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

Açıklama otomatik olarak oluşturuldu

**YILDIZ TEKNİK ÜNİVERSİTESİ**

**KİMYA-METALÜRJİ FAKÜLTESİ**

**MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ**

**MATEMATİK MÜHENDİSLİĞİNDE TASARIM UYGULAMALARI**

**FİNANSAL PİYASA HAREKETLERİNİN MAKİNE ÖĞRENMESİ İLE TAHMİNİ**

Tez Yöneticisi: Doç. Dr. Ülkü YEŞİL

22052610, İlker EFİL

İstanbul, 2023

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

|  |  |
| --- | --- |
| **İÇİNDEKİLER** **Sayfa** | |
| SEMBOL LİSTESİ ................................................................................................. | iii |
| KISALTMA LİSTESİ ............................................................................................. | iv |
| ŞEKİL LİSTESİ ..................................................................................................... | v |
| ÖNSÖZ ................................................................................................................... | vi |
| ÖZET ...................................................................................................................... | vii |
| ABSTRACT ............................................................................................................ | viii |
| 1 GİRİŞ ................................................................................................................. | 9 |
| 1.1 Problem Tanımı ........................................................................................... | 9 |
| 1.2 Amaç ve Hedefler ....................................................................................... | 9 |
| 1.3 Çalışmanın Kapsamı ................................................................................... | 10 |
| 1.4 Literatür Taraması ....................................................................................... | 10 |
| 2 MAKİNE ÖĞRENMESİ ................................................................................... | 12 |
| 2.1 Gözetimli Öğrenme (Supervised Learning) ................................................ | 12 |
| 2.2 Gözetimsiz Öğrenme (Unsupervised Learning) ......................................... | 12 |
| 2.3 Pekiştirmeli Öğrenme (Reinforcement Learning) ....................................... | 13 |
| 2.4 Derin Öğrenme (Deep Learning) ................................................................ | 13 |
| 2.4.1 Yapay Sinir Ağları (YSA) ................................................................. | 14 |
| 2.4.2 Katmanlar .......................................................................................... | 14 |
| 2.4.3 Aktivasyon Fonksiyonları ................................................................. | 15 |
| 3 VERİ SETİ VE YÖNTEMLER ......................................................................... | 17 |
| 3.1 Verilerin Toplanması .................................................................................. | 17 |
| 3.2 LSTM (Long Short Term Memory) ............................................................ | 19 |
| 3.3 Random Forest (Rastgele Orman) ............................................................... | 20 |
| 4 YÖNTEMLERİN UYGULANMASI ................................................................ | 21 |
| 4.1 LSTM Yöntemi ........................................................................................... | 21 |
| 4.1.1 LSTM için Veri Setinin Hazırlanması .............................................. | 21 |
| 4.1.2 LSTM Modelinin Kurulması ............................................................ | 23 |
| 4.1.3 LSTM Değerlendirme ....................................................................... | 24 |
| 4.2 Random Forest Yöntemi ............................................................................. | 25 |
| 4.2.1 Random Forest için Veri Setinin Hazırlanması ................................ | 25 |
| 4.2.2 Random Forest Modelinin Kurulması ............................................... | 26 |
| 4.2.3 Random Forest Değerlendirme ......................................................... | 27 |
| 5 SONUÇ .............................................................................................................. | 28 |
| KAYNAKLAR ....................................................................................................... | 30 |
| EKLER .................................................................................................................... | 32 |
| ÖZGEÇMİŞ ............................................................................................................ | 37 |

**SEMBOL LİSTESİ**

σ(x) Sigmoid fonksiyonu (2.1)

tanh(x) Hiperbolik tanjant fonksiyonu (2.2)

relu(x) ReLU fonksiyonu (2.3)

Σ Toplam sembolü (4.1)

RS Göreceli güç (4.2)

N Hesaplama için kullanılan zaman dilimi (4.3)

**KISALTMA LİSTESİ**

ARIMA Bütünleşik Otoregresif Hareketli Ortalama

BLSTM İki Yönlü Uzun Kısa Süreli Bellek

GRU Kapı Özyinelemeli Geçitler

LSTM Uzun Kısa Süreli Bellek

MSE Ortalama Kare Hatası

ReLU Rectified Linear Unit

RMSE Kök Ortalama Kare Hatası

RNN Tekrarlayan Sinir Ağları

RSI Göresceli Güç Endeksi

SMA Basit Hareketli Ortalama

SVM Destek Vektör Makineleri

YSA Yapay Sinir Ağları

**ŞEKİL LİSTESİ Sayfa**

|  |  |  |
| --- | --- | --- |
| Şekil 2.1 | Gözetimli öğrenme örneği ................................................................ | 12 |
| Şekil 2.2 | Gözetimsiz öğrenme örneği ............................................................. | 12 |
| Şekil 2.3 | Pekiştirmeli öğrenme örneği ............................................................ | 13 |
| Şekil 2.4 | Derin öğrenme ve yapay zekanın makine öğrenimi ile ilişkisi ........ | 13 |
| Şekil 2.5 | Basit ve derin yapay sinir ağları ....................................................... | 14 |
| Şekil 2.6 | Yapay sinir ağı matematiksel modeli ............................................... | 15 |
| Şekil 2.7 | Sigmoid Fonksiyonu ........................................................................ | 15 |
| Şekil 2.8 | Tanh Fonksiyonu .............................................................................. | 16 |
| Şekil 2.9 | ReLu Fonksiyonu ............................................................................. | 16 |
| Şekil 3.1 | Verilerin elde edilmesi ..................................................................... | 17 |
| Şekil 3.2 | Google hisse verileri 2013-12-02 / 2023-11-30 ............................... | 18 |
| Şekil 3.3 | İlgili tarihler arasındaki fiyat değişimi ............................................. | 18 |
| Şekil 3.4 | LSTM yapısı ..................................................................................... | 19 |
| Şekil 3.5 | Random Forest yapısı ....................................................................... | 20 |
| Şekil 4.1 | float64 tipinden float32 veri tipine dönüştürme işlemi .................... | 21 |
| Şekil 4.2 | Veri setinin eğitim ve test olarak ayrımı. ......................................... | 21 |
| Şekil 4.3 | Verilerin ölçeklendirilmesi ............................................................... | 22 |
| Şekil 4.4 | Zaman Aşımı Algoritması ................................................................ | 23 |
| Şekil 4.5 | LSTM Katmanlar ............................................................................. | 23 |
| Şekil 4.6 | LSTM Model .................................................................................... | 24 |
| Şekil 4.7 | LSTM için RMSE değerleri ............................................................. | 24 |
| Şekil 4.8 | RSI ve SMA değerlerinin eklenmesi ................................................ | 25 |
| Şekil 4.9 | Eğitim ve test verisinin ayrılması ..................................................... | 26 |
| Şekil 4.10 | Random Forest Model ..................................................................... | 26 |
| Şekil 4.11 | Random Forest için başarı değerlendirme metrikleri ....................... | 27 |

**ÖNSÖZ**

Bu çalışmanın ana fikrini benimsememde, harekete geçmemde bana destek olan aileme, arkdaşlarıma ve hocalarıma minnettarım.

Bu çalışma süresince her daim yanımda olan, bana motivasyon veren ve desteklerini esrigemeyen annem Songül Efil, babam Durmuş Efil ve kardeşim İrem Efil’e sonsuz teşekkür ederim.

Araştırmalarım süresince bana desteğini esirgemeyen her konuda yardımcı olan, sorularımı her daim cevaplayan tez danışmanım Doç. Dr. Ülkü YEŞİL hocama sonsuz teşekkürlerimi sunuyorum.

İlker EFİL

İstanbul, 2023

**ÖZET**

Çalışmanın temel amacı, farklı makine öğrenimi algoritmalarının finansal varlıkların fiyatlarını tahmin etme yeteneklerini karşılaştırmak ve bu modellerin tahmin başarısını anlamaktır.

Literatür taraması sonucunda finansal piyasalardaki makine öğrenimi uygulamalarının çeşitliliği göz önüne alınmıştır. Ardından, LSTM ve Random Forest gibi önemli makine öğrenimi modelleri kullanılarak Google şirketinin belirli bir döneme ait fiyat tahminleri yapılmıştır. Elde edilen sonuçlar, LSTM modelinin daha etkili bir performans sergilediğini göstermiştir, ancak her iki modelin de eğitim ve test setleri arasındaki performans farklarının dikkate alınması gerektiği vurgulanmıştır.

Bu çalışma, finansal piyasalarda makine öğrenimi ile tahminin önemini ortaya koyarken, aynı zamanda bu modellerin sınırlamalarını ve dikkate alınması gereken faktörleri vurgulamaktadır.

**Anahtar kelimeler:** LSTM, Random Forest, Finansal Piyasalar, Makine Öğrenmesi

**ABSTRACT**

The main goal of this study is to compare how different machine learning algorithms predict financial asset prices and understand how successful these models are in making predictions.

Taking into account the diversity of machine learning applications in financial markets, a review of the literature was conducted. After that, significant machine learning models such as LSTM and Random Forest were used to predict the prices of Google company for a specific period. The results showed that the LSTM model demonstrated more effective performance, emphasizing the importance of considering performance differences between training and test sets for both models.

While highlighting the importance of machine learning in forecasting in financial markets, this study also points out the limitations of these models and factors that need to be taken into account.

**Keywords:** LSTM, Random Forest, Financial Markets, Machine Learning

**1 GİRİŞ**

Teknolojik gelişmelerin hız kazandığı günümüzde, finansal piyasa hareketlerinin makine öğrenmesi ile tahminini anlamak ve uygulamak, finansal karar alma süreçlerinde daha sofistike bir yaklaşım benimsemeyi gerektirmektedir. Bu bağlamda, insan ihtiyaçlarındaki artış ve finansal piyasalardaki karmaşıklığın büyümesi, geleneksel analiz yöntemlerinin sınırlarını ortaya çıkarmaktadır. Makine öğrenimi algoritmaları, özellikle de derin öğrenme yaklaşımları, finansal piyasa hareketlerindeki gizli örüntüleri keşfetme ve gelecekteki trendleri daha hassas bir şekilde tahmin etme potansiyeli sunmaktadır.

Bu bağlamda, finansal piyasa hareketlerinin makine öğrenmesi ile tahminini ele almak, mevcut veri setlerini daha etkili bir şekilde analiz etme ve piyasadaki değişimlere hızla adapte olma kapasitesi sağlamak adına önemlidir. Bu teknikler, gelecekteki trendeleri keşfetme ve kısa vadeli dalgalanmalara anında tepki verme yeteneği ile finansal analistlere ve yatırımcılara daha derin ve hızlı bir analiz fırsatı sunmaktadır.

**1.1 Problem Tanımı**

Finansal piyasalardaki karmaşık ve dinamik yapı, geleneksel finansal analiz yöntemlerinin sıklıkla yetersiz kalmasına neden olmaktadır. Yatırımcılar, hisse senedi fiyatları, döviz kurları ve emtia piyasalarındaki dalgalanmaları anlamak ve gelecekteki fiyat hareketlerini öngörmek için daha güçlü ve etkili araçlara ihtiyaç duymaktadır. Bu noktada, finansal piyasa tahminleme süreçlerine yapay zeka ve makine öğrenmesi tekniklerinin dahil edilmesi, özellikle derin öğrenme algoritmalarının kullanımı, daha güçlü analizlerin önünü açabilir ve tahminleme süreçlerinin performansını artırabilir.

**1.2 Amaç ve Hedefler**

Bu çalışmanın amacı, finansal piyasalardaki karmaşık ve dinamik yapının neden olduğu zorluklara karşı daha etkili bir çözüm sunmak üzere makine öğrenmesi tekniklerini finansal piyasa tahminleme süreçlerine entegre etmektir. Geleneksel finansal analiz yöntemlerinin yetersiz kaldığı bir ortamda, yatırımcılar, hisse senedi fiyatları, döviz kurları ve emtia piyasalarındaki dalgalanmaları daha derinlemesine anlamak ve gelecekteki fiyat hareketlerini daha doğru bir şekilde öngörmek amacıyla daha güçlü araçlara ihtiyaç duymaktadır.

Bu bağlamda, çalışmanın ana hedefi, finansal piyasa verilerini analiz etmek ve tahmin etmek için özellikle uygun makine öğrenimi algoritmalarını kullanarak daha güçlü bir tahminleme modeli geliştirmektir. Bu modellerin başarısını, istatistiksel metriklerle değerlendirip, karşılaştırıp finansal piyasa yatırımcılarına daha etkili ve güvenilir bir analiz aracı sunmaktır. Bu şekilde, makine öğrenimi destekli finansal tahminleme süreçleri, yatırımcıların daha bilinçli ve kararlı yatırım stratejileri oluşturmalarına katkıda bulunabilir.

**1.3 Çalışmanın Kapsamı**

Bu çalışma, bir dizi adımdan oluşan bir metodoloji kullanarak, makine öğrenmesi tekniklerini finansal piyasa verileriyle entegre etmeyi amaçlamaktadır. İlk olarak, finansal varlıkların fiyat hareketlerini tahminlemeyi amaçlayan, hangi algoritmanın nasıl bir sonuç verdiğini inceleyen literatür taraması yapılmıştır. Sonrasında ise makine öğreniminden bahsedilmiştir ve ardından derin öğrenme konusuna değinilmiştir. Son olarak, uygun veri kaynağı seçilmiştir ve bu veriler ön işleme adımlarından geçirilmiştir. Model geliştirme sürecinde uygun algoritma kurulmuştur, eğitim ve doğrulama verileri kullanılarak modelin optimize edilmesi sağlanmıştır.

Çalışmanın kapsamı aynı zamanda elde edilen sonuçların eleştirel bir şekilde değerlendirilmesini içermekte olup, bu değerlendirme finansal piyasa yatırımcılarına ve araştırmacılara gerçek dünya uygulamalarında kullanılabilecek değerli içgörüler sağlamayı amaçlamaktadır.

Bu çalışmanın sonunda, makine öğrenmesi tekniklerinin finansal piyasa tahminleme süreçlerine olan katkıları ve bu bağlamda sahip olduğu avantajlar ile sınırlamalar ele alınmıştır.

**1.4 Literatür Taraması**

Seher Arslankaya ve Şevval Toprak’ın çalışmasında makine öğrenmesi yöntemlerinden Polinom Regresyon ve Random Forest Regresyon, derin öğrenme tekniklerinden Tekrarlayan Sinir Ağları (RNN) ve Uzun-Kısa Süreli Bellek (LSTM) yöntemleri kullanılmıştır. Elde edilen sonuçlara göre en iyi sonucu Random Forest Regresyon modeli, en kötü sonucu ise Polinom Regresyon modeli vermiştir. Tahmin sonuçlarının başarısının istatistik analizi MSE, MAE ve RMSE hata metriklerinden faydalanarak yapılmıştır. Random Forest Regresyon için bakıldığında ise hata değerleri % 0,14 ve % 3,7 arasında değiştiği görülmüştür [1].

Erdoğan Gavcar ve Hüseyin Mustafa Metin’in çalşmasında ise Ocak 2016 ve Aralık 2021 tarihleri arasındaki günlük hisse senedi fiyatlarının veri değerleri kullanılarak, derin öğrenme modellerinden Uzun Kısa Süreli Bellek (LSTM modeli ile hisse senetlerinin açılış fiyat değerleri tutarlı tahminler ile elde edilmiştir. LSTM modelinden korelasyon katsayısı performans değerlendirme ölçütüne göre %95 doğruluk oranı elde edilmiştir [2].

Bu çalışma, hisse senedi piyasası getirilerini tahmin etmeyi amaçlayan bir analizi içermektedir. Şirketin web sitesinde bulunan tarihsel veri seti sınırlı özelliklere sahip olduğundan, mevcut değişkenler kullanılarak yeni değişkenler oluşturularak daha yüksek doğruluk elde edilmeye çalışılmıştır. Tahminleme için Yapay Sinir Ağı (YSA) ve karşılaştırmalı bir analiz için Random Forest kullanılmıştır. Elde edilen sonuçlar, YSA’nın Random Forest’a göre daha iyi hisse senedi fiyatı tahminleri sağladığını göstermiştir. En iyi performansı gösteren YSA modelinin RMSE (0.42) değerleri elde ettiği belirlenmiştir [3].

Adem Üntez ve Mümtaz İpek’in çalışmasında ise derin öğrenme algoritmaları olan LSTM ve ARIMA mimarileri kullanılarak gümüş/ons paritesi üzerinde tahminler yapılmıştır. Elde edilen sonuçlara göre, LSTM mimarisi günlük kapanış değerlerini net bir şekilde tahmin etmiştir. ARIMA mimarisi ile yapılan tahmin çalışması, hem aralık hem de yön açısından grafikle daha yakın bir sonuç üretmiştir. LSTM algoritması, RMSE ve MAE açısından sırasıyla % 5,74 ve % 10,64 oranında ARIMA algoritmasından daha iyi performans göstermiştir [4].

Nesrin Koç Ustalı, Nedret Tosun ve Ömür Tosun’un çalışmasında, Borsa İstanbul'da işlem gören şirketlerin hisse senetlerinin gelecek fiyatlarını tahmin etmeyi amaçlanmıştır. BİST 30 Endeksi'nde yer alan şirketlerin 2010-2019 finansal tabloları ve aylık kapanış fiyatları kullanılarak Yapay Sinir Ağları (YSA), Random Forest ve XGBoost algoritmaları ile tahminler yapılmıştır. Elde edilen sonuçlara göre, XGBoost algoritması en iyi performansı gösterirken, Rastgele Orman ile benzer sonuçlar elde edilmiştir. Her iki modelin de YSA'ya göre daha yüksek performans sergilediği tespit edilmiştir [5].

Bu çalışmada, LSTM, GRU ve BLSTM isimli 3 farklı derin öğrenme modeli kullanılarak bir hisse senedi tahmin sistemi geliştirilip, kullanılan modeller arasında karşılaştırmalı bir analiz yapılmıştır. Spekülatif hareketlerden uzak olması için veri seti olarak 1968'den 2018'e kadar olan New York Borsası'ndan hisse senedinin zaman serisi değerleri kullanılmıştır. Spesifik olarakta IBM hisse senedi ile test çalışmaları yapıldı. Deneysel sonuçlar, hisse senedine ait son 5 günlük işlem verilerinin girdi olarak kullanıldığı BLSTM modeliyle yapılan eğitimin %63,54 lük bir yönsel doğruluk değerine ulaşıldığını göstermişir [6].

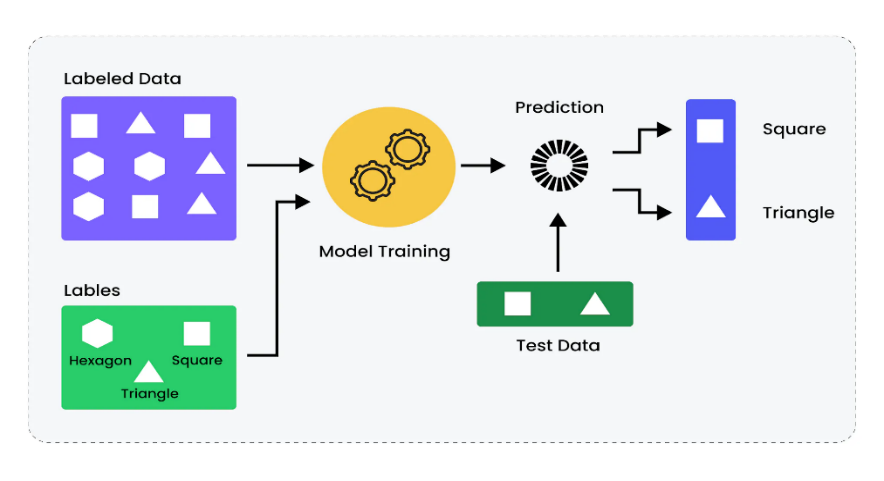
Bu çalışma, Hindistan hisse senedi piyasası için hisse senedi ve hisse senedi endekslerinin hareket yönünü tahmin etme sorununu ele almaktadır. Dört farklı tahmin modeli, Yapay Sinir Ağı (YSA), Destek Vektör Makinesi (SVM), Random Forest ve Naive-Bayes, bu modellere iki farklı giriş yaklaşımıyla karşılaştırılmıştır. İlk yaklaşım, on teknik parametrenin (açılış, yüksek, düşük ve kapanış fiyatları) sürekli değerler olarak temsil edilmesine dayanırken, ikinci yaklaşım bu teknik parametreleri trend belirleyici veri olarak temsil etmeye odaklanmıştır. Değerlendirme, Reliance Industries ve Infosys Ltd. adlı iki hisse senedi ile CNX Nifty ve S&P Bombay Menkul Kıymetler Borsası (BSE) Sensex endekslerinin 2003-2012 yılları arasındaki on yıllık tarihsel verileri üzerinde gerçekleştirilmiştir. Sonuçlar, on teknik parametrenin sürekli değerler olarak temsil edildiği ilk yaklaşımda Random Forest’in genel performansta diğer modelleri geride bıraktığını göstermektedir. Ayrıca, bu teknik parametrelerin trend belirleyici veri olarak temsil edildiğinde tüm tahmin modellerinin performansının arttığı gözlemlenmiştir [7].

**2 MAKİNE ÖĞRENMESİ**

Makine öğrenmesi, bilgisayar sistemlerinin deneyimlerden öğrenme yeteneği kazandığı bir yapay zeka dalıdır. Geleneksel programlamada olduğu gibi belirli bir görev için doğrudan talimatlar vermek yerine, makine öğrenmesinde bilgisayarlar, veri üzerinde örüntüler ve ilişkiler bulma yeteneğine sahiptir. Bu, algoritmaların deneyimlerle geliştirilmesi ve performanslarını zamanla iyileştirmeleri anlamına gelir.

**2.1 Gözetimli Öğrenme (Supervised Learning)**

Bu algoritmalar, veri kategorilerinin veya sonuçların önceden bilindiği durumları içerir. Algoritma, bilinen sonuçlara dayanarak öğrenir ve gelecekteki tahminlerde bu bağlantıları kullanır. Gözetimli öğrenme iki alt kategoriye ayrılır:

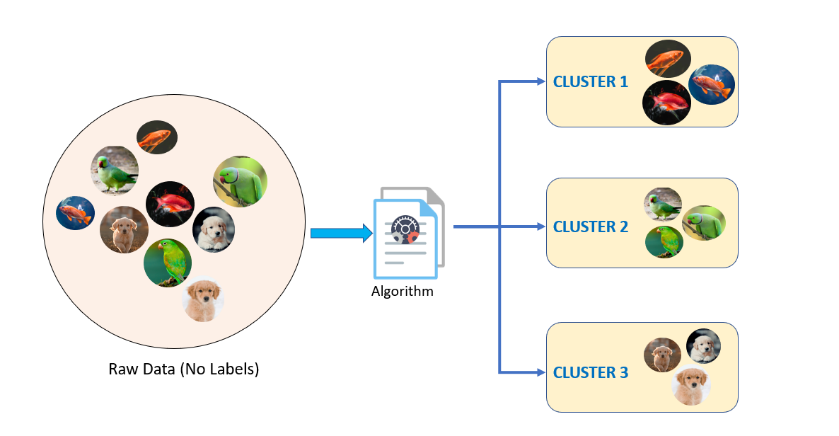
* Regresyon: Sürekli çıktılara odaklanır.
* Sınıflandırma: Belirli kategorilere ait çıktıları sınıflandırır.

**Şekil 2.1** Gözetimli öğrenme örneği [8]

**2.2 Gözetimsiz Öğrenme (Unsupervised Learning)**

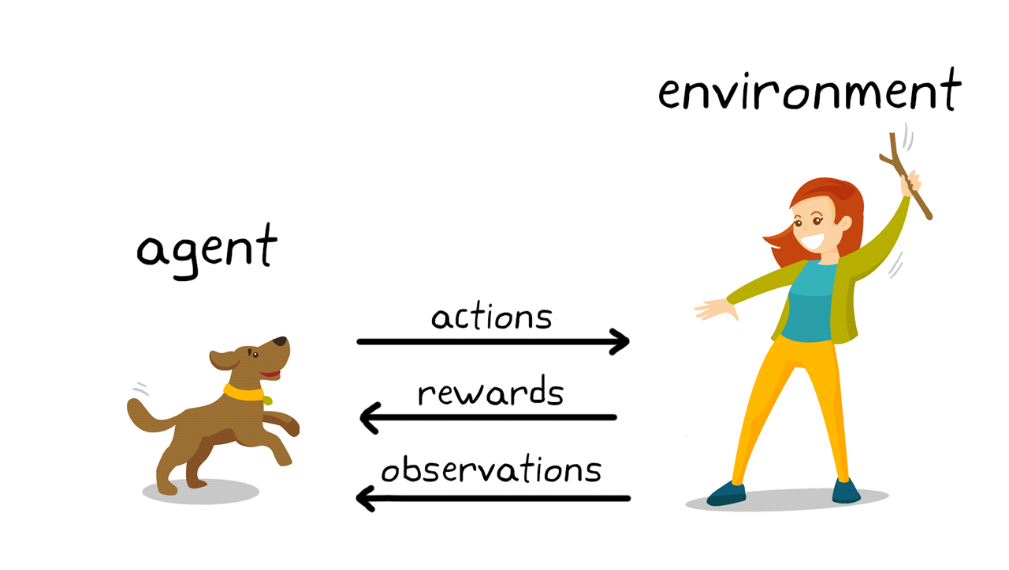
Bu kategori, sonuç bilgisi olmaksızın etiketlenmemiş veriler üzerinde çalışır. Bu algoritmalar, veriler arasında doğal bağlantılar kurarak kümeler oluşturur. Gözetimsiz öğrenme üç alt kategoriye ayrılabilir:

* Kümeleme (Clustering): Benzer özelliklere sahip verileri gruplandırır.
* Birliktelik Kuralı (Association Rule Mining): Veriler arasındaki ilişkileri belirler.
* Boyut Azaltma (Dimensionality Reduction): Veri setinin karmaşıklığını azaltır.



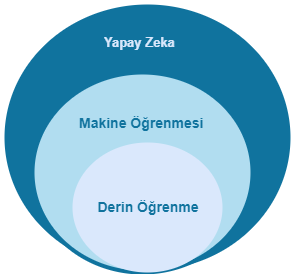
**Şekil 2.2** Gözetimsiz öğrenme örneği [9]

**2.3 Pekiştirmeli Öğrenme (Reinforcement Learning)**

****Bu yöntem, öğrenme sürecini geri bildirimle ilerletir. Algoritma sürekli denemeler yapar, başarılı veya başarısız olduğu durumları değerlendirir ve hedefe ulaşana kadar bu bilgiyi kullanarak devam eder. Hareket değerlendirmesi sayesinde algoritma sürekli olarak gelişir ve yanlış adımlardan kaçınmayı öğrenir.

**Şekil 2.3** Pekiştirmeli öğrenme örneği [10]

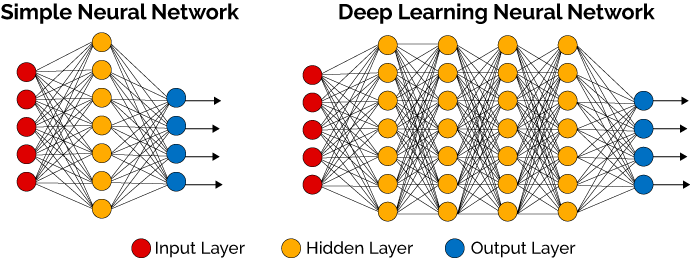
**2.4 Derin Öğrenme (Deep Learning)**

****Derin öğrenme, yapay zeka ve makine öğrenmesinde kullanılan güçlü bir teknik olup, karmaşık ve büyük veri setlerini analiz etme, desenleri tanıma ve karmaşıklığı anlama konularında etkili bir şekilde kullanılmaktadır. Derin öğrenme, yapay sinir ağları kullanarak daha yüksek düzeyde temsil öğrenimine odaklanmaktadır.

**Şekil 2.4** Derin öğrenme ve yapay zekanın makine öğrenimi ile ilişkisi [11]

**2.4.1 Yapay Sinir Ağları (YSA)**

Derin öğrenme, çok katmanlı yapay sinir ağları (neural networks) kullanarak çalışır. Bu ağlar, insan beyninin nöronlarını taklit eden bir mimari kullanarak veriler üzerinde karmaşık örüntüleri öğrenirler. Her katman, belirli bir düzeyde temsil oluşturur ve bu temsiller, öğrenme sürecinin ilerlemesiyle daha karmaşık ve soyut özelliklere dönüşür.



**Şekil 2.5** Basit ve derin yapay sinir ağları [12]

**2.4.2** **Katmanlar**

Bir derin öğrenme modeli genellikle giriş katmanı, gizli katmanlar ve çıkış katmanı olmak üzere birden çok katmandan oluşur. Her bir gizli katman, veriler üzerinde farklı özellikleri çıkarmak ve temsil etmek için tasarlanmıştır. Yeterli sayıda ve düzenlenmiş gizli katmanlar, modelin karmaşık yapıları daha iyi anlamasına olanak tanır.

Giriş katmanı, ağa dış dünyadan gelen verileri tanıtır. Her bir giriş, bir nörona karşılık gelir ve bu nöron, verinin ağırlıklarla çarpılması ve bir aktivasyon fonksiyonundan geçirilmesi ile bir çıkış üretir. Giriş katmanı, genellikle veri setindeki özelliklerin sayısına eşit sayıda nörona sahiptir.

Gizli katmanlar, ağın karmaşıklığını ve öğrenme kapasitesini artıran katmanlardır. Bu katmanlar, giriş verilerinden çeşitli özelliklerin çıkarılmasını sağlar. Her bir nöron, önceki katmandaki tüm nöronlarla bağlantılıdır ve bu bağlantılar ağırlıklarla belirlenir. Gizli katmanlardaki nöronlar genellikle bir aktivasyon fonksiyonundan geçirilerek çıkış üretir.

Çıkış katmanı, ağın nihai çıkışını üreten katmandır. Bu katmanın nöron sayısı, problem türüne bağlıdır. Örneğin, ikili sınıflandırma problemlerinde bir nöron, çok sınıflı sınıflandırmada ise sınıf sayısı kadar nöron kullanılabilir. Çıkış katmanındaki nöronlar genellikle hiberbolik tanjant, sigmoid, ReLu gibi uygun aktivasyon fonksiyonlarından geçirilir.

**2.4.3 Aktivasyon Fonksiyonları**

metin, diyagram, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturulduYapay sinir ağlarında, gerçek dünya özelliklerini etkili bir şekilde temsil etmek ve ağın öğrenme kapasitesini artırmak için aktivasyon fonksiyonları kullanılır. Temelde, basit bir yapay sinir ağında girişler *x* ve ağırlıklar *w* olarak tanımlanır. Bu girişler ve ağırlıklar, bir aktivasyon fonksiyonu *f*(*x*) ile işlenir ve elde edilen çıkış, nihai sonuç ya da bir sonraki katmanın girişi olarak kullanılır. Bu süreç, ağın doğrusal olmayan özellikleri öğrenmesini sağlayarak, daha karmaşık ve gerçek dünya verilerini etkili bir şekilde işlemesine olanak tanır [13].

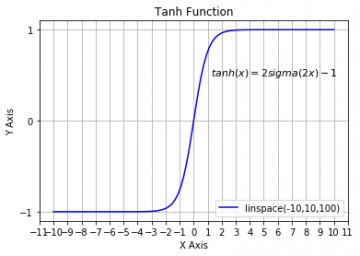
**Şekil 2.6** Yapay sinir ağı matematiksel modeli [14]

metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturulduSigmoid fonksiyonu, genellikle ikili sınıflandırma problemlerinde tercih edilir çünkü çıkışı 0 ile 1 arasında bir değer alır. Bu özelliği sayesinde sonucun 0 veya 1 olarak yorumlanabilmesine olanak sağlar. Bu sayede, sınıflandırma sonuçları daha kesin ve anlaşılır hale gelir, özellikle tek bir sınıfa ait olma olasılığı yüksek olduğunda. Yani, sigmoid fonksiyonu, özellikle "evet" veya "hayır" gibi iki seçenek arasında karar vermek için kullanıldığında daha etkili bir sonuç elde etmemize yardımcı olmaktadır.

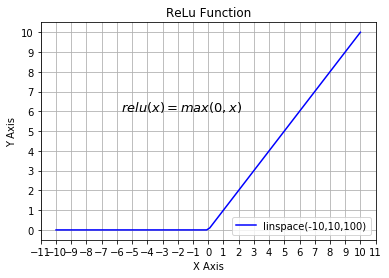
**Şekil 2.7** Sigmoid Fonksiyonu [15]

(2.1)

Hiperbolik tanjant fonksiyonu (tanh), genellikle sinir ağlarının gizli katmanlarında tercih edilir çünkü çıktıları -1 ile 1 arasındadır. Bu özellik, gizli katmanların ortalama değerini 0'a yaklaştırır, bu da ağın daha stabil bir şekilde öğrenmesine katkı sağlar. Ortalama değerlerin yaklaşık olarak sıfır olması, sinir ağının daha etkili bir şekilde öğrenmesine yardımcı olabilir ve bu da genellikle veri setlerinin istatistiksel özelliklerini daha iyi temsil etmesine olanak tanır.

**Şekil 2.8** Tanh Fonksiyonu [16]

(2.2)

****ReLU (Rectified Linear Unit), derin öğrenme modellerinde sıkça kullanılan bir aktivasyon fonksiyonudur. Girdi pozitifse aynı değeri, negatifse sıfırı çıkış olarak verir. Basit hesaplamaları ve türevidirken yaşanan kolaylık, modelin hızlı öğrenmesine katkıda bulunur.

**Şekil 2.9** ReLu Fonksiyonu [17]

(2.3)

**3 VERİ SETİ VE YÖNTEMLER**

Bu çalışma, finansal piyasalardaki gelecekteki fiyat hareketlerini makine öğrenmesi ile tahmin etme amacı güden bir yaklaşımı incelemektedir. Çalışmada kullanılacak teknikler, literatür araştırmaları bağlamında LSTM (Uzun Kısa Süreli Bellek) ve Random Forest (Rastgele Orman) yöntemidir.

LSTM, finansal piyasalardaki belirsizlikleri ve doğrusal olmayan ilişkileri anlama yeteneği sayesinde özellikle bu alanda kullanım potansiyeli taşımaktadır. Bu yaklaşımla, ilgili işlem günününe ait kapanış fiyatı ve tarihi üzerinden tahmin işlemi gerçekleştirilmiştir.

Random Forest, finansal piyasalardaki belirsizlik ve karmaşıklıkları ustaca ele alma yetisiyle öne çıkmaktadır. Bu algoritmayla, geçmiş fiyat hareketleri, alım satım hacmi ve diğer finansal göstergeler üzerinden gerçekleştirdiği öğrenme süreciyle, gelecekteki fiyat eğilimlerini öngörmek amaçlanmıştır.

Elde edilen sonuçlar, finansal piyasalardaki makine öğrenmesi tabanlı tahmin yaklaşımlarının pratik uygulanabilirliğini ve başarı potansiyelini değerlendirmek amacıyla analiz edilmiştir.

**3.1 Verilerin Toplanması**

Verilerin toplanması ve ön işleme aşamaları, finansal piyasaların makine öğrenmesiyle tahmin edilmesi amacıyla gerçekleştirilen araştırmaların temel taşlarından birini oluşturur. Bu aşamalar, modelin güvenilir ve etkili bir şekilde çalışabilmesi için büyük bir öneme sahiptir. Verilerin uygun bir şekilde toplanmaması, gürültülü veya eksik verilere yol açabilir, bu da modelin doğruluğunu ve performansını olumsuz etkileyebilir.

**metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**Bu bağlamda, finansal piyasalardan elde edilen verilerin gürültüsüz ve güvenilir olması kritiktir. Bu nedenle, veri toplama aşamasında Yahoo Finance'e[18] ait yfinance kütüphanesi tercih edilmiştir. Bu kütüphane, finansal piyasa verilerini etkili bir şekilde çekebilen ve kullanıcıya geniş bir veri yelpazesi sunan bir araçtır. Python[19] dilinin esnekliği ve etkileşimli Jupyter Lab[20] ortamının kullanım kolaylığı, veri çekme ve düzenleme süreçlerini daha da optimize etmektedir.

**Şekil 3.1** Verilerin elde edilmesi

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduŞekil 3.2** Google hisse verileri 2013-12-02 / 2023-11-30

Tablo 1’den anlaşıldığı üzere bu araştırmada, örnek olması amacıyla Google şirketinin 2013-12-01 ile 2023-12-01 tarihleri arasındaki finansal veri seti kullanılmıştır, ki bu veriler Yahoo Finance kaynağından temin edilmiştir. Yahoo Finance'in sağladığı geniş kapsamlı veri setleri, hisse senedi fiyatlarından işlem hacmine kadar çeşitli finansal göstergeleri içermekte ve bu nedenle Google'ın ve diğer enstrümanların finansal durumunu detaylı bir şekilde analiz etmek için uygun bir kaynak sağlamaktadır.

* Date, ilgili işlem gününün tarhini temsil etmektedir.
* Open, ilgili işlem günününe ait hisse başlangıç fiyatını temsil etmektedir.
* High, ilgili işlem günününe ait en yüksek fiyatı temsil etmektedir.
* Low, ilgili işlem günününe ait en düşük fiyatı temsil etmekedir
* Close, İlgili işlem günününe ait kapanış fiyatını temsil etmektedir.
* Adj Close, ilgili işlem günününe ait düzeltilmiş kapanış fiyatını temsil etmetedir.
* metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, ekran görüntüsü, çizgi içeren bir resim

  Açıklama otomatik olarak oluşturulduVolume, ilgili işlem günününe ait işlem hacmi bilgisini temsil etmektedir.

**Şekil 3.3** İlgili tarihler arasındaki fiyat değişimi

**3.2 LSTM (Long Short Term Memory)**

ekran görüntüsü, kırpıntı çizim, çizim içeren bir resim

Açıklama otomatik olarak oluşturulduLSTM yinelemeli bir sinir ağı türüdür ve temelde bilgi akışını düzenlemek için özel kapı mekanizmalarını kullanır. Her bir veri noktasında, LSTM, üç ana kapıdan oluşan bir yapıyla bilgi akışını kontrol eder: unutma kapısı, giriş kapısı ve çıkış kapısı. Unutma kapısı, önceki bilgilerin ne kadarının unutulacağını belirler. Giriş kapısı, yeni bilgilerin ne kadarının hafızaya eklenmesi gerektiğini kontrol eder. Çıkış kapısı ise hafızadan alınan bilgileri kullanarak güncellenmiş bir bilgi akışını üretir. Bu kapı mekanizmaları, özellikle uzun vadeli bağımlılıkları daha iyi yöneterek ve gradient kaybını azaltarak LSTM'nin etkili bir şekilde öğrenmesini sağlar.

**Şekil 3.4** LSTM yapısı [21]

Unutma kapısı, hücre durumundan hangi bilgilerin silineceğini belirler. Bu, geçmişteki gereksiz bilgileri atarak modelin sadece önemli bilgileri hatırlamasını sağlar.

Giriş kapısı, hangi yeni bilgilerin hücre durumuna eklenip çıkarılacağını kontrol eder. Bu kapı, yeni bilgileri hücre durumuna ekleyerek önceki bilgilerle günceller.

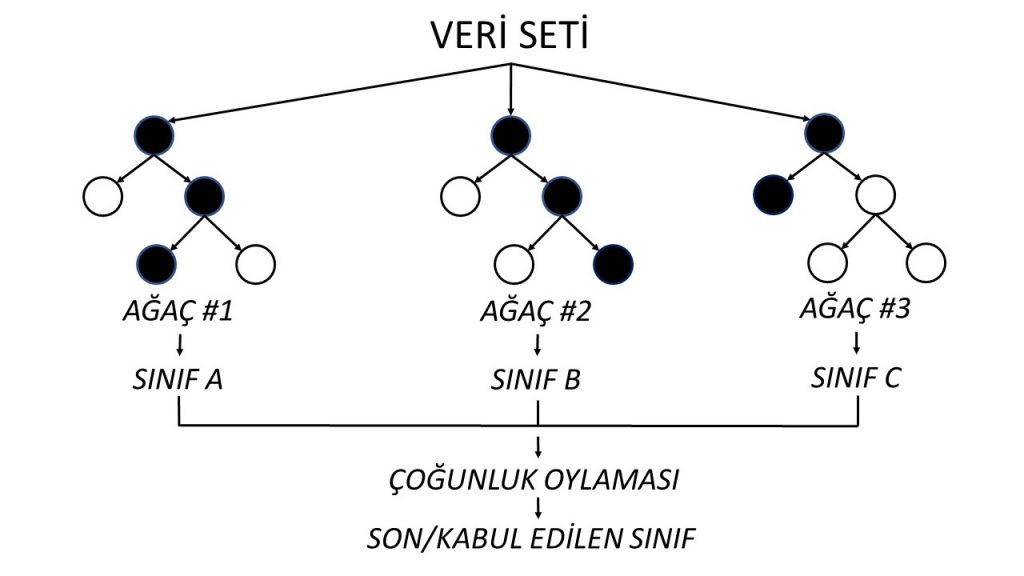
Hücre durumu, LSTM'nin temelinde bulunan bu özellik, bilginin uzun vadeli olarak saklandığı yerdir. Hücre durumu, belirli bir süre boyunca taşınan bilgileri içerir ve zaman içinde güncellenir.

Çıkış kapısı, hücre durumundan çıkan yeni bilgileri belirler. Bu kapı, güncellenmiş hücre durumunu filtreleyerek, belirli bir zaman noktasındaki bilgilerin model tarafından kullanılmasını sağlar.

LSTM, finansal piyasa analizi gibi karmaşık verilerin işlenmesinde, özellikle fiyat hareketlerini tahmin etme, risk yönetimi ve portföy yönetimi gibi finansal uygulamalar için büyük öneme sahiptir. Bu model, zaman içindeki verileri daha etkili bir şekilde analiz edebilme yeteneği sayesinde, finansal profesyonellere, bireysel veya kurumsal yatırımcılara daha güçlü bir araç sunar. LSTM, önceki çalışmalar ve literatürdeki yayınlarla desteklenen bir yöntemdir ve finansal piyasalardaki değişkenlikleri anlama ve gelecekteki fiyat hareketlerini tahmin etme yeteneği ile önemli bir rol oynar.

**3.3 Random Forest (Rastgele Orman)**

Random Forest, karar ağaçları temelli bir makine öğrenimi algoritmasıdır ve özellikle toplu öğrenme (ensemble learning) yöntemlerinden biridir. Random Forest, birden çok karar ağacını bir araya getirerek genel performansı artırır. Her bir karar ağacı, rastgele seçilen örneklem veriler üzerinde eğitilir ve bu sayede farklı bakış açılarından öğrenme gerçekleştirir.

****Random Forest, temelde birden çok karar ağacını birleştirerek güçlü bir model oluşturur. Modelin temel öğeleri arasında belirli bir sayıdaki ağaç sayısı, her ağacın bölünmesinde kullanılacak özelliklerin sayısı ve her bir ağacın maksimum derinliği yer alır. Her ağaç, rastgele seçilmiş özellikler üzerinden veri setini bölerek bir karar ağacı oluşturur. Model, her ağacın tahminlerini birleştirerek çoğunluk oylaması yapar, bu da genel tahminin belirlenmesini sağlar. Random Forest, her bir ağacın kendi içinde overfitting eğiliminde olsa bile, ensemble yapısı genellikle daha istikrarlı ve genelleme yeteneği yüksek bir model elde etmeye yardımcı olur. Bu özellikleri sayesinde Random Forest, finansal piyasa analizi gibi karmaşık verilerin işlenmesinde, fiyat hareketlerini tahmin etme, risk yönetimi ve portföy yönetimi gibi finansal uygulamalarda etkili bir şekilde kullanılır.

**Şekil 3.5** Random Forest yapısı [22]

Başlangıç aşamasında, belirli bir veri seti üzerinde karar ağacı oluşturulmak üzere n adet ağaç belirlenir. Bu, modelin genel karmaşıklığını ve çeşitliliğini kontrol etmek açısından önemlidir.

Her bir karar ağacının oluşturulma sürecinde, her düğümde rastgele seçilen m adet değişken arasından seçim yapılır. Bu değişkenler üzerinden Gini indeksi hesaplanarak en iyi bölünme seçilir. Bu adım, her ağacın farklı özellikleri vurgulamasını sağlayarak çeşitliliği artırır.

En iyi bölünme seçilerek belirlenen düğüm, Gini indeksi sıfır oluncaya kadar iki alt düğüme ayrılır. Bu aşama, her ağacın oluşturulma sürecini yönlendirir ve homojen alt düğümler elde edilene kadar devam eder.

Final aşamada ise n adet karar ağacının ayrı ayrı yaptığı tahminler arasından en fazla oy alan sınıf, modelin genel tahminini belirler. Bu, her ağacın katkısını birleştirerek daha güçlü bir tahmin elde edilmesine katkı sağlamaktadır.

**4 YÖNTEMLERİN UYGULANMASI**

Bu bölümde Random Forest ve LSTM modelinin finansal piyasa tahmini için uygulanış şekilleri ve oluşturulan tahmin modelleri anlatılmıştır. Yöntemin uygulanma şeklini madde madde belirtecek olursak:

* Verilerin hazırlanması,
* İlgili modelinin oluşturulması ve eğitilmesi,
* Modellerin performansının ölçülmesi,
* Modellerin değerlendirilmesi.

**4.1 LSTM Yöntemi**

Bu bölümde, LSTM yöntemi için verilerin hazırlanması, modelin oluşturulması, modelin performansının değerlendirilmesi başlıkları ele alınmıştır.

**4.1.1 LSTM için Veri Setinin Hazırlanması**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**Veri düzenlenmesi işlemleri, elde edilen finansal verilerin analiz ve modelleme süreçlerine uygun hale getirilmesini içerir. Bu aşamada, veri setindeki eksik değerlerin ele alınması, verilerin ölçeklendirilmesi ve düzeltilmesi gibi ön işleme teknikleri uygulanmıştır. İlgili görsellerle desteklenen bu düzenleme adımları, veri setinin daha sağlıklı ve model için uygun hale getirilmesine katkı sağlamıştır. Bu sayede, modelin daha güvenilir ve hassas tahminler yapabilmesi için temel veri ön işleme aşamaları başarıyla tamamlanmıştır.

**Şekil 4.1** float64 tipinden float32 veri tipine dönüştürme işlemi

Sinir ağlarında genellikle float32 veri tipi tercih edilir çünkü bu veri tipi, float64 veri tipine kıyasla daha düşük bellek kullanımına ve daha hızlı işlem yapma yeteneğine sahiptir. float32 veri tipi, her bir sayıyı 32 bitte temsil ederken, float64 bunu 64 bitte yapar. Bu durum, daha fazla sayıda verinin bellekte depolanabilmesi ve işlem yapılabilmesi anlamına gelmektedir.

Bu avantajlar, özellikle büyük veri setleri ve karmaşık sinir ağlarıyla çalışırken önemlidir. float32 veri tipi, genellikle GPU'lar tarafından daha etkili bir şekilde işlenebilir, bu da eğitim sürelerini kısaltabilir. Ayrıca, bellek kullanımındaki tasarruf, daha büyük modellerin eğitilmesine olanak tanır.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduBu bağlamda, Şekil 4.1’de ifade edildiği gibi, veri tipini float64'ten float32'ye çevirmek, bellek kullanımını optimize etmek ve işlem hızını artırmak için yapılmıştır.

**Şekil 4.2** Veri setinin eğitim ve test olarak ayrımı.

Eğitim veri seti 1762 günlük veriden oluşmaktadır. Bu dönem boyunca, modelin öğrenme sürecini etkili bir şekilde gerçekleştirmesi amaçlanmıştır. Öte yandan, test veri seti ise 755 günlük veriden oluşmaktadır. Bu, modelin eğitildiği dönemden farklı bir zaman aralığında performansını değerlendirmemizi sağlar. Bu bölünmüş veri setleri, modelin genelleme yeteneğini test etmek adına seçilmiş bir eğitim ve test dönemini temsil etmektedir.

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduŞekil 4.3** Verilerin ölçeklendirilmesi

Özellikler arasındaki büyüklük farklılıkları veya ölçek farklılıkları, modelin eğitimini olumsuz etkileyebilir. Örneğin, bir özellik 0 ile 1000 arasında değerler alırken, diğer bir özellik -1 ile 1 arasında değerler alabilir. Bu durumda, büyük değerlere sahip özellik, model üzerinde daha fazla etkiye sahip olabilir ve diğer özelliklerin etkisini azaltabilir.

Şekil 4.3’da yer alan MinMaxScaler, veri setindeki her bir özelliği belirli bir aralığa çekerek, genellikle [0, 1] aralığına, ancak isteğe bağlı olarak başka bir aralığa da, ölçekler. Bu, modelin daha dengeli ve etkili bir şekilde öğrenmesine yardımcı olmaktadır. Bu nedenle, MinMaxScaler gibi normalizasyon teknikleri, modelin daha hızlı ve etkili öğrenmesine yardımcı olmaktadır.

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduŞekil 4.4** Zaman aşımı algoritması

Zaman serilerinde etkili bir performans elde etmek amacıyla zaman aşımı (Şekil 3.10) algoritması uygulanmıştır. Bu algoritma (Şekil 3.10), belirtilen veri noktalarını gruplandırır ve önceki değerlerin belirlediği sayıdan sonraki değeri tahmin etmeye odaklanmaktadır. Bu değer 20 olarak belirlenmiştir ve her biri 20’şerli olarak gruplandırılmıştır. Ardından, önceki 20 değerle karşılaştırılarak 21. değer tahmin edilmeye çalışılmıştır. Bu yaklaşım, özellikle LSTM modeli için son derece uygun bir yaklaşım sağlamaktadır, çünkü LSTM, bu gruplamayı uzun süreli bellekte etkili bir şekilde saklamaktadır.

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu4.1.2 LSTM Modelinin Kurulması**

**Şekil 4.5** LSTM katmanlar

Modelde, her bir gizli katmanda, 128 nöron içeren toplam üç LSTM katmanı kullanılmıştır. Bu LSTM katmanlarında, modelin öğrenme yeteneklerini artırmak için hiperbolik tanjant (tanh) aktivasyon fonksiyonu kullanılmştır. Hiperbolik tanjant fonksiyonu, aktivasyon çıkışlarının sıfıra yakın olması ve çıkış aralığının -1 ile 1 arasında olması sayesinde, sinir ağı parametrelerinin daha dengeli güncellenmesine yardımcı olur, bu da eğitim süreci için avantajlıdır.

Yapay sinir ağlarında karşılaşılan temel bir sorun, aşırı öğrenme (overfitting) olabilir. Bu aşırı öğrenmeyi önlemek amacıyla, her LSTM katmanına %20'lik bir dropout eklenmiştir. Dropout, her iterasyonda belirli düğümleri rastgele kaldırarak modelin farklı özelliklere adapte olmasını teşvik eder. Bu sayede aynı değerlerin ezberlenmesi probleminin önüne geçilmiş olur.

**metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**Dense katmanı ise modelin çıkışını üretecek olan katmandır ve tek bir çıkış birimi bulunmaktadır. Bu, regresyon problemleri için yaygın bir yapıdır.

**Şekil 4.6** LSTM model

Şekil 4.6’dq ise "mean\_squared\_error" kayıp fonksiyonu ve "adam" optimize edici ile derlenen model, daha sonra X\_train ve y\_train verileri kullanılarak belirtilen epoch sayısı boyunca eğitilmiştir. epochs parametresi, veri setinin tamamının kaç kez geçileceğini belirler, batch\_size ise her eğitim adımında kullanılacak örnek sayısını tanımlar. Eğitim süreci boyunca, model, belirlenen kayıp fonksiyonu ile kaybını minimize etmeye çabalamaktadır.

**4.1.3 LSTM Değerlendirme**

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduKök Ortalama Kare Hata (RMSE), Ortalama Kare Hata (MSE)'nın karekökü alınarak hesaplanan bir metriktir. RMSE, hataların orijinal ölçeğiyle uyumlu bir şekilde ifade edilir ve MSE'nin büyüklüğünü daha iyi anlamamıza yardımcı olur. Düşük bir RMSE değeri, modelin daha iyi tahminler yaptığını ve daha az hata içerdiğini göstermektedir.

(4.1)

**Şekil 4.7.** LSTM için RMSE değerleri

Bu sonuçlara göre, modelin eğitim setindeki performansı oldukça iyi (RMSE: 1.22), yani model eğitim verilerine uyum sağlıyor ve bu veriler üzerinde başarılı tahminler yapıyor. Ayrıca, test setindeki (RMSE: 2.67) değer de düşük ancak eğitim setine kıyasla biraz daha yüksek. Bu durum, modelin yeni ve görünmeyen verilere karşı hala iyi bir genelleme yapabildiğini, ancak eğitim setine göre beklendiği üzere bir miktar performans kaybı olduğunu göstermektedir.

**4.2 Random Forest Yöntemi**

Bu bölümde, Random Forest yöntemi için verilerin hazırlanması, modelin oluşturulması, modelin performansının değerlendirilmesi başlıkları ele alınmıştır.

**4.2.1 Random Forest için Veri Setinin Hazırlanması**

Bu bölümde, finansal piyasaların analizi için sıkça kullanılan bazı göstergeleri ekleyerek Random Forest modelinin performansını artırmaya odaklanılmıştır. Bu özellikler arasında Göreli Güç Endeksi (RSI) ve Basit Hareketli Ortalama (SMA) bulunmaktadır. Bu göstergeler, modelin finansal piyasalardaki dinamikleri daha iyi anlamasına ve gelecekteki fiyat hareketlerini tahmin etmesine yardımcı olabilir.

Göreli güç endeksi (RSI) finans piyasaların analizinde kullanılan bir teknik indikatördür. 1978 yılında J. Welles Wilder tarafından geliştirilmiştir ve ilk kez isimli kitabında yayınlanmıştır. RSI bir momentum osilatörüdür ve basitçe aşırı alım - aşırı satım sinyalleri üretir [23]. 0 ve 100 arasında değerler alır, genellikle 30’un altındaki değerler aşırı satım iken 70’in üstündeki değerler aşırı alım olarak değerlendirilir.

, (4.2)

Basit hareketli ortalama (SMA), belirli bir zaman dilimindeki fiyatların ortalamasını alarak trendleri gösteren bir göstergedir. SMA, fiyat hareketlerini yumuşatarak trendleri daha net bir şekilde ortaya koyamaktadır.

(4.3)

**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**Bu göstergelere erişebilmek için pandas\_ta[24] teknik analiz kütüphanesi kullanılmıştır.

**Şekil 4.8** RSI ve SMA değerlerinin eklenmesi

İlk satırda, veri seti içindeki kapanış sütununu kullanarak 50 günlük SMA göstergesini hesaplayıp, sonuçları "sma50" adlı yeni bir sütuna ekleyen bir işlem gerçekleştirilmiştir. 50 günlük ortalama genellikle orta vadeli bir perspektf katmaktadır.

İkinci satırda, aynı şekilde kapanış sütununu kullanarak bu sefer 200 günlük SMA hesaplanıp ve sonuçlar "sma200" adlı sütuna eklenmiştir. Bu, daha uzun vadeli bir perspektifteki trendi gösteren bir gösterge olarak kullanılmaktadır.

Üçüncü satırda ise, yine kapanaış sütununu kullanarak 14 günlük RSI hesaplanıp ve sonuçlar "rsi" adlı yeni bir sütuna eklenmiştir. Finansal piyasalarda genellikle 5 işlem günü bulunmaktadır. 14 gün, iki hafta (10 işlem günü) artı iki hafta sonu tatilini içerir. Bu, analistlere her haftanın tam bir döngüsünü değerlendirme ve kapanış fiyatlarını metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldukullanma olanağı tanımaktadır.

**Şekil 4.9** Eğitim ve test verisinin ayrılması

Bağımlı değişken olarak "Close" sütununu (y) seçilmiştir. Bağımsız değişkenler ise "Close" ve "Adj Close" sütunları hariç diğer tüm sütunları içeren bir veri seti (X) olarak belirlenmiştir. Bu, modelin eğitilmesi için kullanılacak girdi özelliklerini oluşturmaktadır.

Veri seti, genellikle makineler arası modelin eğitilmesi ve performansının değerlendirilmesi için kullanılmak üzere iki bölüme ayrılmaktadır: eğitim seti ve test seti. Bu kod parçası, genel veri setinin %70'ini eğitim setine ve geri kalan %30'unu test setine atayarak bu bölümü gerçekleştirmektedir.

Eğitim seti, modelin öğrenme sürecini gerçekleştirmesi ve içsel parametrelerini ayarlaması için kullanılmaktadır. Test seti ise eğitilen modelin genel performansını değerlendirmek ve genelleme yeteneğini test etmek için kullanılmaktadır.

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu4.2.2 Random Forest Modelinin Kurulması**

**Şekil 4.10** Random Forest model

Şekil 4.10’daki kod parçası, random forest algoritması kurulmuştur ve en iyi değerlerini belirlemek amacıyla Grid Search Cross-Validation (GridSearchCV) yöntemini kullanılmıştır.

İlk olarak, basit bir random forest modeli oluşturulmuştur. Bu model, başlangıçta varsayılan değerlerini kullanmaktadır. Ardından, potansiyel değerler belirlenmiştir. Bu değerler, modelin performansını etkileyen max\_depth (ağaç derinliği), max\_features (maksimum özellik sayısı) ve n\_estimators (ağaç sayısı) gibi faktörleri içermektedir.

Grid Search işlemi, belirlenen değerler üzerinde deneme yaparak en iyi performansı sağlayan değerleri bulmaya çalışmaktadır. Bu süreç, 5 katlı çapraz doğrulama kullanılarak gerçekleştirilmiştir. Her bir iterasyonda veri seti beş parçaya ayrılır ve her parça sırayla test verisi olarak kullanılır, geri kalan kısımlar ise eğitim verisi olarak kullanılmaktadır.

Sonuç olarak, eğitim verisi üzerinde en iyi performansı gösteren değerler grid search tarafından belirlenmiştir. Bu değerler ise ağaç derinliği 5, maksimum özellik sayısı 7, ağaç sayısı 200 olmak üzere bulunup, ilgili modelde kullanılmıştır.

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu4.2.3 Random Forest Değerlendirme**

**Şekil 4.11** Random Forest için başarı değerlendirme metrikleri

Modelin eğitim hatası (RMSE: 1.94) ve test hatası (RMSE: 3.73) değerleri arasındaki fark göz önüne alındığında, modelin genel olarak iyi bir performans sergilediği söylenebilir. Düşük eğitim hatası, modelin eğitim verilerine uyum sağladığını gösterirken, makul düzeyde yüksek test hatası, modelin yeni verilerle de başarılı olduğunu ancak genelleme yeteneğinin eğitim hatasına göre düştüğünü gösterebilir.

R-kare skoru ise modelin bağımsız değişkenler tarafından açıklanan toplam varyansın oranını ölçmektedir. Bu değer genellikle 0 ile 1 arasında olup, 1'e ne kadar yakınsa, modelin verileri o kadar iyi açıkladığı anlamına gelmektedir. Bu durumda, R-kare skoru 0.95 olarak hesaplanmıştır. Yüksek bir R-kare skoru, modelin bağımsız değişkenlerle bağımlı değişken arasındaki ilişkiyi iyi açıkladığını göstermektedir. Bu durumda, modelin test setindeki verileri iyi açıkladığı söylenebilir.

**5 SONUÇ**

Bu tez, finansal piyasalarda makine öğrenimi kullanarak fiyat tahmini üzerine odaklanmıştır. Yatırımcıların ve finans profesyonellerinin karar alma süreçlerinde kullanabilecekleri etkili bir araç olarak makine öğrenimi modellerinin değerini incelemiştir. Çalışmanın temel amacı, farklı makine öğrenimi algoritmalarının finansal varlık fiyatlarını tahmin etme yeteneklerini karşılaştırmak ve bu modellerin tahmin konusunda ne kadar başarılı olabileceklerini anlamaktı.

Çalışmanın başlangıcında, literatürdeki araştırmalara dayanarak finansal piyasalardaki makine öğrenimi uygulamaları incelenmiştir. Ardından, LSTM ve Random Forest gibi önemli makine öğrenimi modellerini kullanarak google şirketinin 2013-12-01 ile 2023-12-01 tarihleri arasındaki veri seti kullanılarak fiyat tahminleri gerçekleştirilmiştir.

Bu tez kapsamında gerçekleştirilen çalışmalar sonucunda, LSTM ve Random Forest modelleri kullanılarak gerçekleştirilen finansal piyasa tahminlemede, LSTM modelinin daha etkili bir performans (Test RMSE: 2.67) sergilediği gözlemlenmiştir. LSTM modelinin eğitim ve test setlerinde elde ettiği düşük RMSE değerleri, modelin genel olarak iyi bir tahmin yeteneğine sahip olduğunu göstermektedir. Ayrıca, eğitim ve test setleri arasındaki RMSE değerleri farkı, modelin iyi bir genelleme yeteneğine sahip olduğunu göstermektedir. Öte yandan, Random Forest modelinin performansı (Test RMSE: 3.73) LSTM algoritmasına göre daha düşüktür ve eğitim ile test setleri arasındaki fark daha belirgindir. Bu durum, Random Forest modelinin tahmin yeteneğinin daha düşük olduğunu ve genelleme yeteneğinin LSTM modeline göre daha zayıf olduğunu göstermektedir.

Her iki modelde de eğitim aşamasındaki RMSE değerleri, test aşamasındaki değerlere göre düşüktür ve bu değerler seçilen enstrümana göre farklılık gösterecektir. Bu beklenen bir durumdur, eğer aradaki fark çok yüksek ise bu durum her iki modelde de yüksek performans düşüşleri yaşandığını veya aşırı öğrenme olduğunu göstermektedir. Bu bağlamda, modelin kurulma aşamasında kullanılan parametre değerlerinin seçiminin kritik bir öneme sahip olduğunu göz önünde bulundurmak gerekmektedir. Farklı parametre kombinasyonlarının denenmesi ve modelin genel performansının dikkatlice değerlendirilmesi, elde edilen sonuçların güvenilirliğini artırabilir.

Finansal piyasaların karmaşıklığı ve belirsizlikleri göz önüne alındığında, model performansının yanı sıra dönemsel olarak değişen ekonomik koşulların ve beklenmeyen olayların modeller üzerindeki etkilerini değerlendirmenin kritik bir öneme sahip olduğunu vurgulamak önemlidir. Finansal tahminlerde makine öğrenimi modellerinin kullanımının potansiyelini gösteren bu çalışma, aynı zamanda yatırımcıların karar alma süreçlerinde dikkate almaları gereken riskleri ve sınırlamaları vurgulamaktadır. Bu modellerin, yatırım kararlarına yardımcı olmak amacıyla kullanılmasının ötesinde, tek başına kesin çözümler sunmadığını anlamak önemlidir. Finansal piyasalardaki hızla değişen koşullar, siyasi gelişmeler, küresel olaylar ve diğer etmenler, modellerin performansını etkileyebilir. Bu nedenle, bu modellerin kullanıldığı finansal karar alma süreçlerinde, başka faktörler ve uzman değerlendirmeleriyle entegre edilmiş bir yaklaşım benimsemek kritik öneme sahiptir.

Sonuç olarak, bu tez, finansal piyasalarda makine öğrenimi ile tahminin önemli bir araştırma alanı olduğunu ve doğru bir şekilde uygulandığında değerli bilgiler sağlayabileceğini göstermiştir. Gelecekteki araştırmalarda, daha geniş veri setleri, farklı özellik mühendisliği yaklaşımları ve daha karmaşık modellerin kullanılması gibi konuların derinlemesine incelenmesi önerilebilir.

**KAYNAKLAR**

[1] ARSLANKAYA, S., & TOPRAK, Ş. (2021). Makine Öğrenmesi ve Derin Öğrenme Algoritmalarını Kullanarak Hisse Senedi Fiyat Tahmini. International Journal of Engineering Research and Development, 13(1), 178-192. https://doi.org/10.29137/umagd.771671

[2] GAVCAR, E., & METİN, H. M. (2021). Hisse Senedi Değerlerinin Makine Öğrenimi (Derin Öğrenme) ile Tahmini. Ekonomi Ve Yönetim Araştırmaları Dergisi, 10(2), 1-11.

[3] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock Closing Price Prediction Using Machine Learning Techniques. Procedia Computer Science.

[4] ÜNTEZ, A., & İPEK, M. (2022). Developing Financial Forecast Modeling With Deep Learning On Silver/Ons Parity. Journal of Advanced Research in Natural and Applied Sciences, 8(1), 35-44. https://doi.org/10.28979/jarnas.979429

[5] KOÇ USTALI, N., TOSUN, N., TOSUN, Ö. (2021). Makine Öğrenmesi Teknikleri ile Hisse Senedi Fiyat Tahmini. Eskişehir Osmangazi Üniversitesi İİBF Dergisi, 16(1), 1 – 16.

[6] G. ŞİŞMANOĞLU, F. KOÇER, M. A. ÖNDE, O. K. SAHİNGOZ, Derin Öğrenme Yöntemleri ile Borsada Fiyat Tahmini, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, c. 9, sy. 1, ss. 434–445, 2020, doi: 10.17798/bitlisfen.571386.

[7] Patel, Jigar, Shah, Sahil, Thakkar, Priyank, Kotecha, K. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques, Expert Systems with Applications, 42, 1, 2015, s.259 – 268.

[8] <https://maddevs.io/blog/what-is-semi-supervised-learning/>, 01.11.2023 tarihinde erişildi

[9] <https://www.amygb.ai/blog/unsupervised-learning-in-image-classification>, 01.11.2023 tarihinde erişildi

[10] <https://www.linkedin.com/pulse/mastering-art-learning-comprehensive-guide-jagarlapoodi/>, 01.11.2023 tarihinde erişildi

[11] <https://www.beyaz.net/tr/yazilim/makaleler/derin_ogrenme_deep_learning_nedir.html>, 01.11.2023 tarihinde erişildi

[12] <https://medium.com/@Onurryilmaz/yapay-sinir-a%C4%9Flar%C4%B1-aktivasyon-fonksiyonlar%C4%B1-d964805b189b>, 01.11.2023 tarihinde erişildi

[13] <https://ayyucekizrak.medium.com/derin-%C3%B6%C4%9Frenme-i%C3%A7in-aktivasyon-fonksiyonlar%C4%B1n%C4%B1n-kar%C5%9F%C4%B1la%C5%9Ft%C4%B1r%C4%B1lmas%C4%B1-cee17fd1d9cd>, 30.11.2023 tarihinde erişildi

[14] https://mesutpiskin.com/blog/yapay-sinir-agi-derin-ogrenme.html, 30.11.2023 tarihinde erişildi

[15, 16, 17] <http://buyukveri.firat.edu.tr/2018/04/16/derin-ogrenme-yapay-sinir-aglari-3/>, 12.12.2023 tarihinde erişildi

[18] https://www.python.org/, 01.12.2023 tarihinde erişildi

[19] <https://jupyter.org/>, 01.12.2023 tarihinde erişildi

[20] https://finance.yahoo.com/, 02.12.2023 tarihinde erişildi

[21] https://medium.com/deep-learning-turkiye/uzun-k%C4%B1sa-vadeli-bellek-lstm-b018c07174a3, 30.11.2023 tarihinde erişildi

[22] <https://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi/>, 15.12.2023 tarihinde erişildi

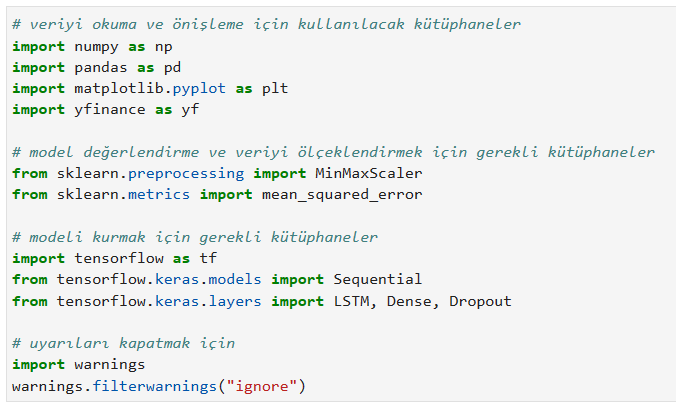
[23] <https://tr.wikipedia.org/wiki/G%C3%B6reli_g%C3%BC%C3%A7_endeksi>, 19.12.2023 tarihinde erişildi

[24] <https://github.com/twopirllc/pandas-ta>, 21.12.2023

**EKLER**

Bu kısımda proje boyunca yazılan tüm kodlar verilmiştir.

LSTM için yazılan kodlar:

****

**metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

****

**metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

Random Forest için yazılan kodlar:

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, yazı tipi, çizgi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**ÖZGEÇMİŞ**

Ad Soyad İlker EFİL

Doğum Tarhi 03.01.2001

Doğum Yeri Ordu

Lise Ankara Anadolu Lisesi (2015-2018)

Uğur Temel Lisesi (2018-2019)