

KADIR HAS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
MÜHENDİSLİK VE DOĞA BİLİMLERİ ANABİLİM DALI

**MAKİNE ÖĞRENMESİ İLE ÇOK TERİMLİ HİSSE
SENEDİ YÖNLÜ TAHMİNİ; BİST100 ÖRNEĞİ**

YASEMİN KOÇ

YÜKSEK LİSANS TEZİ

İSTANBUL, MAYIS, 2021

Yasemin Koç

Yüksek Lisans Tezi

2021



MAKİNE ÖĞRENMESİ İLE ÇOK TERİMLİ HİSSE SENEDİ YÖNLÜ TAHMİNİ; BİST100 ÖRNEĞİ

YASEMİN KOÇ

YÜKSEK LİSANS TEZİ

Mühendislik ve Doğa Bilimleri Anabilim Dalı Finans Mühendisliği Programı'nda Yüksek Lisans derecesi için gerekli kısmi şartların yerine getirilmesi amacıyla Kadir Has Üniversitesi Lisansüstü Eğitim Enstitüsü'ne teslim edilmiştir.

ARAŐTIRMA ETİĐİ
VE
YAYIN YÖNTEMLERİ BİLDİRİMİ

Ben, YASEMİN KOÇ;

- Hazırladığım bu Yüksek Lisans Tezini tamamen kendi çalışmam olduğunu ve başka çalışmalardan yaptığım alıntıların kaynaklarını kurallara uygun biçimde belirttiğimi;
- Yüksek Lisans Tezinin başka bir eğitim kurumunda bir derece veya diplomaya sunulan veya kabul edilen herhangi bir materyal içermediğini;
- "Yükseköğretim Kurulu Etik Davranış İlkeleri" uyarınca hazırlanan "Kadir Has Üniversitesi Akademik Etik İlkelerini takip ettiğimi onaylıyorum.

Buna ek olarak, bu çalışma ile ilgili ortaya çıkabilecek herhangi bir usulsüzlük iddiasının, üniversite mevzuatına uygun olarak disiplin işlemi ile sonuçlanacağını kabul ediyorum.

Ayrıca, çalışmalarımın hem basılı hem de elektronik kopyaları, aşağıda belirtilen şartlar çerçevesinde Kadir Has Bilgi Merkezi'nde saklanacaktır:

Tezimin tamamı her türlü erişime açılabilir.

ÖĐRENCİNİN ADI SOYADI

Yasemin Koç

TARİH VE İMZA

24/05/2021

KADIR HAS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

KABUL VE ONAY

YASEMİN KOÇ tarafından hazırlanan **MAKİNE ÖĞRENMESİ İLE ÇOK TERİMLİ HİSSE SENEDİ YÖNLÜ TAHMİNİ; BİST100 ÖRNEĞİ** başlıklı bu çalışma **24.05.2021** tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafında **YÜKSEK LİSANS** olarak kabul edilmiştir.

ONAYLAYANLAR:

Prof. Dr.Ayşe Hümevra Bilge (Danışman)(Kadir Has Üniv.) _____

Doç.Dr. Tamer Dağ (Kadir Has Üniv.) _____

Dr. Öğretim Üyesi Figen Özen (Haliç Üniv.) _____

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

(Unvanı, Adı ve Soyadı)

Lisansüstü Eğitim Enstitüsü Müdürü

ONAY

TARİHİ:

(Gün/Ay/Yıl)

İÇİNDEKİLER

ABSTRACT	i
ÖZET	ii
TEŞEKKÜRLER	iii
İTHAF	iv
TABLolar LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
1. GİRİŞ	1
2.1 Makine Öğrenmesi Yöntemler	8
2.1.1 Denetimli Öğrenme	9
2.1.2 Denetimsiz Öğrenme.....	9
2.1.3 Takviyeli Öğrenme.....	9
2.2 Metodoloji	10
2.2.1 Yapay Sinir Ağları (YSA).....	10
2.2.2 Uzun Kısa Süreli Bellek (LSTM).....	16
2.3. Literatür İncelemesi	19
2.4. Veri ve Yöntem	24
2.4.1. Veri Seti	24
2.4.2. Makroekonomik Değişkenlerin Hisselere Etkisi.....	24
2.4.3. Kullanılan Yöntemler ve Uygulamaları	26
2.5. Analiz ve Bulgular	28
3. SONUÇ	49
KAYNAKÇA	52
ÖZGEÇMİŞ	55

MULTINOMIAL DIRECTION FORECAST WITH MACHINE LEARNING
ALGORITHMS; BIST100 EXAMPLE

ABSTRACT

In this study, a study was conducted to predict the direction of medium-term stocks. In the course of this study, Pegasus shares from the transportation sector and İşbank stock data from the banking sector, which are included in the BIST100 stock exchange, were used. Using two separate data sets for these two companies in the BIST100 index, the results were compared with each other. These data were compared with two different machine learning algorithms. The models used in the study are artificial neural network and long short-term memory (LSTM) model. The data were obtained from the data base of the JUL Program. For the data covering the last five years, a decomposition was performed in order to make the results more meaningful. The analysis of the data sets was carried out primarily on the Excel platform, and then using the Matlab platform, using the formulas written, by training the data. In the study, machine learning algorithms were used to make directional estimation of stocks based on the closing prices of the data. For Pegasus and İşbank stocks, data were trained using the LSTM model, which is an artificial neural network (ANN) and recurrent neural network (RNN) architecture. The parameters used in the models were changed and the data were continuously trained. The data were continuously trained until a significant result was obtained in the parameters used in the models. At the end of the trained data, long-term short-term memory (LSTM) models and Pegasus and İşbank stocks were compared with each other using the artificial neural network model.

Keywords: BIST100, Machine Learning Algorithms, Artificial Neural Networks, Recurrent Neural Network, LSTM, Matlab, Closing Price, Directional Prediction.

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ÇOK TERİMLİ HİSSE SENEDİ YÖNLÜ TAHMİNİ; BİST100 ÖRNEĞİ

ÖZET

Bu çalışmada orta vadeli hisse senetlerinin yönlü tahmini için çalışma yapılmıştır. Bu çalışmayı yaparken, BİST100 borsası içinde yer alan, ulaşım sektöründen Pegasus hisse senedi ile bankacılık sektöründen İş Bankası hisse senedi verileri kullanılmıştır. BİST100 endeksindeki bu iki şirket için iki ayrı veri seti kullanılarak, çıkan sonuçlar birbiri ile karşılaştırılmıştır. Bu veriler iki farklı makine öğrenmesi algoritması ile karşılaştırılmıştır. Çalışmada kullanılan modeller, yapay sinir ağı ve uzun kısa süreli bellek (LSTM) modelidir. Veriler İdeal Programının veri tabanından temin edilmiştir. Son beş yılı kapsayan veriler için, sonuçların daha anlamlı çıkması adına ayrıştırma yapılmıştır. Veri setlerinin analizi, öncelikli olarak Excel platformunda, sonrasında ise Matlab platformu kullanılarak, yazılan formüller üzerinden, veriler eğitilerek yapılmıştır. Çalışma, verilerin kapanış fiyatları üzerinden hisse senetlerinin yönlü tahminini yapabilmek için, makine öğrenmesi algoritmaları kullanılmıştır. Pegasus ve İş Bankası hisse senedi için, yapay sinir ağı (YSA) ve tekrarlayan sinir ağı (RNN) mimarisi olan LSTM modeli kullanılarak veriler eğitilmiştir. Modellerde kullanılan parametreler değiştirilerek, veriler sürekli olarak eğitilmiştir. Modellerde kullanılan parametrelerde anlamlı sonuç çıktısı alınana dek, veriler sürekli olarak eğitilmiştir. Eğitilen veriler sonucunda, yapay sinir ağı modeli ile uzun kısa süreli bellek (LSTM) modeli ve hisse senetleri çıktıları birbiri ile kıyaslanmıştır.

Anahtar Sözcükler: BİST100, Makine Öğrenmesi Algoritmaları, Yapay Sinir Ağları, Tekrarlayan Sinir Ağı, LSTM, Matlab, Kapanış Fiyatı, Yönlü Tahmin.

TEŞEKKÜRLER

Yapmış olduğum tez çalışmamın başlangıcından, bitişine dek her an destekçi olan, her koşulda yardımlarını sunan Dr. Selçuk Öğrenci Hocam'a sonsuz saygı ve teşekkürlerimi sunuyorum. Aynı zamanda tez sürecimde, yardımlarını eksik etmeyen tez danışmanım Prof. Dr. Ayşe Hümeysra Bilge Hocam'a saygı ve teşekkürlerimi sunuyorum. Bu süreçte en büyük desteği gösteren aileme en çok da sevgili Annem'e sonsuz teşekkür ediyorum. Ve son olarak da çalışmamın en önemli kısmı olan verileri, çok rahat ve hızlı bir şekilde temin etmekte desteğini ve yardımlarını her an hissettiğim teknik analist Umut Gümölcüneli'ye sonsuz saygı ve teşekkürlerimi sunuyorum.



Aileme...

TABLULAR LİSTESİ

Tablo 2.5.a. Pegasus Hisselinin Standart Sapma ve Varyansı.....	27
Tablo 2.5.b. İş Bankası Hisselinin Standart Sapma ve Varyansı.....	28
Tablo 2.5.c. Pegasus, YSA Algoritması İçin TP Değerleri.....	31
Tablo 2.5.d. İş Bankası, YSA Algoritması İçin TP Değerleri.....	32
Tablo 2.5.e. İş Bankası, LSTM Değişken Değerleri.....	39
Tablo 2.5.f. Pegasus, LSTM Değişken Değerleri.....	42



ŞEKİLLER LİSTESİ

Şekil 2.2.1.a Yapay Sinir Ağı Modeli.....	11
Şekil 2.2.1.b Aktivasyon Fonksiyon Grafiği I	12
Şekil 2.2.1.c Aktivasyon Fonksiyonu Grafiği II.....	12
Şekil 2.2.1.d Aktivasyon Fonksiyonu.....	13
Şekil 2.2.1.e Basit Yapay Sinir Ağı Modeli.....	14
Şekil 2.2.1.f Tek Katmanlı YSA Modeli.....	15
Şekil 2.2.1.g T1 ve T2 Değer Grafiği.....	15
Şekil 2.2.2.a Tekrarlayan Sinir Ağları (RNN) Algoritması.....	17
Şekil 2.2.2.b Uzun Kısa Süreli Bellek (LSTM) Algoritması.....	18
Şekil 2.5.a Pegasus Hissenin Confusion Matrisi.....	24
Şekil 2.5.b İş Bankası Hissenin Confusion Matrisi.....	26
Şekil 2.5.a Pegasus Günlük Getiri Volatilitesi.....	29
Şekil 2.5.b Pegasus 21 Günlük Getiri Volatilitesi.....	29
Şekil 2.5.c Pegasus 63 Günlük Getiri Volatilitesi.....	29
Şekil 2.5.ç İş Bankası Günlük Getiri Volatilitesi.....	30
Şekil 2.5.d İş Bankası 21 Günlük Getiri Volatilitesi.....	30
Şekil 2.5.e İş Bankası 63 Günlük Getiri Volatilitesi.....	31
Şekil 2.5.f Pegasus %3 Getiri İçin TP Değerleri.....	33
Şekil 2.5.g Pegasus %5 Getiri İçin TP Değerleri.....	34
Şekil 2.5.h İş Bankası %3 Getiri İçin TP Değerleri.....	35
Şekil 2.5.ı İş Bankası %5 Getiri İçin TP Değerleri.....	35
Şekil 2.5.i Pegasus Hissenin Confusion Matrisi.....	36
Şekil 2.5.j Pegasus Eğitim Verisi ROC Eğrisi.....	37
Şekil 2.5.k İş Bankası Hissenin Confusion Matrisi.....	38
Şekil 2.5.l İş Bankası Eğitim Verisi ROC Eğrisi.....	39
Şekil 2.5.m İş Bankası RMSE Değeri.....	41
Şekil 2.5.n İş Bankası %90-%10 RMSE Değeri(Gizli Birim:40).....	42
Şekil 2.5.o İş Bankası %90-%10 RMSE Değeri(Gizli Birim:40).....	43
Şekil 2.5.ö İş Bankası %90-%10 RMSE Değeri(Gizli Birim:100).....	43

Şekil 2.5.p İş Bankası %90-%10 RMSE Değeri(Gizli Birim:100).....	44
Şekil 2.5.r Pegasus RMSE Değeri.....	45
Şekil 2.5.s Pegasus %80-%20 RMSE Değeri(Gizli Birim:100).....	46
Şekil 2.5.t Pegasus %80-%20 RMSE Değeri(Gizli Birim:100).....	46
Şekil 2.5.u Pegasus %80-%20 RMSE Değeri(Gizli Birim:60).....	47
Şekil 2.5.v Pegasus %80-%20 RMSE Değeri(Gizli Birim:60).....	47



1.GİRİŞ

Bilimin ve teknolojinin her geçen gün geliştiđi dünyada, finans piyasası da gelişen teknoloji ile açığa çıkan ihtiyaçlar doğrultusunda uyum sağlamaktadır. Finansal modellerin yapılan çalışmalarda daha doğru sonuç vermesi için teknolojinin sunduđu çalışmalar vazgeçilmez olmuştur. Her gelişen teknoloji ile birlikte finans sektörü de kendi payına düşeni almaktadır. Finans piyasaları arasında riski en fazla olan borsalar da teknoloji ile birlikte her geçen gün gelişmektedir. Davranışsal finans yönüyle de ele alındığında, hisse senedi getirilerinin tahmin edebilirliđi oldukça önemlidir. Yaşanan teknolojik gelişmeler, Borsa İstanbul'u da etkilemiş olup, 2013 yılında tek bir çatı altında toplamış ve işlem hızını arttırmıştır. Borsa piyasası, işlem gören menkul kıymetler arasında en riskli ve en yüksek getiriye sahip olan yatırım aracıdır. Borsada yer alan varlıkların analizlerini yapabilmek, doğru tahminlerde bulunabilmek yatırımcılar için çok önemli bir yere sahiptir.

Yapılan araştırmalarda, hisse senetleri fiyat tahmininin hiçbir zaman öngörülemeyeceđini, yüksek getiri elde etmenin de mümkün olamayacağı kanılarına rastlanmıştır. Bunlardan bir tanesi; ünlü finans yazarı Fama French'in Etkin Piyasalar Hipotezine göre hisse senedi fiyatları bir önceki fiyat hareketlerinden bağımsızdır. Bu hipoteze göre piyasaları öngörmek mümkün değildir. Etkin piyasalar hipotezine göre piyasada yüksek getiri elde etmek de mümkün olamayacaktır (Nevalsami, 2020).

Şu an ise gelişen makine öğrenmesi yöntemleri ile yön tahmini sanıldığıının aksine çok zor değildir. Hisse senedi getirileri tahmin edilirken, sadece makine öğrenmesi de değil, kullanılan yöntemler, kullanılan modeller, kullanılan değişkenler, verilerin kullanılabilirliđi, sektörlerin ayrıştırılması gibi parametreler yapılan çalışmalara dâhil edilmesi öngörülerini mümkün kılmaktadır. Hisse senedi fiyatlarının ve eğiliminin tahmin edildiđini, yaptıkları çalışmalarda gördüklerini öne sürülmüştür. Hisse senedi getirilerinin tahmin edilebilirliđi sorunu, finans sektöründe en çok araştırılıp, çalışma yapılan alandır. (Özçalıcı, 2016)

Küreselleşen Dünya'da ekonomik, sosyolojik, politik etmenlerin neden olduđu piyasalar, yüksek oynaklık gösterebilmektedir. Bu durum, bilhassa Türkiye için

kaçınılmaz bir sorundur. Bu gibi durumlar yatırımcılara karlı fırsatlar sunar. Yatırımcılar volatilitenin, belirsizliğin ve riskin yüksek olduğu koşullar altında, yüksek getirili hisse senedini seçerek varlık değerlerini tahmin edebilmek istemektedir. Hisse senedi getirilerini tahmin etmek, risklerden kazanç elde etmek tüm yatırımcıların ana hedefidir. Hisse senedi tahmin edilebilirliğinin doğru sonuç vermesi de, makine öğrenmesi yöntemlerinin gelişimini önemli kılmaktadır. Hisse senetleri en riskli yatırım araçları arasında yer aldığından ötürü, yön tahmini yapmak, trendi yakalayabilmek zor bir süreçtir. Finansal piyasalardaki tüm yatırımcılar, hisse senedine yatırım yaparken hangi varlığa yatırım yapılmalı hususunda karar vermede ciddi sorun yaşarlar.

Bu çalışmada, Türkiye için literatür taraması yapıldığında, makine öğrenmesi için yapılan çalışmaların kısıtlı sayıda olduğu sonucuna varılmıştır. Yaptığımız çalışma ile literatürdeki bu eksiği kapatmak öncelikli hedefimiz. Yapay zekânın alt bilimi olan makine öğrenmesi algoritmaları ile hisse senedi yön tahmini için çözüm üretmek Türkiye literatürü için önemli olacaktır (Şahin, 2018).

Politik, ekonomik, sosyolojik dalgalanmalar Türkiye Borsasındaki yerli ve yabancı yatırımcılar için yüksek risk taşıyan bir izlenim vermektedir. Riski yüksek olan piyasa pozitif veya negatif etkiyi de yatırımcıya hissettirir. Yüksek risklerin oluşu bazı yatırımcılar için fırsat kapısı açmaktadır. Bu risklerin iyi bir tahmin yöntemi ile analiz edip, kazanca çevirmek literatürde yapılan çalışmalara bakıldığında çok zor olmadığı yönündedir.

Getiriler risk oranı ile doğrusal çalışan bir parametredir. Getirisi yüksek olan bir varlığın, riski de yüksektir. Portföydeki yatırım aracı çeşitliği ve hisse senedi sayısı, portföy riskini azalttığı, portföy riski ile varlık sayısının artışının doğrusal olmadığı ortadadır (Şahin, 2018). Harry Markowitz yaptığı çalışmalarda portföy riskinin, varlıkların risklerinin ağırlıklı toplamı değil de portföy riskinin iyi bir sepet oluşturma ile yani çeşitlilikle azaltılabileceğini öne sürmüştür. Öne sürdüğü bu çalışma için de varlıklar arasındaki Korelasyon katsayısını kullanarak ispat etmiştir (Yıldırım & Eren, 2020).

Varlık getirilerinin yön tahmini için en doğru sonucu yakalayabilmek ve riskleri azaltmak için yıllardan beri çalışmalar yapılmaktadır. Literatürde yönlü tahmin öngörüsü adına olumlu sonuçlar elde edilmiştir. Günlük hisse senedi getirilerinin genellikle tahmin edilebilir olmadığı kabul edilir. Literatürde yapılan çalışmalarda

piyasa hipotezi veya tüketime dayalı varlık fiyatlandırma modelleri ve Ampirik uygulamalarla günlük hisse senedi getirilerinin iyi bir şekilde sonuçlandığını ileri sürmüş olsalar da günlük getirin yönü bir yere kadar tahmin edilebilir olduğunu yaptıkları çalışma ile görmüşlerdir. Bu doğrultuda yapılan araştırmalara göre kısa vadede varlık fiyatlarının tahmin edilmesinin oldukça zor olduğunu söyleyen Eugene Fama için Nobel ödülü ile sonuçlanmıştır (Becker & Leschinski, 2018).

Yapay zekânın ve makine öğrenimin bu denli geliştiği bir dünyada, riski azaltmak ve belli bir model çerçevesinde analiz yapmak artık çok önemli hale gelmiştir.

TCMB'nin göstergeleri Türkiye borsası için çok önemlidir. Hisse senetlerinin yönlü tahmini yapılırken bir diğer önemli hususta makroekonomik göstergelerdir. Bu göstergeler tüm dünyayı etkileyen, politik, ekonomik, sosyolojik durumları kapsayan değişkenlerdir. Bu değişkenler, hisse senetlerini pozitif veya negatif eksene çekmede önemli etkendir. Siyasi ve sosyolojik durumlar, bilhassa siyasi durumlarda önemli bir sapma olduğunda piyasa müdahalesini beklemek doğaldır. Bu süreçte piyasa gözlemlenir ve ne olabileceği geçmişte yaşanan benzer durumlara bakılarak yön tahmini yapılabilir. Yönlü tahmin edilebilirlik, yatırım yaparken tercih edilen önemli bir adımdır. İç faiz oranı yabancından yüksek, yurt içi enflasyon oranı ise yabancından düşükse bu durum yerel paranın değer kazanması anlamına gelir. Bu gibi durumlar piyasaları yönlü tahmin edebilir olduğumuzu gösterir. Bundan dolayı değişkenler para politikalarında önemli yere sahiptir ve doğru analiz edilmelidir (Chung & Hong, 2007).

Türkiye Borsasını etkileyen makroekonomik değişkenler olan, reel GSYH, reel faiz oranı, reel döviz kuru, enflasyon, işsizlik oranı, emtia fiyatları, değişkenlerinin hisse senetleri üzerindeki etkisi oldukça önemlidir. Bu çalışmada ana değişkenimiz BİST100 verileridir. Diğer değişkenler ise bağımsız değişkenler olarak kabul edilir. Hisse senetleri için fiyat tahmini yapılırken çalışmalar iki grupta toplanabilir. İlk grup makroekonomik verilerin girdi olarak çalışmaya dâhil edilmesidir. (Özçalıcı, 2016.) Makine öğrenimi, yapay zekânın alt bilim dalıdır. Makine öğrenimi, büyük veriler üzerinden çıktılar olarak ve matematiksel işlemler yaparak tahminlerde bulunan sistemlerin, bilgisayarlarla modellenmesidir. Makine öğrenimi birden fazla model ve algoritmadan oluşmaktadır. Makine öğrenimi genel olarak üç gruba ayrılır; Denetimli (Supervised), Denetimsiz (Unsupervised), Takviyeli (Reinforcement). Bu çalışmada

yapay sinir ağı (YSA) modeli ve tekrarlayan sinir ağı (RNN) algoritması olan LSTM algoritması ile analiz yapıp, hisse senedi yönlü tahmin edilebilirliği kıyaslanacaktır.

Yapay sinir ağı (YSA) yöntemi, literatürde en az hata ile ne iyi sonucu veren bir modeldir. Hisse senedi özellikleri belirlenip, aktif bir portföy yaratmak için kullanılan veriler, bu özelliklere uygun bir şekilde kümelendir. Oluşturulan kümede, benzer yakınlığa sahip hisse senetleri bir arada bulunabilir. Benzer yakınlıktaki hisse senetlerinin kümelenebilmesi, aktif Portföy oluşturulması için gerekli bir husustur. Kümelemenin en önemli katkıları hisse senedi piyasalarında varlık tahsis ve risk yönetimidir (Özçalıcı, 2016).

Makine öğrenmesi içerisinde yer alan derin öğrenme, doğrusal olmayan klasik yöntemlerden farklıdır. Derin öğrenmede, her katman kendinden önceki katmandaki çıktıyı girdi olarak almaktadır. Derin öğrenme mimarilerinden olan tekrarlayan sinir ağı (RNN), bağlantılar arasındaki döngüyü oluşturan yapay sinir ağının alt kümesinde yer alan bir mimaridir. İleri beslemeli YSA'lara göre RNN'ler, giriş hafızasındaki giriş verilerini rastgele dizilerini işleyebilmektedir. Tekrarlayan sinir ağında, sıralı bilgileri kullanmak öncelikli düşüncedir (Şeker, Diri, Balık, 2017).

Finansal verilerin gelecekteki tarihsel verilerini tahmin edebilmek oldukça önemli ve bir o kadar da zorlu bir süreçtir. Finansal verileri etkileyen faktörler, mevsimsellik, ekonomik göstergeler, sistematik olmayan politik olaylar gelecekteki verileri tahmin etmede pozitif ya da negatif etkindir. Kısa vadede tahmin edilebilirlik daha kolay olsa da uzun vadeli tahminlerde doğrusal olan yöntemlerde doğru sonuç verme olasılığı düşmektedir. Uzun vadeli tahminler için makine öğrenimi mimarileri uygulanmaktadır. Makine öğrenmesi modellerinden derin öğrenme, zaman serilerinde kullanılan doğrusal olmayan yöntemlerdendir. Derin öğrenme mimarisi olan tekrarlayan sinir ağı (RNN), zaman serisi verilerinde ve analizlerinde oldukça başarılı mimari olarak bilinmektedir. RNN, girdi verilerini hatırlatan mimaridir. Derin öğrenme yöntemleri, verileri beslemeye dayalı öğrenme yöntemleridir. RNN'nin bir türü olan LSTM ağları ise uzun girdi ve çıktıları modellemek için geliştirilmiştir. LSTM mimarisi, geçmiş verilerden öğrenip, hatırlatan bir modelledir. Geçmiş verilerden öğrenen model, gelecekteki verileri de inşa etmektedir.

Tekrarlayan sinir ağlarının (RNN), yapay sinir ağından ayırt eden en önemli özelliği, geçici hafızaya sahip olmasıdır. Yapay sinir ağı modeli, bir hafızaya sahip değildir.

Geçici de olsa hafızaya sahip olan RNN, zaman serisi ile birlikte geleceğe yönelik öngörülerde bulunabilmektedir.

Kümeleme analizi yapılırken, birbirine benzeyen hisse senetleri aynı grupta yer alır. Bu da gösteriyor ki hisse senedi seçmekte zaman harcamak yerine kümelerdeki en iyi performansı gösteren hisse senetlerinden aktif bir portföy oluşturulabilir.

Yapılan çalışmalarda S&P 500 borsa endeksinin günlük hisse senedi getirilerini, birbirinden farklı makine öğrenimi modelleri kullanılarak yeni birçok terimli sınıflandırma yaklaşımı sunmaktadır. Bu araştırmada hisse senedi fiyat tahmin etmek değil de, çok terimli sınıflandırma yaklaşımını kullanarak hisse senedi yönlü tahmini yapabilmektir. Yönlü tahmin denildiği zaman, getirilerin önceden belirlenmiş bazı hisse senetlerinden daha yüksek olup olmadığının tahmin edilmesine dayanan bir eşik değeridir. Hisse senetlerini farklı modeller kullanarak varlıkların getirilerini yönlü tahmin etmektedir. Yaptığı çalışmadaki temel amaç farklı makine kullanım modelleri kullanılarak, S&P500 endeksini günlük hisse senedi getirilerini çok terimli sınıflandırma yaklaşımı ile yönlü tahmin edebilmektir (Nevalsami, 2020).

Biz de bu çalışmada literatürde yapılan çalışmalara karşılık farklı makine öğrenim yöntemleri kullanarak ve günlük hisse senedine karşılık olarak da sadece kapanış verileri sınıflandırma yaparak, karşılaştırmamızı yapacağız.

Hisse senedi tahminlerinde kullanılan değişkenlerin boyutu, alt kümelerin de boyutunu artırmaktadır. Belirli bir kümedeki değişken adedini ve seçmek istediğimiz değişken adedi parametrelerini kullanarak, değişken alt küme sayısını basit bir kombinasyon formülü ile aşağıdaki gibi hesaplayabiliriz.

$$(n/r) = n! / (r!(n-r)!)!$$

n bu formülde değişken kümesindeki değişken adedi, r de kaç adet değişken seçimini olduğunu ifade eder (Özçalıcı, 2016).

Çok terimli sınıflandırma yaklaşımında iki durum söz konusudur. Bunlardan biri gürültü, diğer bir durum ise sinyal olarak karşımıza çıkar. Mutlak küçük getiriler gürültü, mutlak büyük getiriler ise sinyal olarak sunulur. Gürültünün bir kısmının izole edilmesi, mutlak büyük getirileri tahmin etmenin önemini artırır.

Çok terimli yaklaşım, yüksek mutlak hisse senedi getirilerini tahmin etmeye daha fazla vurgu yapar. Burada yapılan bir diğer kıyaslama da ikili yanıt değişkeni olan gürültü ve sinyal değişkenleridir. Bu araştırmada çok terimli yanıt değişkenleri, dönüş serisinin alt

ve üst çeyreklerini temel almış. Farklı makine öğrenim yöntemleri arasındaki karşılaştırmayı basitleştirenin de dengeli sınıflar oluşturma olduğu sonucuna varılmış (Nevalsami, 2020).

Makine öğrenmesi yöntemleri hem sınıflandırma doğruluğu açısından hem de üretme yeteneği açısından karşılaştırılması açısından önemlidir. Bu şekilde gerçek hayattaki ticaretin simülasyonunda ekonomik kâr da gözlenebilir (Özçalıcı, 2016).

Bu çalışmada uzun vadede yatırımcı çekmek için ve tabana yayılan bir borsa yaratmak için ilk adım olarak uygulamamızda BİST100 verilerini içerisinde yer alan ulaşım ve bankacılık sektörü bazında, çok terimli sınıflandırma yöntemlerini kullanarak anlamlı artma, anlamlı düşme nedir sorusuna cevap aramak, borsa piyasasının yönlü tahmin edilebilirliğini incelemek olacaktır.

Literatürde, yönlü tahmin edilebilirliğin daha büyük getiriler için daha güçlü olduğu ve yönlü tahminde farklı faiz oranlarının yararlı bir değişken olduğu vurgulanmıştır (Chung & Hong, 2007). Dinamik bir düşüş ve yükseliş tahmini ya da sınıflandırılması yakalamak temel amacımız. Yatırımcıların yükseliş ve düşüş için örneğin; %3 'lük bir düşüş ya da %5'lik bir artış gibi birbiri ile simetrik olmayan, düşüşleri ve yükselişleri yakalayabilmek. Bu çalışmayı yaparken de son beş yıllık BİST100 verilerinin kapanış seviyeleri temel alınarak sınıflandırma yapılacaktır. Kullanılan verilerin %70'ini eğitim verisi için kullanıp, %30'unu da test için kullanacağız. Bu çalışmada yapay sinir ağı modeli ile uzun kısa vadeli bellek (LSTM) algoritması ile analiz yapıp, hisse senedi yönlü tahmin edilebilirliği kıyaslanacaktır. Gereklilik olmayan değişkenlerin veri setinden çıkarılması, modelin daha hızlı işlem yapmasını sağlayacaktır. Değişken seçiminin üç amacı vardır. Değişkenlerin tahmin performansının güçlendirilmesi, daha işlevsel tahmincileri ayırt edebilmek, bir üçüncüsü de veri setlerindeki süreçleri anlayabilmek (Özçalıcı, 2016).

Verilerin temiz olması, yapılan analizlerin daha anlaşılır ve sonuçların daha net olmasına neden olmaktadır. Farklı veri kaynağı kombinasyonlarını ele alarak ve çok kaynaklı pazar hassasiyeti, hisse senedi getirileri üzerinde doğrusal olmayan etki yaratır. Çok kaynaklı kombinasyon karışık frekanslı olduğundan doğrusal olması beklenemez. Literatürdeki çalışmaların tümünde de olduğu gibi karışık frekanslı verilerde doğrusal olmayan modellemeyi düşünmek en önemli noktalardan biridir.

Para politikalarının bir aracı da müdahaledir. Para politikaları, geleceğe sinyal vermek için fiilen müdahale edilebilir (Chung & Hong, 2007).

Döviz kuru da borsa analizi için bir değişken olduğu için tüm bu değişkenler birbirleri ile orantılı çalışmaktadır. Bu yüzden değişkenlerin etkilendiği nicelikler hisse senetlerini de etkilemektedir.

Veri setleri için kümeleme ve sınıflandırma analizleri yapılması, belli bir süreçten geçmelidir. Bu süreç verilerin temizlenmesi, ön işleme alınması, tahminler ve simülasyon sonuçlarının analizi ile sona erer. Verilerdeki mevsimsellik ve dönemselliğin tespiti ve bu şekilde de modellerin tahmin doğruluğunu değiştirebilir.

Bu yüzden verilerin homojen olması önemlidir. Bu çalışmanın üçüncü adımda ise makine öğrenimi yöntemlerini kullanarak ve bu yöntemleri birbiri ile kıyaslayıp, verilerin getirilerindeki artış ya da azalış yüzdesini yönlü tahmin edebilmektir.

Ampirik analizde makine öğrenmesi algoritmaları seçimi ve tahmin dönemi ayrılarak analiz yapılacaktır.

2. ÇOK TERİMLİ MAKİNE ÖĞRENMESİ ALGORİTMALARI, METODOLOJİ ve LİTERATÜR TARAMASI

Öğrenme, kişinin hayatı boyunca davranışlarında meydana gelen değişimlerdir. Makineleri de insan sistemine benzetirsek, her an gelişen, değişen, sisteme ayak uyduran sistemlerdir. Makine öğrenimi, 1959'da Amerikan Bilgisayar Bilimcisi Arthur Samuel tarafından oluşturulan bir terim olup, verilere dayalı algoritmalar geliştiren bir yapay zekâ alt tipidir. Makine öğrenmesi, verilerin yorumlanmasını amaçlar. Sistemleri ve verileri birbirinden ayırabilir ve sınıflandırabilir. Makine öğrenimi, karmaşık olan yapıyı anlama, verilere dayalı karar verme, sınıflandırma ve yorumlama gücüne sahiptir. Makine öğrenmesi, verilerin çıktılarında doğru tahminler yapabilmemizi sağlayan algoritmalarıdır (Ebru Çağlayan, 2018). Makine öğrenmesi metotları, geçmiş verileri analiz edip, çıktı verilerini tahmin edebilmek için uygun olan modeli bulmaya çalışır. Makine öğrenmesi algoritmaları **2.1.** başlığı altında detaylandırılmıştır.

2.1. Makine Öğrenmesi Metotları

Büyük verilerin anlamlı ve işlenebilir duruma gelebilmesi için makine öğrenmesi algoritmalarını kullanırız. Makine öğrenmesi algoritmaları, verilerin analiz edilmesini ve çıktılarından anlamlı tahmin çıkarılmasını sağlarlar. Makine öğrenmesi algoritmaların en önemli noktası, sınıflandırma ve tahmin edilebilirliktir. Algoritmaların her biri, direktif kümesi olarak kabul edilebilir. Bunun sebebi ise, her makinenin belirli bir hedefi gerçekleştirmek için izleyebileceği farklı adımlarının olmasından kaynaklıdır. Veri seti ve belirli bir algoritma ile kurulmuş olan modelin, en iyi sonucu vermesi beklenmektedir. Her algoritma yapısı ile ortaya çıkan modeller birbirinden farklıdır. Bunun sebeplerinden bir tanesi de her algoritmanın çalışma mantığının farklı olmasıdır. Bundan sebep, birbirinden farklı makine öğrenmesi metotları geliştirilmiştir. Bu metotlar; denetimli, denetimsiz ve pekiştirmeli öğrenme(takviyeli) olmak üzere üç gruba ayrılmıştır.

2.1.1. Denetimli Öğrenme

Girdiler ve çıktılar arasındaki eşleşmedir. Bu öğrenme yönteminde girdilerden oluşan veri seti ile veriler bilinen yanıtları alır. Daha sonra yeni veriler için anlamlı tahminler oluşturulması beklenir. Bu yöntemle, makine girdi ve çıktılar arasındaki ilişkiyi öğrenir. Denetimli öğrenme iki başlık altında incelenir. **Regresyon, sınıflandırma, tahmin** algoritması. Regresyon algoritmasında sonuçlar sürekli olan bir çıktı içinde tahmin edilmeye çalışılır. Sınıflandırma ise, sonuçlar ayrı ayrı çıktılar olarak tahmin edilmeye çalışılır. Sınıflandırma modeli girdileri kategorilere ayırır. Tahmin ise, geçmişteki ve şu anki verilere bağlı olarak gelecek ile ilgili öngörüler yapma sürecidir.

2.1.2. Denetimsiz Öğrenme

Modelde gözetilmeye ihtiyaç duyulmaz. Bu makine öğrenmesi tekniğinde sadece girdiler vardır. Veri setindeki veriler yorumlanır, ortak alan bulunur. Yorumlanan veriler ve bulunan ortak nokta ile kümeleme işlemi yapılarak anlamlı bir veri elde edebilmektedir. Çıktı olmadığı için tahmin sonuçlarına dair bir geri bildirim söz konusu değildir. Denetimsiz öğrenme yöntemi olarak **kümeleme ve boyut küçültme** yöntemi ele alınır. Kümeleme yöntemi, benzer verilerin gruplandırılmasıdır. Veriler birkaç gruba ayrılır ve durumları bulmak için de her kümede analiz yapılır. Boyut küçültme ise, değişken sayısını azaltmaktır.

2.1.3. Takviyeli(Pekiştirmeli) Öğrenme

Hedefe yönelik yapılması gerekeni öğrenen bir öğrenme modelidir. Agent adı verilen öğrenen makine, karşılaştığı her duruma bir tepki verir ve karşılığında bir sonuç sinyali alır. Agent, sonuç puanını en yükseğe çıkartmak için çalışır. Pekiştirmeli öğrenmede, Markov karar süreci modeli kullanılmaktadır. Diğer öğrenmelerden farkı, herhangi bir eğitim verisi olmadan, kesin yöntemlerin yetersiz olduğu zamanlarda kullanılır.

2.2. Metodoloji

Bu tez çalışması, BİST100 borsası içerisinde bulunan Pegasus hisse senedi ve İş Bankası hisse senedinin son beş yıllık kapanış fiyatları ile girdi seti oluşturulmuştur. Bu veri seti, uzun kısa süreli bellek (LSTM) ve yapay sinir ağı mimarisi ile eğitilip, çıkan sonuçları birbiri ile karşılaştırılıp, hisse senedi fiyatının yönlü tahminini yapmaktadır. Yapılan çalışmada aşağıdaki soruların cevabı aranmaktadır;

- Makine öğrenmesi metotları arasında en doğru ve anlamlı tahmini yapan öğrenme modeli hangisidir?
- Yön tahminini kapanış fiyatı üzerinden analiz etmek anlamlı bir sonuç çıkarır mı?

Bu çalışmanın temel amacı yukarıdaki sorulara anlamlı ve doğru cevapları vermek için çalışmalar yapmaktır. Bu çalışmada uygulanan makine öğrenmesi algoritmaları, BİST100 içerisinde yer alan Pegasus ve İş Bankası hisse senetlerinin son beş yıllık kapanış verileri ile eğitilecektir. 1 Ocak 2017 ile 31 Aralık 2019 arasındaki kapanış fiyatları temel alınmıştır. 1 Ocak 2020 ile 30 Nisan 2021 arasındaki veriler ise test verisi olarak eğitilmiştir. Bu bölümde makine öğrenmesi metotları detaylı bir şekilde açıklanmaktadır.

2.2.1. Yapay Sinir Ağları

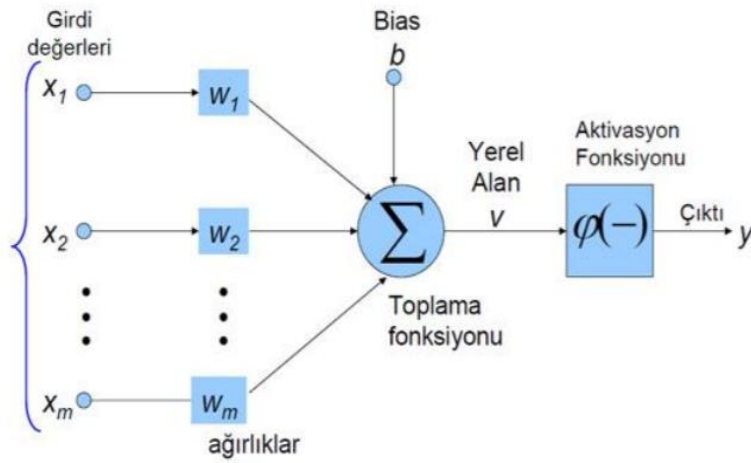
Yapay sinir ağı, insan beyninin bilgi işleme sisteminden yola çıkılarak geliştirilmiş yazılımlardır. Biyolojik sinir ağı, insan ya da hayvan sistemini oluşturan sistemdir. Sinir hücreleri, aksiyon almamızı sağlayan hücrelerdir. Sinir sisteminin çalışma akışı matematiksel bir modele dönüştürülüp, insan gibi düşünen bir model yaratarak makine öğrenmesi algoritması geliştirilmiştir. Sinir hücrelerinin görevi, dışarıdan gelen uyarılarla bilgileri toplamak ve bilgilerin iletilmesi sağlamaktır. Bilgilerin iletilmesi sonucunda refleks dediğimiz tepkiler oluşmaktadır. Refleksin, sinir hücresinin ana fonksiyonu olan nöronlar, karar vermemizi sağlayan birimlerdir. Nöronlar, gelen sinyalleri alır ve bu sinyallere göre eyleme geçer. Burada önemli olan

nokta, sinir hücrelerine birden fazla alandan bilgiler gelir. Bilgiler toplandıktan sonra, karar aşamasına geçilir. Karar sonucuna göre sinyal iletilir ya da iletilmez.

Yapay sinir ağları; girdi katmanı, gizli katman, çıktı katmanı, aktivasyon fonksiyonu, toplam fonksiyonu ve ağırlık adı verilen sayısal değerlerden meydana gelmektedir.

Yapay sinir ağları, girdi verileri toplayıp işleyen ve sonucunda yeni veriler üreten matematiksel modelleme ile çalışan yapay zekâ işleme modelidir. Yapay sinir ağları, girdi ve çıktılardan oluşur. Çıktı sayısı kadar nöron vardır ve bu akışta modellenir.

Makine öğrenmesindeki sinir ağına da girdi değişkenler gelir. Bu değişkenlere göre nöron dizilimi yapılır. Bu nöronlar çıkış seviyesini belirlerler.



Şekil 2.2.1.a:Yapay Sinir Ağı Modeli

Şekil.2.2.1.a’da olduğu gibi giriş değerimiz ile çıkış değerimiz arasında bir iletim yolu vardır. Bu iletim yolu sinyali taşıyıp, çıkış seviyesini belirler. İletim yolundaki sinyal bir direnç değerine sahiptir.

Çıkış seviyesini nöron olarak düşünürsek;

Şekil.2.2.1.a’da bulunan A, B, C, D nöronlarında toplanan giriş değerleri bulunur. Buradaki iletim yolu dediğimiz sinyal direnç değerleri yani ağırlık 12 adettir. Giriş değerleri, ağırlıklarla çarpılıp, toplamı tek bir nörona yani çıktı değerine eşit olur.

Örneğin; Şekil 2.2.1.a’daki olduğu gibi;

$X*WX1 + Y*WY1 + Z*WZ1 = A$ nöronunun toplam değeridir. Toplanan değerler bir değer üstünde mi altında mı diye bakılır. Adım fonksiyonu dediğimiz değer belirli bir değer üstündeyse A, belirli bir değer altında ise 0 değerini alır. Verilen karar neticesinde sinir hücresi üzerinden sinyal iletilir ya da iletilmez.

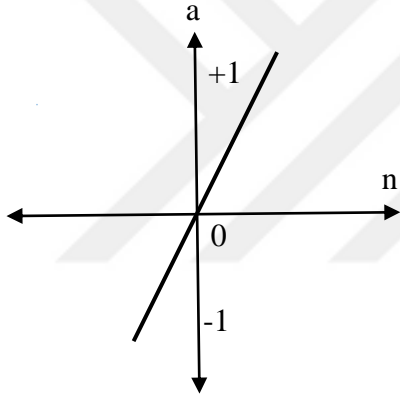
Ağırlıklar **Şekil 2.2.1.a**'da olduğu gibi W olarak tanımlanırsa, X ' den çıkan 4 farklı iletim yolu vardır. Her iletim yolundan çıkan ağırlık farklı olabilir. Her X, Y, Z değeri ağırlıkla çarpılır ve çıktı içerisinde toplanır.

$$X*WX1 + Y*WY1 + Z*WZ1 = A$$

A kümesinin içindeki toplam veri **Şekil 2.2.1.a**'deki formülde olduğu gibidir.

Sistem (Nöron) üzerine gelen girdi değerlerini toplayıp, çıktı değerini oluşturur. Sisteme giren veriler iletim yolu dediğimiz sinyal direnç değerleri yani ağırlıklar ile çarpılıp toplanır. Sonuçta da bir çıktı oluşur. Bu giriş fonksiyonu kısmının izlediği yoldur.

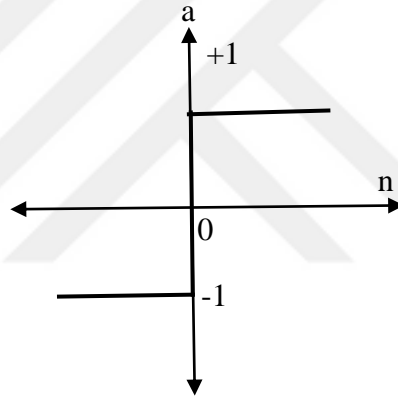
Aktivasyon fonksiyonunun da ise, belli bir değer üstünde ise sinyal iletilir ya da altında ise sinyal iletilmez sürecinin gerçekleştiği kısımdır. Burada temel olan nöronların giriş değerlerine bakılıp sinyal iletimini gerçekleştirilir.



Şekil 2.2.1.b: Aktivasyon Fonksiyon Grafiği I

$$y = k * x \quad a = m * n \quad y = a * x + b$$

Lineer fonksiyondur. Alınan değere göre sonuç döndürebilir. n , aldığı değere göre aktivasyon değeri döndürülür.



Şekil 2.2.1.c: Aktivasyon Fonksiyon Grafiği II

$$y = f(x) \quad a = f(n)$$

0 ve 1 a'nın elemanlarıdır.

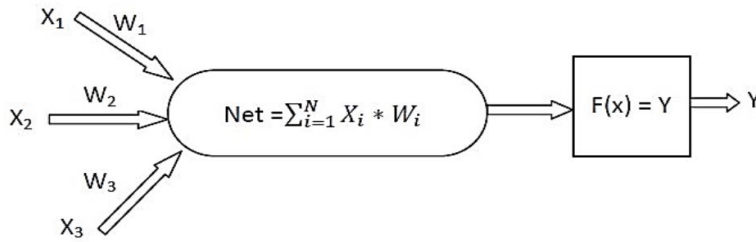
Aktivasyon fonksiyon grafiklerine bakıldığında;

0 değerinin altındaki değerler için -1, 0 değerinin üstündeki değerler için +1 değeri alınır. Fonksiyonda a'nın değerleri n' e bağlıdır. Girdi olarak gelen ve toplanan değerler, negatif değere sahipse aktivasyon fonksiyonu -1, pozitif bir değere sahipse aktivasyon fonksiyonu +1 değerine tekabül eder.

Şekil 2.2.1.b'de görüldüğü üzere her X giriş değerleri ağırlıklarla çarpılır ve çarpılan tüm değerler toplanır. Bias adında b değeri vardır. Ağırlıklara pozitif değer olarak eklenmektedir. Bias dediğimiz b değeri, iki boyutlu düzlemde kaydırma işlemini gerçekleştirmektedir.

Nöronlar üzerindeki bağlı olan sinyalleri, ağırlıklarla çarpıp toplar ve sonrasında çıkan aktivasyon fonksiyonuna göre harekete geçer, sonucunda ise diğer nöronlara bilgi vermektedir.

Aktivasyon Fonksiyonu:



Şekil 2.2.1.d: Aktivasyon Fonksiyonu

Buraya kadar nöronların yaptığı işlemlerden bahsedildi. Nöronların ardından, ağ mimarileri gelmektedir.

Üç farklı ağ mimarisi vardır;

- i. Tek katmanlı ileri besleme**
- ii. Çok katmanlı ileri besleme**
- iii. Geri Besleme**

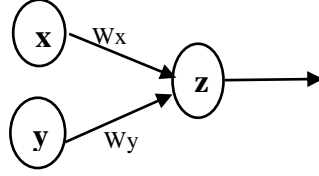
Tek katmanlı YSA, girişlerin nöronlara bağlandığı, nöronlar ile girişler arasında bir ağırlığın olduğu, bu ağırlıkların eğitildiği modeldir. Giriş katmanında giriş verileri sisteme girer. Çıkış katmanında veriler okunur. Burada önemli olan yapay sinir ağı modeli ile istenen çıktıların eğitilmesidir. Aşağıdaki formülle bunu hesaplayabiliriz.

$$\Delta w_{i,j} = a * [\text{öğrenme}_i - \text{çıkış}_i] * \text{giriş}_j$$

$$W_{i,j} = W_{i,j} + \Delta w_{i,j}$$

Yapay sinir ağı modeli eğitilirken, iletim yolundaki direnç değerleri olan ağırlıkların değerleri değiştirilir.

Aşağıdaki şekilde olduğu gibi iki girdi değişkenimiz olsun. X ve Y olarak.



Şekil 2.2.1.e Basit Yapay Sinir Ağı Modeli

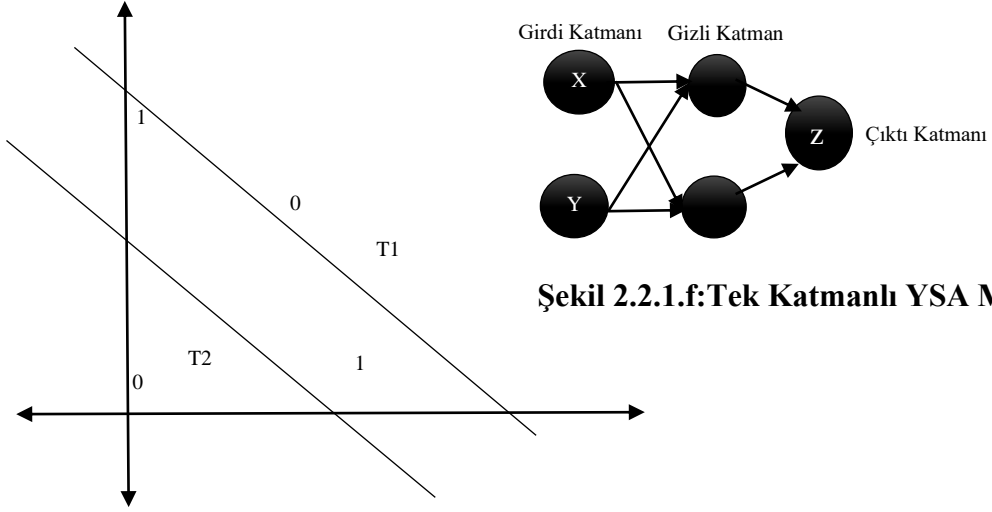
$X*W_x+Y*W_y>Z$ 'den büyükse 1,

$X*W_x+Y*W_y<Z$ 'den küçükse 0 çıkmaktadır.

Yapay sinir ağı modeli eğitilirken, hata değeri meydana çıkmaktadır. Hata değerine göre ağırlık değerleri değiştirilir. Bu hata değeri delta formülü ile bulunur. Çıkan sonuç ile çıkması gereken sonuç arasındaki fark delta değerini yani hata değerini vermektedir. Delta değeri kullanılarak sistem eğitime devam edilmektedir. Ağırlık değerlerini de değiştirmek için alfa değeri kullanılır. Alfa değeri öğrenme oranı olarak tanımlanır.

Alfa değeri ile delta değeri toplanıp yeni ağırlık değerleri belirlenir ve sistem tekrardan eğitilir. Çıkan sonuca göre de yapay sinir ağı modeli doğru ya da hatalı çalışabilir. Bu modelde mühim olan hatayı ortadan kaldırıncaya kadar, ağırlıkların değişmesidir.

Çok katmanlı YSA, gizli katmanlara sahip bir modeldir. Giriş katmanları, gizli katmanlar ve çıktı katmanlarından oluşmaktadır.



Şekil 2.2.1.f: Tek Katmanlı YSA Modeli

Şekil 2.2.1.g: T1 ve T2 Değer Grafiği

Giriş değerlerinden sonra bir katman bulunmaktadır. Bu katman, gizli katmandır. Gizli katman birden fazla olabilir. Sistemin girdi değişkenlerine göre değişir. Burada T1 ve T2 değerlerini belirlemek önemlidir. T1 ve T2 değerleri, ağırlık çarpımından meydana gelen toplam değerlerdir.

Şekil 2.2.1.g'de görüldüğü üzere belli bir değerın altındakiler ve belli bir değerin üstündekilere 0 değeri atanır. T1 ve T2 arasındaki değer de 1 çıkmalıdır. Gizli katman, problemin çözümü için önemlidir. Bir çalışmada kaç adet gizli katman kullanılabileceği konusunda ise literatürde herhangi bir bilgi bulunmamaktadır.

İleri beslemeli yapay sinir ağlarında, giriş katmanlarındaki değerler bir sonraki katman olan gizli katmana verilirken ağırlık değerinin çarpımı ile giriş verisi olarak simüle edilmektedir. Girdi katmanlardan alınan bilgiler, hiçbir şekilde değiştirilmeden gizli katmana verilir. Girdi katmanı ile ağırlık değerinin çarpımı, gizli katmandaki nöron içindeki değerden büyükse 1, değilse 0 değeri atanmaktadır.

İleri beslemeli yapay sinir ağları doğrusal olmayan statik işlemi gerçekleştirir.

Geri beslemeli yapay sinir ağlarında ise geciktirme değişkeni mevcuttur. Bir çıktı değişkeni kendinden sonraki katmana girdi olarak verilmez. Çıktı, kendi katmanında ya da kendisinden önce gelen katmana girdi olarak bağlanır. Geri beslemeli yapay sinir ağları, ileri beslemeli yapay sinir ağları gibi statik bir yapıda değil, dinamik bir yapıdadır.

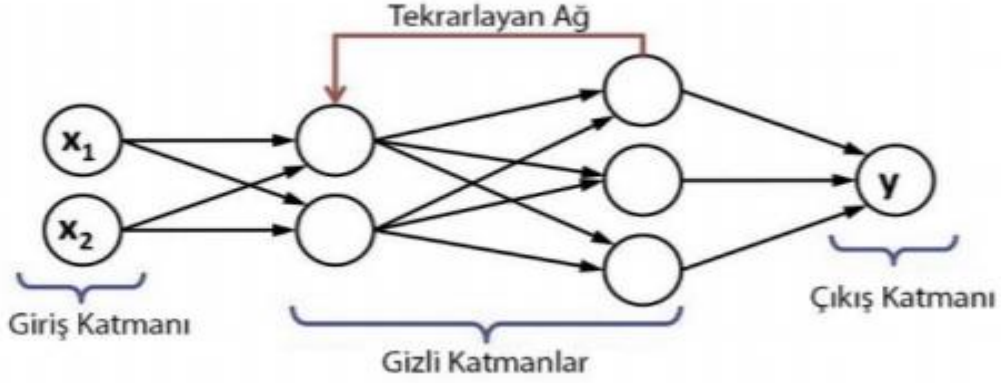
2.2.2. Uzun Kısa Süreli Bellek (LSTM) Mimarisi

Uzun kısa süreli bellek (LSTM) olan derin öğrenme mimarisi, tekrarlayan sinir ağından (RNN) esinlenerek geliştirilmiştir. RNN, ard arda gelen bilgileri kullanan, gizli katmandaki çıktı veriyi, yine gizli katmandaki giriş veri olarak gönderen ve bu şekilde kendi kendini çağıran bir derin öğrenme mimarisidir (Süzen, Kayaalp, 2018).

Tekrarlayan sinir ağı mimarisinde kullanılan, önceki bilgiler mantığından esinlenerek geliştirilmiş bir mimaridir. Tekrarlayan sinir ağı (RNN), yapay sinir ağı modelinde artırılan katman sayısı ile ortaya çıkan mimaridir. Derin öğrenme mimarileri, sınıflandırma ve öngörü işlemleri yapabilmektedir. Finansal verilerin gelecekteki tarihsel verilerini tahmin edebilmek oldukça önemli ve bir o kadar da zorlu bir süreçtir.

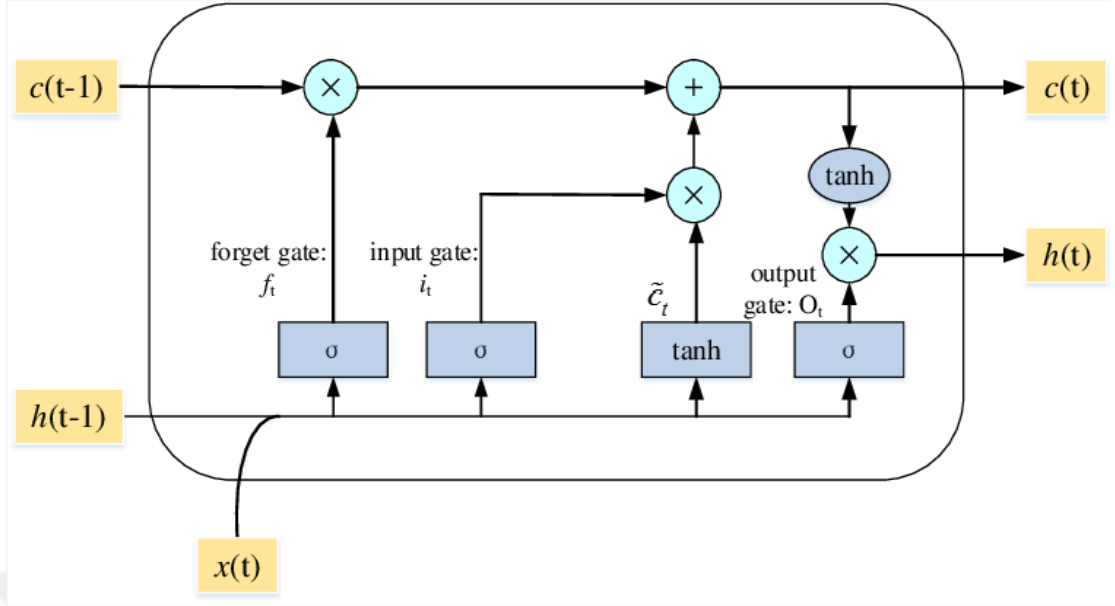
Finansal verileri etkileyen faktörler, mevsimsellik, ekonomik göstergeler, sistematik olmayan politik olaylar gelecekteki verileri tahmin etmede pozitif ya da negatif etkendir. Kısa vadede tahmin edilebilirlik daha kolay olsa da uzun vadeli tahminlerde doğrusal olan yöntemlerde doğru sonuç verme olasılığı düşmektedir. Uzun vadeli tahminler için makine öğrenimi mimarileri uygulanmaktadır. Makine öğrenmesi modellerinden derin öğrenme, zaman serilerinde kullanılan doğrusal olmayan yöntemlerdendir. Derin öğrenme mimarisi olan tekrarlayan sinir ağı (RNN), zaman serisi verilerinde ve analizlerinde oldukça başarılı mimari olarak bilinmektedir. RNN, girdi verilerini hatırlatan mimaridir. Derin öğrenme yöntemleri, verileri beslemeye dayalı öğrenme yöntemleridir. RNN'nin bir türü olan LSTM ağları ise uzun girdi ve çıktıları modellemek için geliştirilmiştir. LSTM mimarisi, geçmiş verilerden öğrenip, hatırlatan bir modelledir. Geçmiş verilerden öğrenen model, gelecekteki verileri de inşa etmektedir (Namini, Tavakoli, Namin, 2019). **Şekil 2.2.2.a'da** tekrarlayan sinir ağı (RNN) modeli yer almaktadır. Bu modelde, x gizli katmanı, h girdi katmanını, y ise çıktı katmanını ifade etmektedir. X gizli katmanından çıkan değer, yine x gizli katmanına girildiği görülmektedir. Tekrarlayan sinir ağı modeline göre, h 'dan giren y 'den giren değerdir. Aradan geçen zaman, zamanla durumların değişmesi, tekrarlı sinir ağının tahmin edilebilir özelliğini zorlaştırmaktadır. Bu tarz sorunlardan dolayı tekrarlı sinir ağı mimarisinin geliştirilmiş mimarisi olan uzun kısa süreli bellek (LSTM)

mimarisi, tahmin edilebilirlik açısından daha doğru ve hızlı sonuçlar elde edildiği görülmektedir.



Şekil 2.2.2.a:Tekrarlayan Sinir Ağları (RNN) Algoritması.

Uzun süreli bağımlılıkları öğrenen ve tekrarlayan sinir ağı (RNN) türüdür. LSTM, bellekten oluşmuş bellek blokları olarak da bilinmektedir. Tekrarlayan sinir ağı (RNN) ile uzun kısa süreli bellek (LSTM) mimarisin farkı, tekrarlayan sinir ağı (RNN) geri bildirim döngüsüne sahip bir mimari iken, uzun kısa süreli bellek (LSTM) ise bellek bloğundan sinir ağı katmanını hücreleyen bir mimaridir. LSTM'de her hücrenin üç farklı kapısı vardır. Bu kapılar, giriş, çıkış, unut kapısı olarak adlandırılır. Blok çıktısı, üç kapıya ve blok girişine bağlanmaktadır. LSTM mimarisinde, giriş, çıkış, unutma/hatırlatma olarak üç farklı noktada aktivasyon fonksiyonu oluşmaktadır. Giriş ve çıkış kapısında tanjant fonksiyonu kullanılırken, unut/hatırlat kapısında sigmoid fonksiyonu kullanılmaktadır. Unut kapısı, geçmişteki gerekli olmayan bilgilerin ağırlıkları azaltmaktadır (Kızırak, Bolat, 2019). **Şekil 2.2.2.b.**'de uzun kısa süreli bellek (LSTM) algoritmasında bloklar gösterilmiştir.



Şekil 2.2.2.b: Uzun Kısa Süreli Bellek (LSTM) Algoritması.

Tekrarlayan sinir ağında, sıralı bilgileri kullanmak öncelikli düşüncedir. Tekrarlayan sinir ağı mimarisi, önceki bilgiye bağlı bir yaklaşımdır. Örneğin; “erik ağaçta yetişir” cümlesindeki erik kelimesini tahmin etmek zor değildir. Bağlantılar arasındaki mesafe arttığında, araya zaman kavramı girdiğinde, RNN mimarisinin geçmişteki bilgiyi kullanabilmesi zorlaşmaktadır. Bu problemleri çözebilmek için de LSTM mimarisi geliştirilmiştir. LSTM mimarisinde, giriş, unut ve çıkış olmak üzere üç farklı kapı, blok girişi, sabit hata döngüsü, çıkış aktivasyon fonksiyonu ve gözetme bağlantıları mevcuttur. Bloğun çıktısı, bloğun girişine ve diğer tüm kapılarına bağlanmaktadır. Unutma kapısı, LSTM'nin kendi halini sıfırlayabilmesi, gözetleme bağlantıları ise zamanlamaları çözümleyebilmek için eklenmiştir (Şeker, Diri, Balık, 2017).

2.3.Literatür İncelemesi

Beta sayıları ile oluşturulan portföyde, kısa ve uzun vadeli performanslar incelenmiştir. Oluşturulan portföyler, pazardan daha yüksek ya da daha düşük getiri elde edip etmedikleri araştırılmıştır. Yapılan çalışmada Borsa İstanbul Sınai endeksindeki firmaların beta katsayıları hesaplanıp, düşük, orta ve yüksek beta katsayısına sahip portföyler kümeleme analizi yapıp, kısa ve uzun vadeli trendler incelenmiştir. Performanslar aylık getiriler üzerinden analiz edilmiştir. Günlük ve haftalık getirilerle, beta katsayılarının analiz edilmesinin daha farklı sonuçlar çıkacağını ve tüm pazar değil de sadece sınai şirketleri dâhil edildiğinden, diğer tüm sektörlerin çalışmaya dâhil edilmesinin de daha anlamlı sonuç çıkaracağı düşünülmektedir. K-means kümeleme algoritmasını kullanarak çalışma yapılmıştır.

Çalışma sonucunda ise orta vadeli performansın daha anlamlı olduğu sonucuna varmışlardır (Yıldırım & Eren, 2020).

BİST100 verilerini kullanarak, makine öğrenmesi metodu olarak da yapay sinir ağları ve bulanık mantık yöntemi ile performans tahmin analizi yapmıştır. Bu çalışma, iki farklı veri setinden oluşturulmuş. İlk set günlük verilerden ve ekonomik parametrelerden, ikinci set ise dış faktörlerden meydana getirilmiştir.

Karşılaştırılan sonuçlarda yapay sinir ağ yönteminin, bulanık mantık yönteminden daha doğru sonuç verdiği görülmüş olup veri seti olarak da oluşturulan ilk set olan günlük verilerden veri üretme yönteminin daha doğru sonuçlar verdiği gözlemlenmiştir. Sonuç itibari ile makroekonomik parametreler ve yapay sinir ağ modelinin, dış faktörler ve bulanık mantık yöntemine göre çok daha iyi getiri elde ettiğini sonucuna varılmıştır (Şahin, 2018).

Yapılan çalışmalarda, 5 farklı makine öğrenmesi algoritması kullanılarak ve S&P500 hisselerinin günlük getirilerini tahmin etmek için yeni bir sınıflandırma yaklaşımı sunulmuştur. Bu yaklaşım da çok terimli (multinomial) 'dır. Multinomial, hisse senedi analiz yapıp, karşılaştırma sonuçları analiz edilmiştir. K-en yakın komşu sınıflandırıcı, gradyan artırma, rastgele orman, nöral ağlar, vektör makineleri kullanılan makine yöntemleridir. Yapılan çalışmada çok terimli yanıt değişkeni, dönüş serisinin alt ve üst çeyrekleri temel alınmış. Beş farklı makine öğrenmesi yöntemleri ve iki farklı yanıt değişkeni kullanılarak yapılan karşılaştırmada büyük hisse senedi getirilerini tahmin

etmek için çalışmalar yapılmıştır. Bu çalışmada kullanılan makine öğrenmesi modellerinden en iyi performansı gradyan artırma modeli ile sağladıklarını göstermişler. Çalışma sonunda ise araştırmanın farklı taraflara genişletilebilir olduğu vurgulanmıştır (Nevalsami, 2020).

BİST50 endeksindeki veriler ve endeksi etkileyen parametreleri ele alarak, yapay sinir ağları ile birlikte makine öğrenmesi yöntemleri kullanılarak sınıflandırmalar yapılmıştır. Çalışmada yapay sinir ağ modelinin diğer yöntemlere göre olan üstün olduğu vurgulanmıştır. Makine öğrenmesi metotları olarak da k-en yakın komşu algoritması (k-NN), Navie Bayes algoritması, C4.5 sınıflandırma algoritması ve yapay sinir ağ modelleri karşılaştırılmış. Burada yapılan ana çalışma, endeksi etkileyen parametrelerin yükseliş ve düşüşleri temel alınarak makine öğrenmesi yöntemleri sınıflandırma yapılarak, karşılaştırılma yapılmıştır. Yapılan çalışma sonucunda ise C4.5 algoritmasının en başarılı sonucu verdiği vurgulanmıştır. Bu çalışma ile yatırımcıların günlük planlama yapması doğru bir yardımcı kaynak olduğu belirtilmiştir (Karaboğa & ark. , 2017).

Ulaşım sektöründen bir şirket temel alınarak yapılan çalışmada en çok tercih edilen makine öğrenimi olan YSA(yapay sinir ağ) modeli kullanılmıştır. Veri olarak da BİST100'de bulunan ulaşım sektöründe yer alan tek bir şirketin 2015-2018 arasındaki verileri kullanılmıştır. Eğitim seti olarak haftanın ilk dört günü, test verisi olarak da Cuma günleri temel alınmıştır. Eğitim seti ile test seti için Korelasyon analizi yapılmıştır. Korelasyon analizini de, eğitim seti ile test setinden oluşan verilerin farklı yapıda olup olmadığını anlayabilmek için yapmışlardır. Modelle oluşturulan tahmini hisse senedi değerleri ile gerçekte olan hisse senedi değerleri karşılaştırılmıştır. En yakın değerli model ele alınıp, bu modelin doğruluğu için başka bir veri seti oluşturulup, çalışmaya eklenmiştir. Ulaşım sektörüne bağlı olarak çalışan bu şirket için yapılan çalışmada girdi değişkenleri olarak, petrol fiyatları, dolar kuru, BİST ulaştırma endeksi ve BİST100 endeksi değerleri olarak ele alınmıştır.

Çıktı değişkeni ise şirketin piyasa değeridir. Çalışma sonunda YSA modelinin anlamlı sonuç verdiği, kullanılan analizler dışında yeni değişkenlerin de belirlenip çalışmaya eklenmesinin daha başarılı sonuçlar çıkaracağı aynı şekilde farklı topolojiler ve farklı makine öğrenmesi modellerinin de kullanılması ile birlikte karşılaştırma yapılması gerektiği düşünülmektedir (Çınaroğlu & Avı, 2020).

Yapılan çalışmalarda, tüm şirketleri kapsayan veri seti kullanılmıştır. Günlük verilerin yönlü tahmini için kullanılan yöntemler ise, lojistik regresyon, sinirsel ağlar, destek vektör makineleri, sınıflandırma ağaçları. Karşılaştırılan modellerde en iyi sonucu, regresyon modelinin en iyi tahminleri ürettiği sonucuna varılmıştır. Günlük hisse senedi yönlü tahmininin bir noktaya kadar tahmin edilebilir olduğunu sonuçlarla sunulmuştur (Becker & Leschinski, 2018).

Hisse senedi getirileri tahmin etmek için piyasa duyarlılığının önemine dikkat çekilmiştir. Davranışsal finansa dikkat çeken çalışmada, U-MIDAS-SVR modeli kullanılmıştır. Veriler iki konu üzerinden incelenmiştir. Karma frekans ve doğrusal olmayan ilişki. Karma frekanslı veri analizini SVR yöntemi ile ele almışlardır. Bununla birlikte (U) MIDAS-SVR modeli ile de çalışmayı detaylandırmışlardır. Burada kuadatik programlamanın, Language Dualite tekniğini ele alarak, tahmin edilebilirliğin olabileceği düşünülmüş. Veri seti olarak, Çin'e ait olan SHSE ve SZSE'nin haftalık getirilerini tahmin edebilmek için (U) MIDAS-SVR modeli uygulanmıştır. Karma frekanslı verilerde hisse senedi getirileri tahmin edilmesi için doğrusal olmayan modellemenin düşünülmesi gerektiği savunulmuştur. Bu modelin de SVR'lerle birlikte (U)MIDAS-SVR modeli. MIDAS (Mixed Data Sampling) yaklaşımı karma frekanslı verilerin bir arada kullanılmasını sağlayan bir modeldir. Bu modelin kullanılmasındaki temel amaç ise, deneysel uygulamalar yapılmasıdır. Öncelikli olarak üç farklı karma frekans kullanılıyor. SF, UMIDAS, MIDAS modelleri. Bu modellerle SHSE ve SZSE'nin haftalık hisse senedi getirileri tahmini karşılaştırıldığında en iyi sonucu (U) MIDAS-SVR modeli olduğu sonucuna varılmıştır. Farklı makine öğrenmesi yöntemleri olarak, MIDAS-SVR, UMIDAS-SVR, SF-SVR, Doğrusal olmayanları anlamak için SF-ANN, SF-RT ve SF-Bagging uygulanmıştır. Yedi farklı senaryo ile hisse senedi tahmin edilebilirliği kıyaslanmıştır. Gelecek araştırmalarda, birkaç olası durumun araştırılması gerektiği, e-SVR ile kayıp fonksiyonunun karşılaştırılmasının anlamlı bir fikir olduğu düşünülmektedir (Xu & ark, 2019).

Yapılan çalışmalarda, bir dakikalık, beş dakikalık ve on dakikalık bir hisse senedinin fiyatını tahmin etmek için uzun ve kısa süreli hafıza teknik göstergeleri üzerinde S&P500 verileri içerisinde Kaggle'den gün içi ticaret verileri alınarak çalışma yapılmıştır. Bu çalışma yaparken, Regresyon uzun-kısa süreli bellek modeli eğitilmiştir. Bu çalışmanın temel amacı, yüksek frekanslı geçmiş veriler kullanılarak, kapanış

fiyatının 10 dakika, 5 dakika önceden tahmin edilebilirliğini LSTM modelini kullanarak ölçebilmektedir.

S&P500 endeksinin SPY günlük hisse senedi verilerinin yön tahmini için bulanık c-ortalama yöntemini, verileri kümelemek için kullanılmıştır. Temizlenen verilerle on iki yeni veri seti oluşturulmuştur. Yapay sinir ağları ve lojistik regresyon modelleri günlük hisse senedi getirisinin yönünü tahmin etmek için PCA modeli ile sınıflandırılma yapılmıştır. Çalışmanın anlamlı sonuç vermesi için daha geniş kapsamlı çalışma yapılması düşünülmektedir (Zhong & Enke, 2017).

Yapılan çalışmalarda, IBM hisse senetlerinin fiyatları veri seti olarak oluşturmuşlardır. Temel amaç ise, farklı öğrenme modelleri ile zaman serisine en uygun modeli ortaya çıkarmaktır. Çalışmada 1968-2018 arasındaki ABD hisse senetlerinin tamamı için günlük fiyat verileri kullanılmıştır. Veri setlerini 5 günlük veri setine bağlı olarak oluşturmuşlar. Derin öğrenme algoritması olarak da LSTM, BLSTM, GRU modelleri kullanılmıştır. Hedef değer olarak kapanış değerleri ele alınmıştır. Hata ölçüm modelleri olarak RMSE, MSE, MAE, DA modellerinden yararlanılmıştır. Yapılan çalışmalarda, al-sat kararı verebilecek şekilde geliştirme yapmışlardır (Şişmanoğlu & ark. , 2020).

Başka bir çalışmada ise hisse senedi fiyat tahmininin öğrenme süreçlerinin yavaş olmasının yatırımcıları olumsuz etkilediği bundan dolayı da bu duruma çözüm olarak da aşırı öğrenme makineleri adı yöntemi geliştirilmiştir. Yapılan çalışmada, Goodyear, Amazon ve WalMart şirketleri ile S&P500 endeksinin geçmiş verileri kullanılarak, 12 farklı teknik gösterge kullanılarak hesaplama yapılmıştır. Bu veriler girdi değişkeni olarak ele alınmış olup, çıktı değişkeni de ertesi güne ait kapanış fiyatıdır. Yapılan çalışmada modeller, yapay sinir ağları modeli ile kıyaslanmıştır. Yapılan çalışma sonucunda, al-tut performansının daha yüksek olduğu ve AÖM'nin hisse senedi fiyat tahmini için önemli ve anlamlı sonuç çıkaran bir model olduğu gözlemlenmiştir. Yön tahmininde YSA'nın endeks temelli daha iyi sonuç verdiği, üç farklı hisse senedi için de en doğru yön tahmini AÖM ile sağlanmıştır (Özçalıcı, 2017).

Literatür incelendiğinde, LSTM mimarisinin sınıflandırma ve tarihsel öngörülerde anlamlı sonuçlar ortaya çıkardığı görülmektedir. Uzun kısa süreli bellek (LSTM) mimarisi, giriş, çıkış, unut kapıları ve hafıza hücrelerinden oluşan bir yapıdır (Süzen, Yıldız, Yılmaz, 2019).

Yapılan başka bir çalışmada ise Covid-19 pandemisinin zaman serisi verileri kullanılarak, kümülatif alanını tahmin etmek için ARIMA yöntemi ile LSTM mimarisi kıyaslanmıştır. ARIMA modelleri ve LSTM tabanlı tekrarlayan nöral ağ, günlük kümülatif tahmini doğruladığı gözlemlenmiştir. LSTM mimarisi, uzun ve kısa vadede zaman serilerinden yararlanılarak zamansal tahmini gerçekleştirmiştir. LSTM mimarisi kullanılarak yapılan tahminlerin, ARIMA modellerinden çok daha doğru sonuçlar ortaya çıkardığı gözlemlenmiştir. Bu sonuçlar da göstermiş oluyor ki, vaka sayılarının tahmin edilmesi, salgın ve politika sürecine zamansal olarak katkıda bulunmuştur (Masum, Shahriar, Haddad, Alam, 2020).

Hindistan borsası (NSE) ve Newyork borsasından (NYSE) hisse senedi fiyat tahmini için dört farklı ağ eğitilmiştir. Kullanılan doğrusal modeller, MLP, RNN, LSTM ve CNN ağlarıdır. Ortaya çıkan sonuçtan modellerin dinamik ağlar olduğu görülmüştür. Doğrusal ağlardan CNN'nin en iyi sonuca sahip olduğu gözlemlenmiştir. Bu dört ağa sahip DL modeli dışında, doğrusal olmayan ve tek değişkenli zaman serisi tahmini için ARIMA modeli de kullanılmıştır. Bu iki model kıyaslanmış ve en iyi sonucun DL modellerine ait olduğu gözlemlenmiştir (Hiransha, Gropalakrishnan, Krishna, Soman, 2018).

Uçak motorlarının kalan süresini tahmini için derin öğrenme mimarisi olan LSTM algoritması kullanılmıştır. Çalışmada kullanılan veriler ise NASA tarafından hazırlanan C-MAPSS veri kümesi içerisinde seçilen yüz adet motor sensör verileri kullanılıp, veriler eğitilmiştir. LSTM mimarisi ve klasik makine öğrenmesi algoritmaları ile kıyaslanmıştır. Sonuç olarak LSTM mimarisinin diğer tüm klasik yöntemlere kıyasla en iyi sonucu verdiği analiz edilmiştir (Kızrak, Bolat, 2019).

Finansal verileri tahmin etmede etkili olan hem sistematik hem de sistematik olmayan birden fazla faktör bulunmaktadır. Kısa vadede tahmin edilebilirlik, uzun vadeli tahmin edilebilirliğe göre daha kolaydır. Literatürde, ARIMA yöntemleri ile birlikte zaman serilerini analiz etmede ve tahmin etmede makine ve derin öğrenmeye bağlı yaklaşımları kıyaslanmıştır. ARIMA, LSTM ve BiLSTM yöntemleri kullanılarak veriler eğitilmiş ve modeller birbirleri ile kıyaslanmıştır. Burada, LSTM'nin ARIMA modelinden, LSTM'nin geliştirilmiş yapısı olan BiLSTM'nin ise LSTM'den daha başarılı sonuçlar verdiği kanısına varılmıştır.

2.4. Veri ve Yöntem

2.4.1. Veri Seti

Bu çalışmada, 1 Ocak 2017 ile 30 Nisan 2021 arasındaki BİST100 borsasında işlem gören, ulaşım sektöründe yer alan Pegasus hisse senedi ile bankacılık sektörü içinde yer alan İş Bankası hisse senedi verileri ele alınmıştır. Veriler İdeal adlı programın veri tabanından temin edilmiştir. Veriler açılış, kapanış, en yüksek, en düşük, işlem hacmi, şirketin borsadaki işlem görmeye başladığı tarih itibari ile text olarak çıktı alınmıştır. İdeal programından alınan veriler öncelikle detaylı bir inceleme ve temizleme sonrası istenen şekle dönüştürülmüştür. Veriler, çalışılacak tarih temeline gelecek şekilde temizlenmiştir. 1 Ocak 2017 ile 31 Aralık 2019 arasındaki veriler training set olarak oluşturulmuştur. 1 Ocak 2020 ile 30 Nisan 2021 arasındaki veriler ise test verisi olarak oluşturulmuştur. Tarih temelli ayrıştırılan veriler için ikinci aşama ise, açılış, en yüksek, en düşük, işlem hacim gibi değerler veri setinden çıkarılmıştır. Veri seti, kapanış fiyatı ve tarih temeline oluşturulmuştur. Temizlenen veri setleri, eğitime hazır hale getirilmiştir. Oluşturulan veri setleri içerisinde sektör ayrımı yapılmıştır.

2.4.2. Makroekonomik Değişkenlerin Hisselerine Etkisi

Ekonomik etkenlerin hisse senedi değerlerinin tahmin edilmesinde büyük bir role sahiptir. Ekonomik değişkenler yalnızca emtialar değerlerini değil, genel itibari ile ülkelerin ekonomisi hakkında da önemli bilgiler sunar. Ekonomik değişkenlerin ileride değerlerindeki değişiminin tahmin edilmesi portföy yönetimi için önemli bir yere sahiptir. Makroekonomik değişkenlerin tüm emtialar üzerinde önemli bir yere sahiptir. Bilhassa hisse senedi piyasası için belirli olan makroekonomik değişkenler mevcuttur.

Hisse senetleri, ekonominin gelişimi ve büyümesi için önemli bir yere sahip piyasadır. Hisse senetlerinin oynaklıkları, sürdürülen ekonominin unsurudur. Hisse senedi piyasası her an değişebilen bir piyasadır. Ülke içindeki beklenmeyen durumlar ve yurt dışındaki beklenmeyen durumlar ekonomik gündemi fazlasıyla etkilemekte ve

piyasalarda ciddi oynaklıklara sebebiyet vermektedir. Bu gündemler içerisinde siyasi olaylar, doğal afetler, dünyayı etkisi altına alan salgın hastalıklar tüm piyasaları etkilemektedir. Bu gibi beklenmeyen ani durumlar, sektörlerin çoğunu etkilemektedir. Bazı sektörleri pozitif yönde, bazı sektörleri ise negatif yönde etkilemektedir. Bu durumlarda hisse senedi piyasasında oynaklıklara sebebiyet vermektedir. Emtiadaki ciddi oynaklıklar, döviz kurunu, enflasyon ve faiz oranında değişimlere neden olmaktadır. Bu gibi değişkenlerin değişimi de direkt olarak hisse senedi piyasalarını etkilemektedir. Bunlar dışında hisse senedi fiyatındaki değişimleri etkileyen bir diğer durum ise, sektördeki artı/eksi değişimler, sektördeki şirketlerin bilançolarını açıklaması, yatırım yapmaları ya da şirketlerin üretim yapıp yapmaması gibi durumlar hisse senedi alımını ve satımını etkilemektedir. Bir hisse senedinin fiyat artışı o hisseye olan talebi artıracaktır. Tam tersi durumda geçerlidir. Fiyat düştükçe ise hisse senedi satışı başlar. Bu durum ise psikoekonomi ile yakından ilgilidir. Bu başka bir araştırma konusudur. Günümüzde hisse senedi piyasaları önem arz etmektedir. Görüldüğü üzere hisse senedi piyasasını etkileyen soyut ve somut birden fazla değişkenler mevcuttur. Yatırımcıların bu durumları yönetebilmesi için portföylerini iyi bir şekilde analiz etmeleri kaçınılmaz olmaktadır. Makroekonomik değişkenlerin hisse senedi piyasaları ile ilişki konusu literatürde hala tartışma konusudur. Piyasalardaki değişimler ve ülke ekonomisindeki durumlar göstermektedir ki, makroekonomik değişkenler her piyasayı ayrı ve farklı bir açıdan etkilemektedir. Makroekonomik değişkenlerin Borsa İstanbul için nasıl bir etki yarattığı ve tahmin yönünü ne tarafa ve nasıl bir oranda değiştirdiği yatırımcılar için çok önemlidir.

Şirketlerin yatırım alması ya da yatırım yapmaları, bilanço tablosu, yatırım yapılan hisse senedi için derin bir bilgiye sahip olmak da yatırımcılar için önemli bir noktadır. Bu gibi durumlardan sonra portföyü yönetebilmek için makroekonomik değerlerin etkilerini ve makroekonomik değişkenlerin ani beklenmedik değişimlerinin nedenlerinin de araştırılması, hisse senedi piyasası için önem arz etmektedir. Bir değişkenin nedeni direkt olarak piyasaları etkilemektedir. Her durum birbirine paralel olduğu için her zaman başlangıç noktasının ne olduğunun bilinmesi gerekmektedir. Makroekonomik değişimler incelendikten sonra, hisse senedi değer ve yön tahmini için analiz yapılmalıdır. Bu analizler, teknik analizler olabileceği gibi, bazı algoritmalar da

olabilir. Yazılımsal olarak düşünürsek, herhangi bir program ya da platform kullanarak, bu platformların toolboxlarında bulunan makine öğrenmesi algoritmalarını da kullanarak, çalışmalar yapılmaktadır. Bu çalışmaların en temel amacı, en iyi performansı ortaya çıkarabilecek, hata oranı düşük algoritmayı bulmaktır. Bunun içinde oluşturulan veri setlerinin, kullanılan platforma bağlı olarak, uygun algoritmalarla, simülasyonlar oluşturup, verileri eğitip, oluşturulan algoritmaları karşılaştırmaktır.

2.4.3. Kullanılan Yöntemler ve Uygulamaları

Bu çalışmada veriler öncelikli olarak hisse senedi kapanış fiyatı bazından ayrıştırılmıştır. Hisse senedi ve tarih temelli ayrıştırılan veriler text formatından Excel formatına dönüştürülmüştür. Burada veri setleri düzenlenirken 1 Ocak 2017 ile 30 Nisan 2021 verileri temel alınmıştır. Analizler, öncelikli olarak Excel platformu kullanılarak günlük, 21 günlük ve 63 günlük kapanış fiyatlarının getirileri hesaplanmıştır. Getiri hesabına bağlı olarak da gerçekleşen getiri değerleri üzerinde yönlü grafikler çizilmiştir. Grafiklerin yönleri ve getiri değerleri birbirleri ile karşılaştırılmıştır. Excel’de yapılan analizi sonrası Matlab platformu ile makine öğrenmesi modelleri kullanılarak veriler eğitilip, analizlerin yapılmasına karar kılınmıştır.

Matlab platformunda öncelikle text formatından Excel formatına dönüştürülen ve ayrıştırılan BİST100 içerisinde yer alan Pegasus ve İş Bankası hisse senedi verilerinin dosyası eklenmiştir. Daha sonra ise Matlab platformunda yer alan ve bizim kullanacağımız algoritmalar için ayrı ayrı komutlar ve fonksiyonlar yazılmıştır. Kullanılan algoritmaların komut kümelerine ait fonksiyonlar yazılarak, analizler yapılmıştır. Matlab’de veri dosyası oluşturulduktan sonra, kullanılacak komutlar için değişken isimleri belirlenmiştir. Komutlar tek tek workspace’e eklendikten sonra fonksiyon içerisine dâhil edilip, algoritmalar çalıştırılmıştır. Hata oranı fazla çıktığında manuel girilen değişkenler sürekli değiştirilip, hata oranı azaltılmaya, eğitilen verilerin doğruluk oranı da yükseltilmeye çalışılmıştır. Veri setine dâhil edilmiş tüm değişkenler üzerinden fonksiyonlar düzenlenerek, algoritmalar sürekli eğitilmiştir. Bu çalışma iki aşamada gerçekleşmiştir. Birinci aşama gerçekleşen değerler üzerinden

eđitilen bir veri setidir, ikinci ařama ise test deęerleri zerinden eđitilen bir veri setidir. Test ve gerekleřen durumlar iin iki ayrı veri seti oluřturulmuřtur. Bu deęiřkenler yazılan fonksiyonlara dâhil edilip, deęiřken deęerleri srekli deęiřtirilip, veri setleri srekli eđitilmiřtir. Eđitilen veri setlerinin sonucundan en az hata oranını ve en yksek doęruluk oranı, model olarak seilmiřtir.

Bu alıřma iki farklı makine ęrenmesi algoritması iin yapılmıřtır. Yapay sinir aęları (YSA) ve uzun kısa sreli bellek (LSTM) algoritması ile. Bu iki algoritma iin ayrı ayrı deęiřkenler tanımlanmıř, fonksiyonlar yazılmıř, veri setleri srekli eđitilmiřtir. İki algoritmada da eđitim sonunda en az hata oranını veren fonksiyon ve deęerler temel alınmıřtır. Bu deęerler iki algoritma iin tek tek karřılařtırılmıřtır. ncelikle, algoritmaların kendi iinde en doęru sonucu veren deęeri bulunup, sonrasında ise iki algoritma iinde en doęru tahmine ulařan oranlar belirlenmiřtir.

2.5. Analiz ve Bulgular

Verilerin analizi, Excel ve Matlab programı kullanılarak yapılmıştır. Öncelikle çalışma, Ocak 2017 ile 31 Aralık 2019 arasındaki BİST100 içerisinde yer alan Pegasus ve İş Bankası hisse senedinin kapanış fiyat verileri alınarak Excel platformu ile analiz edilmiştir. Test verisi ise, 1 Ocak 2020 ile 30 Nisan 2021 arasındaki verilerdir. Hisse senetlerinin günlük getiri hesaplaması yapılmıştır. Günlük getiriler hesaplanırken aşağıdaki formülden yararlanılarak hesaplama yapılmıştır.

$$\frac{[PD(t) - PD(t - 1)]}{PD(t - 1)}$$

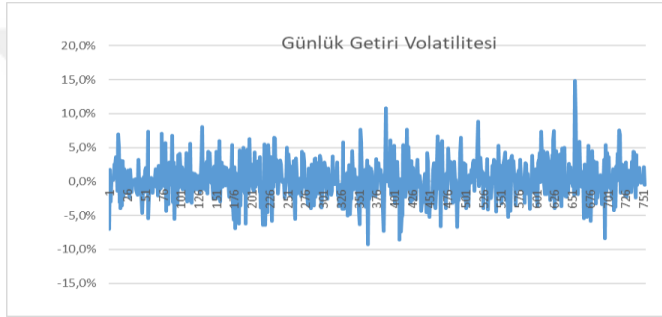
Excel platformu ile Pegasus ve İş Bankası hisse senetlerinin günlük, 21 günlük, 63 günlük getirileri hesaplanmıştır. Bu getirilerin, oynaklığını diğer adıyla volatilitelerini ölçmek için, standart sapma ve varyansı hesaplanmıştır. Bu çalışma ile Pegasus ve İş Bankası hisse senetlerinin davranışlarını analiz etmek, gelecekteki değeri hakkında tahminlerde bulunmak temel amacımızdır. Zaman serilerinin gelecekteki değerlerini tahmin edebilmek için, verilerin geçmişteki değerlerinin volatilitelerini yorumlamak önemli bir süreçtir.

Volatilite, diğer adıyla oynaklık olarak da bilinen bir tür istatistiksel ölçüdür. Volatilite, belirsizliği ya da riski ifade etmektedir. Bir varlığın volatilitesi ne kadar düşükse, o varlığın riskinin düşük ve istikrarlı bir varlık olduğu, bir varlığın volatilitesi ne kadar yüksekse o varlığın yüksek riske sahip olduğu ve istikrarlı bir varlık olmadığı kanıtıdır. Volatilite, varyans ve standart sapma kullanılarak hesaplanmaktadır. Varyansın karekökü sonucunda standart sapma elde edilmektedir. Volatilite, bir varlığın ortalama değerinden, beklenenin üstünde ya da beklenenin altında değişim değerleri gösterir. Volatilite parametresinin değeri, politik ve sosyolojik gündemler, beklenmeyen ani durumlar, şirketlerin bilanço tabloları, merkez bankası kararları gibi etkenlerle yükseliş ya da düşüşe geçen ve volatilitenin değerindeki istikrarı bozan etkenlerdir (Özer, Karcioğlu, 2017).

İş Bankası ve Pegasus hissesi için, 1 Ocak 2017 ile 30 Nisan 2021 veri setinin standart sapma, standart sapmanın karesi olan varyans değerleri aşağıdaki tabloda yer almaktadır.

Getiri (Pegasus)	Standart Sapma	Varyans
Günlük	3%	0,1%
21 Günlük	16%	2,5%
63 Günlük	26%	6,9%

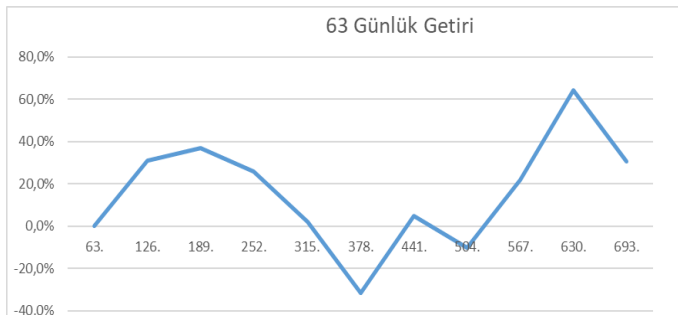
Tablo 2.5.a: Pegasus Hisselinin Standart Sapma ve Varyansı



Şekil 2.5.a: Pegasus Günlük Getiri Volatilitesi



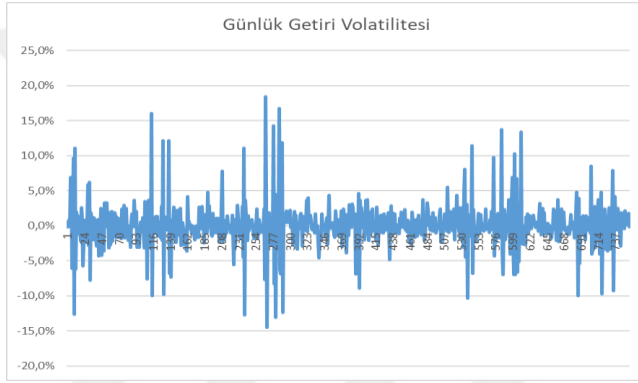
Şekil 2.5.b: Pegasus 21 Günlük Getiri Volatilitesi



Şekil 2.5.c: Pegasus 63 Günlük Getiri Volatilitesi

Getiri (İş Bankası)	Standart Sapma	Varyans
Günlük	3%	0,1%
21 Günlük	8%	0,6%
63 Günlük	12%	1,5%

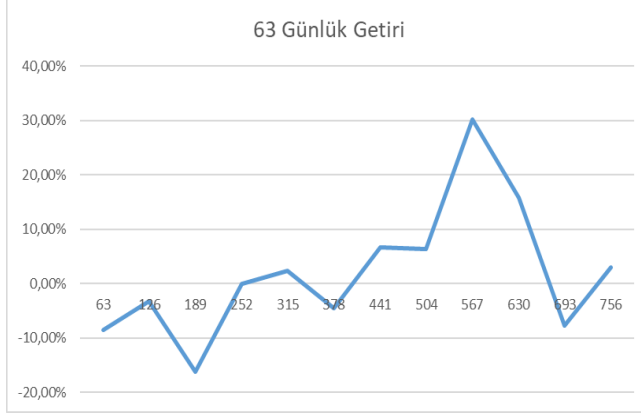
Tablo 2.5.b: İş Bankası Hissenin Standart Sapma ve Varyansı



Şekil 2.5.ç: İş Bankası Günlük Getiri Volatilitesi



Şekil 2.5.d: İş Bankası 21 Günlük Getiri Volatilitesi



Şekil 2.5.e: İş Bankası 63 Günlük Getiri Volatilitesi

Tablo 2.5.a'da ve **Tablo 2.5.b'de** görüldüğü üzere, Pegasus hissesinin standart sapma yüzdesi, İş Bankası hissesine göre çok yüksek olduğu için Pegasusun volatilitesi ve riski, İş Bankasına kıyasla daha yüksektir. Volatitenin yüksek olması, hissenin oynak olmasından kaynaklıdır. Bir hissenin fiyatının hızlı düşüş ya da yükseliş değerine sahip olduğunda tekabül eder.

Yukarıdaki grafiklerinde de görüldüğü üzere İş Bankası ve Pegasus hisse senedinin Excel platformunda günlük, 21 günlük ve 63 günlük volatiliteleri hesaplandığında, volatilitenin bir hesaplama yöntemi olan varyansın en yüksek değerinin her iki hisse için de 63 günlük, ay olarak baz alındığında 3 aylık bir bekleme süresinde gözlemlenmiştir.

İş Bankası ve Pegasus hissesinin davranışı Excel platformunda incelendikten sonra Matlab platformunda yapay sinir ağı (YSA) ve uzun kısa süreli bellek (LSTM) algoritmaları ile fiyat tahmini yapabilmek ve al sat emri yapabilmek için modeller, Pegasus ve İş Bankası hisselerinin kapanış fiyatları ile eğitilmiştir.

Yapay sinir ağı (YSA) modelini eğitirken Matlab platformundaki tahmin sonuçları confusion matris ekranı ile analiz edilmiştir. Confusion matris, yani karışıklık matrisi makine öğrenmesi algoritmalarında sınıflandırma algoritmalarının tahmin performansını analiz etmek için kullanılmaktadır. Confusion matris dört farklı değerlendirme ile analiz edilmektedir. Bu değerlendirmeler şu şekildedir;

- TP (True Positive) → 1 değerini, 1 tahmin etmekte
- TN (True Negative) → 0 değerini, 0 tahmin etmekte
- FN (False Negative) → 1 değerini, 0 tahmin etmekte

- FP (False Negative)→0 değerini, 1 tahmin etmekte

TN+TP toplamının yüksek değer çıkması, FN+FP toplamının da düşük değer çıkması, sınıflandırmanın başarılı sonuç verdiğini göstermektedir. FN+FP/Toplam oranı ile veri setinin hata oranı, TN+TP/Toplam oranı ile sınıflandırmanın doğruluk oranı elde edilmektedir. Confusion matrisinin hassasiyeti ise TP/TP+FP oranı ile elde edilmektedir. Confusion matrisinden sonra ise ROC eğrisine bakılmaktadır. ROC eğrisi, sınıflandırma performansının değerlendirildiği grafikdir. TP ve FP oranı ile çizilen grafikdir. TP oranı x eksenine, FP oranı ise y eksenine tekabül eder. ROC eğrisi altındaki alan AUC olarak tanımlanmaktadır. ROC eğrisi altında kalan eğri ne kadar büyükse, sınıflandırma modelleri sınıfları ayırt etmede daha doğru sonuç vermektedir. TN ve TP kesişmediği takdirde AUC değeri yüksek çıkmaktadır.

Matlab ile yapay sinir ağı (YSA) mimarisi ile çalışmamızda, BİST100 içerisinde yer alan İş Bankası ve Pegasus şirketi hisse senetlerinin tarihsel verileri ideal programından dosyaya aktarıldıktan sonra, tarihsel filtreleme ile veriler temizlenmiştir. Daha sonra veriler, test ve eğitim verisi olarak iki ayrı dosya haline getirilmiştir. Öncelikli olarak eğitim serisi oluşturulmuştur. Eğitim serisi içerisinde 1 Ocak 2017 ile 31 Aralık 2019 arasındaki veriler yer almaktadır. Yapay sinir ağı modeli ile eğitilen verilerin yaklaşık %70'i training olarak eğitilmiş olup, %30'u test veri seti olarak eğitilmiştir.

İlk olarak Pegasus hisse senedinin eğitim verileri yapay sinir ağı modeline uygun yazılan komutlarla eğitilmiştir. Altı farklı simülasyon kurularak, analizler yapılmıştır.

Yapay sinir ağı için; k, N, P değişkenlerine değerler verilerek, doğruluk oranı hesaplanmıştır.

K değişkeni pencere büyüklüğünü, N değişkeni nöron(katman) sayısını, P değişkeni de getirisinin seviyesi olarak tanımlanmıştır. Bu değişkenlere farklı değerler verilerek, yapay sinir ağı algoritması eğitilmiştir. Değişkenlerin aldıkları farklı değerlerde, doğruluk oranları karşılaştırılmıştır. Doğruluk oranı da TN+TP /Toplam ile hesaplanmıştır. Bizim çalışmamızda önem arz eden nokta ise TP (True Positive) değeridir. TP değeri, 1 değerini 1 tahmin eden değerdir.

K değişkeni 5, 7, 15 değerleri ile N değişkeni 30, 40, 60 değerleri ile P değişkeni de %3 ve %5 oranında alınarak, yapay sinir ağı algoritması her iki hisse için de eğitilmiştir

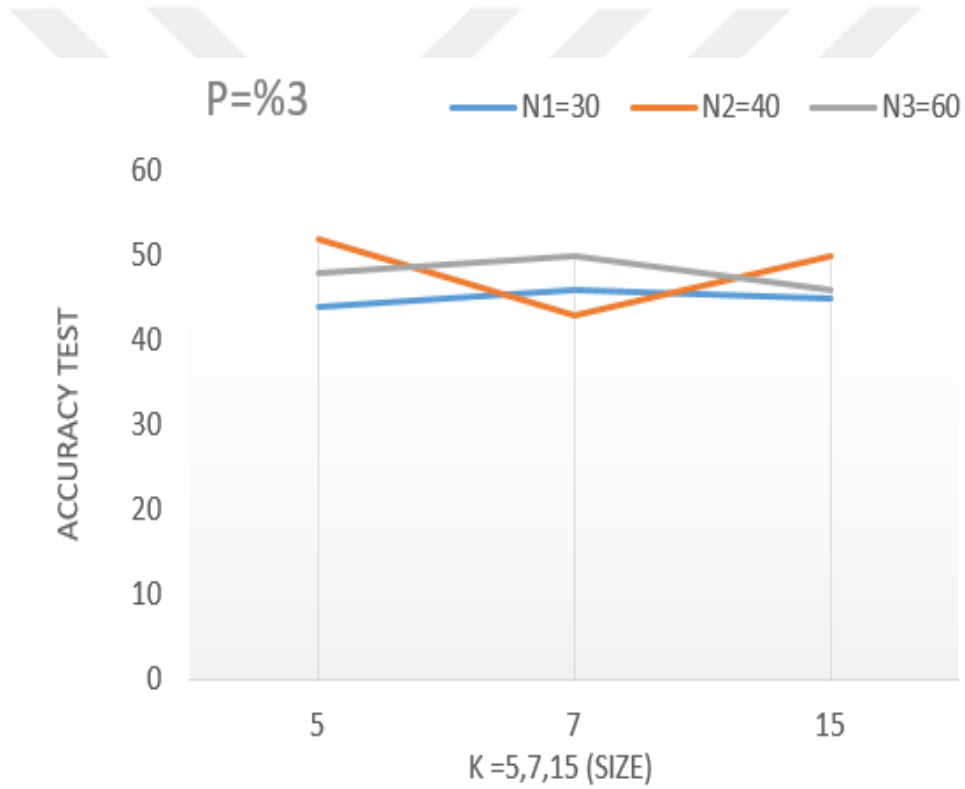
Aşağıdaki tablo ve grafiklerde hem Pegasus hem de İş Bankası için, TP değerleri yer almaktadır.

Pegasus Hisse Senetleri için PN Değerleri;

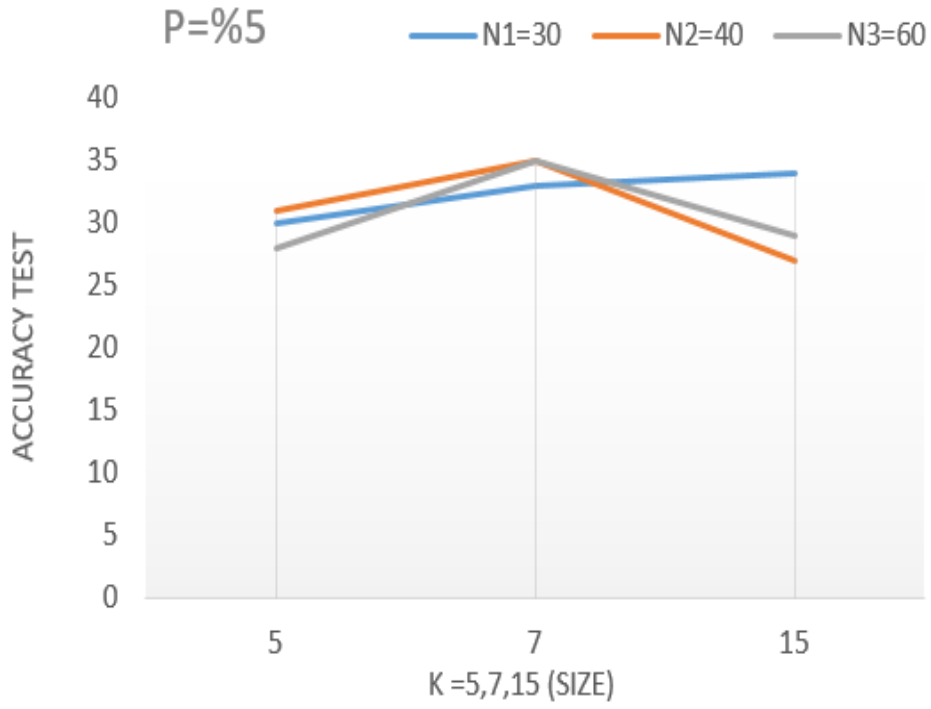
K	N1=30	N2=40	N3=60
5	44	52	48
7	46	43	50
15	45	50	46

K	N1=30	N2=40	N3=60
5	30	31	28
7	33	35	35
15	34	27	29

Tablo 2.5.c: Pegasus, YSA Algoritması İçin TP Değerleri



Şekil 2.5.f: Pegasus %3 Getiri İçin TP Değerleri

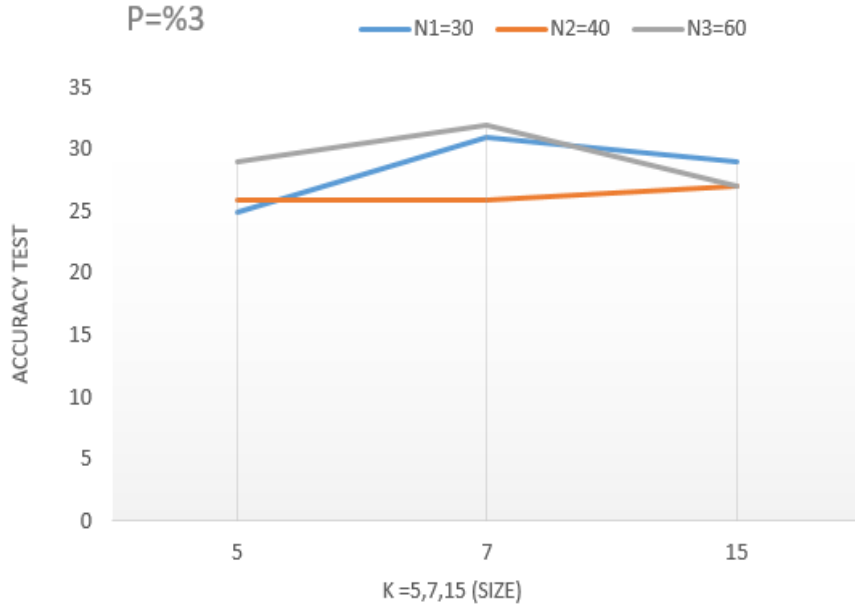


Şekil 2.5.g: Pegasus, %5 Getiri İçin TP Değerleri

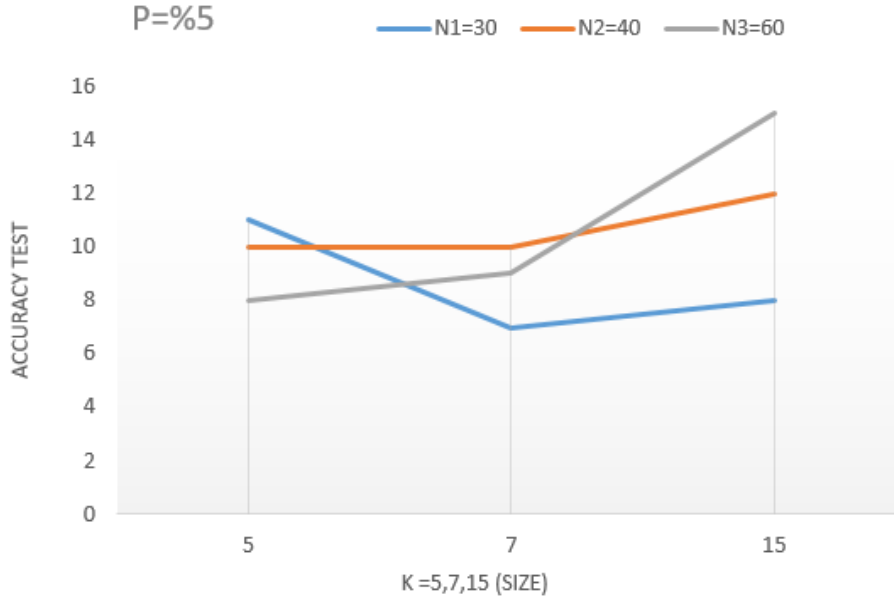
İş Bankası Hisse Senetleri için PN Değerleri;

İş Bankası, P=%3				İş Bankası, P=%5			
K	N1=30	N2=40	N3=60	K	N1=30	N2=40	N3=60
5	25	26	29	5	11	10	8
7	31	26	32	7	7	10	9
15	29	27	27	15	8	12	15

Tablo 2.5.d: İş Bankası, YSA Algoritması İçin TP Değerleri



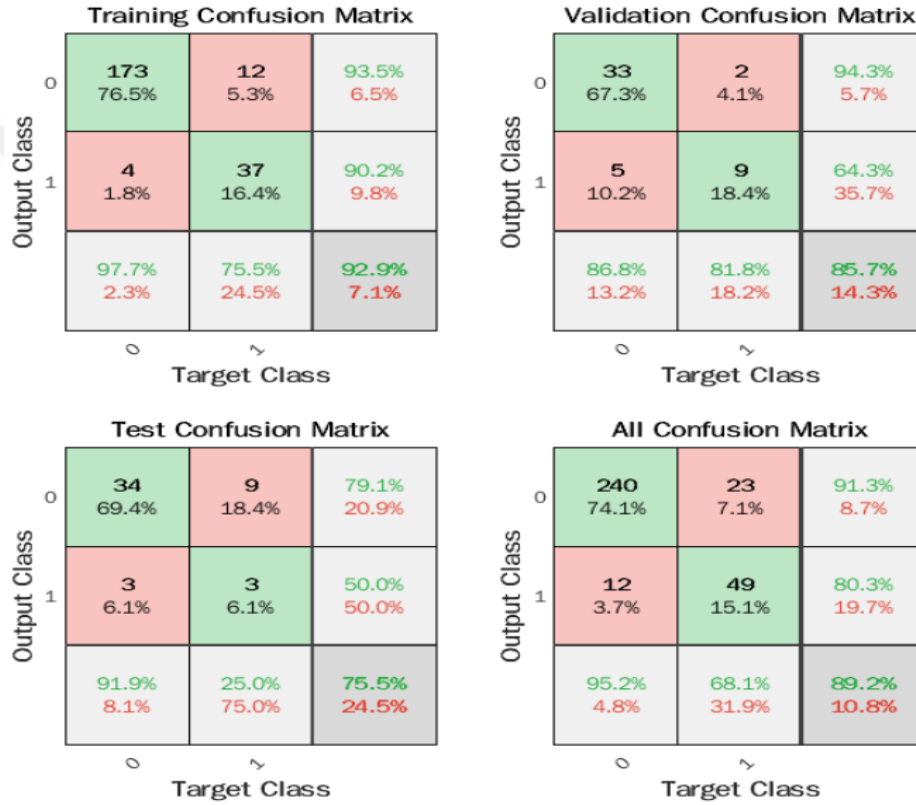
Şekil 2.5.h: İş Bankası, %3 Getiri İçin TP Değerleri



Şekil 2.5.ı: İş Bankası, %5 Getiri İçin TP Değerleri

324 adet eğitilen test verilerinin çıktıları incelendiğinde Pegasus hisse senedinde en yüksek doğruluk değerinin nöron sayısı 40, pencere büyüklüğü 5, P seviyesi de %3 olduğunda gözlemlenmiştir. %3 ile %5 seviyeleri karşılaştırıldığında, %3'lük seviyenin tüm çıktıları, %5 seviyesinden daha yüksek olduğu görülmektedir. İş Bankası için de aynı durum söz konusudur. %3'lük seviye, en yüksek TP değerlerine sahiptir. Eğitim

sonrasında trainig için çıktı sonucuna bakıldığında doğruluk oranı %89.2 iken, hata oranı %10.8 çıkmaktadır. All confusion matraste, 240 adet 0 tahmin edilmesi gereken çıktı 0 tahmin edilip, 49 adet 1 tahmin etmesi gereken çıktı da 1 olarak tahmin edildiği görülmektedir. Yaklaşık olarak %68.1 oranında 1 olan değeri doğru tahmin edip, %31.9 oranında ise 0 değerinin de 1 tahmin edildiği görülmektedir. Yapılan bu çalışmada önemli olan noktalardan biri de 1 olan değerleri 1 tahmin edilmesi, yani TP oranının yüksek olmasıdır. TN+TP oranı ne kadar yüksekse sınıflandırma o kadar hatasız çalışmaktadır.



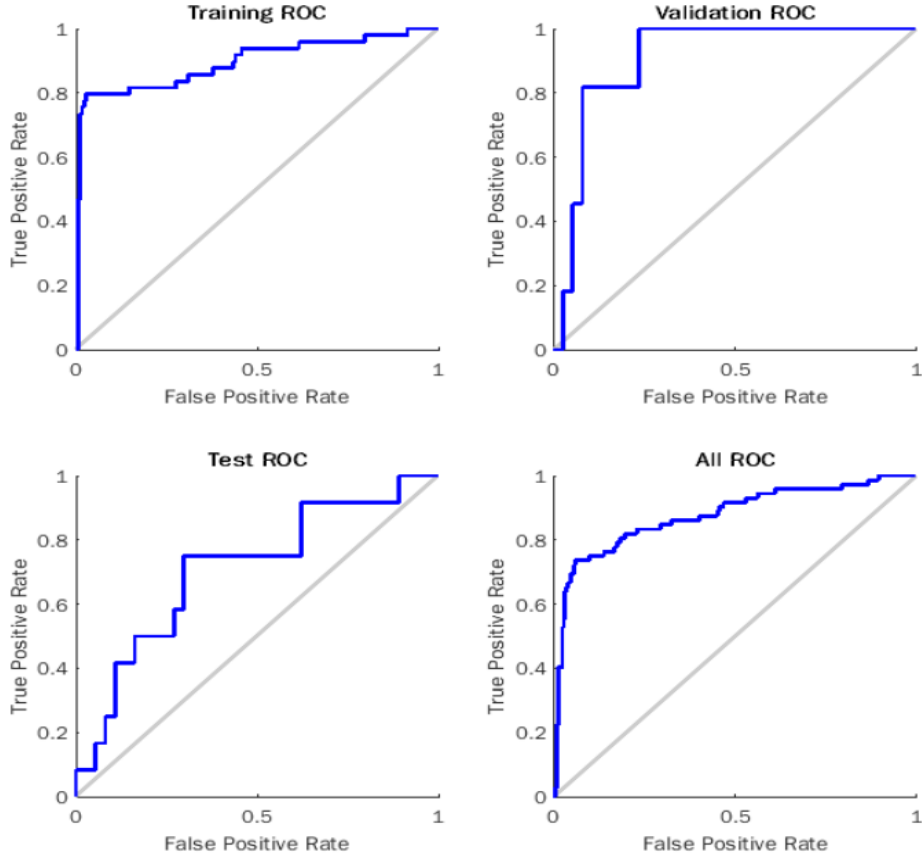
Şekil 2.5.i: Pegasus Hissesinin Confusion Matrisi

Pegasus hissesi için oluşturulan test verileri, 1 Ocak 2020 ile 30 Nisan 2021 arasındaki veriler olarak oluşturulmuştur. Pegasus hisse senedinin test verilerinin confusion matris (karışıklık matrisi) çıktıları detaylandırmıştır. Eğitilen verilerin çıktılarına bakıldığında 324 adet test verisinin 15 günlük pencere büyüklüğünde, nöron sayısının 40 ve getirinin değeri de %3 seçildiğinde diğer k, N ve P değerlerinin çıktılarına göre en iyi sonucu

verdiği

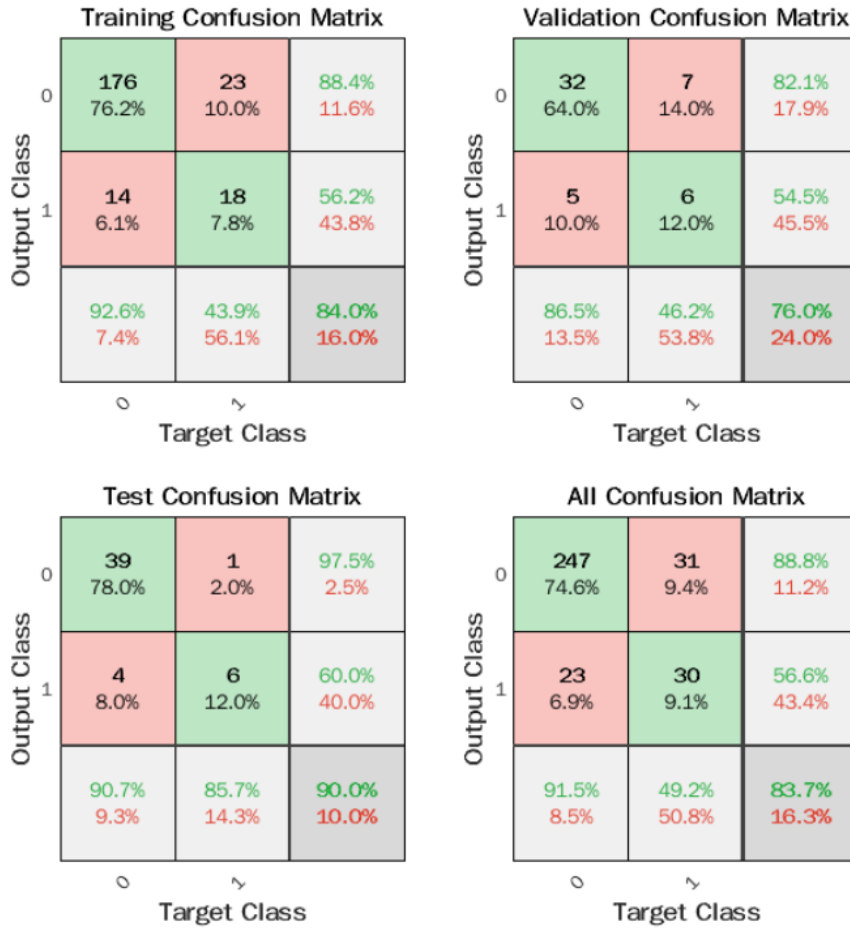
tespit

edilmiştir.



Şekil 2.5.j: Pegasus Eğitim Verisi ROC Eğrisi

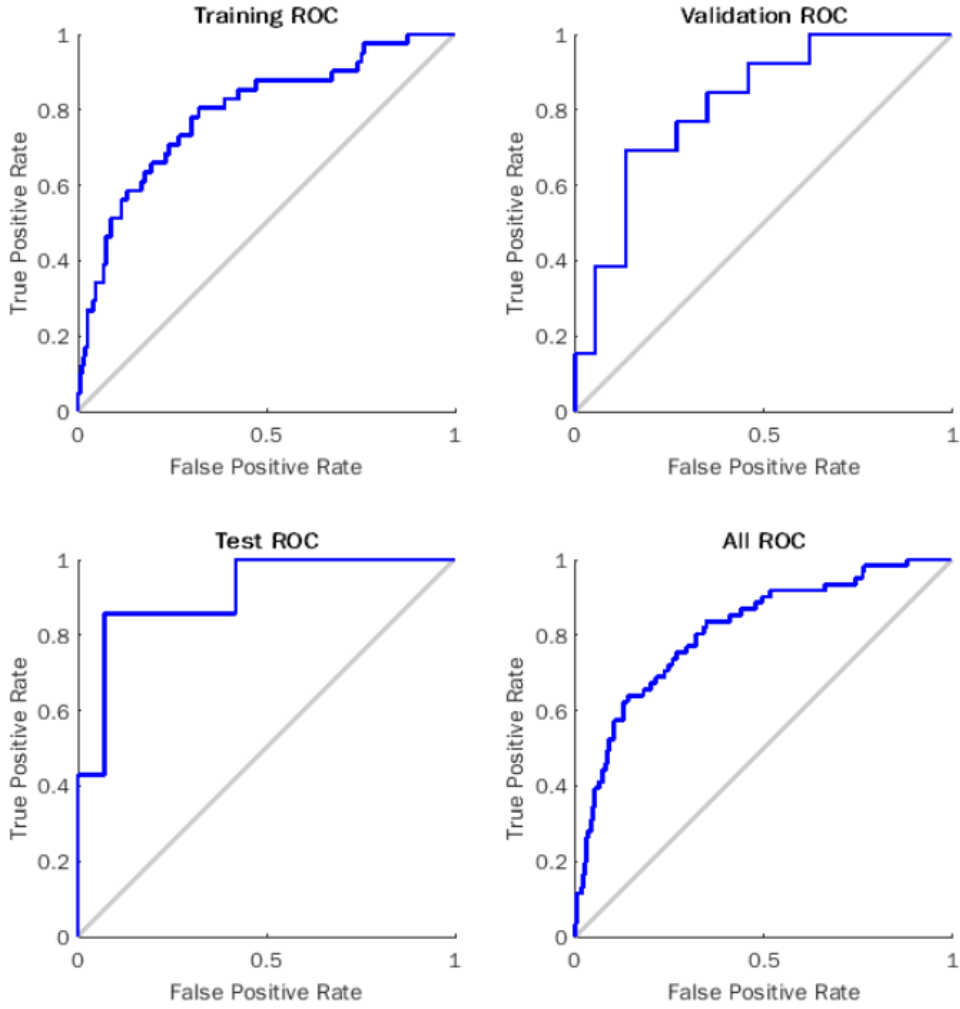
Pegasus hissesinin en iyi sonucu için Roc eğrisine bakıldığında ise training ve all (training, validation, test) için grafiğin altındaki en büyük alanın training ve all roc eğrilerine ait olduğunu görülmektedir. Bunun sebebi, trainingde 0 tahmin etmesi gereken çıktıların %97.7 oranında tahmini doğru yapmış olmasıdır.



Şekil 2.5.k: İş Bankası Hissinin Confusion Matrisi

Aynı çalışma, İş Bankası hissesinin 1 Ocak 2020 ile 30 Nisan 2021 arasındaki verileri için de yapılmıştır. Eğitilen verilerin çıktıklarına bakıldığında 324 adet test verisinin 7 günlük pencere büyüklüğünde, nöron sayısının 60 ve getirinin değeri de %3 seçildiğinde diğer k, N ve P değerlerinin çıktıklarına göre en iyi sonucu vermektedir. Confusion matris çıktısı detaylı incelendiğinde, training matrisindeki doğruluk oranının %84, hata oranının da %16 oranında olduğu görülmektedir. All (Training, validation, test) matrisine bakıldığında ise doğruluk oranının %83.7 oranında olduğu ve hata oranının da %16.3 olduğu görülmektedir. All matrisinde 247 adet 0 değerini 0 tahmin edildiği, 30 adet 1 değerinin de 1 tahmin edildiği ve 31 adet 0 değerinin de 1 tahmin edildiği görülmektedir.

Yapılan çalışmada amaç, TP değerlerinin yani 1 değerinin 1 tahmin edilmesinin yüksek oranda çıkmasını beklemektedir. Her ne kadar TP+TN oranı yüksek olsa da, TP'nin her iki hisse için de en iyi çıkan sonucun bile TP değerlerinin, TN değerlerine göre oldukça düşük olduğu görülmektedir.



Şekil 2.5.1: İş Bankası ROC Eğrisi

ROC eğrisi, y ekseninin 1 değerine yaklaştığı ve eğrinin altındaki alanın yüksek değerde olduğunda anlamlı sonuç vermektedir.

İş Bankasının training ve all (training, validation, test) grafiğın altındaki alanın 7 günlük pencere büyüklüğünde, nöron sayısının 60 ve getirinin değeri de %3 seçildiğinde diğer k, N ve P değerlerinin çıktıklarına göre en iyi sonucu veren eğri çıktısıdır.

Bunun sebebi, trainingde 0 tahmin etmesi gereken çıktıların %74.6 oranında tahmini doğru yapmış olmasıdır.

YSA için Pegasus ve İş Bankası verilerinin 1 Ocak 2020 ve 30 Nisan 2021 arasındaki verileri eğitildiğindeki doğruluk testlerini ve roc eğrilerini analiz ettik. Bu analizler sonucunda tespit edilen şudur ki; yapay sinir ağı algoritmasında 0 değerini 0 tahmin eden değer oldukça yüksek çıkması, 1 değerini 1 tahmin eden değer de TN'e göre daha düşük çıkmasıdır. Bizim beklentimiz ise TP değerinin TN'den daha yüksek çıkmasıydı. Yapay sinir ağı algoritması ile bu mümkün olmadı. İlerleyen çalışmalarda yağı sinir ağı için yazılan algoritmada güncelleme yapılması düşünülmektedir.

Yapay sinir ağı algoritması ile yazılan fonksiyonlar, Pegasus ve İş Bankası için karşılaştırıldıktan sonra, bir diğer modelimiz olan LSTM'de de verilerin analizi gerçekleştirildi.

Uzun kısa vadeli bellek (LSTM) modelini eğitirken, Matlab platformundaki tahmin sonuçları, RMSE değerine göre analiz edilmektedir. RMSE değeri, makine öğrenmesi algoritmalarının tahmin ettiği değerleri ile gerçek değerleri arasındaki uzaklık olarak tanımlanan, modelin hata değerini ölçen kuadratik bir ölçümdür. RMSE değeri 0'dan başlayarak sonsuza kadar tüm değerleri alabilmektedir. RMSE değerinde kök ortalama kare hata, ne kadar düşükse modelin performansı o derece iyidir. RMSE değeri 0 olduğunda, bu değer modelin hatasız çalıştığını göstermektedir. RMSE değerine göre modelin hatalı sonuç verip vermeyeceği analiz edilebilmektedir. RMSE, diğer adıyla karesel ortalama hata, hesaplanırken gözlemin gerçek değeri ile model sonucu elde edilen tahmin arasındaki farkın kareleri, ortalamasının karekökü ile hesaplanmaktadır. RMSE değerinin düşük olması tahmin edilen değer ile gerçek değer arasındaki farkın düşük olduğu göstermektedir.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}$$

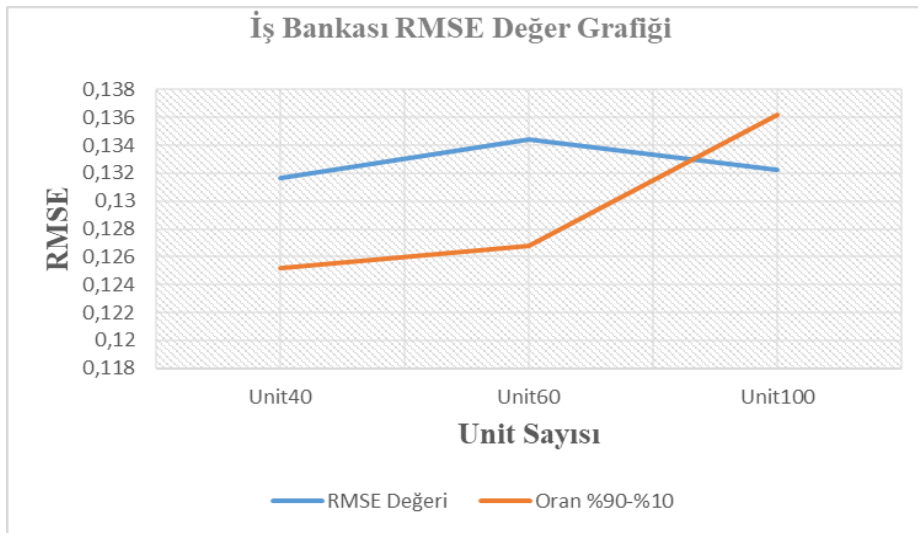
RMSE formülünde görüldüğü üzere, iki vektör arasındaki uzaklığın karesinin 1/n ile çarpımıdır. Burada da tahmin edilen değer ile gerçek değer arasındaki uzaklığın karesini temsil eder.

LSTM modeli ile eğitilen Pegasus ve İş Bankası hisse senetleri için öncelikli olarak, bu modele uygun komutlar yazılmıştır. 1 Ocak 2017 ve 30 Nisan 2021 arasında yer alan 1094 adet verinin tümü alınmış olup, yazılan fonksiyonla, %80 eğitim seti ve %20 test veri seti ile birlikte, %90 eğitim seti ve %10'u test veri seti olacak şekilde komutlar yazılıp, veriler eğitilmiştir. LSTM modelinde, eğitim ve test verilerinin ayrımı yapmak için yazılan komutlarla birlikte, unit sayısı diğer adıyla gizli birim, yapay sinir ağı modelindeki nöron sayısına tekabül eden değişken için değer atanır. Bu çalışmadaki LSTM modelinde üç farklı unit sayısı üzerinden veriler eğitilmiştir. Unit sayıları, 40, 60 ve 100 olarak belirlenmiştir. LSTM modelinden portföy değeri de 100 olarak tanımlanmıştır.

LSTM modelinde fonksiyonlar yazılıp, veriler sürekli olarak eğitilmiştir. Öncelikli olarak İş Bankası hisse senedinin verileri aşağıdaki değişkenler değiştirilerek eğitilmiştir.

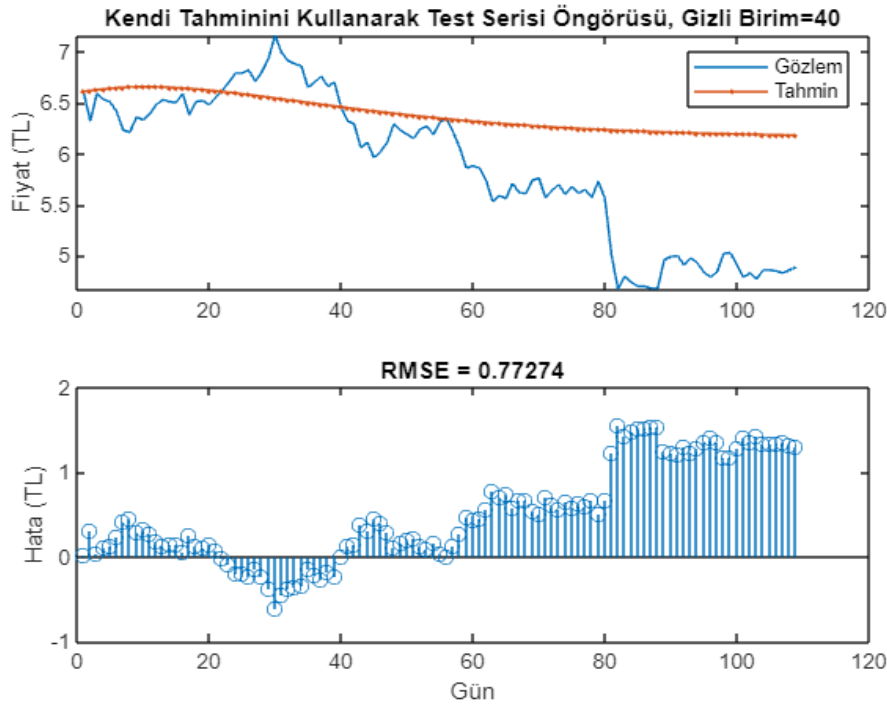
Unit Sayısı	Oran %80-%20	Oran %90-%10
Unit40	0,13167	0,12518
Unit60	0,13443	0,12676
Unit100	0,13223	0,1362

Tablo 2.5.e: İş Bankası, LSTM Değişken Değerleri

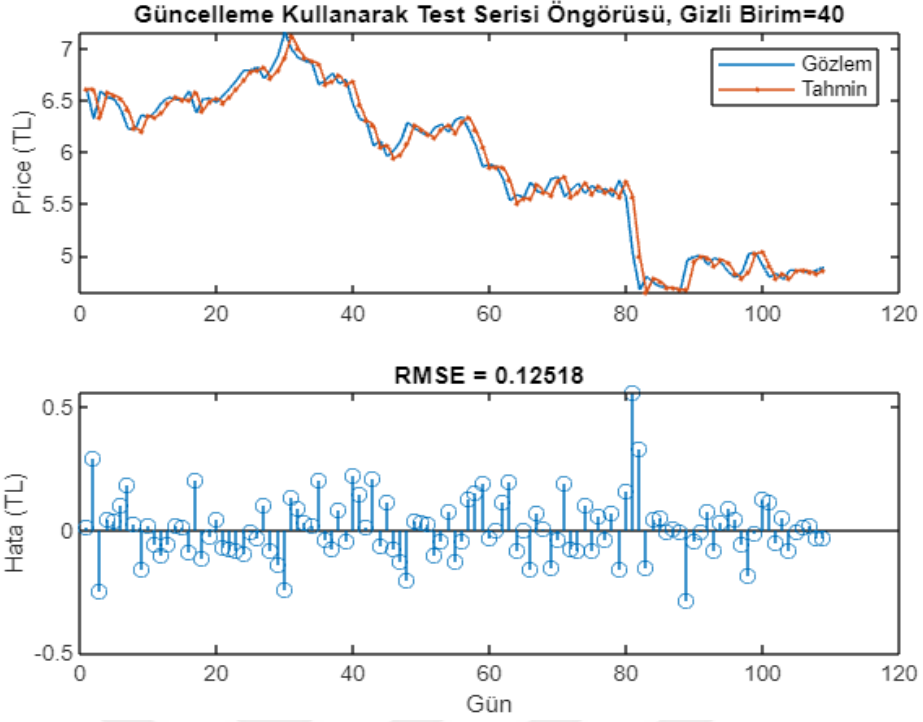


Şekil 2.5.m: İş Bankası, RSME Grafiği

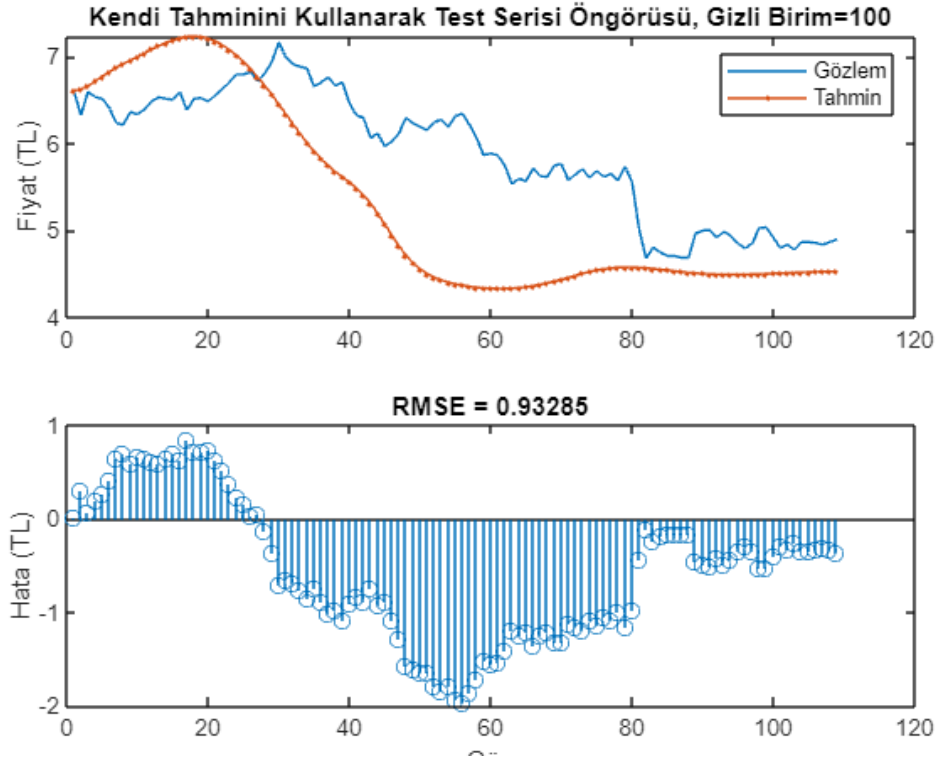
Yukarıdaki grafikte görüldüğü üzere portföy değerinin sabit kalındığı ve 100 alındığı LSTM modeli, ilk olarak İş Bankası hissesi verileriyle eğitildikten sonra, %80 eğitim, %20 test verisi ile %90 eğitim, %10 test verisi seçilerek eğitilen modelde, en düşük RMSE değerine sahip nokta çıktısı analiz edilmiştir. Verilerin eğitilmesi sonucu, %80-%20 aralığı ile birlikte unit sayısı da (gizli katman) 40 değerinde seçildiğinde, İş Bankası için en düşük RMSE değeri olan 0,12518 çıktı değeri bulunmuştur. RMSE değeri, sıfırına ne kadar yakınsa gerçek değer ile tahmini sonuç arasındaki fark o kadar azdır. En yüksek RMSE değeri ise, yine %90-%10 aralığı ile birlikte unit sayısının (gizli katman) 100 değerinde seçilmesi sonucu çıkmıştır. Bu analiz gösteriyor ki, gizli katman sayısı RMSE değerini doğrudan etkilememektedir. LSTM modeli, İş Bankası gibi spekülasyon olmayan bir hisse için riski düşük bir model diye düşünülmektedir. İş Bankası hissesi için en yüksek ve en düşük RMSE değer grafikleri aşağıda yer almaktadır.



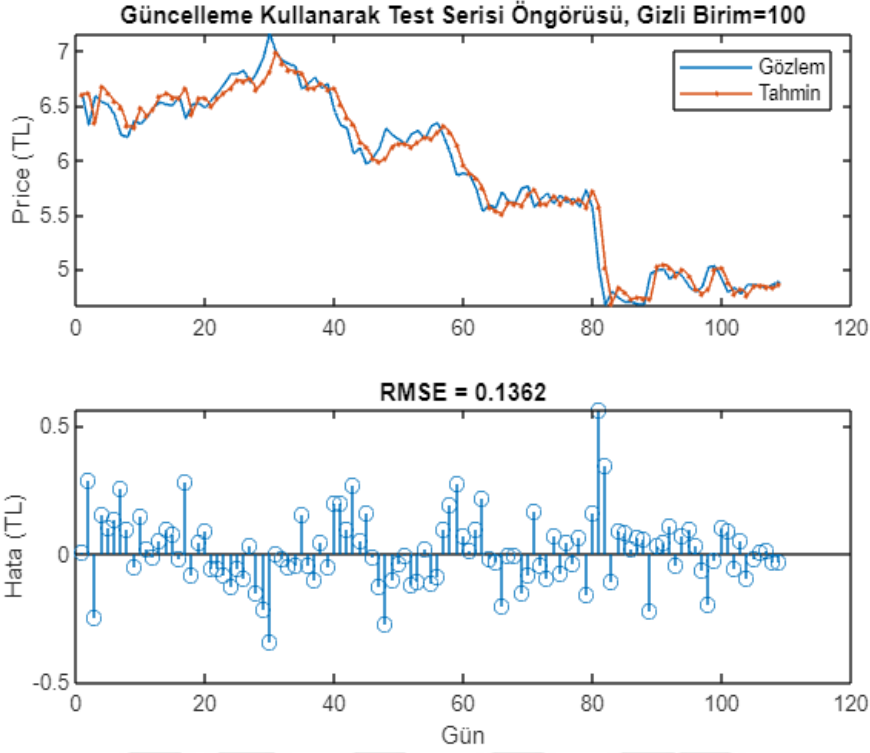
Şekil 2.5.n: İş Bankası %90-%10 RMSE Değeri (Gizli Birim:40)



Şekil 2.5.o: İş Bankası %90-%10 RMSE Değeri (Gizli Birim:40)



Şekil 2.5.ö: İş Bankası %90-%10 RMSE Değeri (Gizli Birim:100)



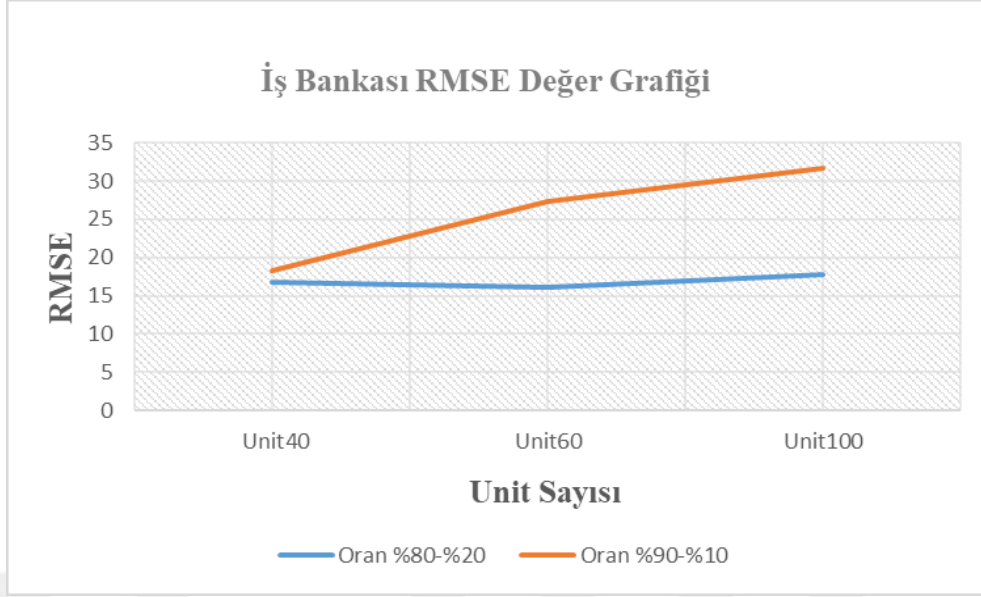
Şekil 2.5.p: İş Bankası %90-%10 RMSE Değeri (Gizli Birim: 100)

Grafik 2.5.o grafiğinde İş Bankası hisse senedinin güncelleme sonrasındaki RMSE değerinin gösterildiği grafik mevcuttur. Tahmin edilen değer ile gerçek değer arasındaki farkın karesinin karekökü olduğundan, hata oranının sifıra yakın olduğunu göstermektedir. Bu da gerçek değer ile tahmini değer arasındaki farkın düşük olmasıdır. Bu çalışmadaki amaç da hata oranının mümkün olduğunca düşük olmasıdır.

Aynı model Pegasus hisse senedi verileri için de eğitilmiştir. Pegasus hisse senedine ait RSME ve unit sayıları aşağıdaki tabloda detaylandırılmıştır.

Unit Sayısı	Oran %80-%20	Oran %90-%10
Unit40	3,3154	3,4393
Unit60	7,6525	3,4198
Unit100	2,5956	3,3405

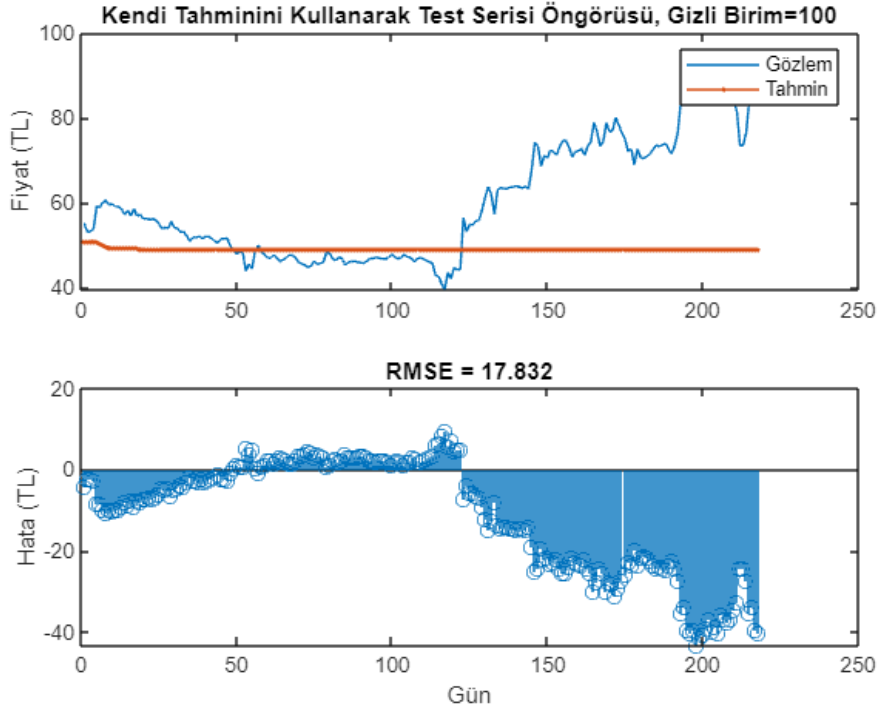
Tablo 2.5.f: Pegasus, LSTM Değişken Değerleri



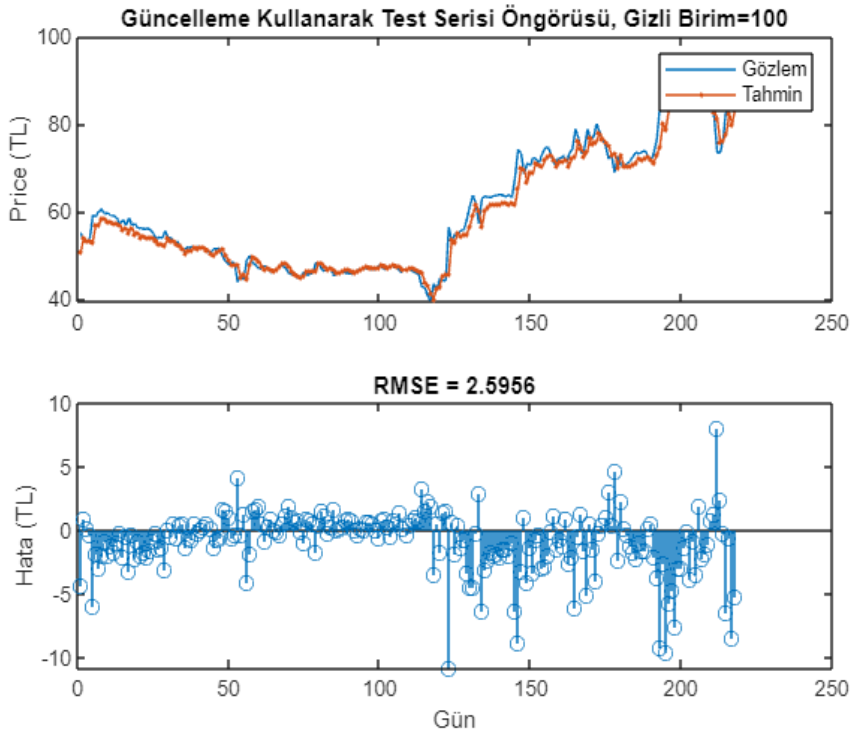
Şekil 2.5.r: Pegasus, RSME Grafiği

Grafik 2.5.s ve **Tablo 2.5.f**'de Pegasus hisse senedi için LSTM modelinin çıktıları ve RMSE değer grafiği yer almaktadır. İş Bankası hisse senedinde olduğu gibi portföy değerinin sabit kaldığı ve 100 alındığı LSTM modeli, %80 eğitim, %20 test verisi ile %90 eğitim, %10 test verisi seçilerek eğitilen modelde en düşük RMSE değerine sahip noktayı bulmak için veriler eğitilmiştir. Eğitim sonucu Pegasus hisse senedi için en düşük değerin 2,5956 olduğu görülmektedir. Bu değer %80-%20 aralığı seçilip, unit sayısı da 100 olduğunda çıkan sonuçtur. Pegasus için en yüksek RMSE değeri ise, 7,6525 değerinde olup, %80-%20 aralığı ile birlikte unit sayısının 60 değerinde seçilmesi sonucunda çıkmıştır.

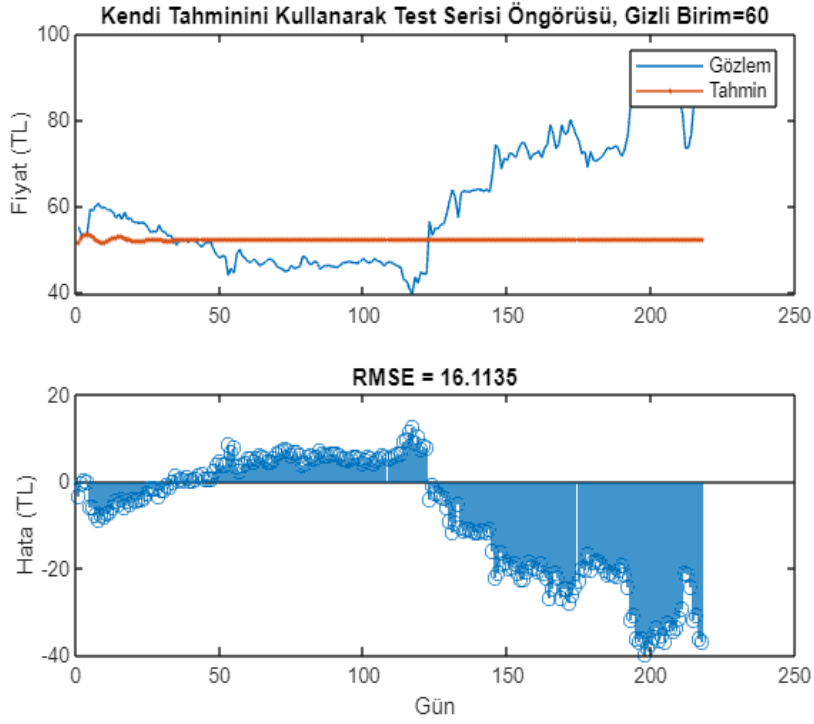
Pegasus hissesi için RMSE değerlerinin, sıfırdan çok uzak değerlerde çıktığı görülmektedir. RMSE değerinin her zaman sıfıra çok yakın olması, LSTM modeli için önem arz etmektedir. Sıfırdan uzak olan RSME değeri, tahmin edilen değer ile gerçekleşen değer arasındaki farkın yüksek olduğunu gösterdiği için, Pegasus hisse senedi için bu modelin istenilen sonucu vermesi düşük ihtimal olarak düşünülmektedir.



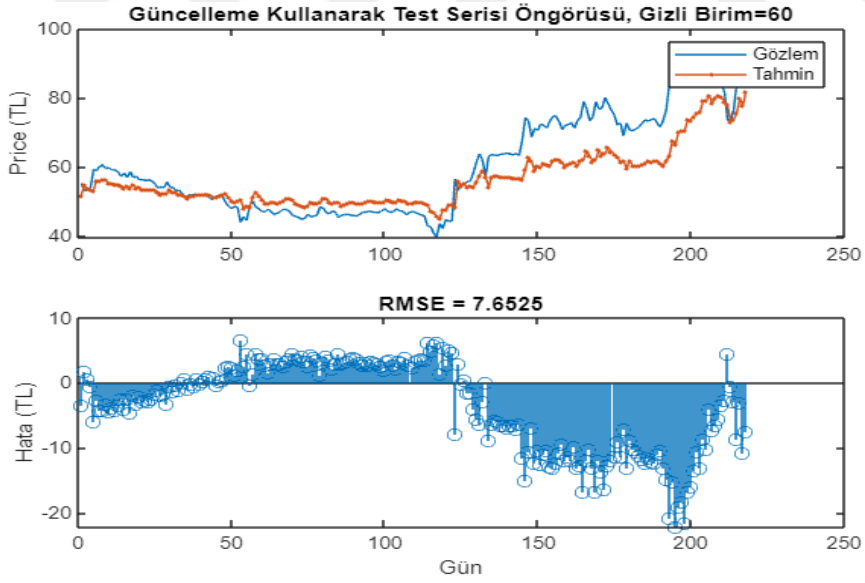
Şekil 2.5.s: Pegasus, %80-%20 RMSE Değeri (Gizli Birim:100)



Şekil 2.5.t: Pegasus, %80-%20 RMSE Değeri (Gizli Birim:100)



Şekil 2.5.u: Pegasus, %80-%20 RMSE Değeri (Gizli Birim:60)



Şekil 2.5.v: Pegasus, %80-%20 RMSE Değeri (Gizli Birim:60)

Pegasus hisse senedine ait en düşük ve en yüksek RMSE deęerlerinin grafiklerinde de grldę zere, olduka yksek ve sıfır deęerinden uzak olduęu grlmektedir. Pegasus hisse senedi iin LSTM modelinin riskli bir model olabileceęi dşnlmektedir. Pegasus hisse senedinin, İř Bankası hisse senedine gre oynaklıęı dięer adıyla volatilitesi yksek olduęu iin, LSTM modelinin Pegasus hisse senedi iin anlamlı bir model olmadıęı dşnlmektedir. İř Bankası hisse senedi iin ise tam aksi dşnlmektedir. İř Bankası hisse senedi verileri, LSTM modelinde eęitildięinde sıfıra yakın deęerde ıktıęı iin LSTM modeli, İř Bankası hisse senedi iin daha anlamlı olduęu ve İř Bankası hisse senedinin volatilitenin dşk olmasının da RSME deęerini etkiledięi dşnlmektedir.



3. SONUÇ

Bu çalışmada yapılan arařtırmalar ve analizlerde en iyi sonucu elde etmek için veriler öncelikli olarak kullanılabilir olması adına istenen formata getirilmiştir. Kapanış fiyatı, temel alınarak BİST100 endeksi içerisinde yer alan Pegasus ve İş Bankası hisse senedi seçilmiştir. Veriler, Excel programında BİST100 endeksinin kapanış fiyatı üzerinden analiz edilmiştir. Volatiliteleri hesaplanarak, hisse senetlerinin yıllar içerisindeki davranışı incelenmiştir. Günlük veri üzerinden strateji oluşturmanın yatırımcının portföyü için oldukça riskli olacağı sonucuna varılmıştır. Doğruluk değeri dengeli olmadığı için, günlük getiriye baėlı portföy oluşturmak, kesin bir sonuç elde edilemeyeceğini göstermektedir. Burada yapılan çalışmada Matlab ve Excel platformları kullanılarak Pegasus ve İş Bankası hisse senedinin gelecekteki fiyat tahmini yapılmaya çalışılıp, al sat emri verebilecek bir algoritma yaratılmaya çalışılmıştır. 1 Ocak 2017 ile 30 Nisan 2021 tarihleri arasındaki veriler eğitilmiştir. Matlab platformunda iki farklı makine öğrenmesi algoritması için model oluşturulmuştur. Bu algoritmalar, yapay sinir aėı (YSA) modeli ve uzun kısa süreli bellek (LSTM) algoritmasıdır. Bu iki model için kodlar yazılıp, ayrıştırılan veriler Matlab platformuna aktarılmıştır. Öncelikli olarak yapay sinir aėı (YSA) modeli eğitilmiştir. Yapay sinir aėı modelinde k (pencere büyüklüğü), N (katman sayısı) deėişkenleri sürekli deėiştirilerek veriler eğitilmiştir. Eğitim sonunda doğruluk oranı yüksek olan ve hata oranı düşük olan modeller seçilmiştir.

Bu çalışmada orta vadeli hisse senetlerinin yönlü tahmini için çalışma yapılmıştır. Eğitilen veri setleri sonucunda tüm algoritmalar ve sonuçlar birbiri ile kıyaslanmıştır. Yapay sinir aėı (YSA) algoritması ve uzun kısa süreli bellek (LSTM) algoritması için farklı fonksiyonlar yazılıp, doğruluk değeri ve grafik sonuçları incelenmiştir. İş Bankası ve Pegasus hisse senedinin volatiliteleri birbiri ile kıyaslandığında, Pegasus hisse senedinin volatilitelerinin, İş Bankasına göre daha yüksek bir değere sahip olduėu görülmüştür. Daha sonra Matlab platformu aracılığı ile yapay sinir aėı için modelleme yapılmış olup, volatilitesi yüksek olan Pegasus hissesinin, İş Bankası hissesine kıyasla confusion matrisindeki TN+TP oranının yüksek olduėu görülmüştür. Yapay sinir aėı algoritmasından sonra, Matlab'de uzun kısa süreli bellek (LSTM) modellemesi için veriler eğitilmiştir. LSTM modelindeki eğitilen verilerin

çıktılarına bakıldığında ise, RMSE değeri olan hata oranı, İş Bankası hissesinde, Pegasus hisse senedine kıyasla çok düşük olduğu sonucuna varılmıştır. RSME değeri, gerçekleşen değer ile tahmin edilen değer arasındaki farka eşit olduğundan, RSME hata oranının düşük olması modelin anlamlı çıktı verdiğini göstermektedir. İş Bankasının hata oranının Pegasus hisse senedine kıyasla düşük bir değerde çıkmasının sebebi, İş Bankası hissesinin volatilesinin düşük olmasının bir nedeni olduğu düşünülmektedir. LSTM modelinde Pegasus hissesinin RMSE değerinin, diğer adıyla hata oranının yüksek çıkması aynı şekilde Pegasus hisse senedinin yüksek volatiliteye sahip olduğundan kaynaklı olduğu düşünülmektedir.

LSTM ve YSA algoritmaları birbiri ile kıyaslandığında, her ikisinin de farklı hisse senetleri için farklı sonuçlar verdiği, yapay sinir ağı algoritmasının hata ve doğruluk oranlarının yaklaşık olarak birbiri ile aynı oranda olduğu ve bu durumdan dolayı da riskli bir model olacağı düşünülmektedir. LSTM modeli ise, volatilesi düşük olan hisse senedi için hata oranı düşük olduğunda anlamlı çıktı verdiği düşünülmektedir. Fakat LSTM modeli, volatilesi yüksek olan hisse senedinde, hata oranı yüksek çıktığı için anlamlı çıktı vermediğini düşünülmektedir.

Yapay sinir ağı modeli için volatilesi yüksek hisseler ele alınıp ve nöron sayısı ile birlikte k değeri olan pencere büyüklüğü değeri de değiştirilerek daha anlamlı sonuçlar elde edileceği düşünülmektedir. Bununla ilgili ilerleyen zamanlarda çalışma yapılması gerekmektedir.

LSTM modeli için ise, İş Bankası ve Pegasus hissesi karşılaştırıldığında RSME değerinin en düşük olan değerinde anlamlı çıktı ortaya çıkmıştır. Pegasus hisse senedi için en yüksek RSME gizli birimin 60 değerinde olup, oranın ise %80-%20 olduğunda, en düşük RSME değerinin de gizli birimin 100 değerinde olup, oranın %80-%20 da olduğu gözlemlenmiştir. İş Bankası için ise en yüksek RMSE değerinin 100 gizli birimde, %90-%10 oranında gerçekleştiği, en düşük RSME değerinin ise 40 gizli biriminde %90-%10 oranında gerçekleştiği gözlemlenmiştir. Pegasus ve İş Bankası hissesinin düşük değerdeki RSME'lerini karşılaştırdığımızda gizli birim sayısının, İş Bankası hissesinde, Pegasus hissesine göre daha düşük bir değere sahip olduğu görülmektedir. Volatilesi yüksek olan Pegasus benzeri hisseler için LSTM modeli değil de yapay sinir ağı modelinin daha mantıklı olduğu, volatilesi düşük olan İş Bankası ve benzeri hisseler için ise LSTM modelinin YSA modelinden daha anlamlı

sonuç verdiđi düşünölmektedir. Buradaki sonuç, verilerin volatilitesine uygun model yaratmak diye düşünölmektedir. Bununla beraber aynı sektöreden birden fazla hisse senedi için portföy yaratılıp, sektör bazlı bir analiz yapılması, verilerin tarih aralığına bađlı, aynı tarih aralığına sahip makroekonomik deđişkenlerle incelenip, verilerin eđitilmesi ve yorumlanması gerekmektedir. İleride bu çalışmaların yapılması düşünölmektedir. Bu çalışmanın temel amacı aşıđıdaki iki soruya yanıt bulmaya çalışmaktı.

- Makine öğrenmesi metotları arasında en dođru ve anlamlı tahmini yapan öğrenme modeli hangisidir?
- Yön tahminini kapanış fiyatı üzerinden analiz etmek anlamlı bir sonuç çıkarır mı?

Bu çalışma sonucunda, hisselerin volatilitelerine göre her iki algoritmanın da, her iki hisse senedi için farklı sonuçlar verdiđi, kullanılan algoritmalarındaki deđişkenlere verilen deđerlerin artırılması ve hisse senetlerinin sektör bazlı ayrıştırılması ile birlikte, yön tahmini için ise kapanış fiyatı yeterli olmakla birlikte, açılış fiyatı üzerinden de çalışma yapılması ilerleyen zamanlarda düşünölmektedir.

KAYNAKÇA

- Akay. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenimi. , *Social Sciences Research Journal*.2. , pp.45-49
- Alpay (2020).LSTM Mimarisi Kullanarak USD/TRY Fiyat Tahmini.,*Avrupa Bilim ve Teknoloji Dergisi* , pp.453-456
- Altaş, Gülpınar. (2012). Karar Ağaçları ve Yapay Sinir Ağlarının Sınıflandırılma Performanslarının Karşılaştırılması: Avrupa Birliği Örneği. , *Trakya Üniversitesi Sosyal Bilimler Dergisi*.14. , pp.3-7
- Aydın. (2019). Finansal Bilgi Manipülasyonunun Denetimli Makine Öğrenmesi Yöntemleri Kullanılarak Tahmin Edilmesi: Destek Vektör Makinesi, Olasılıksal Sinir Ağı, K-en Yakın Komşu ve Karar Ağaçları Kullanımı. , *Sosyal Bilimler Enstitüsü TOBB Ekonomi ve Teknoloji Üniversitesi*. , pp.38-43
- Becker, Leschinski. (2018).Directional Predictability of Daily Stock Returns., *Econstor Make Your Publications Visible*. , pp.1, 9
- Chung, Hong. (2007).Model-Free Evaluation of Directional Predictability in Foreign Exchange Markets. , *Journal of Applied Econometrics*. , pp.856-858
- Constantinides, Jackwerth, Perrakis. (2009).Mispricing of S&P 500 Index Options. , *The Review of Financial Studies*. , pp.12448-1251
- Çınaroğlu, Avcı. (2020).THY Hisse Senedi Değerinin Yapay Sinir Ağları İle Kestirimi. , *Atatürk Üniversitesi İktisadi ve Bilimler Dergisi*.34, pp.3-8
- Dağlıoğlu, Kural. (2018).Hisse Senedi Piyasa Fiyatlarının Saklı Markov Modeli ile Tahmin Edilmesi: Türkiye Örneği. , *Uluslararası Ekonomi ve Yenilik Der*.4. , pp.62-67
- Dai, Dong, Kang, Hong. (2020).Forecasting Stock Market Returns: New technical indicators and two-step economic constraint method. , *North American Journal of Economics and Finance*. , China
- Derin, Türkoğlu. (2018). Derin Öğrenme Modelleri ve Uygulama Alanları İlişkin Bir Derleme. , *DÜMF Mühendislik Dergisi*. , pp.409-410
- Ercan, Irmak, Çevik, Canbazoğlu. (2020).Yapay Sinir Ağları Kullanılarak Konutlarda Elektrik Tüketimi Düzeylerinin Tahmin Edilmesi. , *Sosyoekonomi*.28. , pp. 173-174
- Fahri Keskenler, Furkan Keskenler. (2017). Geçmişten Günümüze Yapay Sinir Ağları ve Tarihçesi. , *Takvim-i Vekayi*. , pp.9-10
- Filiz, Karaboğa, Akoğul. (2017).Bist50 Endeksi Değişimi Değerlerinin Sınıflandırılmasında Makine Öğrenmesi Yöntemleri ve Yapay Sinir Ağları Kullanımı. , *Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi*.26. , pp.232, 235

- Hiransha, Gopalakrishnan, Krishna, Soman. (2018). NSE Stock Market Prediction Using Deep-Learning Models . , *Procedia Computer Science.* , pp.1355-1356
- Jerez, Kristjanpoller. (2020). Effect of the Validation Set on Stock Return Forecasting. , *Expert Systems with Applications.* , pp.1-2
- Karciođlu, Özer. (2017). Hisse Senedi Piyasasında Yılın Ayları Anomalilerinin Getiri ve Volatilite Üzerindeki Etkisinin İncelenmesi: Borsa İstanbul Uygulaması., *Atatürk Üniv. Sosyal Bil.Enst..* , pp.1572-1573
- Kartal. (2015). Sınıflandırmaya Dayalı Makine Öğrenmesi Teknikleri ve Kardiyolojik Risk Değerlendirmesine İlişkin Bir Uygulamalar. , *İstanbul Üniversitesi Fen Bilimleri Enstitüsü Enformatik Anabilim Dalı.* , pp.5-7
- Kayaalp, Süzen. (2018). Derin Öğrenme ve Türkiye'deki Uygulamaları. , *İKSAD.*, pp.16-19
- Kızrak, Bolat. (2019). Uçak Motoru Sağlığı için Uzun Kısa Süreli Bellek Yöntemi ile Öngörücü Bakım. , *Bilişim Teknoloji Dergisi.* , pp. 105-106
- Lanbouri, Achchab. (2020). Stock Market prediction on High frequency data using Long- Short Term Memory., *Procedia Computer Science.175.* , pp.603, 604, 607
- Masum, Shahriar, Haddad, Alam. (2020). r-LSTM: Time Series Forecasting for Covid-19 Confirmed Cases with LSTM based Framework. , *IEEE International Conference on Big Data.* , pp.1376-1378
- Namini, Tavakoli, Akbar. (2019). A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM and BiLSTM. , *arXiv:1911.09512v1.* , pp.3-4
- Nevalsami. (2020). Forecasting Multinormal Stock Returns Using Machine Learning Methods. , *The Journal of Finance and Data Science.6.* , pp.87, 88, 92
- Özçalıcı. (2016). Hisse Senetlerinin Öz düzenleyici Haritalarla Kümelendirilmesi: BİST50 Endeksinde Yer Alan Hisseler Üzerine Bir Uygulama. , *İstanbul Üniversitesi İşletme Fakültesi Dergisi.1.* , pp.23-24
- Özçalıcı. (2016). Yapay Sinir Ağları ile Çok Aşamalı Fiyat Tahmini: BİST30 senetleri Üzerine Bir Araştırma. , *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Dergisi.3.* , pp.210-213
- Özçalıcı. (2017). Aşırı Öğrenme Makineleri İle Hisse Senedi Fiyat Tahmini. , *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi.1.* , pp.69-74
- Özdemir, Aksoy. (2012). Türkiye'de Makroekonomik Değişkenler ve İstihdam Teşviklerinin İstihdam Üzerindeki Etkileri. , *Akdeniz İ.İ.B.F.24.* , pp. 103-105

- Özer, Sarı, Başakın. (2018).Bulanık Mantık ve Yapay Sinir Ağları İle Borsa Endeks Tahmini: Gelişmiş ve Gelişmekte Olan Ülkeler Örneği. , *Hitit Üniversitesi Sosyal Bilimler Dergisi.1.* , pp.111-112
- Öztürk, Şahin (2018).Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. , *Takvim-i Vekayi.2.* , pp.26-32
- Sarıkaya. (2019).Yapay Sinir Ağları ve Regresyon Modelleri ile Bist Ulusal-100Endeksinin Tahmini. , *21.Yüzyılda Eğitim ve Toplum23.* , pp.327-331
- Sülkü, Ürkmez.(2018).Hisse Senedi Getirilerinde Doğrusal Olmayan Dinamikler: Türkiye'den Kanıtlar. , *Uluslararası İktisadi ve İdari İncelemeler Der.18.* , pp.480-481
- Süzen, Yıldız, Yılmaz. (2019).LSTM Tabanlı Derin Sinir Ağı İle Ayak Taban Basınç Verilerinden VKİ Durumlarının Sınıflandırması.,*BEÜ Fen Bilimleri.* , pp.1392-1393
- Şişmanoğlu, Koçer, Önde, Şahingöz. (2020).Derin Öğrenme Yöntemleri İle Borsada Fiyat Tahmini. , *BEÜ Fen Bilimleri Dergisi.9.* , pp.435-440
- Şeker. (2017). Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. ,*Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi* , pp.12-14
- Şeker, Diri, Balık.(2017).Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme., *Gazi Journal of Engineering Sciences.* , pp.51-52
- Xu, Wang, Jiang, Liu. (2019). A Novel (U) MIDAS-SVR Model With Multi Source Market Sentiment For Forecasting Stock Returns. , *Neural Computing and Applications.* , London, pp.5875, 5886
- Yamak, Samut, Koçak. (2018). Farklı Frekanslı Veriler Altında Ekonomik Büyüme Oranının Tahmini. , *Ekonomi Bilimleri Dergisi.* , pp.36
- Yıldırım, Eren. (2020).Beta Portföylerin Performans Analizi: Borsa İstanbul Örneği. , *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Dergisi.38.* , pp.169-175
- Zhong and Enke. (2017). A comprehensive cluster and classification mining procedure for daily stock market return forecasting. , *Neurocomputing267.* , pp.153-165

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı :Yasemin Koç

Eğitim Durumu

Lisans Öğrenimi: İstanbul Üniversitesi / Fizik

Yüksek Lisans Öğrenimi: Kadir Has Üniversitesi / Finans Mühendisliği

Bildiği Yabancı Diller: İngilizce

İş Deneyimi

Çalıştığı Kurumlar ve Tarihleri:

Aktif Bank / 2018 Mayıs - Halen

Bilge Kalem Eğitim ve Danışmanlık / Temmuz 2015 - Aralık 2019

Kadıköy Anadolu Lisesi / Eylül 2016 - Mart 2017

Sınav Koleji / Kasım 2014 - Haziran 2015