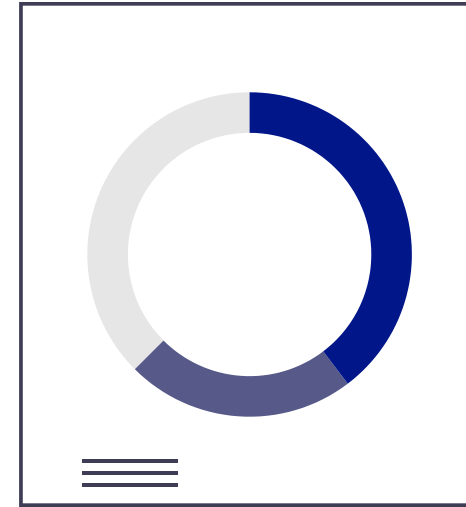
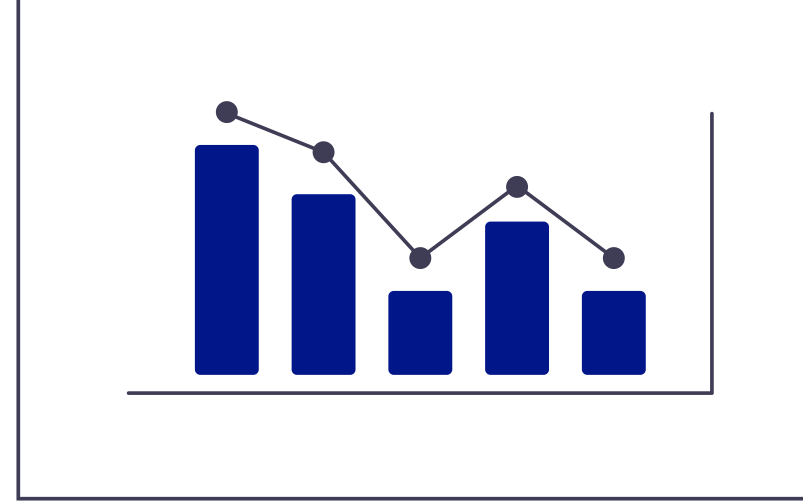


# Patika & EnerjiSA Veri Bilimi ve Analitiği Bootcamp Bitirme Sunumu Grup-2

- Fatma Nur USLUER
- Furkan KILINÇ
- Gülşah UĞUR
- Mustafacan GENÇGÜL
- Uğur Selim ÖZEN



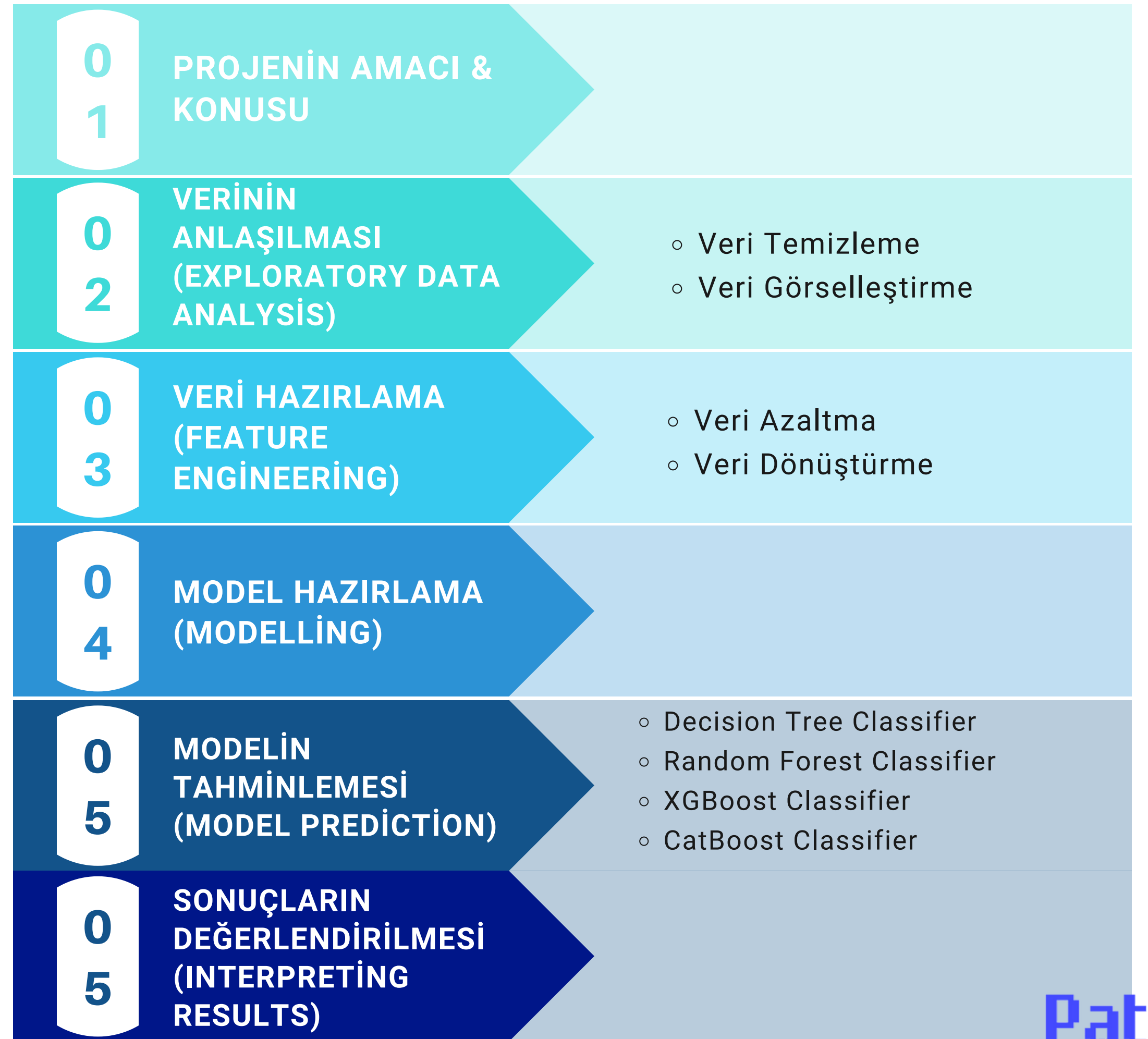
# Kullanılan Teknolojiler



CatBoost



# Neler Anlatacağız?



Grup-2

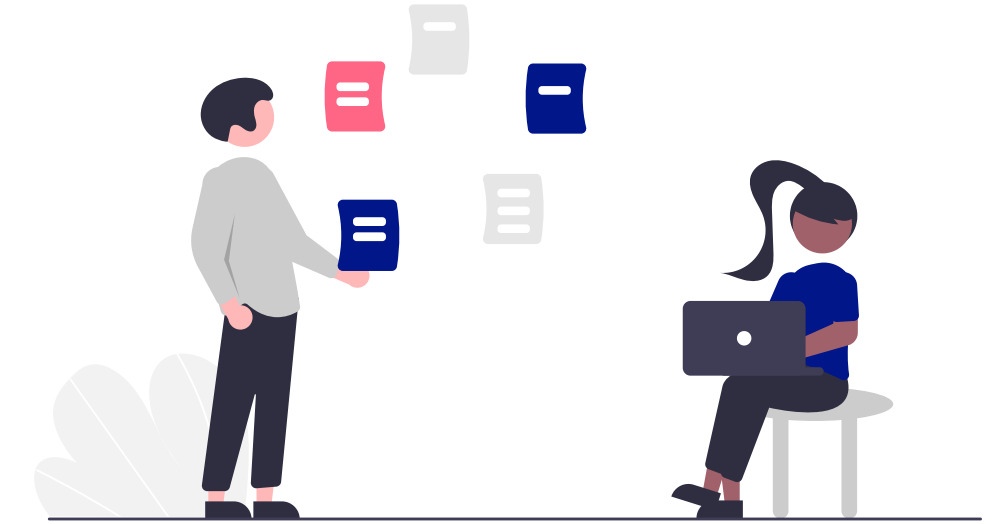
# Projenin Amacı ve Konusu



**Kayıp Kaçak Nedir?**



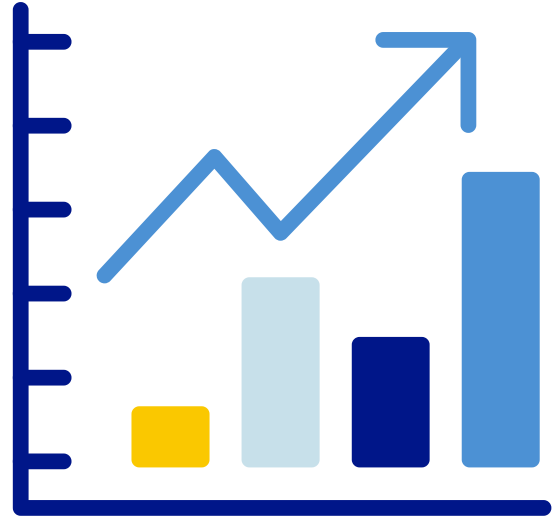
**Kayıp Kaçak  
Yöntemleri Nelerdir?**



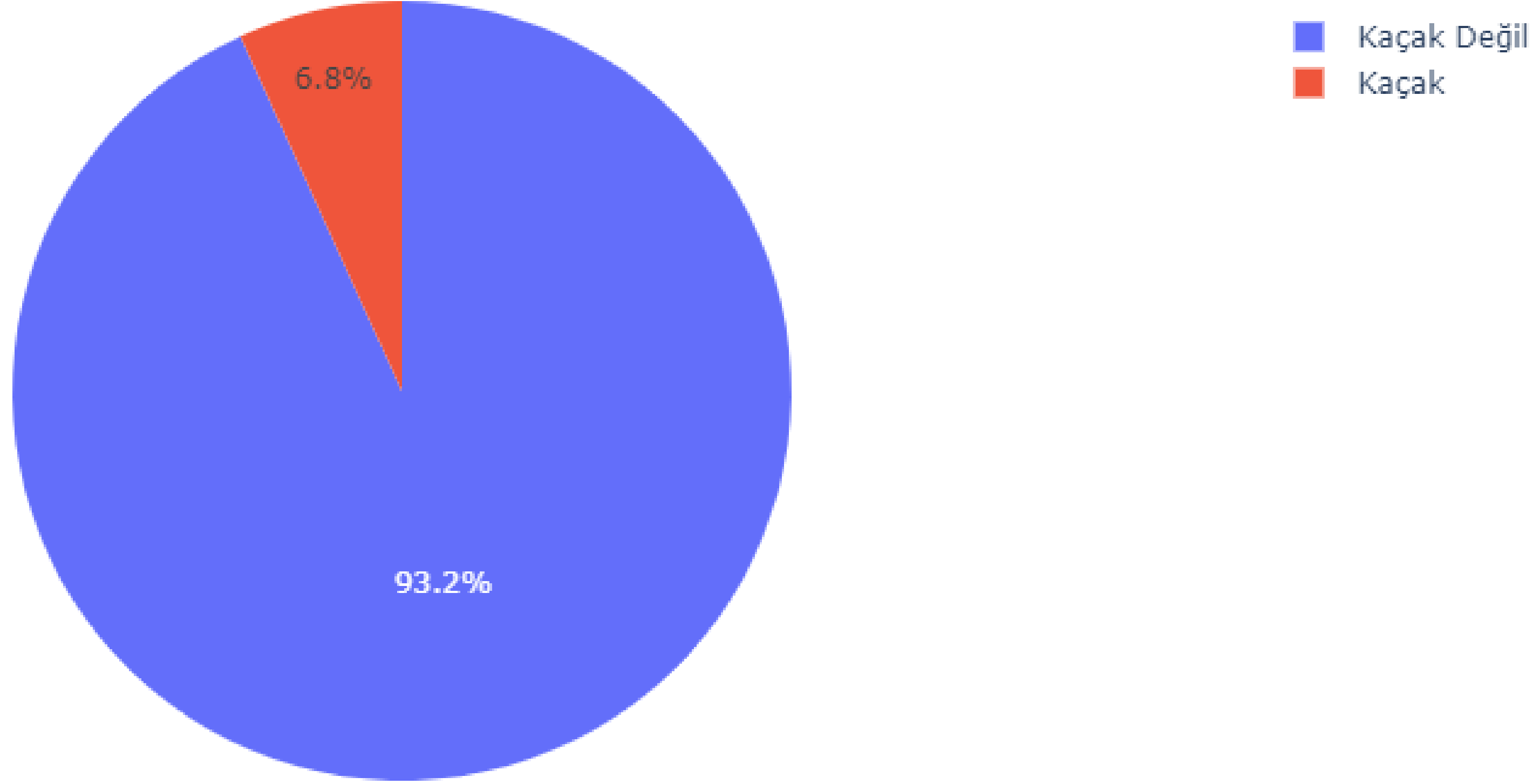
**Biz Ne Yaptık?**

# Verinin Anlaşılması

- Veri: **15.000** satır ve **50** sütun
- Hedef değişken(target): **NK\_FLAG**
- Tüketim ve Demand toplam **36 sütun (%72)**
- 24 ayın ortalama Tüketim Değeri: **200 kWh**
- Eksik veri tespiti
- Sayısal verilen istatistiksel çıkarımı
- Sayısal sütunların birbiri ile olan korelasyonu
- Verilerin görselleştirilmesi

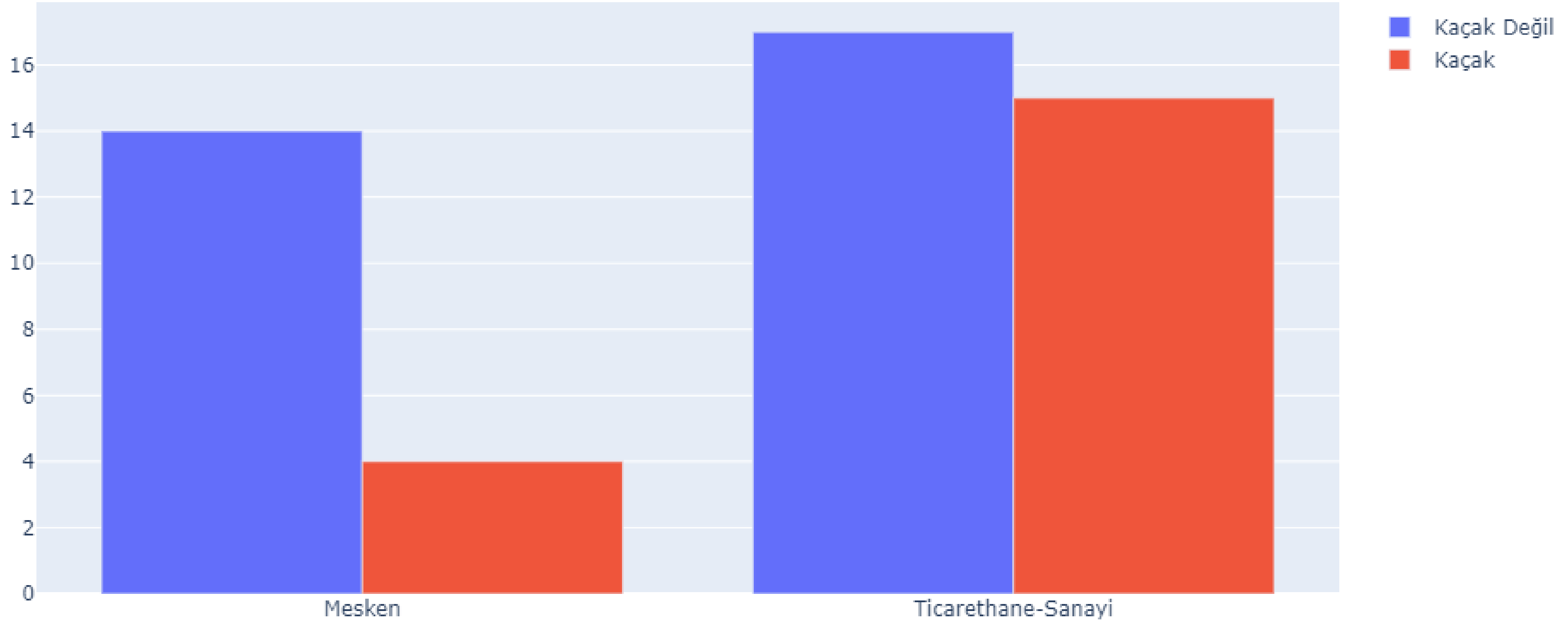


## Toplam Kullanıcıların Kaçak Kullanım Dağılımı

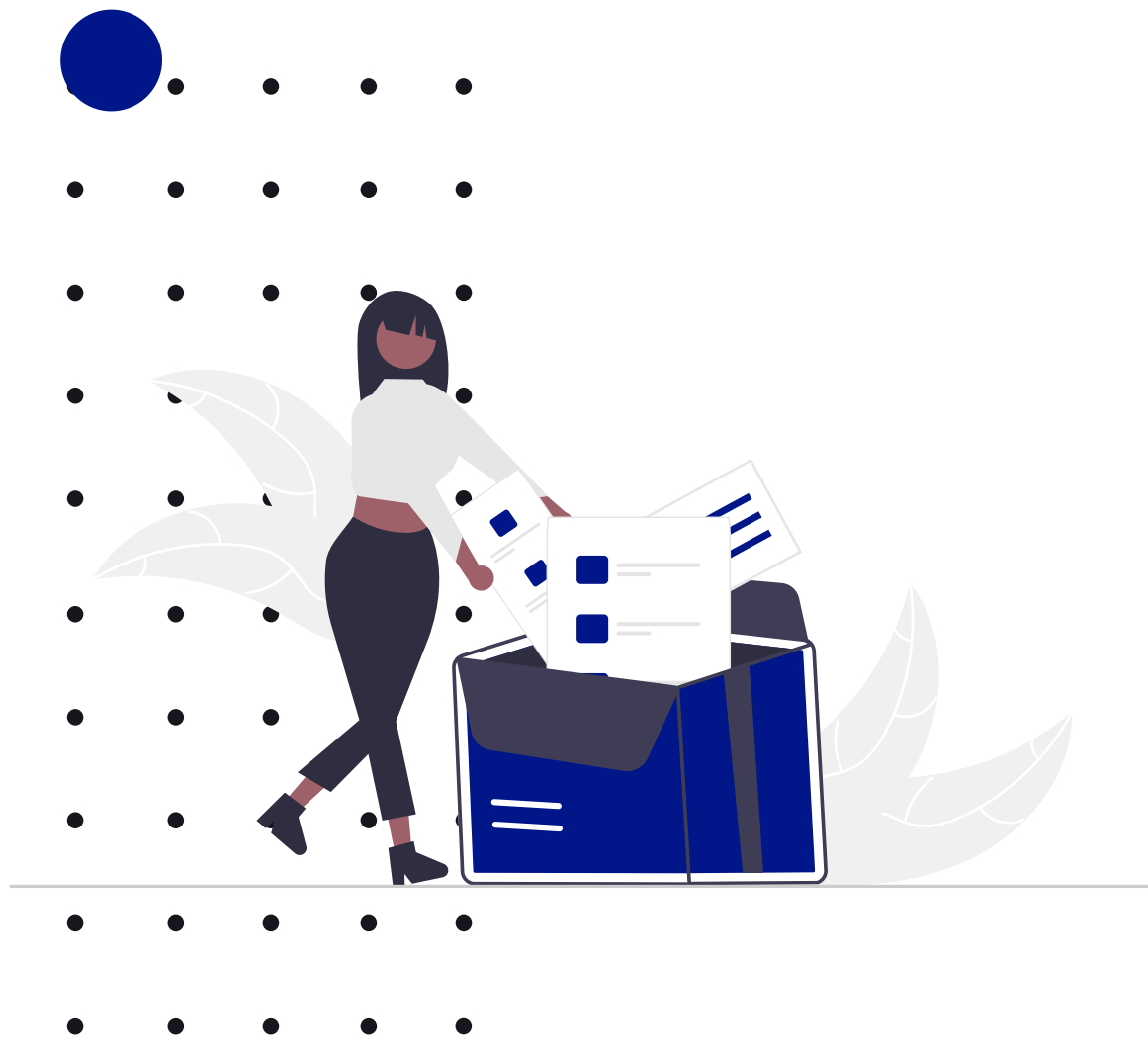


Grup-2

## Eksik Verilerde Kaçakların Tesisat Tipine Göre Dağılımı



Grup-2



# Veri Hazırlama (Feature Engineering)

- Eksik verilerin doldurulması (0, 'EKSİK', yüksek bir sayı)
- Yeni kolonlar oluşturulması
- Mevcut/Oluşturulan kolonların istatistiksel çıkarımlar doğrultusunda gruplanması

## Gruplanan Mevcut Kolonlar



SAYAC\_MODEL\_group

SAYAC\_YAS\_group

## Yeni Oluşturulan Kolonlar



SAYAC\_YAS

ABONELIK\_SURESI

TUKETIM\_std

DEMAND\_std

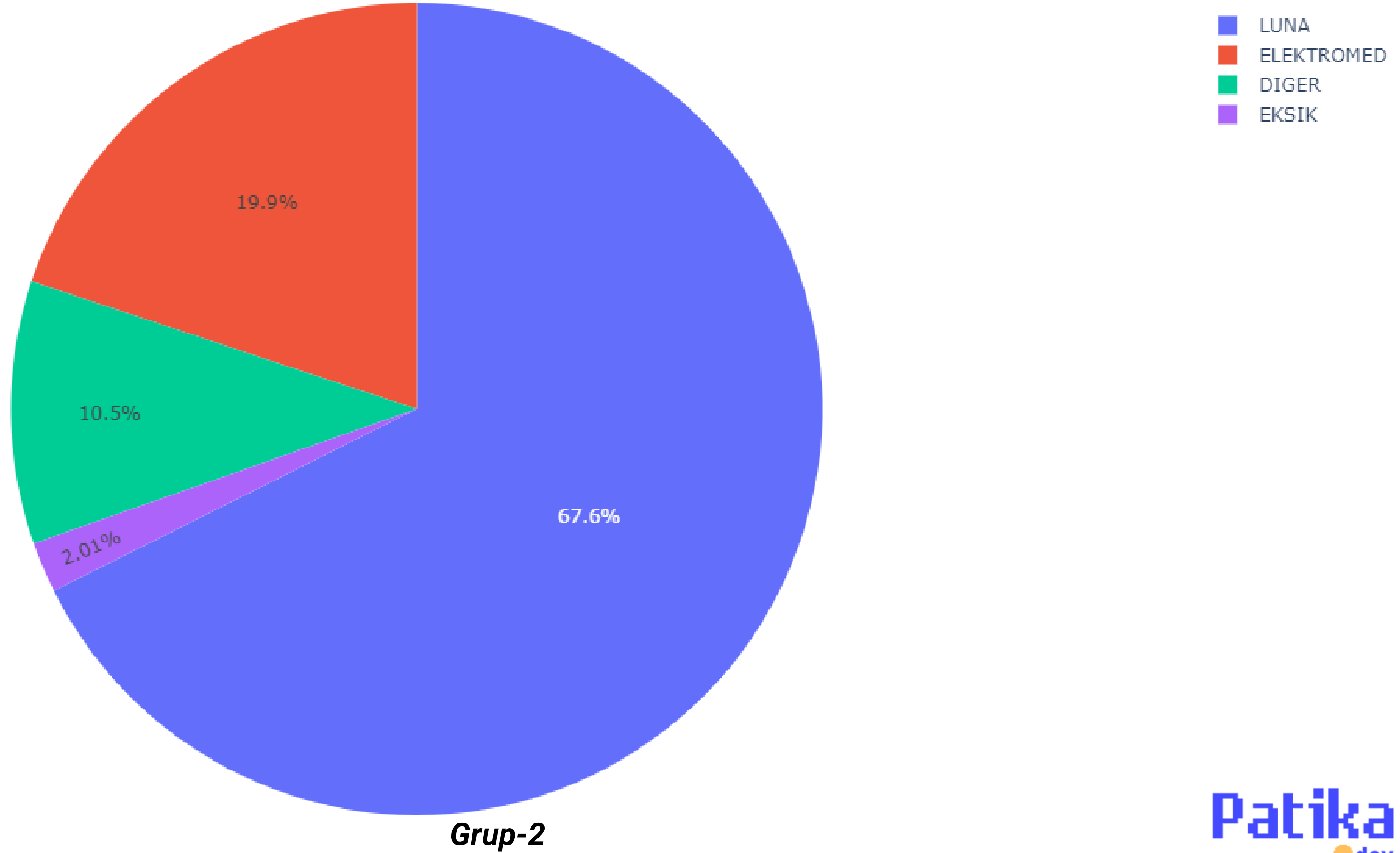
Grup-2



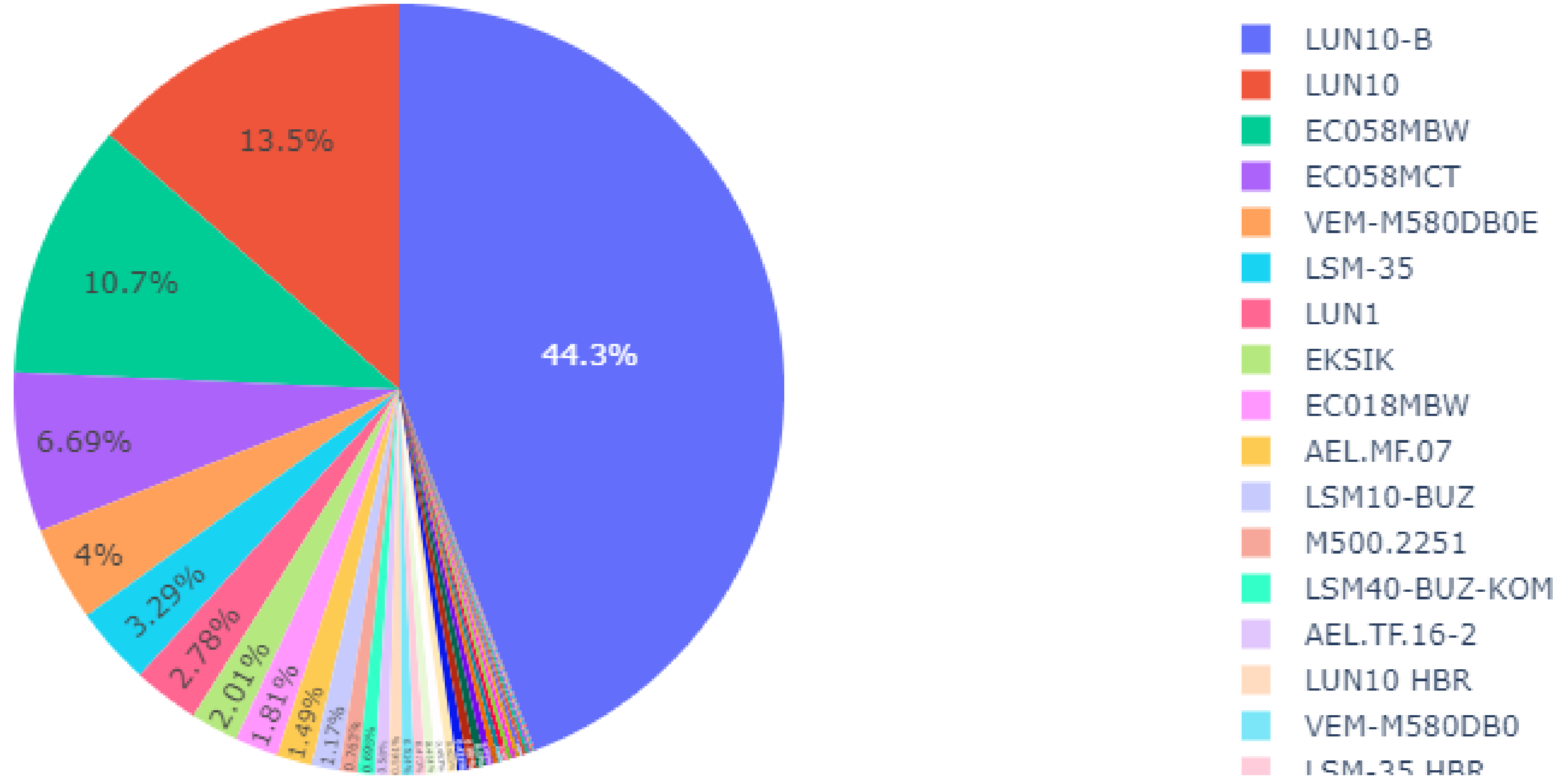
## Sayaç Yaşına Göre Kaçak Oranları

	SAYAC_YAS	KACAK_SAYISI	TOPLAM_ABONE	KACAK_ORANI
0	0-3 YAŞ	400.0	1611.0	24.829299
2	4+ YAŞ	345.0	11039.0	3.125283
1	4 YAŞ	256.0	2300.0	11.130435
3	EKSIK	19.0	50.0	38.000000

# Tespit Edilen Kaçakların Sayaç Markalarına Göre Dağılımı

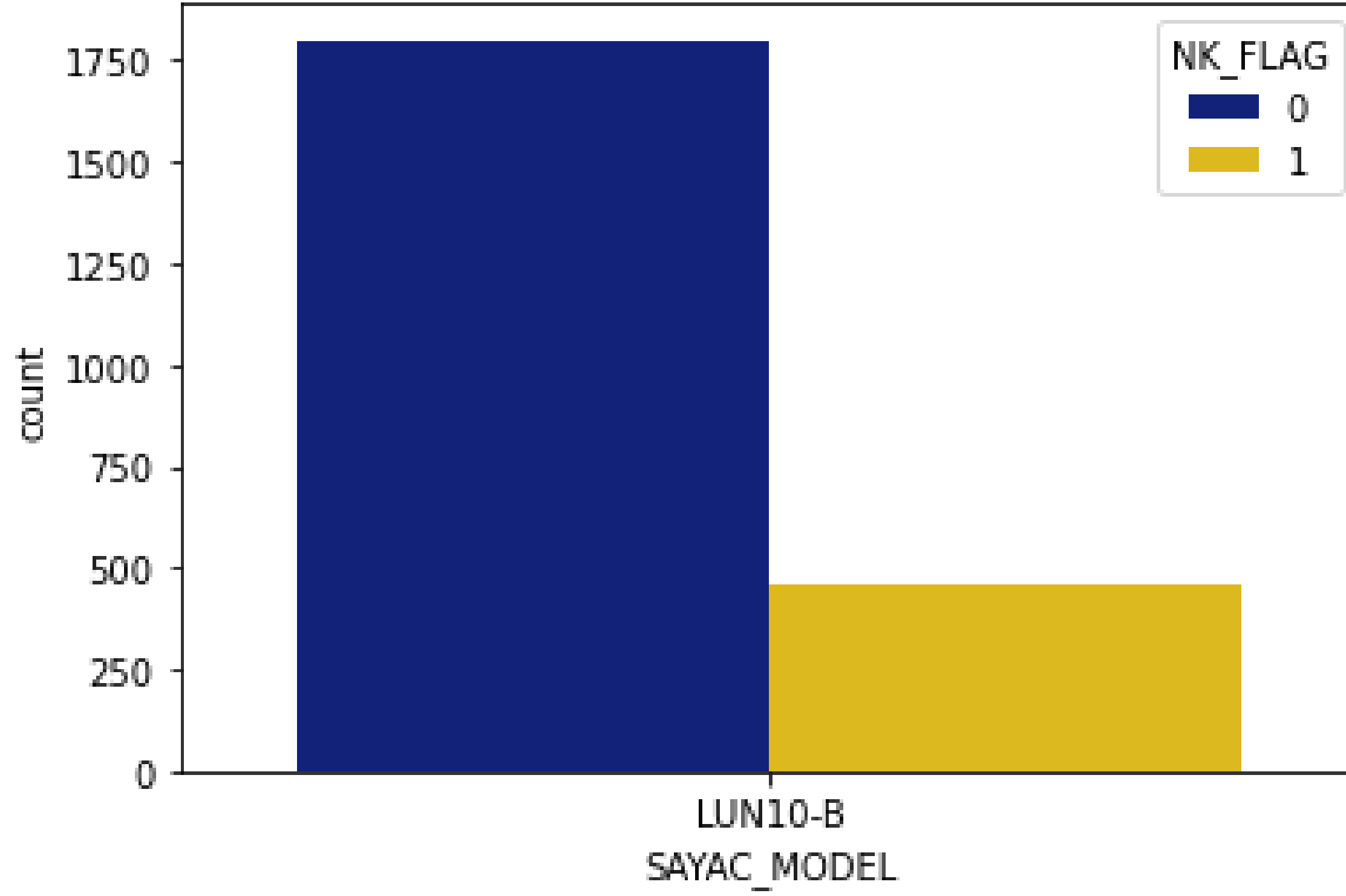


# Tespit Edilen Kaçakların Sayaç Modellerine Göre Dağılımı



Grup-2

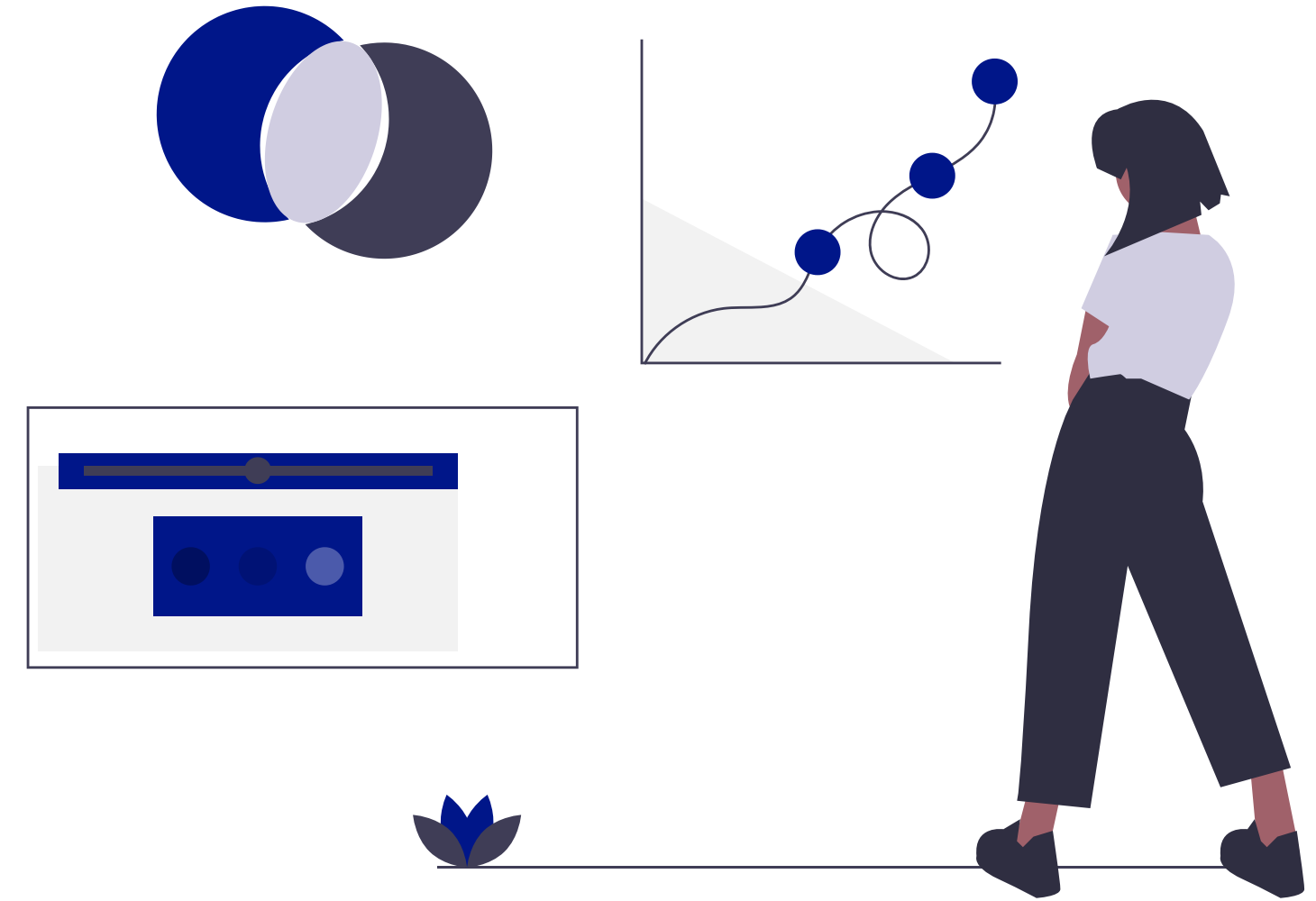
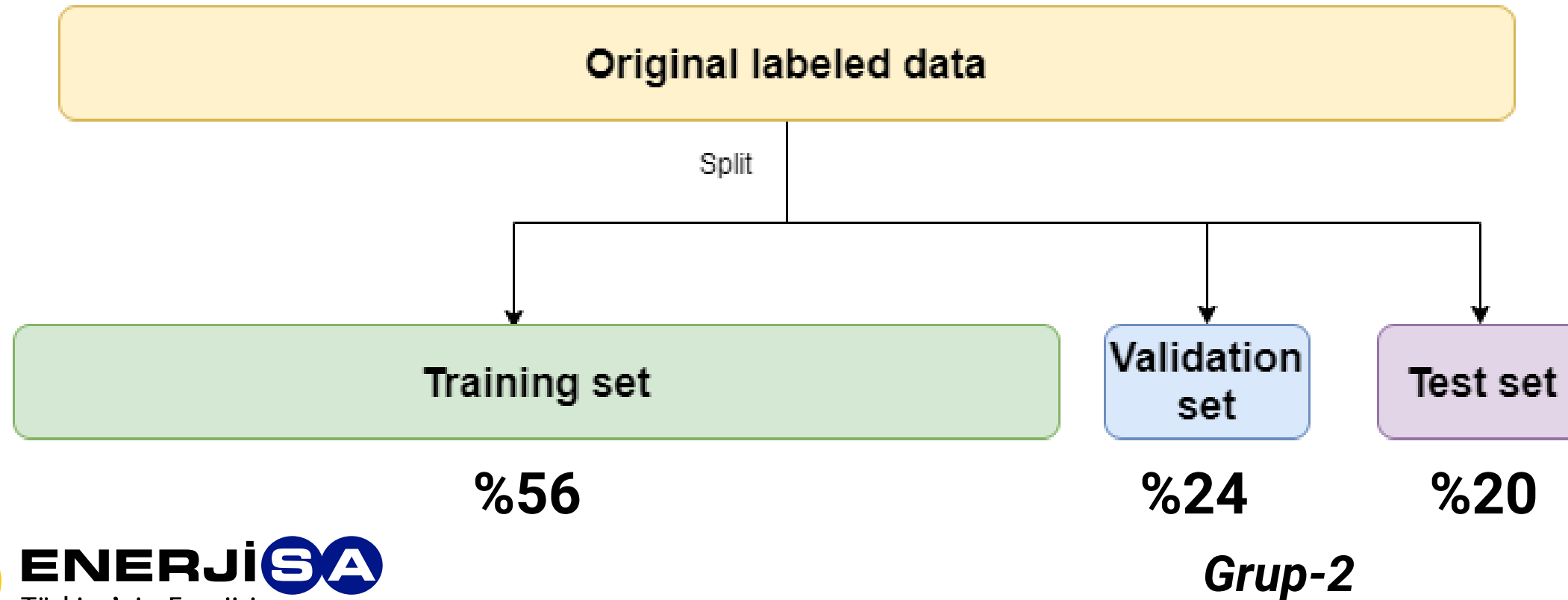
# LUN10-B Modelinin Kaçak Grafiği



Grup-2

# Model Hazırlama (Design Modelling Data )

- Geliştirilen modelin en güncel verilerle test edilmesi için sayaç başlangıç tarihine göre sıralanması
- Verilerin %20'si validasyon, %80'i ise train olarak ayrılması
- Oversampling işlemi
- Kategorik değişkenler için One-Hot Encoding



# Modelleme Algoritmaları

Recall - 0

Recall - 1

F1 Score - 0

F1 Score -1

**Decision Tree**

0.93

0.33

0.89

0.40

**Random Forest Classifier**

0.97

0.52

0.93

0.64

**Logistic Regression**

0.70

0.73

0.79

0.49

**XGBoost Classifier**

0.94

0.75

0.94

0.75

**CatBoost Classifier**

0.92

0.82

0.94

0.77

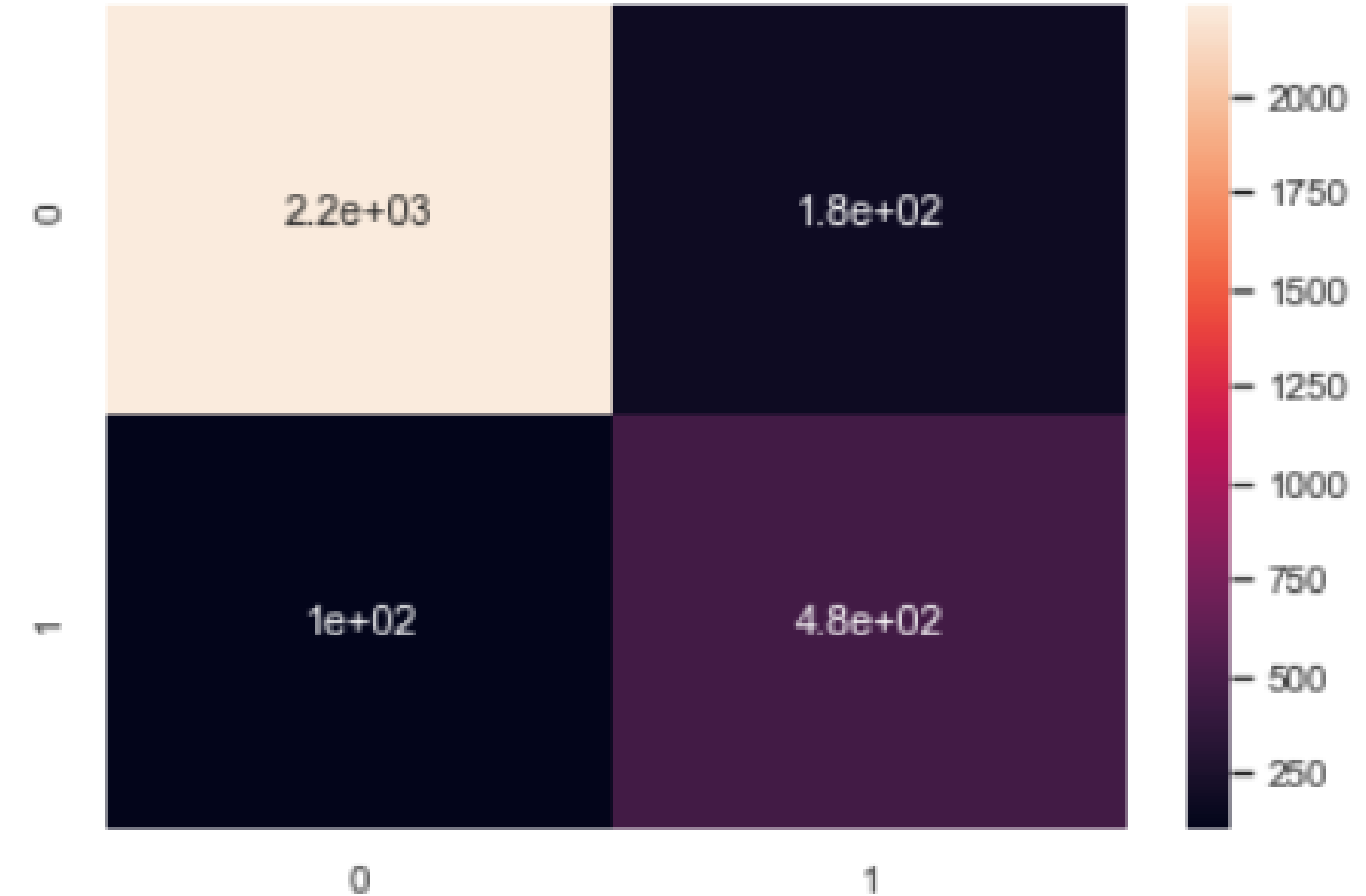
*Grup-2*

# Modelin Tahminlemesi (Model Prediction)

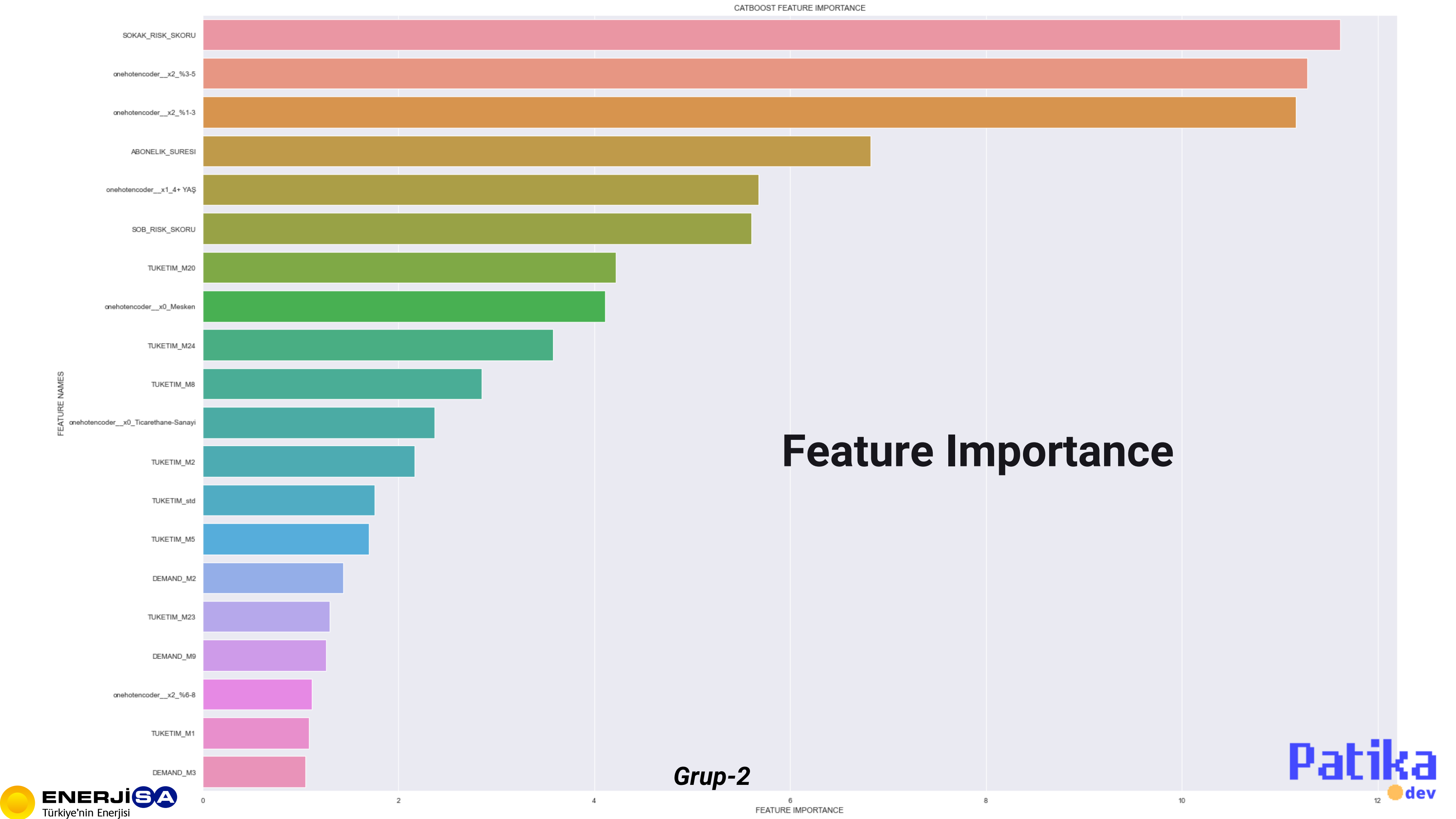
- Verilen data seti üzerinde yapılan işlemlerden sonraki CatBoost Classifier Modellemesi
- Sıfırları tahminlemede %90 üzeri değerler
- 581 kaçak kullanımdan 476 tanesinin doğru tahminlenmesi
- Sokak risk skorunun Sıfırları öğrenmede etkisi

90.4		precision	recall	f1-score	support
0	0.96	0.92	0.94	2419	
1	0.72	0.82	0.77	581	
accuracy				0.90	3000
macro avg		0.84	0.87	0.85	3000
weighted avg		0.91	0.90	0.91	3000

<AxesSubplot:>

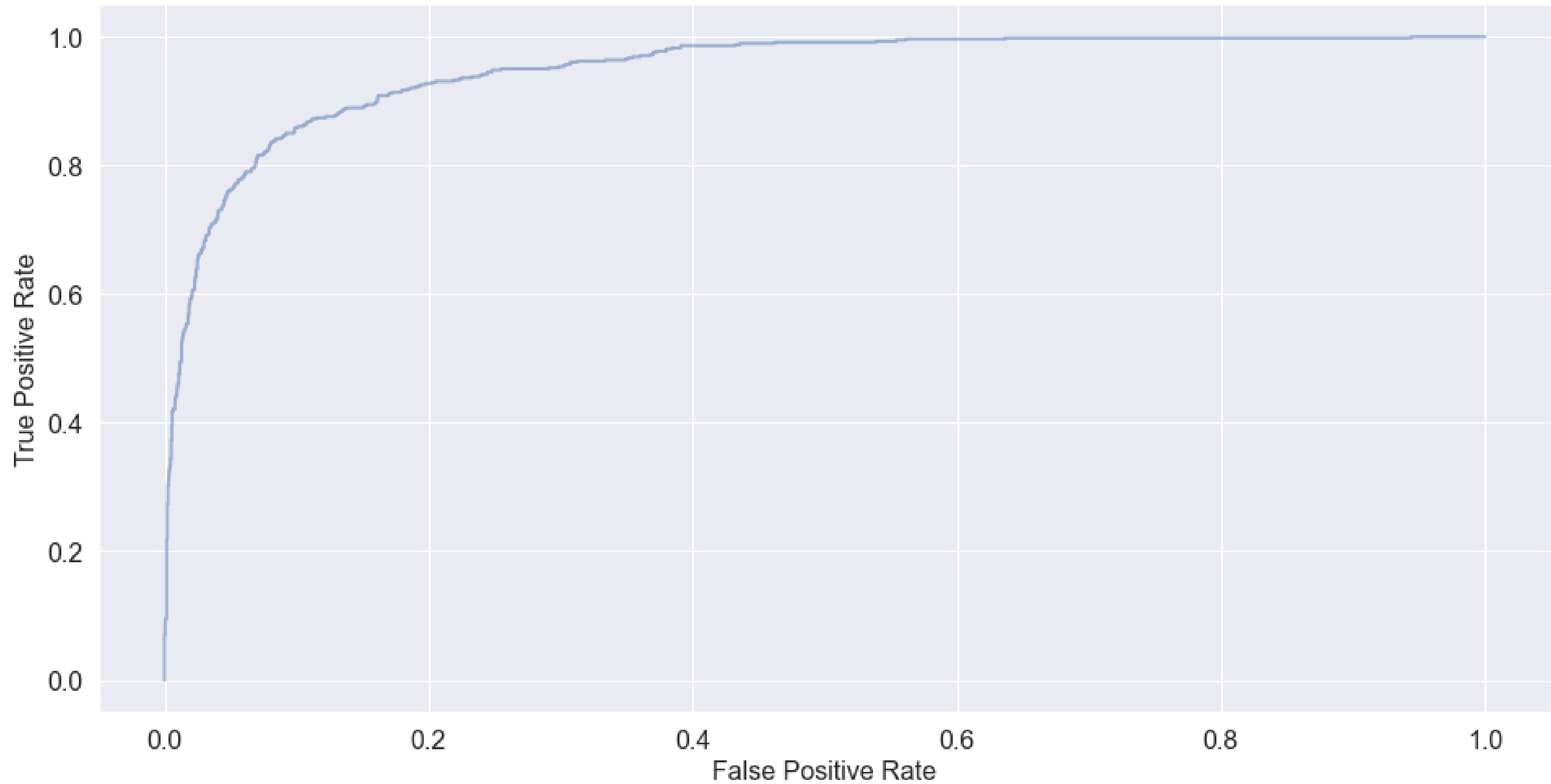


Grup-2





# ROC Curve



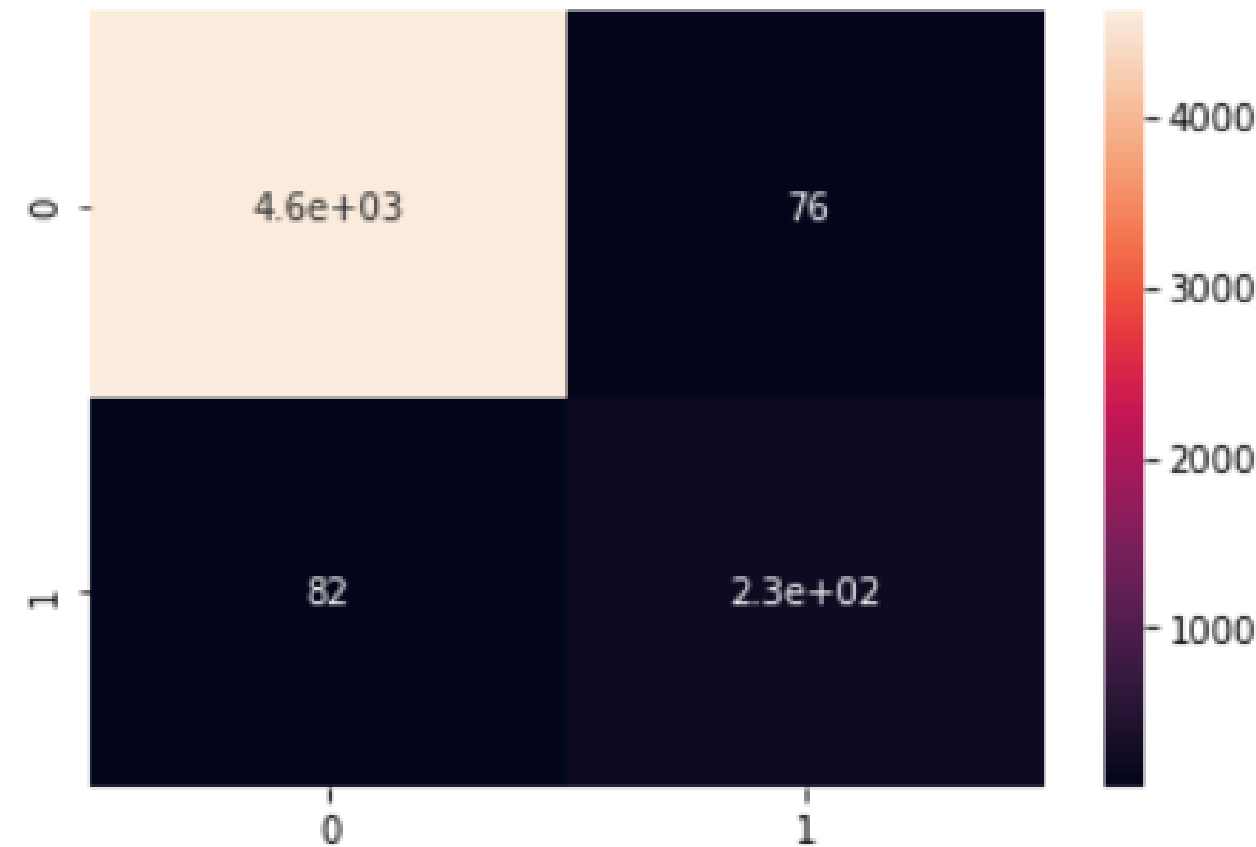
*Grup-2*

# Modelin Validasyon Sonuçları ve Değerlendirilmesi

96.64

	precision	recall	f1-score	support
0.0	0.98	0.98	0.98	4687
1.0	0.75	0.74	0.75	313
accuracy			0.97	5000
macro avg	0.87	0.86	0.86	5000
weighted avg	0.97	0.97	0.97	5000

<AxesSubplot:>

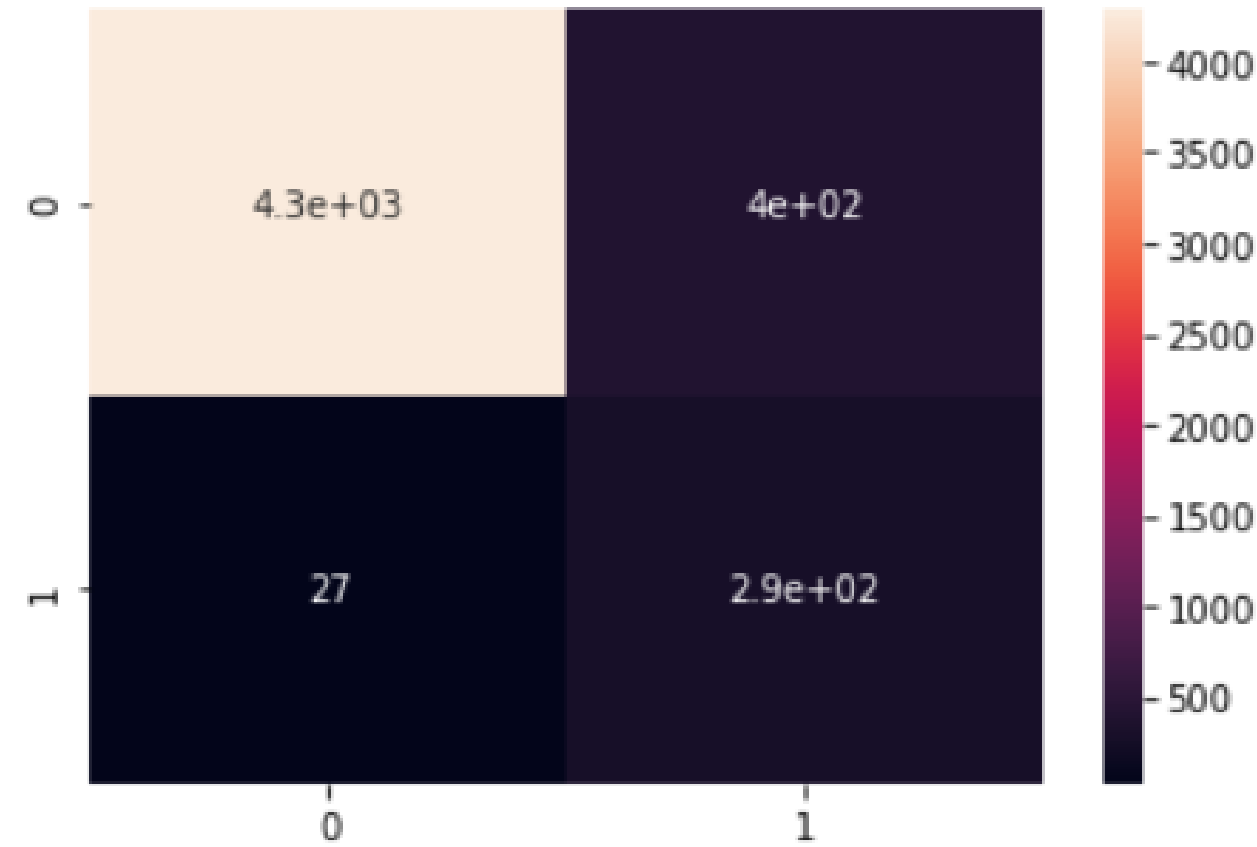


Threshold = 0.5

96.64

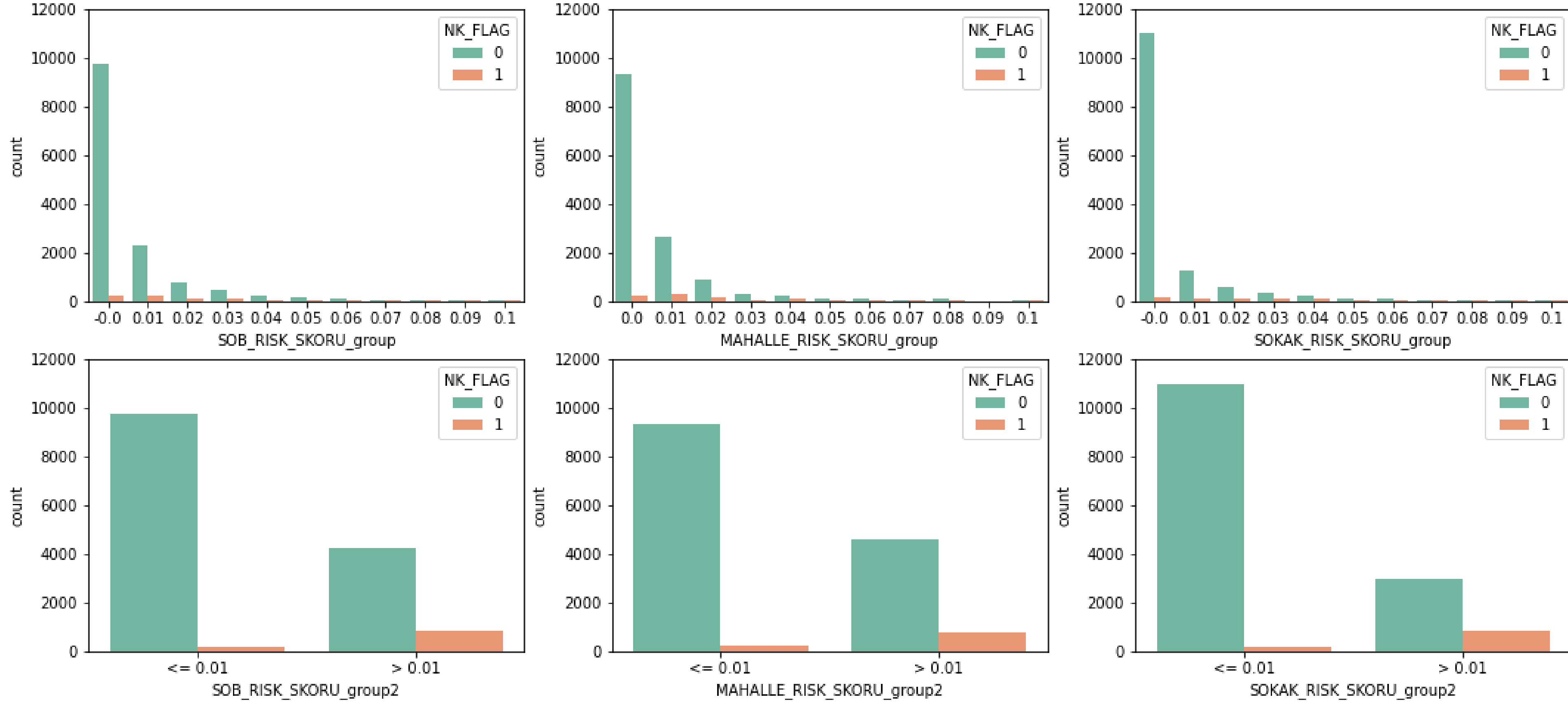
	precision	recall	f1-score	support
0.0	0.99	0.91	0.95	4687
1.0	0.42	0.91	0.57	313
accuracy			0.91	5000
macro avg	0.70	0.91	0.76	5000
weighted avg	0.96	0.91	0.93	5000

<AxesSubplot:>



Threshold = 0.15

# Modelin Validasyon Sonuçları ve Değerlendirilmesi



Grup-2

Bizi dinlediğiniz için  
teşekkür ederiz...



Grup-2