01 23:03:22,b'Sakaryaspor',27,0,Twitter for iPhone  
 iyi yolculuklar favÄt gelir mi sanmam <https://t.co/RWnEHh2HO>,60,2023-03-01 22:58:28,b'Serdivan',1,0,Twitter for iPhone  
 GÄrdÄnÄz dimi nasÄtl yazÄtklar olduÄYunuuuu,41,2023-03-01 22:57:55,b'Serdivan ',7,0,Twitter for Android  
 BeÄYkÄŒprÄ parkÄtnÄtn orda yaptÄtÄYÄtmÄtz 3 saatlik dedikodunun iÄŒinde sÄYlemiÄYtim,85,2023-03-01 22:56:58,b'',2,0,Twitter for iPhone  
 Adam bay yanlÄtÄY kalkmÄtÄY bizim hep doÄYru olmamÄtzÄt istiyor hahahaha evet ya ben hep seni beklicemm hepp,102,2023-03-01 22:56:35,b'Serdivan ',2,0,Twitter for Android  
 bana ve asilliÄYime iyi geceler diliyorum size iyi geceler yok ezikler sizi,74,2023-03-01 22:55:30,b'dnya',6,0,Twitter for iPhone  
 bu saatte canÄtm zeytinyaÄYlÄt yaprak sarmasÄt ÄŒekti gYm yemiÄY gibi hayal ederek uyursam rÄyamda gÄrÄrÄm belki gY'e,105,2023-03-01 22:14:04,b'dnya',8,0,Twitter for iPhone  
 Kaderimde varmÄtÄY de geÄŒ,23,2023-03-01 22:03:19,b'Adapazar ',14,0,Twitter for Android  
 KeÄYke ÄYü an iki ÄÄŒ sene sonrasÄtna ÄtÄYÄtnlansam ve erasmus iÄŒin yurt dÄtÄYÄnda olsam...,82,2023-03-01 21:59:02,b'Akyaz',10,0,Twitter for Android

## Œekil 5.2. İşlenmemiş Tweet Metinleri



### 5.1.2. Veri ön işleme

Genel olarak re,pandas ve NLTK kütüphaneleri kullanılmış olup, ön işleme aşamaları yalnızca metin sütunu için gerçekleştirilmiştir Şekil 5.3'te veri ön işleme aşamalarından geçirilmiş metin sütunu gösterilmiştir.

metni	uzunluk	tarih	konum	beğeni	retweet	kaynak
deme kız ağlarım	25	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
burnum içten kırılmış canımı acıyor	49	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
olum ben niye bu kadar safım ya	32	2023-03-01	b'Serdivan'	2	0	Twitter for Android;
günlük türk kahvesi ihtiyacımızı bu günde attığınız hikayelerden karşılığık çok sağ olun	92	2023-03-01	b'Sakarya / Kocaeli'	0	0	Twitter for Android;
az önce evimin sokağında snap attı	43	2023-03-01	b'Serdivan'	3	0	Twitter for iPhone;
tüm yabancı dil derslerimden bu sene de kalırsam kalbimde kırılacak başka bir yer kalmamış olacak	98	2023-03-01	b''	4	0	Twitter for Android;
erkeğin e si küçük lütfen	40	2023-03-01	b'Adapazar'	1	0	Twitter for Android'';
muğla ülke olmalı	41	2023-03-01	b'Sakarya/Adapazar'	1	0	Twitter for Android;
benim gönülüm sarhoştur yıldızların altındaa	43	2023-03-01	b'Adapazar'	26	0	Twitter for iPhone;
allah razı olsun bizi düşünmussen	46	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
bu gece çok derin gibi uyusam iyi olurdu	40	2023-03-01	b'adapazar'	1	0	Twitter for Android;
olum biye hesap kapatıp kapatıp duruyorsun manitalara mi yakalanıyorsun noluyor	92	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
ne eksik ne fazla aynen bu şekil	56	2023-03-01	b'Sakarya'	10	0	Twitter for iPhone;
yıldızlara bakın lan	20	2023-03-01	b'Adapazar'	35	0	Twitter for Android;
gıdanın sorumluluğu biz gıda mühendislerinde olmalıdır bunun için kamuda gıdamuh alımtarsın	112	2023-03-01	b''	0	93	Twitter for iPhone;
araba plakasını biyoya yazmak nedir	35	2023-03-01	b'Akyaz'	13	0	Twitter for Android;
güzel mimarım benim	19	2023-03-01	b'Sakaryaspor'	6	0	Twitter for iPhone;
sana olan tavırımın mimarımın	28	2023-03-01	b'Sakaryaspor'	27	0	Twitter for iPhone;
çiğ köfteyle sarap içerken dinlemelik şarkılar	70	2023-03-01	b'Serdivan'	4	0	Twitter for iPhone;
o kadar rahatsız insana maruz kaldım ki herkesin her hareketi normal geliyo	75	2023-03-01	b''	6	0	Twitter for iPhone;
yok bıraktım kotu oynadığım için	45	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
iyi yolculuklar favı gelir mi sanmam	60	2023-03-01	b'Serdivan'	1	0	Twitter for iPhone;
gördünüz dimi nasıl yazıklar olduğunuuuuu	41	2023-03-01	b'Serdivan'	7	0	Twitter for Android;
beşköprü parkının orda yaptığımız 3 saatlik dedikodunun içinde söylemiştim	85	2023-03-01	b''	2	0	Twitter for iPhone;
adam bay yanlış kalkmış bizim hep doğru olmamızı istiyor hahahaha evet ya ben hep seni beklicemm heepp	102	2023-03-01	b'Serdivan'	2	0	Twitter for Android;
trileçe tam şu an	43	2023-03-01	b'S4'	7	0	Twitter for Android;
bana ve asiliğime iyi geceler diliyorum size iyi geceler yok ezikler sizi	74	2023-03-01	b'dnya'	6	0	Twitter for iPhone;
bu saatte canım zeytinyağı yaprak sarması çektii yemiş gibi hayal ederek uyusam rüyamda görürüm belki	105	2023-03-01	b'dnya'	8	0	Twitter for iPhone;
kaderimde varmış de geç	23	2023-03-01	b'Adapazar'	14	0	Twitter for Android;
keşke şu an iki üç sene sonrasına işlansam ve erasmus için yurt dışında olsam	82	2023-03-01	b'Akyaz'	10	0	Twitter for Android;

Şekil 5.3. Ön İşleme Yapılmış Tweet Metinleri

### 5.1.3. Modelin çalıştırılması

Öğretimli Türkçe BERT Algoritması, seçilmiş olan modeliyle birlikte tweetleri analiz ederek hem kategori tahminlemesi hemde duygu analizi skorlaması üzerine çalıştırılmaktadır. Şekil 5.4'te işlenmiş veriler üzerinde ilgili model çalıştırıldıktan sonraki etiketlenmiş metinler gösterilmiştir.

Metin	length	date	location	likes	retweet	source;	Kategori	Duygu	Score	Duygu Kategori
yanıtı şimdi gördüm	35	2023-03-0	b'Adapazar'	0	0	0	Twitter fo Teknoloji	Olumlu	0,897470355	Çok Olumlu
silah mehmet	140	2023-03-0	b'Adapazar/Türkiye'	1	0	0	Twitter W Kültür	Olumsuz	0,546974659	Olumsuz
al işte asla yiyeceğim şimdi	42	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Sağlık	Olumlu	0,543362439	Olumsuz
deme kız ağlarım	25	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Teknoloji	Olumlu	0,917253733	Çok Olumlu
burnum için kırılmış canımı acıyor	49	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Sağlık	Olumsuz	0,992679477	Çok Olumlu
olum ben niye bu kadar safım ya	32	2023-03-0	b'Serdivan'	2	0	0	Twitter fo Teknoloji	Olumsuz	0,99430424	Çok Olumlu
günlük türk kahvesi ihtiyacımızı bu günde attığımız hikayelerden karşılığın çok sağ c	92	2023-03-0	b'Sakarya / Kocaeli'	0	0	0	Twitter fo Sağlık	Olumsuz	0,788718045	Olumlu
az önce evimin sokağında snap attı	43	2023-03-0	b'Serdivan'	3	0	0	Twitter fo Siyaset	Olumlu	0,841599464	Olumlu
tüm yabancı dil derslerimden bu sene de kalırsam kalbimde kırılacak başka bir yer	98	2023-03-0	b''	4	0	0	Twitter fo Teknoloji	Olumlu	0,948874414	Çok Olumlu
erkeğin e si küçük lütfen	40	2023-03-0	b'Adapazar'	1	0	0	Twitter fo Sağlık	Olumlu	0,819916844	Olumlu
muğla ülke olmalı	41	2023-03-0	b'Sakarya/Adapazar'	1	0	0	Twitter fo Siyaset	Olumlu	0,914462507	Çok Olumlu
benim gönlüm sarhoştur yıldızların altında	43	2023-03-0	b'Adapazar'	26	0	0	Twitter fo Kültür	Olumsuz	0,937598884	Çok Olumlu
allah razi olsun bizi düşünmüsun	46	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Teknoloji	Olumsuz	0,535199106	Olumsuz
bu gece çok derin gibi uyusam iyi olurdu	40	2023-03-0	b'adapazar'	1	0	0	Twitter fo Kültür	Olumsuz	0,524209678	Olumsuz
olum biye hesap kapatıp kapatıp duruyorsun mantalara mi yakalanıyorsun noluy	92	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Teknoloji	Olumsuz	0,933667958	Çok Olumlu
ne eksik ne fazla aynen bu şekil	56	2023-03-0	b'Sakarya'	10	0	0	Twitter fo Teknoloji	Olumlu	0,71230787	Nötr
yıldızlara bakın lan	20	2023-03-0	b'Adapazar'	35	0	0	Twitter fo Teknoloji	Olumlu	0,898051381	Çok Olumlu
gıdanın sorumluluğu biz gıda mühendislerinde olmalıdır bunun için kamuda gıdam	112	2023-03-0	b''	0	93	0	Twitter fo Ekonomi	Olumsuz	0,885879755	Çok Olumlu
araba plakasını bioya yazmak nedir	35	2023-03-0	b'Akyaz'	13	0	0	Twitter fo Teknoloji	Olumlu	0,849524915	Olumlu
güzel mimarım benim	19	2023-03-0	b'Sakaryaspor'	6	0	0	Twitter fo Kültür	Olumlu	0,916317463	Çok Olumlu
sana olan tavımın mimansın	28	2023-03-0	b'Sakaryaspor'	27	0	0	Twitter fo Kültür	Olumlu	0,986923754	Çok Olumlu
çığ köfteyle şarap içerken dinlemelik şarkılar	70	2023-03-0	b'Serdivan'	4	0	0	Twitter fo Kültür	Olumlu	0,987349033	Çok Olumlu
o kadar rahatsız insana maruz kaldım ki herkesin her hareketi normal geliyo	75	2023-03-0	b''	6	0	0	Twitter fo Sağlık	Olumsuz	0,997407854	Çok Olumlu
çok biraktim kotu oynadığım için	45	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Spor	Olumsuz	0,980891883	Çok Olumlu
iyi yolculuklar favi gelir mi sanmam	60	2023-03-0	b'Serdivan'	1	0	0	Twitter fo Teknoloji	Olumlu	0,925794959	Çok Olumlu
gördünüz dimi nasıl yazıklar olduğunuuuu	41	2023-03-0	b'Serdivan'	7	0	0	Twitter fo Siyaset	Olumlu	0,750526786	Olumlu
beşköprü parkının orda yaptığımız 3 saatlik dedikodunun içinde söylemiştim	85	2023-03-0	b''	2	0	0	Twitter fo Kültür	Olumsuz	0,963459969	Çok Olumlu
adam bay yanlış kalkmış bizim hep doğru olmamız istiyor hahahaha evet ya ben hi	102	2023-03-0	b'Serdivan'	2	0	0	Twitter fo Teknoloji	Olumlu	0,99487865	Çok Olumlu
trileçe tam şu an	43	2023-03-0	b'S4'	7	0	0	Twitter fo Kültür	Olumlu	0,742008388	Nötr
bana ve asiliğime iyi geceler diliyorum size iyi geceler yok ezikler sizi	74	2023-03-0	b'dnya'	6	0	0	Twitter fo Kültür	Olumsuz	0,992773354	Çok Olumlu
bu saatte canım zeytin yağlı yaprak sarması çektii yemiş gibi hayal ederek uyusam	105	2023-03-0	b'dnya'	8	0	0	Twitter fo Sağlık	Olumsuz	0,994242251	Çok Olumlu
kaderimde varmış de geç	23	2023-03-0	b'Adapazar'	14	0	0	Twitter fo Kültür	Olumlu	0,962096095	Çok Olumlu
keşke şu an iki üç sene sonrasına işlansam ve erasmus için yurt dışında olsam	82	2023-03-0	b'Akyaz'	10	0	0	Twitter fo Teknoloji	Olumsuz	0,695021987	Nötr

Şekil 5.4. Tahminleme Yapılmış Metinler

### 5.1.4. ETL İşlemi

Sınıflandırma ve gerekli analiz adımları tamamlanan tweetler yeni tahminlenmiş sütunlarıyla birlikte bir excel olarak kaydedilir. Veri Fabrikası(Azure Data Factory) ile verilerin çıkarım işlemi(extract), dönüştürme işlemi olarak hatalı yada bozuk olarak elde edilmiş ilçe ve kaynak bilgilerinin düzeltilmesi, Veri Ambarı model geçiş mimarisine uygun olarak, ilk adım olarak staging katmanına eklenen tablolar stored procedure(SP) ve view fonksiyonları kullanılarak ods katmanına taşınmıştır. Performansın en iyi derecede tutulması için ods katmanında dimensional(boyutsal) yani tanım bilgileri azaltılmış olup daha çok transactional(işlem) bilgilerine yer verilerek gerekli tablo ilişkileri kurulabilmesi adına tanımlayıcı anahtar sütunlar(ID Key) oluşturulmuştur(transform) ve yükleme işlemi olarak da Azure SQL veri tabanına aktarımı(load) gerçekleştirilmektedir. Veri aktarımı işlemi için 2 ana akış oluşturulmuştur ilk akış boyutsal tablo mimarilerini içinde bulundururken ikinci akış işlem tablo mimarisini içinde bulundurmaktadır.

Şekil 5.5'te tahminleme sonrası ilgili verilerin Azure SQL staging katmanına taşınmış örnek hali gösterilmektedir.

Metin	Uzunluk	Tarih	Lokasyon	BegeniSayisi	RetweetSayisi	Kaynak	KonuKategori	DuyguKategori	DuyguSkoru	DuyguKategori
mutlu olmak her şeye sahip olmak değil sahip olduğun kadımı her şey yapabilecektir	86	2023-03-15 00:00:00.000	Adapazan	2	0	iPhone için Twitter	Sağlık	Çok Olumlu	0.993817269802094	Çok Olumlu
sahibi burlan kim yaptı	47	2023-03-15 00:00:00.000	Serdivan	0	0	iPhone için Twitter	Siyaset	Çok Olumlu	0.952300906181335	Çok Olumlu
başlar başlamaz bir çıkmaza girince bırakmam 10 saniyemi almadi	75	2023-03-15 00:00:00.000	Akyaz	2	0	Android için Twitter	Teknoloji	Çok Olumlu	0.95661461353302	Çok Olumlu
başkılarımız an başkanı yardımcılarımız an an ve an 100 yaşını ağı	140	2023-03-15 00:00:00.000	Serdivan	0	20	iPhone için Twitter	Siyaset	Nötr	0.602455317974091	Olumsuz
allah rahmet eylesin mekanı cennet olsun kemal bey tüm annelerin anneler günü kull	140	2023-03-15 00:00:00.000	Serdivan	1	1	Android için Twitter	Sağlık	Olumlu	0.76342543577	Nötr
suan yan yana olmalıydık	33	2023-03-15 00:00:00.000	Adapazan	1	0	iPhone için Twitter	Kültür	Olumsuz	0.53070741891861	Çok Olumsuz
kolay gelsin	24	2023-03-15 00:00:00.000	Adapazan	1	0	Twitter Web Uyg.	Teknoloji	Çok Olumlu	0.57890234124	Çok Olumsuz
insanların yüzlerini göremiyorum	32	2023-03-15 00:00:00.000	Adapazan	7	0	iPhone için Twitter	Teknoloji	Çok Olumlu	0.80890234124	Nötr
kafamız artık hiç kimseyi kalmıyor	37	2023-03-15 00:00:00.000	Adapazan	24	0	iPhone için Twitter	Teknoloji	Çok Olumlu	0.814721645	Nötr
üzer ama yapma benice	33	2023-03-15 00:00:00.000	Adapazan	5	0	iPhone için Twitter	Teknoloji	Çok Olumlu	0.883451785688	Olumlu
erde kalmaya ve okumaya devam adapazamı çok seviyoruz çünkü takip et gönder	140	2023-03-15 00:00:00.000	Serdivan	0	102	iPhone için Twitter	Kültür	Nötr	0.655816363334656	Olumsuz
ağık 3 gekerli	31	2023-03-15 00:00:00.000	Adapazan	1	0	Twitter Web Uyg.	Sağlık	Çok Olumlu	0.82351617	Nötr
arkadaşlar bi dk suan evim konusuyo konus evim suurvivor2020	106	2023-03-15 00:00:00.000	Serdivan	0	2	iPhone için Twitter	Teknoloji	Çok Olumlu	0.959425449371338	Çok Olumlu
yazacaksan yaz artık hesi gtmiyor bendennnnnn	47	2023-03-15 00:00:00.000	Adapazan	6	0	Android için Twitter	Kültür	Çok Olumlu	0.8769571185112	Olumlu
rahmeti nenen oğlum algış kudu muştan betedir dendi	70	2023-03-15 00:00:00.000	Serdivan	1	0	iPhone için Twitter	Kültür	Nötr	0.557304799556732	Çok Olumsuz
sabri abi o gözkele koparılar	32	2023-03-15 00:00:00.000	Adapazan	54	0	iPhone için Twitter	Siyaset	Çok Olumlu	0.967404246330261	Çok Olumlu
kendin için yaşa onlar doymuyor	32	2023-03-15 00:00:00.000	Erenler	3	0	Android için Twitter	Sağlık	Çok Olumlu	0.98800128698349	Çok Olumlu
bünye konser istiyor yüksek ses istiyor serserlik istiyor	58	2023-03-15 00:00:00.000	Adapazan	8	0	iPhone için Twitter	Kültür	Çok Olumlu	0.862151205539703	Olumlu

Şekil 5.5. ETL İşlemi Sonrası Metinler

Şekil 5.6'da FactTweetler tablosuna ait akış görülmektedir. Bu akışta ilk adımda, FactTweetler tablo bilgisi Excel'den Azure SQL staging.FactTweetler tablosuna aktarılır.

İkinci adımda ise, boyutsal tablolarla ilişkilendirilerek elde edilen ID sütunlarıyla birlikte oluşturulan staging.FactTweetlerView isimli View'dan ods.FactTweetler tablosuna aktarım sağlanır.



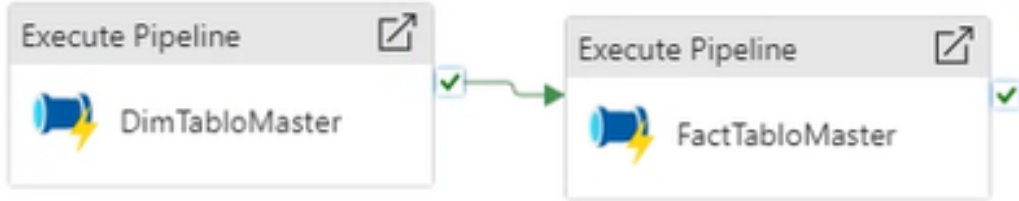
Şekil 5.6. İşlem Tablosu Veri Aktarımı

Şekil 5.7'de boyutsal tabloların Excel'den Azure SQL staging tablolarına ve bu tablolardan ODS tablolarına akışı gösterilmektedir. Lokasyon tablosu stabil bir tablo olduğu için bu akış mimarisine eklenmemiş olup, manuel sorgu ile Azure SQL içerisinde oluşturulmuştur.



**Şekil 5.7.** Boyut Tabloları Veri Aktarımı

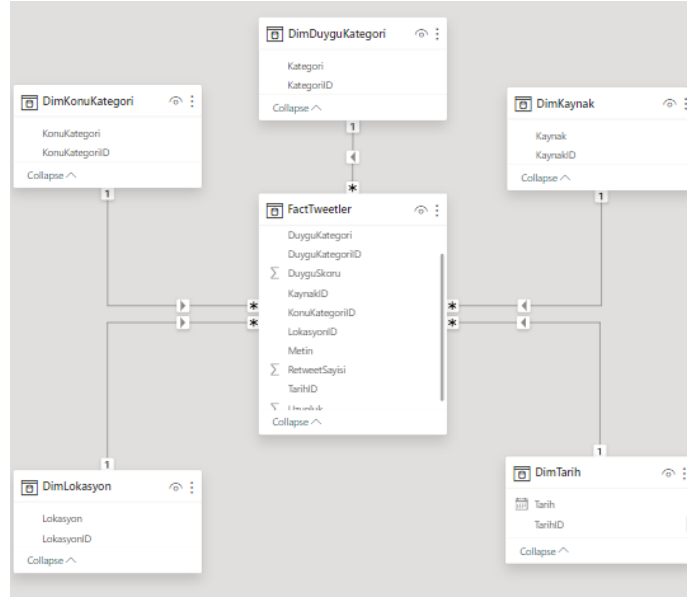
Veri akışlarının çalışması Şekil 5.8’de gösterilmiştir. Dimension ve Fact Master akışları sırasıyla birbirine bağlı olarak çalışmaktadır. İşlem tablosu boyut tablolarından referans bilgiler içerdiğinden dolayı ilk adımda Dimension tablo akışına ikinci adımda Fact tablo akışına yer verilmiştir.



**Şekil 5.8.** Dim ve Fact Tabloları Veri Akışları

### 5.1.5. Veri modelleme ve görselleştirme

Azure SQL Veri tabanına bağlanarak boyut(dimension) ve işlem(transaction) tablolarıyla yıldız şeması veri modeli oluşturulup analizlerin çıktılarını daha anlamlı ve anlaşılabilir hale getirmek için Power BI Veri Görselleştirme aracı ile raporun çıktısının sunulması adına ilgili rapor oluşturulmuştur. Gerekli raporların yapılabilmesi için oluşturulan Yıldız Şeması veri modeli Şekil 5.9’da gösterilmiştir.

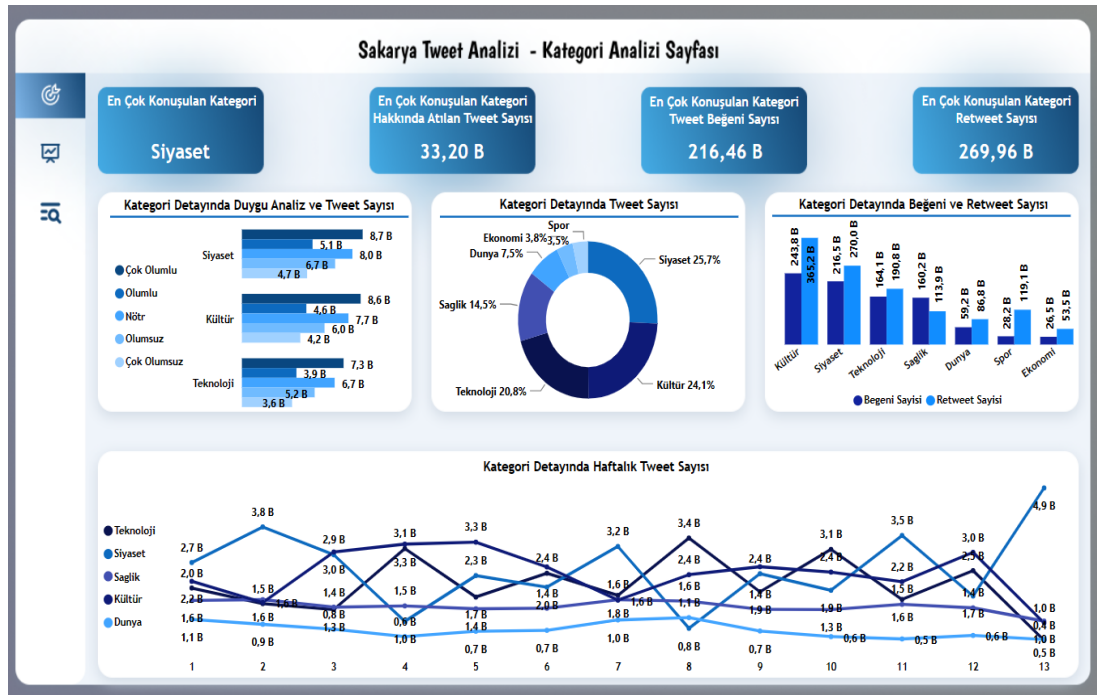


Şekil 5.9. Proje Veri Modeli



## 6. BULGULAR

Twitter API ile, Sakarya ili merkezi (40.771945624249085, 30.388644933700565) 25 kilometre çapında ve 01/01/2023 – 31/03/2023 tarihleri arasında atılan yaklaşık 150.000 tweet metin madenciliği ile elde edilmiş olup, gerekli ön işleme ve temizleme işlemleri yapıldıktan sonra analize hazır 125.000 tweet haline getirilmiştir. Verileri anlamlandırılmak ve edilen bilgi birikiminin bilgeliğe dönüştürülmesi için Power BI Desktop veri görselleştirme uygulaması ile aşağıda bulunan rapor görselleri oluşturulmuştur. Raporlar 3 farklı analiz kategorisine ayrılıp farklı sayfalarda görselleştirilmiş ve analiz edilmiştir. Şekil 6.1’de ilk rapor sayfası olarak kategori bazlı analizleri içermektedir.



Şekil 6.1. Kategori Analizi Sayfası

Kategori Analizi Sayfası grafiklerinde;

Kategori bazlı duygu analiz ve tweet sayısı grafiğinde tweetlerde bulunan kategori ve duygu tahminiyle etiketlenmiş verilerin dağılımı görülmekte. Grafiğe göre en çok tweet paylaşımının Siyaset kategorisiyle ilgili yapıldığı görülmektedir. Duygu analizi sonucunda genel olarak değerlendirildiğinde Çok olumlu ve Olumlu olarak

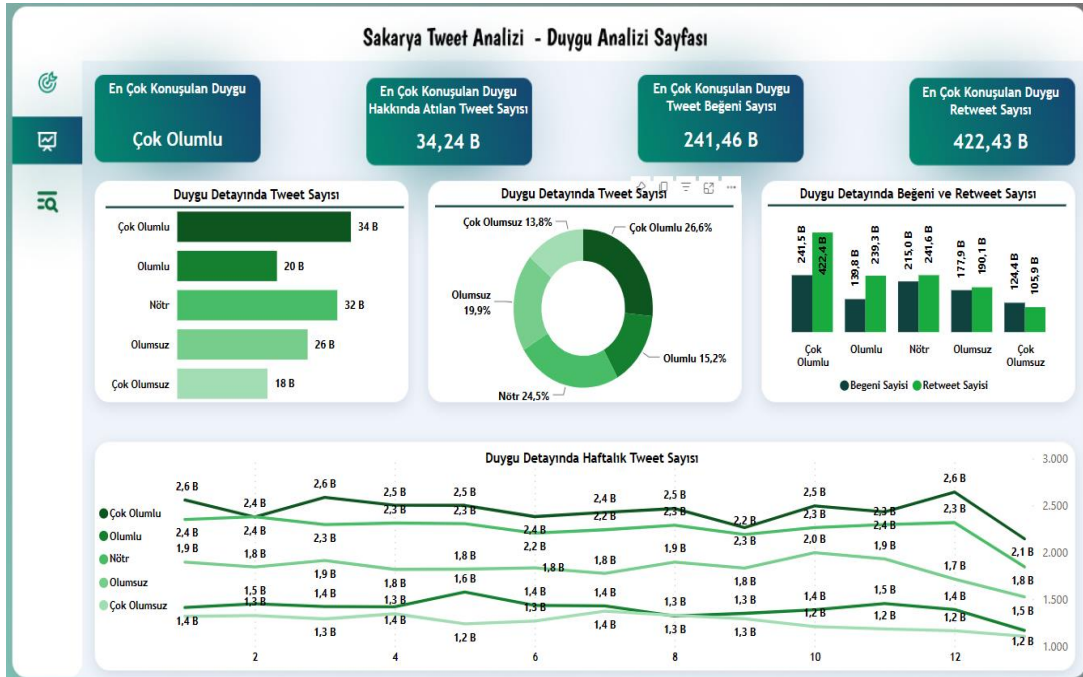
tahminlenen verilerin Çok Olumsuz ve Olumsuz olarak tahminlenen verilerden daha fazla sayıda olduğu sonucuna ulaşılmaktadır.

Kategori Detayında Tweet sayısı grafiğinde ise hangi kategoride ne kadar tweet paylaşıldığı görülmektedir. İlk sırayı Siyaset kategorisi alırken, ikinci sırada Kültür, üçüncü sırada ise Teknoloji kategorileri yer almaktadır.

Kategori Detayında Beğeni ve Retweet sayısı grafiği, tahminleme yapılan kategorilerin beğeni ve retweet sayısını bizlere göstermektedir. En çok tweet paylaşımı Siyaset kategorisinde yapılmasına rağmen en fazla beğeni ve retweet sayısına sahip olan kategorinin içerisinde günlük konuşmalar, kültürel paylaşımlar ve duygusal paylaşımları da bulunduran Kültür kategorisi olduğu gözlemlenmektedir.

Kategori bazında haftalık tweet sayısı grafiğinde kategorilere ait paylaşımların haftalık çizelgesi sunulmuştur. En fazla konuşulan 5 kategori grafiğe dahil edilmiştir. Dünya ve Sağlık kategorisindeki hareketler daha standart bir şekilde ilerlerken özellikle Siyaset ve Kültür kategorilerinde hızlı yükseliş ve düşüşler görülmektedir.

Şekil 6.2’de ise raporun ikinci sayfasında bulunan duygu bazlı analiz sayfası gösterilmektedir.



Şekil 6.2. Duygu Analizi Sayfası

Duygu analizi sayfasında ise;



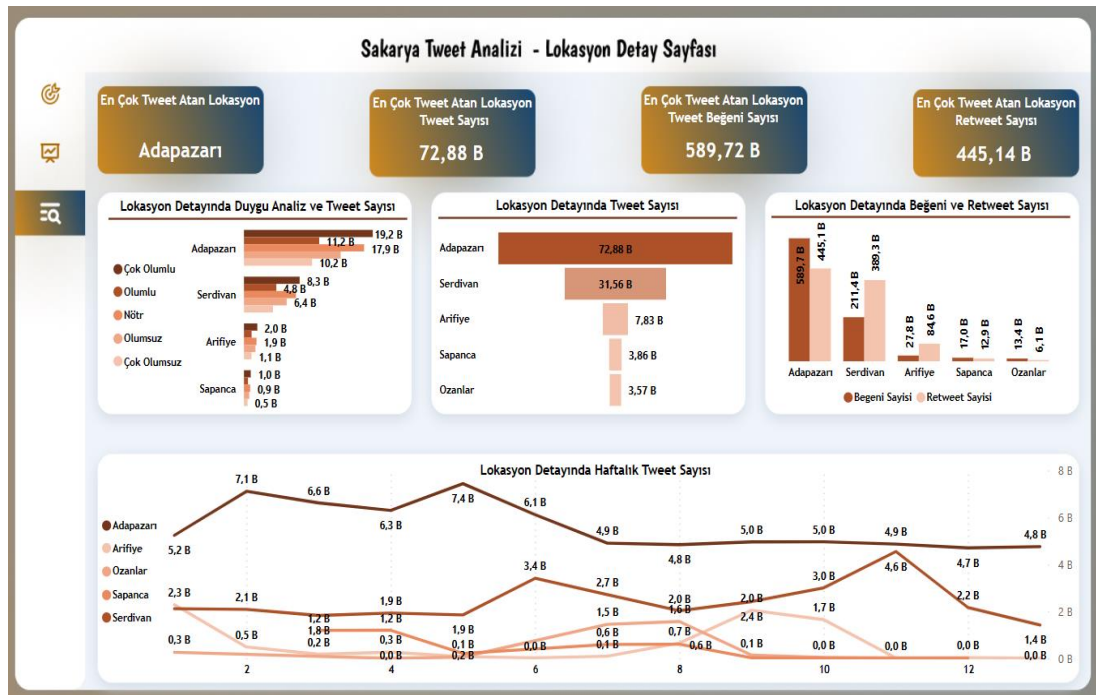
Duygu kategorilerindeki tweet sayıları gösterilmiştir. Yapılan paylaşımlarda en yüksek sayıya sahip olan duygu paylaşımı Çok Olumlu ve Nötr olarak görülmektedir. Olumsuz paylaşımların da üçüncü sırada bulunuyor olması dikkat çekebilir niteliktedir.

Duygu detayında tweet grafiğinde ise etiketlenen duygu kategorilerinin toplam tweet sayısına oranları gösterilmektedir. Çok olumlu, Nötr ve Olumsuz tweet paylaşımlarında sıklık olduğu sonucuna ulaşılabilir.

Duygu detayında beğeni ve retweet sayısı grafiğini incelediğimizde, en çok beğeni ve retweet alan duygumuzun Çok olumlu tweet paylaşımları olduğunu görmekteyiz. İlk üç duygu kategorimizin genel olumlu paylaşımlar olduğunu düşündüğümüzde madenciliğini yaptığımız Sakarya ilindeki paylaşımlarda olumlu tweetlerin daha çok ilgi gördüğü sonucu ortaya çıkmaktadır.

Duygu bazında haftalık tweet sayımızı incelediğimizde verilerin çekildiği dönemde genel olarak benzer bir sürecin ilerlediğini sadece Nötr ve Çok Olumlu duyguya sahip paylaşımlarda bazı zamanlarda hareketlilikler olabildiğini gözlemlemekteyiz.

Şekil 6.3'te üçüncü ve son rapor sayfası olan lokasyon bazlı rapor sayfasına yer verilmiştir.



Şekil 6.3. Lokasyon Detay Sayfası

Lokasyon detay sayfası grafiklerinde ise;

Lokasyon detaylı duygu analiz ve tweet sayısı grafiđimiz en çok tweet paylaşımı yapan ilk 4 ilçemizde paylaşımların içerdiđi duygular gösterilmiştir. İlk iki sırada yer alan Adapazarı ve Serdivan ilçelerinde genel olarak Çok Olumlu ve Nötr duygulu paylaşımların yapıldığını üçüncü sırada ise Olumsuz paylaşımların yapıldığını görebilmekteyiz.

Lokasyon Detayında tweet sayısı grafiđi incelendiđinde, en çok tweet paylaşımı yapan ilçelerin Adapazarı, Serdivan ve Arifiye ilçeleri olduđu görölmektedir.

Lokasyon detayında beđeni ve retweet sayısı grafiđinde görüldüđu üzere, en fazla beđeni ve retweet alan ilçe Adapazarı olmuştur. Tweet sayılarındaki farklılıđa rağmen Serdivan ilçesinden yapılan paylaşımlarda bulunan retweet sayısı da dikkat çekmektedir.

Lokasyon bazında haftalık tweet sayısı grafiđinde çeşitli hareketlilikler görölmektedir. Tweet çekim işleminin yapıldığı dönem ve konuma bađlı olarak, öğrenim döneminin bitişi ve başlangıcının da ilçelerdeki hareketliliđe bulunduđu etki varsayım olarak yorumlanabilmektedir.

## 7. SONUÇLAR

Doğal Dil İşleme ve İş Zekası perspektifinden yürütülen Twitter verileri analizi, özellikle duygu analizi ve kategori tahminleme üzerinden elde edilen sonuçları iş zekasıyla bütünleştirme amacı bu tezin temelini oluşturmaktadır. Sosyal medya platformları, hızla değişen ve geniş veri kümelerini barındıran bir ortam sunarak, iş zekası uygulamaları için zengin bir kaynak haline gelmiştir. Bu noktada, Twitter verileri kullanılarak elde edilen duygu analizi sonuçları iş zekası perspektifiyle değerlendirilerek, bilgiye dayalı kararlar alınması ve öngörülse analizler yapılabilmesi amaçlanmıştır. Bu çalışma, sosyal medya analitiği ile iş zekası arasındaki köprüyü kurarak, bilgi yönetimi ve stratejik planlama süreçlerine katkıda bulunmayı hedeflemektedir.

Kategori analiz çalışmasında, Sakarya ilinde en fazla “Siyaset” ve “Kültür” kategorilerinde paylaşımlar görülmüş olup genel olarak bu kategorilerle ilgili yapılan paylaşımlarda “Çok Olumlu” ve “Nötr” duygularının baskın olduğu sonucuna ulaşılmıştır. Duygu analizi çalışması değerlendirildiğinde, paylaşım yoğunluğunun “Çok Olumlu”, “Nötr” ve “Olumsuz” duygu sınıflarıyla olduğu görülmektedir. Retweet ve beğeni sayıları ise pozitif duygularda daha yüksek değere ulaşmıştır. Lokasyon analizinde ise, en fazla tweet paylaşımının Adapazarı ve Serdivan ilçelerinden olduğu görülmektedir. Tweet paylaşım sayılarının arasında farkın fazla olmasına rağmen Serdivan ilçesinden yapılan tweet paylaşımlarının çokca değer gördüğü ve retweet yapıldığı görülmektedir.

Twitter üzerinden elde edilen tweetler, ilk aşamada sadece bir veri niteliği taşıyorken, çeşitli veri ve metin madenciliği yöntemleri uygulanarak, temiz ve analize hazır metinler haline gelmiş ve bilgi birikimine dönüştürülmüştür. Analiz edilen tweetler anlamlandırılarak bir görsel oluşturulmuş ve Veri – Bilgelik döngüsü başarıyla tamamlanmıştır. Kategori ve duygu analizi madenciliği yöntemleri kullanılarak, Sakarya halkının küçük bir kesiminin duygusal durumu, düşünce içeriği ve hissiyatı üzerine bir analiz gerçekleştirilmiştir. Gelecek çalışmalarda ise Meta şirketinin kullanıma açtığı Threads gibi farklı sosyal medya platformları daha farklı analiz kitlesi

ve karara yönelik analiz çalışmalarla desteklenerek, sürekli entegrasyon ve geliştirme yöntemleriyle sürdürülebilir bir analiz, karar ve aksiyon döngüsüyle yeni bir ürün ortaya çıkartılabilir. Bu sayede analize dahil edilen ortam için düzenlemeler ve iyileştirmeler yapılmak üzere değerli adımlar atılabilir.

## KAYNAKLAR

- Abro, A., Talpur, S., Jumani, A. (2023). Natural Language Processing Challenges and Issues: A Literature Review. *GU J Sci* 36(4): 1522-1536
- ABRO, A. A., TALPUR, M. S. H., & JUMANI, A. K. (2023). Natural Language Processing Challenges and Issues: A Literature Review. *Gazi University Journal of Science*, 1-1.
- Acikalin, U., Bardak, B., Kutlu, M. (2020). Turkish Sentiment Analysis Using BERT, 28th Signal Processing and Communications Applications Conference (SIU).
- Ahmad, S., Miskon, S. Alabdan, R. Tlili, I. (2020). Towards Sustainable Textile and Apparel Industry: Exploring the Role of Business Intelligence Systems in the Era of Industry 4.0. *Sustainability*, No 7, p. 2632
- Aravi, G. (2014). Social Media Analysis with Text Mining, Istanbul Aydin University Institute of Science and Technology
- Arslantekin, S. (2003). Data Mining and Information Centers. *Turkish Librarianship* 17,4 (2003),369-380
- Ay, R. A. (2021). Doğal Dil İşleme Ve Kümeleme Yöntemleri İle Müşteri Yorumlarının İncelenmesi.
- Aydin, C. (2018). Classification of fire station requirement using machine learning algorithms, *European Journal of Science and Technology* No. 14, pp. 169-175
- Bahja, M. (2020). Natural Language Processing Applications in Business. *E-Business - Higher Education and Intelligence Applications*
- Bhargava, M.G., Kiran, K.T., & Rao, D.R. (2018). Analysis and Design of Visualization of Educational Institution database using Power BI Tool. *Global journal of computer science and technology*. Volume 1 Issue 4 Version 1.0
- Breslin, Mary Beth. (2004). DW MODELS Data Warehousing Battle of the Giants : Comparing the Basics of the Kimball and Inmon Models.
- Celik, S., Akdamar, E. (2018). Big Data and Data Visualization, *Academic Perspective International Refereed Journal of Social Sciences*, (65), 253-264
- Choi, L. Panjaitan, S., Apriliasari, D. (2022). The Effectiveness of Business Intelligence Management Implementation in Industry 4.0. *Startupneur Business Digital (SABDA)* Vol. 1 No. 2
- Cover, T., Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27.
- Çelik, D., Karakas, A., Bal, G., Gültunca, C., Elçi, A., Buluz, B., Alevli, M. (2013). Towards an information extraction system based on ontology to match resumes and jobs. *IEEE 37th annual computer software and applications conference workshops* pp 333-338

- Çetin, G., Tanrıöver, Ö. (2020). Personel İş Zekası Sistemi ve Veri Madenciliği ile Personel Memnuniyetinin Ölçülmesi. *Avrupa Bilim ve Teknoloji Dergisi Özel Sayı*, S. 323-334, Kasım 2020
- Dragoni, M., Villata, S., Rizzi, W., Governatori, G. (2016). Combining NLP Approaches for Rule Extraction from Legal Documents. *HAL Open Science*
- Elgamal, Neveen. (2016). Data Warehouse Test Routine Descriptions
- El-Wessimy, M., Mokhtar, H., Hegazy, O. (2013). Enhancement Techniques For Data Warehouse Staging Area. *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.6*
- Feng T, Timmermans HJP. (2010). Comparison of advanced imputation algorithms for detection of transportation mode and activity episode using GPS data. *13th WCTR, Rio de Janeiro, Brazil*
- Grossecq, G., Holotescu, C. (2008). Can We Use Twitter For Educational Activities? The 4th International Scientific Conference eLearning and Software for Education Bucharest.
- Gündüz, D. (2015). İş Zekası Uygulamaları Ve Pazar Sepeti Analizi
- Gürler, A., Yılmaz, A., Tekerek, M. (2018). Data Visualization and Infographics, *KSU Journal of Engineering Sciences Research Paper 21(2):131-148*
- Halim, K.K., Halim, S., & Felecia. (2019). Business Intelligence for Designing Restaurant Marketing Strategy: A Case Study. *Procedia Computer Science. Journal of Content, Community & Communication Vol. 11 Year 6*
- Han, J.; Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, USA
- Hannula, M.; Pirttimäki, V. (2003). Business Intelligence Empirical Study On The Top 50 Finnish Companies. *J. Am.Acad. Bus.,2, 593–599*
- Hastie, T., Tibshirani, R., Friedman, J.H. (2009). *The elements of statistical learning: data mining, Inference and Prediction*, 2nd edn. Springer, New York, USA p. 533.
- Hosmane, S. ve Rakshitha, S. (2021). Overview of Data Warehouse. *International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 05*
- Hotho, A.; Nürnberger, A., Paass, G. (2005). A Brief Survey of Text Mining, *GLDV Journal for Computational Linguistics and Language Technology*.
- Inmon, W.H. (2002). *Building the Data Warehouse (Third Edition)*, New York: John Wiley & Sons.
- Inmon, W.H. and Imhoff, C. (2002). *Corporate Information Factory Components*
- Jadhav, M., Goje, A., Chavan, J. (2021). Data Migration From On Premise Oracle Database To SQL Manage Instance On Azure Cloud Using Azure Data Factory -A Working Approach. *Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 3524-3528*
- Jordan, S., Hovet, S., & Fung, I., Liang, H., Fu, K. & Tse, Z. (2018). Using Twitter for Public Health Surveillance from Monitoring and Prediction to Public Response. *Data 4, no. 1: 6*. <https://doi.org/10.3390/data4010006>

- Kaaur, V. (2016). Business Intelligence And E-banking: A Study Of Bi Importance In Banking Sector. Biz and Bytes, Vol. 7. Issue: 1
- Karabulut, Y. Ve Küçüksille, E. (2018). Twitter Profesyonel İzleme ve Analiz Aracı. Teknik Bilimleri Dergisi Cilt 8, Sayı 2, 17-24, 2018
- Kazan, S., Karakoca, H. (2019). Product Category Classification with Machine Learning, Sakarya University Journal Of Computer And Information Sciences Vol. 2, No. 1
- Kharis, M., Laksono, K., Suhartono, Ridwan, A., Mintowati, Yunissefendri. (2022) Tokenization and Lemmatization on German Learning Textbook Level A1 of CEFR Standard. Journal of Higher Education Theory and Practice
- Khilari, Dr.Sunil and Singh, Chandrani and Mane, Mr.Bharat (2022). Business Intelligence Tool-Power BI for Performance Management.
- Kimball, R. and Ross, M. (2000). The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling (Second Edition), New York: John Wiley & Sons.
- Kumar, R.,& RAVI, Vadlamani. (2015). A Survey On Opinion Mining And Sentiment Analysis: Tasks, Approaches And Applications, Knowledge-Based Systems, 89, 14-46
- Kuyumcu, Birol & Aksakalli, Cuneyt & Delil, Selman. (2019). An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing. 1-4. 10.1145/3342827.3342828.
- Küçükkartal, Hatice K. (2020).Twitter'daki Verilere Metin Madenciliği Yöntemlerinin Uygulanması. ESTUDAM Bilişim Dergisi Cilt 1, Sayı 2 , 10-13, 2020
- MadeSukarsa, I. & Wisswani, Ni & Putra, I. & Linawati, Linawati. (2012). Change Data Capture on OLTP Staging Area for Nearly Real Time Data Warehouse Base on Database Trigger. International Journal of Computer Applications. 52. 32-37.
- Mattila, M., Salman, H. (2018). Analysing Social Media Marketing on Twitter using Sentiment Analysis. EXAMENSARBETE INOM TEKNIK,GRUNDNIVÅ, 15 HP STOCKHOLM, SVERIGE
- Maynard, D., Saggion, H., Yankova, M., Bontcheva, K., & Peters, W. (2007). Natural Language Technology for Information Integration in Business Intelligence. BIS 2007: Business Information Systems pp 366–380
- Nay, J. (2018). Natural Language Processing and Machine Learning for Law and Policy Texts. Legal Informatics. Cambridge University Press.
- Nguyen, Minh, Viet Dac Lai, Amir Pouran Ben Veyseh and Thien Huu Nguyen. (2021). Trankit: A Light-Weight Transformer-based Toolkit for Multilingual Natural Language Processing. Conference of the European Chapter of the Association for Computational Linguistics
- Olexov, C.(2014) Business intelligence adoption: A case study in the retail chain. WSEAS Trans. Bus. Econ.,11,95–106
- Padmanayana, & Varsha, & K, Bhavya. (2021). Stock Market Prediction Using Twitter Sentiment Analysis. International Journal of Scientific Research in Science and Technology. 265-270.

- Pazarçeviren, S., Zor, Ü., Gürbüz, F. (2015). Business Intelligence: Conceptual Framework, Components and Process, *Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi*, 2015, Yıl:3, Cilt:3, Sayı:1.
- Performance Evaluation Utilizing Geological Data. IRMA International Conference
- Rao, B. Jogeswara, Prof. M. Surendra Prasad Babu and Dr. S. Hanumanth Sastry. (2019). A New Methodology to Perform Big Data Analytics on Business Warehouse Data.
- Rennie, J.D., Shih, L., Teevan, J., & Karger, D.R. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. *International Conference on Machine Learning*.
- Romero, C., Ortiz, J., Khalaf, O., Prado, A. (2021). Web Application Commercial Design for Financial Entities Based on Business Intelligence. *Computers Materials & Continua* 67(3):3177-3188
- Sarica, S. and Jianxi, L. (2020) Stopwords in Technical Language Processing. *PloS one* 16 8
- Scott, S., Matwin, S. (1999). Feature Engineering for Text Classification, *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning*, 379–388
- Seyed-Abbasi, B. and Madesi, V. (2015). Data Warehouse Design Using Row and Column Data Distribution. *Int'l Conf. Information and Knowledge Engineering, IKE*
- Seyed-Abbasi, B. and Pusey, L. (2007). XML in a Data Warehouse Design:
- Shakeel, M., Barsaiyan, S., Sijoria, C. (2020). Twitter As A Customer Service Management Platform: A Study On Indian Banks
- Shobirin, K., Iskandar, A., Swamardika, I. (2017). Data Warehouse Schemas using Multidimensional Data Model for Retail. *International Journal of Engineering and Emerging Technology*, Vol. 2, No. 1
- Singh, S. and Mahmood, S. (2021). The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architectures. *IEEE Access* 9, 68675-68702.
- Sverrison, G. ve Hákonarson, H. (2023). Modernizing the Data Integration Process Implementation of Azure Data Factory at Össur.
- Taktak, S., Alshomrani, S., Feki, J. and Zurfluh, G. (2017). The Power of a Model-Driven Approach to Handle Evolving Data Warehouse Requirements. *International Conference on Model-Driven Engineering and Software Development*
- Türk, A. (2020). Numune Üretimi Atölyesinde İş Zekası Uygulaması
- Uddin, M., Bin Hafiz, F., Hossain, S. and Islam, S. (2022). Drug Sentiment Analysis using Machine Learning Classifiers. *International Journal of Advanced Computer Science and Applications(IJACSA)*, 13(1).



- Viswanathan, L., Chandra, B., Lang, W., Ramachandra, K., Patel, J., Kalhan, A., DeWitt, D. and Halverson, A. (2017). Predictive Provisioning: Efficiently Anticipating Usage in Azure SQL Database. IEEE 33rd International Conference on Data Engineering (ICDE) (2017): 1111-1116.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics
- Yang, Qishan & Ge, Mouzhi & Helfert, Markus. (2019). Analysis of Data Warehouse Architectures: Modeling and Classification. 21st International Conference on Enterprise Information Systems 604-611.
- Yıldırım, S (2022). Bert-base Turkish Sentiment Model, <https://huggingface.co/savasy/bert-base-turkish-sentiment-cased>, adresinden 10 Şubat 2023 tarihinde alınmıştır.
- Yıldırım, S (2022). Turkish Text Classification, <https://huggingface.co/savasy/bert-turkish-text-classification>, adresinden 10 Şubat 2023 tarihinde alınmıştır.
- Yüksel, A. ve Tan, F. (2018). Metin Madenciliği Teknikleri İle Sosyal Ağlarda Bilgi Keşfi. Mühendislik Bilimleri ve Tasarım Dergisi 6(2), 324 – 333, 2018



## ÖZGEÇMİŞ

Ad-Soyad : Furkan SARAÇOĞLU

### ÖĞRENİM DURUMU:

- **Lisans** : 2020, Sakarya Üniversitesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Bilişim Sistemleri Mühendisliği

### MESLEKİ DENEYİM VE ÖDÜLLER:

- 2020 – 2021 yılları arasında Toyota Motor Avrupa Sakarya Şubesi'nde İş Zekası Danışmanı olarak çalıştı.
- 2021-2023 yılları arasında Nestle Türkiye Genel Merkezi'nde Veri Analitiği Uzmanı olarak çalıştı.
- 2023 yılında Gunvor Türkiye Destek Hizmetleri Ltd. Şti'de İş Zekası Veri Analisti olarak çalışmaya başladı.