Bilgisayar Bilimleri ve Mühendisliği Dergisi

Cilt-15
Sayı-1
2022

Hisse Senedi Tahmininde Karşılaşılan Veri Dengesizliği Problemi için Yeni Bir Kural Tabanlı Yaklaşım ve 2D-CNN Modeli

A New Rule-Based Approach for Encountered Data Imbalance Problem in Stock Prediction and 2D-CNN Model

Zinnet Duygu AKŞEHİR
Ondokuz Mayıs Üniversitesi
Bilgisayar Mühendisliği Bölümü
Samsun, Türkiye
duygu.aksehir@bil.omu.edu.tr
ORCID: 0000-0002-6834-6847

Erdal KILIÇ
Ondokuz Mayıs Üniversitesi
Bilgisayar Mühendisliği Bölümü
Samsun, Türkiye
erdal.kilic@bil.omu.edu.tr
ORCID: 0000-0003-1585-0991

Öz

Bu çalışmada literatürdeki borsa tahmini kapsamında son yıllarda yapılan çalışmalar detaylı bir şekilde incelenmiştir. İncelenen çalışmalar doğrultusunda evrişimsel sinir ağları (CNN) modelinin borsa tahmini alanına uyarlandığı ve başarılı sonuçlar verdiği gözlemlenmiştir. Bu kapsamda Dow30 endeksinde yer alan hisse senetlerinin bir gün sonraki pozisyonunu (al, sat, tut) tahmin etmek için 2D-CNN tabanlı bir model kullanılmıştır. Bu model için hisse senedi kapanış fiyatları, teknik göstergeler, altın fiyatı, altın oynaklık endeksi, petrol fiyatı ve petrol oynaklık endeksi verileri kullanılarak görüntü tabanlı girdi değişken kümesi oluşturulmuştur. Ayrıca bu çalışmada veri dengesizliği problemini çözmek için yeni bir kural tabanlı etiketleme algoritması önerilmiş ve buna ek olarak elde edilen görüntüler üzerinde döndürme işlemi gerçekleştirilmiştir. Kaydırmalı eğitim-test yaklaşımını kullanan CNN modelinin tahmin performansı literatürdeki diğer çalışmalarla kıyaslanmıştır. Deney sonuçları, veri dengesizliği problemini gidermek için önerilen yaklaşımın CNN modeli ile birlikte kullanıldığında diğer CNN tabanlı çalışmalardan daha yüksek başarı sağladığını göstermiştir. Ayrıca önerilen bu yaklaşımın, modelin tahmin performansını literatürdeki aynı amaçla önerilen Chen ve Huang'ın yaklaşımından daha fazla iyileştirdiği gözlemlenmiştir.

Gönderme, düzeltme ve kabul tarihi: 14.02.2022 - 18.02.2022 - 28.02.2022
Makale türü: Araştırma

Anahtar sözcükler: Hisse senedi tahmini, Evrişimsel sinir ağları, Etiketleme, Veri dengesizliği

Abstract

In this study, studies of stock market forecasting in the literature were examined in detail. We observed that the convolutional neural networks (CNN) model was adapted to stock market forecasting and gave successful results. Considering the studies in the literature, a 2D-CNN-based model was proposed to predict the next day's trade action (buy, sell, hold) of stocks in the Dow30 index in this study. Image-based input variable set was created using the closing price of stocks, technical indicators, gold price, gold volatility index, oil price, and oil volatility index data for this model. In addition, a new rule-based labeling algorithm is proposed to solve the data imbalance problem, and the rotation process is performed on the obtained images. The prediction performance of the CNN model using the sliding-window training-test approach was compared with other studies in the literature. The experimental results showed that the proposed approach in this study to solve the data imbalance problem achieved higher success than other CNN-based studies when used with the CNN model. In addition, it has been observed that our proposed approach improves the model's prediction performance more than the approach proposed by Chen and Huang for the same purpose in the literature.

Keywords: Stock prediction, Convolutional neural networks, Labeling, Data imbalance

1. Giriş

Borsa tahmini hem ekonomi hem de bilgisayar bilimleri alanında ilgi çekici konular arasında yer almaktadır. Bu ilginin asıl nedeni ise oluşturulan modeller ile başarılı tahminlerin gerçekleştirilmesi ve bu durumun yatırımcıların kârlılığını arttıracağı beklentisidir. Fakat geçmiş borsa verilerinden yararlanarak bir model oluşturmak ve bu model üzerinden borsanın gelecekteki davranışını tahmin etmek oldukça zordur. Bunun sebebi olarak borsaların rastgele yürüyüş özelliğine sahip olması, başka bir deyişle zamanın her noktasında bağımsız olarak hareket etmesi gösterilebilir [1]. Ayrıca hisse senedi piyasası olarak da bilinen borsa; siyasi olaylar, şirket politikaları, genel ekonomik durum, yatırımcı beklentileri/psikolojileri, diğer borsa hareketleri gibi çok fazla faktörden etkilenmekte olup kaotik bir yapıya sahiptir. Hisse senedi piyasasındaki bu yüksek belirsizlik de borsa/hisse senedi tahminini zorlaştırmaktadır [2].

Borsa tahmini, bir dizi geçmiş verilerden yola çıkarak gelecekteki olası yönü veya fiyat değerini tahmin eden bir zaman serisi problemi olarak ele alınmaktadır. Bu kapsamda zaman serisi analiz yöntemlerini kullanan borsa/hisse senedi tahmin çalışmaları incelendiğinde önerilen modellerde genellikle hisse senedi veya endeks değerinin belirli bir zaman aralığındaki fiyat verilerinin (açılış, kapanış, gün içi en yüksek, gün içi en düşük, hacim) kullanıldığı görülmüştür. Bu çalışmalarda önerilen modellerin performansını değerlendirmek için literatürde yer alan çeşitli derin öğrenme ve makine öğrenmesi yöntemlerini kullanan yaklaşımlar ile karşılaştırmalar yapılmıştır [3-5]. Yapılan başarım değerlendirmelerinde genellikle makine öğrenmesi ve derin öğrenme tabanlı yaklaşımların istatistiksel yaklaşımlardan daha başarılı tahminler gerçekleştirdiği raporlanmıştır [3, 5]. Bunun sebebi olarak borsa verilerinin çok fazla faktörden etkilenmesi ve doğrusal olmaması bu tahmini geleneksel zaman serileri yöntemleri ile doğru-etkili bir şekilde gerçekleştirmeyi zorlaştırması gösterilebilir.

Zaman serileri analiz yöntemleri dışında destek vektör makineleri, karar ağaçları, yapay sinir ağları gibi çeşitli makine öğrenmesi yöntemleri [4-7] de kullanılmaktadır. Vijh vd., makine öğrenmesi yöntemlerinden yapay sinir ağları (ANN) ve rassal orman yöntemlerini kullanarak New York borsasında yer alan beş hisse senedinin bir gün sonraki kapanış fiyatını tahmin etmişlerdir. Önerdikleri tahmin modeli için hisse senedi fiyat verileri ile haraketli ortalama ve standart sapma değerlerinden yararlanmışlardır. Her iki yöntemle oluşturulan tahmin modellerinin performansı karşılaştırıldığında ANN modelinin daha başarılı olduğu gözlemlenmiştir [7].

Zaman serileri ve makine öğrenmesi yöntemlerinin aksine son yıllarda derin öğrenme yöntemleri [8-17] borsa/hisse senedi tahmini kapsamındaki etkinliğini kanıtlamıştır. Bu çalışmalarda genellikle derin öğrenme yöntemlerinden uzun kısa dönemli bellek ağları (LSTM) ve evrişimsel sinir ağları (CNN) yöntemleri kullanılarak tahmin modelleri oluşturulmuştur. Oluşturulan tahmin modellerinde girdi değişkeni olarak çoğunlukla hisse senedi/borsa endeks geçmiş fiyat verileri ve teknik göstergeler kullanılmıştır. Sezer ve Özbayoğlu, Dow30 hisse senetleri ve borsa yatırım fonunun (ETF) bir gün sonraki pozisyonunu tahmin etmek için 2D-CNN tabanlı bir model önermiş ve bu modelin girdi değişkenlerinde teknik göstergeler kullanmışlardır [10]. Sim vd., S&P500 endeksini tahmin etmek için girdi değişkenlerinde endeksin kapanış fiyatı ve farklı teknik göstergeler kullanan dört tane CNN modeli önermişlerdir. Önerilen modellerin performansı değerlendirildiğinde teknik göstergelerin kapanış fiyatına benzer davranışlarda bulunduğu için teknik gösterge kullanımının model üzerinde olumlu bir etki yaratmadığı belirtilmiştir [11]. Chen ve Huang ise S&P500 endeksinin yönünü ve pozisyonunu (al/sat/tut) tahmin etmek için CNN ile LSTM modellerinin girdi değişkenlerinde Sim vd. 'den farklı olarak altın, altın oynaklık endeksi, petrol ve petrol oynaklık endeksi kapanış fiyat verilerini kullanmışlardır. Modelin performansı değerlendirildiğinde ise girdi değişkenlerinde altın, altın oynaklık endeksi, petrol ve petrol oynaklık endeksi kapanış fiyatlarını ele almanın tahmin performansını arttırdığı belirtilmiştir [13]. Dolayısıyla derin öğrenme yöntemleri borsa tahmini için umut verici bir çözüm olmuştur.

Bu çalışma kapsamında Dow30 endeksinde yer alan hisse senetlerinin bir gün sonraki pozisyonlarını tahmin etmek için 2D-CNN tabanlı bir model [10] kullanılmıştır. Modelde özellik vektörleri oluşturulurken [13] numaralı çalışmada önerilen girdi değişkenleri ele alınmış ve modelin başarımını arttırmak için de yeni bir kural tabanlı etiketleme algoritması önerilmiştir.

1.1 Motivasyon ve Katkılar

Yapılan literatür taraması sonucunda son yıllarda CNN tabanlı hisse senedi fiyat/borsa endeks tahmini çalışmalarında önerilen tahmin modellerinin başarılı sonuçlar elde ettiği gözlemlenmiştir. Ancak CNN tabanlı çalışmalara baktığımızda her ne kadar sonuçlar tatmin edici görünse de halen düzeltilmesi gereken veri dengesizliği problemi için yeni yaklaşımlara ihtiyaç olduğu görülmektedir. Bu çalışmanın temel motivasyonu hisse senedi tahmini kapsamında kullanılan veri kümelerindeki dengesizliği gidermek için farklı bir çözüm yöntemi geliştirmektir.

Bu çalışmada Dow30 endeksinde yer alan hisselerin bir gün sonraki hisse ile ilgili alınacak pozisyonu (al/sat/tut) tahmin etmek için 2D-CNN tabanlı bir model [10] kullanılmıştır. Veri dengesizliğini gidermek için geliştirilen algoritmanın başarısı bu model üzerinde test edilmiştir. Çalışmaya yapılan katkılar aşağıda verilmiştir:

- İncelenen literatür çalışmaları sonucunda tahmin modeli için sadece teknik göstergelerin kullanmanın model performansına olumlu bir katkı sağlamadığı görülmüştür. Dolayısıyla tahmin modelinde teknik göstergelere ek olarak [13] çalışmasında önerilen altın, altın oynaklık endeksi, petrol ve petrol oynaklık endeksi kapanış fiyat verileri de kullanılmıştır.
- Chen ve Huang'ın S&P500 endeksini tahmin etmek için [10] numaralı çalışmada önerilen CNN modeli

kullanmış olup bu çalışmada da Dow30 hisse senetlerinin bir gün sonraki pozisyonuna karar vermek için aynı modelden yararlanılmıştır.

- Veri dengesizliğini gidermek için literatürdeki yaklaşımlardan farklı olarak kural tabanlı yeni bir algoritma önerilmiştir.
- Veri dengesizliğine neden olan al/sat/tut etiketli verileri tam anlamıyla dengeleyebilmek için ilk olarak önerilen kural tabanlı algoritma uygulanmış, ardından al ve sat etiketli verilerden oluşturulan resimler saat yönünde 30 derecelik açıyla döndürülmüştür. Böylece al/sat/tut verilerinin sayısı yaklaşık olarak eşitlenmiştir.

1.2. Organizasyon

Çalışmanın geri kalanı şu şekilde organize edilmiştir: Bölüm 2'de kullanılan CNN modeli için veri kümesinin nasıl oluşturulduğu ve bunların görüntülere nasıl dönüştürüldüğü detaylandırılmıştır. Bölüm 3'te uygulamada kullanılan model hakkında detaylı bilgi verilmiştir. Bölüm 4'te elde edilen sonuçlar tablolar halinde verilerek yorumlanmıştır. Son bölümde ise sonuçlar tartışılarak gelecek çalışmalar değerlendirilmiştir.

2. Veri Kümesi

Bu bölümde çalışmada kullanılan 2D-CNN tahmin modeli için seçilen özniteliklerin neler olduğu, veri kümesinin ve görüntülerin nasıl oluşturulduğu detaylandırılmıştır.

2.1. Özniteliklerin Belirlenmesi

Yapılan literatür araştırmaları sonucunda borsa endeksi/hisse senedi tahmin çalışmalarında önerilen modeller için genellikle geçmiş fiyat verileri ile birlikte teknik analiz göstergelerinin kullanıldığı görülmektedir. Fakat son yıllarda yapılan çalışmalarda ise teknik göstergelerinin kullanımının model üzerinde olumlu bir etki yaratmadığı ve tahmin edilecek olguyu etkileyen faktörlerin iyi analiz edilmesi gerektiği belirtilmiştir [11]. Bu kapsamda modelin girdi değişkenlerinde altın, petrol gibi varlıkların fiyat verilerinin ele alınması önerisi Sim vd. tarafından ortaya atılmış ve [13] numaralı çalışmada bu öneri değerlendirilmiştir. Bu çalışmada ise Chen ve Huang'ın S&P500 tahmin modeli için girdi değişkenlerinde kullandığı 14 öznitelik ele alınmıştır. Dolayısıyla girdi değişkenlerinde hisse senetlerinin kapanış fiyatları, bu fiyat verilerinden hesaplanan 9 teknik gösterge (SMA, EMA, ROC, MACD, Fast %K, Slow %D, Upper band, Lower band, %B), altın fiyatı, altın oynaklık endeksi, petrol fiyatı ve petrol oynaklık endeksi günlük kapanış fiyat verileri kullanılarak Dow30 hisselerinin bir gün sonraki pozisyonu 2D-CNN tabanlı bir model ile tahmin edilmiştir.

2.2. Veri Kümesinin Oluşturulması

Bu çalışma kapsamında, Dow30 endeksinde yer alan hisse senetlerinin 4 Haziran 2008 – 31 Aralık 2021 tarihleri arasındaki işlem günlerini kapsayan günlük fiyat verileri (açılış, kapanış, gün içi en yüksek, gün içi en düşük ve işlem hacmi) finance.yahoo.com sitesinden elde edilmiştir. Bu günlük fiyat verilerinden 9 teknik gösterge değeri hesaplanmıştır. Ayrıca altın fiyatı, altın oynaklık endeksi, petrol fiyatı ve petrol oynaklık endeksinin kapanış fiyat verileri de aynı site üzerinden elde edilmiştir.

Verilerin elde edilmesinin ardından kullanılan kaydırmalı eğitim-test yaklaşımı için eğitim ve test kümeleri oluşturulmuştur. Bu veri kümeleri oluşturulurken ilk beş yıl eğitim, ardından gelen bir yıllık periyot ise test süreci için seçilmiştir. Örneğin, 2008-2012 yılını kapsayan veriler eğitim sürecinde kullanılırken 2013 yılına ait bir yıllık veriler ise modelin test işlemi için kullanılmıştır. Daha sonra bu periyot değerleri bir yıl şeklinde kaydırılmıştır (eğitim: 2009-2013, test: 2014). Dolayısıyla bu kaydırmalı yaklaşımla dokuz eğitim-test kümesi oluşturulmuştur.

2.3. Hisse Senetlerinin Pozisyonlarının Belirlenmesi

Tahmin modeli için veri kümesi oluşturulduktan sonra Sezer ve Özbayoğlu'nun [10] önerdiği etiketleme algoritması temel alınmış ve hisse senetlerinin her bir işlem günü "Al", "Sat" ve "Tut" olarak etiketlenmiştir. Önerilen bu algoritmada 11 günlük kayan pencere yaklaşımı kullanılmıştır. Bu yaklaşımda 11 günlük pencerenin orta noktasına, başka bir deyişle 6. günün kapanış fiyatına göre pozisyon etiketlemesi gerçekleştirilmiştir. Eğer 6. günün kapanış fiyatı bu 11 günlük pencerenin maksimum değeri ise "Sat", minimum değeri ise "Al", diğer durumlarda ise "Tut" olarak etiketlenmiştir.

Veri kümesinin Sezer ve Özbayoğlu'nun önerdiği etiketleme algoritması ile etiketlenmesinin ardından kaydırmalı eğitim-test yaklaşımında kullanılacak eğitim ve test kümeleri incelenmiştir. Her bir eğitim ve test kümesinde "Tut" olarak etiketlenen verilerin sayısının "Al" ve "Sat" olarak etiketlenenlerin yaklaşık 15 katı olduğu gözlemlenmiştir. Bu durum bir veri dengesizliği problemine işaret etmektedir. Bu problemi çözmek için [13] numaralı çalışmadan farklı olarak Sezer ve Özbayoğlu'nun önerdiği etiketleme algoritmasını temel alan ve Algoritma 1'de verilen kural tabanlı yeni bir etiketleme algoritması bu çalışmada önerilmiştir. Önerilen bu yeni etiketleme algoritması kullanıldığında her bir eğitim ve test kümesinde "Tut" olarak etiketlenen verilerin sayısının "Al" ve "Sat" olarak etiketlenenlerin yaklaşık 1,8 katına gerilediği gözlemlenmiştir. Dolayısıyla Sezer ve Özbayoğlu'nun önerdiği etiketleme algoritması ile karşılaşılan veri dengesizliği problemi burada büyük oranda giderilmiştir.



Al Tut Sat

Şekil-1: "Al", "Tut" ve "Sat" etiketine sahip örnek görüntüler

2.4. Görüntülerin Oluşturulması

Tahmin modeli için girdi değişken kümesinin oluşturulmasının ardından her bir işlem gününe ait görüntüler elde edilmiştir. Önerilen yeni etiketleme algoritmasına göre hisse senedinin bir gün sonraki pozisyonunu belirlemek için çeşitli denemeler sonucunda geçmiş beş günü ele almanın daha başarılı

olduğunu gözlemlenmiştir. Hisse senedinin pozisyonunu tahmin etmek için geçmiş beş işlem gününe ait veriler birleştirilmiş ve Algoritma 1 kullanılarak etiketlenmiştir. Ayrıca algoritmanın akış diyagramı Şekil-2'de verilmiştir. Dolayısıyla veri kümesinde her bir işlem günü için 5x14'lük matrisler elde edilmiş ve bunlar gri seviyeli görüntülere dönüştürülmüştür. Şekil-1'de elde edilen örnek görüntüler verilmiştir. Ayrıca "Tut" etiketine sahip verilerin "Al" ve "Sat"'tan yaklaşık olarak 1,8 kat daha fazla olmasından dolayı veri eşitliğini tam olarak sağlayabilme yoluna gidilmiştir.

Algoritma 1 Kural Tabanlı Etiketleme Algoritması

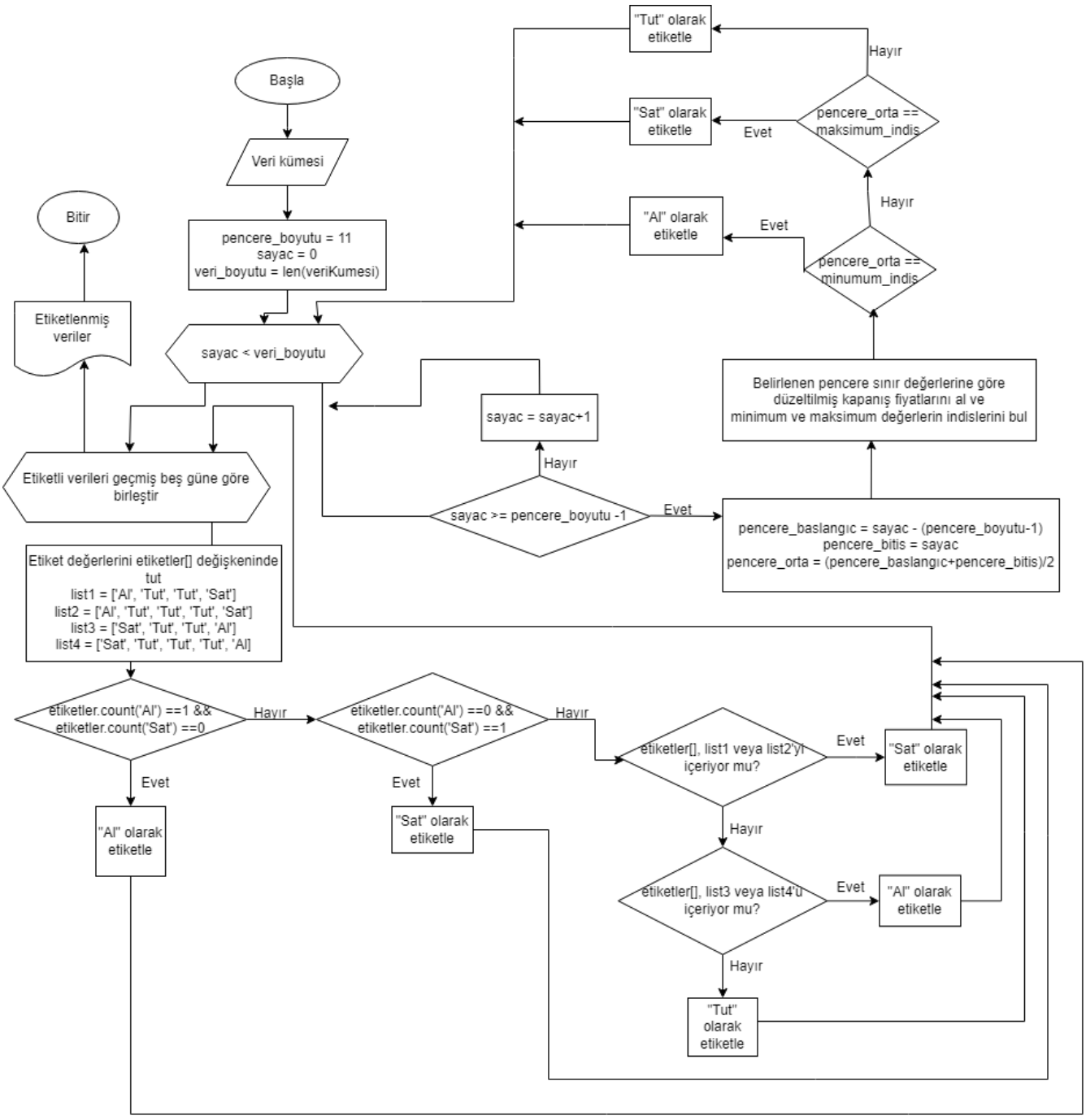
```
pencereBoyutu = 11
sayac = 0
veriBoyutu = len(veriKumesi)
while sayac < veriBoyutu do
    if sayac >= pencereBoyutu − 1 then
        pencereBaslangıcı = sayac − (pencereBoyutu − 1)
        pencereBitisi = sayac
        pencereOrtası = (pencereBaslangıcı + pencereBitisi)/2
        for (i = pencereBaslangıcı; i <= pencereBitisi; i + +) do
            kapanısFiyatı = veriKumesi.iloc[i]['AdjClose']
            if kapanısFiyatı < minimum then
                minimum = kapanısFiyatı
                minimumIndex = i
            else if kapanısFiyatı > maksimum then
                maksimum = kapanısFiyatı
                maksimumIndex = i
            end if
        end for
        if maksimumIndex == pencereOrtası then
            etiket = "Sat"
        else if minimumIndex == pencereOrtası then
            etiket = "Al"
        else if  then
            etiket = "Tut"
        end if
    end if
    sayac+ = 1
end while
etiketler[] ← geçmiş beş günün etiket değerleri
list1 = ['Al', 'Tut', 'Tut', 'Sat']
list2 = ['Al', 'Tut', 'Tut','Tut', 'Sat']
list3 = ['Sat', 'Tut', 'Tut', 'Al']
list4 = ['Sat', 'Tut', 'Tut','Tut', 'Al']
if ((etiketler.count('Al') == 1)&(etiketler.count('Sat') == 0)) then
    etiket = 'Al'
else if ((etiketler.count('Sat') == 1)&(etiketler.count('Al') == 0)) then
    etiket = 'Sat'
else if  ((is_sublist(list1, etiketler)|is_sublist(list2,etiketler) ) then
    etiket = 'Sat'
else if  ((is_sublist(list3, etiketler)|is_sublist(list4,etiketler))  then
    etiket = 'Al'
else
    etiket = 'Tut'
end if
```

Bu amaçla "Al" ve "Sat" etiketine ait görüntüler saat yönünde 30 derecelik açıyla döndürülmüştür. Böylece her bir etiketten eşit sayıda görüntü oluşturulmuş ve bu sayede veri dengesizliği problemi tam anlamıyla giderilmiştir.

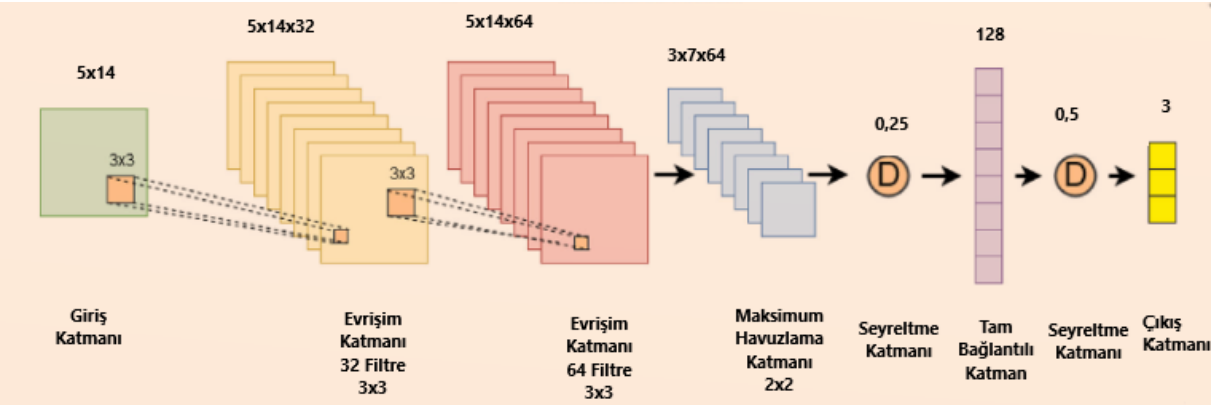
3. Yöntem

Bu çalışmada, birebir karşılaştırma yapabilmek için Şekil 3'te verilen [10, 13] numaralı çalışmalardaki CNN mimarisi benimsenmiş ve bu model oluşturduğumuz görüntü kümesine göre uyarlanmıştır. Kullanılan CNN mimarisi giriş, iki evrişim, bir maksimum havuzlama, iki seyreltme, bir tam bağlantılı ve çıkış katmanı olmak üzere toplam sekiz katmandan oluşmaktadır. Katlama işleminin gerçekleştirildiği ilk iki evrişim katmanı için filtre boyutu 3x3 ve filtre sayısı ise sırasıyla 32 ve 64 olarak seçilmiştir. Bu katmanlardan sonra, görüntü boyutunu azaltmak için 2x2 filtre boyutuna sahip maksimum havuzlama katmanı kullanılmıştır. Ayrıca aşırı ezberlemenin önüne geçmek için seyreltme oranları sırasıyla 0,25 ve 0,5 olarak belirlenen seyreltme katmanları eklenmiştir. Bu katmanlardan sonra elde edilen matris verisini vektör haline dönüştürmek için tam bağlantılı katman kullanılmıştır. Son olarak sınıflandırma işlemini gerçekleştirmek için çıkış katmanı ve bu katmanda da softmax aktivasyon fonksiyonu kullanılmıştır.

Bölüm 2'de anlatılan görüntü oluşturma yöntemi kullanılarak elde edilen resimler CNN modeline girdi olarak verilmiş ve hisse senetlerinin bir gün sonraki pozisyonu tahmin edilmiştir. Elde edilen sonuçlar bir sonraki bölümde tartışılmış ve modelin başarısı istatistiki yöntemler kullanılarak değerlendirilmiştir. Ayrıca önerilen etiketleme algoritmasının etkinliği de bu sayede değerlendirilmiştir.



Şekil-2: Kural tabanlı etiketleme algoritmasının akış diyagramı



Şekil-3: CNN modeli4. Deneysel Sonuçlar ve Tartışma

4. Deneysel Sonuçlar ve Tartışma

Belirlenen özellik vektörleri ve önerilen etiketleme algoritması ile Dow30 endeksindeki hisse senetleri için 5x14 boyutunda beş yıllık eğitim kümesi için yaklaşık 50.000, bir yıllık test kümesi için ise 10.600 görüntü oluşturulmuştur. Bu görüntüler ile kaydırmalı eğitim-test yaklaşımını kullanan CNN modeli eğitilerek test edilmiştir. Literatürdeki çalışmalarla birebir karşılaştırma yapmak için hisse senetlerinin son 10 yıllık verilerini (yaklaşık 112.000 görüntü) de içeren test kümesi oluşturulmuştur. Bu veri kümesi kullanılarak daha önce eğitilmiş CNN modelinin performansı değerlendirilmiş ve literatürdeki diğer çalışmalarla [10, 13] karşılaştırılmıştır.

Modelin başarısını değerlendirmek için literatürde sıklıkla kullanılan doğruluk, anma, kesinlik ve F1 skor metrikleri kullanılmıştır. Bu metrikler Çizelge-1'de örnek olarak verilen karmaşıklık matrisindeki ifadelerden yararlanarak hesaplanmakta olup sırasıyla 1-4 numaralı denklemlerde verilmiştir.

Çizelge-1: Karmaşıklık matrisi

		Tahmin Edilen Sınıf Etiketleri	
		Negatif	*Pozitif*
Gerçek Sınıf Etiketleri	*Negatif*	Doğru Negatif (DP)	Yanlış Pozitif (YP)
	Pozitif	Yanlış Negatif (YN)	Doğru Pozitif (DP)

$$Doğruluk = \frac{DN + DP}{DN + YP + YN + DP} \quad (1)$$

$$Anma = \frac{DP}{DP + YN} \quad (2)$$

$$Kesinlik = \frac{DP}{DP + YP} \quad (3)$$

$$F1\ skor = \frac{2 \times anma \times kesinlik}{anma + kesinlik} \quad (4)$$

2D-CNN modeli için elde edilen karmaşıklık matrisi ve istatistiksel metriklere ait sonuçlar sırasıyla Çizelge-2 ve 3'te verilmiştir. Sonuçlar incelendiğinde önerilen yöntemin doğruluk değerinin %70 olduğu görülmekte ve bu model hisse senedinin bir gün sonraki pozisyonunu yüksek doğrulukta tahmin etmektedir. Çizelge-3 incelendiğinde modelin anma ve kesinlik değerlerinin dengeli olduğu bunun da F1 skor değerini arttırdığı görülmektedir. Bu, hisse senedi pozisyon tahmininde istenilen bir durumdur.

Çizelge-2: Test veri kümesinin karmaşıklık matrisi

		Tahmin Değerleri		
		Al	*Sat*	*Tut*
Gerçek Değerler	*Al*	*26928*	*7635*	*3757*
	Sat	*3914*	*26613*	*4279*
	Tut	*7379*	*6949*	*24774*

Çizelge-3: Test veri kümesinin değerlendirilmesi

	Al	**Sat**	**Tut**
Anma	0,70	0,63	0,76
Kesinlik	0,70	0,76	0,65
F1 skor	0,70	0,69	0,70

Literatürdeki benzer çalışmalarla yapılan karşılaştırma sonuçları Çizelge-4'te verilmiştir. Bu sonuçlar irdelendiğinde Sezer ve Özbayoğlu'nun çalışmasındaki veri dengesizliği nedeniyle elde edilen anma-kesinlik değerleri arasındaki büyük farkların düşük F1 skor değerine sebep olduğu dikkat çekmektedir. Bu değeri iyileştirmek için bu çalışmada önerdiğimiz kural tabanlı yaklaşımın Chen ve Huang'ın yaklaşımından daha iyi performans gösterdiği görülmektedir. Tablodaki değerler dikkatli bir şekilde incelendiğinde [13] numaralı çalışmadaki doğruluk değerinin, önerilen yöntemden daha yüksek olduğu gözlemlenmektedir. Fakat, bu durum modelin tahmin performansını değerlendirmede tek başına bir anlam ifade etmeyip F1 skor değerine bakmak gerekmektedir. Zira kesinlik ve anma değeri arasındaki dengesizlik modelin tahmin yeteneğinin her ne kadar doğruluk yüksek olsa da kısıtlı olduğunu göstermektedir. Bu durum da F1 skor değerlerine yansımıştır. Aynı şeyleri [10] çalışması için de söylemek mümkündür.

[10] ve [13] numaralı çalışmaların aksine bu çalışmada önerilen yöntemle elde edilen kesinlik ve anma değerlerinin

birbirine çok yakın olduğu görülmektedir. Dolayısıyla bu durum F1 skor değerlerini iyileştirmiş ve modelin tahmin performansının diğer çalışmalara göre daha başarılı olmasını sağlamıştır.

Çizelge-4: Önerilen tahmin modelinin performansının diğer modellerle karşılaştırılması

	CNN8 [13]	*CNN-TA [10]*	*Önerilen model*
Doğruluk	**%78**	%58	%70
Kesinlik (Al)	0,30	0,22	**0,70**
Anma (Al)	**0,92**	0,80	0,70
F1 skor (Al)	0,45	0,34	**0,70**
Kesinlik (Tut)	**0,98**	0,95	0,65
Anma (Tut)	0,77	0,55	0,76
F1 skor (Tut)	**0,86**	0,70	0,70
Kesinlik (Sat)	0,29	0,18	**0,76**
Anma (Sat)	0,77	**0,81**	0,63
F1 skor (Sat)	0,43	0,29	**0,69**

5. Sonuç ve Gelecek Çalışmalar

Bu çalışma kapsamında derin öğrenme yöntemleri ile oluşturulan modellerin finansal piyasa verileri üzerinde uygulanabilirliği araştırılmış ve bu doğrultuda Dow30 endeksindeki hisse senetlerinin bir gün sonraki pozisyon değerini (al, sat, tut) tahmin etmek için 2D-CNN tabanlı bir yaklaşım kullanılmıştır. Literatürdeki çalışmalar incelendiğinde veri kümeleri oluşturulurken karşılaşılan en önemli problemlerden bir tanesi veri dengesizliği problemi olduğu tespit edilmiştir. Karşılaşılan bu problem bu çalışmada önerilen kural tabanlı bir yaklaşımla büyük oranda giderilmiştir. Etiketli verilerin sayısını kendi içinde tam olarak eşitleyebilmek için basit bir manipülasyonla "Al" ve "Sat" etiketli görüntülere saat yönünde 30 derecelik döndürme işlemi uygulanarak bunlar diğer görüntülerle birleştirilmiştir. CNN modelinin girdi değişkenleri için [13] numaralı çalışmada belirtilen hisse senedi kapanış fiyatları, 9 teknik gösterge, altın fiyatı, altın oynaklık endeksi, petrol fiyatı ve petrol oynaklık endeksi olmak üzere toplam 14 öznitelik kullanılmıştır. Oluşturulan veri kümesi kaydırmalı eğitim-test yaklaşımı için belirlenen periyot değerlerine göre (eğitim için 5, test için 1 yıl) eğitim ve test kümelerine ayrılmıştır. Ardından bu veri kümeleri kullanılarak CNN modeli eğitilmiştir. Ayrıca literatürdeki çalışmalarla [10, 13] birebir karşılaştırma yapmak için eğitilmiş CNN modelinin performansı hisse senetlerinin son 10 yıllık verileri üzerinde test edilmiştir. Yapılan karşılaştırmalar sonucunda önerilen kural tabanlı etiketleme algoritmasının modelin tahmin performansını iyileştirdiği görülmüştür. Veri dengesizliği problemini gidermek için önerdiğimiz yaklaşımın [13] çalışmasında önerilen yaklaşımdan daha iyi performans gösterdiği tespit edilmiştir.

Gelecek çalışma olarak hisse senetlerinin fiyatlarını tahmin etmek amacıyla geliştirilen CNN tabanlı yaklaşımların başarısını arttırmak için hibrit yöntemler kullanılabilir. Özellikle literatürde başka alanlara uygulanan dikkat (attention) tabanlı derin öğrenme ve derin q ağları yaklaşımlarının bu alana uygulanabilirliği araştırılabilir.

Kaynakça

[1] Fama, E. F. *Random Walks in Stock Market Prices*, Financial Analysts Journal, 1995, 51(1), pp. 75-80.

[2] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. *Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The Sample of The Istanbul Stock Exchange*, Expert Systems with Applications, 2011, 38(5), pp. 5311-5319.

[3] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. *Stock Price Prediction Using LSTM, RNN and CNN-Sliding Window Model*, In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1643-1647. IEEE.

[4] Tekin, S., & Çanakoğlu, E. *Analysis of Price Models in Istanbul Stock Exchange*, In 2019 27th Signal Processing and Communications Applications Conference (SIU), 2019, pp. 1-4. IEEE.

[5] Unal, B., & Aladag, C. H. *Stock Exchange Prediction via Long Short-Term Memory Networks*, Proceedings Book, 2019, 246.

[6] Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. *Stock Market Prediction Using Machine Learning*, In 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), 2018, pp. 574-576. IEEE.

[7] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. *Stock Closing Price Prediction Using Machine Learning Techniques*, Procedia Computer Science, 2020, 167, pp. 599-606.

[8] Du, J., Liu, Q., Chen, K., & Wang, J. *Forecasting Stock Prices in Two Ways Based on LSTM Neural Network*, In 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2019, pp. 1083-1086. IEEE.

[9] Alhazbi, S., Said, A. B., & Al-Maadid, A. *Using Deep Learning to Predict Stock Movements Direction in Emerging Markets: The Case of Qatar Stock Exchange*, In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 440-444. IEEE.

[10] Sezer, O. B., & Ozbayoglu, A. M. *Algorithmic Financial Trading with Deep Convolutional Neural Networks: Time Series to Image Conversion Approach*, Applied Soft Computing, 2018, 70, pp. 525-538.

[11] Sim, H. S., Kim, H. I., & Ahn, J. J. *Is Deep Learning for Image Recognition Applicable to Stock Market Prediction?*, Complexity, 2019.

[12] Thakkar, A., & Chaudhari, K. *A Comprehensive Survey on Deep Neural Networks for Stock Market: The Need, Challenges, and Future Directions*, Expert Systems with Applications, 2021, 177, 114800.

[13] Chen, Y. C., & Huang, W. C. *Constructing A Stock-Price Forecast CNN Model with Gold and Crude Oil Indicators*, Applied Soft Computing, 2021, 112, 107760.

[14] Chandar, S. K. *Convolutional Neural Network for Stock Trading Using Technical Indicators*, Automated Software Engineering, 2022, 29(1), pp. 1-14.

[15] Jiang, W. *Applications of Deep Learning in Stock Market Prediction: Recent Progress*, Expert Systems with Applications, 2021, 184, 115537.

[16] Hu, Z., Zhao, Y., & Khushi, M. *A Survey of Forex and Stock Price Prediction Using Deep Learning*, Applied System Innovation, 2021, 4(1), 9.

[17] Ji, Y., Liew, A. W. C., & Yang, L. *A Novel Improved Particle Swarm Optimization with Long-Short Term Memory Hybrid Model for Stock Indices Forecast*, IEEE Access, 2021, 9, pp. 23660-23671.