



**Team 7**

**Финал**

## **Анализ контрагентов**



Виктор Ожерельев – tg: @ozhereliev\_v

Алексей Байдин

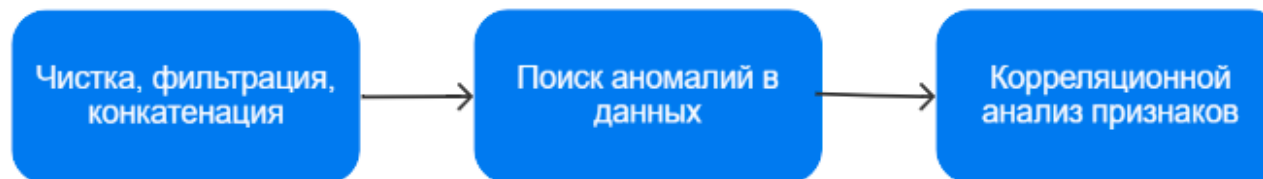
Любовь Лазарева

Дмитрий Сорока

# 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



# Используемые экономические показатели



## Баланс:

- **non\_current\_assets** - внеоборотные активы
  - **intangible\_assets** - нематериальные активы
  - **property\_plant\_equipment** – основные средства
- **current\_assets** - оборотные активы
  - **accounts\_receivable** - дебиторская задолженность
- **capital\_and\_reserves** - капитал и резервы
  - **share\_capital** - уставный капитал
- **non\_current\_liabilities** - долгосрочные обязательства
  - **nc\_borrowings** - заёмные средства (долгосрочные)
- **current\_liabilities** - краткосрочные обязательства
  - **c\_borrowings** - заёмные средства (краткосрочные)
  - **accounts\_payable** - кредиторская задолженность

## Финансовый результат:

- **revenue** - выручка
- **cost\_amount** – себестоимость
- **sales\_profit** - прибыль (убыток) от продажи
- **profit\_before\_tax** - прибыль (убыток) до налогообложения

## Показатели оценки ликвидности и устойчивости:

- **OWL** - собственные оборотные средства (COC)
- **CWC** - чистый оборотный капитал (ЧОК)
- **WC** - рабочий капитал (РК)
- **kti** - коэффициент текущей ликвидности ( $K_{тл}$ )
- **kfn** - коэффициент финансовой независимости ( $K_{фн}$ )
- **kfu** - коэффициент финансовой устойчивости ( $K_{фy}$ )
- **kfl** - финансовый ливеридж ( $K_{фл}$ )
- **km** - коэффициент маневренности ( $K_m$ )
- **kowc** - коэффициент обеспеченности собственными оборотными средствами  $KCOCKCOC$  ( $kowc$ )
- **ROA\_ncl** - рентабельность внеоборотных активов ( $ROA_{вн}$ )
- **ROA\_ca** - рентабельность оборотных активов ( $ROA_{oa}$ )
- **ROA** - рентабельность суммарных активов ( $ROA$ )
- **Rsale** - рентабельность продаж ( $R_{sale}$ )
- **ROCP** - рентабельность прямых затрат (ROCP)
- **ROCS** - рентабельность полных затрат на продажу (ROCS)

# Как читать показатели в решении



## Сдвиг:

В решении используется относительная нумерация исторических данных и показатели смещены относительно «текущего» года:

Обозначение	bef_04 (-4й)	bef_03 (-3й)	bef_02 (-2й)	bef_01 (-1й)	Текущий
Пример	2018	2019	2020	2021	2022

## Виды показателей:

- Прямой
- Абсолютная разница год-к-году
- Относительная разница год-к-году

### Прямой показатель:

Статья, раздел баланса, или рассчитанный коэффициент, выраженный в своей единице изм.

Пример: `current_assets_bef_01`

### Абсолютная разница год-к-году:

Разница значений показателя соседних лет (динамка)

Пример: `current_assets_diff`

Расчет: `current_assets_bef_01 - current_assets_bef_02`

### Относительная разница год-к-году:

Отношение прироста к показателю предыдущего года (темп):

Пример: `current_assets_diff_frac`

Расчет: `(current_assets_bef_01 - current_assets_bef_02) / current_assets_bef_02`

## Примеры:

`current_assets_diff_bef_01_02`

Разница стоимости оборотных активов прошлого года к позапрошлому.

`current_assets_bef_01 - current_assets_bef_02`

`current_assets_diff_frac_bef_01_02`

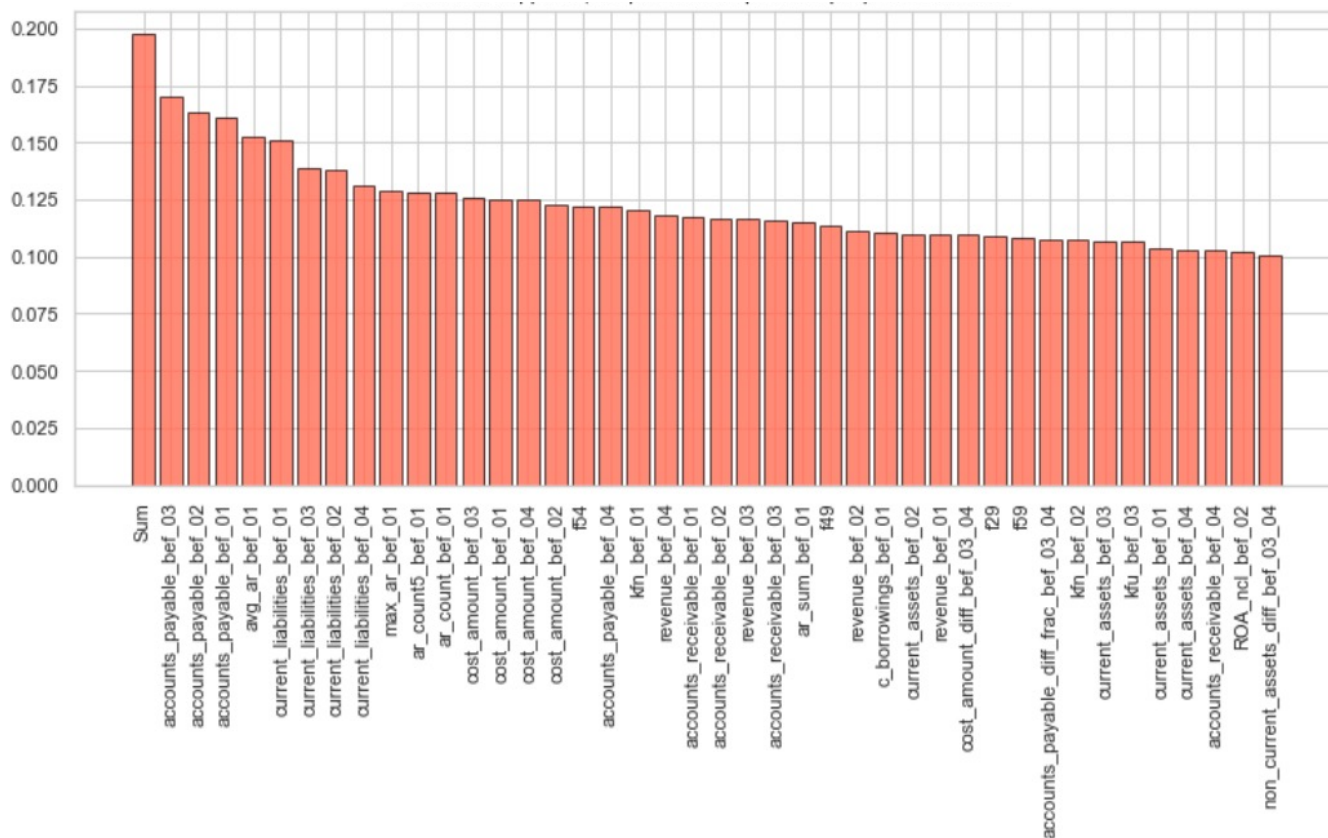
Отношение прироста оборотных активов прошлого года к позапрошлому.

$$\frac{(current\_assets\_bef\_01 - current\_assets\_bef\_02)}{current\_assets\_bef\_02}$$

# Feature Engineering



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

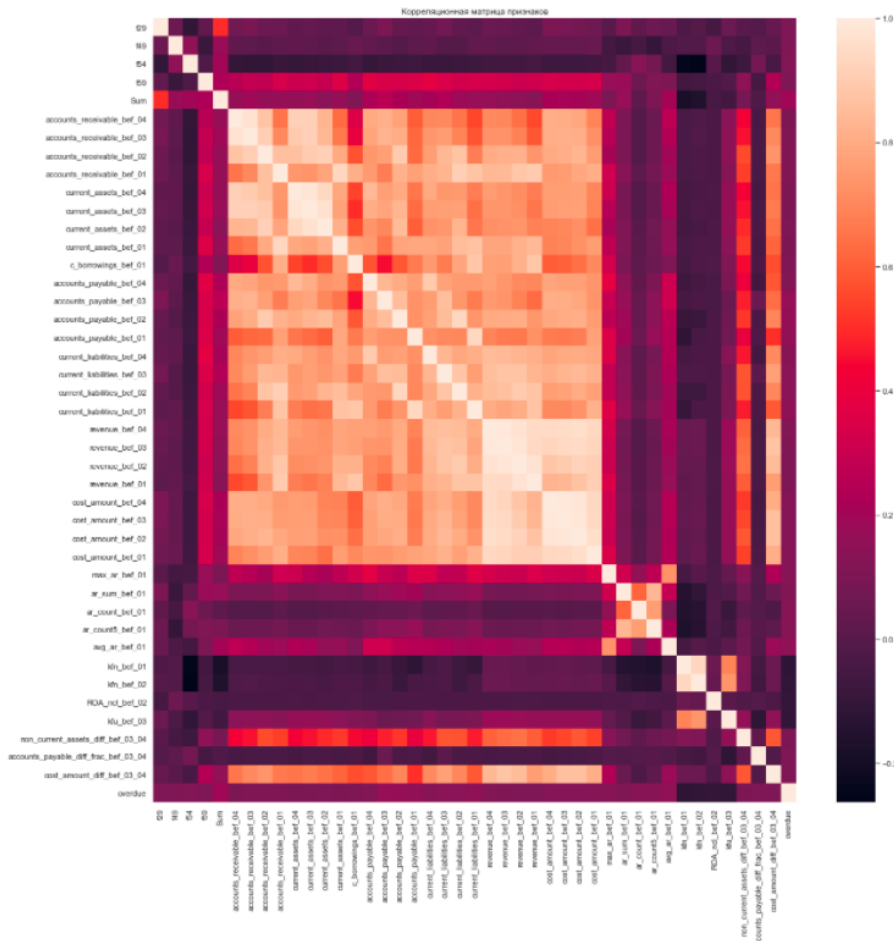


На графике отражены признаки, сила корреляции которых более 0.1

# Feature Engineering



Проверка корреляции между признаками



Статьи и разделы баланса из года в год коррелируют между собой.

Признаки, коррелирующие между собой с силой более 0.7 - удалим.

Количество признаков после предобработки

13

# Feature Engineering



## Проверка корреляции оставшихся признаков



Затем мы удалили «пустые» признаки и признаки с бесконечными значениями, и построили график Heatmap.

Высокой силы корреляции между отобранными признаками нет

Количество признаков после  
предобработки **7**



# Модель предсказания просрочки

Используемый модуль -

**LGBMClassifier**



Для старых контрагентов

Weight	Feature
0.1381 ± 0.0356	Sum
0.1000 ± 0.1061	kfu_bef_03
0.0905 ± 0.0819	ar_sum_bef_01
0.0810 ± 0.0571	ROA_ncl_bef_02
0.0762 ± 0.0555	non_current_assets_diff_bef_03_04
0.0667 ± 0.0555	cost_amount_diff_bef_03_04
0.0667 ± 0.0356	accounts_payable_diff_frac_bef_03_04
0.0238 ± 0.0426	max_ar_bef_01

Наиболее значимым признаком для модели, обученной на полном множестве отобранных признаков, получился признак "Sum" - сумма факторных признаков.

Для новых контрагентов

Weight	Feature
0.1143 ± 0.1360	kfu_bef_03
0.0905 ± 0.1292	Sum
0.0333 ± 0.1069	ROA_ncl_bef_02
0.0238 ± 0.0602	non_current_assets_diff_bef_03_04
0.0095 ± 0.0883	cost_amount_diff_bef_03_04
-0.0524 ± 0.0632	accounts_payable_diff_frac_bef_03_04

Наиболее значимыми для модели оказались признаки "kfu\_bef\_03" и "Sum".

Weight	Feature
0.2857 ± 0.1313	Sum
0.1571 ± 0.0883	kfu_bef_03

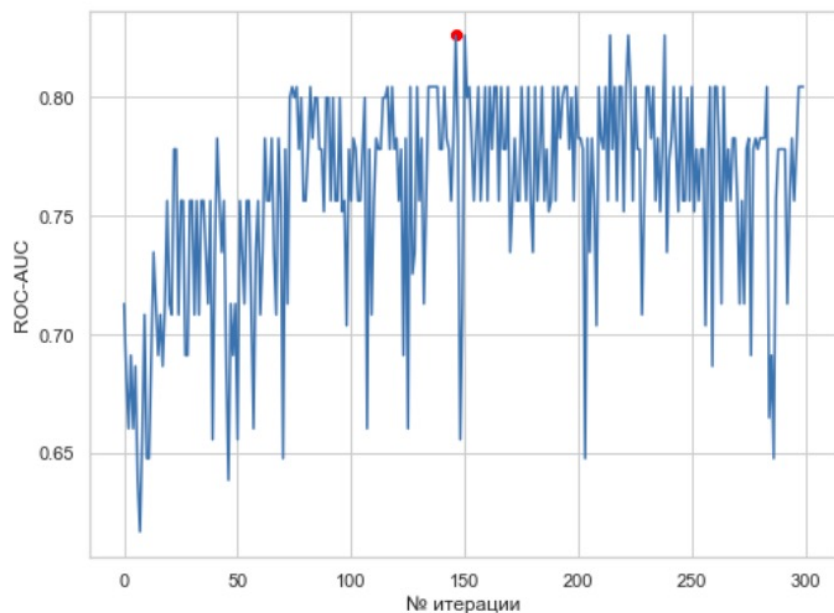
Важность признаков при использовании попарной комбинации 6 признаков



# Модель предсказания просрочки

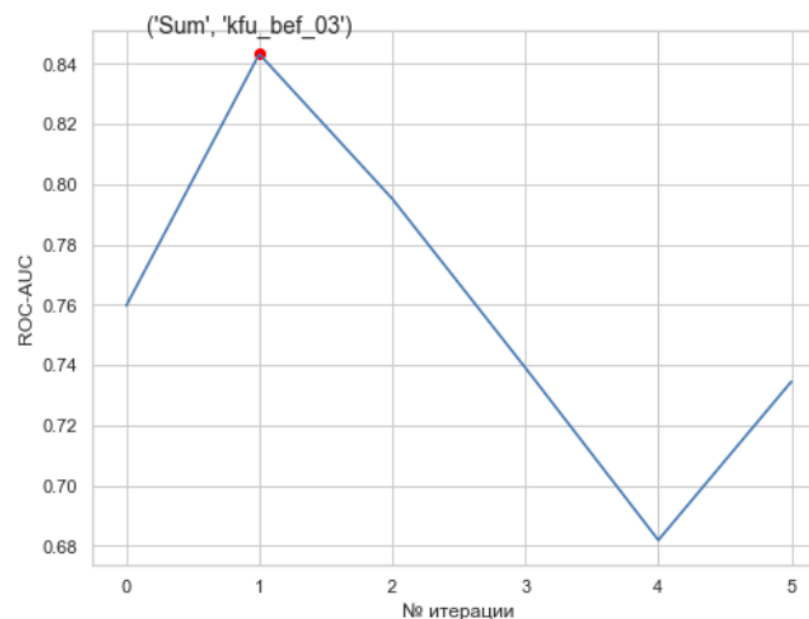


## Для старых контрагентов



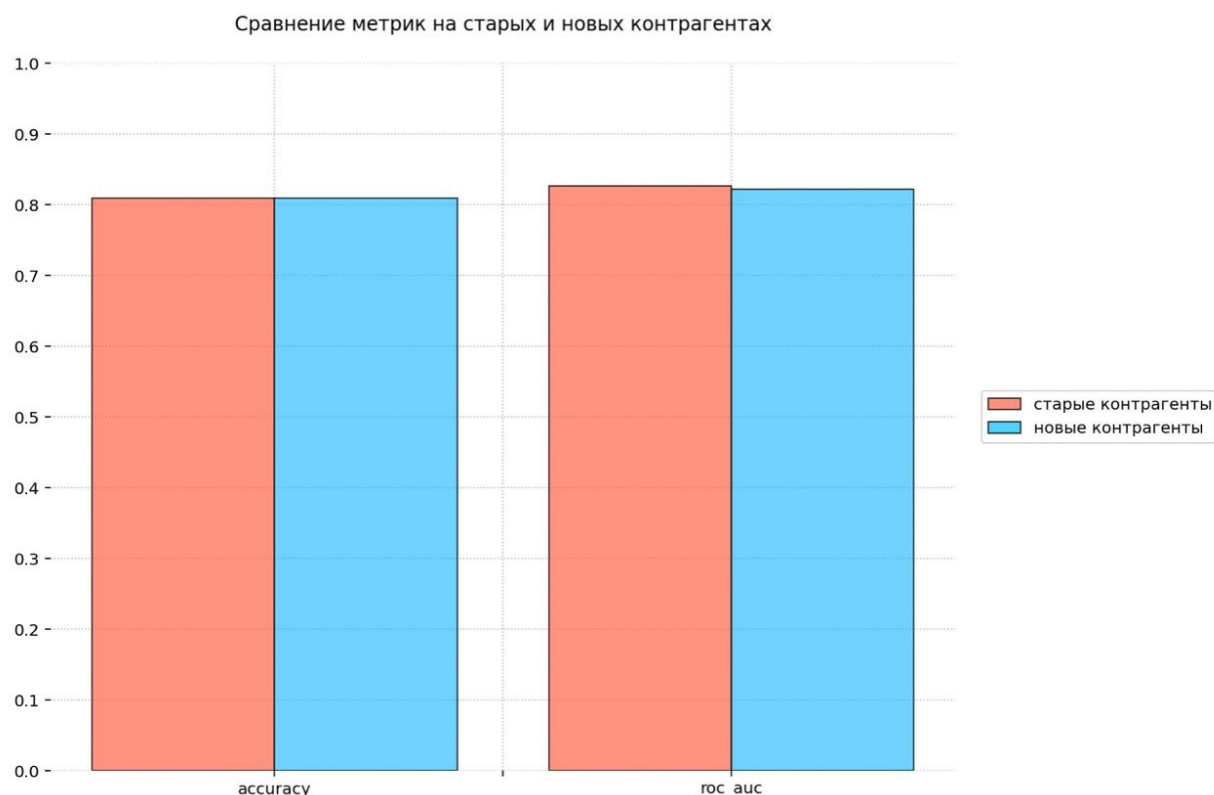
**Вывод:** для предсказания факта ПДЗ по старым контрагентам необходимо использовать все отобранные признаки, потому что они отлично дополняют друг друга, создавая синергетический эффект. **Точность предсказаний ~81%.**

## Для новых контрагентов



**Вывод:** для предсказания факта ПДЗ для новых контрагентов лучше использовать комбинацию следующих признаков: "Sum" и "kfu\_bef\_03"; эта комбинация позволяет получить **Точность предсказаний ~81%.**

# Модель предсказания просрочки - Вывод



В ходе проделанной работы по анализу контрагентов и выявлению факторов просрочки дебиторской задолженности были промоделированы две ситуации:

- взаимодействие со старыми контрагентами, т.е. с таким, с которым было сотрудничество ранее;
- взаимодействие с исключительно новыми контрагентами.

Эти ситуации влияют на признаки, которые брались для обучения модели.

## Результаты

[Код на GitHub](#)



>350

Признаков проанализировано

7

Признаков выбрано для обучения модели



Предложено 3 подхода к моделированию

82%

Сформирован список дополнительных данных для повышения точности прогноза



**Спасибо  
за внимание**

