

Explainability(Açıklanabilirlik) Raporu

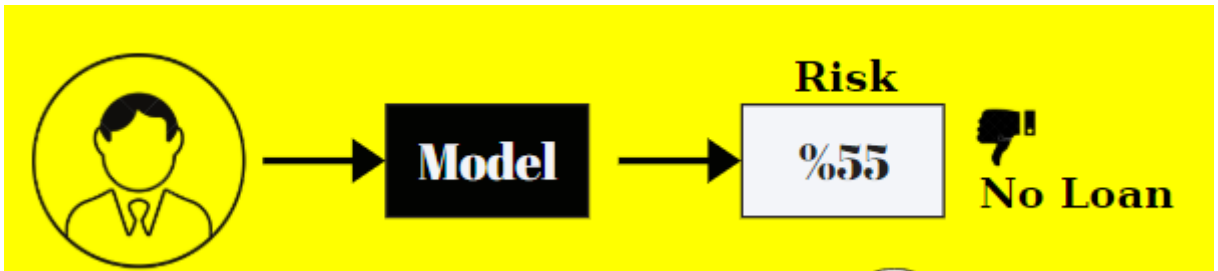
Explainability(açıklanabilirlik), bir makinenin veya derin öğrenme sisteminin iç mekaniğinin insani terimlerle ne ölçüde açıklanabileceğidir.

Sağlık hizmeti ve banka gibi riskli alanlar, hesap verebilirlik ve şeffaflık sorularının özellikle önemli olduğu yapay zeka ve derin öğrenme sistemleri çok önemlidir.

Kısaca buradaki temel amaç siyah kutu diye tabir ettiğimiz makine öğrenme algoritmaların verdiği çıktının hangi değişkenlere göre değiştiğini vs. açıklamaktır.

Bir banka müşterilerini tanımak istiyor müşterilerinin iki yıl içinde herhangi bir gelir riski taşıyıp taşımayacağına bakarak ona göre müşterilerine kredi verip veremeyeceğini karar veriyor.

Müşterinin daha önceki ödemlerdeki gecikme, aldığı konut kredileri sayısı, aylık geliri, yaş bilgisi vs. göre eğer müşterinin risk olasılığı %50 üstünde ise kredi vermiyor.



Burada biz kendimize şu soruyu sormalıyız yazdığımız algoritma ne kadar güvenilir ya da müşteri neden kredi alamadığını sorduğunda nasıl bir cevap vereceğiz. Müşteriye elbette oluşturduğumuz makine öğrenme algoritmalarından bahsetmeyeceğiz büyük ihtimalle makine öğrenme algoritmasının ne olduğunu bilmiyordur. Bu yüzden modeller açıklama yoluna giriliyor.

Explainability Uygulama

Explainability temel adımlar:

1. Verisetini okuma

```
test_data = pd.read_csv("cs-test.csv")  
train_data = pd.read_csv("cs-training.csv")
```

2. Bir model oluşturma

```
gbm_model = GradientBoostingClassifier(learning_rate= 0.05,max_depth=4,n_estimators=200)  
gbm_model = gbm_model.fit(X_train,y_train)
```

3. Explainability yöntemlerini kullanmak. LIM, SHAP burada SHAP kullanıldı.

```
xgb_model_explainer = shap.TreeExplainer(gbm_model)  
gbm_shap_values = xgb_model_explainer.shap_values(X_test)
```

Bu arada XGBoost algoritması ağaç temelli bir algoritma olduğu için TreeExplainer kullanıldı. Ağaç temelli olmayan algoritmalarda KernelExplainer kullanıldı. (SVM)
İkinci satırda ise test verisetinin shapley değerleri hesaplanıyor.

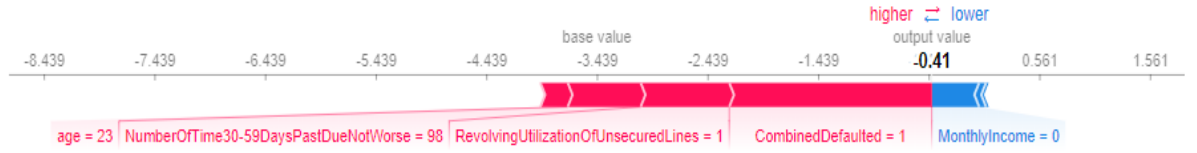
Shapley değerleri kısaca her özelliğin çıktı üzerinde ne kadar etkisi olduğunu hesaplıyor. Bu adım sonunda elimizde bir açıklanabilirlik modeli oluşturmuş olduk.

- ## 4. Bu adımda oluşturulan açık kutu modeli için bazı görselleştirmelere bakalım
- Force plot(kuvvet grafiği), her özelliğin, karar puanının değerinin taban değerden (tüm girdiler eksikse karar puanının tahmini) sınıflandırıcı tarafından öngörülen değere taşınması sürecine katkısını gösterir.

```
shap.force_plot(xgb_model_explainer.expected_value, gbm_shap_values[9,:], X_test.iloc[9,:])
```

Satırın shapley
değerleri

Satırın normal
değerleri



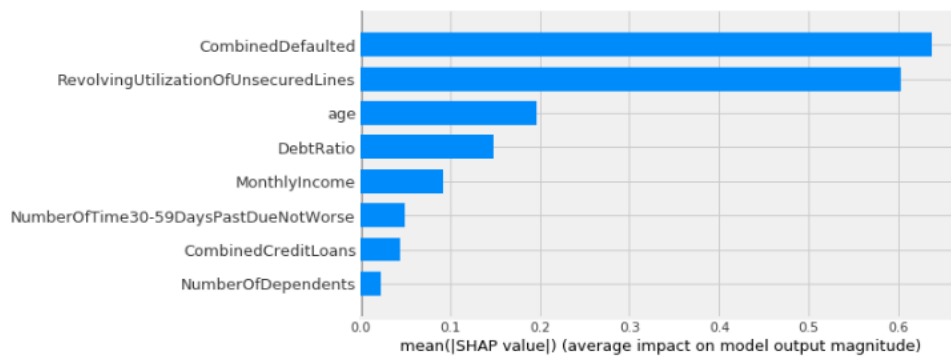
b. Summary plot özelliklerin model üzerindeki etkileri en fazladan en aza doğru yukarıdan aşağıya sıralanır.

```
shap.summary_plot(gbm_shap_values, X_test, show=False)
```



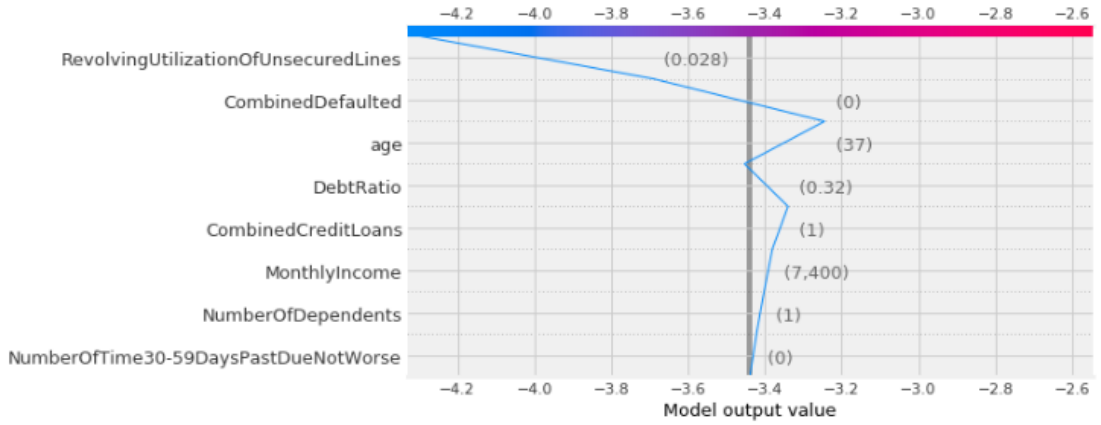
Bu grafik aynı zamanda bar plot olarak da gösterilebilir.

```
shap.summary_plot(gbm_shap_values,sample, plot_type="bar",show=False)
```



- c. Decision plot bu grafikte summary plot ile çok benzerdir özellikleri önem derecelerine göre yukarıdan aşağıya doğru sıralar. Bu grafikte katkı değerini de gösterir.

```
shap.decision_plot(xgb_model_explainer.expected_value, gbm_shap_values[10,:], X_test.iloc[10,:])
```



Örnek olarak 10 ID kişinin bilgilerine bakıp model nasıl bir sonuç vermiş ona bakalım.

```
X_test.iloc[10]
```

```
RevolvingUtilizationOfUnsecuredLines    0.028485
age                                     37.000000
NumberOfTime30-59DaysPastDueNotWorse    0.000000
DebtRatio                               0.319687
MonthlyIncome                          7400.000000
NumberOfDependents                      1.000000
CombinedDefaulted                       0.000000
CombinedCreditLoans                     1.000000
Name: 10, dtype: float64
```

Kişinin kredi sonucuna bakalım

```
Id          10.000000
Probability  42.403987
Name: 9, dtype: float64
```

```
shap.force_plot(xgb_model_explainer.expected_value, gbm_shap_values[10,:], X_test.iloc[10,:])
```



Mavi değerler risk oluşturmuyor demek. Kırmızı ise risk var oluşturuyor demek oluyor.

Risk olasılığı 42 yani kredi alabilir.

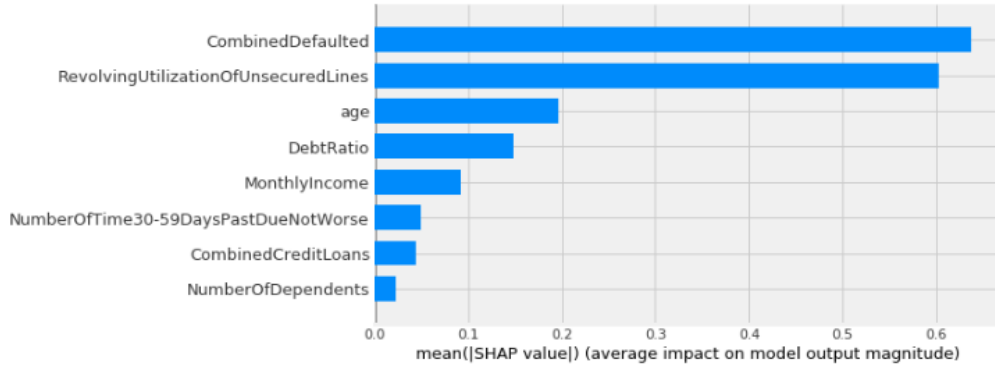
Başka kişinin kredisine bakalım

```
X_test.loc[101283]
```

```
RevolvingUtilizationOfUnsecuredLines    1.117520
age                                       51.000000
NumberOfTime30-59DaysPastDueNotWorse    5.000000
DebtRatio                                0.807154
MonthlyIncome                           4500.000000
NumberOfDependents                       1.000000
CombinedDefaulted                       1.000000
CombinedCreditLoans                     1.000000
Name: 101283, dtype: float64
```

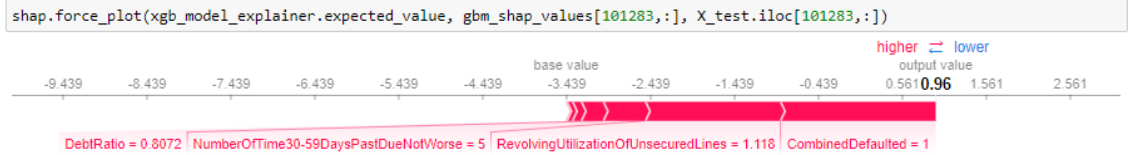
```
Id            101284.000000
Probability    70.96765
Name: 101283, dtype: float64
```

Risk olasılığı çok yüksek bu yüzden kredi alamaz almama sebeplerine bakalım.



Kredi sonucu üzerinde 3 önemli özellik var:

Dilerseniz kuvvet grafiğine bakalım



Mavi değerler risk oluşturmuyor demek. Kırmızı ise risk var oluşturuyor demek oluyor.

Ödemelerde yapılan gecikmeler ve Güvenli Olmayan Hatların Döner Kullanımı çok olduğu için kredi alamadı.

