

Task #1 Модель відтоку

Бізнес ціль:

Утримання абонентів за 2 місяці до їх фактичного відтоку/неактивності (без вхідних, вихідних та платних подій протягом місяця)

Задача: Визначити абонентів, які через 2 місяці підуть і відток

Model Target: (target)

1- Активний абонент через 2 місяці став неактивним

0- Активний абонент через 2 місяці залишився активним

Evaluation:

Метриками якості моделі визначені показники AUC, Classification report, Confusion matrix на Test Sample.

За базову метрику для порівняння якості моделей студентів обрана AUC з базовим результатом

(Baseline model) 0,89 на Test Sample.

Опис даних в файлі «Описание данных.xlsx»

Sample size

Train
50%
150K

Test
50%
150K



Task #1 Модель відтоку

Дані:

Train Sample

TRAIN_CHURN_MODEL_FE (Target - розмітка) – абонентська вітрина

TRAIN_CHURN_MODEL_B_NUM – дані про короткі номери

TRAIN_CHURN_MODEL_DPI – трафік мобільних додатків

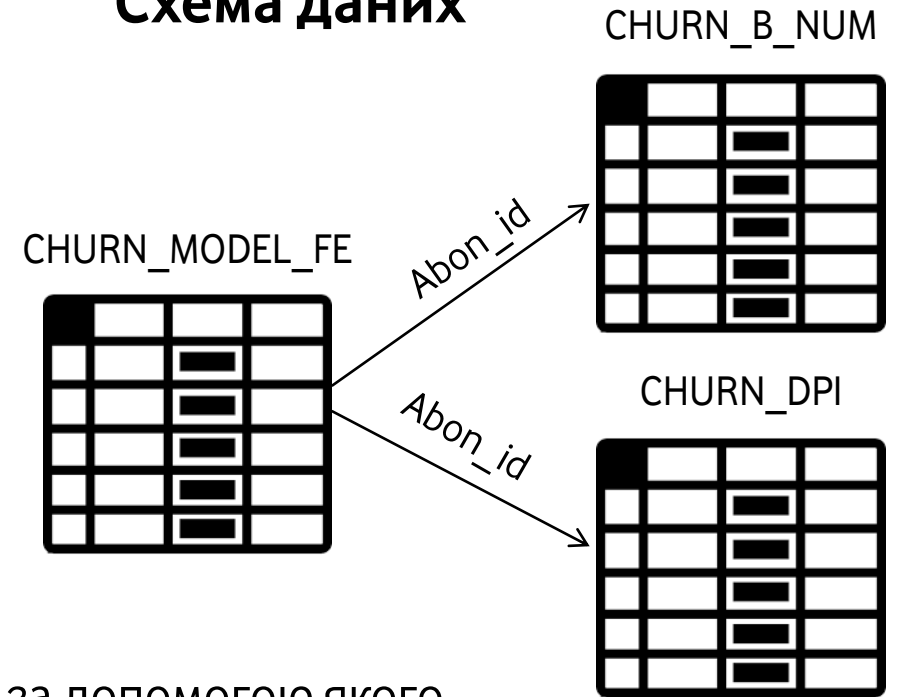
Test Sample

TEST_CHURN_MODEL_FE (Target - розмітка) – абонентська вітрина

TEST_CHURN_MODEL_B_NUM – дані про короткі номери

TEST_CHURN_MODEL_DPI – трафік мобільних додатків

Схема даних



Результатом роботи є Jupyter Notebook з коментарями до коду, за допомогою якого прогнозується результат на Test Sample та **презентації** (графіки, таблиці) результатів з орієнтацією на бізнес (не технічних) користувачів, що має включати наступні пункти: feature explorer, feature encoding, спосіб валідації, best model (best algorithm/best metrics), підхід до підбору hyperparameters model, features importance, shap



Task #2 Модель визначення вікової категорії

Бізнес ціль:

Визначення вікової категорії для доповнення портрету абонентів-користувачів послуг та для використання в рамках пропозицій цільового маркетингу

Задача: Визначити вікову категорію абонентів

Model Target: (target)

- 1 – Age group <20 year
- 2 – Age group 20-30 year
- 3 – Age group 30-40 year
- 4 – Age group 40-50 year
- 5 – Age group >50 year

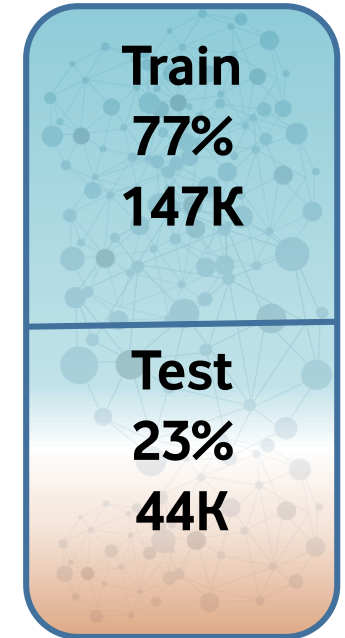
Evaluation:

Метриками якості моделі визначені показники AUC, Classification report, Confusion matrix на Test Sample.

За базову метрику для порівняння якості моделей студентів обрана Accurasy з базовим результатом (Baseline model) 0,45 на Test Sample.

Опис даних в файлі «Описание данных.xlsx»

Sample size



Task #2 Модель визначення вікової категорії

Дані:

Train Sample

TRAIN_AGE_MODEL_FE (Target - розмітка) – абонентська вітрина

TRAIN_AGE_MODEL_B_NUM – дані про короткі номери

TRAIN_AGE_MODEL_DPI – трафік мобільних додатків

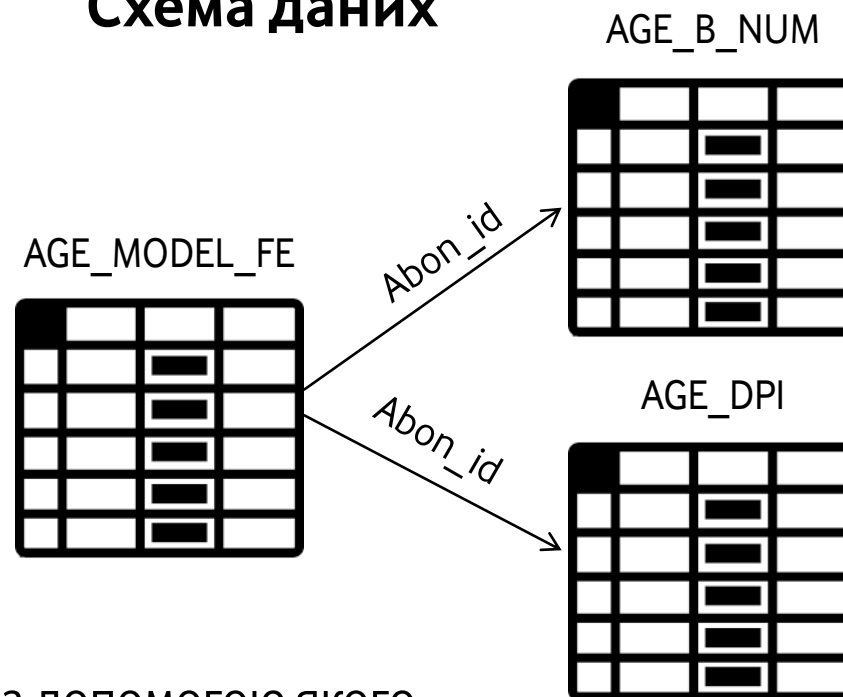
Test Sample

TEST_AGE_MODEL_FE (Target - розмітка) – абонентська вітрина

TEST_AGE_MODEL_B_NUM – дані про короткі номери

TEST_AGE_MODEL_DPI – трафік мобільних додатків

Схема даних



Результатом роботи є Jupyter Notebook з коментарями до коду, за допомогою якого прогнозується результат на Test Sample та **презентації** (графіки, таблиці) результатів з орієнтацією на бізнес (не технічних) користувачів, що має включати наступні пункти: feature explorer, feature encoding, спосіб валідації, best model (best algorithm/best metrics), підхід до підбору hyperparameters model, features importance, shap



Task #3 Модель прогнозу приросту річних витрат абонентів

Бізнес ціль:

Визначити динаміку зміни річних витрат абонентів (збільшення, зменшення, стабільність) для підготовки індивідуальних пропозицій

Задача: Спрогнозувати приріст річних витрат абонентів

Model Target: (target)

Відношення річних витрат в наступному році до річних витрат минулого року

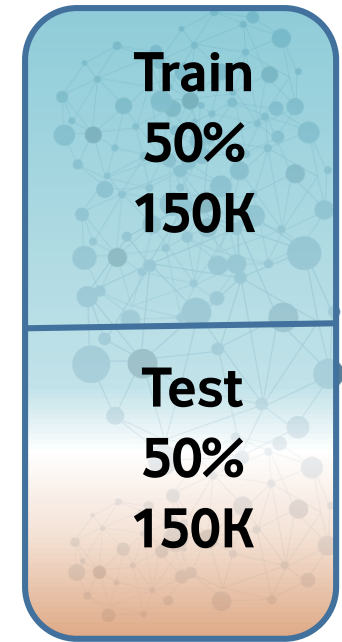
Evaluation:

Метриками якості моделі визначені показники Root Mean Squared Error (RMSE), Mean Absolut Error (MAE), R2 на Test Sample.

За базову метрику для порівняння якості моделей студентів обрана RMSE з базовим результатом (Baseline model) 0,58 на Test Sample.

Данные описаны в файле «Описание данных.xlsx»

Sample size



Task #3 Модель прогнозу приросту річних витрат абонентів

Дані:

Train Sample

TRAIN_LTV_MODEL_FE (Target - розмітка) – абонентська вітрина

TRAIN_LTV_MODEL_B_NUM – дані про короткі номери

TRAIN_LTV_MODEL_DPI – трафік мобільних додатків

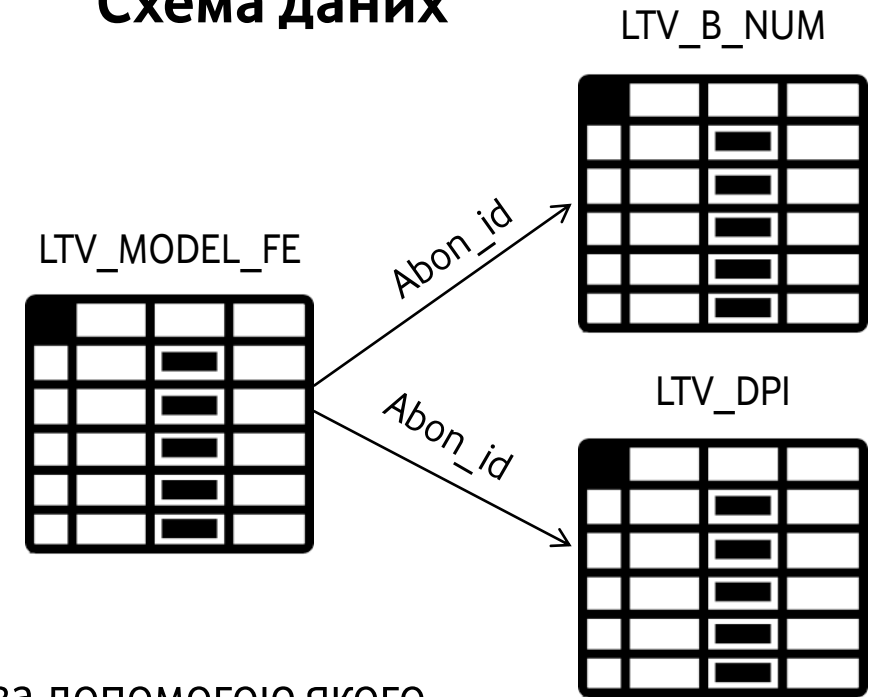
Test Sample

TEST_LTV_MODEL_FE (Target - розмітка) – абонентська вітрина

TEST_LTV_MODEL_B_NUM – дані про короткі номери

TEST_LTV_MODEL_DPI – трафік мобільних додатків

Схема даних



Результатом роботи є Jupyter Notebook з коментарями до коду, за допомогою якого прогнозується результат на Test Sample та **презентації** (графіки, таблиці) результатів з орієнтацією на бізнес (не технічних) користувачів, що має включати наступні пункти: feature explorer, feature encoding, спосіб валідації, best model (best algorithm/best metrics), підхід до підбору hyperparameters model, features importance, shap

