

Makine Öğrenmesi Kullanarak Müşterilerin Ürün Tercihi Tahmini

Grup-10



İçindekiler

01

Projemizin Amacı

02

Veri Setini Tanıyalım

03

Veri Hazırlığı

04

CLTV Analizi

05

SVM ile Prediction

01



Projenin Amacı

Projenin Amacı

Proje kapsamında bir ayakkabı mağazasından alışveriş yapan müşterilerin verileri ile SVM Modeli (Support Vector Machine) kullanılarak, müşterilerin hangi ürünleri satın alacağını tahmini yapılması amaçlanmıştır.



02



Veri Setini Taniyalım

19,945

Adet Müşteri

12

Veri Seti Değişkeni

14.01.2013 – 30.05.2021

Tarih Aralığı

Veri setindeki toplam gözlem ve değişken sayısı

```
df.shape
```

```
(19945, 12)
```



Veri Değişkenleri

master_id = Müşteri Kimliği

order_channel = Sipariş verilen kanal

last_order_channel = Son sipariş verilen kanal

first_order_date = İlk sipariş tarihi

last_order_date = Son sipariş tarihi

last_order_date_online = Online olarak verilen son sipariş tarihi

last_order_date_offline = Mağazadan yapılan son alışveriş tarihi

order_num_total_ever_online = Online olarak alınan toplam ürün adedi/alışveriş sayısı

order_num_total_ever_offline = Mağazadan alınan toplam ürün adedi/alışveriş sayısı

customer_value_total_ever_offline = Online'da ödenen toplam ücret

customer_value_total_ever_online = Mağazadan ödenen toplam ücret



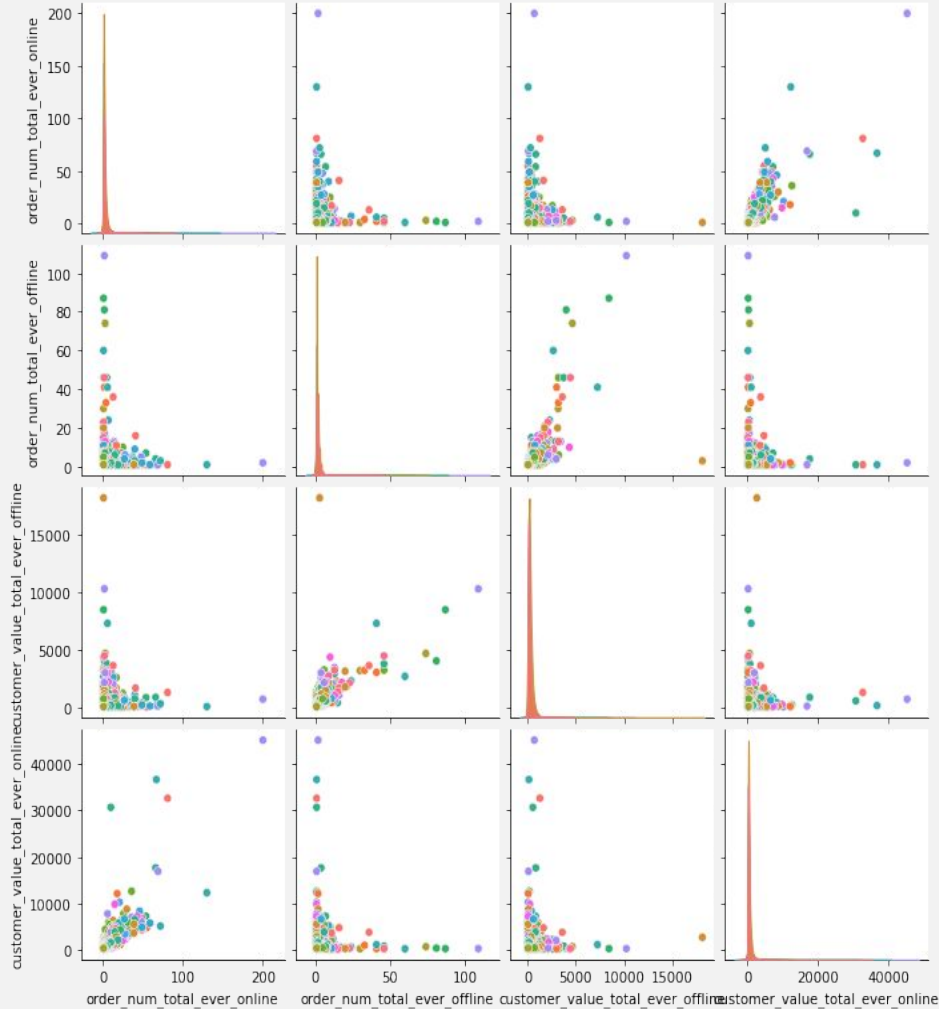
Numerik Değişkenlerinin Betimsel İstatistikleri

	count	mean	std	min	25%	50%	75%	max
order_num_total_ever_online	19945.0	3.110855	4.225647	1.00	1.00	2.00	4.00	200.00
order_num_total_ever_offline	19945.0	1.913913	2.062880	1.00	1.00	1.00	2.00	109.00
customer_value_total_ever_offline	19945.0	253.922597	301.532853	10.00	99.99	179.98	319.97	18119.14
customer_value_total_ever_online	19945.0	497.321690	832.601886	12.99	149.98	286.46	578.44	45220.13

Veri Setindeki Değişken Tipleri

#	Column	Non-Null Count	Dtype
0	master_id	19945 non-null	object
1	order_channel	19945 non-null	object
2	last_order_channel	19945 non-null	object
3	first_order_date	19945 non-null	object
4	last_order_date	19945 non-null	object
5	last_order_date_online	19945 non-null	object
6	last_order_date_offline	19945 non-null	object
7	order_num_total_ever_online	19945 non-null	float64
8	order_num_total_ever_offline	19945 non-null	float64
9	customer_value_total_ever_offline	19945 non-null	float64
10	customer_value_total_ever_online	19945 non-null	float64
11	interested_in_categories_12	19945 non-null	object

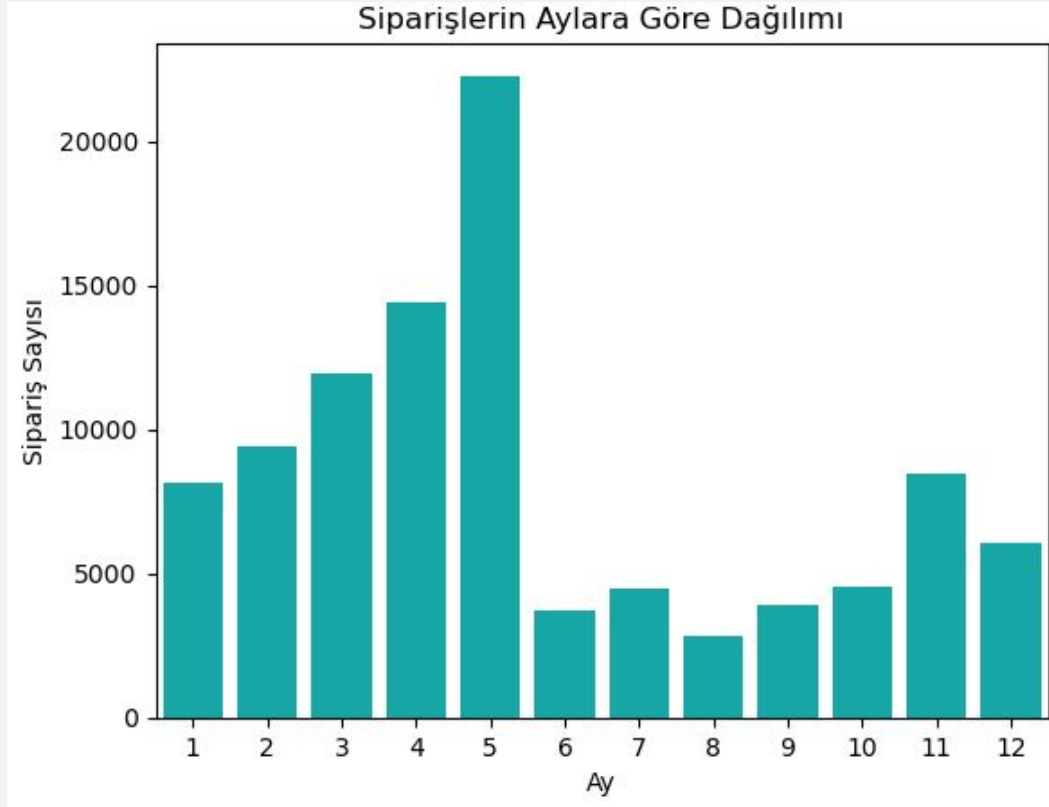
Veri setinde bulunan kategorilerin omnichannel olarak satın alınma adetleri ve parasal değerleri



interested_in_categories_12

- [KADIN]
- [ERKEK, COCUK, KADIN, AKTIFSPOR]
- [ERKEK, KADIN]
- [AKTIFCOCUK, COCUK]
- [AKTIFSPOR]
- [COCUK]
- [ERKEK, COCUK, KADIN]
- [KADIN, AKTIFSPOR]
- [AKTIFCOCUK, COCUK, KADIN]
- [COCUK, KADIN, AKTIFSPOR]
- [ERKEK, AKTIFSPOR]
- [AKTIFCOCUK, COCUK, AKTIFSPOR]
- [ERKEK]
- [COCUK, KADIN]
- [ERKEK, KADIN, AKTIFSPOR]
- [AKTIFCOCUK]
- [COCUK, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, KADIN, AKTIFSPOR]
- [ERKEK, COCUK]
- []
- [AKTIFCOCUK, COCUK, KADIN, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, COCUK]
- [AKTIFCOCUK, ERKEK]
- [AKTIFCOCUK, KADIN]
- [AKTIFCOCUK, ERKEK, COCUK, KADIN, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, COCUK, KADIN]
- [AKTIFCOCUK, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, COCUK, AKTIFSPOR]
- [AKTIFCOCUK, KADIN, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, AKTIFSPOR]
- [ERKEK, COCUK, AKTIFSPOR]
- [AKTIFCOCUK, ERKEK, KADIN]

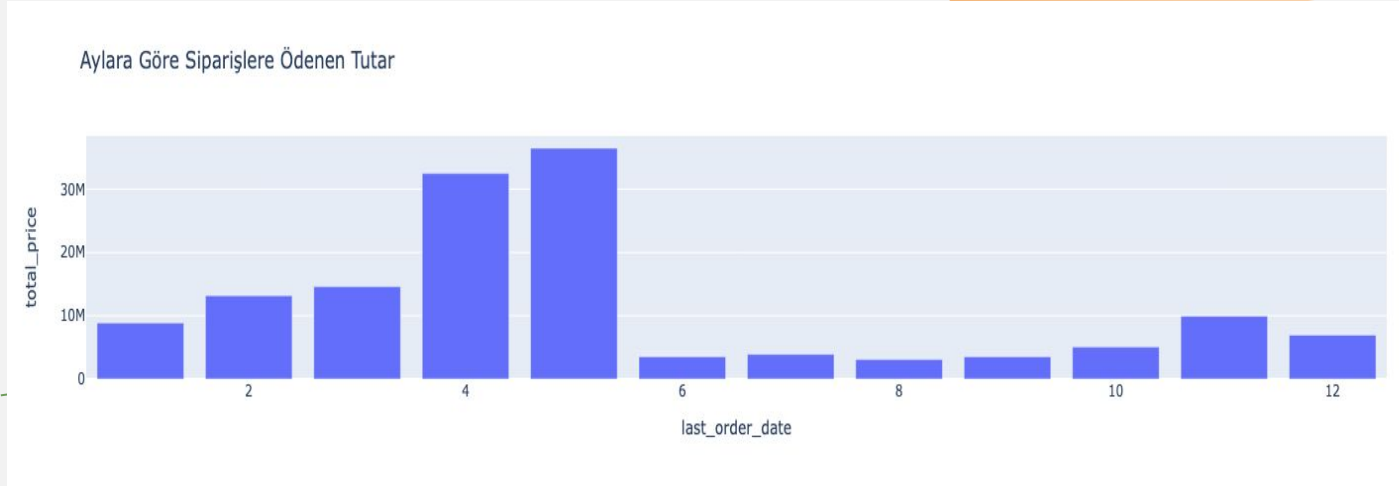
Aylara Göre Verilen Sipariş Adetleri



Veri Değişkenleri

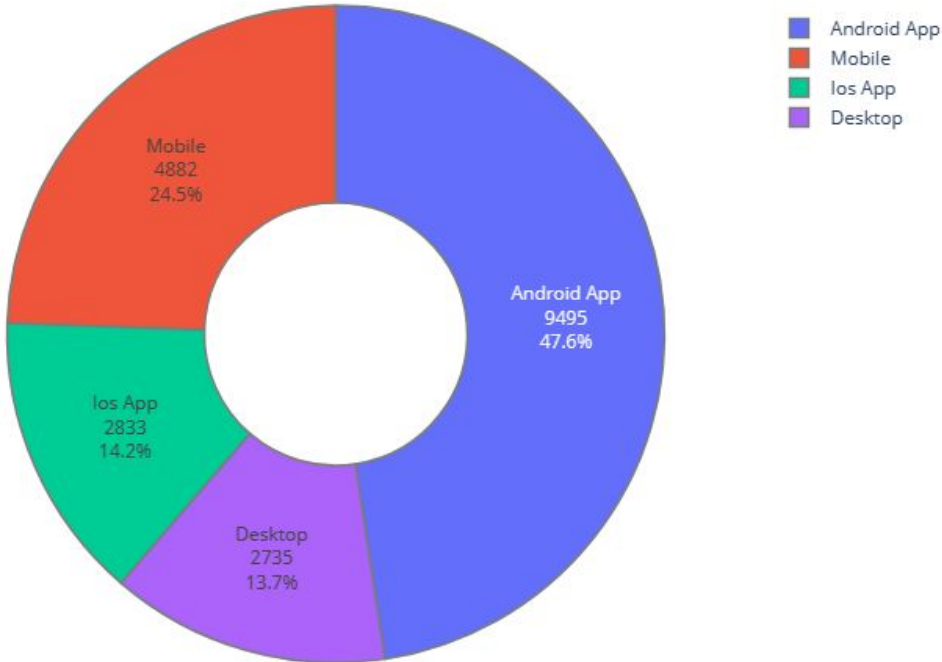
Aylara göre siparişlere ödenen tutar

En yüksek kazanç
5. ayda sağlanmıştır.



Online Kanal Siparişleri

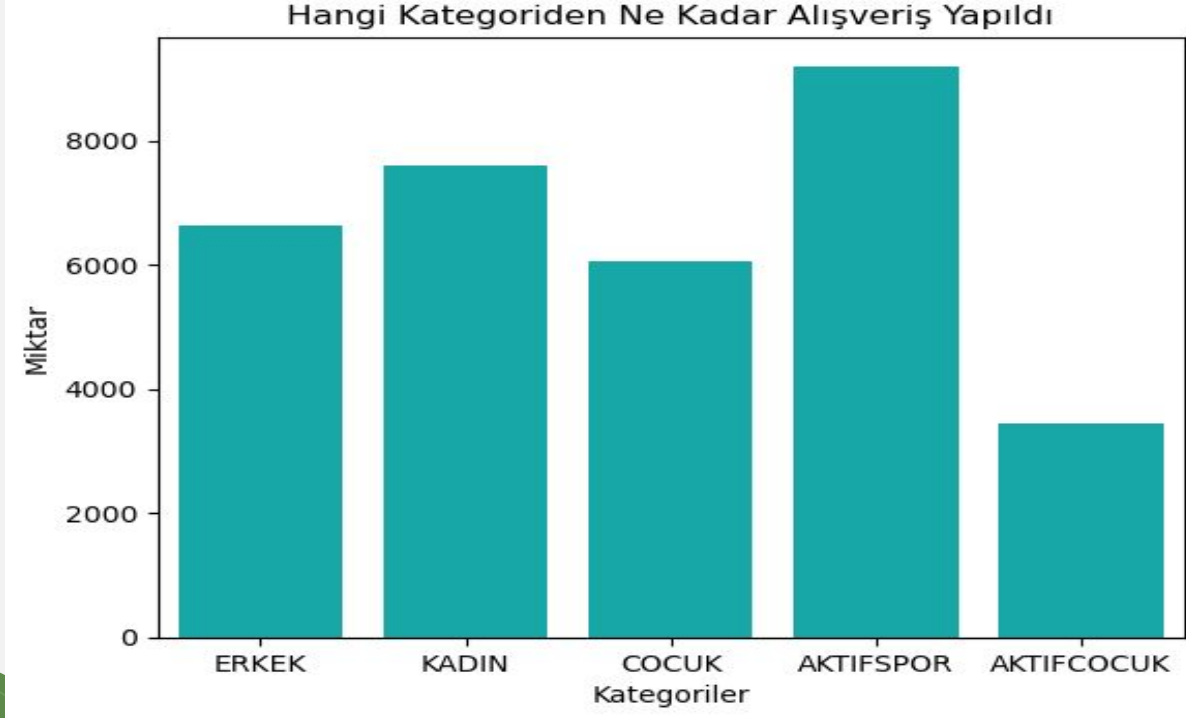
Order Channel



Adroid App kanalı, 9,495 siparişe en çok tercih edilen kanal olmuştur.

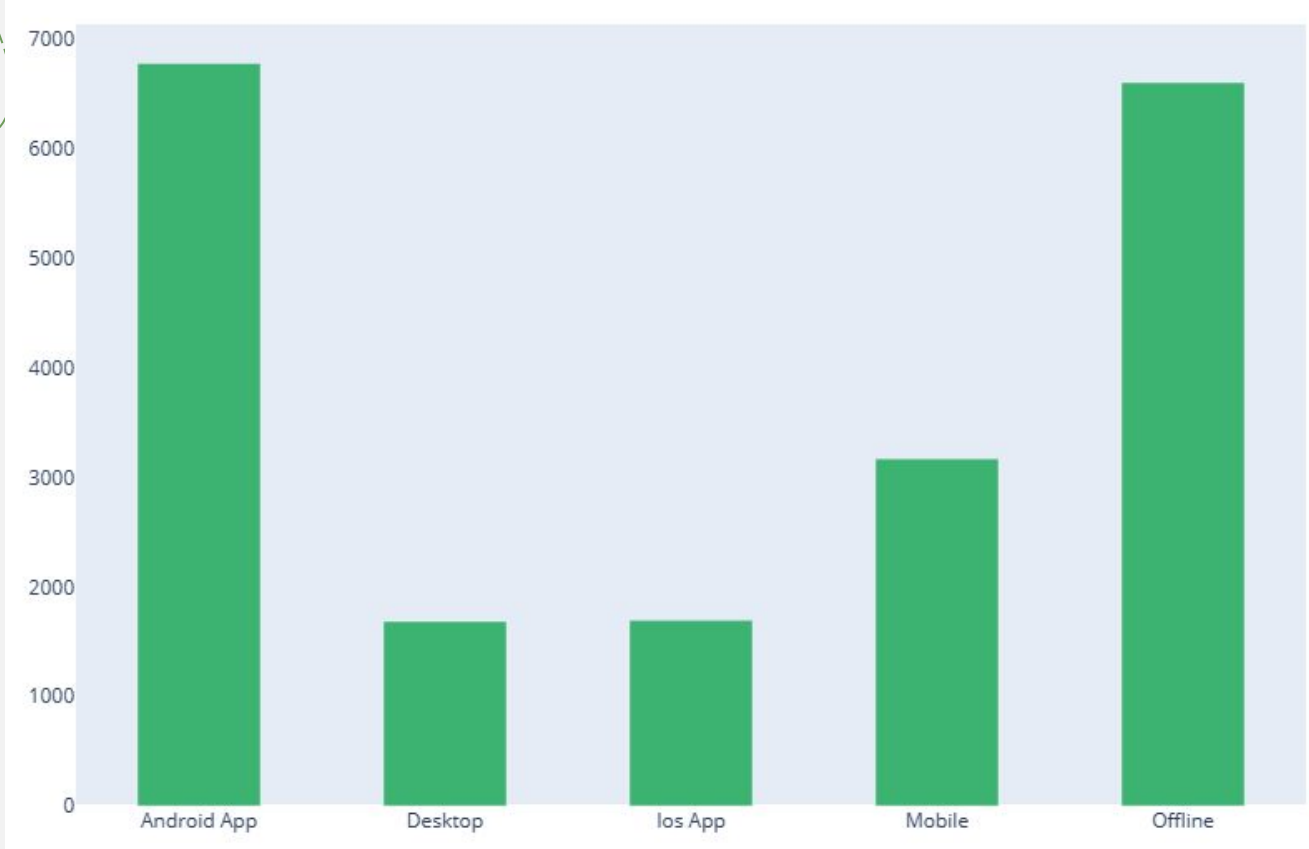


Ürün Kategorileri



Müşteriler tarafından en fazla 'AKTIFSPOR' kategorisi tercih edilirken en az 'AKTIFCOÇUK' kategorisindeki ürünler tercih edilmiştir.

Son Sipariş Kanalı



03

Veri Hazırlığı



Veri Hazırlığı

Veri setinde, tarih verileri object tipinde verilmiştir. Hazırlık aşamasında öncelikle bu düzenlenmiştir.

#	Column	Non-Null	Count	Dtype
0	master_id	19945	non-null	object
1	order_channel	19945	non-null	object
2	last_order_channel	19945	non-null	object
3	first_order_date	19945	non-null	object
4	last_order_date	19945	non-null	object
5	last_order_date_online	19945	non-null	object
6	last_order_date_offline	19945	non-null	object
7	order_num_total_ever_online	19945	non-null	float64
8	order_num_total_ever_offline	19945	non-null	float64
9	customer_value_total_ever_offline	19945	non-null	float64
10	customer_value_total_ever_online	19945	non-null	float64
11	interested_in_categories_12	19945	non-null	object

3	first_order_date	19945	non-null	datetime64[ns]
4	last_order_date	19945	non-null	datetime64[ns]
5	last_order_date_online	19945	non-null	datetime64[ns]
6	last_order_date_offline	19945	non-null	datetime64[ns]

Veri Hazırlığı

```
#Analizimizin gerçekçi olması için en son girilen veri tarihine yakın bi zamanı seçmek için son tarihe bakıyoruz (2021-05-30)  
df["last_order_date"].max()
```

```
Timestamp('2021-05-30 00:00:00')
```

Modelin 'Accuracy' değerinin yüksek olması için yeni değişkenler eklenmelidir. Bunun için de analiz yapılan günün tarihi gereklidir.

Analizin gerçekçi sonuç verebilmesi için veri setindeki son alışveriş tarihine bakılmış ve bu tarihe yakın tarih seçilmiştir.



04

CLTV Analizi



CLTV Analizi

Veri setinin daha doğru sonuçlar vermesi için alışveriş tarihlerinden müşterilere özel yeni değişkenler oluşturuldu.

Örneğin ilk alışverişinden analiz yapılan zamana ne kadar zaman geçtiğini ifade eden bu veriye CLTV analizinde müşterinin yaşı denir ve "T" olarak ifade edilir.

Bu ve buna benzer bilgiler de veri setine eklendi.

T	hang	recency	total_price	total_order	average_order_value	customer_value
215	119	96	4696.85	5.0	939.37	4720.116422
1575	1469	106	42284.55	21.0	2013.55	10117.621833

05



SVM ile Prediction

SVM

Veri setindeki müşterilerin satın aldığı ürünlerin kategorisi bu şekilde toplu ve object tipinde verilmekteydi.

Bu şekilde bir tahmin yapıldığında, accuracy oranının yeterli düzeyde olmadığı görüldü.

Model Accuracy: 0.2149588931221175

Bu sonuçtan yola çıkarak her bir kategori için ayrı feature oluşturulup boolean değerlerle ifade edildi.

Bu yolla accuracy oranı oldukça yükseldi.

interested_in_categories_12
[KADIN]
[ERKEK, COCUK, KADIN, AKTIFSPOR]
[ERKEK, KADIN]
[AKTIFCOCUK, COCUK]
[AKTIFSPOR]

SVM

Kategori değişkenleri, tabloda görüldüğü gibi tanımlandı ve her biri için model birer kez, toplam da 5 kez çalıştırıldı.

Her bir tahminde bağımsız değişken olarak diğer 4 kategori de dahil edildi. Böylece modelin doğruluk oranını daha da yükselmiş oldu.

Veri seti hazırlığının son aşaması olarak da kullanılmayacak değişkenler veri setinden silindi.

customer_value	ERKEK	KADIN	COCUK	AKTIFSPOR	AKTIFCOCUK
4720.116422	False	True	False	False	False
10117.621833	True	True	True	True	False
2941.097271	True	True	False	False	False
612.870967	False	False	True	False	True
1055.100808	False	False	False	True	False

SVM ile Kategori Tahmini

Modele örnek olarak öncelikle ERKEK kategorisinde tahmin yapıldı ve 0.72 oranında doğru sonuç elde edildi. Daha sonra tekrara düşmemek adına fonksiyonlaştırılarak diğer kategoriler için de tahminler yapıldı.

Model Accuracy: 0.7224784439542812

```
[[ 3115 168]  
 [1216 488]]
```



Model Accuracy

0.72247844

ERKEK Model Accuracy: 0.722478444

KADIN Model Accuracy: 0.692600762

COCUK Model Accuracy: 0.872067375

AKTIFSPOR Model Accuracy: 0.709444556

AKTIFCOCUK Model Accuracy: 0.868257469

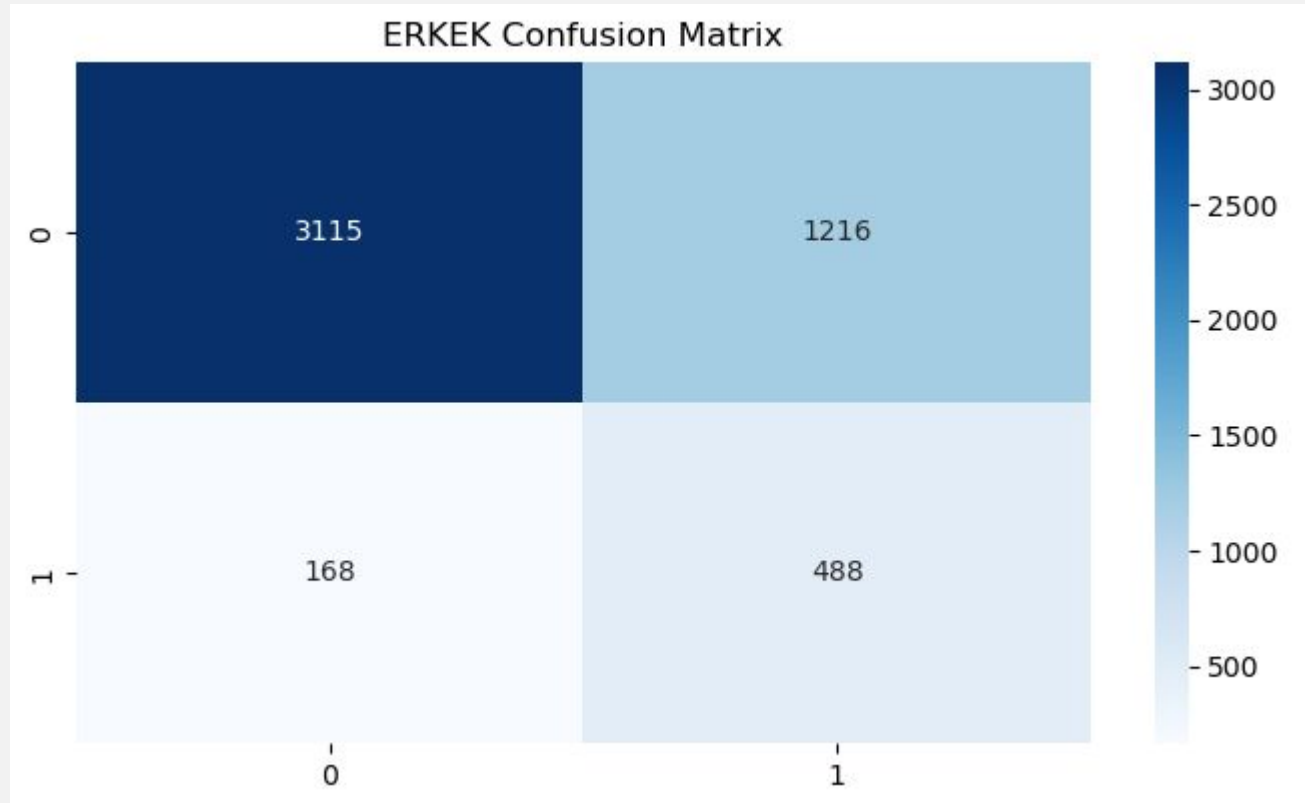


Tahmin

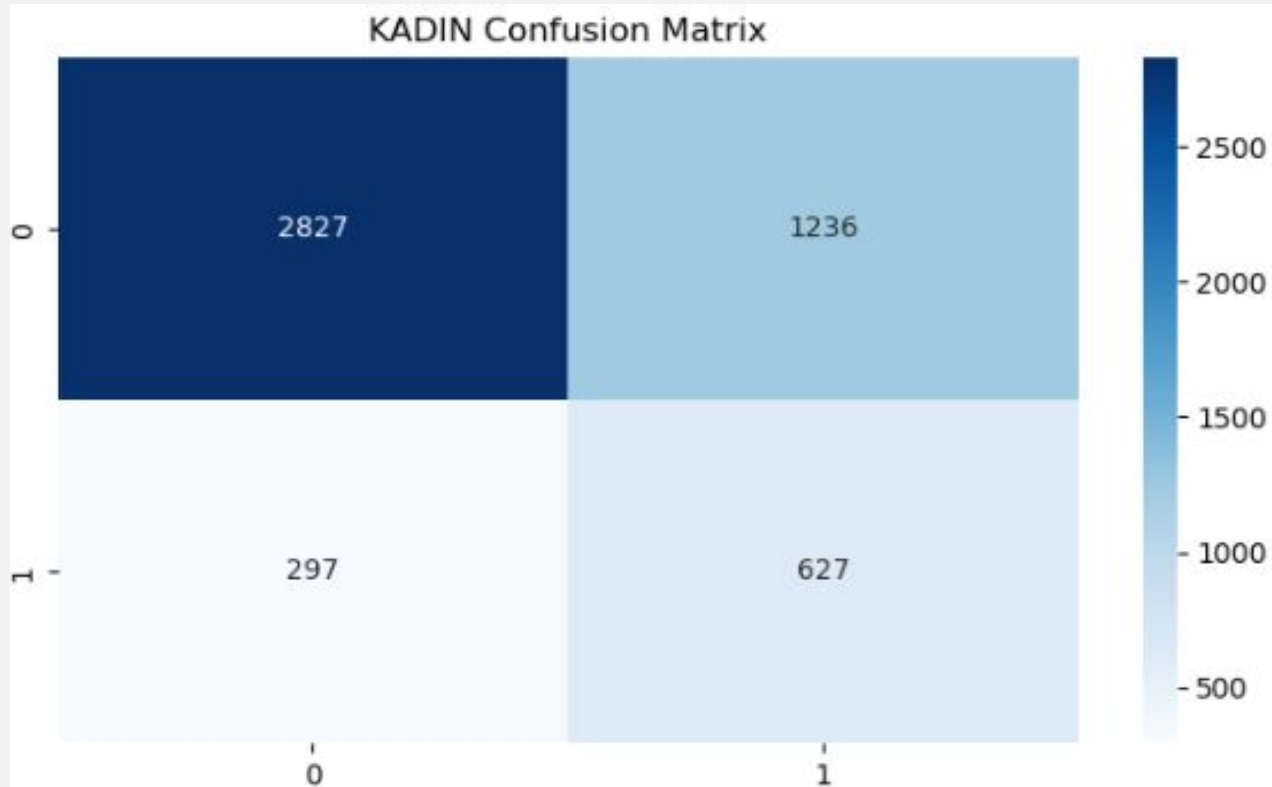
Sonuçları



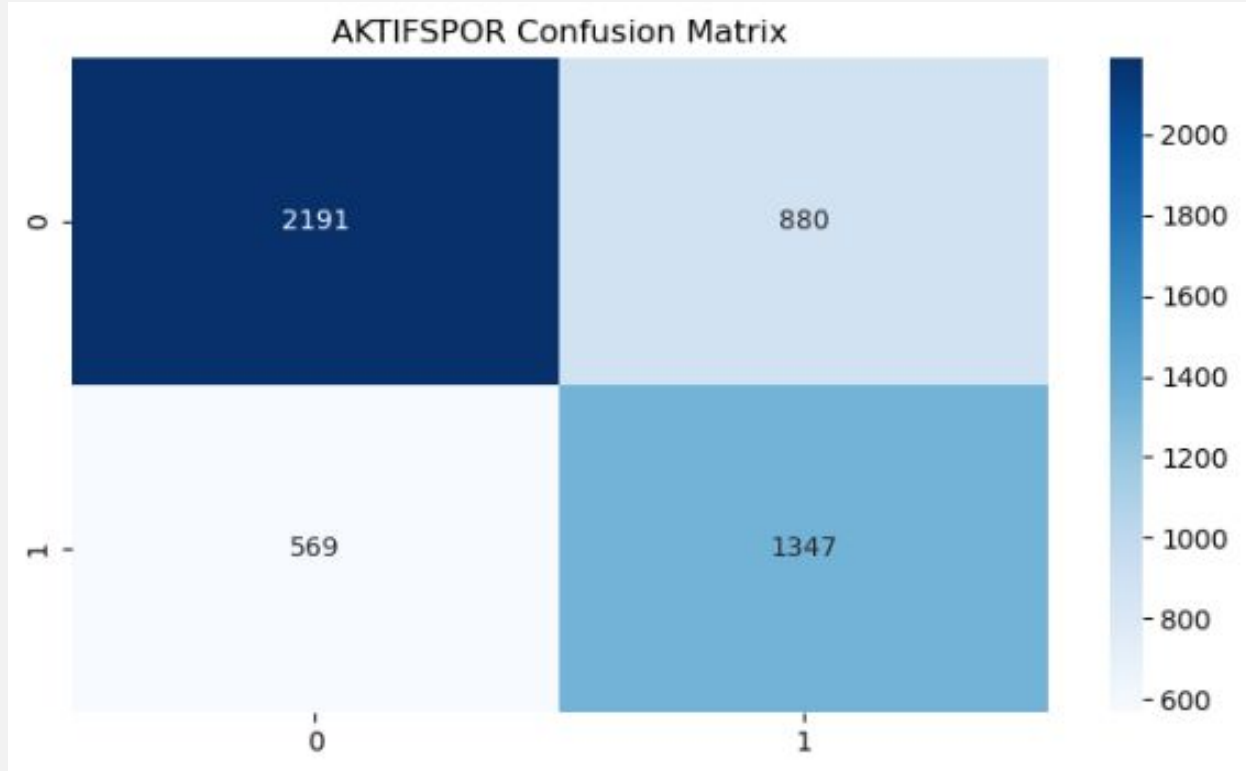
ERKEK True Positive Rate: 0.72
ERKEK False Positive Rate: 0.26



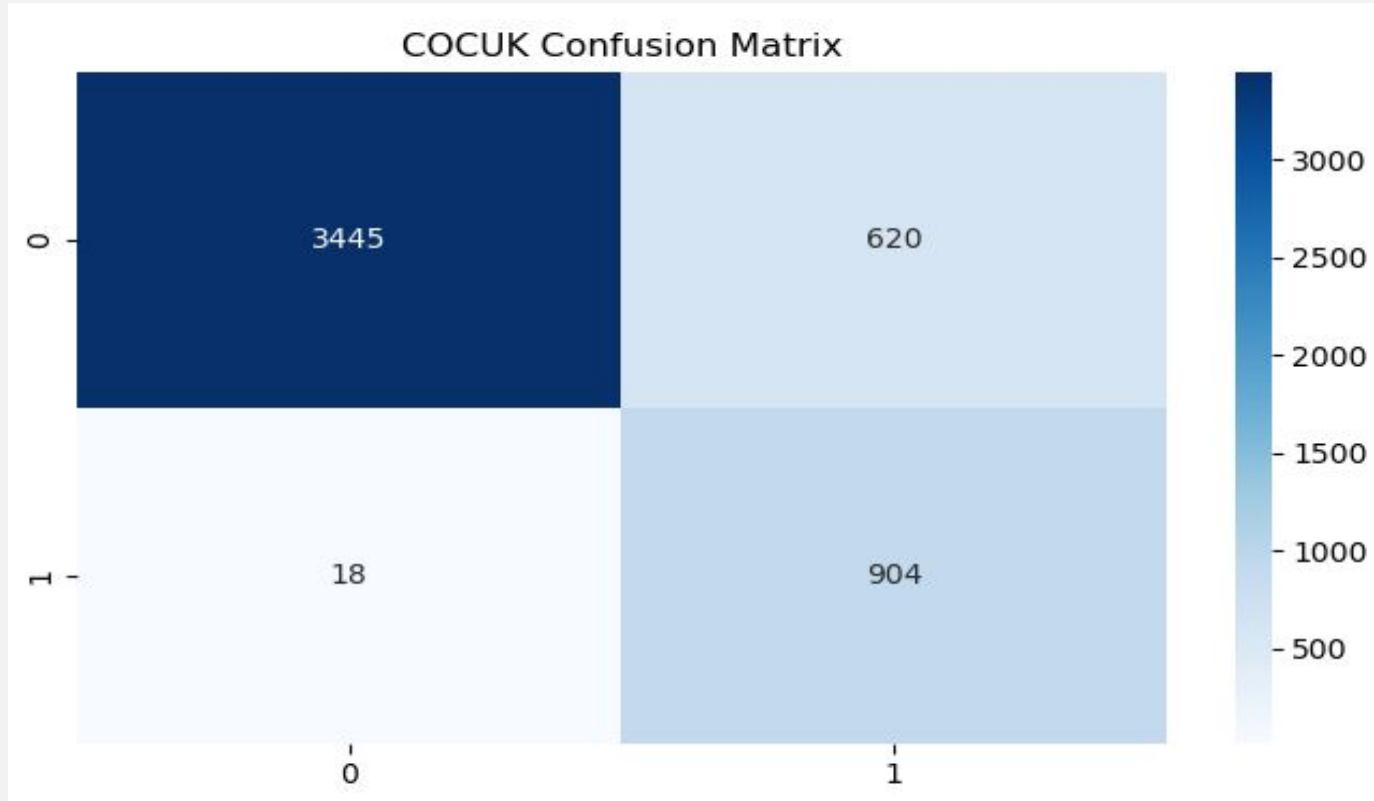
KADIN True Positive Rate: 0.70
KADIN False Positive Rate: 0.32



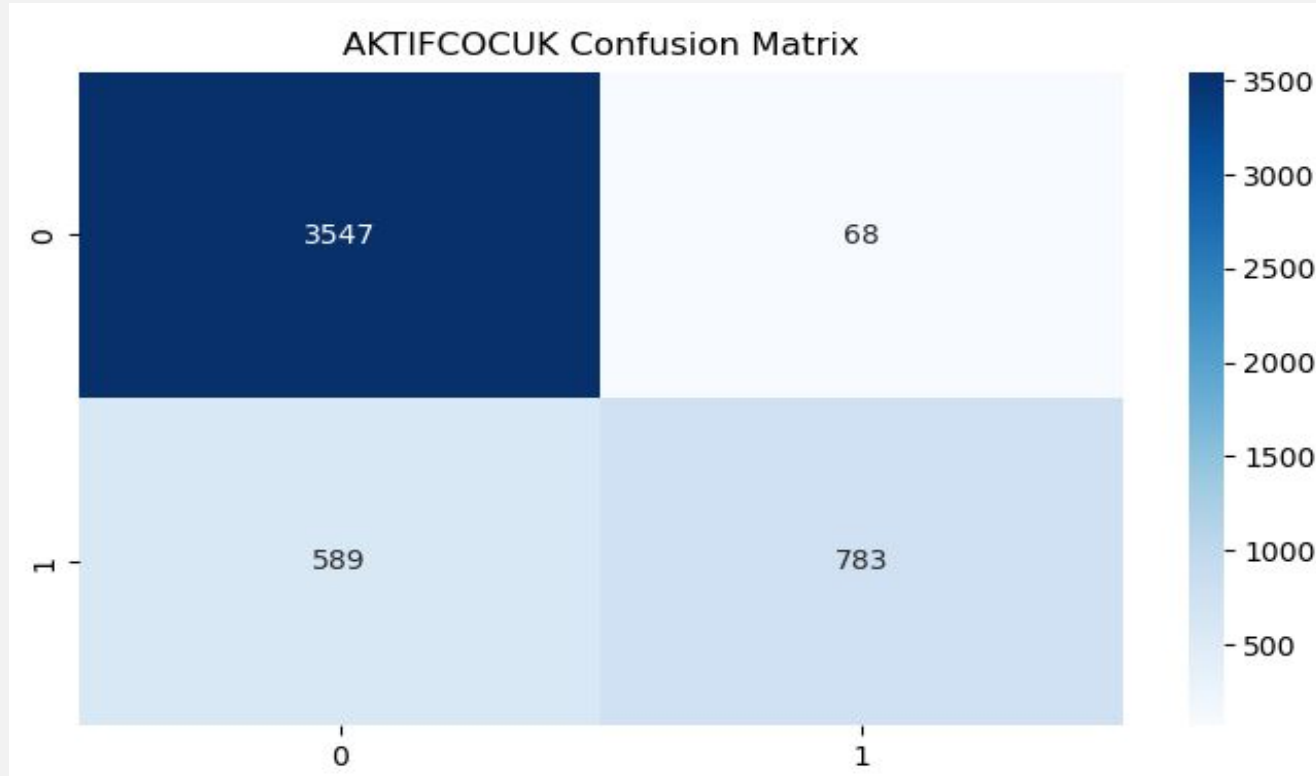
AKTIFSPOR True Positive Rate: 0.71
AKTIFSPOR False Positive Rate: 0.30



COCUK True Positive Rate: 0.85
COCUK False Positive Rate: 0.02



AKTIFCOCUK True Positive Rate: 0.98
AKTIFCOCUK False Positive Rate: 0.43



Teşekkürler!

Grup- 10

Hilal Emin

Hüseyin Efkan Alp

Berna Fevzioğlu

