**Aylık Satış Tahmini (Linear Regression ile)**

*Ayşegül SARIKAYA – 202113171059*

*Kaggle Linki:* [*https://www.kaggle.com/code/sarkayas/predict-future-salas-linear-regression*](https://www.kaggle.com/code/sarkayas/predict-future-salas-linear-regression)

**1. Giriş**

**Proje Amacı:**  
Bu projede, geçmiş satış verilerine dayanarak gelecekteki satış miktarlarını tahmin etmek için bir Lineer Regresyon modeli geliştirilmiştir. Verilerdeki aykırı değerler ve veri temizliği gibi aşamalar dikkate alınarak modelin performansı değerlendirilmiştir.

**Uygulama Alanları:**

* Stok yönetimi
* Satış tahmini
* Gelir optimizasyonu

**2. Kullanılan Veri Seti**

**Veri Kaynağı:**  
[Competitive Data Science Predict Future Sales](https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales)

**Özellikler (Features):**

* **date\_block\_num:** Ay numarası (zaman serisi).
* **shop\_id:** Mağaza kimliği.
* **item\_id:** Ürün kimliği.
* **item\_price:** Ürün fiyatı.
* **item\_cnt\_day:** Günlük satış miktarı.

**Hedef Değişken:**

* **item\_cnt\_month:** Aylık toplam satış miktarı.

Gerekli kütüphanelerin kurulumu:

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Veriyi yükleme ve gözlemleme:

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

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

**3. Veri Ön İşleme**

**a. Eksik ve Hatalı Değerlerin İncelenmesi:**

Negatif fiyatlar ve satış değerleri ayıklandı.

Tarihler datetime formatına dönüştürüldü.

**b. Aykırı Değerlerin Çıkarılması:**

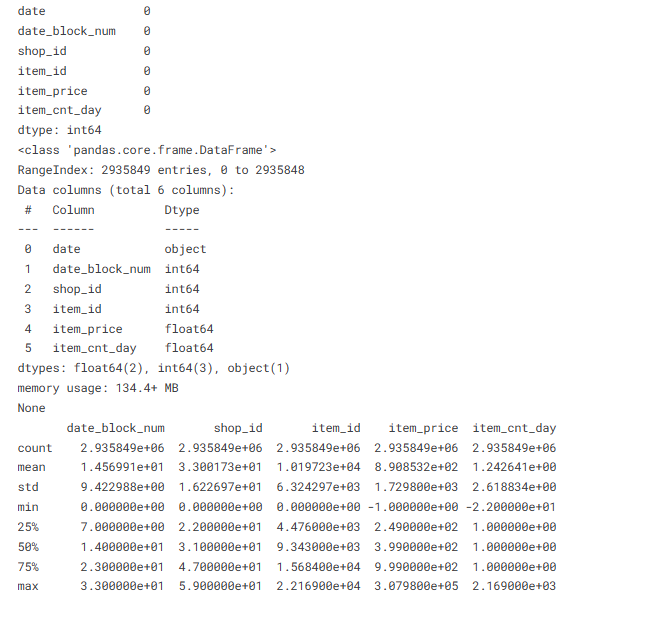
Çok büyük veya anlamlı olmayan değerler kaldırıldı.

**c. Veri Özet Bilgileri:**

Toplam 2.9 milyon veri noktası temizlendikten sonra işlemeye alındı.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu



**4. Görselleştirme**

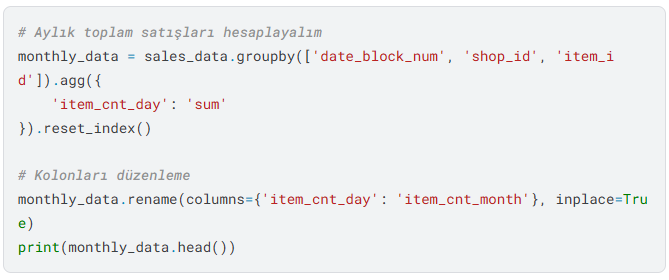
**Aylık Toplam Satışlar:**  
Aylık satışların zaman içindeki değişimi incelendi ve satışların belirgin bir sezonluk eğilim gösterdiği gözlemlendi.

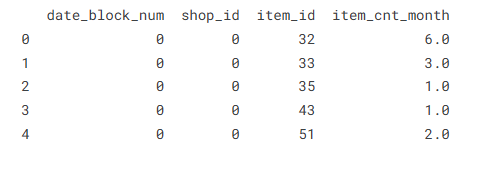
metin, çizgi, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

**5. Özellik Mühendisliği**

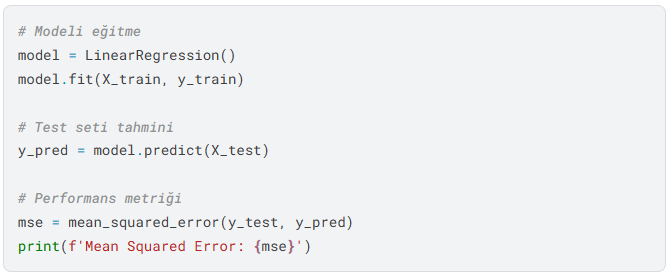
* **Aylık Toplam Satışlar:**  
  Mağaza ve ürün bazında aylık toplam satışlar hesaplandı.
* **Yeni Değişkenler:**  
  item\_cnt\_month isimli hedef değişken oluşturuldu.





**6. Model Eğitimi ve Testi**

1. **Model Seçimi:**  
   Lineer Regresyon modeli kullanıldı.
2. **Veri Bölünmesi:**
   * Eğitim verisi: %80
   * Test verisi: %20
3. **Modelin Eğitilmesi:**  
   Eğitim verisi üzerinde model eğitildi ve test verisi ile değerlendirildi.



Mean Squared Error: 79.69708835181014

**7. Model Performansı**

**Performans Metrikleri:**

**metin, yazı tipi, beyaz, cebir içeren bir resim

Açıklama otomatik olarak oluşturuldu**

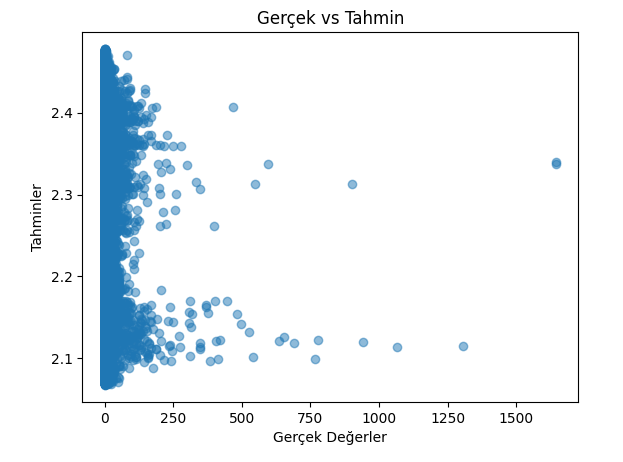
**Analiz:**  
Model, basit bir doğrusal ilişkiyi yakalamakta zorlanmıştır. Verilerin karmaşıklığı ve sezonsallık etkileri doğrusal modellerin performansını sınırlamış olabilir.

**Hata Dağılımı:**

* Hataların büyük bir kısmı düşük frekansta, ancak bazı uç hatalar gözlemlenmiştir.
* Görselleştirme: Gerçek ve tahmin edilen değerler ile hata histogramı analiz edildi.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu



metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, ekran, görüntüleme, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**8. Çıkarımlar ve Gelecekteki Çalışmalar**

1. **Çıkarımlar:**
   * Aylık toplam satış tahmininde, verilerin doğrusal olmayan yapısı model performansını etkiledi.
   * Daha karmaşık modeller (ör. XGBoost, LSTM) ile performans artırılabilir.
2. **Gelecekteki Çalışmalar:**
   * Zaman serisi analizi için ARIMA, Prophet gibi modellerin denenmesi.
   * Sezonsallık ve trendlerin daha iyi yakalanması için veri mühendisliği iyileştirmeleri.

**Ekler**

* **Kod Dosyası:**  
  Tüm veri işleme, model eğitimi ve görselleştirme adımları içeren Python kodları.
* **Kullanılan Kütüphaneler:**
  + Pandas, Numpy
  + Scikit-learn
  + Matplotlib