



**KÜTAHYA DÜMLUPINAR ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ BİLGİSAYAR
MÜHENDİSLİĞİ**

YÜKSEK DÜZEY PROGRAMLAMA PROJE ÖDEVİ

PREDICT FUTURE SALES

ÜMMÜGÜL TÜRK 202213171010

1. Projenin Amacı

Bu projede, "**Predict Future Sales**" veri

seti(<https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales/data>) kullanılarak gelecekteki satışların tahmini yapılmıştır. Amaç, özellikle belirli bir zaman aralığında ürünlerin satış miktarlarını tahmin ederek işletmelerin satış stratejilerini optimize etmelerine yardımcı olmaktır. Tahmin için LSTM (Long Short-Term Memory) modeli kullanılmıştır.

2.Projede Kullanılan Kütüphaneler

pandas: Veri manipülasyonu ve analizleri için kullanılır. Özellikle veri okuma (CSV dosyaları) ve DataFrame yapıları ile veri üzerinde işlemler yapmak için kullanılır.

numpy: Sayısal hesaplamalar için kullanılır. Array manipülasyonu ve matematiksel işlemler için yararlıdır.

matplotlib: Veri görselleştirme için kullanılır. Çeşitli grafikler ve görseller oluşturulabilir.

seaborn: matplotlib'in üzerine kurulu olan ve daha estetik görseller oluşturmak için kullanılan bir kütüphanedir.

warnings: Uyarıları kontrol etmek ve filtrelemek için kullanılır. Özellikle gereksiz uyarıları gizlemek için yararlıdır.

```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

sklearn.preprocessing:

- **LabelEncoder**: Kategorik verileri sayısal verilere dönüştürmek için kullanılır.
- **OneHotEncoder**: Kategorik verileri bir sıcak kodlama formatında dönüştürmek için kullanılır.
- **StandardScaler**: Verileri standartlaştırmak için kullanılır (ortalama 0, standart sapma 1).

scipy.stats: İstatistiksel fonksiyonlar için kullanılır. Aykırı değerlerin tespit edilmesi ve veri seti üzerinde istatistiksel işlemler yapmak için kullanılır.

```

▶ from sklearn.preprocessing import LabelEncoder, OneHotEncoder
  from sklearn.preprocessing import StandardScaler
  from scipy import stats

```

tensorflow.keras:

- **Sequential:** Derin öğrenme modelini katman katman oluşturmak için kullanılır.
- **LSTM:** Uzun Kısa Süreli Bellek ağları (Long Short-Term Memory) ile zaman serisi tahminleri yapmak için kullanılır.
- **Dense:** Tam bağlantılı katmanlar eklemek için kullanılır.

sklearn.metrics:

- **mean_squared_error:** Modelin tahmin hatasını hesaplamak için kullanılır, genellikle regresyon modellerinin doğruluğunu ölçmek için.

sklearn.preprocessing.MinMaxScaler: Verileri [0, 1] aralığına normalize etmek için kullanılır.

```

▶ import numpy as np
  from tensorflow.keras.models import Sequential
  from tensorflow.keras.layers import LSTM, Dense
  from sklearn.preprocessing import MinMaxScaler
  from sklearn.metrics import mean_squared_error

```

3. Veri Setinin İncelenmesi

Projede kullanılan veri seti şu dosyalardan oluşmaktadır:

- **sales_train.csv:** Satış işlemlerini içeren ana veri seti.
- **test.csv:** Test amaçlı kullanılan veri seti.

Veri Setinin Genel Özellikleri


Eğitim veri seti 2935849 satır ve 6 sütundan oluşmaktadır.

Veri sütunları:

- **date:** Satış tarihi.
- **date_block_num:** Ay bazında bir numaralandırma.
- **shop_id:** Mağaza kimliği.
- **item_id:** Ürün kimliği.

- item_price: Ürün fiyatı.
- item_cnt_day: Günlük ürün satış miktarı.


Veri setinin ilk birkaç satırı aşağıda gösterilmiştir:



	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0

İstatistikler

Veri setinin istatistiksel değerleri şu şekildedir:



memory usage: 134.4+ MB

	date_block_num	shop_id	item_id	item_price	item_cnt_day
count	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06
mean	1.456991e+01	3.300173e+01	1.019723e+04	8.908532e+02	1.242641e+00
std	9.422988e+00	1.622697e+01	6.324297e+03	1.729800e+03	2.618834e+00
min	0.000000e+00	0.000000e+00	0.000000e+00	-1.000000e+00	-2.200000e+01
25%	7.000000e+00	2.200000e+01	4.476000e+03	2.490000e+02	1.000000e+00
50%	1.400000e+01	3.100000e+01	9.343000e+03	3.990000e+02	1.000000e+00
75%	2.300000e+01	4.700000e+01	1.568400e+04	9.990000e+02	1.000000e+00
max	3.300000e+01	5.900000e+01	2.216900e+04	3.079800e+05	2.169000e+03

Eksik Değerler

Veri setinde eksik değer bulunmamaktadır:



	0
date	0
date_block_num	0
shop_id	0
item_id	0
item_price	0
item_cnt_day	0

dtype: int64

4. Veri İşleme

Veri seti üzerinde aşağıdaki işlemler gerçekleştirilmiştir:

1. **Tarih Formatı Dönüşümü:** date sütunu, **datetime** formatına dönüştürülmüştür.

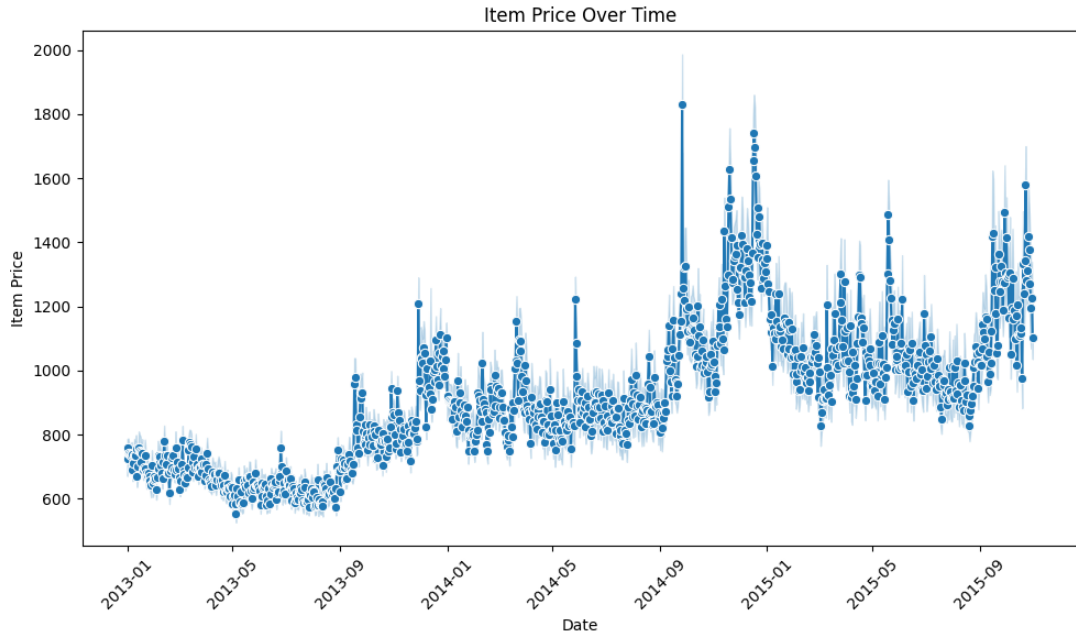


	date	date_block_num	shop_id	item_id	item_price	item_cnt_day	Year	Month	Day	Day_of_Week
0	2013-01-02	0	59	22154	999.00	1.0	2013	1	2	Wednesday
1	2013-01-03	0	25	2552	899.00	1.0	2013	1	3	Thursday
2	2013-01-05	0	25	2552	899.00	-1.0	2013	1	5	Saturday
3	2013-01-06	0	25	2554	1709.05	1.0	2013	1	6	Sunday
4	2013-01-15	0	25	2555	1099.00	1.0	2013	1	15	Tuesday

2. **Ek Özellikler Eklenmesi:** Yıl, ay, gün ve haftanın günü gibi bilgiler eklenmiştir.
3. **Negatif Değerlerin Çıkarılması:** Ürün satış adetlerinde (item_cnt_day) negatif değerler tespit edilip çıkarılmıştır.
4. **Aykırı Değerlerin Çıkarılması:** Z-skoru yöntemiyle aykırı değerler temizlenmiştir. Bu işlem sonrası veri seti 2855775 satıra düşmüştür.

5. Veri Görselleştirme

Veri setindeki bazı sütunların analizine yönelik görselleştirmeler yapılmıştır. Örneğin, ürün fiyatlarının zaman içerisindeki değişimi şu şekilde görselleştirilmiştir:



Grafik Açıklamaları:

- **X Eksen:** Satış tarihleri (date).
- **Y Eksen:** Ürün fiyatları (item_price).

- Zaman içerisinde ürün fiyatlarındaki artış ve düşüş trendleri belirgin bir şekilde görülmektedir.

6. Modelleme

Proje kapsamında, zaman serisi verilerine yönelik bir LSTM (Long Short-Term Memory) modeli oluşturulmuştur.

Model Özellikleri:

- **Giriş Özelliği:** Son 30 günlük satış adetleri kullanılmıştır.
- **Hedef Özelliği:** Bir sonraki günün satış adedinin tahmini.

Modelin Yapısı:

- **Katmanlar:**
 - 1 adet LSTM katmanı (50 nöronlu).
 - 1 adet Dense (Tam Bağlantılı) katman.
- **Optimizasyon:** Adam optimizasyon algoritması.
- **Kayıp Fonksiyonu:** Ortalama Kare Hatası (Mean Squared Error).

Modelin Eğitilmesi:

- Model, 5 epoch boyunca eğitim aldı. Her epoch, modelin veriler üzerinde yaptığı iyileştirmeleri gösterir. Eğitim sırasında batch_size=32 parametresi kullanıldı, bu da her seferinde 32 örneğin model tarafından işlenmesini sağlar.
- Eğitim sırasında her epoch sonunda doğrulama (validation) kaybı da hesaplanmış, bu sayede modelin aşırı öğrenme (overfitting) yapıp yapmadığı kontrol edilmiştir.
- Eğitim ve doğrulama hataları şu şekildedir:

```

⇒ Epoch 1/5
71394/71394 ————— 1255s 18ms/step - loss: 0.0102 - val_loss: 0.0091
Epoch 2/5
71394/71394 ————— 1186s 16ms/step - loss: 0.0101 - val_loss: 0.0091
Epoch 3/5
71394/71394 ————— 1158s 16ms/step - loss: 0.0102 - val_loss: 0.0092
Epoch 4/5
71394/71394 ————— 1195s 17ms/step - loss: 0.0101 - val_loss: 0.0091
Epoch 5/5
71394/71394 ————— 1179s 17ms/step - loss: 0.0102 - val_loss: 0.0091

```

Model Değerlendirme:

Modelin doğruluğu **Root Mean Squared Error (RMSE)** ile ölçülmüştür. RMSE değeri, tahmin hatasının karekökünü alarak modelin ne kadar doğru tahminlerde bulunduğunu gösterir.

- **Validation Loss:** 0.0091
- **RMSR (Kök Ortalama Kare Hatası):** 0.10

⇌ 17849/17849 ————— 111s 6ms/step
17849/17849 ————— 110s 6ms/step - loss: 0.0089

Bu sonuçlar, modelin iyi bir performans gösterdiğini ve satışları doğru tahmin etme kapasitesine sahip olduğunu göstermektedir. Modelin düşük bir kayıp ve düşük RMSR değeri, yüksek doğrulukla tahmin yaptığını gösterir.

Bu eğitim aşaması, modelin zaman serisi verilerini analiz etme ve satış tahminlerini doğru bir şekilde yapma yeteneğini kazandığı aşamadır.

7. Sonuç ve Değerlendirme

Bu proje kapsamında:

1. Veri seti temizlenmiş ve öznitelikler eklenmiştir.
2. Zaman serisi tahmini için LSTM modeli oluşturulmuş ve başarıyla eğitilmiştir.
3. Model, düşük bir doğrulama hatası (Validation Loss) ile gelecekteki satışları tahmin etmek için kullanılabilir bir yapı sunmuştur.