

ОТТОК ПОЛЬЗОВАТЕЛЕЙ ОТЧЕТ

Николай Крамаренко



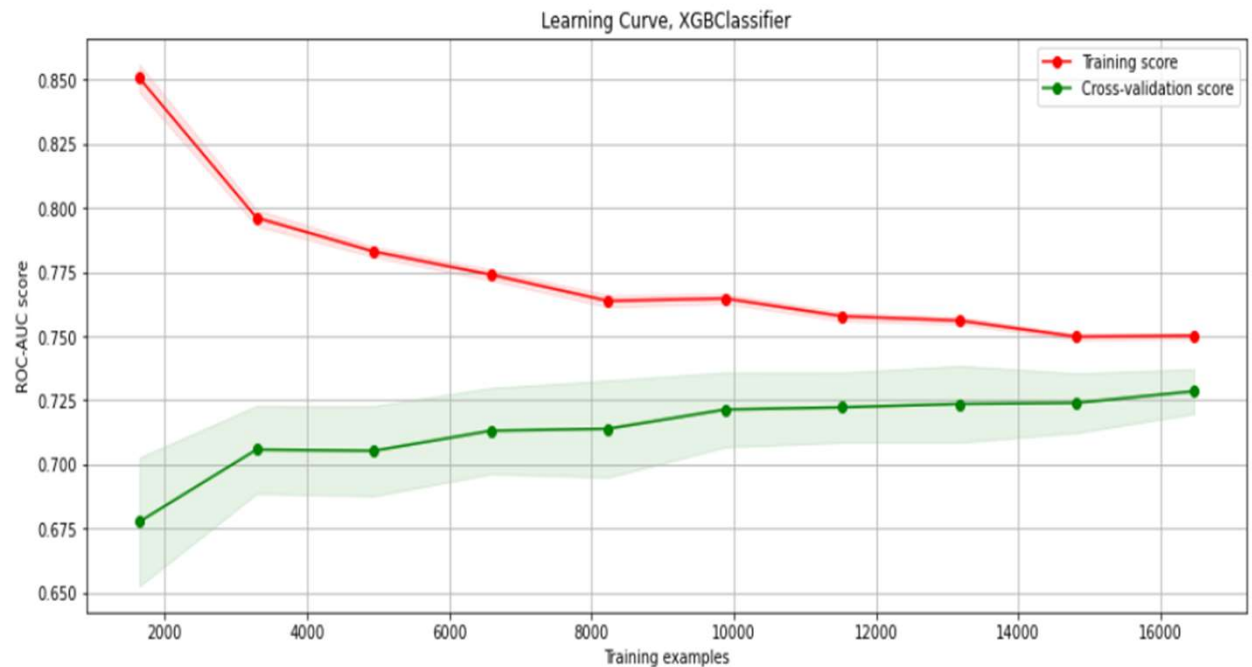
ЦЕЛИ И ЗАДАЧИ ПРОЕКТА

Отток пользователей - одна из наиболее актуальных задач в областях, где распространение услуги практически полностью охватывает всех пользователей. Поэтому особенно важным становится именно удержание клиентов. Яркий примеры таких областей - телекома, банковская сфера.

Цель проекта - научиться находить пользователей, склонных к оттоку. Если научиться находить таких пользователей с достаточной точностью заблаговременно, то можно эффективно управлять оттоком: например, выявлять причины оттока; помогать пользователям, попавшим в группу риска, решать их проблемы и задачи; проводить кампании по удержанию. Для того, чтобы находить пользователей, склонных к оттоку, строят прогнозные модели - модели, позволяющие прогнозировать вероятность того, что пользователь покинет сервис.

ОЦЕНКА МОДЕЛИ И КРИТЕРИЙ УСПЕШНОСТИ

Из-за сильного дисбаланса данных, в качестве основной метрики для оценки качества модели будем использовать величину ROC-AUC.



Критерием успешности будет повышение доходов за счет использования модели.

ТЕХНИЧЕСКОЕ ОПИСАНИЕ РЕШЕНИЯ

В ходе построения модели были опробованы следующие стратегии:

1. Балансировка выборки: задание различных весов объектам, undersampling.
2. Заполнение пропущенных значений: нулями, средними значениями, наиболее часто встречающимися значениями.
3. Обработка категориальных признаков: Label encoding, One-hot encoding, смешанная обработка.
4. Регуляризация: Lasso.
5. Классификаторы: RidgeClassifier, SGDClassifier, RandomForestClassifier, XGBClassifier.
6. Настройка с помощью GridSearchCV гиперпараметров классификатора.

```
weights = 1.00, ROC-AUC = 0.7285  
weights = 0.29, ROC-AUC = 0.7277  
weights = 0.08, ROC-AUC = 0.728  
weights = 0.03, ROC-AUC = 0.7267
```

```
cv_scores_mean = cross_val_score(  
    xgb_class, X_train_mean, y,  
    cv=cv, scoring='roc_auc', n_jobs=-1)  
print('ROC-AUC', cv_scores_mean.mean())
```

ROC-AUC 0.7326105158464555

```
parameters_grid = {  
    'n_estimators': [50, 100, 200],  
    'max_depth': [1, 2, 3],  
    'learning_rate': [0.1, 0.2]  
}
```

```
{'learning_rate': 0.1, 'max_depth': 2, 'n_estimators': 100}
```

ИТОГОВАЯ МОДЕЛЬ:

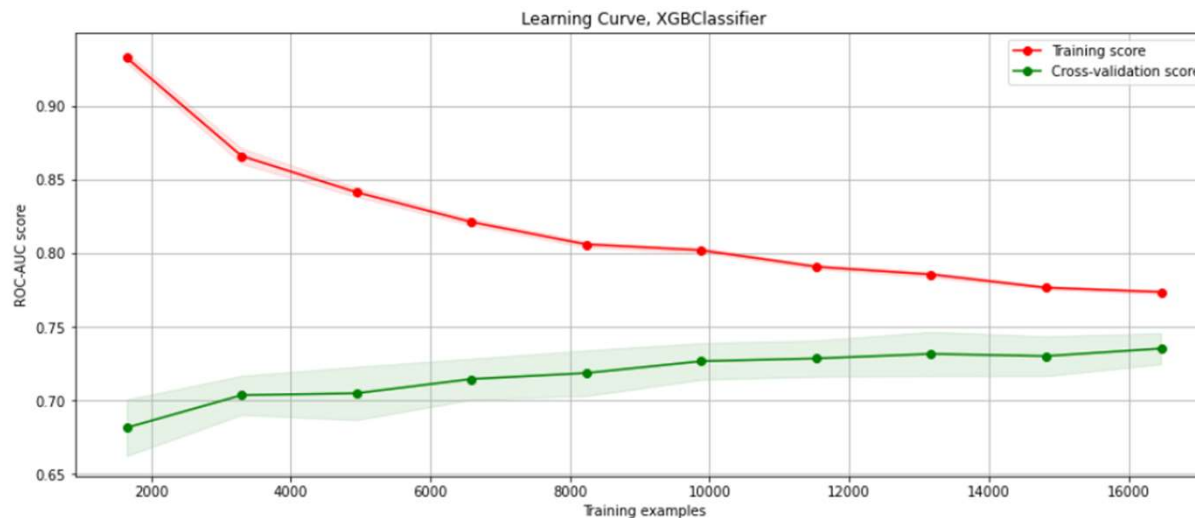
98

Nikolay Kramarenko



0.70460

- Пропуски в числовых признаках заменены на средние значения
- удалены неинформативные признаки
- категориальные признаки частично закодированы с помощью метода Label encoding, а частично с помощью OneHotEncoding
- выбран XGBClassifier с параметрами: {'booster': 'gbtree', 'learning_rate': 0.1, 'max_depth': 2, 'n_estimators': 100}



24.09.2020

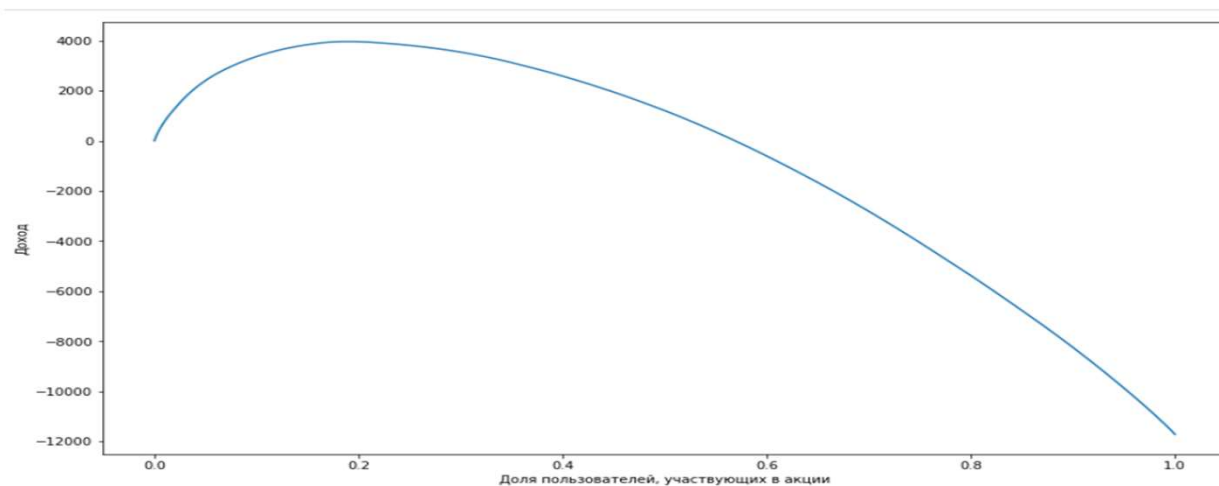
ЭКОНОМИЧЕСКАЯ МОДЕЛЬ

Для оценки успешности эксперимента, была построена экономическая модель зависимости дохода от различных показателей:

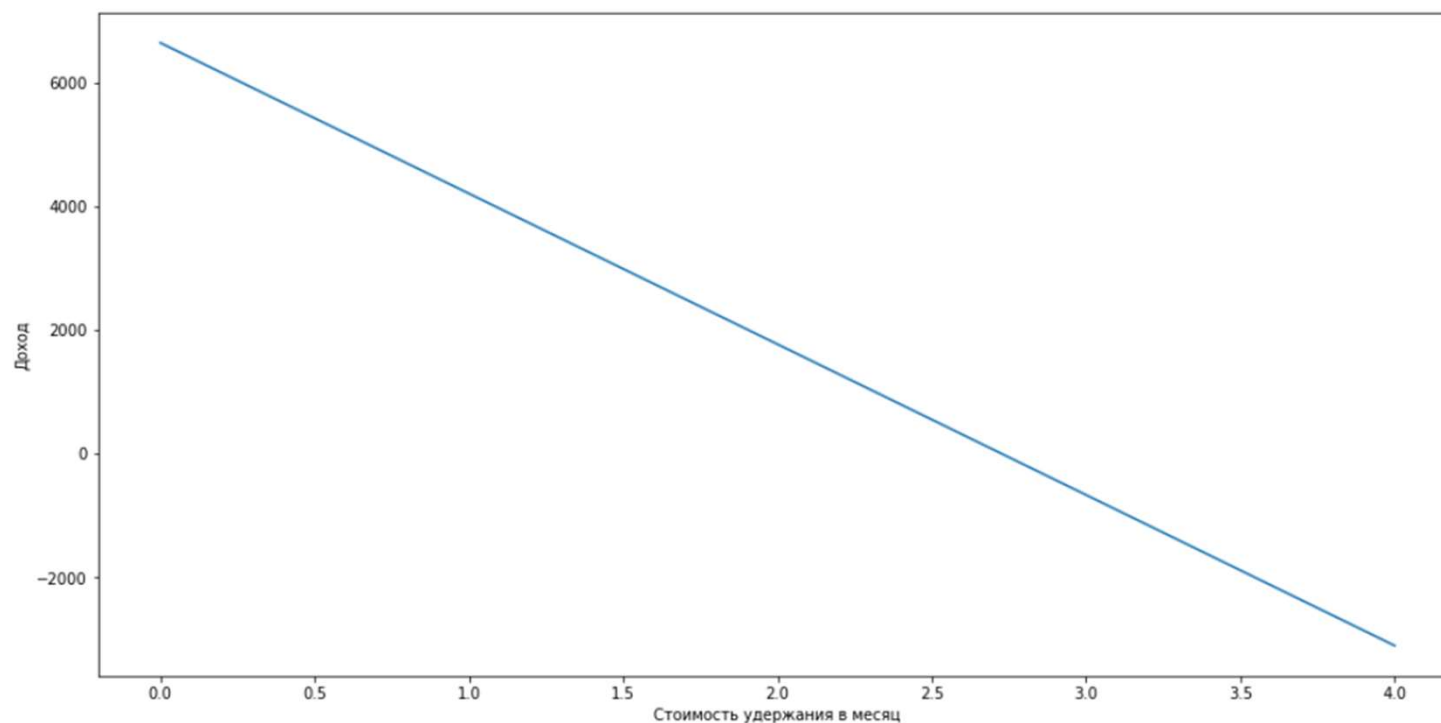
- сколько денег в среднем приносит один пользователь в месяц, условных единиц
- сколько денег в среднем вкладывается в удержание одного пользователя в месяц
- с какой вероятностью пользователь, собиравшийся уйти примет предложение
- сколько пользователей будет участвовать в кампании
- продолжительность скидки для пользователя, месяцы
- вероятность, что пользователь останется клиентом после окончания акции и до конца прогнозного периода
- длина прогнозного периода, месяцы

ЭКОНОМИЧЕСКАЯ МОДЕЛЬ

С помощью этой модели можно очень гибко оценивать доход в зависимости от входных данных. Зависимость дохода от доли пользователей, кому предложена акция:



ЗАВИСИМОСТЬ ДОХОДА ОТ СТОИМОСТИ УДЕРЖАНИЯ В МЕСЯЦ:



ВЫВОДЫ И ПРЕДЛОЖЕНИЯ

Уже в текущем виде модель показывает экономическую целесообразность, но при дальнейшем улучшении можно получить еще больший эффект.

Например, одним из вариантов по улучшению модели, можно предложить улучшение качества входных данных. Для этого желательно раскрытие значений признаков, чтобы можно было более качественно их предобработать.

Некоторые из параметров модели, например вероятность принятия предложения, вероятность продолжения пользования услугами после окончания акции и тд можно подобрать с помощью А/Б тестирования, т.к. эти параметры заметно влияют на итоговый доход.