



FINAL DATATHON SUNUMU

"I have seen the future and it is very much like the present, only longer."

-Kehlog Albran, *The Profit*

Wizard of Stocks : Alperen Demirci

İÇERİK

- Problem Tanımı
- Araştırmalar
- Çözüm Önerisi
- Yorumlama
- Sonuç

PROBLEM TANIMI

Amaç

- o Borsa hareketlerini tahmin etmek için bir **öngörü modeli** geliştirmek.

Arkaplan

- o Borsa, **hareketliliğiyle** ve **karmaşıklığıyla** bilinir, bu da yatırımcıların karar vermelerini zorlaştırmaktadır.
- o Fiyat hareketlerinin **isabetli** tahmini, yatırımcıların alım satım stratejilerine katkıda bulunur ve risk yönetimini kolaylaştırır.

Zorluk

- o Borsa verileri birçok faktöre duyarlıdır, bu yüzden karmaşık ve çözülmeli **zor desen/düzenlere** sahiptir.
 - o Güvenli bir öngörü modeli geliştirmek için bu karmaşıklıkları aşmayı ve temel trendleri doğru yakalamayı gerektirir.
- o **Çözüm** için seçilen özellikler : Açıılış - Yüksek - Düşük - Kapanış (OHLC) fiyatları

Çözüm

- o Hisse senedi için tahmin edilecek özelliğin doğru şekilde seçilmesi.
- o Çeşitli **Istatistik ve Yapay Zeka** teknikleri kullanarak hisse senedinin seçilen özelliği/ özellikleri hakkında tahmin yapmak.

ARAŞTIRMA

- Araştırma kısmı **3** parçadan oluşmaktadır.

1. Akademik Tarama

- Bu kısımda farklı yayın kanallarından faydalılmıştır. Örnek olarak dünya çapındaki bilim insanların katkı yaptıkları **IEEE Research** makaleleri, **Stanford Uni.'nin** CS231N ders özetleri verilebilir. Kullanılan tüm makaleler Kaynakça kısmında bahsedilmiştir.

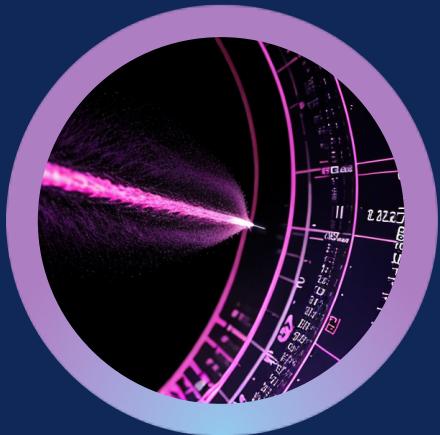
2. Geçmiş Çalışmalar

- Burada daha önceden, bizim amacımızla örtüşen projeler incelenmiştir. Bunlara insanların çalışmalarını paylaşabildiği **Medium** ve **GitHub** gibi web siteleri örnek verilebilir.

3. Teknik Araştırmalar

- Bu kısım daha çok kodlama sürecinde rol almıştır. Kod kütüphanesi kullanımı, model kullanımı gibi çözümün kodlama tarafına yöneliktedir. Kullanılan web siteleri ise **Tensorflow**, **NumPy**, **Pandas** gibi kullanılan teknolojilerin dokümantasyonlarıdır.

ÇÖZÜM ÖNERİSİ



Veriyi Anlamak



Veri İşleme



Model Oluşturma



Yorumlama

Note: Yukarıdaki fotoğraflar **RealDream12 Stable Diffusion** Model(Text to Image) kullanılarak oluşturulmuştur.

VERİYİ ANLAMAK

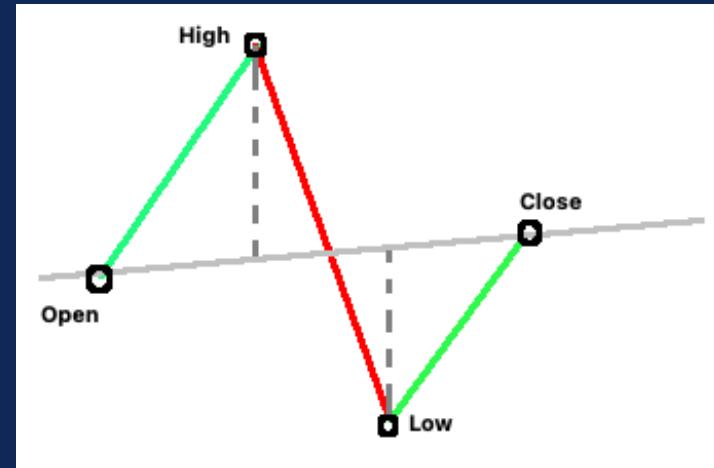
- Elimizdeki 4 farklı grafik bulunmaktadır. Bunlar sırasıyla **AIL-FRX**, **BK-FRX** , **CRP-AIL** ve **ZZC-AIL**'dır.
- Veriler CSV dosyası olarak verilmiştir.
- Verinin özellikleri sırasıyla 'Tarih,Açılış,Yüksek,Düşük,Kapanış'dır.
- Her bir kayıt(entry) 1 saat aralıklarla, 12/12/2012'den 30/11/2014'e kadar tutulmuştur.
- Bazı kayıtlarda **eksik veriler** bulunmaktadır.
- Aşağıdaki tablo hangi verilerin hangi özelliklerinde eksik veri olduğunu belirtmektedir.

	AIL-FRX	BK-FRX	CRP-AIL	ZZC-AIL
Açılış	Tam	● Eksik	● Eksik	Tam
Yüksek	Tam	● Eksik	Tam	● Eksik
Düşük	Tam	● Eksik	Tam	Tam
Kapanış	● Eksik	Tam	● Eksik	● Eksik

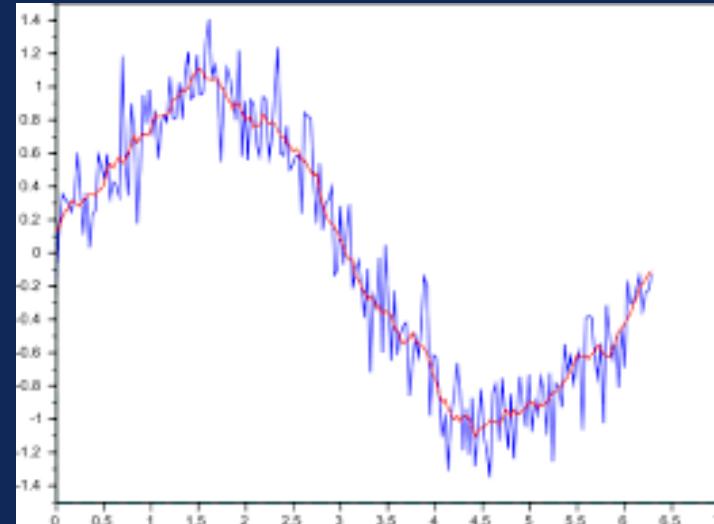
VERİYİ İŞLEMEK

- İlk olarak 'Veriyi Anlama' aşamasında eksik değerlerimizin olduğunu görmüştük.
- Bu boşlukları doldurmak için izlenecek yöntem şudur:
 1. Eksik olan değerleri tespit et.
 2. Boşlukları, yandaki şekli koruyacak şekilde doldur.
 - Örnek olarak eğer Yüksek ve Düşük değerleri boşsa, bu değerleri Açıılış ve Kapanış değerlerinin mutlak farkını kullanarak doldur.
 - Eğer ki sadece Kapanış değeri boşsa, bu değeri şu formüle göre doldur : $Kapanış = Açıılış + Ortalama(Yüksek, Düşük)$
- Yukarıdaki aşamalardan sonra, doldurulan suni verilere Hareketli Ortalama (Moving Average) yumusatması yapılmaktadır.
- Bu yöntem doldurduğumuz suni verileri, geçmişteki değerlerle de bağıdaştırarak daha kararlı bir veri oluşmasını sağlar.
- Bu yöntemde pencere sayısı 8 olarak belirlenmiştir.
- Yumusatma formülü : $Kapanış = Ortalama(Kapanış, HareketliOrtalama(8))$

- Doldurma Şekli



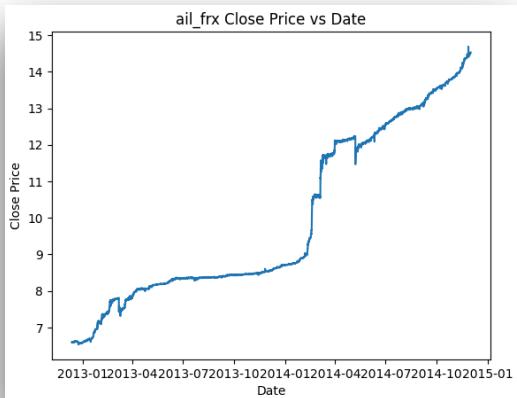
- Hareketli Ortalama Yumuşatması



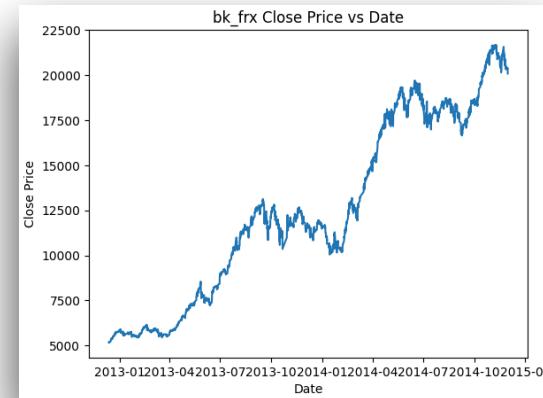
VERİ GÖRSELLEŞTİRME

- Veri Anlama ve Veri İşleme adımlarından sonra Veri Görselleştirme kısmında, farklı hisselerin grafikleri görselleştirildi.

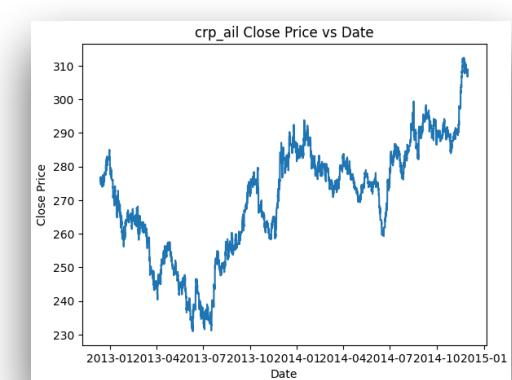
AIL-FRX Grafiği



BK-FRX Grafiği



CRP-AIL Grafiği

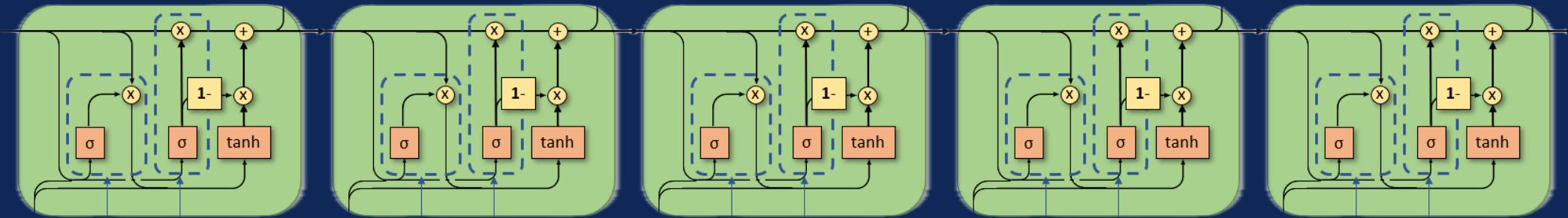


ZZC-AIL Grafiği



VERİ BÖLME

- Öncelikle modelimizi eğitip değerlendirebilmemiz için verimizi test ve eğitim olarak belirli bir yüzdeliğe bölmemiz gerekiyor.
- Ayrıca RNN mimarilerini kullanmak için hem test hem eğitim verimizi sekans oluşturacak şekilde bölmemiz gerekmektedir.
- Bu adımlardan sonra verimiz model eğitimi ve tahmini için hazır hale gelecektir.

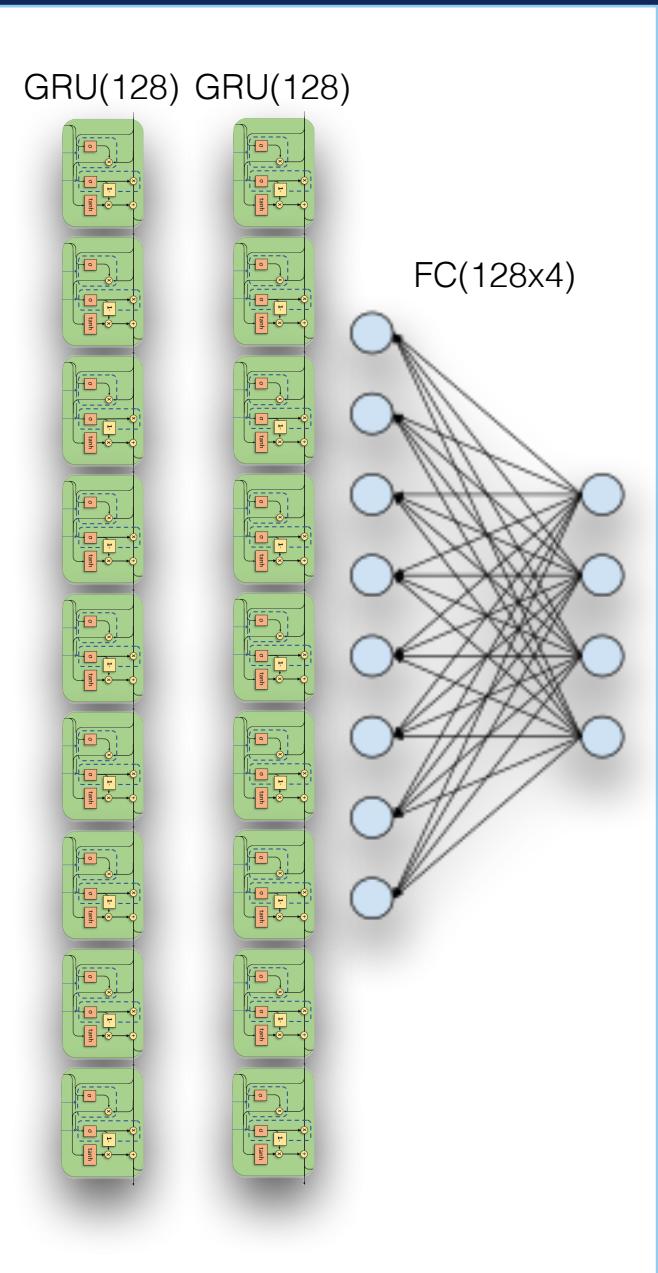


- Verimizi 98:2 oranı ile eğitim ve validasyon verisi olarak böldük. **98:2** oranı, çeşitli oranlar denendikten sonra en efektif oran olarak bulunmuştur. (80:20, 70:30, 75:25, 90:10 vb.)
- Verimizle oluşturduğumuz sekanslar 45 günlük sekanslardır. Yani modelimiz son **45** güne bakıp sonraki günü tahmin etmektedir. 45 gün olmasının sebebi farklı sekans uzunlıklarının denenip, en efektifinin bu bulunmasıdır. (14, 30, 60, 90, 200)

MODEL OLUŞTURMA

- Veri ile ilgili neredeyse tüm aşamalar bitirildi, sıra Model Oluşturma kısmında.
- Öncelikle veriyi bir Zaman Serisi (Time Series) verisi olarak işlendi. Bunun nedeni ise veri belirli bir zaman aralığıyla bölünmüştür.
- Bu zaman serisini tahminlemek içinse RNN yani Özyinelemeli Derin Ağlar yöntemini kullanacağız.
- Sebebi ise bu modellerin belirli bir sekans'ı alıp, bu sekansın sonraki değerini tahminlemesiyle çalışır. Bu veri için örnek verecek olursak son 45 güne bakıp 46. gündeki değerlerin tahmin edilmesidir.

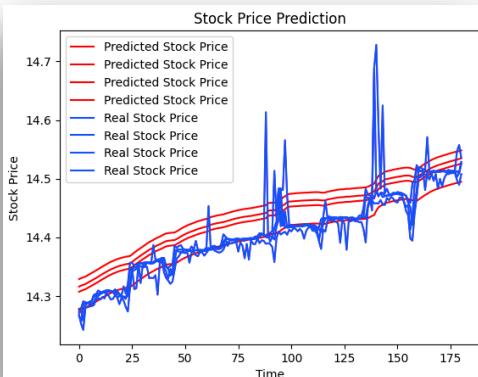
- RNN mimarilerinden **LSTM**, **GRU**, **BiDirectional LSTM** ve **Convolutional LSTM** denendi.
- Bu mimarilerden en iyi sonucu **Stacked GRU** adını verdığımız model verdi. Ayrıca GRU'ların daha az parametre gerektirmesi nedeniyle LSTM ile çalışan modellerden daha hızlı eğitiliyor ve çalışıyor.
- Kullanılan mimari: GRU(128)->Dropout(0.18)->GRU(128)->Dropout(0.18)->FullyConnected(4). Yandaki şekilde mimari görselleştirilmiştir.



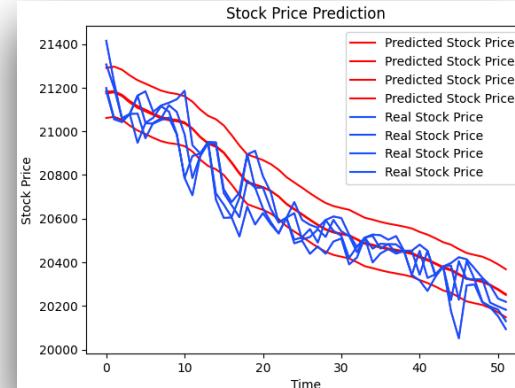
YORUMLAMA

- Model mimarisini önceki sayfada incelemiştik, modelimizi belirlediğimiz parametrelerle çalıştığımızda test verisindeki performansı aşağıda görülebilir.
- Burada gerçek verideki son 45 güne bakılıp tahmin yapılmaktadır. Tahmin yapılırken **sadece gerçek veriler** üzerinden yapılmaktadır.

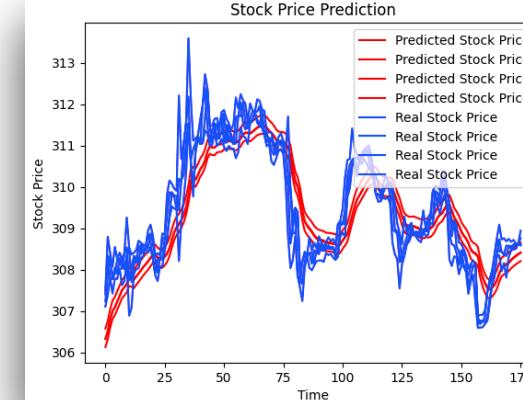
AIL-FRX Grafiği



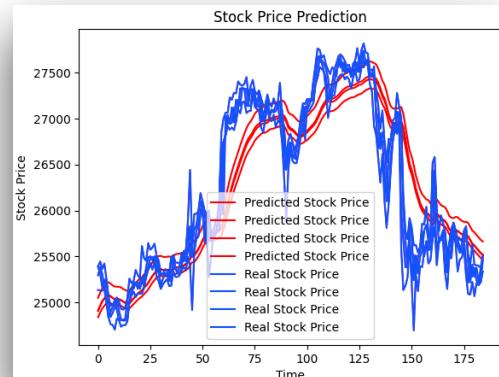
BK-FRX Grafiği



CRP-AIL Grafiği



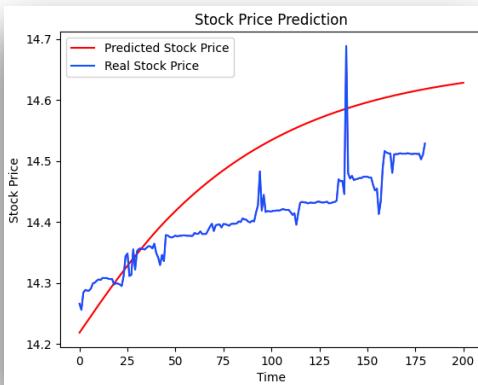
ZZC-AIL Grafiği



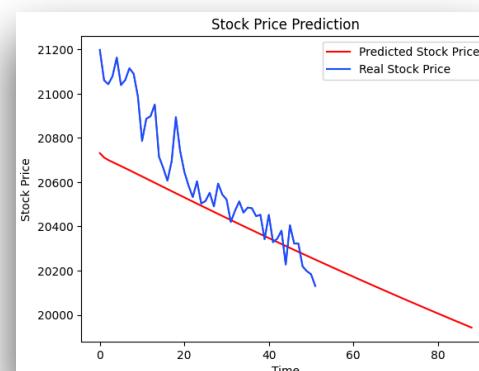
YORUMLAMA

- Modelimizi belirlediğimiz parametrelerle çalıştığımızda validasyon setindeki performansı aşağıdaki gibidir.
- Bu tahminde Kayan Pencere (Rolling Windows) yaklaşımı kullanılmıştır. Bu yaklaşım önceki 45 güne bakıp sonraki günü tahminledikten sonra, pencereyi bir gün kaydırıp bu şekilde gelecekteki verileri tahmin edilebilir.
- Bu yaklaşım en son yapacağımız tahminde kullanılacak olup, **sadece gerçek veriler değil tahmin edilen verileri de kullanarak** tahmin yapmaktadır.

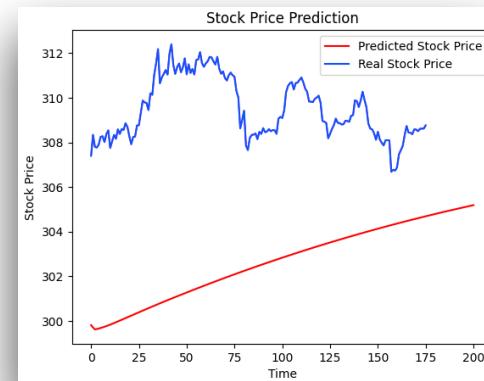
AIL-FRX Grafiği



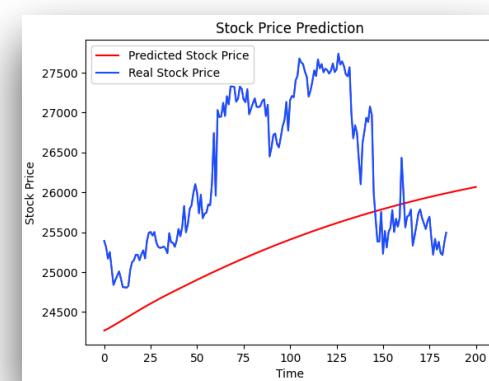
BK-FRX Grafiği



CRP-AIL Grafiği



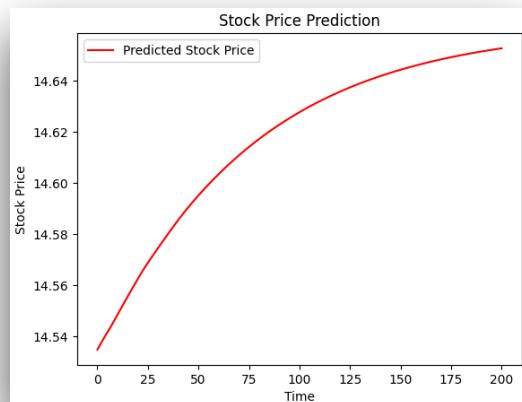
ZZC-AIL Grafiği



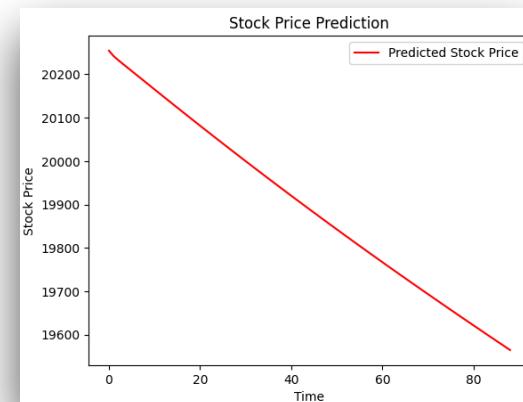
FİNAL TAHMİNİ

- Modelimizi final tahmini için çalıştığımızda elde ettiğimiz grafikler aşağıdaki gibidir.
- Bu tahminler Kaggle'da bulunan değerlendirme verisiyle karşılaştırılıp, MAE kayıp fonksiyonu ile [253.61615](#) Private Score elde edilmiştir.

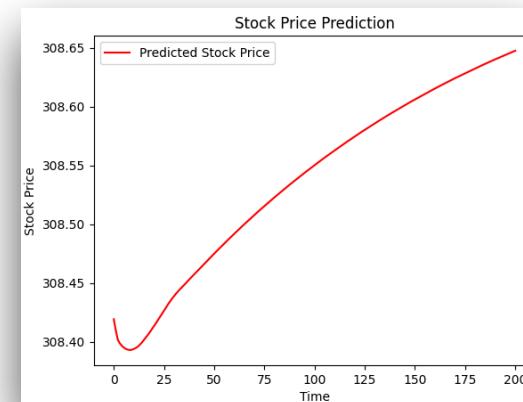
AIL-FRX Grafiği



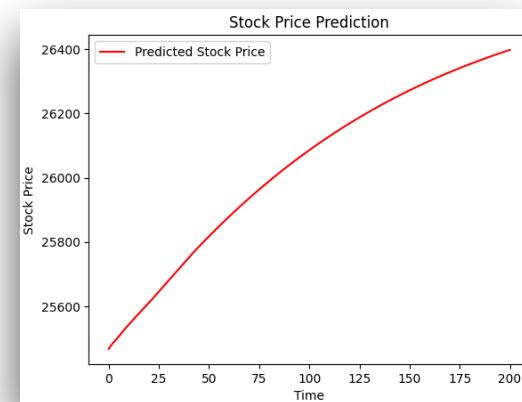
BK-FRX Grafiği



CRP-AIL Grafiği



ZZC-AIL Grafiği



SONUÇ

- Veriyi bahsettiğimiz tekniklerle ve bazı indikatörleri kullanarak ön işledik.
- Tahmin için RNN mimarisi olan Stacked GRU kullandık.
- Tahminlerimizi gerçek verilerle karşılaştırarak modelimizin hiper parametrelerini düzenledik.
- En sonunda da sırasıyla 90-200-200-200 günlük tahminlerle yarışmada yerimizi aldık. MAE kayıp fonksiyonu ile [253.61615](#) Private Score elde ettik.

KAYNAKÇA

Medium and Other Websites

- <https://medium.com/towards-data-science/reading-charts-with-convolutional-neural-networks-cbaabdd5f478>
- <https://medium.com/@sakshisanghi0001/stock-price-prediction-with-lstm-3784c246530b>
- <https://medium.com/@mroko001/rnn-vs-lstm-vs-transformers-unraveling-the-secrets-of-sequential-data-processing-c4541c4b09f>
- <https://towardsdatascience.com/lstm-for-google-stock-price-prediction-e35f5cc84165>

IEEE Papers

- <https://ieeexplore.ieee.org/document/10366646>
- <https://ieeexplore.ieee.org/document/10140896>
- <https://ieeexplore.ieee.org/document/9015840>
- <https://ieeexplore.ieee.org/document/10129490>
- <https://ieeexplore.ieee.org/document/10486391>
- <https://ieeexplore.ieee.org/document/10140728>

Documentations

- <https://www.tensorflow.org/guide/migrate>
- <https://numpy.org/doc/stable/>
- https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html

TEŞEKKÜRLER

Hazırlayan - Sunan: Alperen Demirci

[Mail](#) [LinkedIn](#) [GitHub](#)