

DATA MADRID

 ENES BERKE KARAOĞLAN

 FURKAN DEMİR

 EMİRHAN TOZLU



Proje Tanımı

Bu proje, finansal varlıkların, özellikle hisse senetleri ve kripto paraları gibi gelecekteki fiyat hareketlerini tahmin etmeyi amaçlamaktadır. Geçmiş verilerin analizi ve çeşitli makine öğrenimi tekniklerinin kullanımıyla, yatırımcılara gelecekteki piyasa trendleri hakkında bilgi sağlayarak daha bilinçli yatırım kararları almalarına yardımcı olmayı hedeflemektedir.

Proje Ve Araştırma Basamakları

01

İlk aşamada projenin adımları ve içeriği, bununla birlikte atılması gereken adımlar belirlenmiş olup bir yol haritası oluşturulmuştur. (Neler yapılabilir? Ne şekilde ve hangi uygulamalar ile yapılabilir gibi sorular bu aşamada cevaplanmıştır.)

02

İkinci aşamada genel literatür taraması yapılmıştır. İlk etapta bir finansal varlığın gelecekteki fiyatını tahmin etmekten bağımsız bir şekilde genel olarak makine öğrenimi ve derin öğrenme projeleri araştırılmıştır. Konuya daha vakıf olunduktan sonra alan daraltılmış ve hisse, para birimi gibi finansal varlıklar ile ilgili projeler araştırılmıştır.

Bu adımda incelenen kaynaklar:

- >GitHub Projeleri
 - > Youtube Videoları
 - > Çeşitli Makaleler Ve PDF Dokümanları
- (Sunum sonunda kaynakça kısmında bağlantılar paylaşılmıştır.)

03

İkinci adıma ek olarak yapılan örnek projelerde kullanılan makine öğrenimi ve derin öğrenme yöntemleri analiz edilmiş ve bir araştırma listesi oluşturulmuştur.

04

Gerekli altyapı oluşturulduktan sonra sıra verilerin incelenmesi ve işlenmeye uygun hale getirilmesi aşamasına gelmiştir.

05

Hazırlanan ve işlenmeye uygun hale getirilen veriler ile analizi yapılan modeller gerekli programlama dili (Python) bir araya getirilmiştir. Bu aşamada yer alan en önemli kısım modellerin MAE - RMSE değerlerinin ölçülmesidir. Bu değerler tahminlerimizin gerçeğe ne kadar yakın olduğunun bir tespitidir.

06

Son adım ise en uygun modelin belirlenmesi ve bu model aracılığı ile tahminlerin yapılmasıdır.

Veri Setlerinin Hazırlanması

String to float

Veri setlerinde yer alan string ifadelerin modele uygun hale getirilmesi için bu ifadeler float türüne dönüştürüldü.

NaN Değerleri

Veri setlerinde yer alan NaN değerleri modelin daha istikrarlı bir performans verebilmesi için sütunların ortalama değerleri ile dolduruldu. Bu yaklaşım, eksik veri noktalarının etkili bir şekilde ele alınmasını sağlayarak modelin eğitim sürecini stabilize eder ve sonuç olarak daha tutarlı tahminler elde edilmesini destekler.

```
veri['open'] = veri['open'].str.replace("", "").astype(float)
veri['high'] = veri['high'].str.replace("", "").astype(float)
veri['low'] = veri['low'].str.replace("", "").astype(float)
veri['close'] = veri['close'].str.replace("", "").astype(float)

veri['open'].fillna(veri['open'].mean(), inplace=True)
veri['high'].fillna(veri['high'].mean(), inplace=True)
veri['low'].fillna(veri['low'].mean(), inplace=True)
veri['close'].fillna(veri['close'].mean(), inplace=True)
```

Uygun Modele Giden Yol

Hazırlık

Hazırlık adımında, veri setinin analizi ve ön işleme adımları gerçekleştirilir. Veri seti incelenir, eksik veriler doldurulur veya çıkarılır, gerekiyorsa özellik mühendisliği yapılır ve veriler ölçeklendirilir. Ayrıca, modelin kullanılacağı bağımsız değişkenler ve hedef değişken belirlenir.

Değerlendirme

Değerlendirme aşamasında, farklı modellerin eğitilmesi ve performanslarının değerlendirilmesi sağlanır. Modellerin eğitimi genellikle eğitim ve doğrulama kümeleri üzerinde gerçekleştirilir, ardından modelin test kümesi üzerinde performansı ölçülür. Bu aşamada, belirlenen performans ölçütlerine (RMSE, MAE vb.) göre modeller karşılaştırılır.

Karar

Karar aşamasında, farklı modellerin performansı karşılaştırılır ve en uygun olanı seçilir. Model seçimi, belirlenen kriterlere ve proje gereksinimlerine dayanarak yapılır. En uygun model seçildikten sonra, son adımda modelin kullanımı veya geliştirilmesi sürecine geçilir.

Bazı Modeller Ve Başarı Oranı



Random Forest Regressor:

Birçok karar ağacının bir araya gelerek oluşturduğu bir topluluk modeli. Çeşitli özelliklerin etkisini yakalamak için kullanılır ve aşırı öğrenmeye karşı dayanıklıdır.



SVR:

Karar ağaçları kullanarak veri setini bölerek bir tahmin modeli oluşturur. Karmaşık karar yapılarını öğrenme yeteneği vardır, ancak aşırı öğrenmeye eğilimlidir.



TensorFlow - Yapay Sinir Ağı:

Yapay sinir ağı tabanlı bir derin öğrenme modeli. Karmaşık ilişkileri öğrenebilir ve genellikle büyük veri setlerinde iyi performans gösterir.



Decision Tree Regressor:

Karar ağaçları kullanarak veri setini bölerek bir tahmin modeli oluşturur. Karmaşık karar yapılarını öğrenme yeteneği vardır, ancak aşırı öğrenmeye eğilimlidir.



Gradient Boosting:

Zayıf tahminicileri bir araya getirerek güçlü bir tahmin modeli oluşturur. Zaman içinde hata azaltma stratejisini uygular ve karmaşık ilişkileri yakalama yeteneğine sahiptir.

Çözüm Aşaması - 1

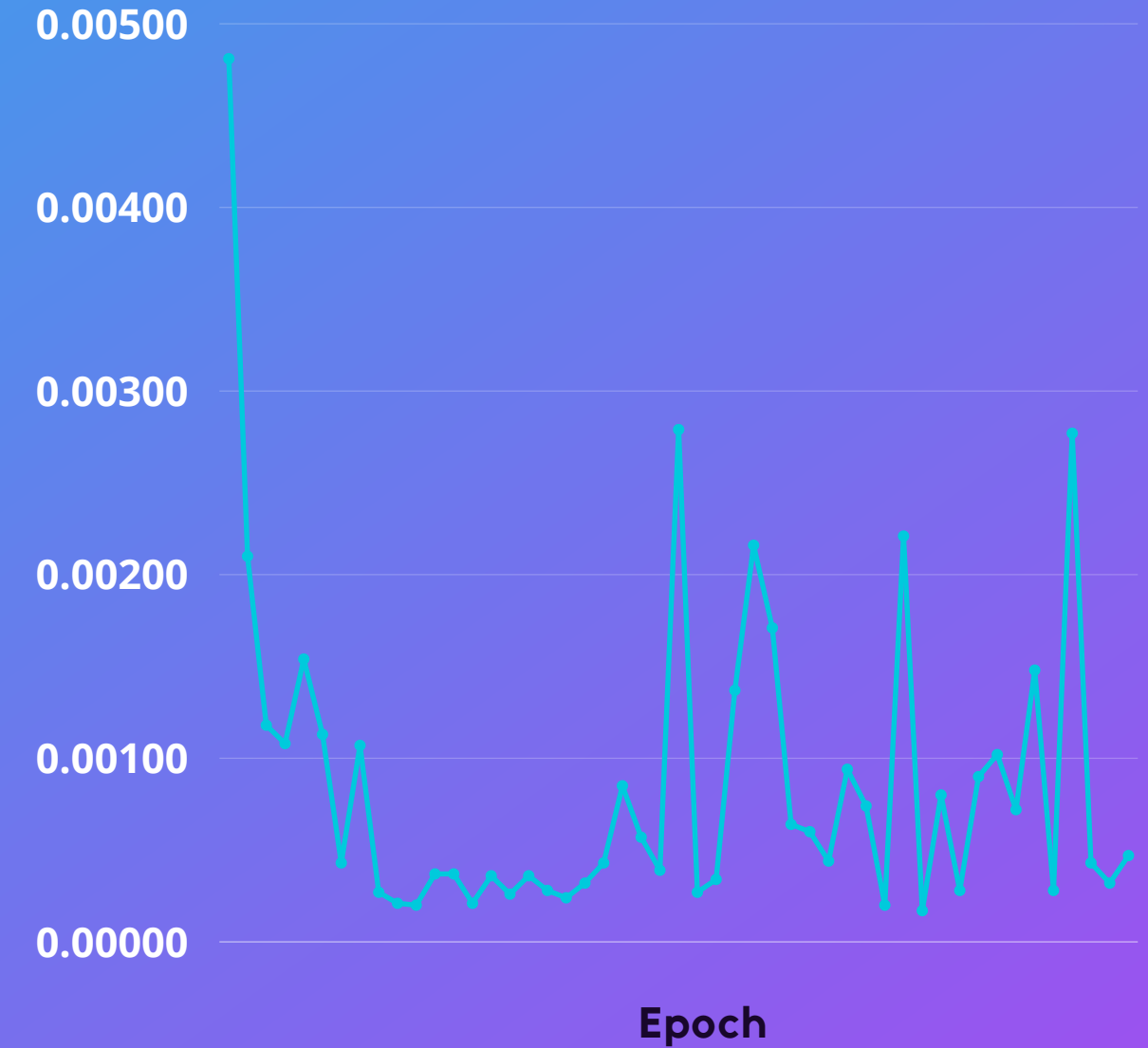
Yaptığımız analizler sonucunda, farklı regresyon ve tahmin modelleri arasında belirli bir performans karşılaştırması yaptık. En düşük kök ortalama kare hatası (RMSE) değerine sahip olan model, TensorFlow tabanlı Yapay Sinir Ağı modeli oldu. Bu sonuçlar, finansal varlık fiyatlarını tahmin etmek için derin öğrenme yöntemlerinin etkili olabileceğini göstermektedir.

Yapay Sinir Ağı - En Başarılı Sonucun Belirlenmesi

Epoch, bir yapay sinir ağı modelinin eğitim sürecindeki bir adımdır. Epoch sayısı, modelin ne kadar iyi öğrendiğini ve ne kadar iyi genelleştirme yaptığını belirler. Eğitim süreci boyunca model daha fazla epoch ile daha fazla veriye maruz kalır ve bu da genellikle daha iyi sonuçlar elde etmek için gereklidir. Ancak, çok fazla epoch kullanmak aşırı öğrenmeye (overfitting) yol açabilir

Oluşturulan RMSE - Epoch grafiğinden anlaşıyor ki en verimli sonucu verecek epoch sayısı 30 olarak ölçülmüştür.

RMSE



Çözüm Aşaması - 2

bk_frx veri setindeki veri miktarının diğer veri setlerine göre daha az olduğu dikkate alınarak, TensorFlow yapay sinir ağı yerine RandomForestRegressor modeli tercih edildi. Bu seçim, daha az veriyle daha iyi bir performans elde etmeyi hedefler. RandomForestRegressor modeli, karar ağaçlarından oluşan bir enseble yöntemi kullanarak veri setindeki karmaşık ilişkileri daha iyi yakalayabilir ve aynı zamanda aşırı öğrenmeye daha dirençlidir. Bu değişiklik, veri setinin boyutuyla daha uyumlu bir modelin seçilmesini sağlar, böylece daha güvenilir tahminler elde edilir.

SONUÇ

Bu projede, finansal varlıkların fiyatlarını tahmin etmek için çeşitli makine öğrenimi modellerini kullandık. Veri setlerini anlama, ön işleme, model seçimi ve performans değerlendirmesi gibi aşamalardan geçerek, her bir modelin gelecekteki fiyatları tahmin etme yeteneğini değerlendirdik. Yapay sinir ağı, RandomForestRegressor, Lineer Regresyon gibi farklı modelleri denedik ve her birinin avantajlarını ve dezavantajlarını değerlendirdik.

Sonuç olarak, veri setlerindeki örüntülerin modelleme sürecinde önemli olduğunu gözlemledik. Hangi modelin en iyi performansı gösterdiği, veri setinin özelliklerine, boyutuna ve kullanım senaryosuna bağlı olarak değişebilir. Dolayısıyla, doğru modelin seçimi ve uygulanması, finansal tahminlerin doğruluğunu ve güvenilirliğini sağlamak açısından kritiktir. Bu projenin sonuçları, finansal varlıkların gelecekteki fiyat hareketlerini tahmin etme konusunda fikir verici olabilir ve gelecekteki benzer projeler için yol gösterici olabilir.



KULLANILAN KAYNAKLAR

- > <https://github.com/afnhsn/deepPrediction>
- > https://github.com/zkhotanlou/LSTM_and_GRU_Stock_Prediction/blob/main/README.md
- > <http://adudspace.adu.edu.tr:8080/jspui/bitstream/11607/4518/1/3154.pdf>
(DERİN ÖĞRENME İLE HİSSE SENEDİ PİYASASI TAHMİNİ)
- > <https://www.youtube.com/watch?v=XZMZrFJE4z8>
(Yapay Zeka Borsa Hisse Senedi Tahmin Uygulaması -V1)
- > <http://iibfdergisi.ksu.edu.tr/en/download/article-file/334511>
(HİSSE SENEDİ KAPANIŞ FİYATLARININ YAPAY SİNİR AĞLARI VE BULANIK MANTIK ÇIKARIM SİSTEMLERİ İLE TAHMİN EDİLMESİ)
- > <https://dergipark.org.tr/en/download/article-file/960735>
(THY Hisse Senedi Değerinin Yapay Sinir Ağları İle Kestirimi)
- > Machine Learning (Unlocking Insights for Heart Failure Prediction)
- > Top 10 Machine Learning Algorithms (Bosscoder Academy)
- > Top 10 Machine Learning Algorithms (tutort Academy)
- > ChatGPT - Copilot ve benzeri yapay zeka araçlarından yardım alınmıştır.



Data Madrid Ekibi



**BİZİ DİNLEDİĞİNİZ İÇİN
TEŞEKKÜRLER**

İLETİŞİM

Enes Berke Karaoğlu



Furkan Demir



Emirhan Tozlu

