**logo, simge, sembol, amblem, ticari marka içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Sakarya Uygulamalı Bilimler Üniversitesi**

**Bilgisayar Mühendisliği**

**Yapay Öğrenmenin Temelleri**

**Proje Ödevi**

**Semih Dursun**

**B210109018**

**Facebook Gönderi Ölçüleri**

**1.Giriş**

Bu projede, Facebook Metrics veri setini kullanarak Facebook'a ait gönderilerin çeşitli özelliklerini analiz ettim. Veri seti, gönderi türü, gönderi zamanı, beğeni sayısı, paylaşım zamanı gibi birçok özelliği içermektedir. İstenilen özellikleri dikkate alarak, bu gönderilerdeki toplam etkileşimi belirlemek amacıyla lineer regresyon modeli uyguladım. Ardından, analizimi daha da geliştirmek ve optimize etmek için gradient descent algoritmasını kullanarak modeli iyileştirdim. Bu yaklaşım, veri setindeki özelliklerle toplam etkileşim arasındaki ilişkiyi daha iyi bir şekilde anlamamı sağladı.

**2 .Metodoloji**

**2.1 Veri Önişleme**

Projemizin ilk aşamasında, Facebook Metrics veri setini pandas kütüphanesini kullanarak başarılı bir şekilde okuyup elde ettik. Veri setinde "paid", "like" ve "share" sütunlarında eksik veriler bulunmaktaydı. Eksik veri sayısına bakarak, bu eksikleri doldurmayı tercih ettim, çünkü sayıları çok fazla değildi. Ancak, eğer çok fazla eksik veri bulunsaydı, sütunları silme işlemi düşünülebilirdi, çünkü bu durumda veri doğruluğunda ciddi bir sapma olabilir. Analiz için belirlenen sütunlar üzerinde çeşitli işlemler gerçekleştirdim. Eksik veri düzeltme adımından sonra, Exploratory Data Analysis (EDA) analizi uygulayarak verilerin etkileşim üzerindeki etkisini inceledim. Sütunları kategorik verilere dönüştürmek amacıyla get\_dummies fonksiyonunu kullanarak sütunları ayrıştırdım. Bu işlemi, sütunlardaki farklı veri türlerinin etkileşimdeki etkisini daha iyi anlamak ve modelin karmaşıklığını azaltmak için gerçekleştirdim. Ayrıca, "Type" sütununu kategorik veriden sayısal forma dönüştürerek regresyon modeli için daha uygun bir formata getirdim. Bu dönüşüm, modelin daha etkili ve doğru bir şekilde öğrenmesine yardımcı oldu. Bu aşamadan sonra, veri seti analizine daha hazır hale getirilmiş olarak devam edebildim**.**

**2.2 Veri Bölme ve Standartlaştırma**

Veri setimi, modelin eğitim ve test aşamalarında kullanılması için %80 eğitim ve %20 test oranında olacak şekilde bölerek veri setimi iki ayrı alt küme haline getirdim. Ardından, özelliklerin birbirleriyle uyumlu bir ölçeğe sahip olmasını sağlamak amacıyla StandardScaler fonksiyonunu kullanarak verilerimi standartlaştırdım. Bu standartlaştırma işlemi, regresyon modeli için daha sağlıklı ve doğru sonuçlar elde etmeme yardımcı oldu. Verilerin standartlaştırılması, her özelliğin ortalamasını sıfır ve standart sapmasını bir yaparak, modelin öğrenme sürecini optimize etmeye katkıda bulundu. Bu adımlar, modelin daha güvenilir ve genelleştirilebilir sonuçlar üretmesini sağlamak amacıyla gerçekleştirildi.

**2.3 Model Oluşturma**

Projede, öncelikle kendi oluşturduğum Lineer Regresyon ve gradyan iniş algoritmalarını kullanarak modelleme işlemlerini gerçekleştirdim. Bu adımların ardından, elde ettiğim sonuçları değerlendirmek ve kendi implementasyonumun performansını kontrol etmek amacıyla sklearn kütüphanesine ait olan LinearRegression() fonksiyonunu kullanarak bir model daha oluşturdum. Bu şekilde, benim geliştirdiğim model ile kütüphanedeki hazır model arasındaki farkları anlamaya çalıştım ve tahmin sonuçlarını karşılaştırdım. Daha sonra, analizimi genişletmek ve model seçimindeki etkenleri anlamak adına sklearn kütüphanesinin Random Forest regresyon modelini kullandım. Bu model ile hangi sütunların etkileşim açısından daha önemli olduğunu belirlemek adına özellik önem sıralamasını inceledim. Bu adımlar, projenin sonuçlarını daha kapsamlı bir şekilde değerlendirmeme ve hangi modellerin daha iyi performans gösterdiğini anlamama yardımcı oldu.

**2.4 Eğitim ve Tahmin**

Model, eğitim veri seti üzerinde belirtilen iterasyon sayısı kadar eğitildi. Ardından, eğitilen model kullanılarak test veri seti üzerinde tahminler yapıldı ve elde edilen sonuçlar detaylı bir değerlendirmeye tabi tutuldu. Bu adımlar, modelin genelleme yeteneğini ve gerçek dünya verileri üzerindeki performansını anlamak amacıyla gerçekleştirildi. Modelin eğitim ve test aşamalarındaki başarıları, projenin amacına uygun bir şekilde incelendi ve sonuçlar detaylı bir analizle değerlendirildi.

**3 Sonuçlar**

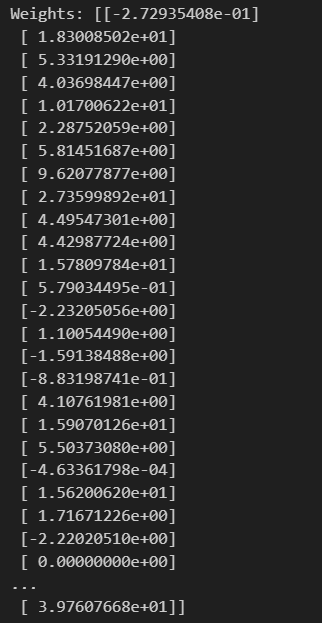
Kendi oluşturduğum class ile çıktılar ;

1→ modelin ağırlık matrisi

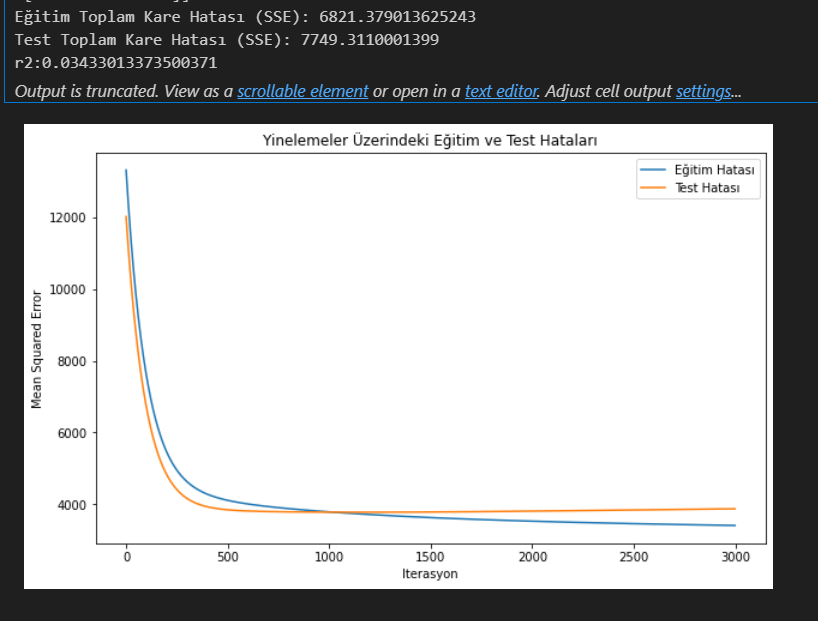
2→Eğitim ve test modeline ait Toplam kare hatası (SSE) ve r2 sonuçlarının çıktısı ve eğitim ve test hatalarının grafiği

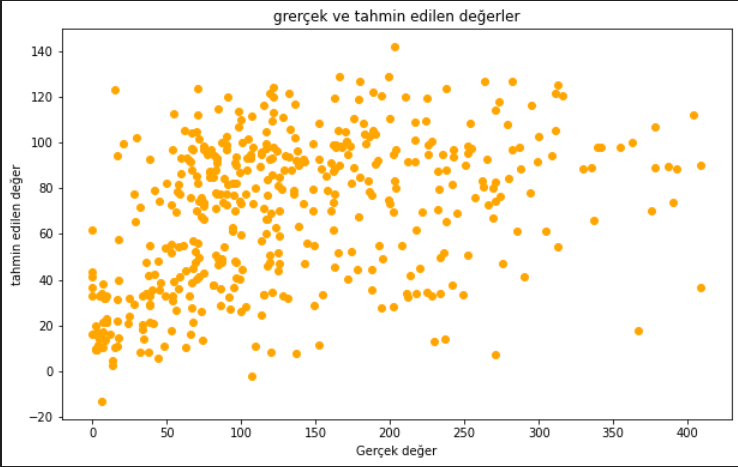
3→ Gerçek verileri ile Tahmini verilerin karşılaştırmasının grafiği

**(1)**

****

**(2)**

****

**(3) **

Sklearn kütüphanesi ile alınan sonuçlar

1→ Linear regresyon ile alınan R2 çıktısı

2→ Linear regresyona ait gerçek değerlere göre tahmin grafiği

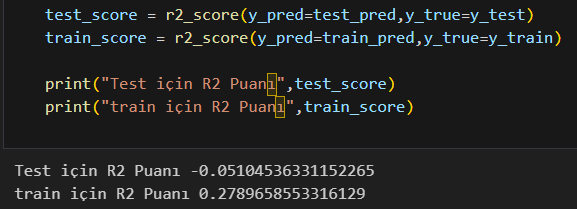
3→ Elastik ağ modeli uygulanrak alınan r2, MSE ve RMSE değerleri

4→ LightGBM regresyon modeli kullanılarak alınan RMSE ,R2 ve gerçek değerlere karşılaştırma grafiği

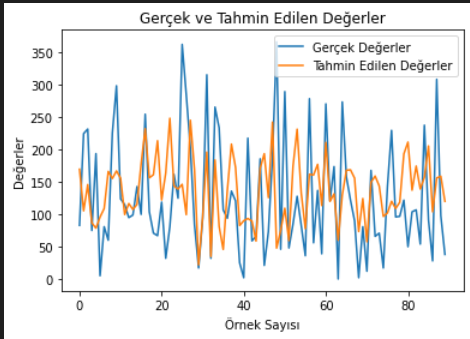
5→ Random forest algoritması kullanılarak alınan sonuçlar

6→Hangi sütun(özellik) etkileşim daha etkili olduğu sonucu

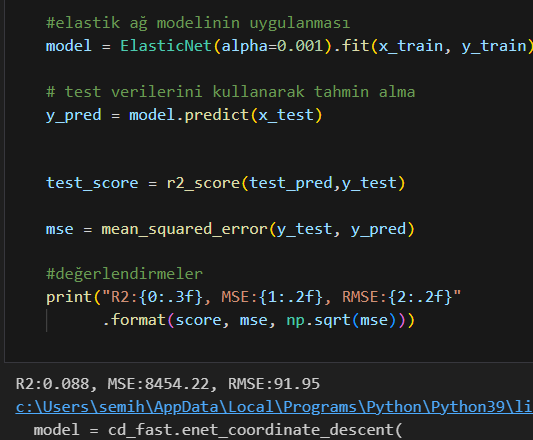
**(1)**

****

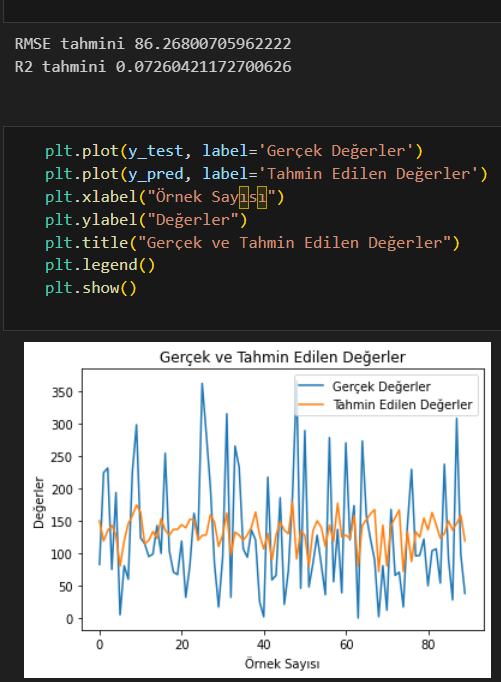
**(2)**

****

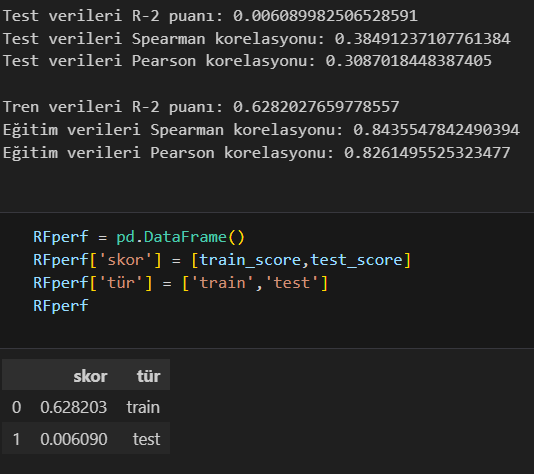
**(3)**

****

**(4)**

****

**(5)**

****

**(6)**

**metin, ekran görüntüsü, yazı tipi, menü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**4 Sonuçlar**

Yapılan analizler sonucunda, Lineer Regresyon ve Gradyan İniş modellerinin Facebook gönderi performansını tahmin etmede mükemmel çalıştığı iddia edilemez. Veri setindeki veriler arasında belirgin bir uyumsuzluk bulunması nedeniyle, bu modeller uyumu yakalamakta zorlandılar. Ancak, bu modellerle gerçekleştirilen işlemler sonucunda elde edilen tahminlerin doğruluğu göz önüne alındığında, yine de kabul edilebilir düzeyde başarı elde edildi. Sklearn kütüphanesinde kullaranak denenen regresyon, Elastik Ağ modeli, LightGBM Regresyon Analizi ve Random Forest algoritmaları arasında yapılan karşılaştırmada ise, normal regresyon modelinin en iyi sonuçları verdiği görüldü. Bu durum, veri setinin özellikleriyle uyumlu olarak, basit bir regresyon modelinin diğer daha karmaşık modellere üstünlük sağladığını işaret edebilir. Son olarak, etkileşim analiziyle belirlenen en önemli sütunun ne olduğu, model performansını etkileyen kritik bir faktör olarak belirlendi. Bu bulgu, Facebook gönderilerinin performansını tahmin etmede kullanılan modellerde, özellikle belirli bir etkileşim özelliğine odaklanmanın önemini vurgular. Bu çeşitli modeller ve analizler, veri setinin karmaşıklığına ve karakteristiğine bağlı olarak farklı sonuçlar elde etmenin mümkün olduğunu göstermektedir. Model seçimi ve performans değerlendirmesi, veri setinin özelliklerini anlamak ve uygun algoritma veya teknikleri seçmek açısından kritik öneme sahiptir.

Referanslar:

Andrew NG → <https://www.coursera.org/learn/machine-learning/>

Musa Arda → <https://www.udemy.com/course/makine-ogrenmesi-egitimi/>