

Задача №4: Анализ и прогнозирование дебиторской задолженности

КОМАНДА JEDAI

Анализ контрагентов

Цель: научиться определять контрагента, у которого может возникнуть ПДЗ

Данные за 3 года были объединены в одну таблицу

Пропуски заполнены медианными значениями

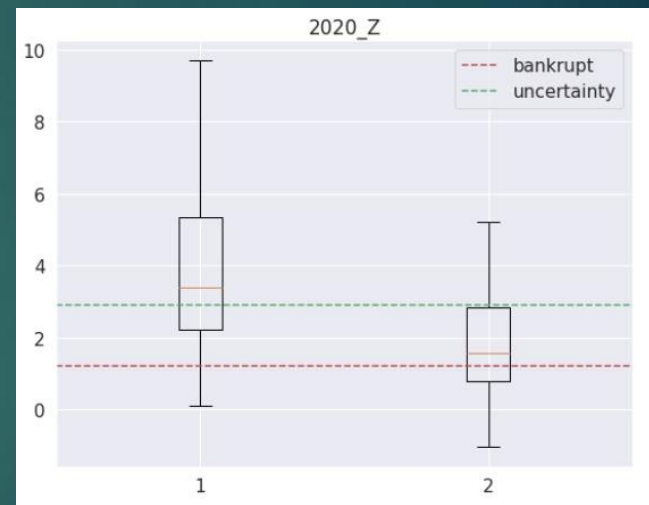
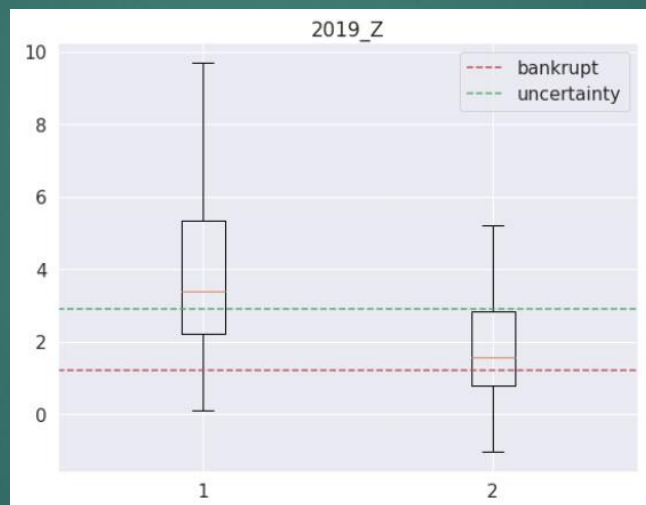
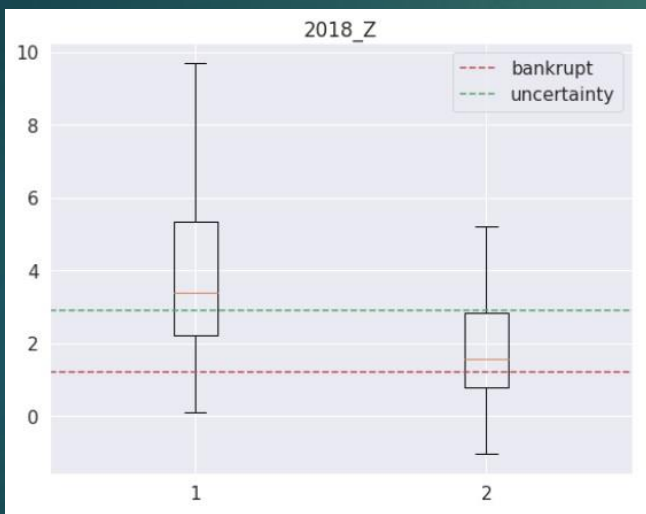
	YY	YY_1_внеоборотные_активы_rub	YY_1_выручка_rub	YY_1_ддзоа	YY_1_дебиторская_задолженность_rub	YY_1_долгосрочные_обязательства_rub	YY_1_заёмные_средства_долгосрочные_rub	YY_1_заёмные_средства_краткосрочные_rub
0	2021	4.379874e+09	2.794788e+10	0.806640	4.977055e+09	2.174739e+08	0.000000e+00	0.000000e+00
1	2021	4.369299e+07	5.798854e+08	0.405340	8.003121e+07	8.257962e+06	0.000000e+00	0.000000e+00
2	2021	1.301418e+09	2.583063e+09	0.492774	3.714420e+08	5.629299e+06	0.000000e+00	0.000000e+00
3	2021	4.369299e+07	5.798854e+08	0.405340	8.003121e+07	8.257962e+06	0.000000e+00	0.000000e+00
4	2021	1.252995e+11	9.629957e+10	0.773734	1.860993e+10	9.076590e+10	8.802936e+10	8.802936e+10
...
1588	2019	8.407643e+05	1.154089e+08	0.295420	2.566242e+07	3.194904e+07	3.194904e+07	3.194904e+07
1589	2019	1.452439e+08	1.188227e+09	0.217893	8.277452e+07	2.354777e+06	0.000000e+00	0.000000e+00
1590	2019	0.000000e+00	0.000000e+00	0.405340	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
1591	2019	4.922949e+08	4.278682e+08	0.182059	7.949618e+07	1.903682e+08	1.666369e+08	1.666369e+08
1592	2019	0.000000e+00	1.448790e+07	0.943162	1.137261e+07	0.000000e+00	0.000000e+00	0.000000e+00

1593 rows × 106 columns

Дополнительные признаки

- Оборачиваемость дебиторской задолженности (ОДЗ)
- Период погашения дебиторской задолженности (ППДЗ)
- Доля дебиторской задолженности в оборотных активах (ДДЗОА)
- Модель Альтмана (5-тифакторная, для оценки банкротства)

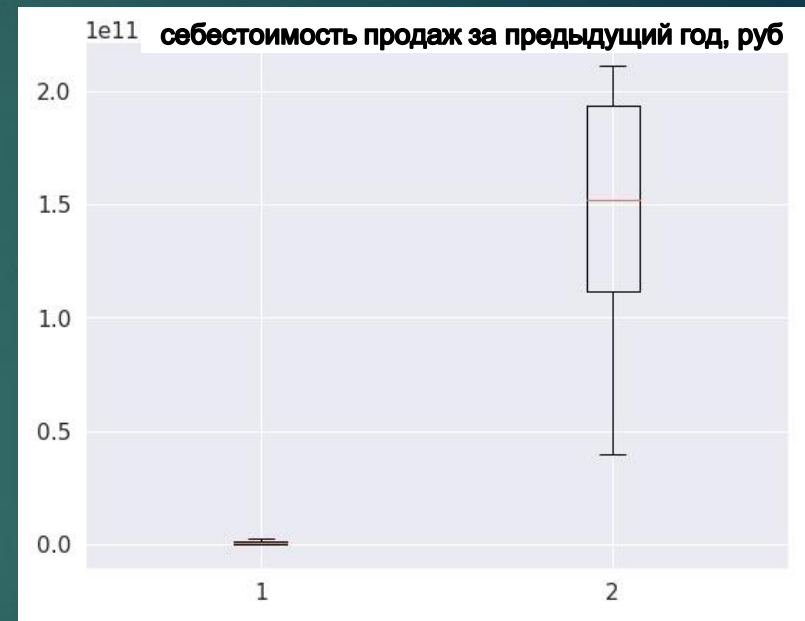
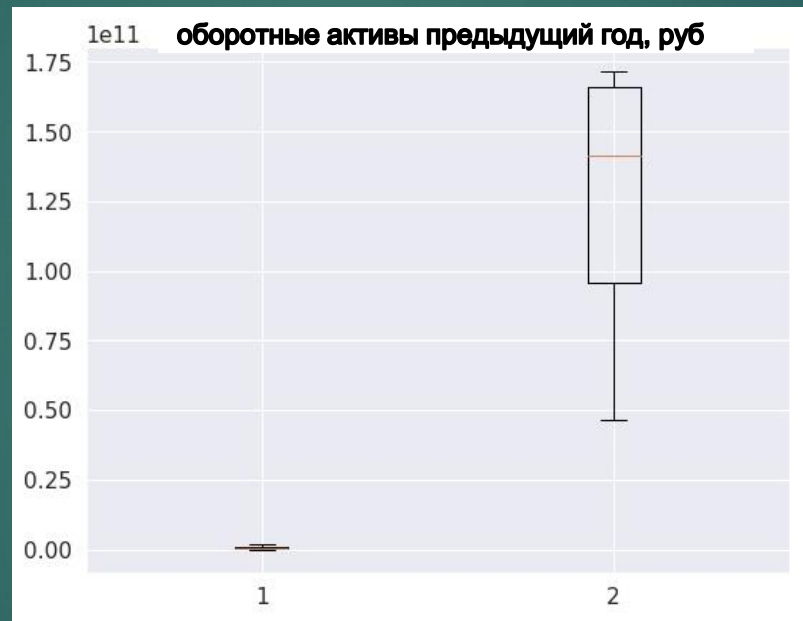
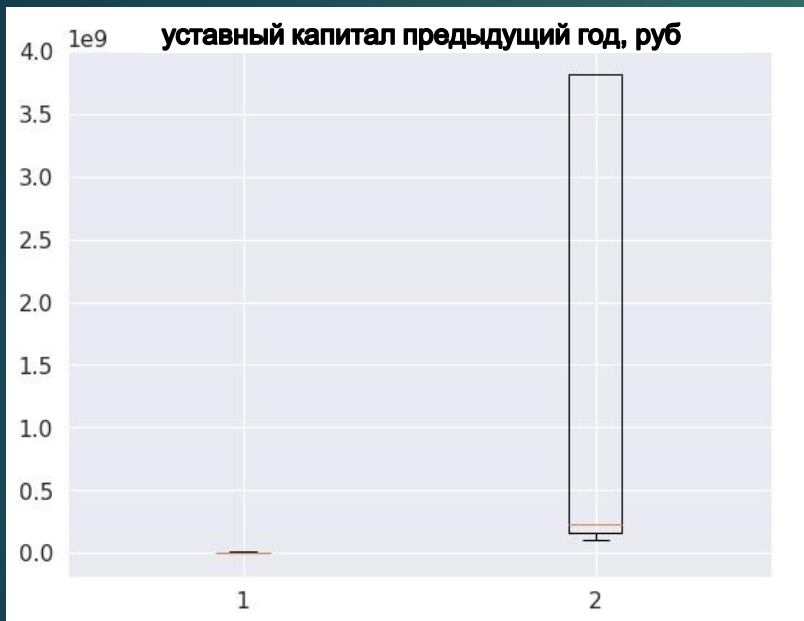
1 – «хорошие» контрагенты
2 – «плохие» контрагенты



Модель Альтмана позволяет выделить потенциально «плохих» контрагентов и может служить критерием для быстрой проверки

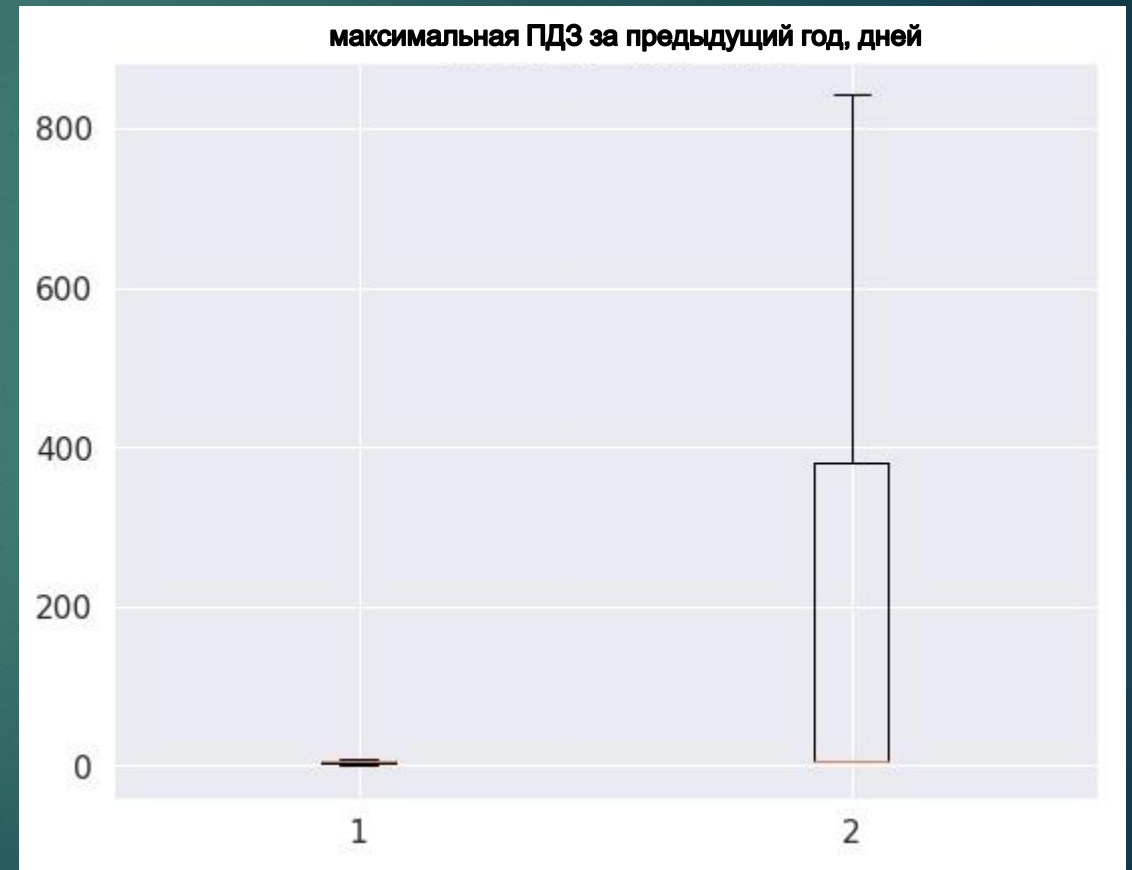
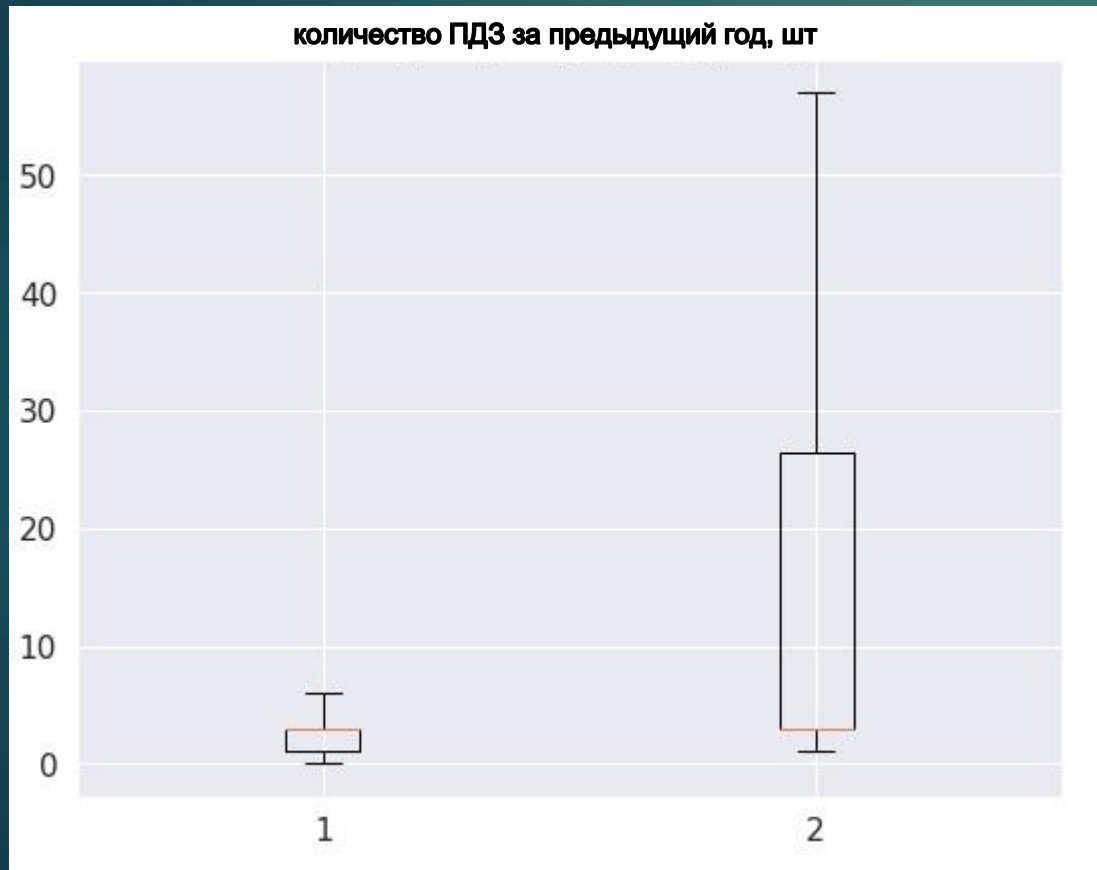
Портреты контрагентов (2019, 2020)

Небольшие компании с большей вероятностью погашают ДЗ в срок



Портреты контрагентов (2019, 2020)

Если у контрагента ранее была небольшая ПДЗ, скорее всего он и дальше будет возвращать ДЗ в срок



Анализ факторов (2021)

Correlation matrix	
макс_пдз_за_2019_год_дней	0.35
макс_пдз_за_2020_год_дней	0.49
сред_пдз_за_2020_год_дней	0.36
2020_дебиторская_задолженность_gub	0.34
2020_капитал_и_резервы_gub	0.31
2020_кредиторская_задолженность_gub	0.42
2020_краткосрочные_обязательства_gub	0.33
2020_себестоимость_продаж_gub	0.31
факт_59	0.3
факт_60	0.32

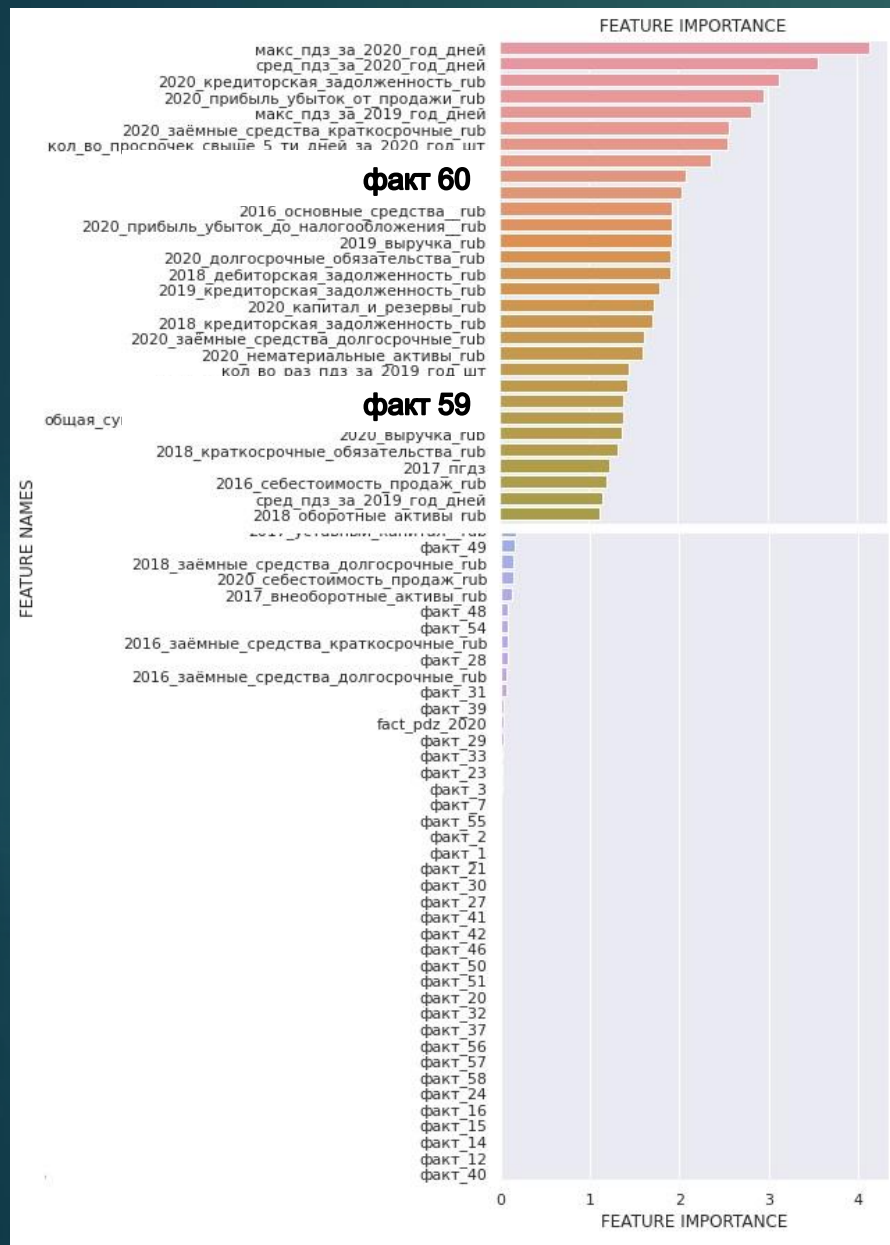
кол_во_раз_пдз_за_2021_год_шт

кол_во_раз_пдз_за_2021_год_шт =

$$\begin{aligned} &= \text{пдз_1_30} + \\ &\quad + \text{пдз_31_90} + \\ &\quad + \text{пдз_91_365} + \\ &\quad + \text{пдз_более_365} \end{aligned}$$

Имеет смысл обратить внимание на факторы 59, 60

Анализ факторов (2021)



Модель прогноза факта наличия ПДЗ

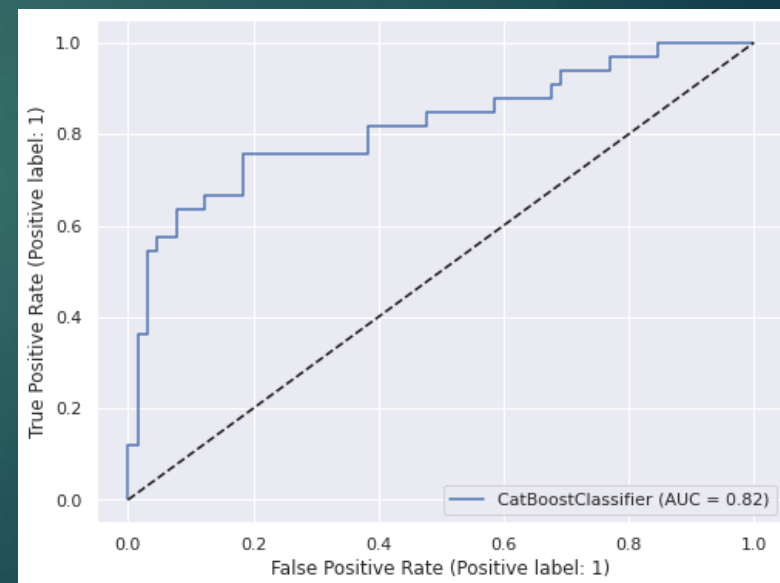
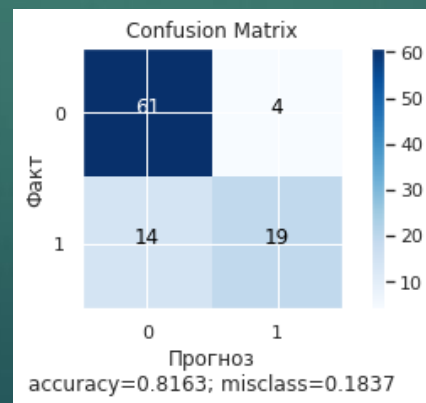
f1: 0.679

Precision: 0.826

Recall: 0.576

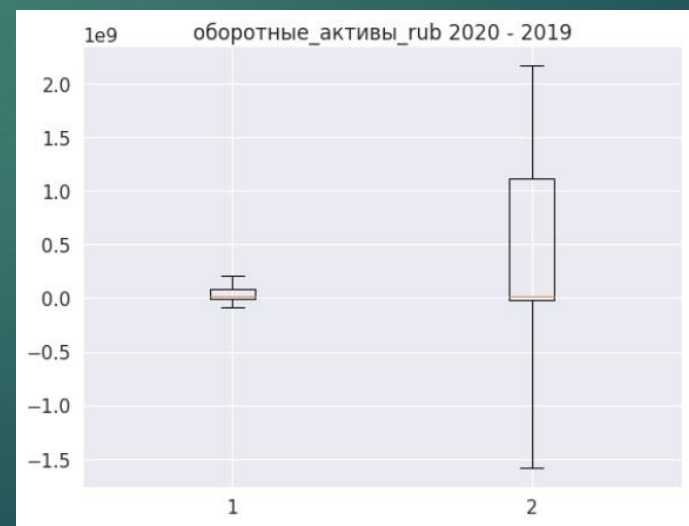
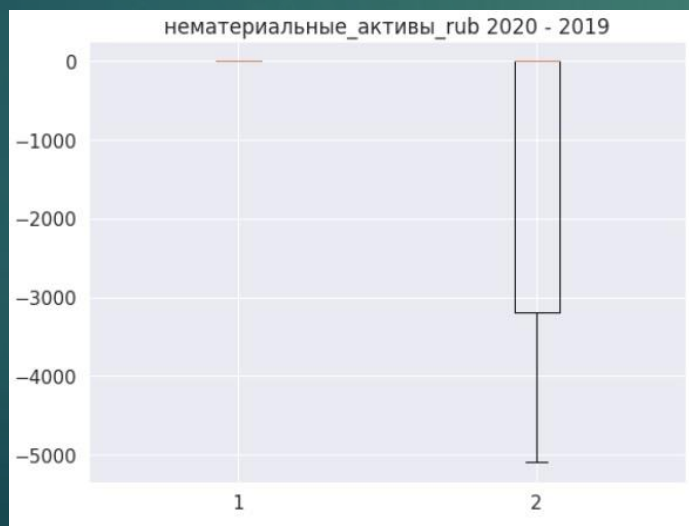
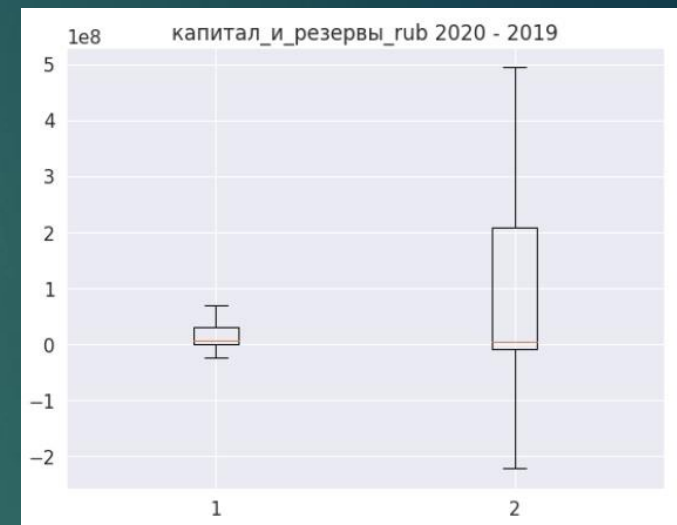
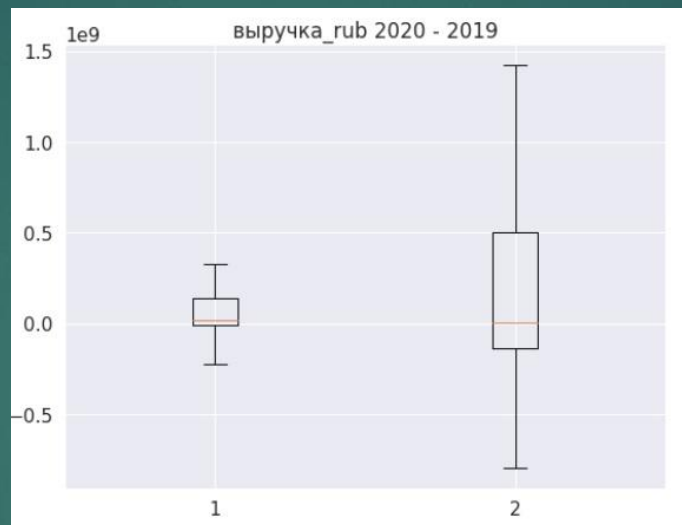
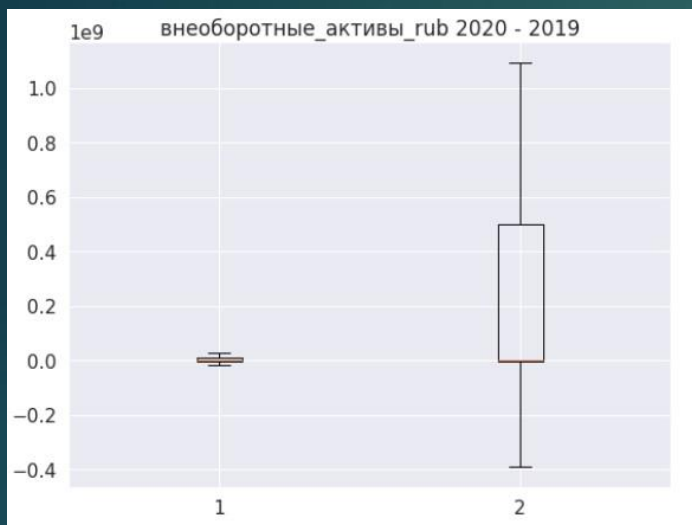
ROC_AUC: 0.757

Mean f1_scores (CV) = 0.642



Изменения финансовых показателей (2021)

Чем меньше изменений в финансовых показателях, тем меньше ПДЗ

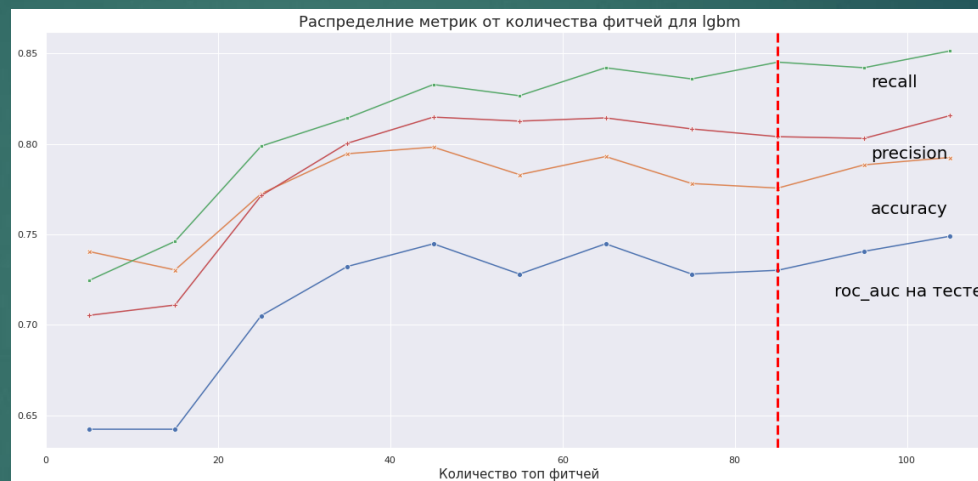


Общий подход к моделированию

1. Несколько вариантов моделей

Линейные	Метрические	Деревья & Ансамбли деревьев	Градиентный бустинг
Logistic Regeression	Kneighbors Classifier	Decision Tree Classifier Random Forest Classifier	LGBM Classifier Catboost

2. Разные наборы фичей



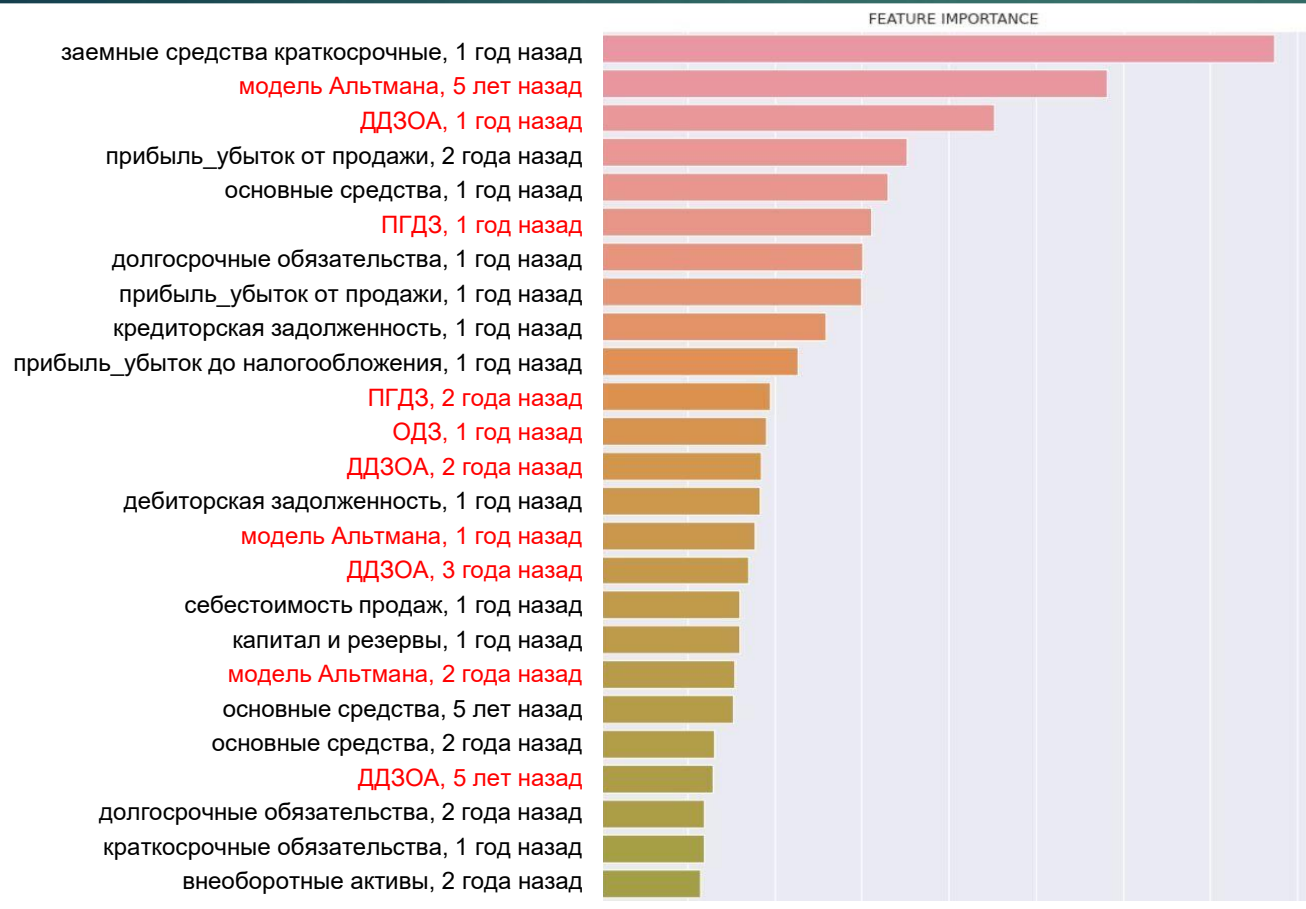
- Для каждой комбинации подбираются оптимальные гиперпараметры (Grid Search)
- Оценка нескольких метрик в совокупности на кросс-валидации

Accuracy	Precision	Recall	Roc_auc valid	Roc_auc test	Модель
0.726	0.778	0.833	0.802	0.791	Catboost
0.734	0.780	0.845	0.809	0.803	LGBM Classifier

Модель прогноза факта ПДЗ (2019 - 2021)

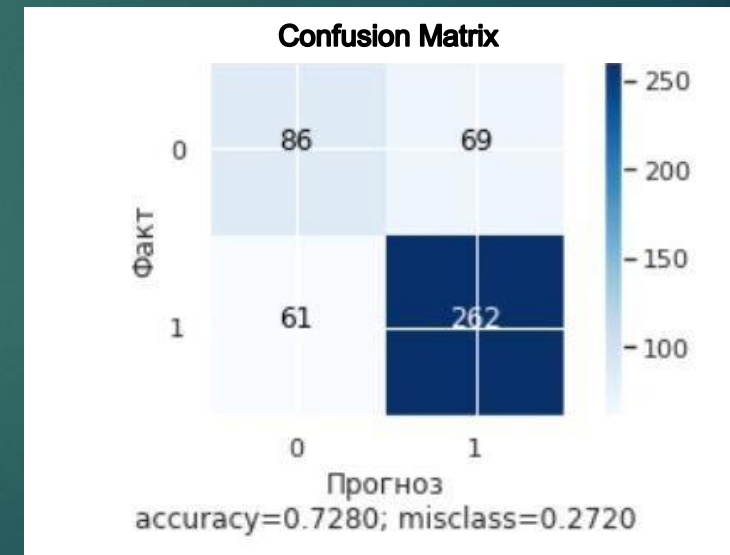
Старые контрагенты

Модель CatBoost, 25 фичей



Метрики:

- accuracy: 0.74
- precision: 0.80
- recall: 0.82
- roc_auc (test): 0.79



Модель прогноза факта ПДЗ (2019 - 2021)

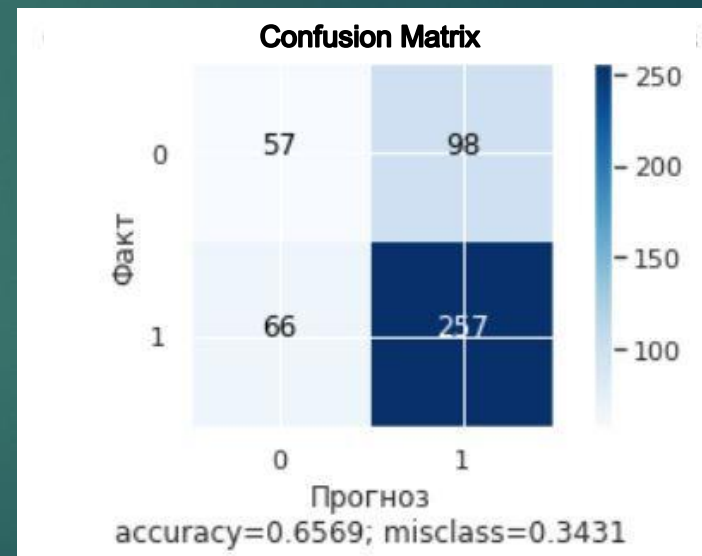
Новые контрагенты

Модель CatBoost, 11 фичей



Метрики:

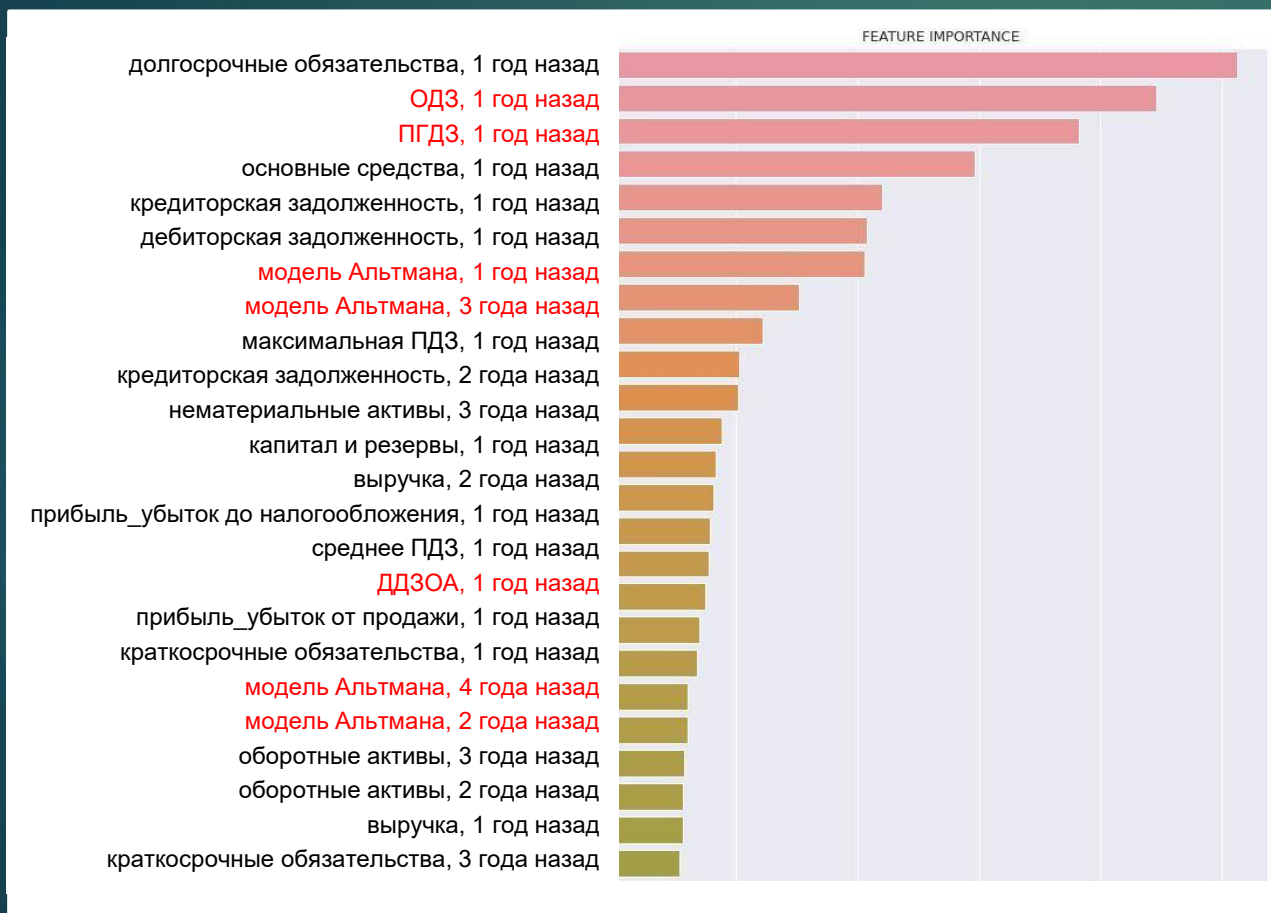
- accuracy: 0.65
- precision: 0.71
- recall: 0.81
- roc_auc (test): 0.71



Модель прогноза факта ПДЗ > 30 дней (2019 - 2021)

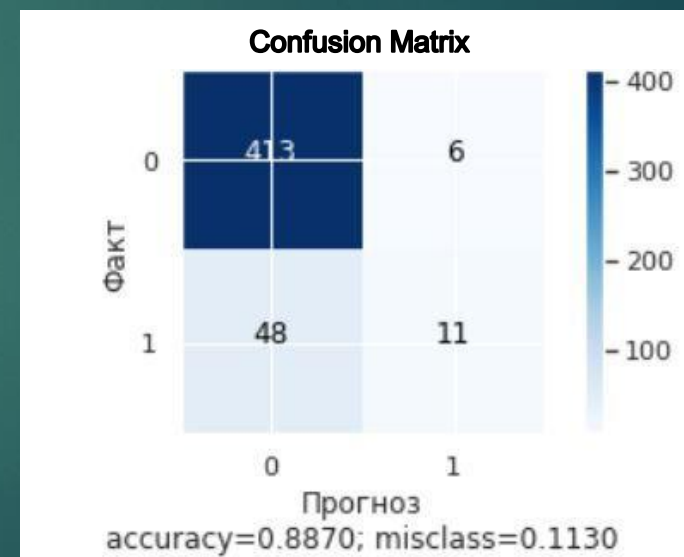
Старые контрагенты

Модель Random Forest, 25 фичей



Метрики:

- accuracy: 0.87
- precision: 0.53
- recall: 0.14
- roc_auc (test): 0.84



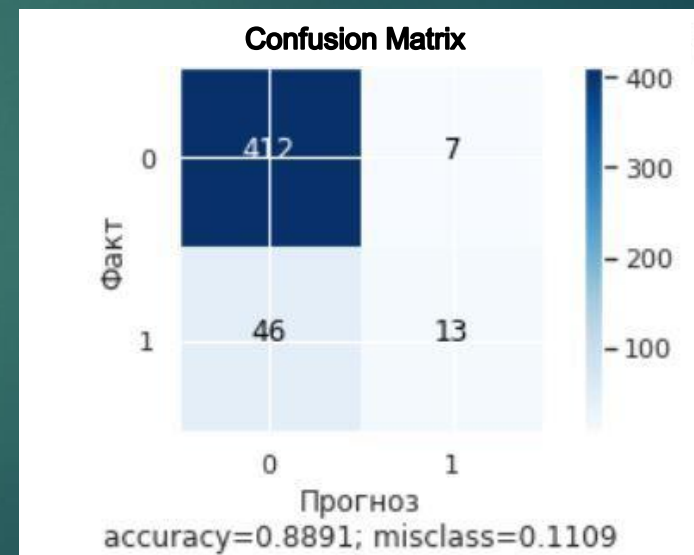
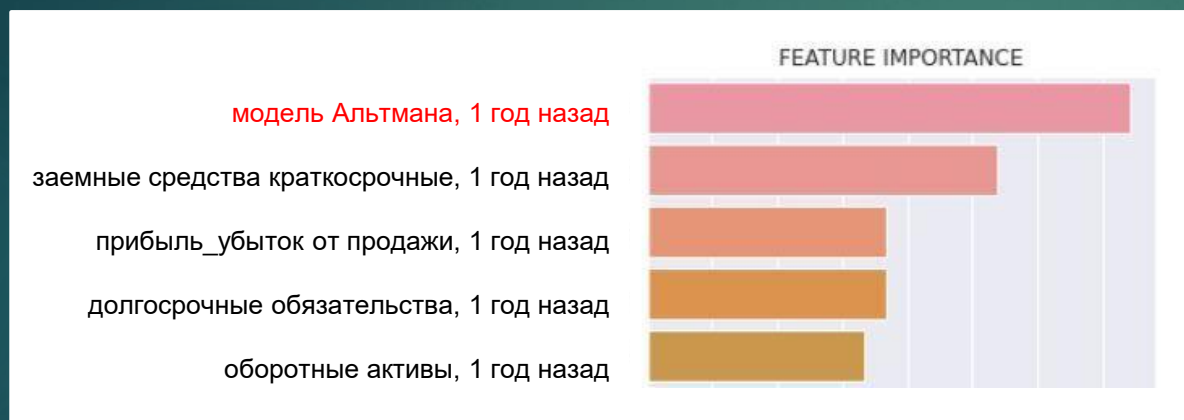
Модель прогноза факта ПДЗ > 30 дней (2019 - 2021)

Новые контрагенты

Модель CatBoost, 5 фичей

Метрики:

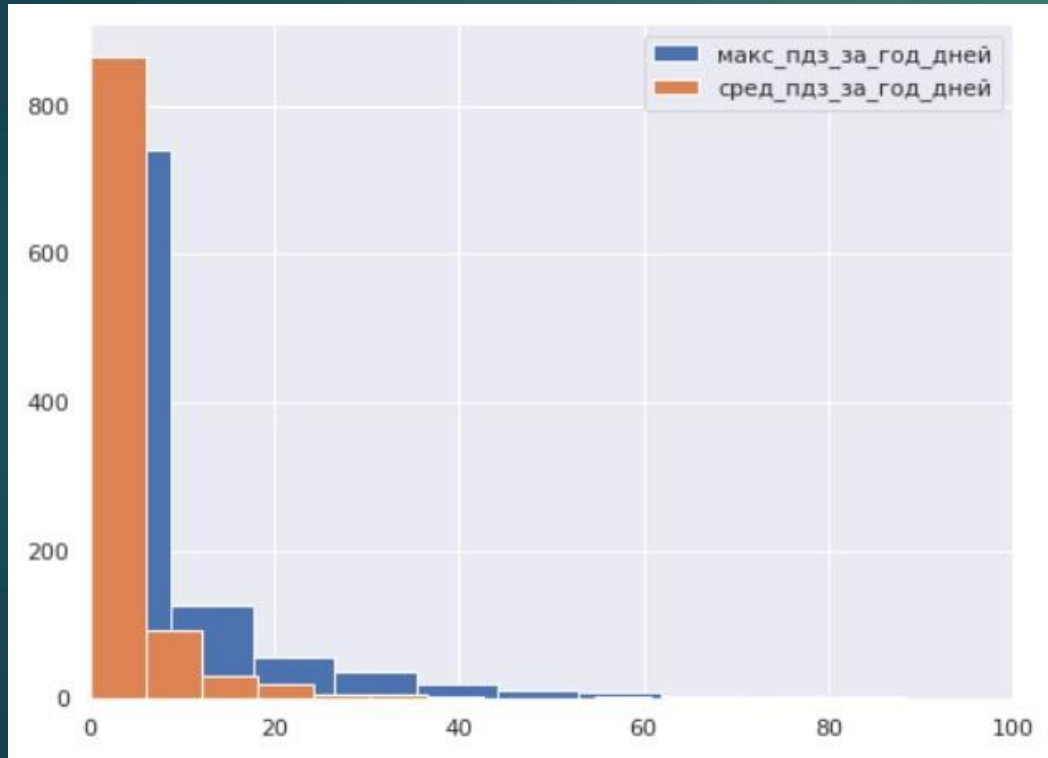
- accuracy: 0.88
- precision: 0.62
- recall: 0.08
- roc_auc (test): 0.80



Модель прогноза факта ПДЗ от 60 до 90 дней (2019 - 2020)

Старые контрагенты

- Нет возможности получить данные за 2021 год
- Сильный дисбаланс классов



Контрагенты с ПДЗ > 60 дней – около 60 строк
Контрагенты с ПДЗ < 60 дней – около 1000

Идеи по расчету экономической эффективности

Без использования модели:

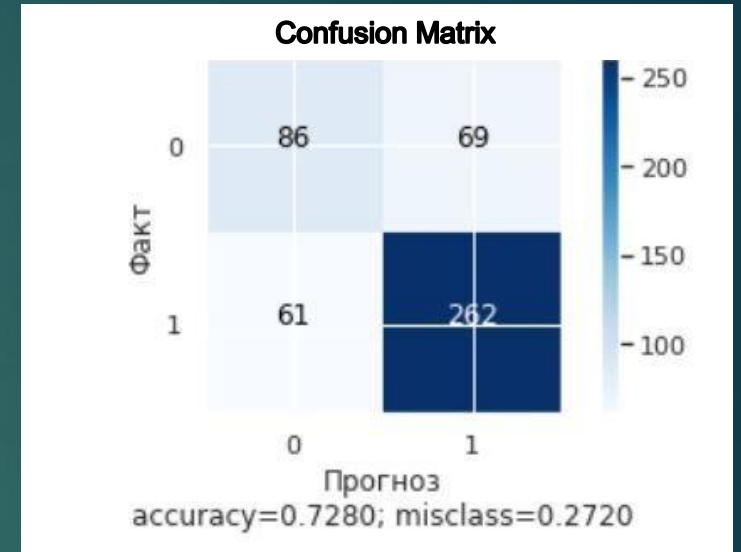
- 155 контрагентов – вернули в срок
- 323 контрагента – не вернули в срок

При использовании модели:

- 86 контрагентов – вернули в срок
- 262 контрагента – «должники», им не предоставили ДЗ
- 61 контрагент – не вернули в срок
- 69 контрагентов – не предоставили ДЗ

Что с этим делать:

- Проанализировать данные по «должникам», которых определили верно, оценить общую сумму «выгоды» в случае не предоставления им ДЗ
- Оценить риски возникновения ПДЗ для конкретных контрагентов



Выводы

- Найдены интересные зависимости в данных
- Для быстрой оценки контрагента (КА) можно использовать портреты «хороших» и «плохих» КА, а также модель Альтмана
- Подготовлена **жизнеспособная модель прогноза факта ПДЗ** – как для старых, так и для новых контрагентов
- Подготовлена **модель прогноза ПДЗ свыше 30 дней** – оценивается как удовлетворительная (лучше, чем ничего)
- Модель **прогноза ПДЗ от 60 до 90 дней** сделать не удалось – очень сильный дисбаланс классов
- Датасет был расширен дополнительными признаками, которые оказались очень полезны при обучении моделей