

Анализ модели классификации для Яндекс.Директа

1. Общая информация

Размер датасета: 500000 записей

Количество признаков: 100 (изначально 1376)

Целевая переменная 'label':

0 - 258499 записей (51.7%)

1 - 241501 записей (48.3%)

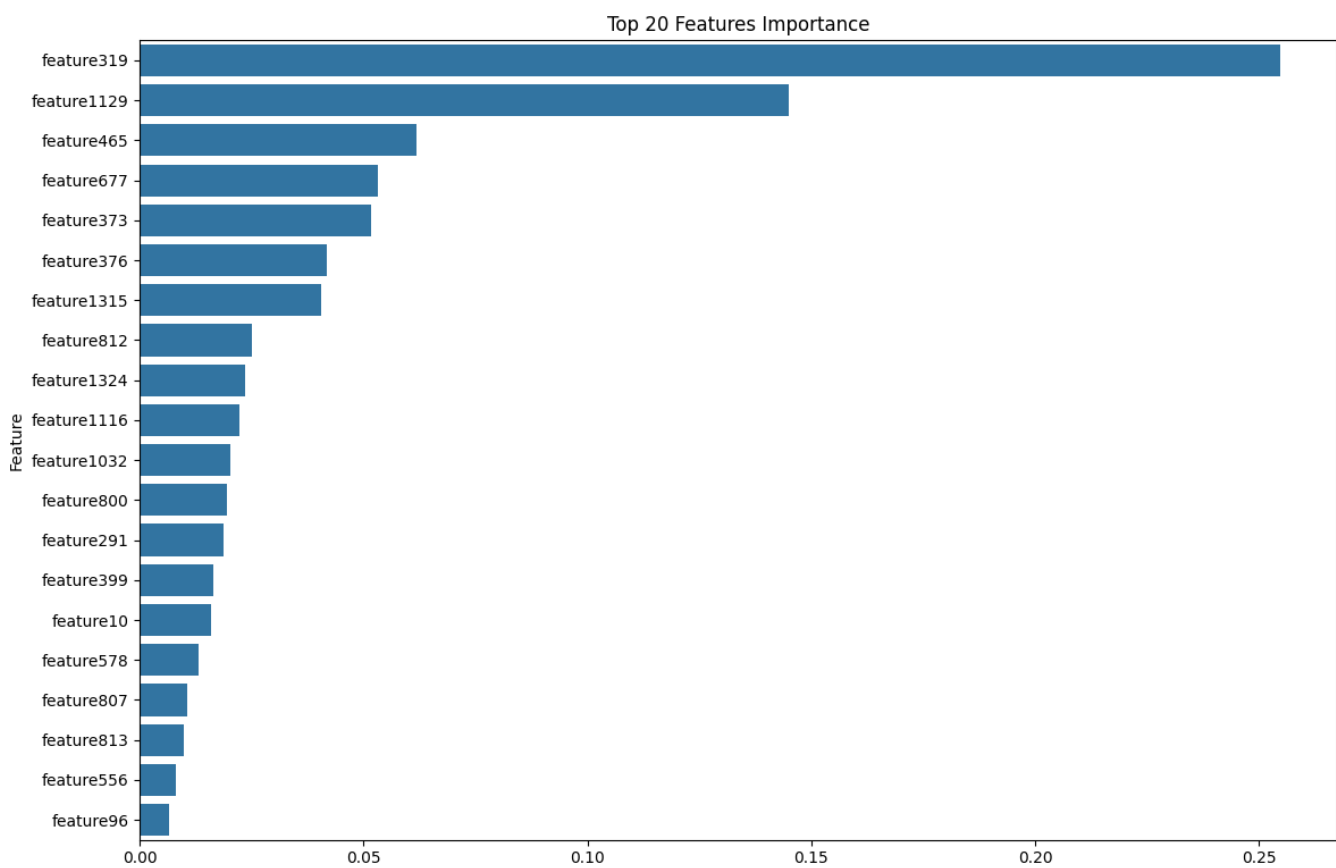
2. Out-of-Time валидация

Средний ROC-AUC на 3 фолдах: 0.9862

Значения по фолдам: ['0.9875', '0.9862', '0.9849']

PSI для предсказаний модели: 0.1009

3. Важность признаков



Анализ модели классификации для Яндекс.Директа

4. PSI анализ признаков

Топ-10 признаков с наибольшим PSI:

Feature	PSI
feature96	0.013957
feature666	0.013245
feature419	0.012485
feature465	0.011911
feature623	0.011668
feature1165	0.011602
feature1002	0.011488
search	0.009972
feature578	0.008643
feature1324	0.007522

5. Метрики по срезам

Результаты классификации для различных сегментов данных:

Срез	Тип	Размер	Precision	Recall	F1	AUC		
2	Клик SDK He	Search	3934	0.931818	0.763146	0.839089	0.938982	
3	Клик He	SDK Search	4582	0.981854	0.976593	0.979217	0.993874	
4	Клик He	SDK He	Search	2469	0.961877	0.941852	0.951759	0.986086
6	Показ SDK He	Search	89004	0.912850	0.917847	0.915342	0.991229	
7	Показ He	SDK Search	151469	0.988878	0.989488	0.989183	0.997226	
8	Показ He	SDK He	Search	239388	0.938159	0.904719	0.921135	0.986266

6. Выводы и рекомендации

1. Качество модели: ROC-AUC = 0.9862 указывает на хорошее качество модели.
2. Стабильность: PSI = 0.1009 (< 0.1) говорит о стабильности модели во времени.
3. Важные признаки: feature96, feature666, feature419 оказались наиболее значимыми.
4. По срезам: Наибольшие проблемы с качеством в срезе Клик|SDK|He Search.
5. Рекомендации: Модель можно использовать в продакшене, но требуется мониторинг срезов с низким F1.