




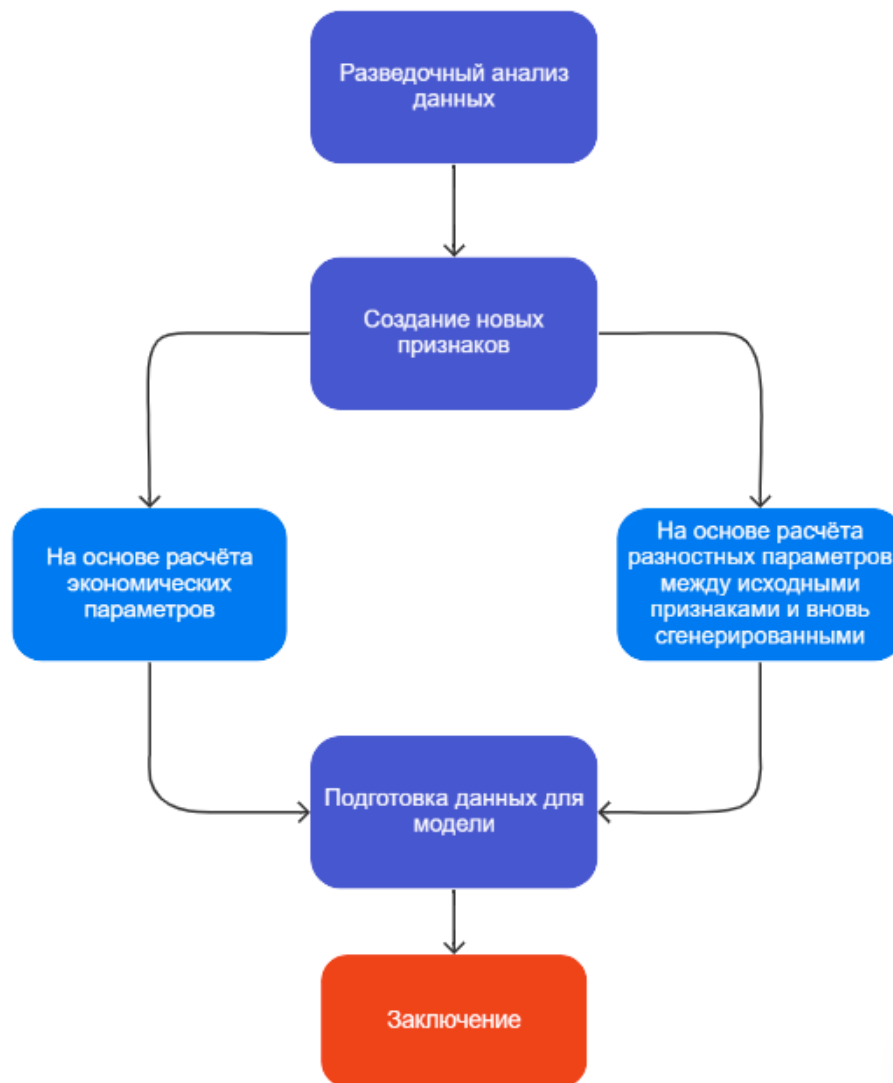
Team 7

Финал

Анализ контрагентов

 Виктор Ожерельев – tg: @ozhereliev_v
Алексей Байдин
Любовь Лазарева
Дмитрий Сорока

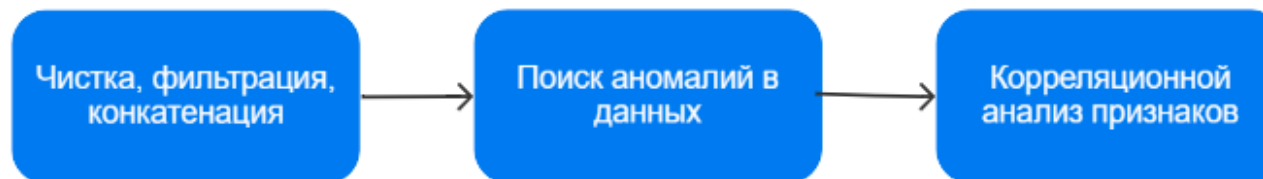
Road map



EDA



- Переименование признаков
- Обработка факторных признаков: замена значений 'X', 'Н/Д' на -1, удаление пустых признаков и признаков с высокой корреляционной связью
- Поиск уникальных признаков по годам
- Конкатенация данных за все года
- Анализ балансовых показателей
- Анализ прибыли и убытков



Экономические данные



Коэффициент текущей ликвидности

$$K_{\text{ТЛ}} = \frac{\Sigma \text{ Быстрореализуемые активы}}{\Sigma \text{ Срочные и среднесрочные обязательства}}$$

Рекомендуемые значения 1.3 - 2.5

Рентабельность продаж

$$R_{\text{sale}} = \frac{\text{sale profit}}{\text{revenue}}$$

Коэффициент финансовой независимости

$$K_{\text{ФН}} = \frac{\text{Капитал и резервы}}{\Sigma \text{ Активы и пассивы}}$$

Рекомендуемые значения 0.4 – 0.6

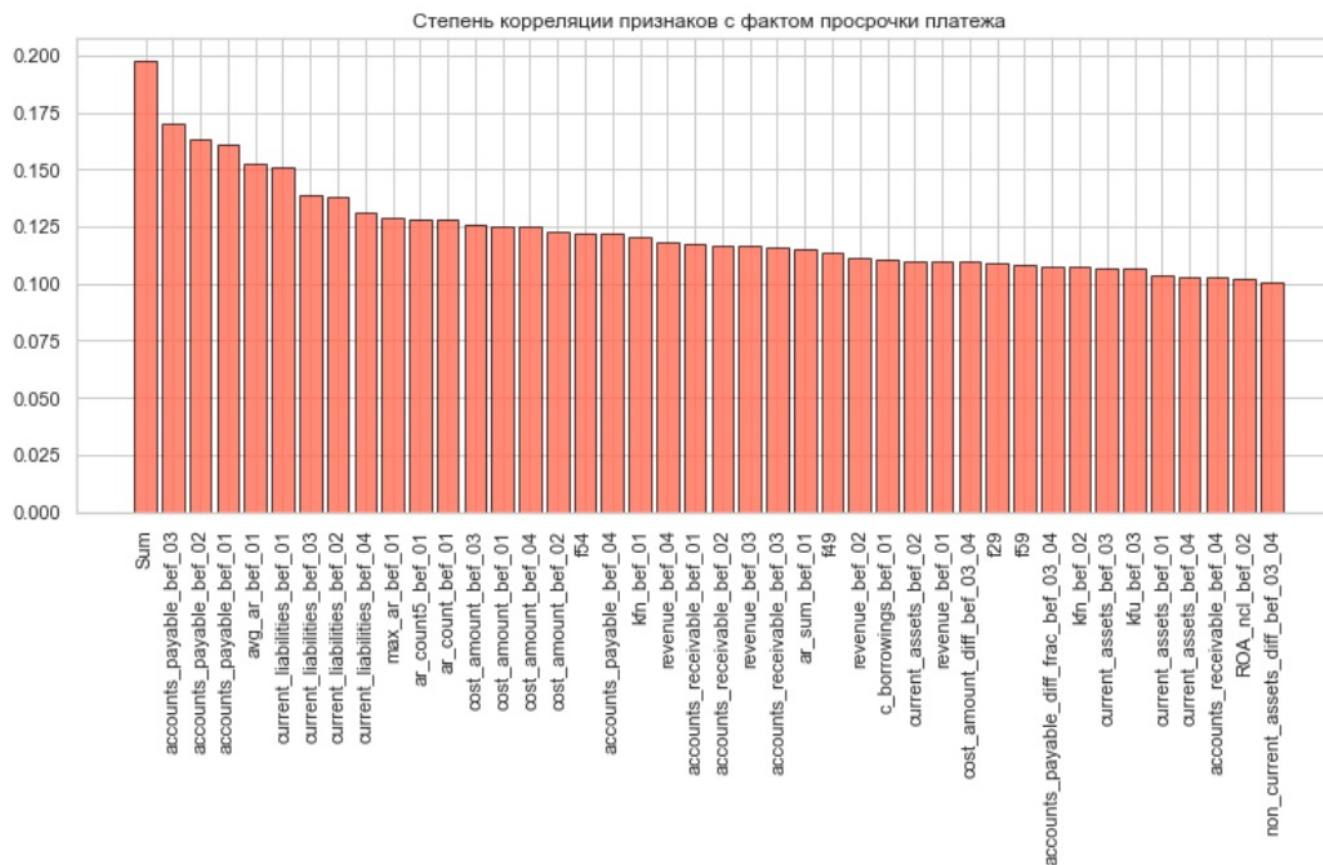
Рентабельность затрат

$$ROCS = \frac{\text{sale profit}}{\text{revenue} - \text{sale profit}}$$

Feature Engineering



Проверка корреляции признаков с целевым - фактом ПДЗ – “overdue”

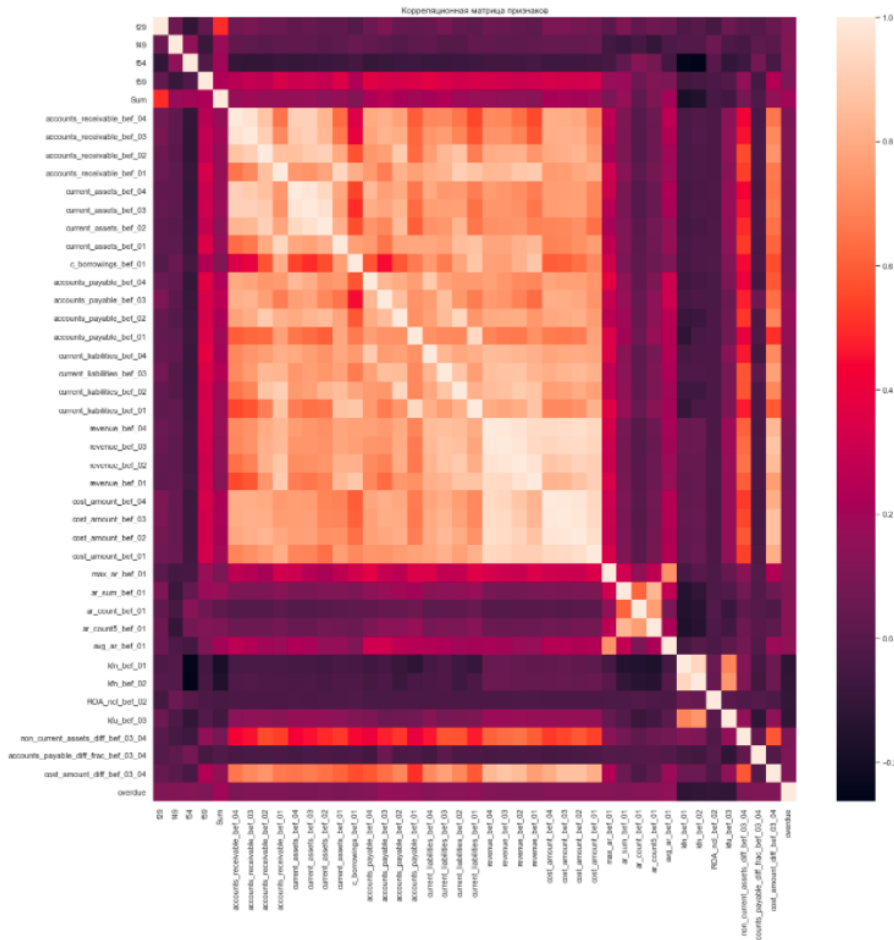


На графике отражены признаки, сила корреляции которых более 0.1

Feature Engineering



Проверка корреляции между признаками



Видно, что статьи и разделы баланса из года в год коррелируют между собой. Для обучения оставим только те, которые коррелируют с целевой переменной наиболее сильно. Остальные признаки, коррелирующие между собой с силой более 0.7 - удалим.

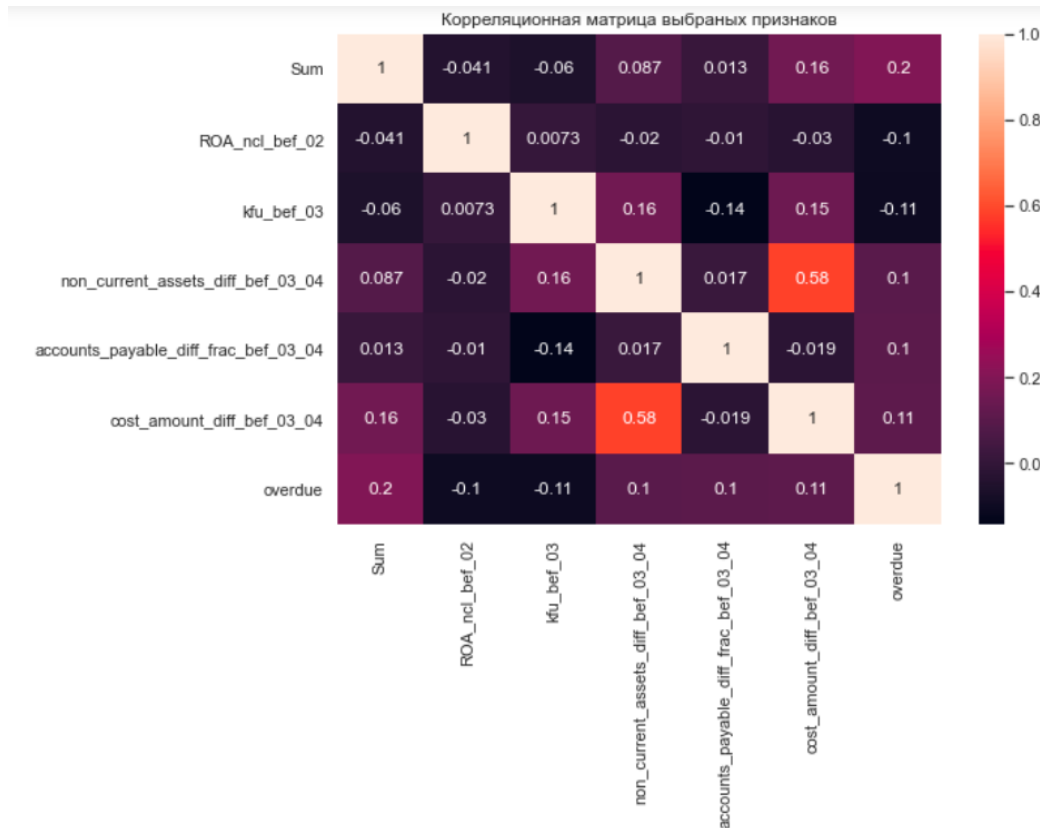
Количество признаков после
предобработки

13

Feature Engineering



Проверка корреляции оставшихся признаков



В ходе удаления «пустых» и признаков с бесконечными значениями, а также признаков, содержащих любую информацию о ПДЗ, т.к. для нового контрагента их не будет, а для контрагента, с которым ранее взаимодействовали интересна будет информации об исполненных старых обязательствах, динамика изменения внутренних параметров -

Количество признаков после предобработки

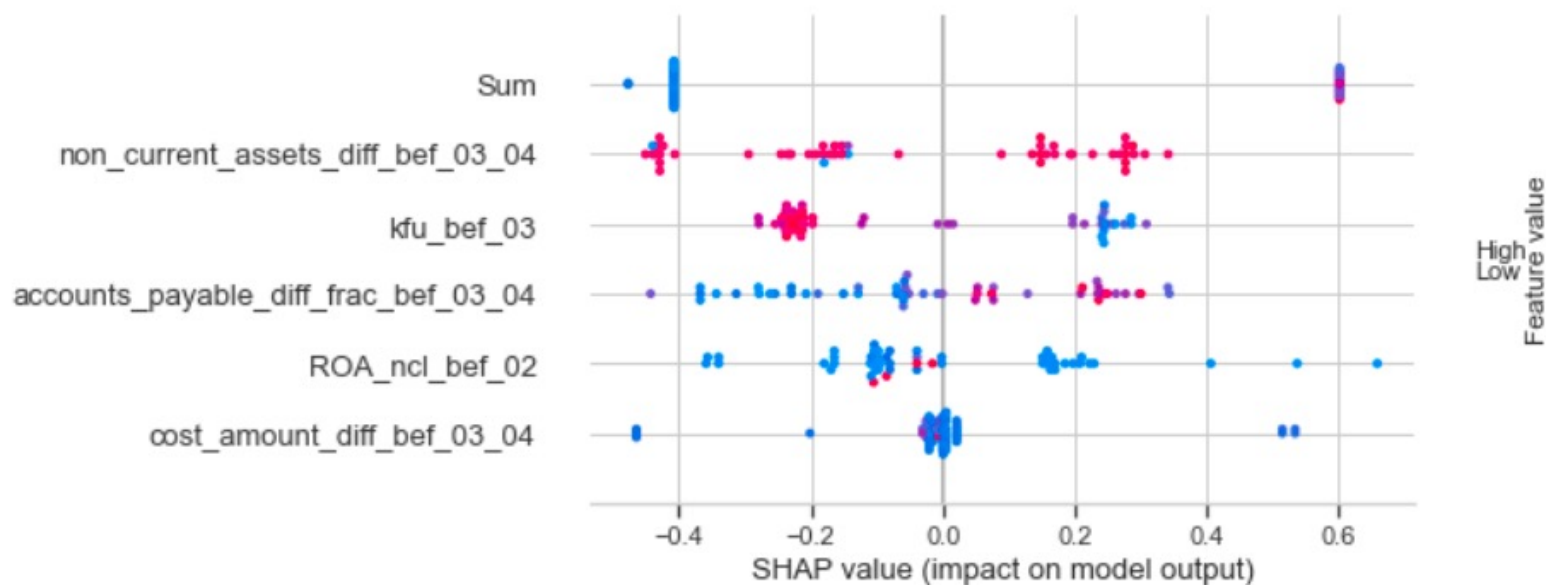
7

Модель предсказания просрочки

Используемая модель

lightGBM

Влияния каждого признака на
предсказания факта ПДЗ



Модель предсказания просрочки



Наиболее значимым признаком для модели, обученной на полном множестве отобранных признаков, получился признак "Sum" (сумма факторных признаков). Примерно одинаковую значимость имеют три параметра следующие за "лидером":

- Коэффициент финансовой устойчивости (по отчетности -3 года, относительно текущего)
- Абсолютное изменение внеоборотных активов -3 года к -4 (относительно текущего)
- Относительное изменение уровня кредиторской задолженности -3 года, к -4 (относительно текущего)

Незначимыми оказались последних два параметра:

- Абсолютное изменений себестоимости -3 года к -4 (относительно текущего)
- Рентабельность внеоборотных активов (по отчетности -2 года, относительно текущего)

Их значения никак не влияют на предсказания модели.

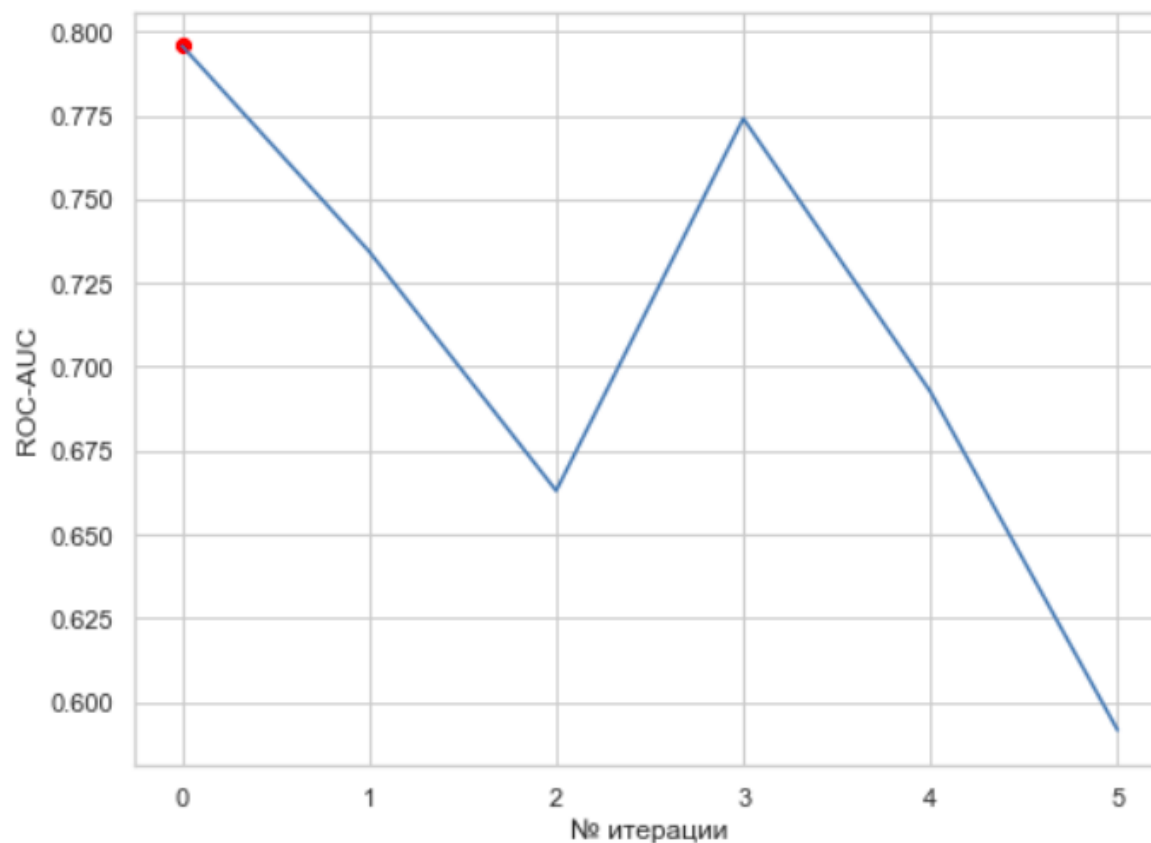
Модель предсказания просрочки



Обучение модели на парах признаков

Наилучшая модель была получена для первой пары признаков: ('Sum', 'kfu_bef_03').

ROC-AUC на тесте составил - 0.82151, доля верных ответов - 80,95238%.



Результаты

[Код на GitHub](#)



>100

Признаков проанализировано и сгенерировано

2

Признака выбрано для обучения модели

81%

Точность предсказаний факта просрочки



Сформирован список дополнительных данных для повышения точности прогноза



**Спасибо
за внимание**

