



MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
FİNANSAL TEKNOLOJİLER VE AI DERSİ FİNAL PROJESİ

SBFIn Proje Raporu

Hazırlayan Öğrenciler:

Samet Elmalı, İbrahim Emre Çelenli, Enhar Apuhan, Mirza Etkä Topal

Ders Sorumlusu

Zeynep AYDEMİR

Aralık, 2024

İstanbul Medeniyet Üniversitesi, İstanbul

İÇİNDEKİLER

Sayfa

No:

1-) Özet.....	3
2-) Literatür Taraması.....	3
LSTM (long short-term memory) ve Finanstaki Kullanımı.....	3
Optimizasyon Algoritmaları ve Kayıp Fonksiyonları.....	3
3-) Proje İçeriği.....	4
3.1 Kullanıcıya Yönelik tasarım.....	4
3.2 Algoritmik Tasarım.....	5
a-) Problem Tanımlama.....	5
b-)Veri Toplama.....	5
c-)Veri Ön İşleme.....	6
d-) Veri Keşfi ve Görselleştirme.....	7
e-) Veri Modelleme.....	8
f-) Model Değerlendirme.....	9
4-) Sonuç ve Öneriler.....	10
5-) Kaynakça.....	11

1-) Özet

Projede Borsa İstanbul (BİST100) hisse değerlerinin geçmişteki değişimlerini analiz ederek geleceğe yönelik fiyat hareketlerini tahmin etmeyi amaçlayan basit düzey bir yapay zeka modelinin geliştirilmesi hedeflenmiştir. Projede hisselerin geçmiş değerleri halka açık kaynaklardan alınıp gerekli ön işleme adımları ile temizlenip analiz edilmiştir. Daha sonra LSTM (uzun kısa vadeli bellek) algoritması gerekli optimizasyon işlemleri yapılarak veriler üzerinde zaman serisi analizi ve gelecek fiyat değerleri hakkında tahminler yapılmıştır. Yapılan bu tahminler bir grafik üzerinde gösterilerek modelin tahminleri ve gerçek tahmin değerlerinin gözlemlenmesi sağlanmıştır. Modelin doğruluk oranı yüzdesel olarak MAPE(mean absolute prediction error) kullanılarak hesaplanmıştır. LSTM algoritması kullanılarak bir sonraki günün kapanış değerleri hakkında tutarlı bazı sonuçlar elde edilmiştir.

2-) Literatür Taraması

LSTM (long short-term memory) ve Finansdaki Kullanımı

LSTM bilgisayarlar ve diğer geleneksel araçların aksine bir işlem sırasında gerçekleşen değişim verilerini ve değişimler arasındaki ilişkileri öğrenebilen ve bu öğrendiklerini kullanarak analiz, tahmin gibi işlemleri gerçekleştiren bir yapay sinir ağıdır. Örneğin kitap okuma esnasında bir kitabı anlayabilmek için önceki sayfalardaki karakterler ve yaşanan olayları aklımızda tutarız. Bu bilgileri bir sonraki sayfaya geçtiğimizde o sayfayı anlamak için kullanırız. LSTM algoritması en basit şekilde bu örnekteki gibi veriler arasında ilişki kurup, hafıza hücresi sayesinde bu ilişkileri tutarak öğrenme işlemini gerçekleştirir. LSTM önemli verileri hatırlayabildiği gibi önemsiz verileri unutma, yeni öğrendiği şeylerle hafızasını güncelleme gibi özellikleri de vardır. Bu özellikleri sayesinde finansal piyasadaki fiyat hareketleri gibi doğrusal olmayan ve karmaşık yapılara sahip zaman serisi verileri üzerinde iyi sonuçlar ortaya koymuştur. Tasarımı nedeniyle zaman serisi analizlerinde iyi sonuçlar ortaya koyan LSTM finansal piyasa analizinde yaygın olarak kullanılan algoritmalarından biridir.

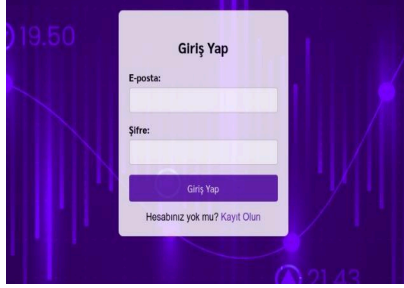
Optimizasyon Algoritmaları ve Kayıp Fonksiyonları

Derin öğrenme ve yapay sinir ağları modellerinin temel iki taşı vardır. Bunlar Kayıp Fonksiyonları (loss function) ve Optimizasyon Algoritmalarıdır. Kayıp fonksiyonları bir modelin tahmin ettiği değerler ve gerçek değerler arasındaki farkı bularak modelin başarısını hesaplar. Örneğin bir hissenin kapanış fiyatı tahmin ediliyorsa kayıp fonksiyonu bu tahminin hata oranını hesaplar. Optimizasyon Algoritmaları ise kayıp fonksiyonlarını tarafından hesaplanan hata oranını minimize etmek için ağırlıkları değiştirerek modelin doğruluğunu arttırmayı hedefler.

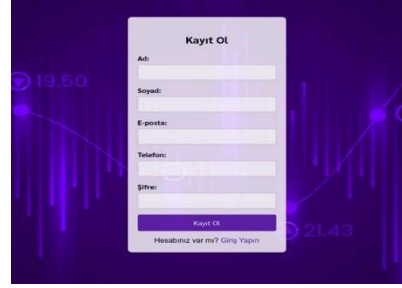
3-) Proje İerięi

3.1 Kullanıcıya Yönelik tasarım

Projede kullanımı ve anlaşılabilirlięi kolaylařtırmak için bir web uygulaması oluşturulmuřtur.

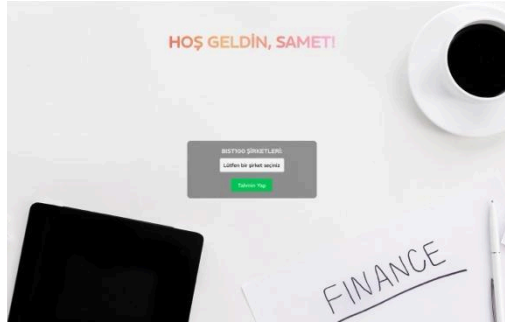


Görsel 1.0

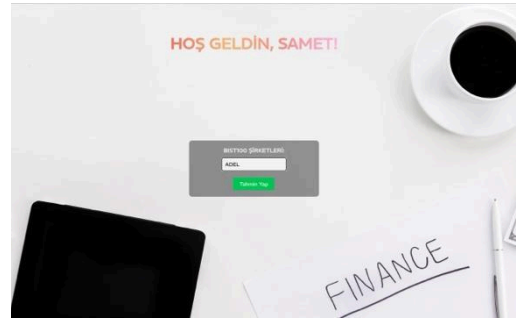


Görsel 1.1

Bu web uygulamasında karşımıza ilk olarak giriş sayfası çıkmaktadır. Görsel 1.0 ve 1.1 de giriş yap ve kayıt ol sayfaları gösterilmiştir.



Görsel 2.0



Görsel 2.1

Web uygulmasına giriş yapıldıktan sonra karşımıza Görsel 2.0 da gözükeceęi üzere BİST100 hisselerinde herhangi birini seçebileceğimiz bir web sayfası gelecektir. Burada Görsel 2.1 de gözükeceęi üzere bir seçim yapıp tahmin yap butonuna basıldıktan sonra tahminlerin görüntüleneceęi dięer web sayfası açılacaktır.



Görsel 3.0



Görsel 3.1

Bir hisse seçilip tahmin yap butonuna basıldıktan sonra seçilen hisse hakkında modelin yaptığı tahmin ve bu tahminlerle gerçek değerler arasındaki ilişki gözlemlenmek için bir grafiğin olduğu web sayfası açılacaktır.

3.2 Algoritmik Tasarım

Projenin algoritmik içeriği standart veri bilimi iş akışına göre açıklanacaktır.

a-) Problem Tanımlama

Amaç: Finansal hisse verilerinde zaman serisi analizi yaparak hisselerin fiyatlarının gelecekte nasıl değişeceğini tahmin etmeye çalışmaktır.

b-)Veri Toplama

Veri kaynağı olarak www.kap.org.tr web sitesi ve yfinance(yahoo finance) pyhton kütüphanesi kullanılmıştır. İlk olarak BİST 100 hisselerinin sembolleri çekilmiş ve bir csv dosyasına kaydedilmiştir. Çekilen sembollere ait hisse verileri yfinance kütüphanesi kullanılıp her hisselerin verileri ayrı dosyalarda olacak şekilde 5 yıllık bir periyot için indirilmiştir.

Veri toplama adımında kullanılan fonksiyonlar

getTickers() Bist 100 hisselerinin sembollerini indirir bir diziye koyar.

saveTickers() Tickers dizisinde tutulan semboller dataframe haline getirir. Daha sonra bu semboller bir csv dosyasına yazılır.

getData() önceden kaydedilmiş semboller yardımıyla hisselerle ait 5 yıllık verileri indirip csv olarak depolar.

c-)Veri Ön İşleme

Önişleme adımında ilk olarak hisselerle ait her bir csv dosyasındaki price başlıkları date olarak güncellenir. Bu işlem tamamlandıktan sonra model eğitimi sırasında kullanılmayacak bilgilerin yer aldığı 2. Ve 3. satırlar her dosyadan kaldırılır. Volume (işlem hacmi) ve close (kapanış fiyatı) arasında finans piyasasında önemli bir ilişki vardır. Bu ilişki trend doğrulaması, volalite seviyesi gibi durumların tespit edilmesinde kullanılır. Bu nedenden dolayı veri dosyalarından volume close ve date değerleri dışındaki diğer değerler kaldırılmıştır. Modeli eğitmek için yeterli veriye sahip olmayan bir kaç tane hisse verisinin csv dosyaları da kaldırılmıştır. Veri önişleme başlığı altında veri normalizasyonu adımıda uygulanmıştır. Daha sonradan değerler kullanılacağı için dönüşüm kolaylığı olması açısından min-max scaling model eğitime başlamadan önce test ve eğitim verileri üzerinde uygulanmıştır.

Veri temizleme adımında kullanılan fonksiyonlar

formatData() Parametre olarak daha önceden indirilmiş csv dosyalarını alır her bir dosyada ki price başlığını date olarak değiştirir ve kullanılmayacak olan 2. ve 3. satırdaki verileri siler.

removingUselessdata(csv_files) Close(kapanış değeri), date(tarih), volume(işlem hacmi) değerleri dışındaki veriler ve eğitim için yeterli verisi olmayan hisseleri siler.

scaleData() Veriler üzerinde min-max normalizasyonu ile[0,1] rasında ölçekleme işlemi yapar.

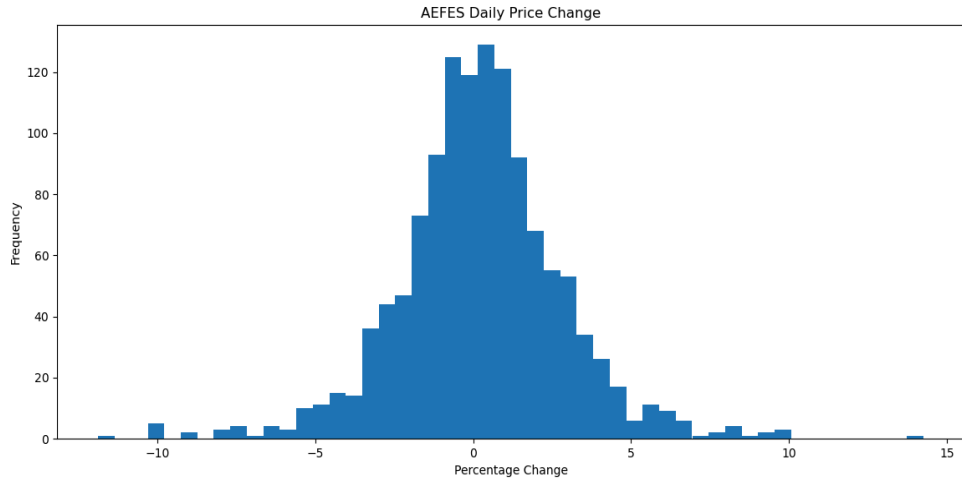
d-) Veri Keşfi ve Görselleştirme

Bu adımda hisse dataFrameleri matplotlib ve seaborn python kütüphaneleri kullanılarak çeşitli görseller oluşturulmuştur.

Train ve Test setleri dağılımı

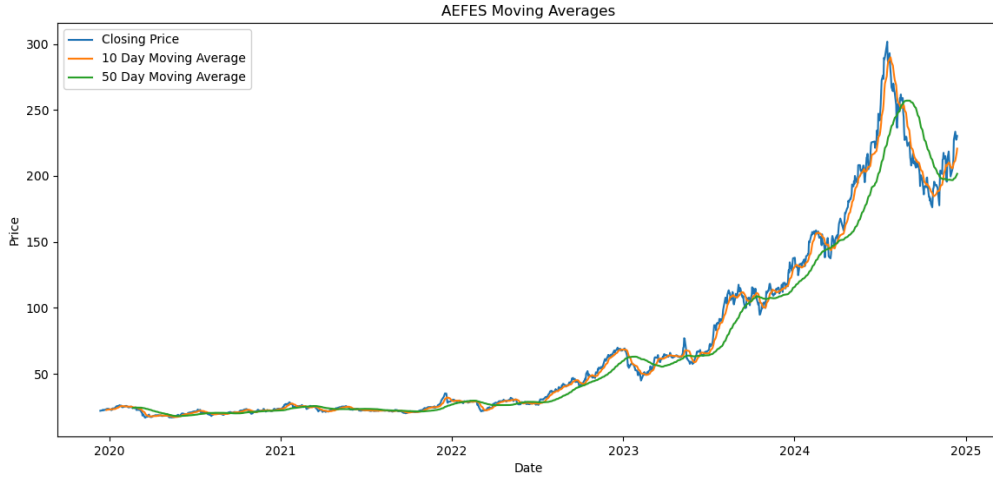


Görsel 4.0



Günlük fiyat değişim yüzdesi dağılımı

Görsel 4.1



50 ve 10 günlük hareketli ortalama dağılımı

Görsel 4.2

Veri modelleme aşamasında model hakkında değişiklikler yapılırken 4.x görsellerinde bulunan değerler göze alınarak yapılmıştır.

e-) Veri Modelleme

Zaman serisi analizlerinde iyi sonuçlar veren lstm derin öğrenme algoritması tahmin yapmak için model olarak seçilmiştir. İlk olarak ön işleme adımları uygulanıp csv formatında saklanan veriler gerekli dizinden bulunur ve bu veriler pandas kütüphanesi yardımıyla dataframe haline getirilir. Bu işlemler yapıldıktan sonra veriler eğitim(%70) ve test(%30) verisi olarak bölünür. Min-max scaling işlemi bu adımdan sonra gerçekleşir. Eğitim ve test verileri hazırlandıktan sonra modeli eğitmek için x(girdi) modelin tahmin yaparken kullanacağı girdi verileri, y(çıktı) modelin tahmin etmeye çalışacağı gelecek değeri olmak üzere bu veriler gerekli işlem adımlarına tabi tutulur. Veriler hazır hale getirildikten sonra artık modelin derlenmesi aşamasına geçilir. Lstm modeli dört katmanlı olacak şekilde oluşturulur. Bu katmanlardan üç tanesi giriş ve bir tanesi çıkış katmanıdır. Giriş katmanlarının her birinden sonra dropout(bırakma) katmanı uygulanarak her katmanın rastgele %20 lik nöronları etkisiz hale getirilerek aşırı öğrenme engellenmiştir. Model hazırlandıktan sonra Adam optimizasyon fonksiyonu ve MSE kayıp fonksiyonu ile derlenip tahminler yapılmıştır.

Veri Modelleme adımımda kullanılan fonksiyonlar:

csvNameList() Daha önce indirilmiş olan her csv dosyasındaki .csv uzantılarını kaldırıp dosya isimlerini bir diziye kaydeder. Daha sonra bu diziyi döndürür.

getDataFrame() Her bir csv dosyasını date sütununu index olarak seçerek pandas kütüphanesi yardımıyla dataFrame' e dönüştürür ve bu dataFramelerin listesini döndürür.

splitTrainAndTest() Dataframe halindeki verileri %70 eğitim %30 test verisi olacak şekilde böler.

splitDataXy() Eğitim ve test verilerin x(girdi) ve y(çıkı) değerleri şeklinde böler.

initModel() LSTM modelini üç girdi ve bir çıkı katmanı olarak bölüp Adam ve MSE algoritması ile derler.

f-) Model Değerlendirme

Eğitim tamamlandıktan sonra, modelin test seti üzerinde tahmin yapması sağlanır. Tahmin aşaması tamamlandıktan sonra modelin trend(artış/azalış) yönünü tahmin etme oranı ölçülür. Trend yönü başarı oranı hissenin bir önceki gün değeri(a) , bir sonraki günün değeri(b) ve bir sonraki gün için modelin yaptığı tahmin değeri(c) kullanılarak hesaplanmıştır. Eğer tahmin edilen değışim(a-c) ve gerçek değışim(a-b) aynı yönde ise başarı, değışim yönleri birbirini tersi ise başarısızlık olarak değerlendirilir. Bu yöntem kullanılarak trend yönünü tahmin etme oranı %45- %53 arasında bulunmuştur. Eğer modelin y(çıkış) değerine artış/azalış değeri de bulunsaydı (0 ve 1 şeklinde) ve loss algoritmamız da bunu göze alan bir özel algoritma olsaydı bu metrikte daha iyi bir sonuca ulaşılabilirdi.

$$MAPE = \frac{\sum \frac{|A - F|}{A} \times 100}{N}$$

Görsel 5.0

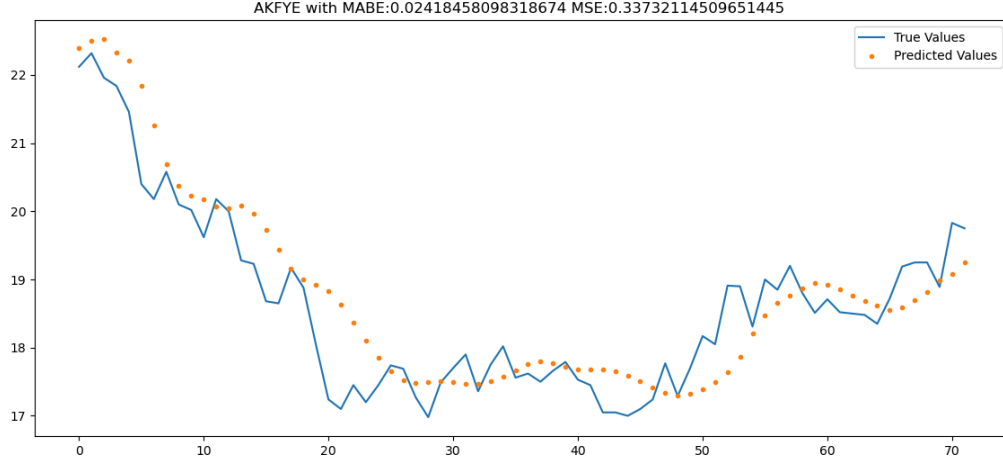
A: Gerçek değeri

F: Tahmin edilen değeri

N: Tahmin yapılan veri sayısı

Model trendleri bilmede çok başarılı olmasada modelin hata oranı MAPE(mean absolute percentage error) kayıp fonksiyonu ile %2.2 - %3 arası hata oranı gibi iyi bir

oranla hesaplanmıştır. MSE(mean squared error) kayıp fonksiyonu MAPE kayıp fonksiyonun oranla daha büyük bir hata oranı çıkarmıştır. Bunun nedeni MSE fonksiyonun MAPE den farklı olarak gerçek değer ile tahmin edilen değerin arasındaki farkların karelerini alarak hesaplama yapmasıdır.



Görsel 6.0

Görsel 6.0'da modelin hissenin fiyatlarını MAPE ile %2.24 hata oranı ile hesapladığı görülmektedir.

4-) Sonuç ve Öneriler

Bu projede derin öğrenme modellerinde biri olan LSTM kullanılarak BIST100 hisselerinin kapanış fiyatlarını tahmin edilmesine yönelik bir model geliştirilmiştir. Hissenin eğer geçmiş 5 yıllık verisi varsa bu veriler, daha az verisi varsa tamamı kullanılarak bir model eğitilmiştir. Eğitilen model belirli bir güne ait kapanış fiyatlarını MAPE (mean absolute percentenge error) fonksiyonu ile ortalama %2.25 hata oranı ile tahmin etmiştir. Modelin eğitimi için kullanılan dropout(bırakma) oranı %20'den daha düşük değerler olarak kullanıldığında modelin eğitim verisini ezberlemeye ve değişik kalıptaki hareketler gördüğünde uyum sağlayamadığı, %50 gibi yüksek değerler kullanıldığında ise modelin dengesizleştiği ve alakasız/beklenmedik değerler oluşturma olasılığının daha da arttığı tespit edilmiştir.

Model bazı hisselerde çok başarılı, bazı hisselerde ise bir o kadar başarısız olmuştur. Bunun nedeni modeli eğitirken close volume gibi niceliklerin kullanılmasıdır. Finans piyasasında manipülasyon, piyasa koşulları, ekonomik faktörler, sektör performansı vb. gibi durumlar fiyatlar üzerinde büyük etkilere sahiptir. Bu nedenden dolayı model geliştirmek için işlem hacmi, kapanış fiyatı gibi nicelikler kadar nitelikler de önemlidir. Daha fazla başarı oranına sahip model geliştirilmek istenirse bu gibi durumlar da analiz edilip model geliştirilmelidir.

5-) Kaynakça

Fischer, T., & Krauss, C. (2018). Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions. Proceedings of the International Conference on Machine Learning (ICML).

Krauss, C., Do, X. A., & Huck, N. (2017). Deep Learning for Finance: Deep Portfolios. Journal of Financial Data Science, 1(1), 1-25.

Gavcar, E., & Metin, H. M. (2024). Hisse Senedi Değerlerinin Makine Öğrenimi (Derin Öğrenme) ile Tahmini. Muğla Sıtkı Koçman Üniversitesi, İktisadi ve İdari Bilimler Fakültesi.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization.