

Курсовой Проект

ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ ПОДКЛЮЧЕНИЯ УСЛУГИ

*Группа GU 1695
студент Бакеев Алексей*



ПОСТАНОВКА ЗАДАЧИ

на основе предоставленной информации об отклике абонентов на предложение о подключении услуг и анонимизированном наборе данных об абонентах, необходимо построить алгоритм, определяющий факт (или вероятность) подключения услуги абонентом.

ОБЗОР ДАТАСЕТА

Для обучения модели предоставлен датасет, содержащий историю предложения услуг пользователям компании Мегафон за 6 месяцев.

При анализе признаков оказалось, что для одного id могут быть несколько строк данных. Объединение признаков по времени покупке buy time было невозможно, так как признаки собирались в разное время.

Для определения наиболее актуальных признаков был разработан алгоритм

ОБРАБОТКА ПРИЗНАКОВ

При анализе данных было выявлено

дисбаланс значений целевой переменной как суммарно по всем услугам, так и по каждой в отдельности

наличие пяти постоянных признаков не представляющих ценности

двадцать бинарных признаков, которые были закодированы

десять категориальных признаков, с количеством категорий менее 10, которые также были закодированы

вещественные признаки, которые были стандартизованы

ИСПОЛЬЗОВАННЫЕ МОДЕЛИ

Были применены градиентные бустинги

LGBMClassifier

XGBClassifier

Для удобства работы применялись пайплайны

ОЦЕНКА ТОЧНОСТИ

Скоринг осуществлялся функцией *f1*

```
In [103]: %%time
```

```
# Прогноз на валидационной выборке.
```

```
y_pred_valid = pipe_model_lgbm.predict(X_valid)
```

```
Wall time: 5.32 s
```

```
In [104]: # Вывод метрик качества прогноза на валидационной выборке.
```

```
print(classification_report(y_valid, y_pred_valid))
```

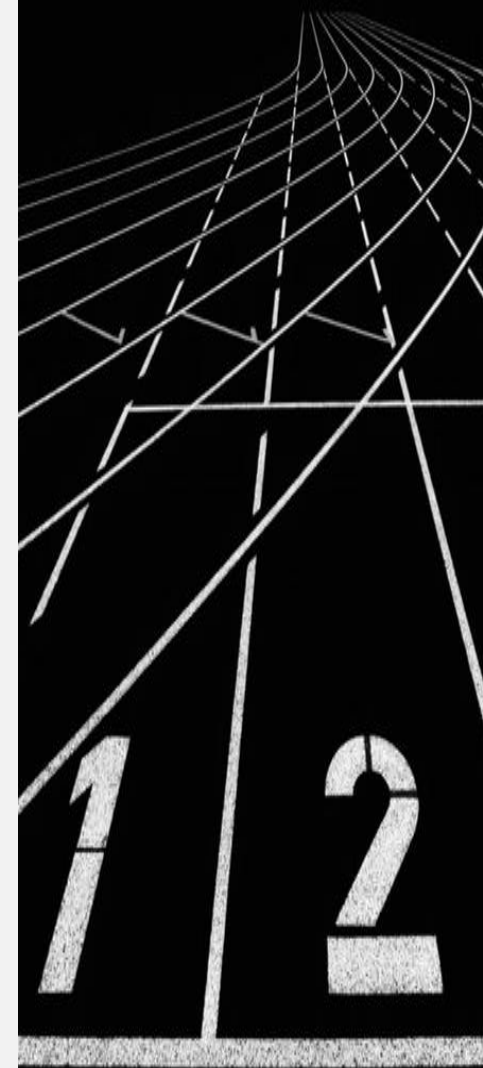
	precision	recall	f1-score	support
0	0.98	0.87	0.92	192867
1	0.32	0.79	0.45	15047
accuracy			0.86	207914
macro avg	0.65	0.83	0.69	207914
weighted avg	0.93	0.86	0.89	207914

РЕЗУЛЬТАТЫ

Результаты предсказаний дают высокое значение метрики Recall = 0.8 что позволяет надеяться на то, что определено большинство потенциальных клиентов

Достаточно низкое значение Precision = 0.3 указывает на то, что не более трети потенциальных клиентов определены верно

При проведении серьезной рекламной компании можно рассчитывать на 30 % подключений от всей целевой аудитории



СПАСИБО

Бакеев Алексей 📱 +7 (531) 987-65-43

✉ *ab@dot.com*