

# Alman Kredi Risk Analizi

Samet Turgut



# İçerik

- Veri Seti Hikayesi
- Çalışmanın Amacı ve Hedefi
- Değişkenler
- Veri Analizi
- Base Model
- Final Model Çıktısı

# Veri Seti Hikayesi

- Alman Kredi Veri Seti, *UCI Machine Learning Repository'* den indirilebilen halka açık bir veri setidir.
- Veri Seti, 20 farklı değişkenden oluşur.
- 1000 kredi başvurusu sahibinin başvurularının, **iyi veya kötü** kredi riski değeri olarak değerlendirilmesi içerir.



# Değişkenler

#	Column	Non-Null Count		Dtype
---	-----	-----		-----
0	existingchecking	1000	non-null	object
1	duration	1000	non-null	int64
2	credithistory	1000	non-null	object
3	purpose	1000	non-null	object
4	creditamount	1000	non-null	int64
5	savings	1000	non-null	object
6	employmentsince	1000	non-null	object
7	installmentrate	1000	non-null	int64
8	statussex	1000	non-null	object
9	otherdebtors	1000	non-null	object
10	residencesince	1000	non-null	int64
11	property	1000	non-null	object
12	age	1000	non-null	int64
13	otherinstallmentplans	1000	non-null	object
14	housing	1000	non-null	object
15	existingcredits	1000	non-null	int64
16	job	1000	non-null	object
17	peopleliable	1000	non-null	int64
18	telephone	1000	non-null	object
19	foreignworker	1000	non-null	object
20	classification	1000	non-null	int64

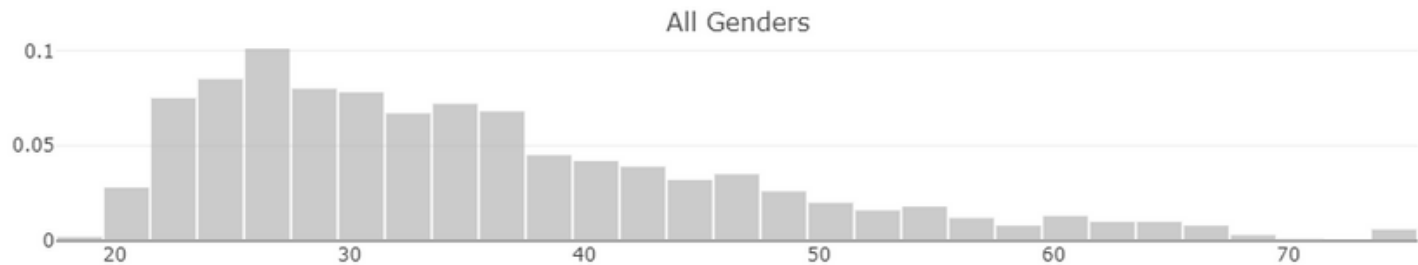
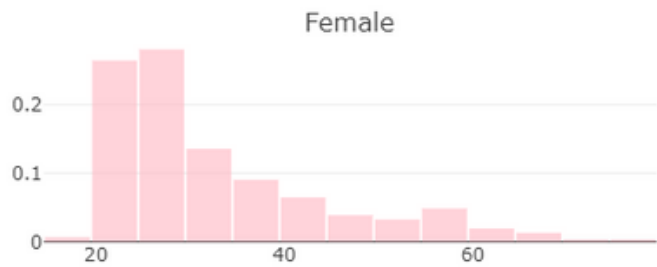
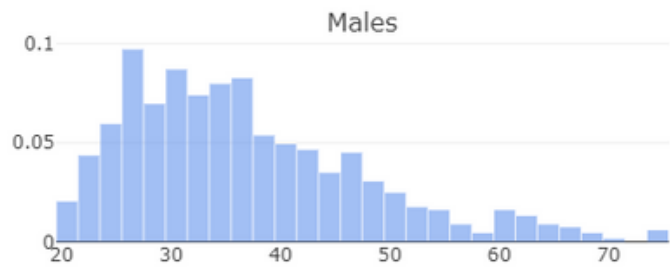
# Çalışmanın İçeriği ve Hedefi

Borçluların davranışlarının incelenmesi.

Bir kredinin riskli olup olmayacağının makine öğrenmesi modelleri ile tahmin edilebilmesi.

# Veri Analizi

job	management/ self-employed/highly qualified employee/ officer	skilled employee / official	unemployed/ unskilled - non-resident	unskilled - resident
purpose				
business	10.135	9.524	9.091	10.000
car (new)	22.973	20.635	50.000	29.500
car (used)	23.649	9.524	4.545	3.500
domestic appliances	0.000	1.587	4.545	0.500
education	4.730	5.079	4.545	5.000
furniture/equipment	14.189	20.000	4.545	16.500
others	6.081	0.159	4.545	0.500
radio/television	17.568	30.952	9.091	28.500
repairs	0.000	2.063	9.091	3.500
retraining	0.676	0.476	0.000	2.500



German Male  
German Female  
Overall Gender

statussex	female	male
job		
management/ self-employed/highly qualified employee/ officer	11.935	16.087
skilled employee / official	63.548	62.754
unemployed/ unskilled - non-resident	3.871	1.449
unskilled - resident	20.645	19.710

statussex	female	male
purpose		
business	6.129	11.304
car (new)	22.581	23.768
car (used)	7.742	11.449
domestic appliances	1.935	0.870
education	6.774	4.203
furniture/equipment	23.871	15.507
others	0.968	1.304
radio/television	27.419	28.261
repairs	1.613	2.464
retraining	0.968	0.870

# Veri Analizi

Cinsiyet ve  
Meslek



Veri setimizde kadınlardan 2 kat daha fazla erkek var.

Kredi başvurusunda bulunan kadınların çoğu 30 yaşın altındadır.

Krediye başvuran erkeklerin çoğu 20-40 yaşlarındadır.

Kadınların mobilya ve ekipman satın almak için krediye başvurma oranları daha yüksektir. (erkeklerden %10 daha fazla)

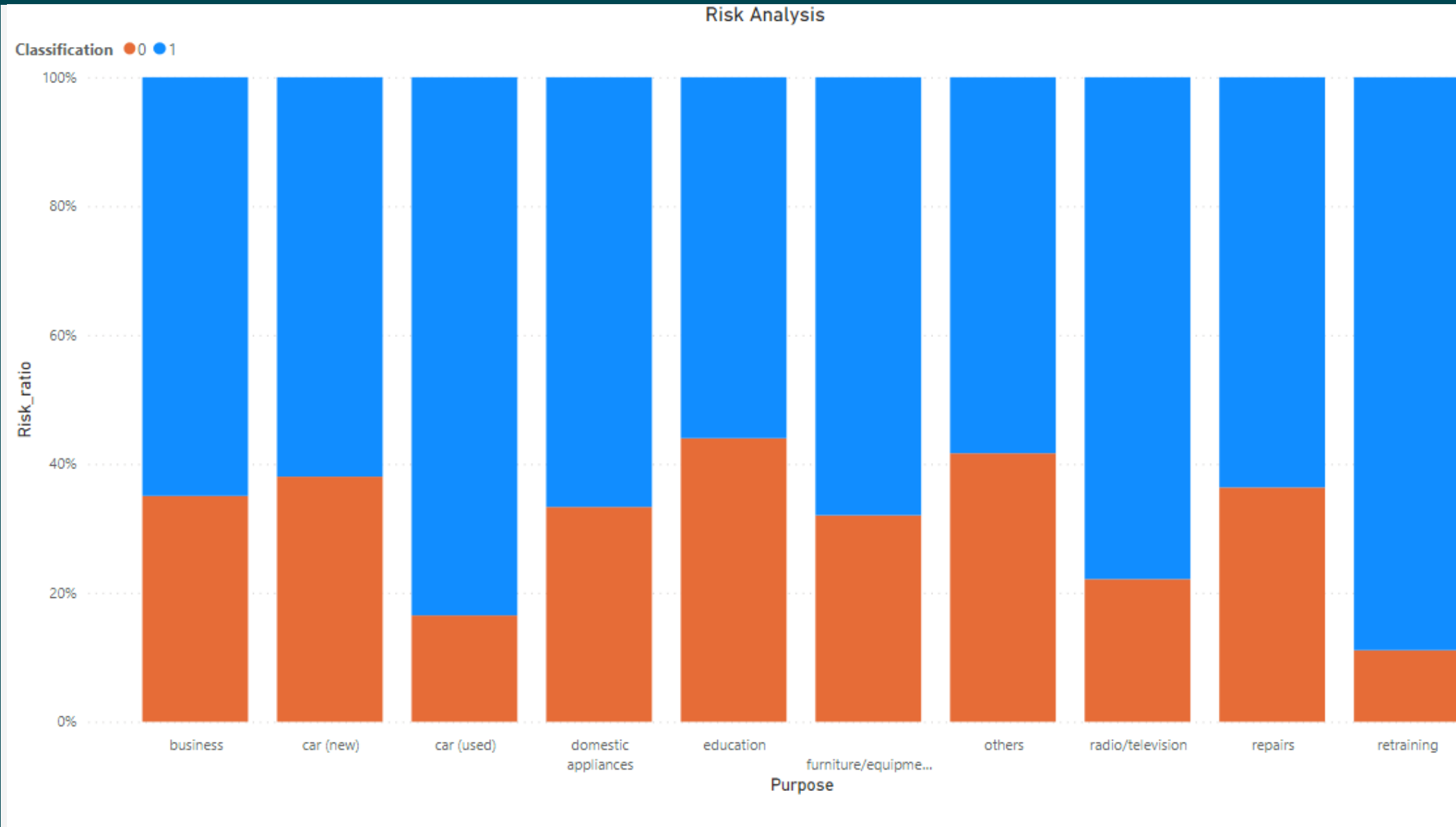
Erkekler, bir işletmeye yatırım yapmak için krediye kadınlardan 2 kat daha fazla başvurmuş.

Erkeklerle kıyasla kadınların 2 katı işsizdir.

Krediye başvuranların büyük çoğunluğu vasıflı çalışanlardır.



# Veri Analizi



classification	0	1	percentage
purpose			
retraining	1	8	89%
car (used)	17	86	83%
radio/television	62	218	78%
furniture/equipment	58	123	68%
domestic appliances	4	8	67%
business	34	63	65%
repairs	8	14	64%
car (new)	89	145	62%
others	5	7	58%
education	22	28	56%

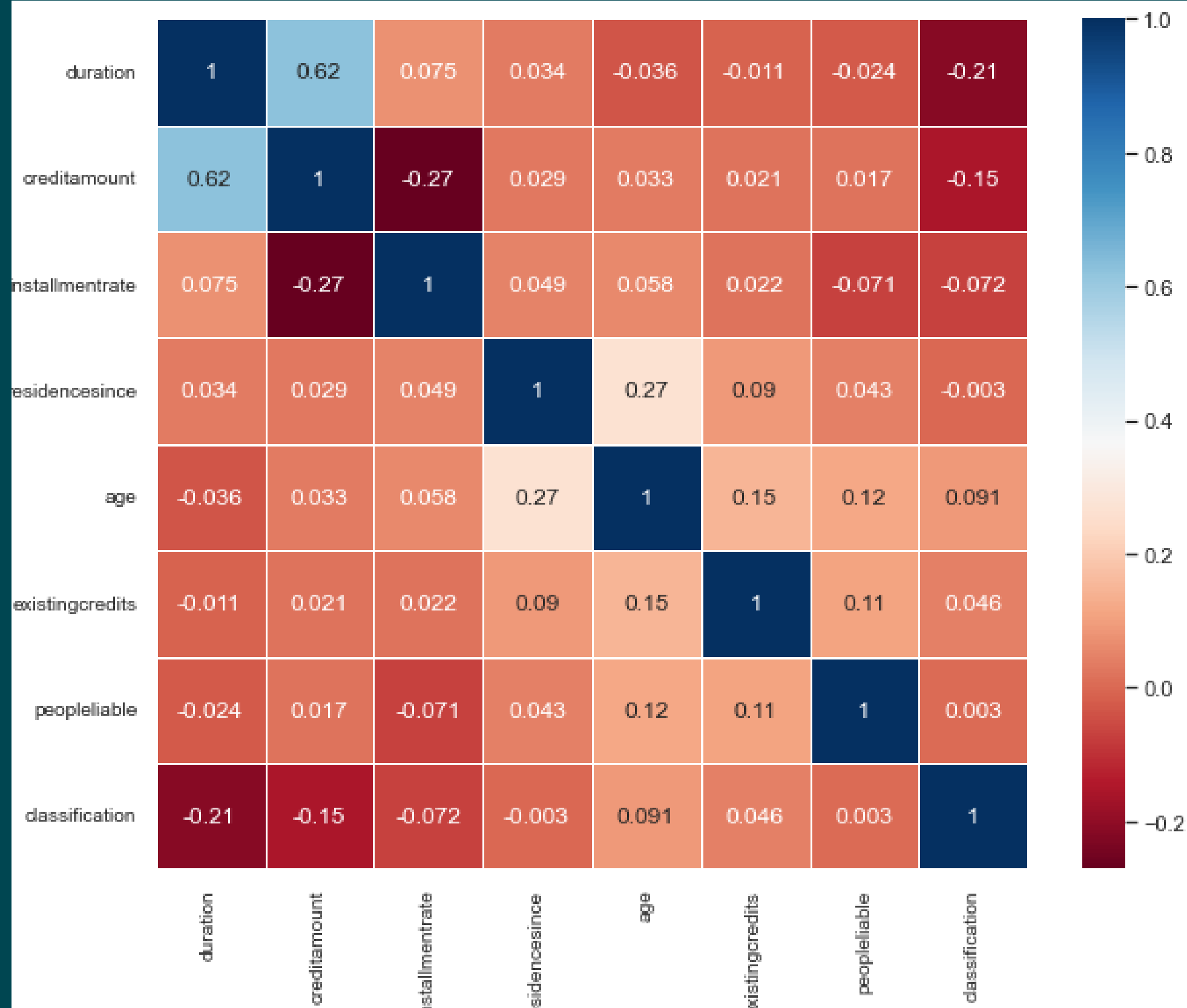


2. El araba ve eğitim kredilerinin diğerlerine göre risk oranı daha fazladır.

Araba, radyo/televizyon ve mobilya kredileri kredilendirilen ürünlerin % 80'ini oluşturmaktadır.



# Korelasyon



Alınan kredi tutarı ne kadar yüksek olursa, kredinin kötü sonuçlanma olasılığı o kadar yüksek olur.

Kredinin süresi ne kadar uzun olursa, kredinin kötü olma olasılığı o kadar yüksek olur.

# Base Model

```
Base Models....
accuracy: 0.749 (LR)
accuracy: 0.694 (KNN)
accuracy: 0.733 (SVC)
accuracy: 0.668 (CART)
accuracy: 0.743 (RF)
accuracy: 0.742 (Adaboost)
accuracy: 0.74 (GBM)
accuracy: 0.751 (XGBoost)
accuracy: 0.75 (LightGBM)
```

```
Base Models....
roc_auc: 0.7853 (LR)
roc_auc: 0.6896 (KNN)
roc_auc: 0.7735 (SVC)
roc_auc: 0.6081 (CART)
roc_auc: 0.7715 (RF)
roc_auc: 0.7759 (Adaboost)
roc_auc: 0.7755 (GBM)
roc_auc: 0.771 (XGBoost)
roc_auc: 0.782 (LightGBM)
```



XG Boost  
LightGBM

# Hiper Parametre Optimizasyonu Sonucu

```
Hyperparameter Optimization....  
##### XGBoost #####  
accuracy (Before): 0.751  
accuracy (After): 0.761  
XGBoost best params: {'colsample_bytree': 1, 'learning_rate': 0.1, 'max_depth': 8, 'n_estimators': 100}  
  
##### LightGBM #####  
accuracy (Before): 0.75  
accuracy (After): 0.753  
LightGBM best params: {'colsample_bytree': 1, 'learning_rate': 0.1, 'n_estimators': 300}
```

```
Hyperparameter Optimization....  
##### XGBoost #####  
roc_auc (Before): 0.771  
roc_auc (After): 0.7802  
XGBoost best params: {'colsample_bytree': 1, 'learning_rate': 0.1, 'max_depth': 8, 'n_estimators': 100}  
  
##### LightGBM #####  
roc_auc (Before): 0.782  
roc_auc (After): 0.7794  
LightGBM best params: {'colsample_bytree': 1, 'learning_rate': 0.1, 'n_estimators': 300}
```

# Voting Classifier

```
Voting Classifier...  
Accuracy: 0.754000000000000001  
F1Score: 0.8321570744977566  
ROC_AUC: 0.7830238095238096
```

# TEŞEKKÜRLER

Abdül Samet Turgut