logo, metin, simge, sembol, grafik içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Facebook Verileriyle Etkileşim Tahmini Modeli**

Yazar Adı SOYADI Selin Alp, Ezgi Gül

1. **Giriş**

İçinde bulunduğumuz son 20 yılda teknolojinin inanılmaz hızlı gelişmesi ile beraber birbirimizle olan etkileşimlerimiz son derece artmıştır. Özellikle sosyal medyanın hayatımıza girmesiyle birlikte hem kilometrelerce ötedeki insanlarla etkileşimde bulunabiliyor, hem de büyük insan kitlelerinin verilerine ulaşabiliyoruz. Bu kadar büyük veri havuzlarının elbette ki ayrıştırılması ve yorumlanması kaçınılmaz bir sonuçtu. Özellikle bu konuda yapay zeka tekniklerinin de kullanılması; devasa veri setlerinin yorumlanmasında, model eğitilip tahminde bulunulmasında çok işe yaramıştır. Model oluşturmak, eğitilmesi için verilen değerler kullanılarak yeni değerimize en az hatayla yaklaşımda bulunabilecek bir fonksiyon üretmektir.

Biz bu projede ünlü bir kozmetik markasının 2014 yılı boyunca Facebook sayfasında yayınlanan gönderilerine ilişkin oluşturulmuş bir veri setini kullanarak doğrusal regresyon modeli oluşturacağız. Bu veri setindeki seçilen bazı özellikleri (sütunları) kullanarak bir model eğiteceğiz. Bu özellikleri seçerken dikkatli olunması gerekir. Çünkü aşırı öğrenme ya da öğrenmeme problemleriyle karşılaşabiliriz. Bu gibi durumlarda modelimizin tahmini gerçek değerinden çok farklı çıkacaktır, istenmeyen bir durumdur.

Doğrusal, bir diğer adıyla Lineer Regresyon, ilgili ve bilinen başka bir veri değeri kullanarak bilinmeyen verilerin değerini tahmin eden bir veri analizi tekniğidir.

1. **Metodoloji**

**Araştırma Problemi:**

Python programlama dilini kullanarak verilen veri setiyle Doğrusal Regresyon kodu yazılacaktır.

**Veri Seti:** Facebook Posts Metrics

**Veri Seti Öznitelikleri:**

• Kategori,

• Sayfa toplam beğenisi: Şirketin sayfasını beğenen kişi sayısı),

• Tür: İçerik türü (Bağlantı, Fotoğraf, Durum, Video),

• Gönderi ayı: Gönderinin yayınlandığı ay (Ocak, Şubat, Mart,…, Aralık),

• Gönderim saati: Gönderinin yayınlandığı saat (0, 1, 2, 3, 4, …, 23),

• Hafta içi gönderi: Gönderinin yayınlandığı hafta içi (Pazar, Pazartesi, …, Cumartesi),

• Ücretli: Şirketin reklam için Facebook'a ödeme yapması durumunda (evet, hayır)

**Modelimizde ilk kullanacağımız özellikler:**

Modellemeye çalışabileceğimiz birçok olası özellik var ancak biz “Total Interactions” ‘e odaklanacağız. Özellik alanımız şunları içerecektir: “Category”, “Page total likes”, “Post month”, “Post hour”, “Post weekday”, “Type” ve “Paid”.

**Takip Ettiğimiz Adımlar:**

• Verinin indirilmesi

• Verinin okunması: Veri seti modelimize aktarılarak, istediğimiz özelliklere göre şekillendirildi. Null değerleri dolduruldu.

• Verileri Bölme (X ve Y (X\_train, X\_Test ve y\_train, y\_test))

a. Veri kümesini X ve Y'ye bölündü.

b. Sağlanan yüzdesel bölmeyi kullanarak X ve Y'yi eğitim ve test setlerine ayrıldı (varsayılan, %80 eğitim ve %20 testtir).

• Doğrusal regresyon fonksiyonunu oluşturabilmek için gerekli araştırmalar yapıldı. Matematiksel olarak anlaşılarak gerekli sınıf içinde fonksiyonlar oluşturuldu.

• Modelin eğitimi ve test edilmesi

a. train\_test\_split() öğesini çağrılarak eğitim ve test seti alındı.

b. Bir ağırlık (θ) vektörü tanımlandı.

c. Yukarıdaki bilgiler kullanılarak Gradyan İnişini (Gradient Descent) uygulandı.

d. Eğitim ve test verileri için Toplam Kare Hata (Sum Squared Error) kaydedildi.

e. Ağırlık matrisi, eğitim hataları ve test hataları döndürüldü.

f. Eğitim ve test hataları çizildi. Grafikte aşırı öğrenme gerçekleştiği için farklı sütunlarla yeniden denendi.

1. **Bulgular ve Tartışma**

Araştırma verilerinin analiz edilmesiyle ortaya çıkan çıktılar, araştırmanın bulguları olarak ifade edilir. Bulgular; tablo, şekil, grafik ya da hesaplamalar yoluyla ortaya konur.

Lineer regresyonda tahminleme yapabilmemiz için bir model kurmamız gerekmektedir. Kuracağımız model aşağıdaki gibi olacaktır:





* "b0" değeri bizim sabit değerimizdir
* "b1" değeri ise modelimizin eğimini temsil eder
* "x1" değeri ise tahminleme yapılacak noktayı temsil eder

Python’da yapacağımız kısım ise buradaki "b0" ve "b1" değerlerini bulmak olacak. Buradaki değerleri bulmanın birçok yolu olsa da "En Küçük Kareler Yöntemini kullanacağız. Bu yönteme göre





Denklemini kullanarak "b1" ve "b0" değerlerini bulabiliyoruz. Buradaki ^T işareti o matrisin transpozunun alınmasını, ^-1 işareti de o matrisin tersinin alınması gerektiğini söylüyor.

Grafikler hakkında yorum yapabilmek için öncelikle modelin aşırı öğrenmesi ve öğrenememesi durumları hakkında bilgi sahibi olmalıyız.

**Overfitting:**

* Bir model çok fazla veriyle eğitilirse bazen aşırı öğrenmiş yani ezberlemiş oluyor. Bu duruma overfitting deniliyor.
* Modelin eğitim verilerinden "aşırı öğrendiği" durumdur.
* Sadece öğrendiği verilerden başarı elde edebilir.
* Modelin özellik sayısının gözlem sayısına kıyasla çok olduğu durumlarda gözlemlenir.
* Overfitting, parametrik olmayan ve doğrusal olmayan yöntemlerdir.
* Bir overfit modeli, eğitim verilerinde çok düşük bir tahmin hatası verir.
* Test verilerinde çok yüksek bir tahmin hatası verir.
* Yüksek varyans ve düşük önyargı vardır.

**Aşırı uyumu azaltmak için yapılabilecekler:**

* Eğitim verileri artırılabilir.
* Model karmaşıklığını azaltılabilir.
* Birbiri ile korelasyon içinde olan sütunlar silinebilir.
* Eğitim ve test verilerinde hatalar arasındaki fark belli bir seviyedeyken eğitim durdurulabilir.

**Underfitting:**

* Bir makine öğrenme algoritmasının veya bir modelin verilerin temelindeki eğilimi yakalayamaması durumunda ortaya çıkan durumdur.
* Modelin eğitim verilerinden "yeterince öğrenemediği" durumdur.
* Eğitim verilerinden öğrendiklerini görmez bu şekilde girdiler ile çıktılar arasındaki temel ilişkiyi öğrenemez.
* Makine öğrenimi modelimizin doğruluğunu yok eder.
* Modelimizin veya algoritmanın verilere yeterince uymadığı anlamına gelir.
* Genellikle doğru bir model oluşturmak için daha az veriye sahip olduğumuzda ortaya çıkmaktadır.
* Girdi özellikleri hedefi iyi tanımlayacak kadar açıklayıcı olmadığında söz konusu olabilmektedir.
* Yüksek önyargı ve düşük varyans vardır.

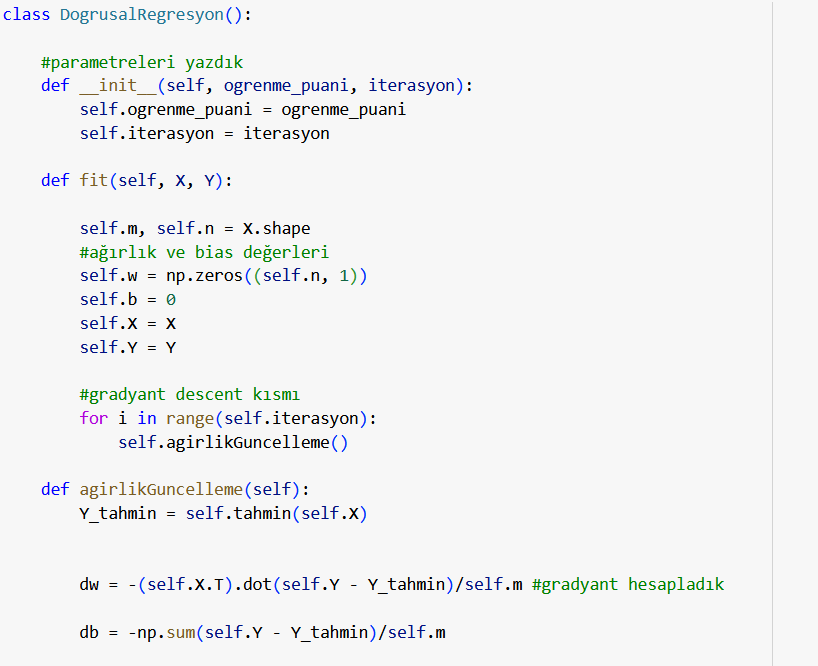
**Yetersiz uyumu azaltmak için yapılabilecekler:**

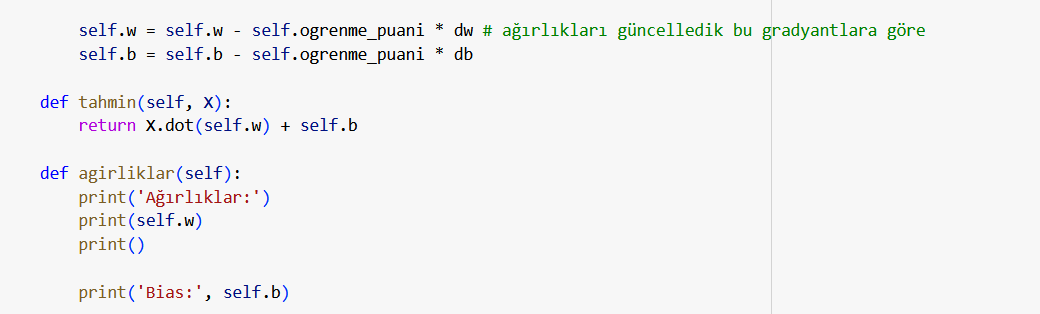
* Model karmaşıklığı arttırılabilir.
* Özelliklerin sayısı arttırılabilir.

metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Kullandığımız lineer regresyon sınıfı ve içindeki fonksiyonlar şu şekildedir:





Gerekli işlemleri yaptığımızda grafiklerimiz bu şekilde gözüküyor.

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, yazılım içeren bir resim

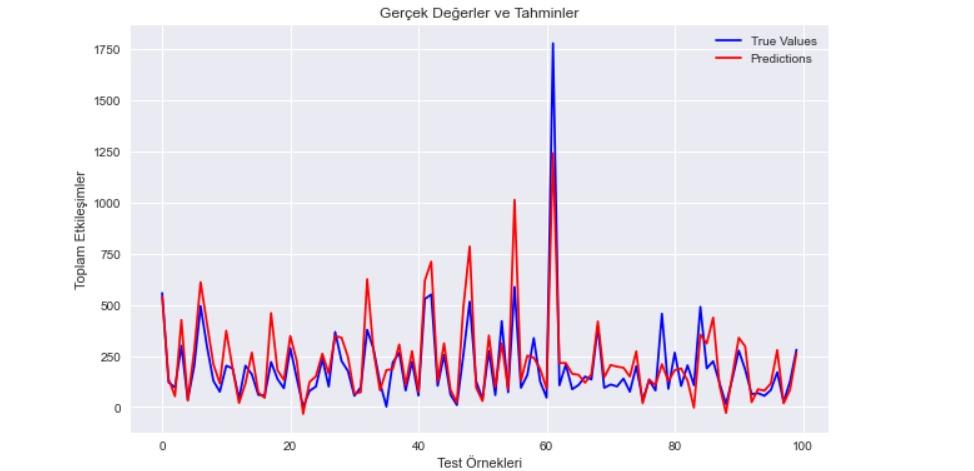
Açıklama otomatik olarak oluşturuldu

Özellikle birinci grafiğe baktığımızda çizgimizin çoğu değer ile arasında büyük bir hata döndereceğini, karşılamadığını görüyoruz.

**MSE (Mean Squared Error - Ortalama Kare Hata):** Modelin tahmin ettiği değerler ile gerçek değerler arasındaki ortalama karesel farkı ölçer. MSE ne kadar düşükse, modelin daha iyi performans gösterdiği söylenebilir. Ancak, MSE çok düşükse (eğitim setine aşırı öğrenmişse), model test setinde kötü performans gösterebilir. Burada MSE’miz aslında sıfıra o kada da yakın değil.

**SSE (Sum of Squared Errors - Karelerin Toplamı):** MSE'nin öğrenme setindeki tüm örnekler üzerinden toplamıdır. Düşük SSE, iyi bir uyum anlamına gelir, ancak çok düşük SSE (örneğin, sıfır) aşırı öğrenmeyi gösterir. SSE değerimiz çok yüksek, o halde burada aşırı öğrenme yok. Öğrenememe durumu var. Bunun sebebi çok fazla özelliğin olması olabilir. Bu gibi durumlarda nitelikler azaltılarak daha az hatalı bir sonuca ulaşılmaya çalışılmalıdır.

Şimdi bu doğruluk oranını arttırabilmek bazı işlemler uygulayacağız. İkinci dosyamızdaki kodumuzda da gösterebileceğimiz gibi veri setinde birtakım değişiklikler yaptık. Aynı zamanda aldığımız özellikleri azalttık. Bu işlemler sonucunda elde ettiğimiz grafikler aşağıdaki gibidir:



**ekran görüntüsü, metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**ekran görüntüsü, metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**ekran görüntüsü, metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**SONUÇLAR:**

İstatistikte, aşırı uyma, üretilen bir analizin belirli bir veri kümesine aşırı oranda uyum göstermesi, dolayısıyla bu veri kümesinde yer almayan yeni verilere uyum sağlayamaması problemidir. İlk elde ettiğimiz öğrenememe durumuydu. Ancak gerekli değişiklikleri yaparak öğrenmesini arttırmaya çalıştık. Son grafikte ise görüldüğü gibi overfitting yani aşırı öğrenme gerçekleşmiştir.

**Kaynakça**

1. Yanto, M., Sanjaya, S., Yulasmi, Guswandi, D., & Arlis, S. (2021). Implementation multiple linear regresion in neural network predict gold price. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(3), 1635–1642. <https://doi.org/10.11591/ijeecs.v22.i3.pp1635-1642>
2. Putra, J. S., Ramadhani, R. D., & Burhanuddin, A. (2022). Prediksi Harga Saham Bank Bri Menggunakan Algoritma Linear Regresion Sebagai Strategi Jual Beli Saham. *Journal of Dinda : Data Science, Information Technology, and Data Analytics*, *2*(1), 1–10. <https://doi.org/10.20895/dinda.v2i1.273>
3. Alfaris, L., Siagian, R. C., Nasution, B., Sinaga, G. H. D., & Indah, I. (2022). Non-Linear Regresion And Bisection Method Numerical Analysis of Humidity And Temperature Relationships. *Jurnal Pendidikan Fisika Dan Teknologi*, *8*(2), 238–244. <https://doi.org/10.29303/jpft.v8i2.4394>
4. Erwin, R., & Feriyana, W. (2019). KOMPETENSI PROFESIONAL, DAN SUPERVISI PENDIDIKAN SERTA DISIPLIN KERJA DALAM MEMPENGARUHI KINERJA GURU SMP NEGERI 1 BUMIAGUNG WAYKANAN LAMPUNG. *Jurnal AKTUAL*, *17*(1), 53. <https://doi.org/10.47232/aktual.v17i1.33>
5. Suwaldiyana, S. (2021). The Increasing of Employee Performance Through Discipline, Motivation, and Organizational Culture. *INNOVATION RESEARCH JOURNAL*, *2*(1), 65. <https://doi.org/10.30587/innovation.v2i1.2398>

1. Trisnawati, N. (2017). PENGARUH PENGETAHUAN KEWIRAUSAHAAN DAN DUKUNGAN SOSIAL KELUARGA PADA MINAT BERWIRAUSAHA SISWA SMK NEGERI 1 PAMEKASAN. *JURNAL EKONOMI PENDIDIKAN DAN KEWIRAUSAHAAN*, *2*(1), 57. <https://doi.org/10.26740/jepk.v2n1.p57-71>
2. Naelufar, Y., Wijayanti, A., & Fajri, R. N. (2021). FAKTOR YANG MEMPENGARUHI HARGA SAHAM PADA PERUSAHAAN MANUFAKTUR SUB SEKTOR OTOMOTIF. *Jurnal Akuntansi Dan Pajak*, *22*(1), 296. <https://doi.org/10.29040/jap.v22i1.1899>
3. Lestari, S. (2016). Pengaruh Tingkat Profitabilitas, Likuiditas, Leverage, Ukuran Perusahaan dan Umur Perusahaan Terhadap Pengungkapan Islamic Social Reporting pada Perbankan Syariah Indonesia Tahun 2010-2014. *Jurnal Akuntansi Universitas Negeri Surabaya*, *4*(2), 1–24. Retrieved from <https://jurnalmahasiswa.unesa.ac.id/index.php/jurnal-akuntansi/article/view/14722>
4. Musri, S., M, I. S., & Huda, N. (2021). ANALISIS FAKTOR YANG MEMPENGARUHI RETURN SAHAM DALAM ISSI SEKTOR CONSUMER GOODS INDUSTRY. *Media Ekonomi*, *28*(1), 1–14. https://doi.org/10.25105/me.v28i1.7264