

T.C.
Yalova Üniversitesi Mühendislik Fakültesi Bilgisayar
Mühendisliği Bölümü
Yapay Sinir Ağları Dersi
Proje Raporu



Dr. Öğretim Üyesi Güneş HARMAN

Konu: Ev Fiyatı Tahmin edilmesi

Hazırlayan

Şaban Emre Çat – 160101007

GİRİŞ

PROJE TANIMI

Veri setinde bulunan verilere göre çok katmanlı yapay sinir ağı yardımı ile makine öğrenimi gerçekleştirilerek bir evin fiyatını tahmin edilmiştir. Giriş olarak 21 farklı giriş kullanılmıştır. Çıkış olarak 'price' değeri olan ücret tahmin edilmeye çalışılmıştır.

VERİ SETİ

Veri setinin alındığı adres:

<https://www.kaggle.com/shivachandel/kc-house-data>

Veri seti daha önce kullanıldı mı?

Kaggle'dan elde edilen bilgiye göre data set üzerinde Lineer Regression, Random Forest, SVR ve Artificial Neural Network gibi yöntemler kullanılarak on adet proje gerçekleştirilmiştir.

PROJENİN FAYDASI

İnsanlar için kolaylık sağlanarak bölgede bulunan evler ve durumları göz önüne alındığı zaman evlerin yaklaşık fiyatlarını kolaylıkla öğrenmelerine destek vererek konu hakkında fikri olmayan kişilere yardımcı olmak.

METOT

KULLANILAN ARAÇLAR

Proje yapım aşamasında programlama dili olarak Python 3.8.7 kullanılmıştır. Ortam olarak Anaconda Spyder tercih edilmiştir.

KULLANILAN PARAMETRE DEĞİŞKENLERİ

Aktivasyon fonksiyonu: Relu

Test_size=0.3

Random_state=101

Optimizer=Adam

Loss=mean_absolute_error

Batch_size=256

Validation_split=0.2

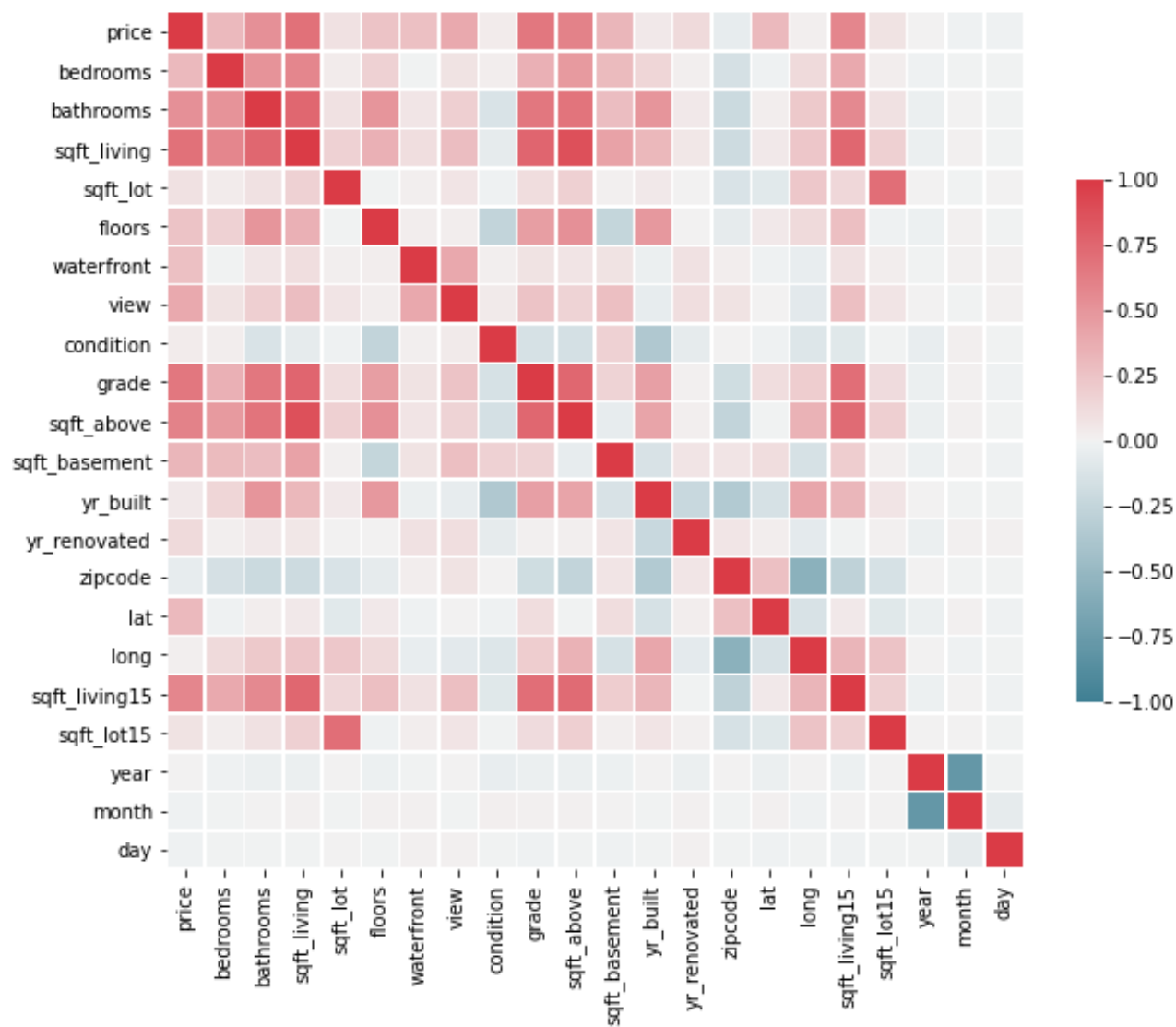
Epoch=300

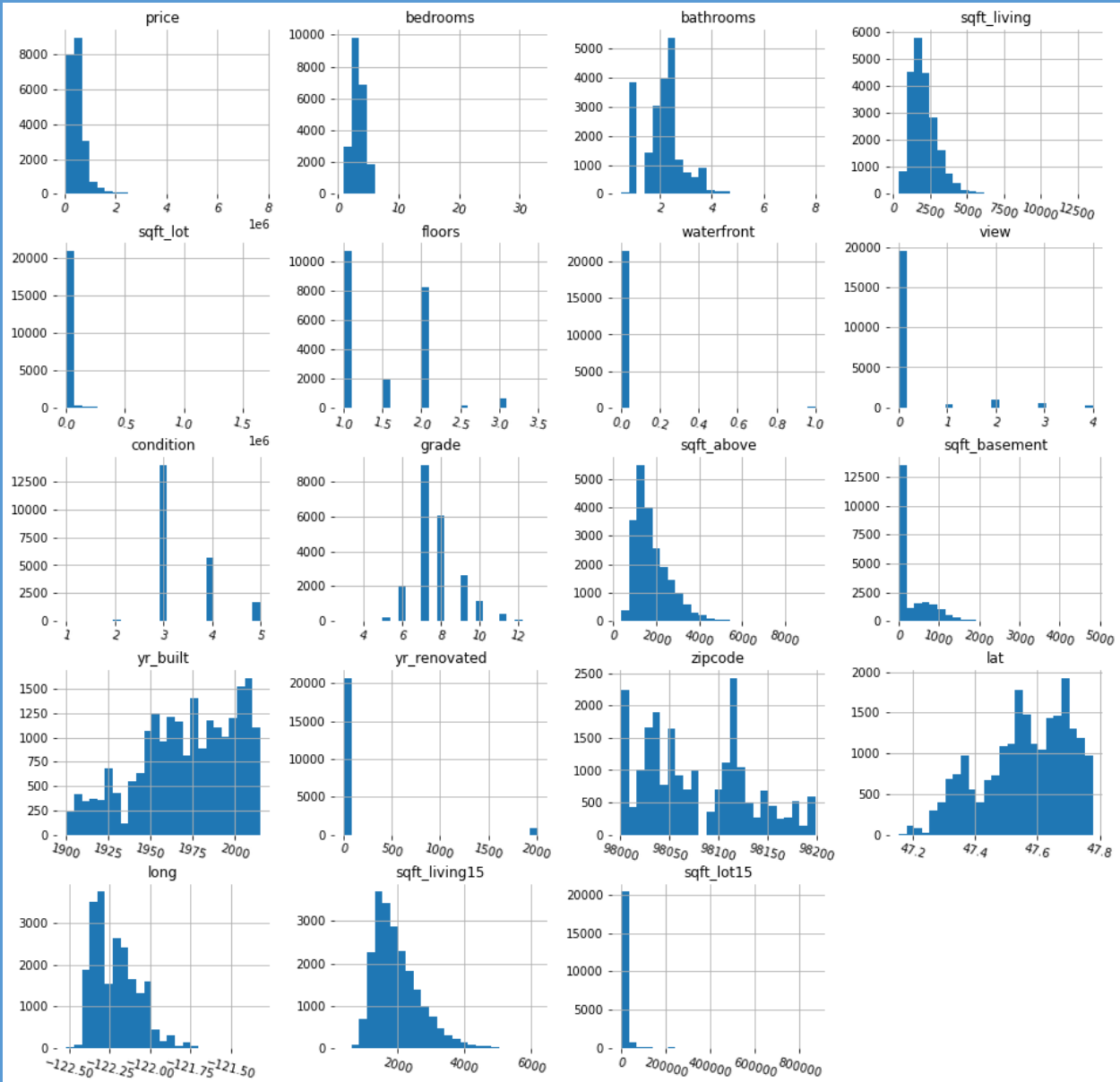
Başarı kriteri= Explained Variance Score, R2 Score

Hata oranı: Mean Absolute Error

Bu proje kapsamında yazılan kodların kısaca anlatımı;

Csv formatındaki veri seti okunup giriş ve çıkış değerleri arasında düzenlemeler yapıldı. Date tipindeki 'date' sütunu gün ay yıl olarak yeni üç sütun haline getirildi. 'Id' sütunu aşırı öğrenme problemi yaratmamak için çıkarıldı. Veri giriş ve çıkışlar ayrılırken %70 eğitim verisi ve %20 validation verisi olarak ayrıldı ve çok katmanlı yapay sinir ağı modeli oluşturuldu Tüm katmanlarda relu aktivasyon fonksiyonu kullanıldı. Optimizer olarak Adam tercih edildi ve gözlemlere göre öğrenme katsayısında değişiklikler yapıldı. Epoch değeri 300 ve batch_size 256 olarak belirlenerek eğitim uygulandı. Korelasyon matrisi ve özelliklerin sayısal dağılımları aşağıdaki gibi gözlendi.





SONUÇ

1. GRAFİK

Explained Variance Score: 0.8433968504980434

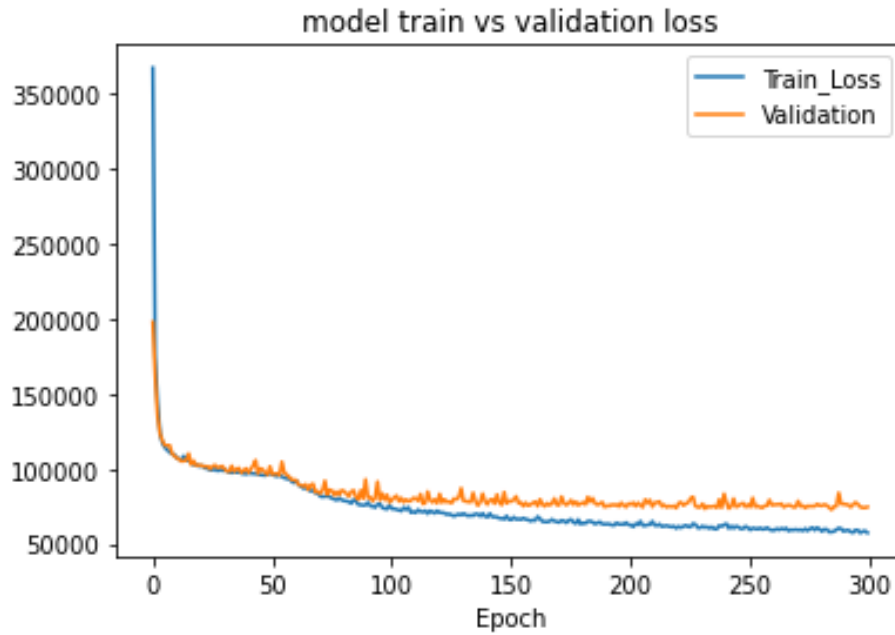
Gizli katman sayısı: 3

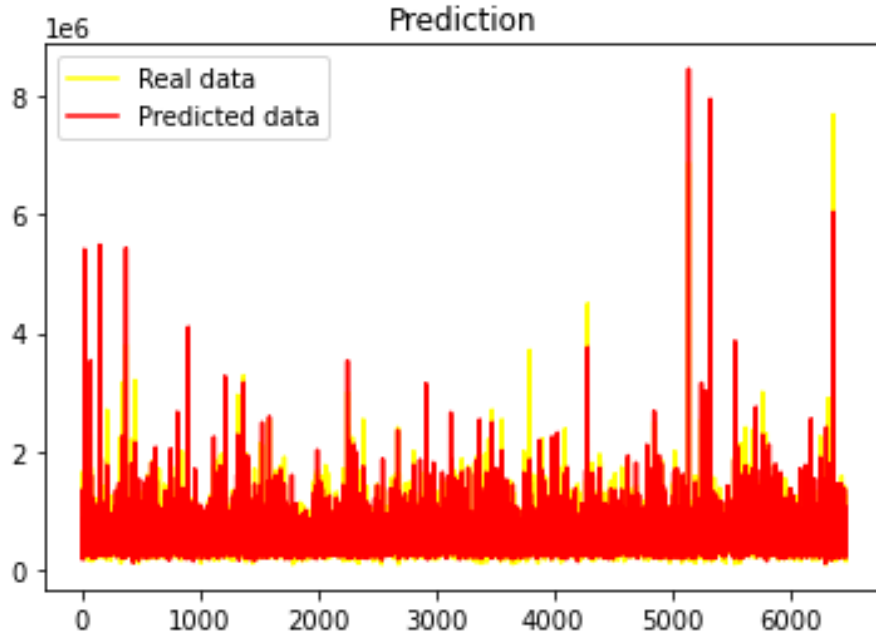
Nöron sayısı: 3 katmanda da 150

Learning rate= 0.01

Epoch sayısı: 300

Aktivasyon fonksiyonu: tüm katmanlarda relu





Epoch sayısı 50-100 arasında bir değerde iken valudation loss değeri ve train loss değeri birbirinden ayrılmıştır. Bu durumda overfitting problemi meydana gelmiştir.

2. GRAFİK

Explained Variance Score: 0.8345985091081982

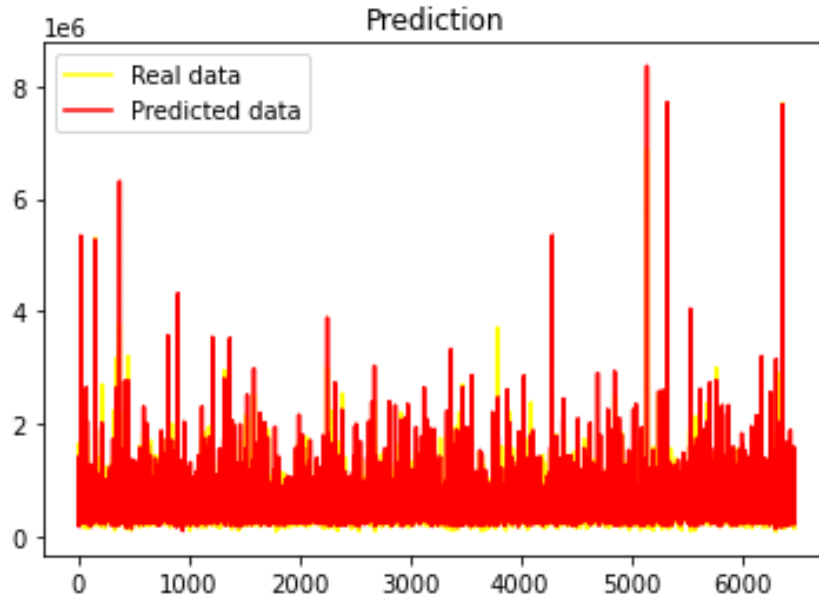
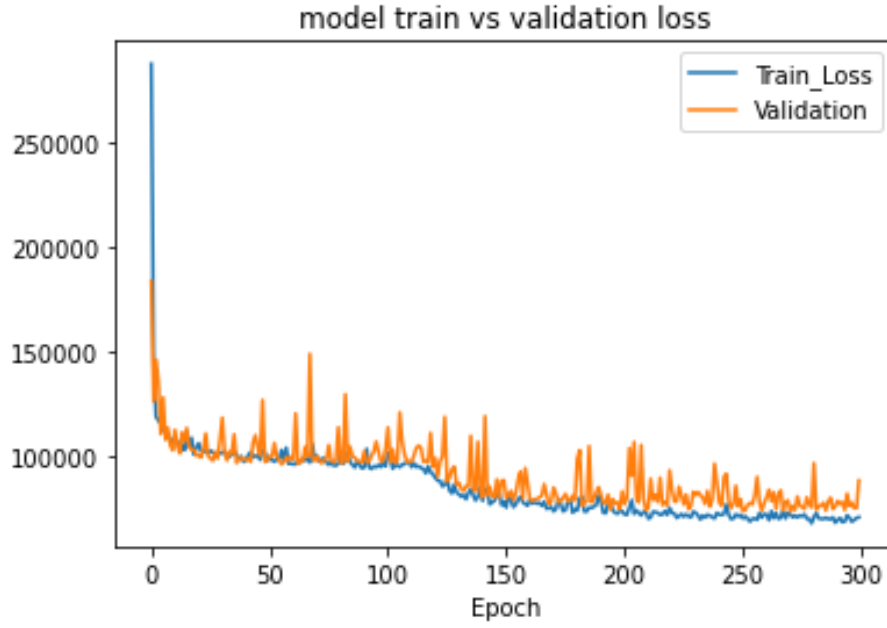
Gizli katman sayısı: 7

Nöron sayısı: 7 katmanda da 20

Learning rate= 0.05

Epoch sayısı: 300

Aktivasyon fonksiyonu: tüm katmanlarda relu



Yapılan deęişikler sonucu overfitting durumu devam etmektedir.

3. GRAFİK

Explained Variance Score: 0.8849226934754193

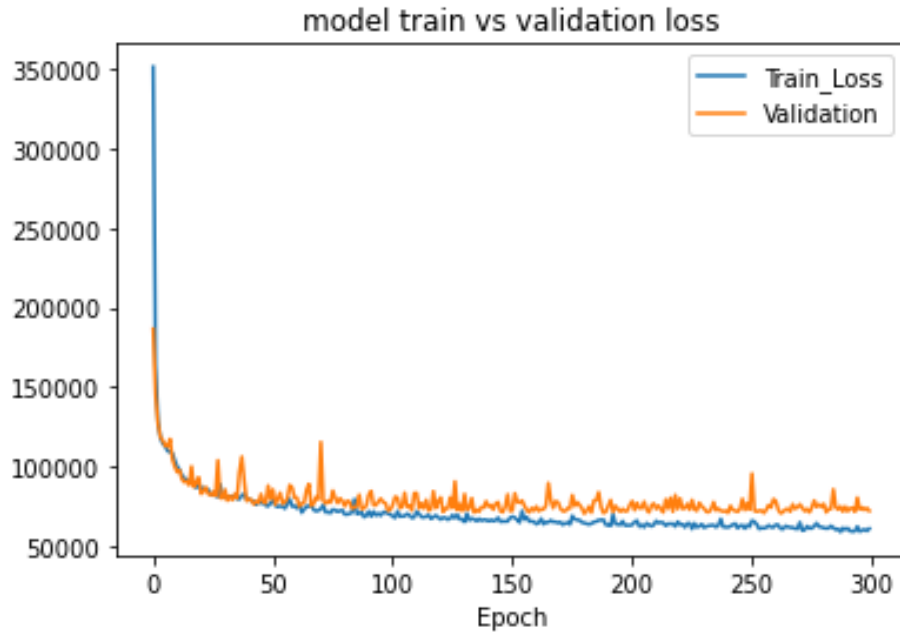
Gizli katman sayısı: 7

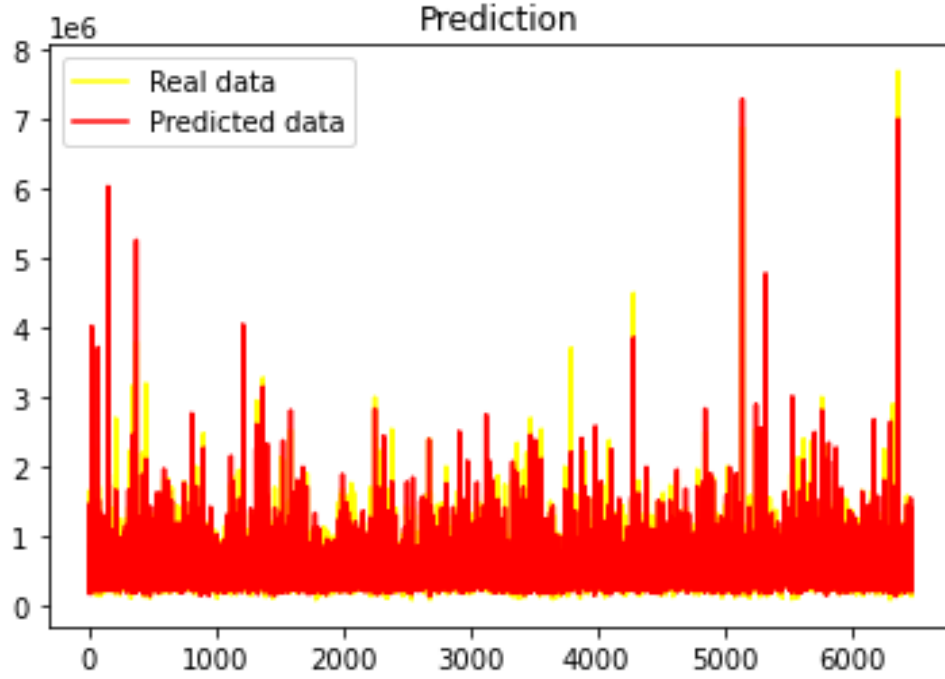
Nöron sayısı: ilk 6 katmanda da 50, 7. Katmanda 20

Learning rate= 0.05

Epoch sayısı: 300

Aktivasyon fonksiyonu: tüm katmanlarda relu





Nöron sayılarında yapılan değişiklikler çizimdeki anlık değişikliklerin önüne geçse de overfitting durumuna çözüm olamamıştır.

4. GRAFİK

Explained Variance Score: 0.8903495872959877

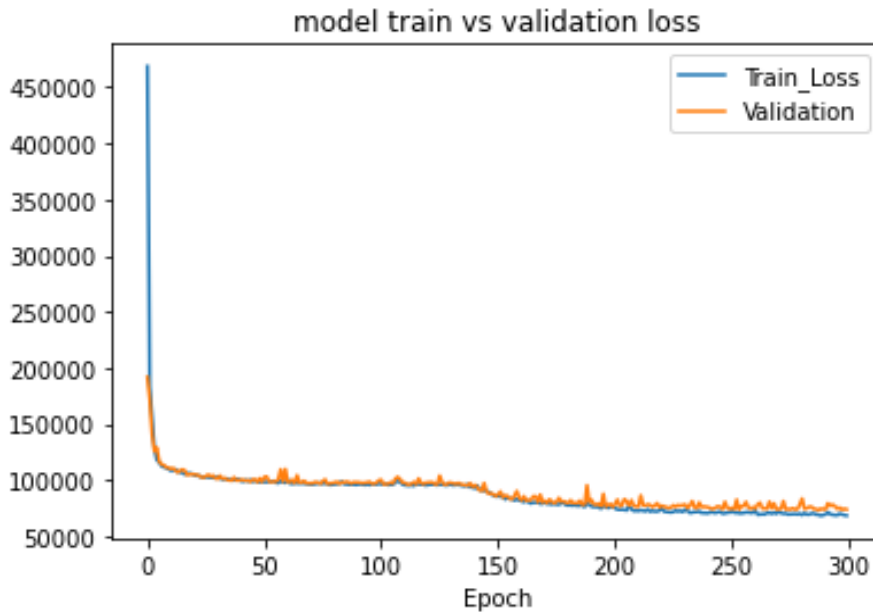
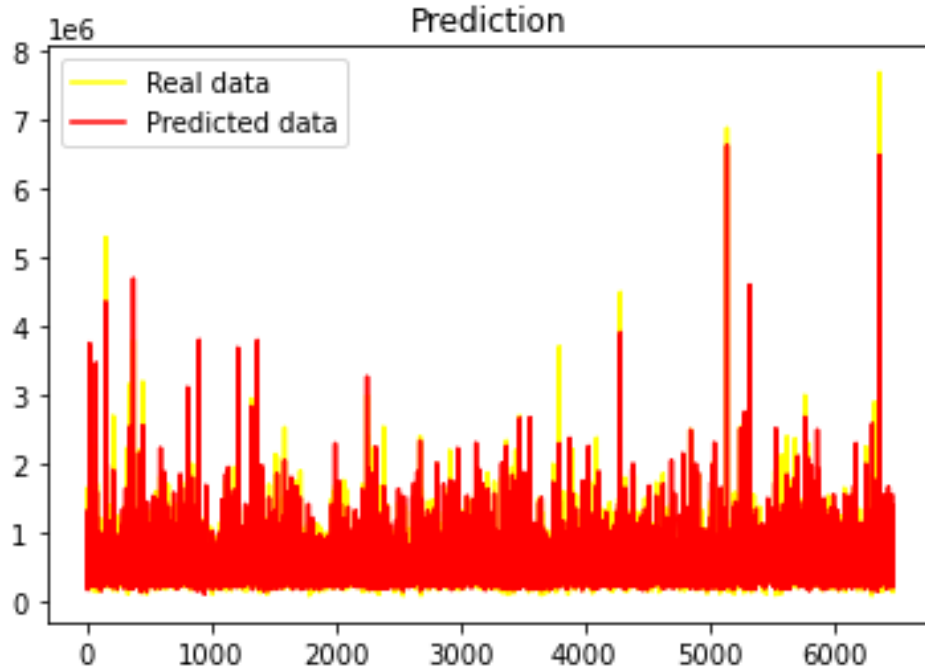
Gizli katman sayısı: 7

Nöron sayısı: 7 Katmanda 20

Learning rate= 0.01

Epoch sayısı: 300

Aktivasyon fonksiyonu: tüm katmanlarda relu



Nöron sayılarında ve learning rate değerinde yapılan azaltma sonucunda oluşan grafikte overfitting problemi hala şüphe uyandırıcı ufak farklarla göze çarpıyor.

Explained Variance Score: 0.8597290835345583

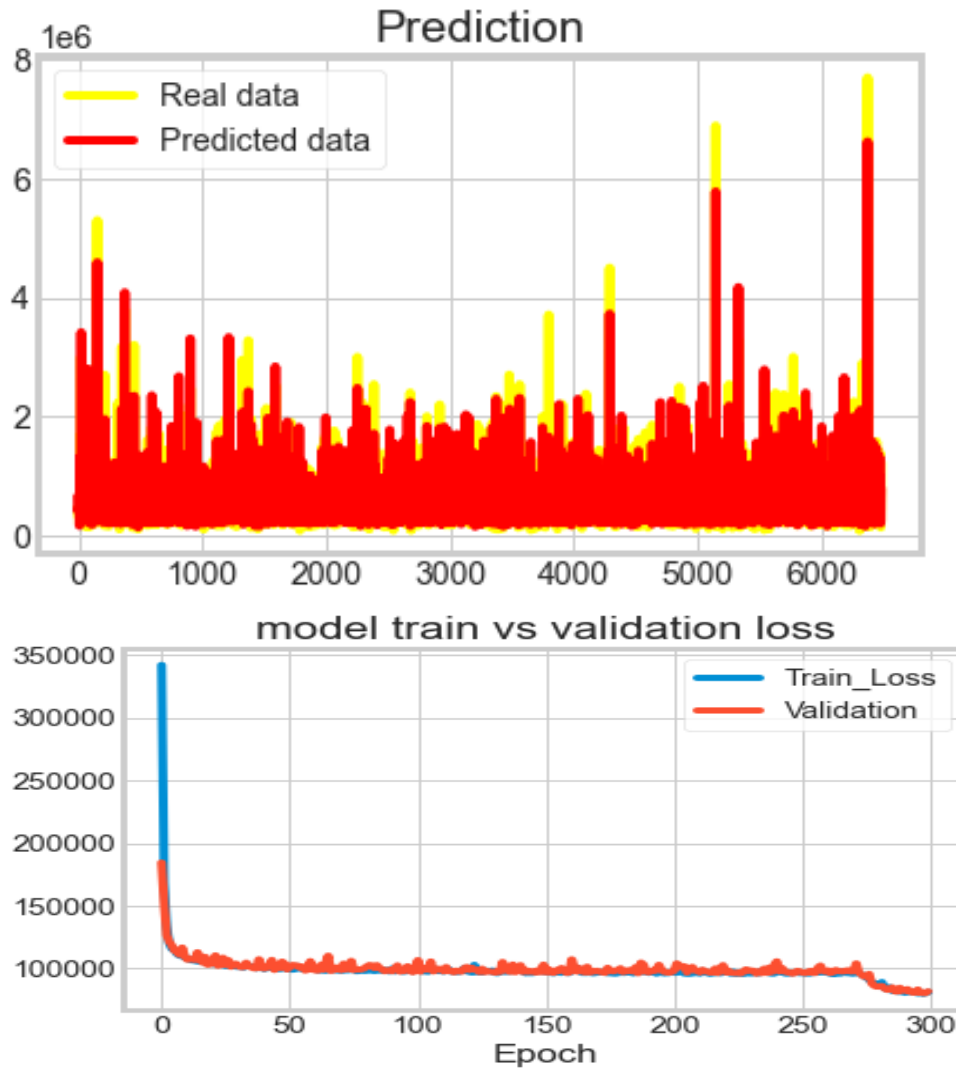
Gizli katman sayısı: 3

Nöron sayısı: 3 Katmanda 20

Learning rate= 0.07

Epoch sayısı: 300

Aktivasyon fonksiyonu: tüm katmanlarda relu



Sonuç kısmında gösterilen grafikler dışında farklı değerler ile yapılan gözlemler sonucunda aşırı öğrenme problemine maruz kalmadan ortaya çıkan en yüksek score değerlerinden birisi olarak 0.8597290835345583 olarak bulunmuştur. Giriş katmanı 20 nöron, 3 gizli katman 20 nöron ve çıkış katmanından oluşmuştur. Öğrenme oranı olarak 0.07 tercih edilmiştir. Eğitim 300 adımda gerçekleştirilmiştir ve batch_size 256 olarak alınmıştır.