

# **Курсовой проект по курсу «Видеокурс от Мегафон»**

# 1. Информация о модели, ее параметрах, особенностях и основных результатах.

В качестве рабочей модели выбрана продвинутая библиотека градиентного бустинга на деревьях решений с открытым исходным кодом от компании Яндекс – **CatBoost**.

**CatBoost** выбран потому, что по сравнению с другими рассматриваемыми моделями показал наивысший показатель целевой метрики **f1\_score(..., average='macro') = 0,69**

Данный результат был достигнут путём подбора оптимальных параметров при помощи **RandomizedSearchCV**, а также балансировки данных.

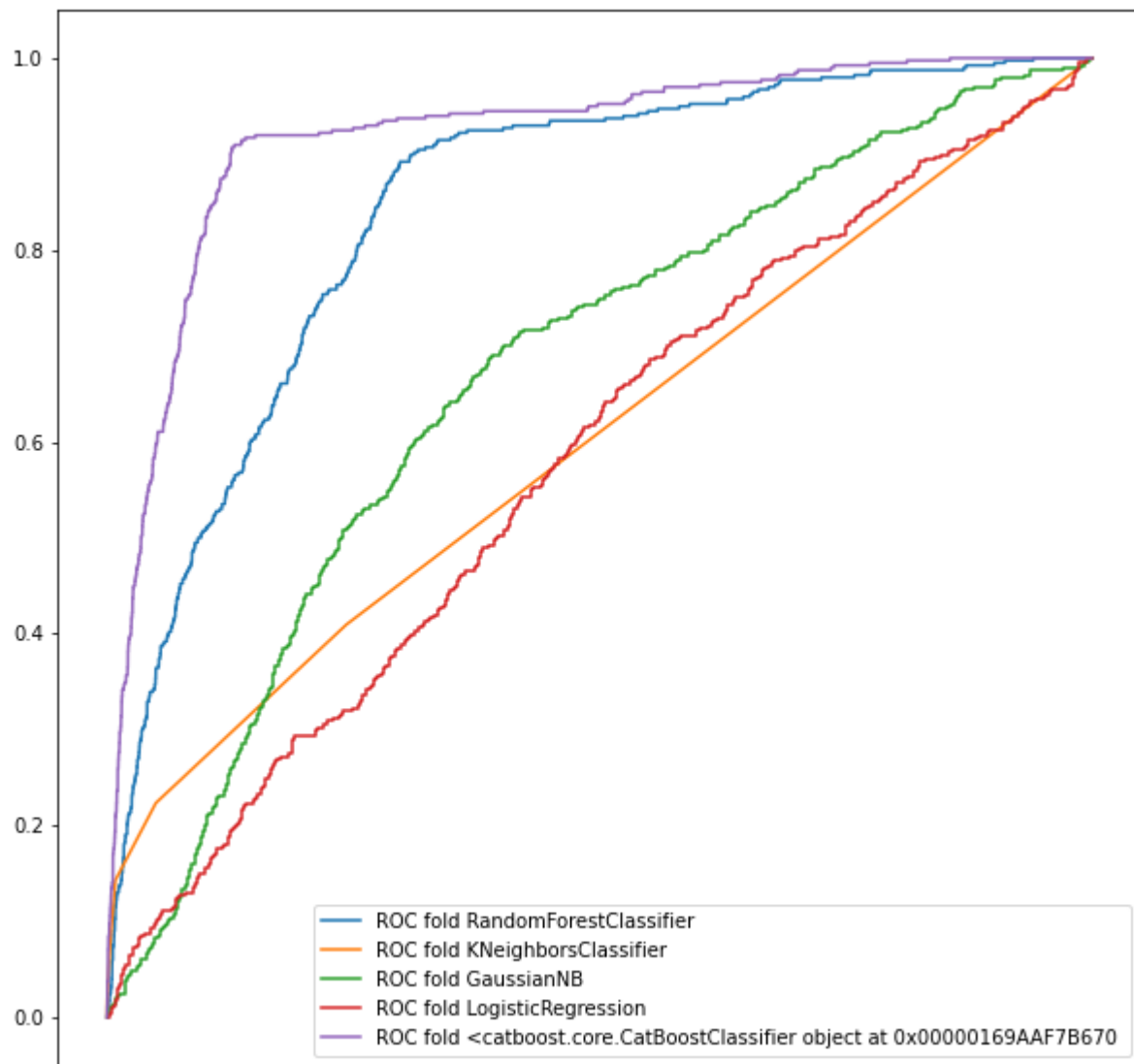
Оптимальные параметры модели: {'n\_estimators': 200, 'max\_depth': 5, 'learning\_rate': 0.1, 'l2\_leaf\_reg': 7}

Оптимальный коэффициент балансировки: ADASYN(sampling\_strategy=0.6)

Более подробно можно посмотреть в файле **Megafon\_course\_project.ipynb**

## 2. Обоснование выбора модели и ее сравнение с альтернативами.

Сравнение рассматриваемых моделей по ROC-кривым



Сравнение по f1\_score на тестовой выборке

MODEL NAME: \*\*\*\*\* RandomForestClassifier \*\*\*\*\*

	precision	recall	f1-score	support
macro avg	0.47	0.50	0.48	5729

MODEL NAME: \*\*\*\*\* KNeighborsClassifier \*\*\*\*\*

	precision	recall	f1-score	support
macro avg	0.75	0.57	0.60	5729

MODEL NAME: \*\*\*\*\* GaussianNB \*\*\*\*\*

	precision	recall	f1-score	support
macro avg	0.53	0.50	0.07	5729

MODEL NAME: \*\*\*\*\* LogisticRegression \*\*\*\*\*

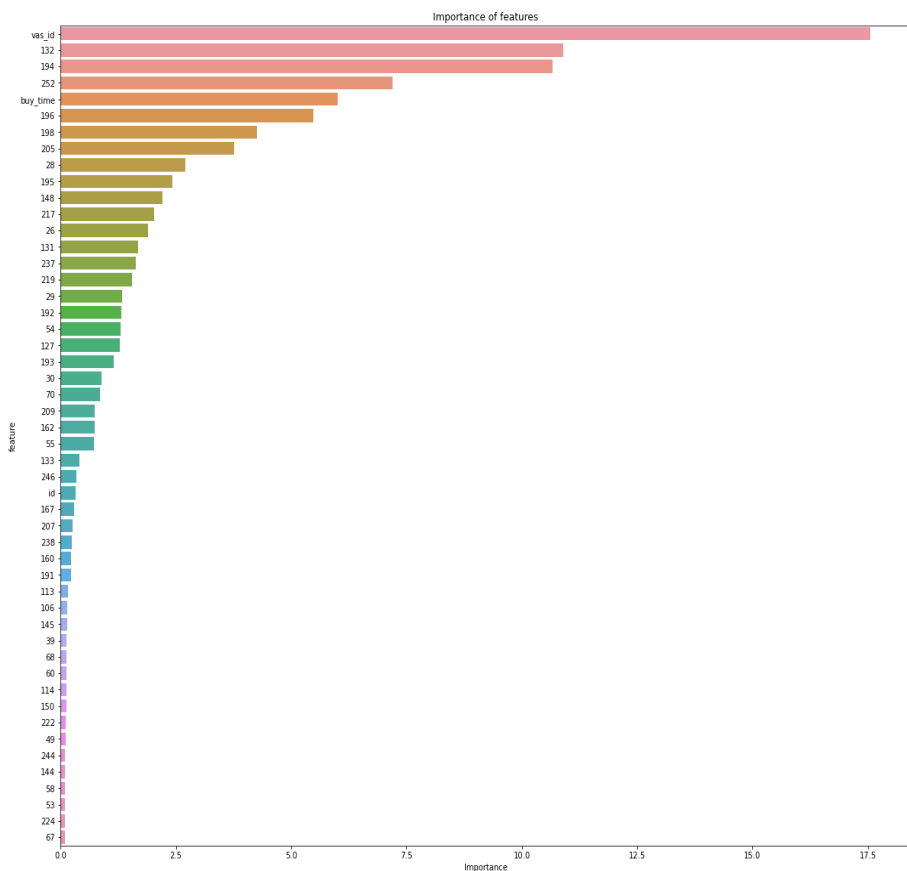
	precision	recall	f1-score	support
macro avg	0.47	0.50	0.48	5729

MODEL NAME: \*\*\*\*\* CatBoostClassifier \*\*\*\*\*

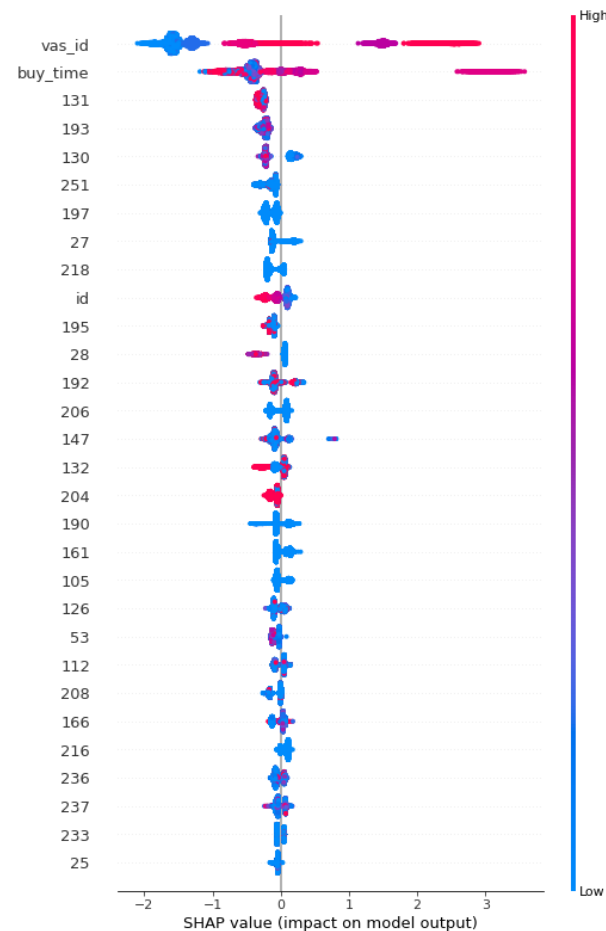
	precision	recall	f1-score	support
macro avg	0.79	0.59	<b>0.63</b>	5729

Модель на основе CatBoost показала наилучшие результаты в обоих случаях (ROC-кривые и метрика f-1), что послужило основанием для выбора её как рабочей модели. Ниже приведены некоторые характеристики модели:

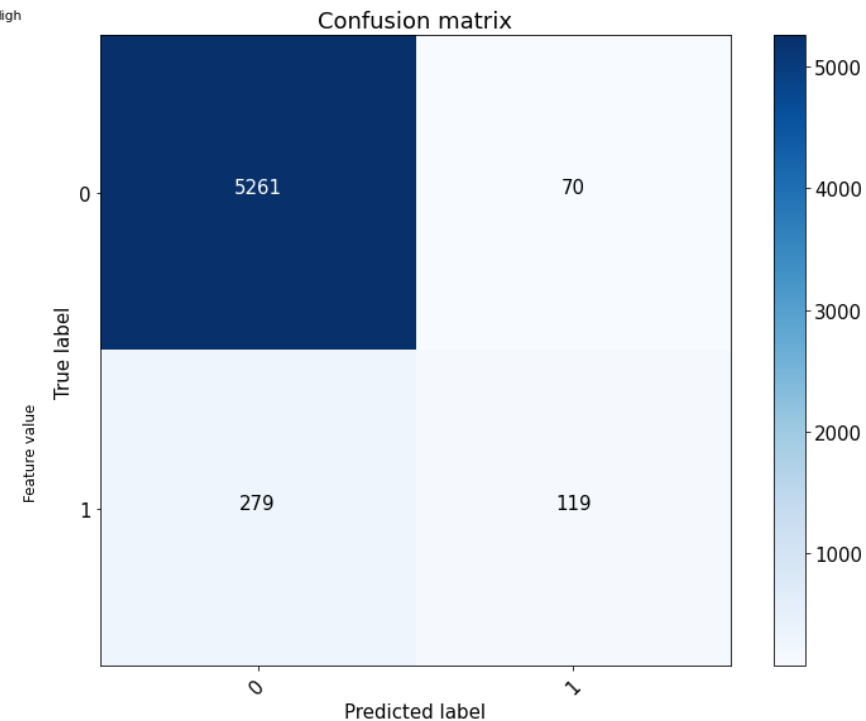
**Большинство признаков не важны**



**Услуга и время покупки – основное влияние**



**Высокий показатель TN**



### 3. Принцип составления индивидуальных предложений для выбранных абонентов.

В задании к курсовому проекту требовалось записать в файл “**answers\_test.csv**” и в качестве таргета использовать вероятность подключения услуги, что и было сделано.

На основании полученных вероятностей можно составить индивидуальные рекомендации. Для этого вероятности из столбца “target” были преобразованы в двоичный вид, где 1 – отправляем рекомендацию, 0 – не отправляем.

Кроме того была введена переменная “threshold” (по умолчанию threshold = 0,5) эта та вероятность выше которой столбец “target” принимает значение 1, а ниже 0. Всё что требуется это определиться с порогом.

Целесообразным видится использование порога вероятности 0,3 т.к. высокий TN данной модели эффективно отсеивает отправку рекомендаций негативно настроенным клиентом. В то же самое время вероятность 0,3 показывает как минимум незначительную заинтересованность клиента в услуге и отправка 1-2 смс навряд ли повлечёт за собой отказ клиента от услуг компании Мегафон, но однозначно проинформирует, например, о новом тарифном плане и возможно клиент сочтёт эту информацию полезной и подключит услугу.