AGE MODEL

By Marharyta Babych

AGENDA

- Business problem
- Executive summary
- Data understanding
- Data preprocessing
- Model
- Lessons-learned report

BUSINESS PROBLEM

The main goal – create a **model** which **forecasts the age-bins** of customers.
Age bins: <20, 20-30, 30-40, 40-50, >50

The model will facilitate work for many departments.

EXECUTIVE SUMMARY

- FE
- DPI
- BNUM

3 Datasets

≈ 13% used features

Catboost

50%

DATA UNDERSTANDING

Datasets description

FE

- Usage Metrics,
- Financial Metrics
- Plan Information
- User Activity
- Interaction and Contacts
- Device Information

SIZE: 120 060 601
MISSING: 11%

DPI

- Application
- Duration of Sessions
- Volume of Data
- Number of Sessions
- Number Using Days

SIZE: 42 350 676
MISSING: 0%

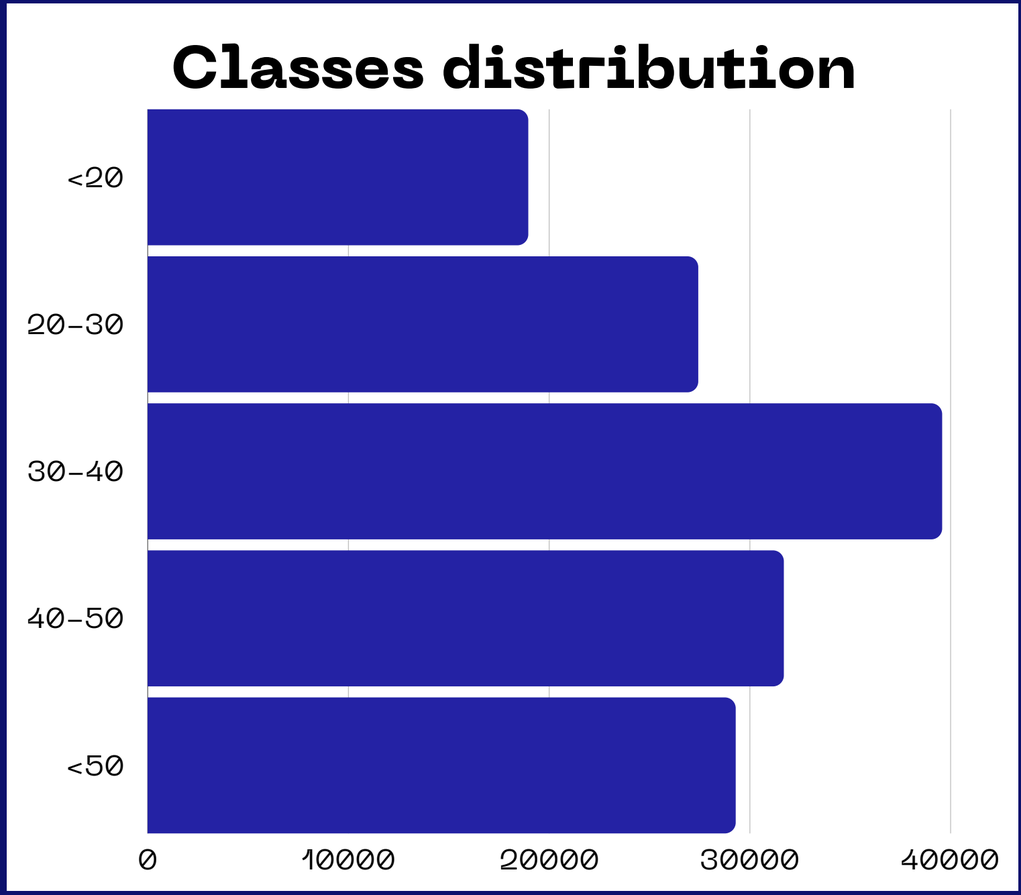
BNUM

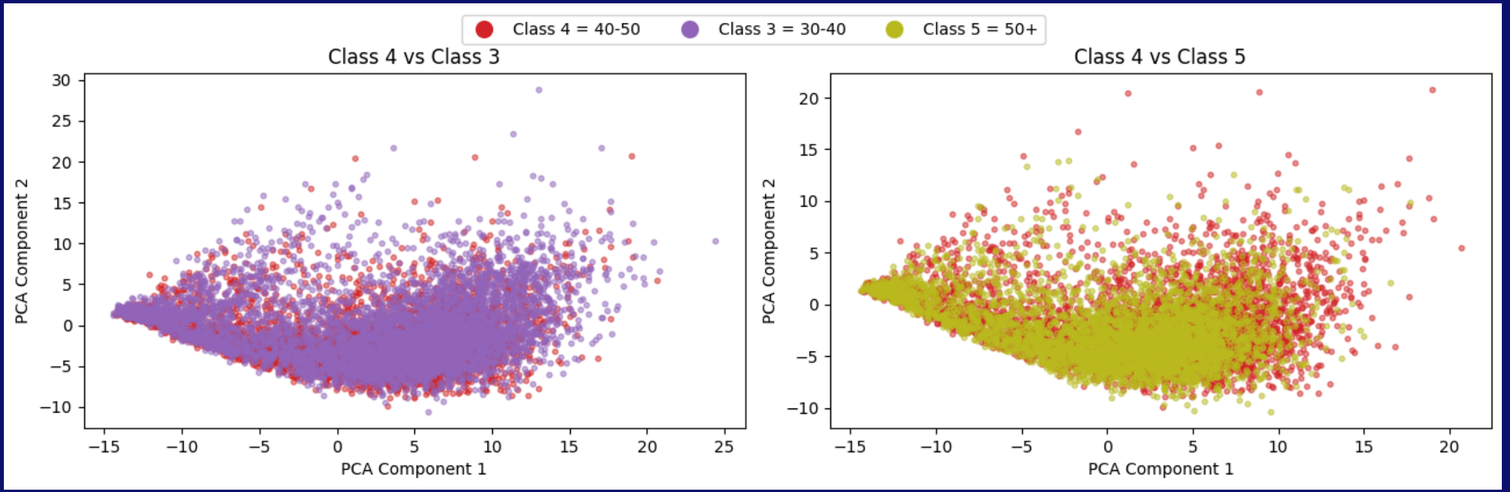
- Contact Information
- Call Data
- Message Data

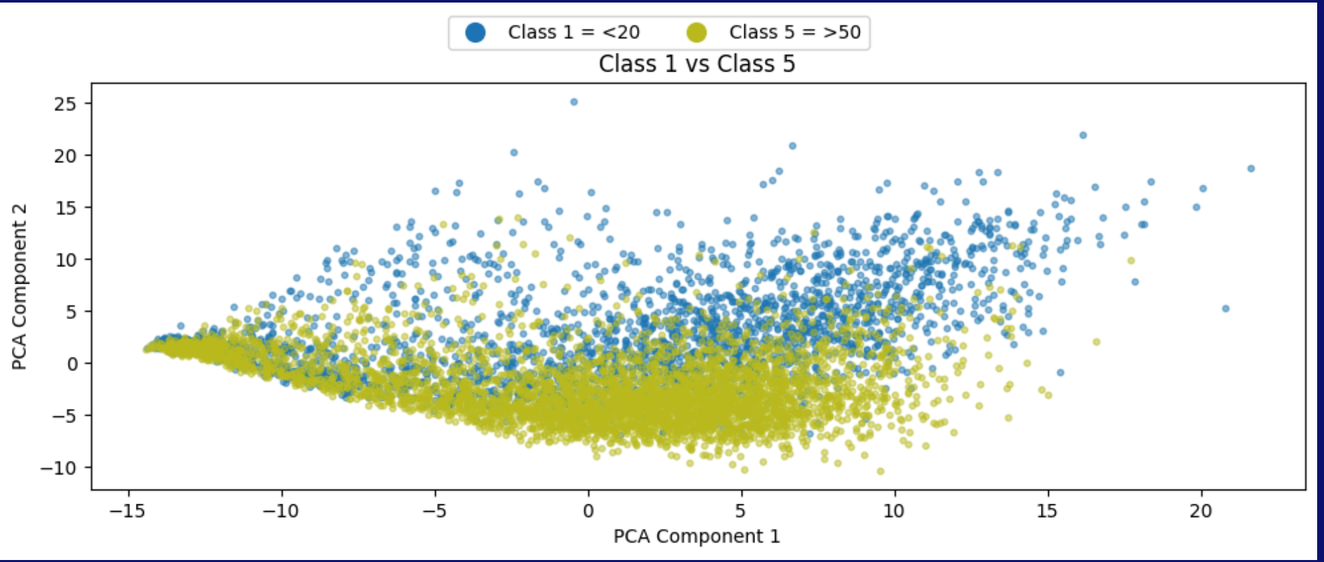
SIZE: 6 426 144
MISSING: 0%

Target overview

Classes visualization







DATA PREPROCESSING

CLEAN FEATURE:

- >75% MISSING
- VARIOUS < 1%

FILL MISSING:

- MODE
- MEDIAN

FEATURE SELECTION

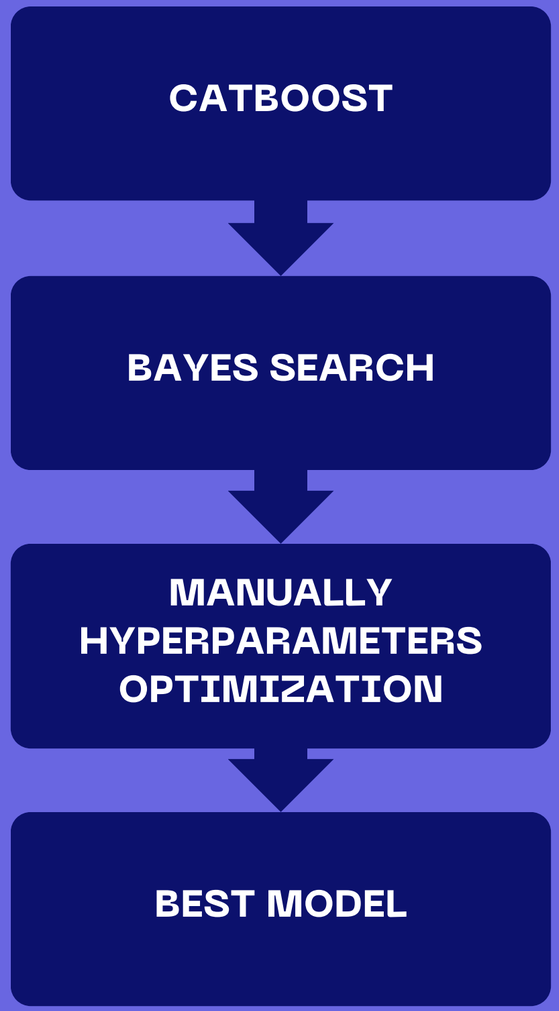
- ANOVA
- MUTUAL INFORMATION
- RANDOM FOREST

RESULT: 224 FEATURES SELECTED (THE BIGGEST PART IS INFORMATION ABOUT APPLICATIONS)

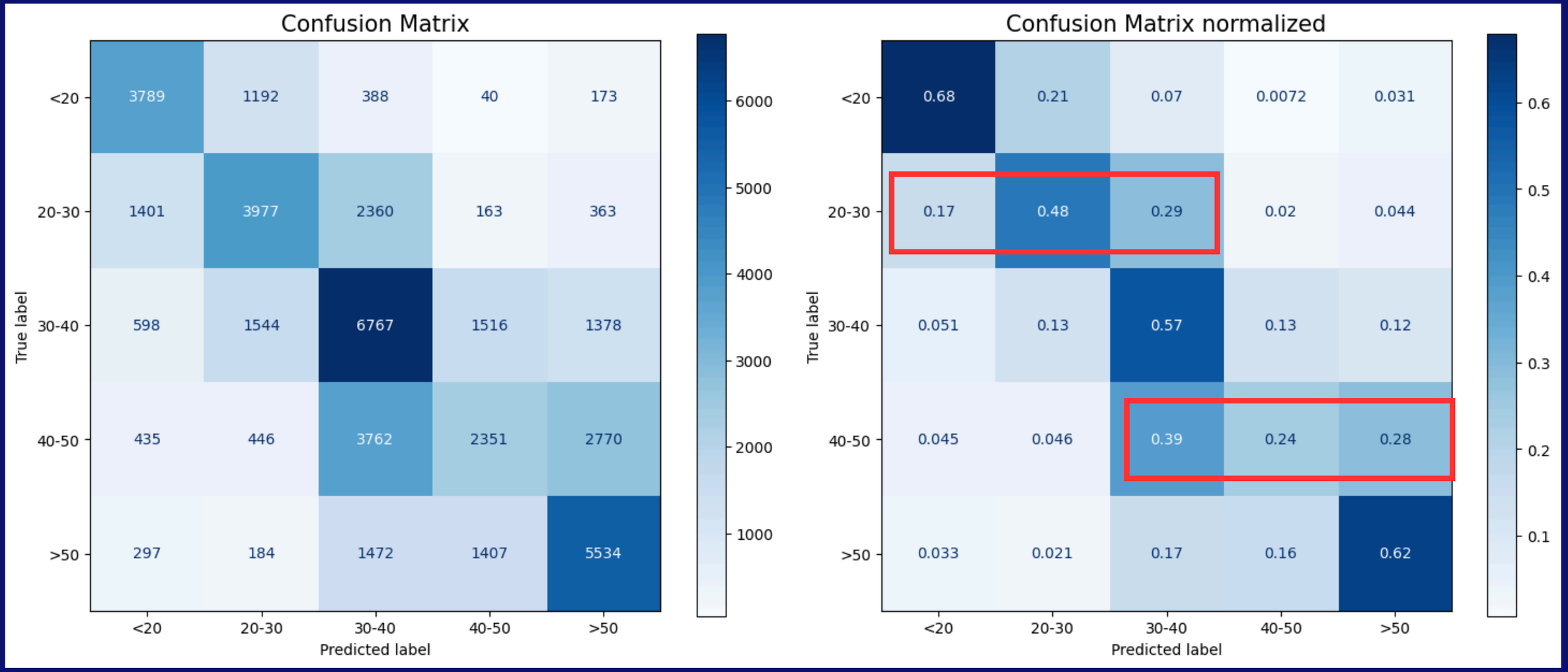
MODEL

PROBLEM DEFINITION

- ACCURACY MIN. 0.45
- AVOID OVERFITTING
- HIGH DISTINGUISH BETWEEN 1 AND 5 CLASSES
- GENERAL CLASSIFICATION ABILITY BETWEEN CLASSES(ROC CURVE)

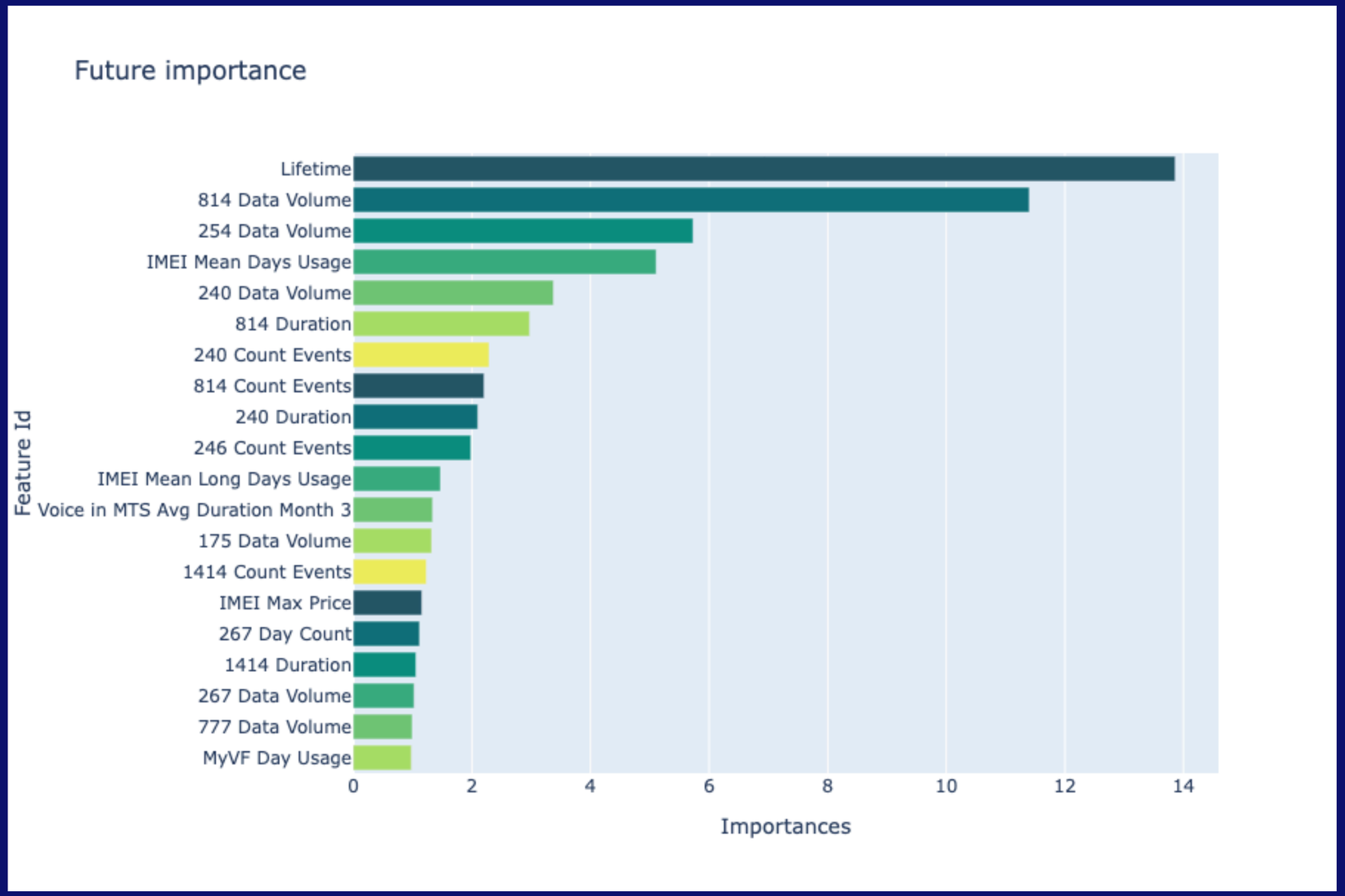


MODEL EVALUATION



ACCURACY: 50%
BALANCED ACCURACY: 52%

MODEL ANALYSIS



FEATURES WITH A NUMBER IN THE NAME MEANS APPLICATIONS CODES

- DATA VOLUME = APP DATA VOLUME IN/OUT
- DURATION = SESSION DURATION IN APP
- COUNT EVENTS = SESSION COUNT IN APP
- DAY COUNT = DAYS IN APP

HOW FEATURES DEFINE CLASS

younger 20	20–30	30–40	40–50	older 50
Life time – low	IMEI Mean Days Usage –low	Life time – high	Life time – high	Life time – high
Apps 240, 246 – low	**IMEI max price – high**	IMEI Mean Days Usage –high	**App 254 – high**	**Balance – high**
App 814 – high	App 814 – high	**App 254 – high**	App 814 – low	IMEI Mean Days Usage –high

WAYS TO IMPROVE RESULTS

1. CHANGE GROUP IN LESS AMOUNT: MERGE CLIENTS 30-40 AND 40-30 YEARS.

2. TRY TO CLASSIFY CLIENTS BY BEHAVIOUR, NOT BY AGE(CLUSTERIZATION).

THANK YOU

EMAIL: MARHARYTABABYCH1@GMAIL.COM
LINKEDIN: MARHARYTA BABYCH