**Unilever Verileriyle Net Maliyet Analizi**

**Net Cost Analysis With Unilever Data**

**Yazar Adları:**

Muhammet Mustafa Uçar - 202802020,

Osman Tonbul - 202804012,

Muhammet Müslüm Ökkeş Kazıcı - 202803017

**Özet:**

Projemiz, Unilever’in 8 aylık süre zarfında Food (Gıda), HC (Ev Bakımı) ve PC (Kişisel Bakım) kategorilerindeki satış verilerini temel alarak, çeşitli düzeltmelerle (örneğin KDV, Fatura Altı gibi) ve ürün gruplarına özgü değişkenlikleri içeren net maliyeti tahmin etmeyi amaçlamaktadır. Bu tahminler, satış verilerindeki çeşitli faktörleri dikkate alarak, kategorilere ve ürün gruplarına özgü olarak düzenlenmiş ve düzeltilmiş maliyetleri içermektedir. Bu analiz, işletme için maliyet etkinliği ve karlılık açısından önemli bir bilgi sağlayabilir.

**Anahtar Kelimeler:**

Kategoriler, Net Maliyet, Regresyon, Unilever, Veri Seti

1. **Giriş:**

Makalemiz, Unilever’in 8 aylık ticari faaliyet dönemine odaklanarak, Food (Gıda), HC (Ev Bakımı) ve PC (Kişisel Bakım) kategorilerindeki satış verilerini temel alarak Regresyon Modeli ile net maliyeti tahmin etmeyi amaçlamaktadır [1]. Bu analiz, şirketin maliyet etkinliği ve karlılık performansını değerlendirmek adına kritik bir rol oynamaktadır. Unilever gibi büyük ölçekli ve çok kategorili bir şirketin maliyet dinamiklerini anlamak, stratejik planlamada ve kaynak tahsisinde önemli bilgiler sağlayabilir.

Makalemiz, özellikle gıda, ev bakımı ve kişisel bakım kategorilerindeki satış verilerini kullanarak net maliyet tahminini ele alacak ve bu süreçte çeşitli düzeltmeleri (örneğin KDV, Fatura Altı gibi) ve ürün gruplarına özgü değişkenlikleri hesaba katacaktır. Bu, şirketin marketlerden bakkallara kadar sattığı ürünlerin net maliyeti hakkında genel bir öngörüde bulunup bu öngörüde özgün bir bakış sunmayı amaçlamaktadır.

Analizimizin temel odak noktası, Unilever’in farklı kategorilerindeki ürün gruplarına özgü maliyet değişkenliklerini anlamak ve bu bilgileri şirketin hedefleri doğrultusunda kullanmaktır. Bu bağlamda, giriş bölümümüzde analizimizin önemini vurgulayacak ve okuyucuya makalenin genel yapısı hakkında bir ön izlenim sunacağız.

1. **Verilerin Anlatımı:**

Aşağıda açıklayacak olduğumuz veri setlerinin kaynağı Unilever Şirketi’nin satış verileridir [2].

* **KOD (Ürün Kodu/Referans Numarası):** Her bir ürünün benzersiz içsel kodu veya referans numarasıdır. Bu, ürünlerin takibi ve tanımlanması için kullanılır.
* **BARKOD (Barkod Numarası):** Ürünün satış noktalarında tanımlanmasına ve takibine olanak tanıyan benzersiz barkod numarasıdır. Bu, işlemleri otomatikleştirmek ve hızlandırmak için kullanılır.
* **MAMUL (Ürün Adı/Tanımı):** Satılan her bir ürünün adı veya tanımını içerir. Bu, müşteriler ve personel tarafından ürünü tanımlamak için kullanılır.
* **K.İÇİ (Birim İçeriği):** Birim miktarı satılan her ürünün koli içinin adet sayısını belirtir. Örneğin, 1 adet veya 1 paket gibi.
* **GR (Ağırlık):** Satılan her bir ürünün ağırlığını gram cinsinden belirtir. Bu, envanter yönetimi ve lojistik süreçlerde önemli bir faktördür.
* **LİSTE (Liste Fiyatı):** Ürünün liste fiyatını ifade eder. Bu, satış fiyatının belirlenmesinde temel bir referans noktasıdır.
* **FATURA ALTI % (Fatura Altı İndirim Oranı):** Fatura altında uygulanan indirim oranını belirtir. Bu, müşterilere özel indirimlerin hesaplanmasında kullanılır.
* **DÖNEM SONU % (Dönem Sonu İndirim Oranı):** Müşterilerin belirli kriterleri yerine getirmesi sonucu şirketin ekstra vermiş olduğu indirim oranını belirler.
* **KDV (Katma Değer Vergisi Oranı):** Ürün üzerinden alınan Katma Değer Vergisi oranını belirtir. Vergilendirme politikalarını ve maliyet hesaplamalarını etkiler.
* **NET MALİYET:** Ürünün net maliyetini temsil eder. Bu değer, maliyet düzeltmeleri ve vergilerle düzeltilmiş toplam maliyeti içerir.
* **Ö.RAF (Önerilen Raf Fiyatı):** Ürünün önerilen raf fiyatıdır. Bir satış yapılacak kuruma (market / bakkal gibi) ürüne koyması gereken öneri fiyatını belirtir.
* **KATEGORİ:** Ürünün ait olduğu kategoriyi belirtir. Örneğin, “FOOD” kategorisi gıda ürünlerini, “HC” ev bakımı ürünlerini ve “PC” kişisel bakım ürünlerini içerebilir.
* **AY (Satışın Gerçekleştiği Ay):** Satışın gerçekleştiği tarih veya ayı belirtir. Bu, satış verilerini zaman içinde izlemek ve analiz etmek için kullanılır.

Verilerin istatistiksel analizini gösteren grafikler [6] :

Şekil a) Ay’a Göre Net Maliyet [7] Şekil b) Kategoriye Göre Net Maliyet Dağılımı [8]

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

Açıklama otomatik olarak oluşturuldu metin, diyagram, daire, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

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

Açıklama otomatik olarak oluşturuldumetin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturulduŞekil c) Net Maliyet Dağılımı [9] Şekil d) Kategorilerin Aylara Göre Net Maliyetleri [10]

**Veri Setinin İstatistiksel Gösterimi**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Adet | Ortalama | Standart sapma | Minimum | 1.çeyrek | Medyan | 3.çeyrek | Maksimum |
| MAMUL | 3820.00 | 8.13 | 5.20 | 1.00 | 3.00 | 9.00 | 12.00 | 17.00 |
| K. İÇİ | 3820.00 | 31.81 | 42.07 | 1.00 | 12.00 | 16.00 | 24.00 | 288.00 |
| GR | 3820.00 | 495.06 | 860.76 | 1.90 | 68.75 | 165.00 | 500.00 | 7500.00 |
| LİSTE | 3820.00 | 54.25 | 32.86 | 2.00 | 29.65 | 54.50 | 69.34 | 320.86 |

Bu istatistiksel gösterim, her bir kategori için merkezi eğilim, dağılım ve ekstrem değerleri özetler. Böylece veri setinin genel özellikleri hakkında bize bilgi sağlar. Örneğin, “GR” kategorisi yüksek standart sapma ve büyük maksimum değeri ile geniş bir dağılıma sahiptir, “LİSTE” kategorisi ise daha düşük standart sapma ve maksimum değerle daha sınırlı bir dağılıma işaret etmektedir.



**Formun Üstü**

**Önerilen Yöntemler:**

* 1. **Regresyon**

Regresyon analizi, bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modelleme amacı taşır. Regresyon modelleri, veri setindeki değişkenler arasındaki ilişkiyi anlamak ve gelecekteki değerleri tahmin etmek için kullanılır.

Örnek regresyon türleri:

* Lineer Regression
* Polynomial Regression
* Decision Tree Regression

Bu projede kullanılan regresyon türü lineer regresyondur. Lineer regresyon, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi ifade eden doğrusal bir denklem kurarak, bağımlı değişkenin değerini bağımsız değişkenlerin değerlerine göre tahmin etmeye çalışır.

Projemizde kullanılan veri seti 2023 yılında üretilen mamullerin maliyetlerini içermektedir. Veri setindeki bağımlı değişken NET MALİYET, bağımsız değişkenler ise KATEGORİ, AY, MAMUL, K. İÇİ, GR, LİSTE’dir.

* 1. **Veri Ön İşleme**
* ColumnTransformer (Sütun Dönüştürücü): Sütun bazlı (column-wise) dönüştürmeleri kolaylaştırmak için kullanılan bir araçtır. Verisetinde farklı özellik türleri (sayısal ve kategorik özellikler) üzerinde ayrı ayrı dönüştürme işlemleri yapılmasına olanak sağlar.
* Projemizde sütun dönüştürücü olarak OneHotEncoder kullanılmıştır [3].
  + OneHotEncoder, Kategorik özellikleri sayısal özelliklere dönüştürmek için kullanılan bir araçtır.
* Projemizde sayısal özelliklerin ölçeklendirilmesi için StandardScaler kullanılmıştır.
  + StandardScaler: Sayısal özelliklerin ölçeklendirilmesi için kullanılan bir araçtır. Veri setindeki sayısal özelliklerin ortalaması 0, standart sapması 1 olacak şekilde ölçeklendirilmesini sağlar.
* Projemizde aykırı değerleri tespit etmek için Z-score yöntemi kullanılmıştır.

Şekil e) Kategori bazında net maliyet kutu grafiği

diyagram, metin, ekran görüntüsü, teknik çizim içeren bir resim

Açıklama otomatik olarak oluşturuldu

* + Outlires için kullanılan Z-score yöntemi:
  + Z-score yöntemi: Bu yöntem, veri setindeki değerlerin standart sapmalarını kullanarak aykırı değerleri tespit etmek için kullanılır.
  + Z-score, bir değerin ortalamadan kaç standart sapma uzaklıkta olduğunu gösteren bir ölçüdür ve şu formülle hesaplanır:
  + X: Değer, mean: Ortalama, std: Standart sapma
  + Aykırı değerler, genellikle belirli bir Z-score sınırının üzerinde veya altında olan değerler olarak kabul edilir. Bu sınır genellikle ± 3 olarak belirlenir.
  + Aykırı değerleri tespit etmek için kullanılan eşik değerler:

Alt eşik değeri: mean−3×std

Üst eşik değeri: mean+3×std

Projemizde aykırı değerlerin alt ve üst eşik değerleri hesaplanarak aykırı değerler tespit edilmiş ve aykırı değerler veri setin den çıkarılmıştır.

* 1. **Veri Görselleştirme**

Bu projede veri görselleştirme için kullanılan kütüphane ve fonksiyonlar:

* + - Veri görselleştirme için Seaborn kütüphanesi kullanıldı [4]. Aşağıda da bu kütüphanenin fonksiyonları yer alıyor:
  + Barplot: Kategorik özelliklerin sayısal özelliklerle ilişkisini görselleştirmek için kullanılan bir çubuk fonksiyondur.
  + Histplot: Sayısal özelliklerin dağılımını görselleştirmek için kullanılan bir fonksiyondur.
  + Boxplot: Sayısal özelliklerin yüzdelik dilimlerini görselleştirmek için kullanılan bir fonksiyondur.
  + Scatterplot: Sayısal özelliklerin dağılımını görselleştirmek için kullanılan bir fonksiyondur.
  + Lineplot: Sayısal verilerin bir eksen üzerinde çizgilerle temsil edildiği bir görselleştirme fonksiyonudur.
    - Veri görselleştirme için kullanılan bir diğer kütüphane ise Matplotlib’dir [5] :
      * Pie Chart: Kategorik özelliklerin dağılımını görselleştirmek için kullanılan bir grafiktir.
  1. **Kaba Kod(Pseudo Code)**

1.Adım: Kütüphaneleri içe aktar

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

import seaborn as sns

2.Adım: Veri setini içe aktar

dataset = pd.read\_excel(“vm\_dataset.xlsx”)

3.Adım: Veri seç ve temizle

* + Veri setinin kategorik ve sayısal sütunlarını belirleme

categorical\_columns = [‘KATEGORİ’, ‘AY’, ‘MAMUL’]

numeric\_columns = [‘K.İÇİ’, ‘GR’, ‘LİSTE’]

* + Z-score yöntemi ile aykırı değerleri tespit et

from scipy.stats import zscore

def detect\_outliers(data, columns):

outliers = pd.DataFrame()

for column in columns:

z\_scores = zscore(data[column])

abs\_z\_scores = np.abs(z\_scores)

trashold = 3

filtered\_entries = (abs\_z\_scores < trashold)

outliers = pd.concat([outliers, data[~filtered\_entries]])

return outliers

outliers = detect\_outliers(data\_selected, numeric\_columns)

* + Aykırı değerleri veri setinden çıkar

data\_selected = data\_selected.drop(outliers.index, axis=0)

4.Adım: Modelin oluşturulması

* + Sayısal ve kategorik sütunları işlemek için sütun dönüştürücü oluştur

preprocessor = ColumnTransformer( transformers=[ (‘num’, StandardScaler(), numeric\_columns), (‘cat’, OneHotEncoder(), categorical\_columns) ])

* + Pipeline’ da LinearRegression modelini kullanarak bir model oluştur

model = Pipeline(steps=[(‘preprocessor’, preprocessor),(‘regressor’, LinearRegression())])

5.Adım: Veri setinin bağımlı ve bağımsız değişkenlerini belirle

X = data\_selected.drop(‘NET MALİYET’, axis=1)

y = data\_selected[‘NET MALİYET’]

6.Adım: Verileri eğitim ve test kümelerine böl

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

7.Adım: Modeli eğitim verileri ile eğit

model.fit(X\_train, y\_train)

8.Adım: Model ile test verileri üzerinde tahmin yap

y\_pred = model.predict(X\_test)

9.Adım: MSE, MAE, RMSE ve R2 değerlerini hesapla

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.metrics import mean\_absolute\_error

def model\_performance(y\_test, y\_pred):

mse = mean\_squared\_error(y\_test, y\_pred)

mae = mean\_absolute\_error(y\_test, y\_pred)

rmse = np.sqrt(mse)

r2 = r2\_score(y\_test, y\_pred)

model\_performance(y\_test, y\_pred)

10.Adım: Gerçek ve tahmin edilen değerleri görselleştir

sns.scatterplot(x=y\_test, y=y\_pred, alpha=0.9, s=50)

sns.lineplot(x=y\_test, y=y\_test, color=‘red’)

1. **Deneysel Sonuçlar ve Tartışma:**

Şekil f) Tahmin ve Gerçek Değerlerin Karşılaştırılması [11].

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

Açıklama otomatik olarak oluşturulduEğitim Seti Sayısı: 2950**

**Test Seti Sayısı: 738**

Deneysel sonuçlar ve tartışma kısmında projemizde kullanılan regresyon modelinin performansı değerlendirilmiştir.

Projemizde kullanılan regresyon modelinin performansı değerlendirilirken kullanılan metrikler:

* MSE (Mean Squared Error): Hata Kareler Ortalaması (0’a ne kadar yakınsa o kadar iyi)
* RMSE (Root Mean Squared Error): Hata Kareler Ortalamasının Karekökü (0’a ne kadar yakınsa o kadar iyi)
* R2 (R Squared): Determinasyon katsayısı (1’e ne kadar yakınsa o kadar iyi)
* MAE (Mean Absolute Error): Ortalama Mutlak Hata (0’a ne kadar yakınsa o kadar iyi)

Projemizde kullanılan regresyon modelinin performansı değerlendirilirken kullanılan metriklerin değerleri:

|  |  |  |
| --- | --- | --- |
| Değer\Yöntem | Z-score veri ön işlemesi uygulandığında bulunan değerler | IQR veri ön işlemesi uygulandığında bulunan değerler |
| MSE | 32.6917 | 37.8279 |
| RMSE | 5.7176 | 6.1504 |
| R2 | 0.9470 | 0.9194 |
| MAE | 4.1301 | 4.603 |

Z-score uygulandığında modelin daha iyi performans gösterdiği görülmektedir. Düşük MSE, RMSE, yüksek R2 ve düşük MAE değerleri, modelin tahminlerinin genel olarak daha doğru olduğunu ve veriye daha iyi uyum sağladığını gösterir.

1. **Sonuç**

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

Açıklama otomatik olarak oluşturuldu**Şekil g) Tahmin ve Gerçek Farkı ile Hata Dağılımı [9].

Yaptığımız bu çalışma ile Unilever’in 8 aylık ticari faaliyet dönemine odaklanarak Food (Gıda), HC (Ev Bakımı) ve PC (Kişisel Bakım) kategorilerindeki satış verilerini temel alarak net maliyet tahmini yaptık. Bu bağlamda şirketin maliyet etkinliği ile karlılık performansını değerlendirmeyi amaçladık. Elde edilen sonuçlara baktığımızda genelde beklediğimiz sonuçları aldık. Elde ettiğimiz bu bilgiler bize şirketin mali yönetimi, stratejik planlaması ve kaynak tahsisi konularında önemli içgörüler sunmaktadır.

Projemiz, Unilever şirketinin büyüklüğü ve çok kategorili yapısı göz önünde bulundurulduğunda, Unilever şirketine stratejik yönetimde daha bilinçli kararlar alma konusunda katkıda bulunabilir. Bu bağlamda yaptığımız bu çalışmalar, şirketin sürdürülebilir büyüme ve rekabet avantajı elde etme çabalarına rehberlik edebilir.

Sonuçları grafik üzerinden değerlendirecek olursak :

Yukarıda yer alan grafik bize hata dağılımını göstermektedir. Grafikte görüldüğü üzere hata dağılımı normal dağılıma yakın bir dağılım göstermektedir. Bu da modelimizin iyi bir performans sergilediğinin yanı sıra modelimizin tahminlerinin de gerçek değerlere yakın olduğunu göstermektedir. Sonuç olarak baktığımızda, projemizde kullanılan regresyon modelinin performansı değerlendirilmiş ve modelimizin iyi bir model olduğu yorumlanmıştır.

**Yazarların Katkıları:**

Muhammet Mustafa Uçar (%40) :

* Veri Setinin Ön İşlemesi,
* Regresyon Modelinin Eğitilmesi,
* Model Performansından Sonuçların Çıkarılıp Excel Dosyasına Kaydedilmesi,
* Raporlama

Osman Tonbul (%30) :

* Veri Setinin Alınması,
* Veri Setinin İncelenmesi,
* Raporlama

Muhammet Müslüm Ökkeş Kazıcı (%30) :

* Veri Görselleştirme,
* Veri Yorumlama,
* Model Değerlendirilmesi,
* Raporlama

**Kaynaklar:**

[1] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), 140-147.

[2] Dentoni, D., & Veldhuizen, M. (2012). Building capabilities for multi-stakeholder interactions at global and local levels: the case of Unilever. International Food and Agribusiness Management Review, 15(1030-2016-82857), 95-106.

[3] Lv, Z., Ding, H., Wang, L., & Zou, Q. (2021). A convolutional neural network using dinucleotide one-hot encoder for identifying DNA N6-methyladenine sites in the rice genome. Neurocomputing, 422, 214-221.

[4] Visualization in Seaborn for Data Science: Create plots using single line of code in Python. (2023). (n.p.): Partha Mishra.

[5] Tosi, S. (2009). Matplotlib for Python developers. Packt Publishing Ltd.

[6] Geleßus, D., & Leuschel, M. (2020, May). ProB and Jupyter for logic, set theory, theoretical computer science and formal methods. In International Conference on Rigorous State-Based Methods (pp. 248-254). Cham: Springer International Publishing.

[7] Isihara, P. (2021). A College Teacher’s Introduction to Jupyter Notebooks. International Journal For Technology In Mathematics Education, 28(4).

[8] Bernstein, M. N., Gladstein, A., Latt, K. Z., Clough, E., Busby, B., & Dillman, A. (2020). Jupyter notebook-based tools for building structured datasets from the Sequence Read Archive. F1000Research, 9.

[9] Bilheux, J. C., Bilheux, H., Lin, J., Lumsden, I., & Zhang, Y. (2019). Neutron imaging analysis using jupyter Python notebook. Journal of Physics Communications, 3(8), 083001.

[10] Larson–Hall, J. (2017). Moving beyond the bar plot and the line graph to create informative and attractive graphics 1. The Modern Language Journal, 101(1), 244-270.

[11] Weiss, C. J. (2022). Visualizing protein big data using Python and Jupyter notebooks. Biochemistry and Molecular Biology Education, 50(5), 431-436.