daire, meneviş mavisi, kalıp, desen, düzen, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

**YILDIZ TEKNİK ÜNİVERSİTESİ**

**KİMYA-METALÜRJİ FAKÜLTESİ**

**MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ**

**BİTİRME ÇALIŞMASI**

**DOĞAL DİL İŞLEME TEKNİKLERİ İLE FİNANSAL PİYASA DUYGU ANALİZİ**

Tez Yöneticisi: Prof. Dr. Coşkun GÜLER

22052610 İlker EFİL

İstanbul, 2024

**© Bu tezin bütün hakları Yıldız Teknik Üniversitesi Matematik Mühendisliği Bölümü’ne aittir.**

İÇİNDEKİLER

[KISALTMA LİSTESİ 4](#_Toc187020521)

[ŞEKİL LİSTESİ 5](#_Toc187020522)

[TABLO LİSTESİ 6](#_Toc187020523)

[ÖNSÖZ 7](#_Toc187020524)

[ÖZET 8](#_Toc187020525)

[ABSTRACT 9](#_Toc187020526)

[1 GİRİŞ 10](#_Toc187020527)

[**1.1 Geçmiş Araştırmaların İncelenmesi 11**](#_Toc187020528)

[2 YÖNTEM 13](#_Toc187020529)

[**2.1 Doğal Dil İşleme (NLP) 13**](#_Toc187020530)

[**2.1.1 Doğal Dil İşleme Tarihçesi 13**](#_Toc187020531)

[**2.1.2 Doğal Dil İşleme Kullanım Alanları 14**](#_Toc187020532)

[**2.1.3 Doğal Dil İşlemede Karşılaşılan Sorunlar 15**](#_Toc187020533)

[**2.1.4 Doğal Dil İşleme Süreci 16**](#_Toc187020534)

[**2.2 Doğal Dil İşleme Teknikleri 17**](#_Toc187020535)

[**2.2.1 Metin Ön İşleme 17**](#_Toc187020536)

[**2.2.2 Özellik Çıkarımı 18**](#_Toc187020537)

[**2.2.3 Metin Sınıflandırma 21**](#_Toc187020538)

[**2.2.4 Varlık Adı Tanıma 21**](#_Toc187020539)

[**2.2.5 Dil Modellemesi 22**](#_Toc187020540)

[**2.2.6 Makine Çevirisi 22**](#_Toc187020541)

[**2.2.7 Metin Üretimi 22**](#_Toc187020542)

[**2.2.8 Duygu Analizi 23**](#_Toc187020543)

[**2.3 Granger Nedensellik Testi 24**](#_Toc187020544)

[**2.3.1 Verinin Hazırlanması 25**](#_Toc187020545)

[**2.3.2 Durağanlık Testi 25**](#_Toc187020546)

[**2.3.3 Gecikme Sayısının Belirlenmesi 25**](#_Toc187020547)

[**2.3.4 Hipotez Testi 26**](#_Toc187020548)

[**2.3.5 Sonuçların Yorumu 27**](#_Toc187020549)

[3 UYGULAMA 28](#_Toc187020550)

[**3.1 Veri Setinin Hazırlanması 28**](#_Toc187020551)

[**3.2 Metin Ön İşleme 28**](#_Toc187020552)

[**3.3 Duygu Etiketleme 30**](#_Toc187020553)

[**3.4 Granger Nedensellik Testi 31**](#_Toc187020554)

[**3.5 Duygu Skorları ile Fiyat Arasındaki İlişki 33**](#_Toc187020555)

[4 SONUÇ VE ÖNERİLER 35](#_Toc187020556)

[KAYNAKLAR 36](#_Toc187020557)

[EKLER 38](#_Toc187020558)

[ÖZGEÇMİŞ 41](#_Toc187020559)

# **KISALTMA LİSTESİ**

NLP Natural Language Processing (Doğal Dil İşleme)

TF-IDF  Term Frequency-Inverse Document Frequency

BoW Bag of Words (Kelime Torbası)

Word2Vec Word to Vector (Kelimeden Vektöre)

GloVe Global Vectors for Word Representation

USE Universal Sentence Encoder (Evrensel Cümle Kodlayıcı)

BERT Bidirectional Encoder Representations Transformers

GPT Generative Pre-trained Transformer

FinBERT Financial Bidirectional Encoder Representations Transformers

VADER Valence Aware Dictionary and Sentiment Reasoner

NER Named Entity Recognition (Varlık Adı Tanıma)

NMT Neural Machine Translation (Nöral Makine Çevirisi)

ADF Augmented Dickey-Fuller

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

MA Moving Average (Hareketli Ortalama)

# **ŞEKİL LİSTESİ**

Şekil 2.1 NLP kullanım alanları

Şekil 2.2 NLP’de karşılaşılan sorunlar

Şekil 2.3 Doğal dil işleme sürecine bir örnek

Şekil 2.4 Bag of Words için bir örnek

Şekil 2.5 TF-IDF için örnek bir işlem

Şekil 2.6 CBOW ve Skip-gram örnek

Şekil 2.7 Durağan iki veri setinin grafiği

Şekil 3.1 Veri setinin ilk 5 haberi

Şekil 3.2 Temizlenmiş ver seti

Şekil 3.3 Duraksama kelimeleri kaldırılmış veri seti

Şekil 3.4 Tokenize edilmiş veri seti

Şekil 3.5 Lemmatize edilmiş veri seti

Şekil 3.6 Duygu skoru hesaplanması için fonksiyon kodu

Şekil 3.7 Duygu skorları ve ilgili haber tarihleri

Şekil 3.8 Parite günlük kapanış fiyatı ve değişim oranı

Şekil 3.9 Augmented Dickey-Fuller testi (p = 0.05 olarak kabul edildi)

Şekil 3.10 Parite fiyatı ve duygu skorları arasındaki ilişki

# **TABLO LİSTESİ**

Tablo 2.1 Modellerin karşılaştırılması

Tablo 2.2 Hipotez testi

Tablo 3.1 Granger nedensellik testi 5 gecikme ile sonuçları

# **ÖNSÖZ**

Bu tez çalışması, finansal haberlerden elde edilen duygu skorlarının piyasa fiyatları üzerindeki etkisini analiz etmek ve duygu analizi yöntemlerinin piyasa hareketlerini tahmin etme sürecindeki rolünü değerlendirmek amacıyla gerçekleştirilmiştir. Özellikle, finansal haberlerin içerdiği duyarlılığın piyasa fiyatlarındaki kısa ve uzun vadeli dalgalanmalar üzerindeki etkisini anlamak, yatırımcı davranışlarını analiz etmek ve bu etkilerin finansal piyasa dinamiklerine olan katkılarını belirlemek hedeflenmiştir. Çalışma, duygu analizi tekniklerinin finansal veri analitiği içindeki yerini ve önemini vurgulamayı amaçlamaktadır.

Araştırmalarım süresince bana desteğini esirgemeyen her konuda yardımcı olan, sorularımı her daim cevaplayan tez danışmanım Prof. Dr. Coşkun GÜLER hocama, her daim yanımda olan, bana motivasyon veren ve desteklerini esrigemeyen aileme ve arkadaşlarıma teşekkürlerimi sunarım.

İlker EFİL

İstanbul, 2024

# **ÖZET**

Bu çalışma, finansal haberlerden elde edilen duygu skorlarının piyasa fiyatları üzerindeki etkisini analiz etmek ve duygu analizi yöntemlerinin piyasa hareketlerini tahmin etmedeki rolünü değerlendirmek amacıyla gerçekleştirilmiştir. Çalışmada, Kaggle platformundan elde edilen 2021-2023 yılları arasında Yahoo Finance ve Binance kaynaklarından toplanan Bitcoin ile ilgili haber metinleri kullanılmıştır. Metinler, doğal dil işleme (NLP) teknikleriyle işlenmiş ve duygu etiketleme için finansal metinlere özel olarak geliştirilmiş FinBERT modeli kullanılarak -1 ile 1 arasında duygu skorları atanmıştır. Duygu skorlarının Bitcoin fiyatları üzerindeki nedensel etkisini belirlemek için Granger nedensellik testi uygulanmıştır.

Test sonuçları, duygu skorlarının 1, 2, 3 ve 4 günlük gecikmelerde Bitcoin fiyatları üzerinde anlamlı bir etkisi olduğunu ortaya koymuş, ancak 5 günlük gecikmelerde bu ilişkinin anlamlı olmadığı belirlenmiştir. Grafiksel analizler, duygu skorlarının hareketli ortalamalarının Bitcoin fiyatlarındaki uzun vadeli eğilimlerle genel olarak paralel bir şekilde hareket ettiğini göstermektedir. Bu durum, piyasa duyarlılığının hem kısa hem de orta vadeli fiyat hareketlerinde etkili olabileceğini, ancak daha uzun vadede piyasa dinamiklerinin makroekonomik koşullar, haber akışı ve diğer faktörler tarafından şekillendiğini düşündürmektedir.

Elde edilen bulgular, duygu analizi yöntemlerinin yatırım stratejileri geliştirilmesinde ve piyasa hareketlerinin tahmin edilmesinde potansiyel bir araç olarak kullanılabileceğini göstermektedir. Bununla birlikte, daha kapsamlı analizler için sosyal medya duyarlılığı, zincir üstü veriler ve makroekonomik göstergeler gibi farklı veri kaynaklarının birleştirilmesi önerilmektedir. Çalışma, finansal piyasalarda duygu analizi kullanımına yönelik önemli bir temel sunmuş ve bu alanda yapılacak gelecekteki araştırmalara ışık tutmuştur.

**Anahtar Kelimeler:** Finansal Haberler, Duygu Analizi, Duygu Skorları, Piyasa Fiyatları, Doğal Dil İşleme (NLP), FinBERT, Granger Nedensellik Testi, Yatırım Stratejileri, Piyasa Duyarlılığı, Zaman Serisi Analizi

# **ABSTRACT**

This study was conducted to analyze the impact of sentiment scores derived from financial news on market prices and to evaluate the role of sentiment analysis methods in predicting market movements. The study used Bitcoin-related news articles collected from Yahoo Finance and Binance between 2021 and 2023, obtained from the Kaggle platform. The texts were processed using natural language processing (NLP) techniques, and sentiment labeling was performed with the FinBERT model, which was specifically developed for financial texts, assigning sentiment scores ranging from -1 to 1. The Granger causality test was applied to determine the causal impact of sentiment scores on Bitcoin prices.

The test results revealed that sentiment scores had a significant impact on Bitcoin prices with 1, 2, 3, and 4-day lags, while no significant relationship was found for 5-day lags. Graphical analyses showed that the moving averages of sentiment scores generally aligned with the long-term trends of Bitcoin prices. This indicates that market sentiment could influence both short- and medium-term price movements; however, in the longer term, market dynamics appear to be shaped by macroeconomic conditions, news flow, and other factors.

The findings indicate that sentiment analysis methods could be a valuable tool for developing investment strategies and predicting market trends. However, integrating additional data sources such as social media sentiment, on-chain metrics, and macroeconomic indicators is recommended for more comprehensive analyses. This study provides a foundation for using sentiment analysis in financial markets and offers insights for future research in this field.

**Keywords:** Financial News, Sentiment Analysis, Sentiment Scores, Market Prices, Natural Language Processing (NLP), FinBERT, Granger Causality Test, Investment Strategies, Market Sentiment, Time Series Analysis

# **1 GİRİŞ**

Finansal piyasalar, ekonomik gelişmelerin, sosyal faktörlerin ve bireysel yatırımcı davranışlarının sürekli etkileşim içinde bulunduğu dinamik ve karmaşık bir ekosistemdir. Bu piyasalarda fiyat hareketlerini anlamak ve öngörmek, bireysel yatırımcılardan kurumsal fon yöneticilerine, politika yapıcılardan akademisyenlere kadar geniş bir yelpazede birçok paydaş için hayati bir öneme sahiptir. Ekonomik belirsizliklerin arttığı dönemlerde bu ihtiyaç daha da belirgin hale gelirken, fiyat hareketlerini etkileyen unsurları doğru bir şekilde anlamak, etkili karar alma süreçleri için kritik bir gereklilik haline gelmiştir.

Teknolojinin hızla gelişmesi ve veri işleme kapasitesindeki artış, büyük ölçekli verilerin analiz edilmesini mümkün kılarak finansal piyasaların dinamiklerini daha derinlemesine incelemek için güçlü bir araç sunmuştur. Özellikle dijitalleşmenin artmasıyla birlikte haber metinleri, sosyal medya gönderileri, blog yazıları ve forum tartışmaları gibi metin tabanlı veri kaynakları, piyasa duyarlılığını yansıtan önemli göstergeler haline gelmiştir. Bu tür verilerin manuel olarak işlenmesi oldukça zaman alıcı ve yorucu olabileceğinden, doğal dil işleme (NLP) teknikleri, bu büyük veri yığınlarından anlamlı içgörüler çıkarma sürecinde etkili bir çözüm olarak öne çıkmıştır.

Doğal dil işleme, dil tabanlı veri kaynaklarının analizi için kullanılan bir yapay zeka alt disiplinidir ve son yıllarda finansal piyasalarda duyarlılık analizi için yaygın bir şekilde kullanılmaktadır. Haber bültenleri, sosyal medya gönderileri ve ekonomik raporlar gibi metin tabanlı veri kaynakları, piyasa katılımcılarının duygu ve düşüncelerini yansıtarak, fiyat hareketleri üzerinde doğrudan etkili olabilir. NLP teknikleri, bu metinlerden anlamlı bilgiler çıkararak, yatırım kararlarını desteklemek ve piyasa trendlerini tahmin etmek için kullanılabilmektedir.

Bu çalışmada, doğal dil işleme teknikleri kullanılarak finansal piyasalardaki duygu analiz edilmekte ve bu duyarlılığın piyasa dinamikleri üzerindeki etkisi incelenmektedir. Çalışma, finansal haber gönderilerinden elde edilen duygu verilerinin, fiyat hareketlerini anlamada nasıl kullanılabileceğini ele alırken, bu verilerden elde edilen duygu skorlarının piyasalar üzerindeki etkisini ortaya koymayı amaçlamaktadır. Araştırmada, duygu analizi modelleri ve teknikleri kullanılarak piyasa duyarlılığı ile fiyat hareketleri arasındaki ilişki detaylı bir şekilde inclenmiştir. Ayrıca, Granger nedensellik testi gibi ileri istatistiksel yöntemlerle duygu ve fiyat ilişkisi üzerinde analiz yaplımıştır.

Bu bağlamda tez, hem teorik bir temel sunmayı hem de uygulamalı bir yaklaşım ile finansal piyasalarda duygu analizi konusuna katkı sağlamayı amaçlamaktadır. Araştırmanın bulguları, yatırım stratejileri geliştirmek ve piyasa risklerini yönetmek isteyen bireyler ve kurumlar için değerli içgörüler sağlayabilir. Aynı zamanda, NLP tekniklerinin finansal analizdeki rolü üzerine yeni perspektifler sunarak akademik literatüre katkıda bulunmayı hedeflemektedir.

## 1.1 Geçmiş Araştırmaların İncelenmesi

Finansal piyasa metinlerinin analizinde doğal dil işleme (NLP) tekniklerinin etkili bir araç olduğu gösterilmiştir. Örneğin, Tetlock (2007) tarafından yapılan bir çalışma, gazetelerdeki olumsuz haberlerin piyasa volatilitesini artırdığını ve yatırımcı davranışlarını etkilediğini ortaya koymuştur. Bu çalışmada kullanılan temel yöntem, metinlerin duygu sınıflarına ayrılmasıdır ve doğruluk oranı %75 olarak rapor edilmiştir. Bu araştırma, duygu analizi ile piyasa hareketleri arasındaki bağlantıyı başarılı bir şekilde göstermesi açısından öncüdür[1].

Sosyal medya verileri üzerinden duygu analizi yapılması, piyasa tahmini çalışmalarında popüler bir yöntem haline gelmiştir. Bollen (2011) tarafından yapılan bir çalışmada, Twitter verilerinden elde edilen duygu skorlarının borsa endeksleri ile %86 korelasyon gösterdiği bulunmuştur. Bu çalışmada, özellikle olumlu duygu değişimlerinin piyasa yükselişlerini öngörmede etkili olduğu rapor edilmiştir. Çalışma, sosyal medya tabanlı duygu analizinin finansal tahminlerde güçlü bir araç olduğunu kanıtlamıştır[2].

Loughran ve McDonald (2011), finansal metinlerde kullanılan kelime türlerinin duyarlılık analizi üzerindeki etkilerini inceleyen bir başka önemli çalışmayı yürütmüşlerdir. Çalışmada, geleneksel duygu sözlüklerinin finansal bağlamda yetersiz kalabileceği belirtilmiş ve finansal terminolojiye özgü kelime listeleri geliştirilmiştir. Bu çalışma, finansal metinlerin analizinde bağlama duyarlı yaklaşımların önemini vurgulamıştır[3].

Xing ve arkadaşları (2018), haber metinlerini kullanarak hisse senedi fiyatlarının kısa vadeli tahminine odaklanan bir çalışma gerçekleştirmiştir. Araştırmada, destek vektör makineleri (SVM) ve latent Dirichlet allocation (LDA) gibi makine öğrenimi teknikleri kullanılarak, haber metinlerinin hisse senedi getirileri üzerindeki etkisi incelenmiştir. Çalışma, haberlerin duygu tonunun piyasa tahminlerinde önemli bir gösterge olabileceğini ortaya koymuştur[4].

Zhang ve Skiena (2010), duygu analizi ve haber başlıklarının etkisini bir araya getirerek, medya duyarlılığı ile hisse senedi fiyatları arasındaki ilişkiyi incelemiştir. Araştırmada, büyük ölçekli haber başlıkları veri seti kullanılmış ve medya duyarlılığı endeksi oluşturulmuştur. Bu endeks, hisse senedi piyasalarındaki dalgalanmaların öngörülmesinde etkili bir araç olarak değerlendirilmiştir. Araştırma, metin tabanlı duygu analizinin pratik finansal uygulamalar için kullanılabilirliğini güçlendiren önemli bulgular sunmuştur[5].

Chen ve arkadaşları (2018), derin öğrenme yöntemlerini finansal duyarlılık analizi için uygulamış ve özellikle LSTM gibi modellerin performansını değerlendirmiştir. Çalışma, finansal metinlerden duygu çıkarımında derin öğrenme tekniklerinin geleneksel yöntemlere göre daha yüksek doğruluk sağladığını göstermiştir. Bu sonuç, doğal dil işleme tekniklerindeki yeniliklerin finansal analiz süreçlerini nasıl dönüştürebileceğine dair güçlü bir örnek sunmaktadır[6].

Transformer tabanlı modellerin finansal duyarlılık analizinde üstün performans sağladığı birçok çalışmada vurgulanmıştır. Örneğin, Araci (2019) tarafından geliştirilen FinBERT modeli, finansal metinlerin duygu analizi için optimize edilmiş bir BERT versiyonudur ve olumlu, olumsuz ve nötr sınıflandırma görevlerinde %90'ın üzerinde bir başarı oranı sağlamıştır. Bu çalışma, finansal metinlerdeki karmaşık dil yapılarının anlaşılmasında derin öğrenme modellerinin gücünü kanıtlamaktadır[7].

Rodriguez Inserte ve arkadaşları (2024), büyük dil modellerinin finansal duygu analizi için uyarlanmasını incelemiştir. Çalışmada, 1,5 milyar parametreden az olan iki temel model, finansal belgeler ve talimatlarla dikkatlice ince ayar yapılarak hedef alana uyarlanmıştır. Sonuçlar, küçük dil modellerinin daha büyük modellerle karşılaştırılabilir performans sergilediğini ve parametreler ile veri açısından daha verimli olduğunu göstermektedir[8].

Asgarov (2023), Twitter duyarlılıklarını kullanarak büyük şirketlerin hisse senedi fiyatlarını tahmin etmeyi amaçlamıştır. Çalışmada, Long-Short Term Memory (LSTM) modeli kullanılarak, tweet'lerdeki duygular ile hisse senedi fiyatları arasındaki ilişki incelenmiştir. Sonuçlar, tweet'lerdeki olumlu ve olumsuz duyguların hisse senedi fiyatları üzerinde önemli bir etkiye sahip olduğunu ve LSTM modelinin bu tahminlerde yüksek doğruluk sağladığını göstermektedir[9].

Bu çalışmalar, doğal dil işleme tekniklerinin finansal piyasa analizindeki kullanım alanlarını geniş bir perspektiften ele alarak, bu alandaki teorik ve uygulamalı çalışmalara zengin bir kaynak sunmaktadır.

# **2 YÖNTEM**

Bu çalışmanın amacı, önceki bölümlerde belirtildiği üzere, doğal dil işleme teknikleri ve duygu analizi yöntemlerini kullanarak finansal piyasalardaki duygu değişimlerini analiz etmek ve bu değişimlerin piyasa hareketleri üzerindeki etkisini incelemektir. Çalışmada, haber metinleri verilerinden duygu analizi yapılmak üzere bir tahmin modeli oluşturulmuştur. Sonrasında, belirlenen model kullanılarak duygu analizi gerçekleştirilmiş ve bu duygu skorlarının fiyat değişimi ile ilişkisi incelenmiştir. Çalışmada, Granger nedensellik testi gibi ileri düzey analiz yöntemleriyle duygu ve fiyat ilişkisi araştırılmıştır.

Bu bölümde, doğal dil işleme teknikleri, nedensellik testi ve genel NLP teknikleri hakkında bilgiler verilmiştir.

## 2.1 Doğal Dil İşleme (NLP)

Doğal dil işleme (Natural Language Processing - NLP), bilgisayarların insan dilini anlamasını, işlemesini ve analiz etmesini sağlayan bir yapay zeka alt disiplinidir. Bu alan, dilbilim, bilgisayar bilimi ve makine öğreniminin kesişiminde yer alır ve dil tabanlı veri kaynaklarının otomatik olarak analiz edilmesine olanak tanır. NLP'nin temel amacı, metin ve konuşma verilerini anlamlı bilgilere dönüştürmektir. Özellikle son yıllarda, büyük veri ve gelişmiş makine öğrenimi algoritmalarının gelişmesiyle NLP, birçok sektörde geniş bir uygulama alanı bulmuştur. Manning ve Schütze (1999), NLP'yi "bilgisayarların dil verileri üzerinde anlamlı işlemler yapmasına olanak tanıyan bir teknoloji ve bilim dalı" olarak tanımlamaktadır[10]. Bu tanım, NLP'nin hem teorik temellerine hem de pratik uygulamalarına ışık tutmaktadır. Finansal piyasalarda NLP'nin kullanımı, yatırımcı duyarlılığı, haber etkileri ve piyasa trendlerini anlamak için yeni bir bakış açısı sunmaktadır.

### **2.1.1 Doğal Dil İşleme Tarihçesi**

Doğal dil işleme çalışmalarının kökleri, 1950'li yıllara kadar uzanmaktadır. Alanın ilk adımları, Claude Shannon’un bilgi teorisi ve Alan Turing’in "Turing Testi" kavramıyla şekillenmiştir. Bu dönemde, bilgisayarların dilsel girdiyi anlaması ve yanıt üretmesi üzerine teorik temeller oluşturulmuştur. Özellikle, 1954 yılında gerçekleştirilen Georgetown-IBM Deneyi, bilgisayar destekli çeviri çalışmalarında önemli bir mihenk taşı olmuştur. Bu deneyde, Rusça'dan İngilizce'ye yapılan otomatik çeviri, NLP'nin pratik uygulamalarını ortaya koymuştur. Ancak bu dönemde dil işleme, sınırlı kurallarla ve temel istatistiksel yöntemlerle yapılmıştır[11].

1980’lerden itibaren, NLP alanında istatistiksel yöntemlerin ve veri odaklı yaklaşımların benimsenmesiyle önemli bir dönüşüm yaşanmıştır. Büyük veri setlerinin kullanılabilir hale gelmesi ve bilgisayarların işlem gücünün artışı, makine öğrenimi algoritmalarının dil işleme süreçlerine entegre edilmesini sağlamıştır. Örneğin, 1990’larda Hidden Markov Models (HMM) ve n-gram modelleri gibi teknikler, dil modelleme ve duygu analizi gibi uygulamalarda büyük başarı sağlamıştır. 2000’li yıllara gelindiğinde, yapay sinir ağları ve derin öğrenme yöntemleri NLP'de devrim niteliğinde bir ilerleme kaydetmiştir. Bu dönemde, word2vec ve transformer tabanlı modellerin (örneğin, BERT ve GPT) geliştirilmesi, NLP’nin doğruluğunu ve verimliliğini önemli ölçüde artırmıştır[12].

### **2.1.2 Doğal Dil İşleme Kullanım Alanları**

Doğal dil işleme, geniş bir uygulama yelpazesiyle birçok sektörde yenilikçi çözümler sunmaktadır. En yaygın kullanım alanlarından biri, müşteri hizmetleri ve destek süreçleridir. Chatbotlar ve sanal asistanlar, kullanıcıların dilini anlayarak sorulara yanıt verebilmekte ve müşteri taleplerine hızlı çözümler sunmaktadır. Örneğin, Siri ve Alexa gibi sanal asistanlar, NLP algoritmalarını kullanarak doğal dildeki komutları anlamakta ve ilgili eylemleri gerçekleştirmektedir. Bu teknolojiler, işletmelerin müşteri deneyimini iyileştirmesine ve maliyetlerini düşürmesine olanak sağlamaktadır.



**Şekil 2.1** NLP kullanım alanları

Bir diğer önemli kullanım alanı, sağlık sektörü olarak öne çıkmaktadır. NLP, sağlık verilerinin analizinde büyük bir rol oynamaktadır. Elektronik sağlık kayıtlarından bilgi çıkarımı, hasta geri bildirimlerinin analizi ve klinik metinlerin sınıflandırılması gibi uygulamalar, sağlık hizmetlerini daha verimli hale getirmektedir. Özellikle, hastalıkların erken teşhisi ve tedavi süreçlerinin iyileştirilmesi için kullanılan NLP tabanlı sistemler, hem hastalar hem de sağlık uzmanları için kritik bir değer sunmaktadır. Ayrıca, klinik araştırmalar için yapılan metin madenciliği çalışmaları, tıbbi literatürden anlamlı bilgiler çıkarılmasını sağlamaktadır.

Finans sektörü, NLP'nin diğer bir önemli kullanım alanıdır. Finansal metinlerin analizi, piyasa duyarlılığını anlamak ve öngörülerde bulunmak için kullanılmaktadır. Haber bültenleri, raporlar ve sosyal medya verileri, yatırımcı davranışlarını etkileyen önemli bilgi kaynaklarıdır. NLP teknikleri, bu metinlerden duygu ve eğilim çıkarımı yaparak, hisse senedi fiyatları ve piyasa hareketleri gibi göstergeleri tahmin etmek için kullanılmaktadır. Bunun yanı sıra, dolandırıcılık tespiti ve müşteri risk analizleri gibi uygulamalar, finansal süreçlerin daha güvenilir ve verimli bir şekilde yönetilmesine katkı sağlamaktadır. NLP’nin sektörel kullanımı, giderek daha geniş bir alana yayılmakta ve günlük yaşamın birçok noktasında etkisini göstermektedir.

### **2.1.3 Doğal Dil İşlemede Karşılaşılan Sorunlar**

Doğal dil işleme uygulamalarında karşılaşılan en büyük sorunlardan biri dilin doğasından kaynaklanan karmaşıklıktır. İnsan dili, bağlama, kültüre ve kişisel tercihlere bağlı olarak değişiklik gösterebilir. Dilin çok anlamlı yapısı, aynı yazılışa sahip farklı anlamdaki kelimeler ve bağlama göre farklı anlamlar taşıyan ifadelerin yorumlanmasını zorlaştırmaktadır. Örneğin, "bank" kelimesi hem finansal bir kurum hem de nehir kıyısı anlamına gelebilir. Bu tür durumlar, NLP algoritmalarının doğru tahmin yapmasını zorlaştırır. Ayrıca, ironi ve mecaz gibi duygusal ve bağlamsal unsurların analizi, NLP modelleri için önemli bir zorluk teşkil etmektedir.

**A diagram of different colored gears

Description automatically generated**

**Şekil 2.2** NLP’de karşılaşılan sorunlar[13]

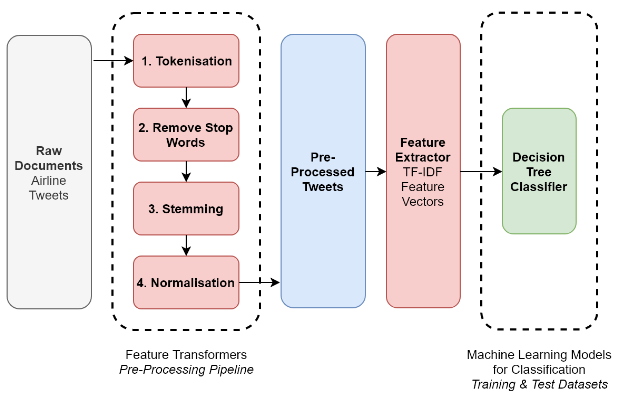
Bir diğer önemli sorun, büyük ve çeşitlilik içeren veri setlerine olan ihtiyaçtır. NLP modellerinin doğru ve etkili bir şekilde çalışabilmesi için, dilin farklı yönlerini kapsayan büyük veri setlerine ihtiyaç vardır. Ancak, özellikle nadir diller veya teknik terimlerin yoğun olduğu alanlar için yeterli veri bulunamayabilir. Veri yetersizliği, modellerin düşük performans göstermesine veya yalnızca belirli bağlamlarla sınırlı kalmasına neden olabilir. Ayrıca, etik kaygılar da önemli bir sorundur; modeller, önyargılı veri setleriyle eğitildiklerinde, sonuçlar ayrımcı veya hatalı olabilir. Örneğin, toplumsal cinsiyet veya ırksal önyargılar, NLP tabanlı sistemlerin yanlış ve zararlı sonuçlar üretmesine yol açabilir. Bu sorunların çözümü, daha kapsayıcı ve dikkatli bir veri toplama ve modelleme sürecini gerektirmektedir.

### **2.1.4 Doğal Dil İşleme Süreci**

Doğal dil işleme, ham dil verilerinin anlamlı bilgilere dönüştürülmesini sağlayan bir dizi aşamadan oluşmaktadır. Bu süreçler ihtiyaçlar doğrultusunda farklı sıralama ile ilerleyebilmektedir. Genel olarak ilk aşama, metinlerin veri kaynağından toplanması ve temizlenmesidir. Bu süreçte, metinlerdeki gereksiz karakterler, boşluklar, durak kelimeler ve analiz için anlam ifade etmeyen diğer unsurlar çıkarılır. Aynı zamanda, eksik veriler tamamlanır ya da çıkarılır, metinler standart bir formata dönüştürülür. Bu adım, verilerin sonraki aşamalarda doğru bir şekilde analiz edilmesi için gerekli olan temel hazırlık aşamasını oluşturur.

Metin temizliğinden sonra, metinlerde kelimelerin köklerine indirgenmesi amacıyla stemming veya lemmatization işlemleri gerçekleştirilir. Stemming, kelimelerin kök formuna indirgenmesi için basit algoritmalar kullanırken, lemmatization daha gelişmiş bir yaklaşımla dilbilgisel analiz yaparak kelimenin sözlükteki gerçek formunu bulur. Daha sonra tokenization adı verilen bir işlem gerçekleştirilir. Bu adımda, metin belirli birimlere, örneğin kelimelere veya cümlelere ayrılır ve analiz edilebilir hale getirilir. Bu süreç, metni yapılandırılmış bir formata dönüştürmek için kritik bir adımdır.

Bir sonraki aşamada, metinlerdeki dilsel özellikler belirlenir. Bu, kelime türlerinin (isim, sıfat, fiil vb) tanımlandığı POS tagging (kelime etiketleme), cümlelerin yapısal analizi için sentetik analiz ve ad öbeklerinin belirlenmesi gibi tekniklerle gerçekleştirilir. Bu aşama, metnin anlamını daha derinlemesine anlamak için kullanılır ve metinlerin bağlamına uygun bir şekilde analiz edilmesine olanak tanır.

****

**Şekil 2.3** Doğal dil işleme sürecine bir örnek[14]

Son aşama ise spesifik NLP görevlerinin gerçekleştirilmesidir. Bu aşamada duygu analizi, metin sınıflandırma, konu modelleme gibi görevler uygulanır. Bu tür analizler genellikle makine öğrenimi veya derin öğrenme algoritmaları ile gerçekleştirilir. Elde edilen modeller, doğruluk, hassasiyet ve genelleme kapasitesi gibi metriklerle değerlendirilir. Bu sürecin her aşaması, doğal dil işlemenin farklı uygulamalarında kullanılan güçlü bir temel oluşturmaktadır. Ancak farklı uygulamalar, farklı teknikler gerektirebilmektedir.

## 2.2 Doğal Dil İşleme Teknikleri

Doğal dil işleme, çeşitli teknikler kullanılarak metin tabanlı verilerin işlenmesini ve analizini sağlar. Bu teknikler, verinin doğasına ve analizin amacına göre farklılık gösterebilir. Genel olarak kullanılan NLP teknikleri alt başlıklarda açıklanmaktadır.

### **2.2.1 Metin Ön İşleme**

Ham metin verilerinin doğal dil işleme yöntemleriyle analiz edilmesi için belirli adımlar doğrultusunda hazırlanması gereklidir. Bu adımlar, metnin yapısal olarak standartlaştırılması ve daha sonraki analiz aşamaları için anlamlı bir forma dönüştürülmesini sağlar. Alt başlıklarda, bu süreçte kullanılan temel teknikler açıklanmıştır.

#### **2.2.1.1 Tokenizasyon**

Tokenizasyon, metinleri daha küçük birimlere, yani kelimelere, cümlelere veya ifadelere ayırma işlemidir. Bu adım, metni analiz edilebilir hale getirmek için atılan ilk adımdır. Örneğin, "The market is volatile today." cümlesi tokenization işlemi sonucunda ["The", "market", "is", "volatile", "today"] şeklinde ayrı kelimelere bölünür.

Matematiksel olarak tokenizasyon, metni belirli bir ayırıcıya göre parçalara ayırır:

Burada metni; , ayırıcı karakteri ifade eder (genellikle boşluk veya noktalama işaretleri). Kelime bazlı tokenizasyon, özellikle kelime sıklığı analizi ve duygu analizi gibi uygulamalarda kritik bir rol oynar.

#### **2.2.1.2 Lemmatizasyon ve Stemming**

Lemmatizasyon, kelimeleri dilbilgisel bağlamlarına göre köklerine indirgerken, kök bulma (stemming) daha basit bir işlemle kelimenin son eklerini keserek kökünü elde eder. Örneğin, "running", "ran" ve "runs" kelimeleri lemmatizasyon ile "run" köküne indirgenir.

Lemmatizasyon, bağlamsal bir işlem olduğu için kelimenin ait olduğu dilbilgisel kategoriye göre çalışır. Matematiksel olarak şu şekilde ifade edilebilir:

Burada kelimeyi, kelimenin temel halini, ise kelimenin dilbilgisel kategorisini ifade eder. Bu işlem, bağlam analizini güçlendirmek ve veri setindeki kelime çeşitliliğini azaltmak için önemlidir.

#### **2.2.1.3 Durak Kelimelerin (Stopwords) Çıkarılması**

Durak Kelimeler (stopwords), metinlerde anlam taşımayan ve sıklıkla kullanılan kelimelerdir (örneğin, "ve", "bu", "o"). Bu kelimelerin analizden çıkarılması, metnin daha anlamlı hale gelmesini ve analiz doğruluğunun artmasını sağlar. Durak kelime çıkarma işlemi şu şekilde ifade edilir:

Burada , metindeki tüm kelimelerin kümesini; , durak kelimelerin kümesini; ise durak kelimeler çıkarıldıktan sonra kalan kelimeleri ifade eder. Bu işlem, modelin yalnızca bilgi taşıyan kelimelere odaklanmasını sağlar ve işlem yükünü azaltır.

#### **2.2.1.4 Küçük Harfe Çevirme**

Metinlerdeki harflerin küçük harfe dönüştürülmesi, veri standardizasyonunun önemli bir parçasıdır. Örneğin, "Price" ve "price" gibi farklı yazımlar aynı kelimeyi ifade eder, ancak analiz sürecinde farklı olarak algılanabilir. Bu işlem, şu şekilde tanımlanabilir:

Burada , metindeki kelimeler kümesini; , her bir kelimenin küçük harfe dönüştürülmesini ifade eder. Bu işlem, metindeki anlamsal birliği artırır ve modelin hata oranını düşürür.

#### **2.2.1.5 Noktalama İşaretlerini Kaldırma**

Noktalama işaretlerinin kaldırılması, metni gereksiz semboller ve özel karakterlerden arındırarak temiz bir veri seti oluşturur. Bu işlem, yalnızca kelimelere odaklanarak analiz sürecini sadeleştirir. Matematiksel olarak:

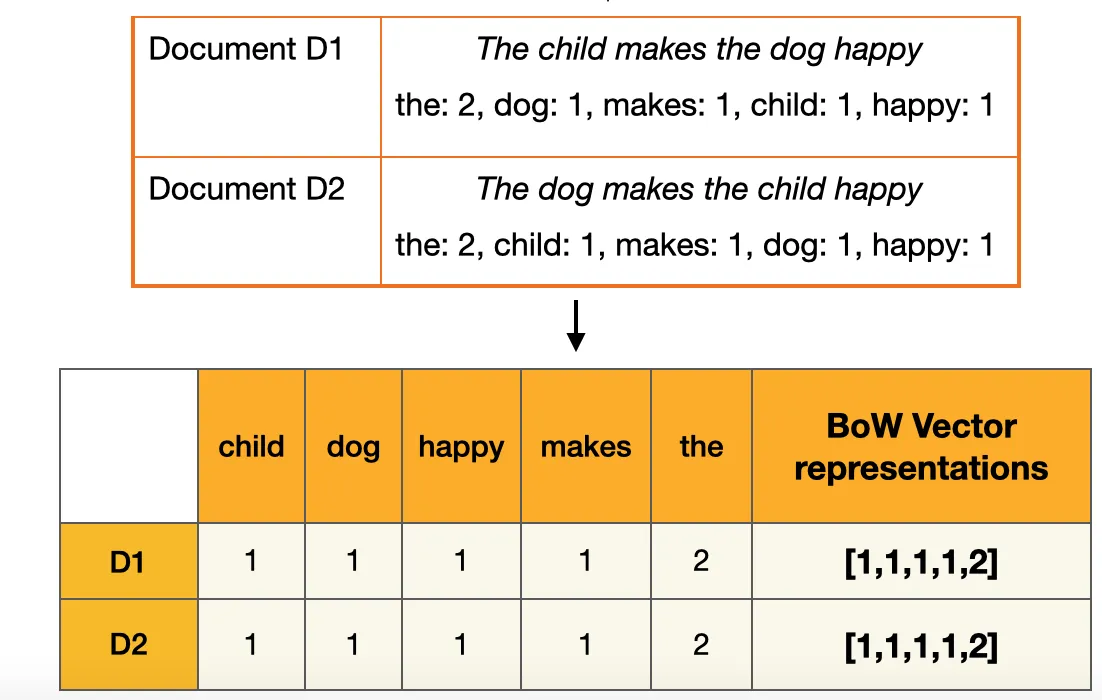
Burada , metindeki tüm kelimeler kümesini; , noktalama işaretleri kümesini; , noktalama işaretleri çıkarıldıktan sonra kalan kelimeler kümesini ifade eder. Örneğin, "Google’s stock price increased!" ifadesi, bu işlem sonucunda "Google stock price increased" haline gelir.

### **2.2.2 Özellik Çıkarımı**

Ham metinlerin analiz edilebilir bir formata dönüştürülmesi, doğal dil işleme süreçlerinin temel gerekliliklerinden biridir. Bu süreçte metin verileri, makine öğrenimi algoritmaları için sayısal temsillere dönüştürülür. Özellik çıkarımı aşamasında kullanılan teknikler, kelimelerin veya cümlelerin temsillerini anlamlı hale getirerek analiz doğruluğunu artırmayı hedefler. Alt başlıklarda, bu aşamada kullanılan yöntemler detaylı bir şekilde açıklanmıştır.

#### **2.2.2.1 Kelime Torbası**

Kelime Torbası (Bag of Words), metindeki her bir kelimenin sıklığını dikkate alarak metni sayısal bir vektör haline dönüştüren temel bir tekniktir. Bu yöntem, kelime sırasını göz ardı eder ve yalnızca kelimenin dokümanda kaç kez geçtiğini kaydeder. Örneğin, Şekil 2.4’teki cümleler, aynı kelimelere sahip olduğu için benzer vektörlere sahiptir.

****

**Şekil 2.4** Bag of Words için bir örnek[15]

Matematiksel olarak, bir metni kelime torbası modeliyle temsil etmek için şu işlem gerçekleştirilir:

Burada , dokümanı; , belirli bir kelimeyi; , kelimesinin dokümanındaki frekansını ifade eder.

#### **2.2.2.2 TF-IDF (Term Frequency-Inverse Document Frequency)**

TF-IDF (Terim Frekansı-Ters Doküman Frekansı), kelimelerin bir dokümandaki önemini ölçmek için kullanılan istatistiksel bir yöntemdir. Kelimenin doküman içerisindeki sıklığını ve tüm doküman kümesindeki nadirliğini dikkate alır. TF-IDF, sık kullanılan ancak anlam taşımayan kelimelerin etkisini azaltırken nadir ve bilgi taşıyan kelimeleri ön plana çıkarır. Özellikle finansal haberlerde sıkça geçen "market", "price" gibi genel terimlerin etkisini azaltarak, nadir ama bilgi taşıyan terimleri ön plana çıkarır.

Matematiksel olarak TF-IDF şu şekilde hesaplanır:

Terim Frekansı (TF), bir kelimenin bir dokümanda geçiş sıklığı:

Burada , kelimesinin dokümanındaki frekansını, ise dokümandaki toplam kelime sayısını ifade eder.

Ters Doküman Frekansı (IDF), bir kelimenin tüm dokümanlar arasındaki nadirliğini ölçer:

Burada , toplam doküman sayısını;, kelimesini içeren doküman sayısını ifade eder, 1 eklenmesi, bir kelimenin hiçbir belgede geçmeme durumunda sıfır bölme hatasını önler.

**A table with numbers and letters

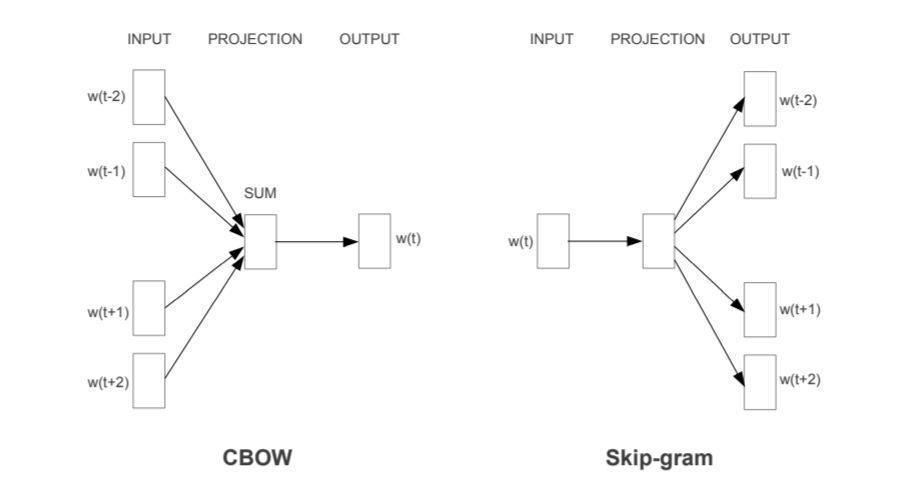
Description automatically generated**

**Şekil 2.5** TF-IDF için örnek bir işlem[16]

#### **2.2.2.3 Kelime Gömme**

Kelime gömme (Word Embeddings), kelimeleri sürekli bir vektör uzayında temsil eden yöntemlerdir. Bu yöntemler, kelimeler arasındaki anlamsal ilişkileri matematiksel olarak ifade eder. Örneğin, Word2Vec ve GloVe gibi algoritmalar, "price" ve "value" gibi kelimeleri benzer vektörlerle temsil ederek, bu kelimelerin semantik yakınlığını gösterir. Bu teknik, finansal metinlerde sıkça geçen kelimeler arasındaki anlamsal ilişkileri çıkararak, duygu analizi ve konu modelleme gibi süreçlere önemli katkılar sağlamaktadır.

Kelime gömme yöntemlerinden biri olan Word2Vec, iki temel modelden oluşur: CBOW (Continuous Bag of Words), çevredeki kelimelere bakarak eksik bir kelimeyi tahmin etmeye çalışırken; Skip-Gram, bir kelimeye bakarak çevresindeki kelimeleri tahmin etmeye odaklanır.

****

**Şekil 2.6** CBOW ve Skip-gram örnek[17]

### **2.2.3 Metin Sınıflandırma**

Metin sınıflandırmasının iki ana türü vardır: denetimli (supervised) ve denetimsiz (unsupervised). Denetimli metin sınıflandırma, etiketlerin önceden bilindiği bir veri seti üzerinde bir modelin eğitilmesini içerir. Denetimsiz metin sınıflandırma ise etiket gerektirmez; bunun yerine, model veriler üzerinde eğitilir ve belgeleri benzerliklere dayanarak kategorilere ayırmayı öğrenir.

Denetimli metin sınıflandırma, denetimsiz metin sınıflandırmadan daha doğrudur, ancak daha zaman alıcı ve maliyetlidir, çünkü etiketlenmiş verilere ihtiyaç duyar. Denetimsiz metin sınıflandırma ise daha az doğru olabilir, ancak etiketlerin mevcut olmadığı durumlarda kullanılabilir.

Hem denetimli hem de denetimsiz metin sınıflandırma, duygu analizi, konu tespiti, spam tespiti gibi çeşitli görevler için kullanılabilir. Algoritma seçimi, ele alınan göreve bağlıdır. Örneğin, eğer doğruluk çok önemliyse, denetimli bir yaklaşım tercih edilir. Ancak zaman veya kaynak kısıtlamaları varsa, denetimsiz bir yaklaşım daha uygun olabilir[18].

### **2.2.4 Varlık Adı Tanıma**

Varlık adı tanıma (Named Entity Recognition), metinlerdeki belirli isimleri, organizasyonları, yerleri, tarihleri ve diğer önemli varlıkları tanımlamak için kullanılan bir doğal dil işleme yöntemidir. Bu yöntem, metinlerin anlamını derinlemesine analiz ederek, kritik bilgi noktalarını ön plana çıkarmaktadır. Özellikle finansal haberlerde, NER kullanılarak şirket isimleri, hisse senetleri, tarihler ve ekonomik olaylar gibi önemli unsurlar tespit edilebilmektedir. Örneğin, "Tesla reported a record profit in New York on August 15" ifadesinde; "Tesla" bir şirket adı olarak, "New York" toplantı yeri olarak, "August 15" bir tarih (raporun yayınlandığı gün) olarak etiketlenmektedir.

Bu sınıflandırma, genellikle denetimli öğrenme algoritmaları veya derin öğrenme modelleriyle gerçekleştirilir. Model, daha önce etiketlenmiş veri setlerini kullanarak eğitilir ve yeni metinlerde varlıkları tanımayı öğrenmektedir. Ayrıca, BERT gibi bağlam tabanlı modeller, kelimelerin çevresindeki kelimelere bağlı olarak doğru sınıflandırma yapmayı sağlamaktadır.

### **2.2.5 Dil Modellemesi**

Dil modellemesi, bir cümledeki kelimeler arasındaki ilişkileri öğrenerek bir sonraki kelimeyi tahmin etmeyi amaçlayan bir NLP tekniğidir. N-gramlar, bir cümledeki önceki kelimeye dayanarak 'inci kelimeyi tahmin etmekedir. Örneğin, "The market is" ifadesinden sonra "volatile" kelimesinin gelme olasılığını belirlemektedir. Bu yöntem, kısa mesafeli kelime ilişkilerini anlamada etkili ve hesaplama açısından hızlıdır ancak uzun mesafeli bağlamları anlamakta yetersizdir.

Sinirsel dil modelleri, derin öğrenme yöntemleriyle dilin yapısını anlamaya yönelik daha gelişmiş bir tekniktir. BERT gibi modeller, bir kelimenin hem önceki hem de sonraki kelimelerle olan bağlamını anlamada başarılıdır. Örneğin, "The stock rose sharply yesterday" cümlesinde, "rose" kelimesinin anlamını çevresindeki kelimelerle birlikte değerlendirir. Bu modeller, uzun mesafeli ilişkileri anlamak ve bağlamı çözümlemek için idealdir.

Finansal haberlerde, dil modellemesi metinlerin analizinde ve anlamlandırılmasında kullanılır. Örneğin, bir finansal haberin eksik cümlelerini tamamlamak veya bir haberin teması doğrultusunda metin üretmek için uygulanabilmektedir. Bu, yatırım stratejileri veya piyasa analizlerini desteklemek için güçlü bir araç sunmaktadır.

### **2.2.6 Makine Çevirisi**

Makine çevirisi, bir metni bir dilden başka bir dile otomatik olarak çeviren bir NLP uygulamasıdır. Bu süreç, kelimeler arasındaki bağlamı ve dil bilgisel yapıları anlamayı gerektirir. Modern makine çeviri sistemleri, Nöral Makine Çevirisi (Neural Machine Translation) tekniklerini kullanır. Örneğin, Google Translate[19] gibi araçlar, Transformer tabanlı modellerle bir cümlenin bağlamını dikkate alarak yüksek doğrulukta çeviriler sağlar. Finansal haberlerde makine çevirisi, farklı dillerdeki piyasa verilerini analiz etmek için kullanılabilir ve yatırımcıların küresel ölçekte bilgiye erişimini kolaylaştırır.

### **2.2.7 Metin Üretimi**

Metin üretimi, insan benzeri metinlerin otomatik olarak oluşturulmasını sağlayan bir NLP tekniğidir. Bu işlem, GPT (Generative Pre-trained Transformer) gibi modellerle gerçekleştirilir. Örneğin, bir finansal rapor özetinin veya piyasa tahminiyle ilgili metinlerin üretilmesi bu teknikle mümkündür. Bu modeller, girdiye dayalı olarak mantıksal ve bağlam açısından anlamlı metinler üretir. Finansal analizlerde metin üretimi, piyasa tahmin raporları veya haber özeti gibi süreçlerde zaman kazandırır ve etkili bir şekilde bilgi sumaktadır.

### **2.2.8 Duygu Analizi**

Duygu analizi, metinlerdeki duygusal tonun belirlenmesi için kullanılan bir doğal dil işleme tekniğidir[20]. Bu yöntem, metinleri pozitif, negatif veya nötr kategorilere ayırır ve özellikle sosyal medya, ürün incelemeleri ve finansal haberlerde yaygın olarak kullanılır. Finansal haberlerde duygu analizi, piyasa duyarlılığını anlamak, yatırım stratejilerini şekillendirmek ve piyasa trendlerini tahmin etmek için kritik bir rol oynar.

#### **2.2.8.1 Kural Tabanlı Modeller**

Kural tabanlı modeller, metinlerdeki kelimelere veya ifadelere belirli kurallar ve sözlükler aracılığıyla duygu atayan basit ama etkili bir duygu analizi yöntemidir. Bu modeller, önceden tanımlanmış bir duygu sözlüğü veya kelime listesi kullanır ve metinlerdeki kelimeleri bu sözlükteki pozitif veya negatif skorlarına göre değerlendirir. Genellikle, bir metnin genel duygusal tonu, sözlükteki kelimelerin toplam skoruna veya ağırlıklı ortalamasına göre belirlenir.

#### **2.2.8.2 Makine Öğrenimi Tabanlı Modeller**

Makine öğrenimi modelleri, etiketlenmiş veri setleri üzerinde eğitilerek metinlerin duygusal tonlarını veya kategorilerini anlamlandıran algoritmalardır. Bu modeller, metinleri sayısal temsillere dönüştürüp sınıflandırma yapar. Makine öğrenimi tabanlı modellerin avantajı, esnek ve özelleştirilebilir olmalarıdır; belirli bir bağlama veya uygulama alanına özel veri setleriyle eğitildiklerinde yüksek doğruluk sağlayabilirler.

#### **2.2.8.3 Önceden Eğitilmiş Modeller**

Önceden eğitilmiş modeller, büyük ölçekli veri setleri üzerinde önceden eğitilmiş ve belirli bir bağlamda güçlü sonuçlar verebilen modellerdir. Bu modeller, kelimelerin bağlamını anlamak için modern derin öğrenme yöntemlerini kullanır ve birçok farklı alanda etkili bir şekilde uygulanabilir. Örneğin, FinBERT, finansal metinlere özgü olarak eğitilmiş bir modeldir ve "bull market" veya "earnings report" gibi ifadelerin finansal bağlamdaki duygusal tonlarını doğru bir şekilde analiz edebilir. VADER ise sosyal medya metinleri gibi kısa ve yoğun içeriklerde başarılıdır. Önceden eğitilmiş modellerin en büyük avantajı, kullanıcıların sıfırdan bir model eğitmek zorunda kalmadan, doğrudan yüksek doğrulukta analiz yapmalarına olanak tanımasıdır. Bu modeller, bağlamsal anlamı güçlü bir şekilde yakalayarak, finansal haberler gibi spesifik alanlarda zamandan tasarruf sağlar ve güvenilir sonuçlar üretir.

**Tablo 2.1** Modellerin karşılaştırılması

|  |  |  |  |
| --- | --- | --- | --- |
| Özellik | Kural Tabanlı Modeller | Makine Öğrenmesi Modelleri | Önceden Eğitilmiş Modeller |
| Bağlam Anlama | Zayıf | Güçlü | Çok güçlü |
| Dil Karmaşıklığı | İroni, olumsuzluk gibi unsurları anlamakta zorlanır | Karmaşık dil yapılarını çözebilir | İroni, deyimler ve karmaşık bağlamları iyi anlar |
| Hız | Daha hızlı | Büyük veride daha yavaş | Orta hızda |
| Özelleştirilebilirlik | Kolayca özelleştirilebilir | Daha fazla eğitim ve optimize gerektirir | Önceden eğitildiği için özelleştirme kısıtlı olabilir |
| Doğruluk | Basit bağlamlarda yeterli | Karmaşık bağlamlarda üstün | Özel bağlamlarda çok yüksek doğruluk sağlar |
| Kullanım Kolaylığı | Kolay | Veri seti ve özellik mühendisliği gerektirir | Veri seti olmadan doğrudan kullanılabilir |
| Eğitim Gereksinimi | Gerektirmez | Gerektirir | Gerekmez |

## 2.3 Granger Nedensellik Testi

Granger nedensellik testi, bir zaman serisinin, olayların kronolojik sırasına dayanarak başka bir zaman serisini tahmin etmeye yardımcı olup olmadığını değerlendiren istatistiksel bir yöntemdir ve nedenin sonuçtan önce geldiğini vurgular[21]. Bu test, iki zaman serisi arasında doğrusal bir ilişki olup olmadığını araştırır ve "nedensellik" kavramını tahmin edilebilirlik bağlamında ele alır. Burada nedensellik, bir değişkenin (​) diğer bir değişkenin (​) gelecekteki değerlerini tahmin etmede katkıda bulunması anlamına gelir. Örneğin, ekonomik analizlerde, faiz oranlarının enflasyonu tahmin edip edemeyeceğini belirlemek için bu test uygulanabilir.

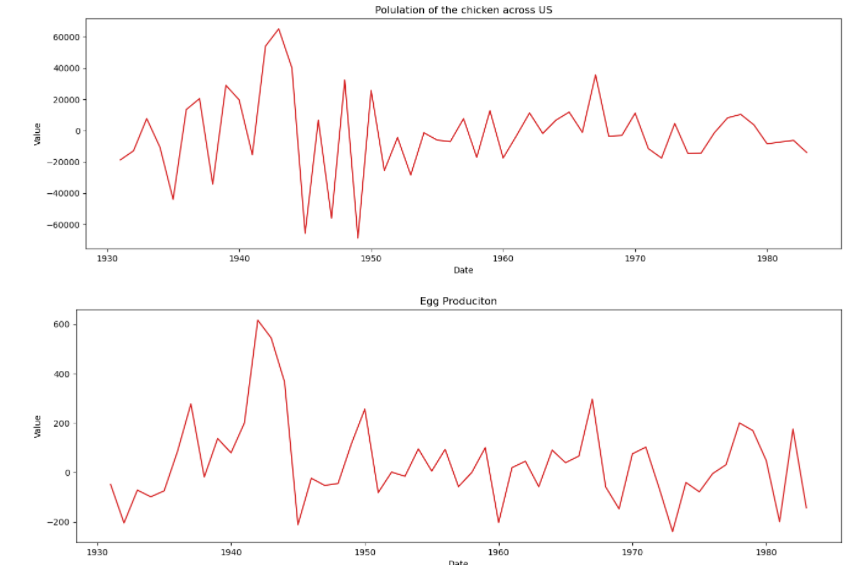
Testin temel prensibi, iki zaman serisi arasında bir "Granger nedensellik" ilişkisi olup olmadığını anlamaktır. Eğer ​'nin geçmiş değerleri, ​'yi, yalnızca ​'nin kendi geçmiş değerlerine dayanarak yapılacak tahminlerden daha iyi bir şekilde tahmin etmemizi sağlıyorsa, bu durumda , 'ye Granger nedenidir. Ancak bu, klasik anlamda bir sebep-sonuç ilişkisini değil, yalnızca istatistiksel bir bağı ifade eder. Bu nedenle, Granger nedensellik testi daha çok bir tahmin gücünü değerlendirme yöntemi olarak kabul edilir. Finansal piyasalarda, döviz kurları ile hisse senedi fiyatları arasındaki ilişkileri anlamak veya haber duyarlılığı ile piyasa dalgalanmaları arasındaki bağları analiz etmek için sıkça kullanılmaktadır.

### **2.3.1 Verinin Hazırlanması**

Öncelikle, testin uygulanacağı iki zaman serisi ( ve ​) düzenlenmelidir. Veriler aynı zaman diliminde ve eşit aralıklara sahip olmalıdır (örneğin, dakikalık, saatlik, günlük fiyatlar). Eksik veriler varsa, bunlar doldurulmalı veya çıkarılmalıdır. Ayrıca, testin doğruluğu için serilerin durağan olup olmadığı kontrol edilmelidir. Durağanlık, zaman serisinin ortalama ve varyansının zaman içinde değişmemesi anlamına gelir.

### **2.3.2 Durağanlık Testi**

Granger nedensellik testinin güvenilir sonuçlar üretebilmesi için analiz edilen zaman serilerinin durağan olması gerekmektedir. Durağanlık, bir zaman serisinin ortalama, varyans ve kovaryansının zaman içerisinde değişmemesi anlamına gelir. Bu koşulu sağlamak için, seriler üzerinde birim kök testleri uygulanır. Yaygın olarak kullanılan testlerden Augmented Dickey-Fuller (ADF) testi[22], serilerin durağan olup olmadığını belirlemek için tercih edilmektedir. Eğer seriler durağan değilse, fark alma () işlemi uygulanarak seriler durağan hale getirilmektedir. Bu adım, Granger testinin sağlıklı sonuçlar üretebilmesi için kritik bir ön koşuldur.

****

**Şekil 2.7** Durağan iki veri setinin grafiği[23]

### **2.3.3 Gecikme Sayısının Belirlenmesi**

Granger testi için hangi geçmiş değerlerin (gecikmelerin) analize dahil edileceği belirlenmelidir. Bunun için Akaike Bilgi Kriteri (AIC) veya Bayesian Bilgi Kriteri (BIC) gibi yöntemler kullanılarak optimal gecikme sayısı seçilebilir. Örneğin, ​'nin ​'ye etkisini analiz ederken gecikmesi kullanılıyorsa, aşağıdaki denklemler oluşturulur:

Model 1 (sadece geçmişi, temel model):

Model 2 (ve geçmişi, genişletilmiş model):

Burada; ​, bağımlı değişkenin mevcut değeri; ,’nin ’inci geçmiş değeri; ve​ regresyon katsayıları;, ’nin ’inci geçmiş değeri; , ​’nin katsayıları; hata terimi.

### **2.3.4 Hipotez Testi**

Granger nedensellik testi, iki zaman serisi arasındaki nedensellik ilişkisini anlamak için bir hipotez testi çerçevesinde uygulanır. Test, aşağıdaki iki temel hipotez üzerinden gerçekleştirilir.

**Tablo 2.2** Hipotez testi

|  |  |
| --- | --- |
| Null Hipotezi (​) | Alternatif Hipotez (​) |
| Birinci zaman serisi (​) ikinci zaman serisine ( ​) Granger neden değildir. Bu durumda, ’nin geçmiş değerlerinin ​’nin gelecekteki değerlerini tahmin etmede ek bir bilgi sağlamadığı kabul edilir. | Birinci zaman serisi () ikinci zaman serisine (​) Granger nedenidir. Bu durumda, ​’nin geçmiş değerlerinden elde edilen bilgilerin,’nin gelecekteki değerlerini tahmin etmede istatistiksel olarak anlamlı bir katkı sağladığı varsayılır. |

Hipotez testi, genellikle F-istatistiği kullanılarak gerçekleştirilir. Eğer F-testinden elde edilen -değeri, belirlenen anlamlılık seviyesinden () küçükse, null hipotez reddedilir ve ​’nin ​’ye Granger nedeni olduğu sonucuna varılır. Aksi takdirde, null hipotez kabul edilir ve ’nin ​’nin gelecekteki değerlerini tahmin etmekte anlamlı bir etkisi olmadığı kabul edilir.

F-istatistiği şu şekilde hesaplanır:

​

Burada , Model 1’in hata kareler toplamı; , Model 2’nin hata kareler toplamı; , gecikme sayısı; , gözlem sayısı; , Model 2’nin toplam parametre sayısıdır.

### **2.3.5 Sonuçların Yorumu**

Granger nedensellik testi sonuçları, iki zaman serisi arasındaki ilişkinin yönü ve gücü hakkında önemli bilgiler sağlar. Eğer , ​’ye Granger nedeniyse, ​’nin geçmiş değerleri, ’nin gelecekteki değerlerini tahmin etmekte anlamlı bir katkı sağlar. Bu durum, ​’nin üzerinde güçlü bir nedensellik ilişkisi oluşturduğu şeklinde yorumlanır.

Eğer hem ​’nin ​’ye hem de ’nin ​’ye Granger nedeni olduğu sonucuna varılırsa, iki yönlü bir ilişki söz konusu olabilir. Bu durumda, iki değişkenin birbirini etkilediği kabul edilir. Örneğin, haber duyarlılığı ile piyasa fiyatları arasında iki yönlü bir Granger nedensellik ilişkisi, haberlerin piyasa üzerinde etkili olduğu kadar, piyasa hareketlerinin de haberlerin tonunu etkileyebileceğini gösterebilir. Bu tür bulgular, yatırım stratejileri veya ekonomik politika analizi gibi uygulamalarda önemli çıkarımlar sağlar.

# **3 UYGULAMA**

Bu bölümde, finansal piyasalarda duygu analizinin etkisini anlamak ve zaman serileri arasındaki nedensellik ilişkilerini belirlemek amacıyla gerçekleştirilen analizler açıklanmıştır. Çalışmada, finansal haber metinlerinden duygu skorları çıkarılmış ve bu skorlar ile piyasa hareketleri arasındaki ilişkiler Granger nedensellik testi ve grafik kullanılarak incelenmiştir.

Uygulama sürecinde Python[24] yazılım dili kullanılmıştır. Veri toplama ve ön işleme aşamalarından başlanarak, duygu analizi ve zaman serisi analizine kadar olan adımlar detaylı bir şekilde ele alınmıştır.

## 3.1 Veri Setinin Hazırlanması

Veri seti[25], Bitcoin ile ilgili haber metinlerinden oluşmaktadır. 2021-12-27 ile 2023-12-19 tarihleri de dahil olmak üzere bu tarihler arasındaki haberlerden derlenmiştir ve 9842 adet haber içermektedir. Kullanılan veri seti Kaggle[26] platformundan elde edilmiştir, burada veri setinde bulunan duygu skorları kolonu ve bazı tarih aralıkları veri setinde gürültü oluşturmaması için projeye dahil edilmeden işlemler yapılmıştır.

Web kazıma ile veri seti elde etme işlemi zaman maliyeti, sunucu ve proxy gibi altyapı ihtiyaçları, özellikle büyük ölçekli bir kazıma projesinde maliyet yaratabildiğinden tercih edilmemiştir. Bu nedenle hazır veri seti kullanılmıştır.

A screenshot of a black screen

Description automatically generated

**Şekil 3.1** Veri setinin ilk 5 haberi

## 3.2 Metin Ön İşleme

Bu süreçte, ham metinlerde yer alan gürültüler temizlenmiş ve duygu analizi için standart bir yapıya dönüştürülmüştür. Bu adımlar, NLP yöntemlerinin daha etkili uygulanabilmesi ve metinlerdeki anlamlı bilgilerden yararlanılabilmesi için kritik bir ön işleme sürecini oluşturmuştur.

İlk olarak, metinlerden noktalama işaretleri, özel karakterler ve sayılar kaldırılmıştır. Ardından, metin içerisindeki tüm harfler küçük harflere dönüştürülerek dil bilgisel farklılıkların etkisi azaltılmış ve metinler analiz için tutarlı bir hale getirilmiştir.

A screenshot of a computer

Description automatically generated

**Şekil 3.2** Temizlenmiş ver seti

Metin ön işleme sürecinin bir sonraki adımında, analiz için anlam taşımayan ve sıkça kullanılan kelimeler, yani stopwords (Şekil 3.3), metinlerden çıkarılmıştır. Bu adım, metinlerin daha anlamlı hale getirilmesini ve modelin yalnızca bilgi taşıyan kelimeler üzerinde çalışmasını sağlamak amacıyla gerçekleştirilmiştir. Bu işlem, kelime frekansı ve bağlam analizi sırasında önemli bilgilerin öne çıkarılmasını ve analiz doğruluğunun artırılmasını sağlamıştır.

A black screen with white text

Description automatically generated

**Şekil 3.3** Duraksama kelimeleri kaldırılmış veri seti

Ön işleme sürecinin bir diğer adımı olarak, metinler tokenizasyon (Şekil 3.4) işlemine tabi tutulmuştur. Bu adımda, haber metinleri kelime veya cümle düzeyinde küçük birimlere ayrılmıştır. Kelime bazlı tokenizasyon, metinlerin her bir kelimesinin bağımsız bir analiz birimi olarak ele alınmasını sağlamıştır. Tokenizasyon, metinlerin daha sistematik bir şekilde işlenmesine olanak tanımıştır.

A screenshot of a computer

Description automatically generated

**Şekil 3.4** Tokenize edilmiş veri seti

Metin ön işleme sürecinin bir sonraki adımında, kelimelerin kök formlarına indirgenmesi için lemmatizasyon (Şekil 3.5) işlemi gerçekleştirilmiştir. Bu işlem, kelimelerin bağlamına bağlı olarak doğru dilbilgisel formunu belirlemeyi ve analizde dilbilgisel çeşitliliğin etkisini azaltmayı hedeflemiştir. Bu işlem, özellikle aynı anlamı taşıyan ancak farklı biçimlerde kullanılan kelimelerin analizlerde bir bütünlük sağlaması için kritik bir adım olmuştur. Böylelikle, metinlerin daha sade ve anlamlı bir yapıya kavuşması sağlanmış ve analizlerin doğruluğu artırılmıştır.

A screenshot of a computer

Description automatically generated

**Şekil 3.5** Lemmatize edilmiş veri seti

## 3.3 Duygu Etiketleme

Metin ön işleme adımlarının ardından, haber metinlerine duygu etiketleri atanması işlemi gerçekleştirilmiştir. Bu adımda, finansal metinler için özel olarak geliştirilmiş bir model olan FinBERT[27] kullanılmıştır. FinBERT modeli ile metinlerdeki bağlamsal anlamı dikkate alarak her bir haberin pozitif, negatif ve nötr olma olasılıkları hesaplanmıştır, bu doğrultuda haberlere -1 ve 1 aralığında duygu skoru ataması yapılmıştır.

A computer code on a black background

Description automatically generated

**Şekil 3.6** Duygu skoru hesaplanması için fonksiyon kodu

Model, finansal haberlerde sıkça geçen "profit", "loss", "growth" gibi terimlerin duygusal tonlarını doğru bir şekilde analiz edebilme yeteneğine sahiptir. FinBERT'in kullanımı, finansal haberlerin piyasa duyarlılığı üzerindeki etkisinin ölçülmesinde yüksek doğruluk sağlamış ve sonraki analizler için kritik bir veri seti oluşturmuştur.

A screenshot of a black screen

Description automatically generated

**Şekil 3.7** Duygu skorları ve ilgili haber tarihleri

Şekil 3.7’de bulunan Sentiment\_Score, metnin duygusal tonunu ifade etmektedir: 1’e yaklaştıkça pozitif, -1’e yaklaştıkça negatif, ve 0’a yakın değerler nötr bir duygu durumu göstermektedir.

## 3.4 Granger Nedensellik Testi

Duygu etiketleme işleminden elde edilen skorların piyasa hareketleri üzerindeki etkisini incelemek amacıyla Granger nedensellik testi uygulanmıştır. Bu test, haber metinlerinden elde edilen duygu skorlarının (bağımsız değişken) Bitcoin fiyatlarının değişim oranını (bağımlı değişken) tahmin etmede anlamlı bir katkı sağlayıp sağlamadığını değerlendirmek için gerçekleştirilmiştir.

A screenshot of a black screen

Description automatically generated

**Şekil 3.8** Parite günlük kapanış fiyatı ve değişim oranı

Şekil 3.8'de görüldüğü gibi, Bitcoin ile ilgili haberlerin tarih aralığında kapanış fiyatları ve değişim oranları, Yahoo Finance[28] piyasa verilerini içeren "yfinance" kütüphanesi kullanılarak elde edilmiştir. Ayrıca, aynı gün içerisinde birden fazla haber çıkma durumu gözlemlendiğinden, o gün çıkan haberler için duygu skorlarının ortalamaları hesaplanmıştır. Bu işlem, verilerin zaman serisi formunda sunulmasını sağlamıştır.

A screen shot of a computer

Description automatically generated

**Şekil 3.9** Augmented Dickey-Fuller testi (p = 0.05 olarak kabul edildi)

Granger nedensellik testinden önce, zaman serilerinin durağan olup olmadığını kontrol etmek için ADF testi (Şekil 3.9) uygulanmıştır. ADF testi sonuçlarına göre hem “Sentiment\_Score” hem de “Change” zaman serilerinin durağan olduğu görülmüştür. Bu, serilerin ortalama, varyans ve kovaryansının zamanla sabit olduğu anlamına gelmektedir ve Granger nedensellik testi için gerekli olan durağanlık koşulunu sağlamaktadır.

Granger nedensellik testinin güvenilir bir şekilde uygulanabilmesi için gerekli olan ön koşul tamamlanmıştır. Bir sonraki adımda, bu seriler arasında nedensellik ilişkisini değerlendirmek için Granger testi uygulanmıştır.

**Tablo 3.1** Granger nedensellik testi 5 gecikme ile sonuçları

|  |  |  |
| --- | --- | --- |
| Gecikme(lag) | F-Test | p-değeri |
| 1 | 3.628085 | 0.057211 |
| 2 | 3.953125 | 0.019615 |
| 3 | 2.757982 | 0.041451 |
| 4 | 2.082520 | 0.081408 |
| 5 | 1.678895 | 0.137347 |

Granger nedensellik testi sonuçlarına göre, farklı gecikme seviyelerinde Bitcoin fiyat değişim oranları ile duygu skorları arasındaki nedensellik ilişkisi incelenmiştir. Testte anlamlılık seviyesi α=0.1 olarak belirlenmiştir. Bu eşik değeri, finansal zaman serilerinde ilişkilerin genellikle zayıf olmasından dolayı tercih edilmiştir.

Sonuçlara göre, lag=1 için p-değeri 0.057211 olarak hesaplanmış ve α=0.1 anlamlılık seviyesinin altında kalmıştır. Bu, duygu skorlarının Bitcoin fiyat değişim oranları üzerinde 1 günlük gecikmede Granger nedenselliği olduğunu göstermektedir. Lag=2, 3 ve 4 için ise p-değerleri sırasıyla 0.019615, 0.041451 ve 0.081408 olarak hesaplanmış, bu değerler de anlamlılık seviyesinin altında kalmıştır. Bu, duygu skorlarının Bitcoin fiyatları üzerindeki etkisinin yalnızca kısa vadeli olmadığını, aynı zamanda 2, 3 ve 4 günlük gecikmelerde de anlamlı bir ilişkiye işaret ettiğini göstermektedir. Ancak lag=5 için p-değeri 0.137347 olarak hesaplanmış, bu değer anlamlılık seviyesini aşmış ve bu gecikmede duygu skorlarının fiyatlar üzerindeki etkisinin anlamlı olmadığı belirlenmiştir.

Bu sonuçlar, duygu skorlarının Bitcoin fiyat değişim oranları üzerindeki etkisinin özellikle 1, 2, 3 ve 4 günlük gecikmelerde ortaya çıktığını, ancak daha uzun vadede bu ilişkinin zayıfladığını göstermektedir. α=0.1'in seçilmesi, finansal piyasalardaki kısa vadeli etkilerin ve gizli ilişkilerin daha iyi tespit edilmesine olanak sağlamış ve bu analizde anlamlı bir katkı sunmuştur. İlk 4 günlük gecikmelerde gözlemlenen nedensellik, piyasa duyarlılığı ile fiyat hareketleri arasındaki ilişkinin kısa ve orta vadede tahmin gücüne sahip olabileceğini işaret etmektedir.

## 3.5 Duygu Skorları ile Fiyat Arasındaki İlişki

Şekil 3.10’daki grafik, Bitcoin fiyatı ile duygu skorları arasındaki ilişkiyi zamana göre iki farklı eksende detaylı bir şekilde ortaya koymaktadır. Mavi çizgi, Bitcoin'in kapanış fiyatını temsil ederken, turuncu çizgi duygu skorlarının ham verilerini, kırmızı kesikli çizgi ise duygu skorlarının hareketli ortalamasını (MA) göstermektedir. Grafik, Bitcoin fiyatlarının 2022 yılından itibaren belirgin bir düşüş trendine girdiğini ve 2023 ortalarında toparlanmaya başladığını göstermektedir. Duygu skorlarının ham verileri (turuncu çizgi), bu süreçte oldukça yüksek bir volatilite sergilemiş ve piyasa katılımcılarının duyarlılık seviyelerinin sürekli değişken olduğunu ortaya koymuştur. Bu dalgalanmalar, duygu analizinin anlık piyasa tepkilerini anlamada etkili olabileceğini göstermektedir.

A graph of red and blue lines

Description automatically generated

**Şekil 3.10** Parite fiyatı ve duygu skorları arasındaki ilişki

Duygu skorlarının hareketli ortalaması (kırmızı kesikli çizgi), duyarlılık eğilimlerini daha net bir şekilde ortaya koymaktadır. Fiyat artışları sırasında duygu skorlarının ortalama olarak pozitif değerlere yaklaştığı, fiyat düşüşlerinde ise genellikle nötr veya negatif seviyelere gerilediği dikkat çekmektedir. Ancak, belirli zaman dilimlerinde fiyatlar ile duygu skorları arasında doğrudan bir paralellik gözlemlenememektedir. Örneğin, bazı dönemlerde fiyat düşüşlerine rağmen duygu skorlarının pozitif kalmaya devam ettiği görülmektedir. Bu durum, piyasa dinamiklerinin yalnızca duygu skorlarından değil, aynı zamanda makroekonomik gelişmeler, regülasyonlar ve haber akışı gibi diğer faktörlerden de etkilendiğini göstermektedir. Grafik, duygu skorlarının genel piyasa eğilimlerini anlamak için önemli bir araç olduğunu vurgulamakla birlikte, bu analizlerin tek başına yeterli bir gösterge olmayabileceğini de ortaya koymaktadır.

# **4 SONUÇ VE ÖNERİLER**

Bu çalışmada, finansal haberlerden elde edilen duygu skorlarının günlük Bitcoin kapanış fiyatları üzerindeki etkisi analiz edilmiş ve duygu analizinin kısa vadeli piyasa hareketlerini tahmin etmede belirli bir güce sahip olduğu tespit edilmiştir. Granger nedensellik testi sonuçları, duygu skorlarının 1, 2, 3 ve 4 günlük gecikmelerde Bitcoin fiyatları üzerinde anlamlı bir etkisi olduğunu göstermiştir. Ancak, 5 günlük ve sonrasındaki gecikmelerde bu etkinin anlamlı olmadığı belirlenmiştir. Grafiksel analizler ise duygu skorlarının hareketli ortalamalarının Bitcoin fiyatlarının uzun vadeli eğilimleriyle genel olarak paralel bir seyir izlediğini ortaya koymuştur. Bu bulgular, piyasa duyarlılığının kısa ve orta vadeli fiyat hareketlerinde etkili olabileceğini, ancak daha uzun vadeli fiyat dinamiklerinde diğer faktörlerin daha belirleyici bir rol oynadığını göstermektedir.

Yalnızca Bitcoin ve finansal haberler ile sınırlı kalmadan, çalışmanın kapsamı farklı kripto para birimlerine, geleneksel piyasa varlıklarına ve daha geniş veri kaynaklarına genişletilebilir, ayrıca farklı zaman dilimlerindeki anlık haber akışı ile anlık fiyat değişimleri de incelenebilir. Duygu analizinin sonuçlarının yalnızca haber veya sosyal medya duyarlılığı ile sınırlı kalmayıp, makroekonomik veriler, piyasa oynaklık endeksleri ve zincir üstü (on-chain) analizlerle birleştirilmesi, piyasa trendlerini ve uzun vadeli fiyat hareketlerini tahmin etme noktasında güçlü bir strateji oluşturabilir. Örneğin, ekonomik büyüme oranları, merkez bankası kararları ve küresel piyasa risk algısı gibi makroekonomik göstergeler, piyasalardaki dalgalanmalarla ilişkili olabilir. Bu tür bir çoklu veri entegrasyonu, yalnızca geçmiş veriler üzerinden tahmin yapmak yerine, daha kapsamlı bir bağlamsal çerçeve sunarak piyasa dinamiklerini daha doğru anlamamıza olanak tanıyabilir.

Bu bulguların yatırım stratejilerinde uygulanabilirliği de dikkat çekicidir. Ancak, yatırım kararlarında duygu analizinin yalnızca piyasanın bir yönünü yansıttığı unutulmamalı ve diğer temel ve teknik analiz faktörleri de dikkate alınmalıdır. Son olarak, gerçek zamanlı duygu analizi sistemlerinin geliştirilmesi, finansal piyasalarda hızlı karar verme süreçlerine destek olabilir. Bu tür sistemler, yatırımcıların anlık duyarlılık değişimlerini yakalamalarına ve stratejik kararlar almalarına yardımcı olabilir. Çalışmanın sonuçları, duygu analizi ile finansal piyasa hareketleri arasındaki ilişkiyi anlamada önemli bir temel oluşturmakla birlikte, bu alandaki araştırmaların derinleştirilmesi gerektiğini de göstermektedir.

# **KAYNAKLAR**

[1] Tetlock, P. C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." Journal of Finance.

[2] Bollen, J., Mao, H., & Zeng, X. (2011). "Twitter Mood Predicts the Stock Market." Journal of Computational Science.

[3] Loughran, T., & McDonald, B. (2011). "When is a liability not a liability? Textual analysis, dictionaries, and 10‐Ks." Journal of Finance, 66(1), 35-65.

[4] Xing, F., Cambria, E., & Welsch, R. E. (2018). "Natural language based financial forecasting: A survey." Artificial Intelligence Review, 50(1), 49-73.

[5] Zhang, L., & Skiena, S. (2010). "Trading strategies to exploit blog and news sentiment." Proceedings of the International AAAI Conference on Web and Social Media, 4(1), 375-378.

[6] Chen, T., Xu, R., He, Y., & Wang, X. (2018). "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN." Expert Systems with Applications, 72, 221-230.

[7] Araci, D. (2019). "FinBERT: A Pretrained Language Model for Financial Communications." arXiv preprint.

[8] Rodriguez Inserte, D., Pagliardini, M., & Hürlimann, M. (2024). "Financial Large Language Models: Improving performance and parameter efficiency with domain-specific pretraining." arXiv preprint.

[9] Asgarov, S. (2023). "Stock Price Prediction Using Twitter Sentiment Analysis and LSTM." arXiv preprint.

[10] Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.

[11] Hutchins, W. J. (1995). "Machine translation: A brief history." In Concise history of the language sciences (pp. 431-445). Pergamon.

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). "Attention is all you need." Advances in Neural Information Processing Systems, 30.

[13] http://datasciencedojo.com/blog/natural-language-processing-applications/

[14] http://www.almabetter.com/bytes/articles/getting-started-with-nlp

[15] http://aiml.com/what-are-the-advantages-and-disadvantages-of-bag-of-words-model/

[16] http://www.linkedin.com/pulse/creating-tf-idf-model-from-scratch-python-noor-saeed/

[17] http://towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314

[18] http://www.linkedin.com/pulse/understanding-text-classification-natural-language-david-adamson-mbcs/

[19] http://translate.google.com/?hl=tr&sl=auto&tl=tr&op=translate

[20] http://aws.amazon.com/tr/what-is/sentiment-analysis/

[21] http://www.sciencedirect.com/topics/earth-and-planetary-sciences/granger-causality-test

[22] http://en.wikipedia.org/wiki/Augmented\_Dickey%E2%80%93Fuller\_test

[23] http://www.analyticsvidhya.com/blog/2021/08/granger-causality-in-time-series-explained-using-chicken-and-egg-problem/

[24] http://www.python.org/

[25] http://www.kaggle.com/datasets/imadallal/sentiment-analysis-of-bitcoin-news-2021-2024

[26] http://www.kaggle.com

[27] http://huggingface.co/mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis

[28] http://finance.yahoo.com/

# **EKLER**

Bu kısımda proje boyunca Yazılan tüm kodlar verilmiştir.

A computer screen shot of a program code

Description automatically generated

A computer screen with white text and red text

Description automatically generated

A computer screen shot of a computer code

Description automatically generated

A black screen with white text

Description automatically generated

A screen shot of a computer program

Description automatically generated

A computer screen shot of text

Description automatically generated

A screen shot of a computer code

Description automatically generated

A black screen with colorful text

Description automatically generated

A computer screen shot of a program code

Description automatically generated

A black screen with white text

Description automatically generated

A screen shot of a computer program

Description automatically generated

A computer code on a black background

Description automatically generated

# **ÖZGEÇMİŞ**

Ad Soyad İlker EFİL

Doğum Tarhi 03.01.2001

Doğum Yeri Ordu

Lise Ankara Anadolu Lisesi (2015-2018)

Uğur Temel Lisesi (2018-2019)

Staj Yaptığı Yerler TSP Teknoloji – DWH Developer (2 ay)

IBTECH – İş Analisti (3 ay)