

Курсовой проект

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

Используемая модель – CatBoost.

Данная модель выбрана исходя из наилучшего показателя целевой метрики – 0,69.

Для достижения данного показателя использовались:

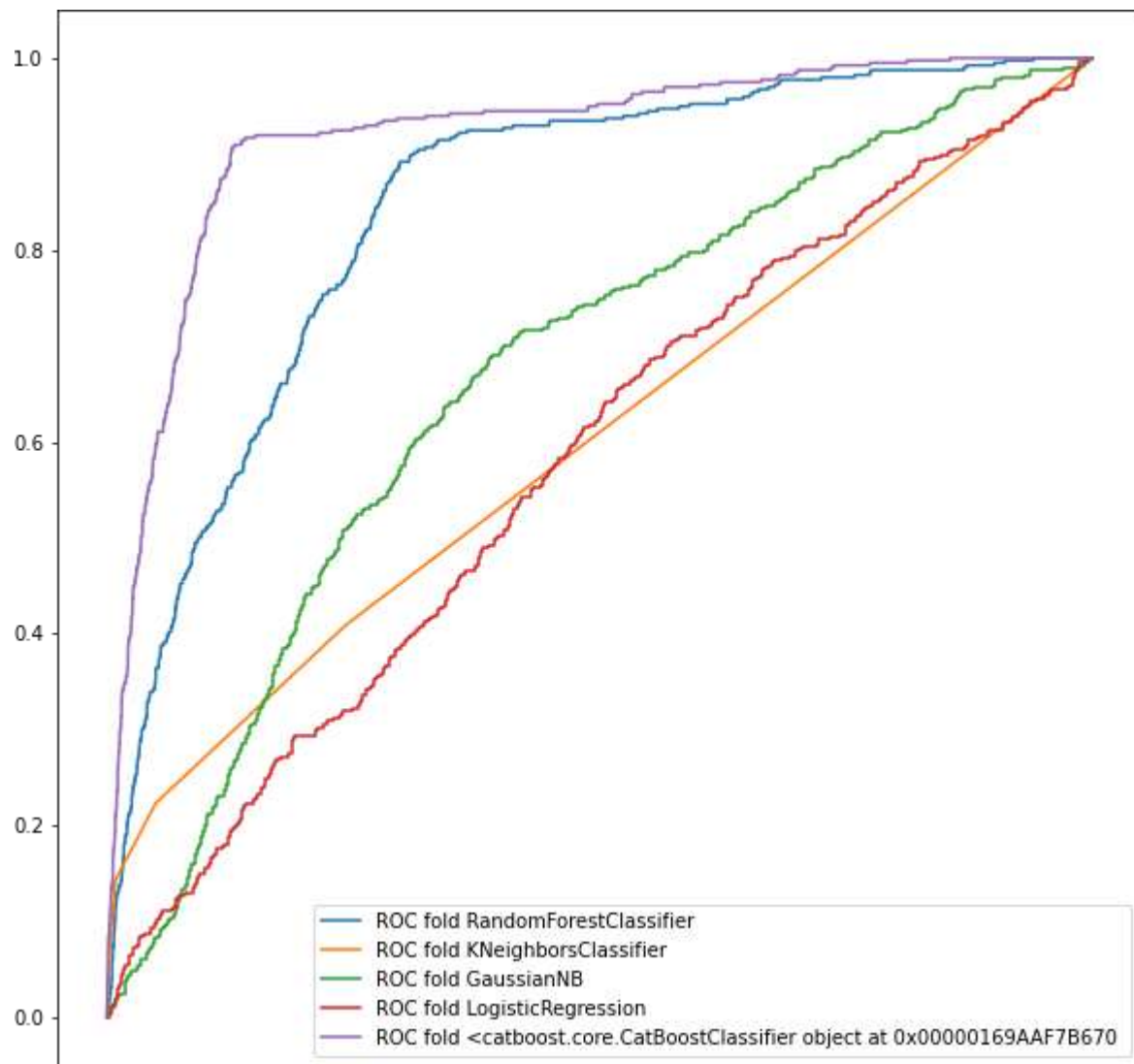
- подбор оптимальных параметров при помощи RandomizedSearchCV;
- балансировка данных.

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)

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 *******

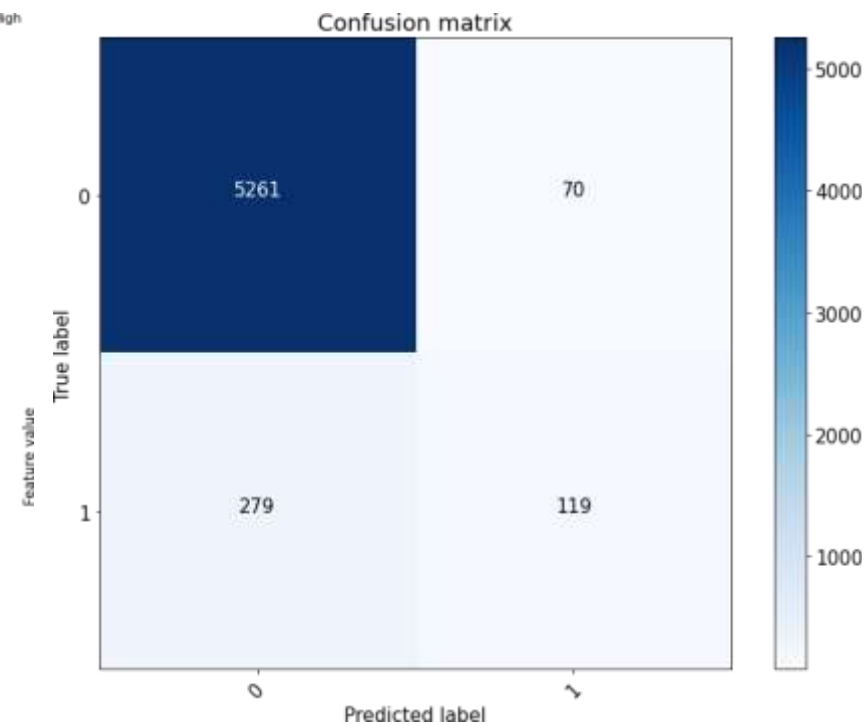
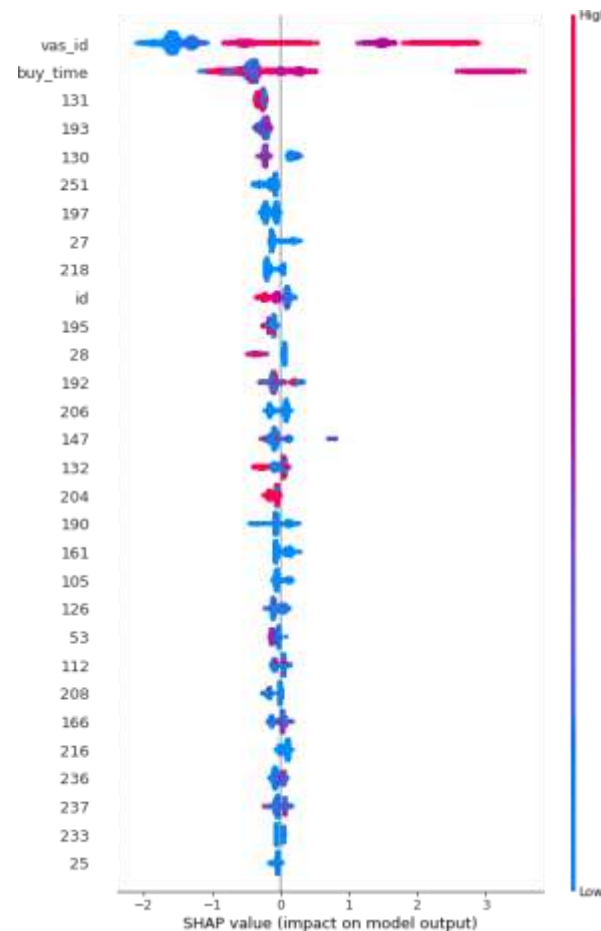
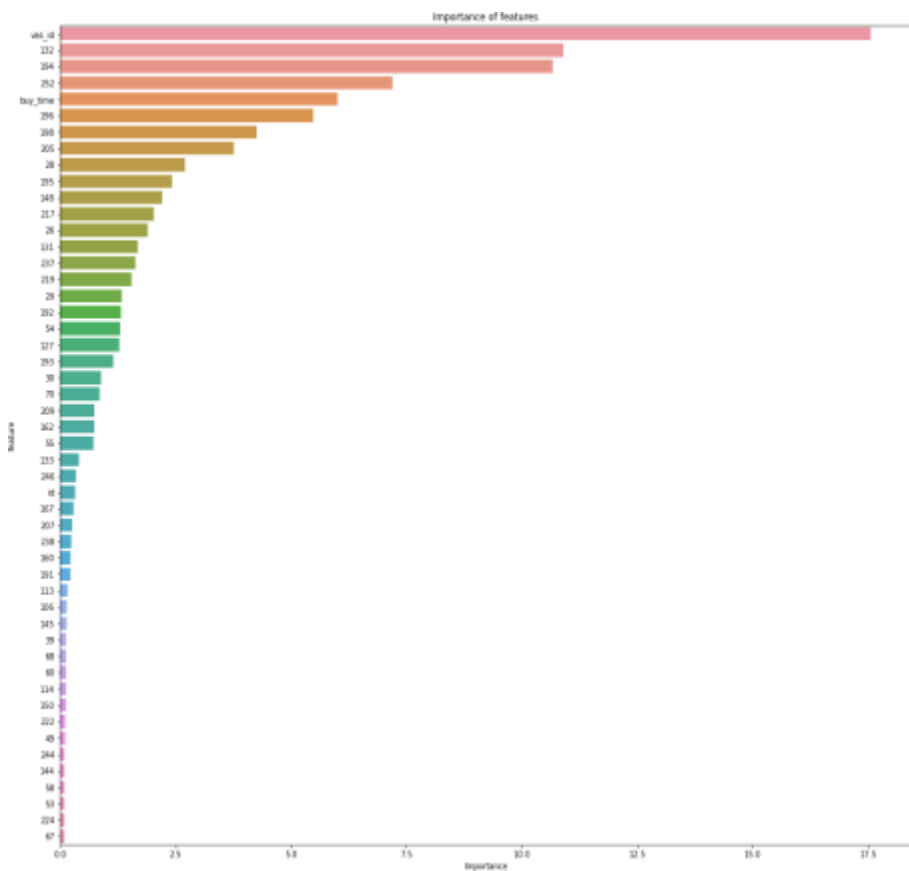
	precision	recall	f1-score	support	macro avg
	0.79	0.59	0.63	5729	

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

Большинство признаков не имеют критического значения

Сильное влияние на значения: услуги и время покупки

Высокий показатель TN



3. Принцип составления индивидуальных предложений для выбранных абонентов.

На основании полученных вероятностей можно составить индивидуальные рекомендации:

Столбец `target` был преобразован в бинарный вид, где 1 – отправка рекомендации, 0 – рекомендация не отправляется.

Использование пороговой переменной `threshold`, в зависимости от значения которой в столбце `target` выявляется необходимость отправки рекомендаций.

По умолчанию значение `threshold = 0,5`, но целесообразнее уменьшение данного значения до 0,3, что позволит увеличить охват абонентов.