



**CustomsDev**

# **CustomsDev team**

**CBU Coding Challenge**



# Muammo hamda ma'lumotlar holati



## Muammo

Mijoz berilgan kreditni  
to'lay olish qibilyatini  
to'g'ri aniqlash



Bankning moliyaviy  
yo'qotishni oldini olish



**90** ming  
kreditga arizalar

**5.1** %  
4,594 default natija



## Maqsad

**Metrika:** AUC-ROC Score

**Data:** 6 formatda fayllar

### Dataset tarkibi

**CSV files**  
application,  
demographics

**Parquet**  
credit\_history

**JSONL**  
financial\_ratios

**XML**  
geographic\_data

**XLSX**  
loan\_details

**Target**  
default (0/1)



## Data cleaning

- **Ma'lumotlarni umumlashtirish ustun ma'lumotlarini birxillashtirish**
- **Tushirib qoldirilgan ustun ma'lumotlarni to'ldirish va standartlashtirish**
- **Datatype'ga tekshirish hamda formatlarini standartlashtirish**
- **Model uchun kerakli ustunlarni ajratib olish**



# Model tanlash va o'qitish



CatBoost

## Kategorik ma'lumotlar va tezlik.

- Kategorik xususiyatlarni avtomatik ishlatish.
- Tez o'qish va yuqori samaradorlik.
- Overfittingni kamaytirish.
- Xususiyat ahamiyatini intuitiv tushunish.
- Kuchli parallelizatsiya.

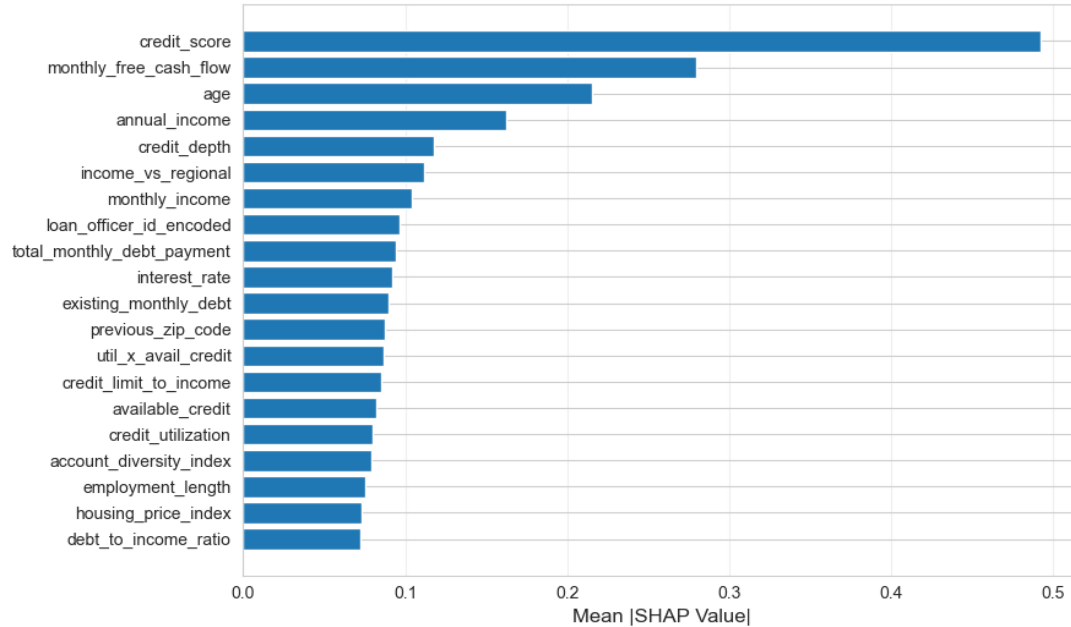
*dmlc*  
**XGBoost**

## Aniqlik, imbalansli ma'lumotlar va regularizatsiya.

- Imbalance ma'lumotlar bilan yaxshi ishlash.
- Regularizatsiya (overfittingni kamaytiradi).
- Qisqa va uzoq muddatli ma'lumotlar bilan samarali ishlash.
- Xususiyatlar o'rtasidagi o'zaro aloqalarni o'rganish.



**XGBoost - Top 20 Features by SHAP Importance**



### --- Evaluating XGBoost

**Accuracy:** 0.9257

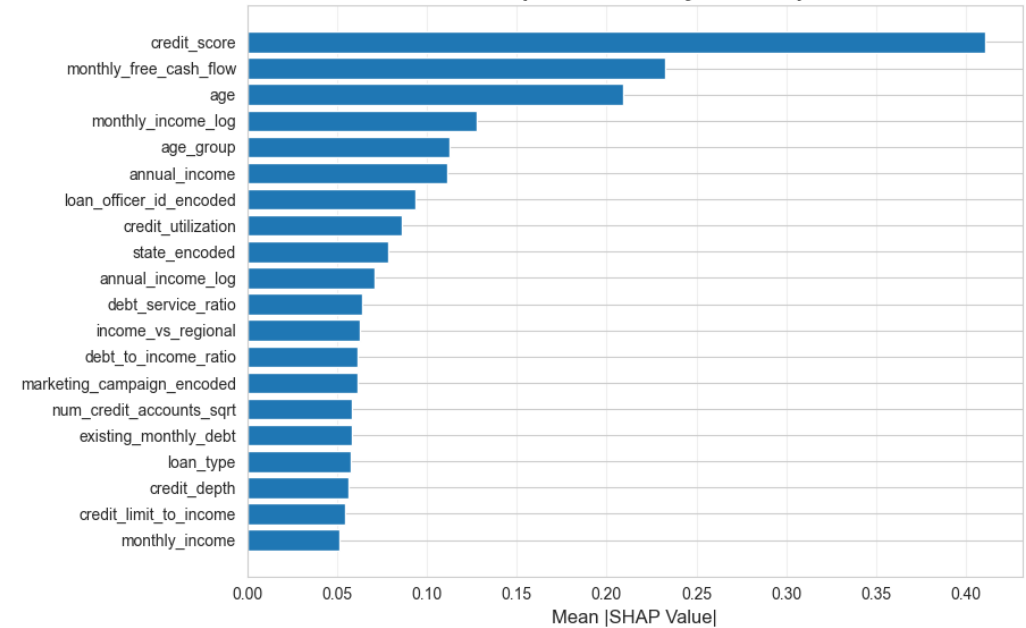
**Precision:** 0.2825

**Recall:** 0.2960

**F1-Score:** 0.2891

**ROC-AUC:** 0.7863

**CatBoost - Top 20 Features by SHAP Importance**



### --- Evaluating CatBoost

**Accuracy:** 0.8888

**Precision:** 0.2160

**Recall:** 0.4483

**F1-Score:** 0.2916

**ROC-AUC:** 0.7960



Jamoa



Data  
Scientist



ML  
Engineer



Data  
Analitic



ML  
Engineer



Data  
Engineer



ML  
Engineer



CustomsDev