

# CBU coding challenge

**Maqsad:**Ma'lumotlar bo'yicha defolt ehtimolini prognoz qilish modelini ishlab chiqish.

# CatBoost

- LR 
- RF 
- XGBoost 
- LightGBM 

- Aralash ma'lumotlar bilan ishlash
- AUC ~ 80
- Overfitting kam
- Imbalanced data
- Murakkab bog'liqliklarni aniqlaydi
- Outlierlar bilan yaxshi ishlaydi
- Yetarlich tez

# Model qanday ishlaydi ?

- Ma'lumotlarni tozalaydi, kategoriyalarni tartibga keltiradi, ekstremal qiymatlarni nazorat qiladi
- Eng muhim feature(belgilar)ni 25 tasini tanlab oladi
- CatBoost shu belgilar o'rtasidagi murakkab bog'liqlikni avtomatik o'rganadi
- Har bir mijoz uchun 0-1 oraliqdagi default ehtimolini hisoblaydi
- Isotonic Calibration orqali ehtimollarni real hayotdagi risk darajasi bilan moslashtiradi
- Yakunda tizim har bir mijoz uchun aniq, tez va barqaror risk bahosini beradi

# Muammolar va yechimlar:

- Imbalanced ma'lumotlar to'plami  $\rightarrow$  class\_weight + threshold optimization
- Outliers  $\rightarrow$  IQR(Inter Quartal Range)
- Noisy data  $\rightarrow$  Ordered Boosting(NOISE-ROBUST)
- Ehtimollik noaniqligi  $\rightarrow$  Calibration

# **Bu model qanday muammolarga yechim bo'лади:**

- Subyektiv-kredit qarorlar
- Default yuqori bo'lishi
- Risk segmentatsiyasi sustligi
- Limit va Stavka noto'g'ri hisoblanishi
- Tezlik pastligi