

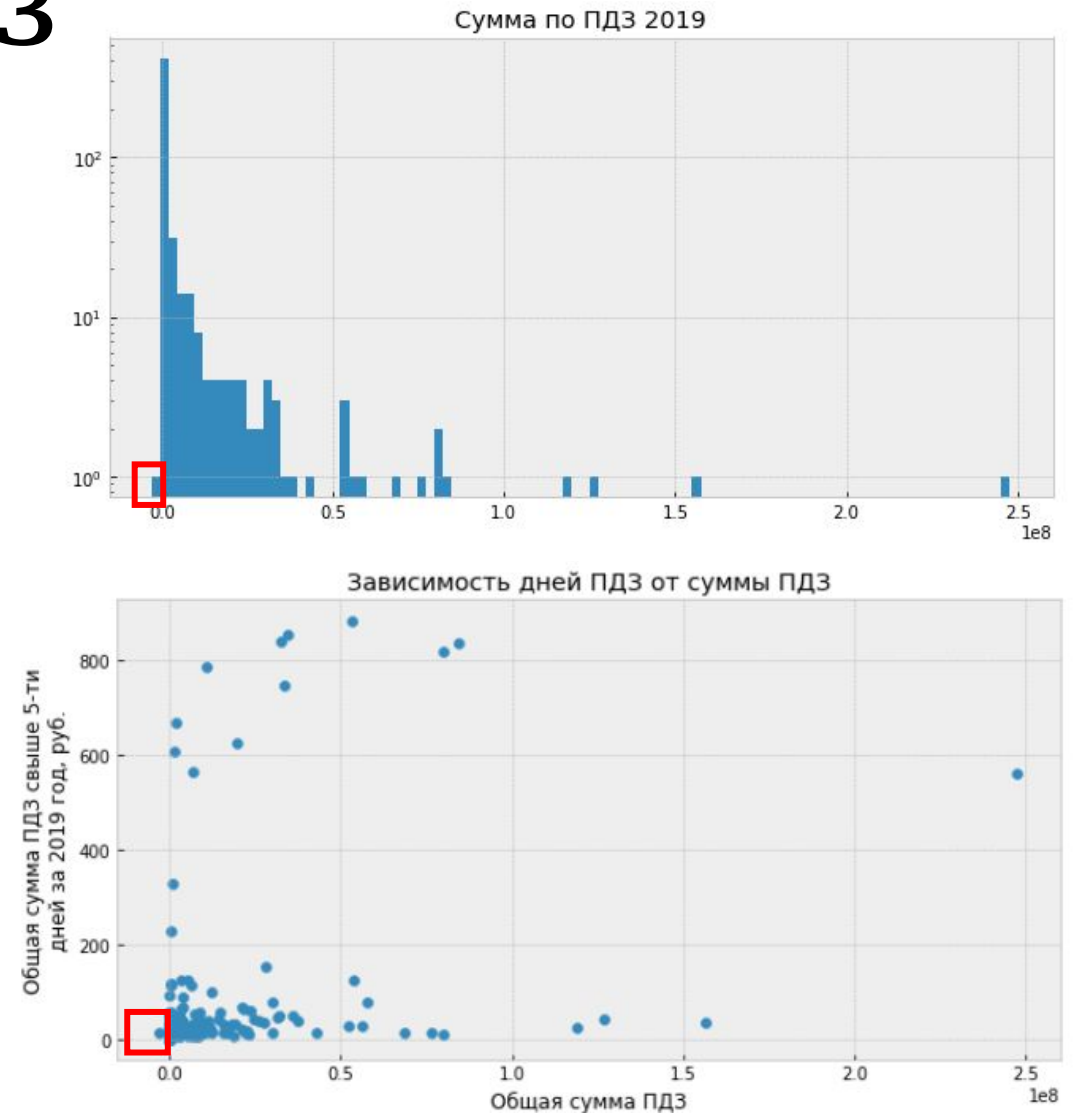
# Task 4. Анализ Контрагентов

# 1. Первичный анализ

- Отсутствие явных корреляций с целевыми признаками,
- Очистка данных:
- замена отрицательных
- выделение пустых значений в категории
- заполнение значений с учетом тренда

## Выводы:

- использование внешних источников
- синтез признаков отражающих тенденции с течением времени (тренд, ускорение, итд.)



## 2. Внешние источники /дополнительные признаки

Для каждого признака, имеющего историю (три года):

1. Добавляем категорию «нет сведений о признаке»
2. Добавляем значение динамики признака – получение тренда по историческим данным

Признаки из внешних источников:

1. Индекс цен производителей по отраслям: производство деталей из порошковой металлургии
2. Рост ВВП(реальный) в процентах (со смещением)
3. Уровень безработицы в % (со смещением)
4. Уровень инфляции в % (со смещением) Стоимость акции (CHMF) на начало года
5. Индекс (LMEX) Лондонская биржа металлов

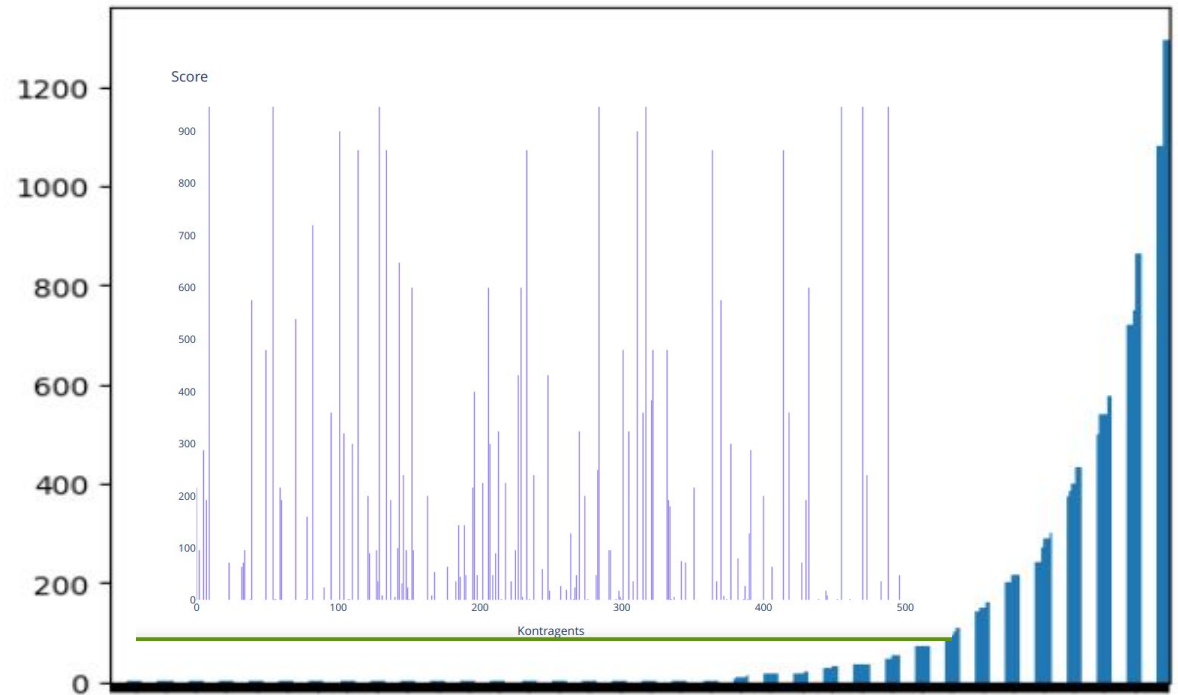


### 3. Скоринг контрагента

Одним из характеристик при принятии решения может служить построенный мультипликатор характеризующий степень благонадежности контрагента с течением времени на основе сведений:

- Макс. ПДЗ,
- Общая сумма ПДЗ свыше 5-ти дней,
- Кол-во просрочек свыше 5-ти дней,
- % Длительных просрочек

*Реализация системы балльной оценки контрагента позволит для ЛПР доступно интерпретировать благонадежность и степень риска, а также принять адекватное решение о возможности выдачи займа.*



Распределение благонадежности компаний в балльной оценке из компонент разбитых на бины.<sup>4</sup>

## 4. Выбор и обучение модели

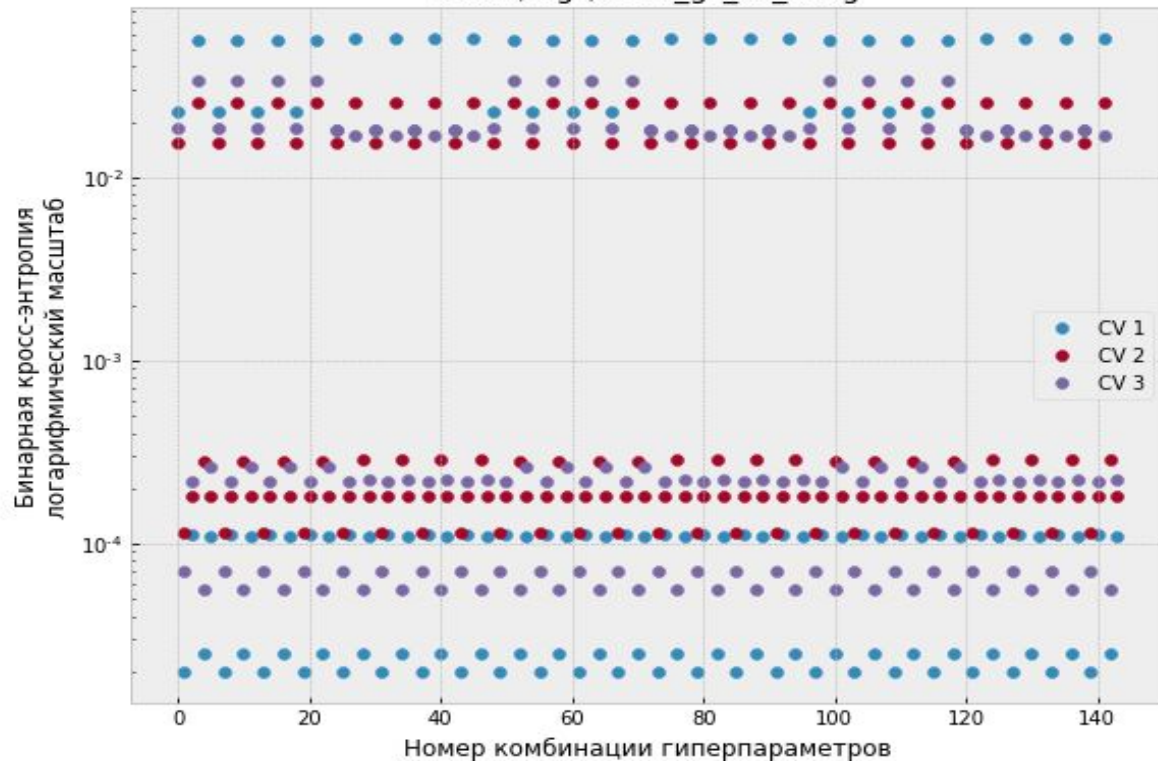
- От задачи регрессии переходим к задаче бинарной классификации - по признаку «Макс ПДЗ» устанавливаем порог, в решении были применены следующие пороги: 0, 30, 60. Данные классификаторы возможно использовать для предсказания факта задолженности, а также для задолженности меньше 30 дней, 30-60 дней и более 60 дней.
- После этого был произведен поиск «по сетке» гиперпараметров, и построено 3 классификатора (CatBoost).

**Accuracy** (0.70, **0.91** и **0.90** соответственно)

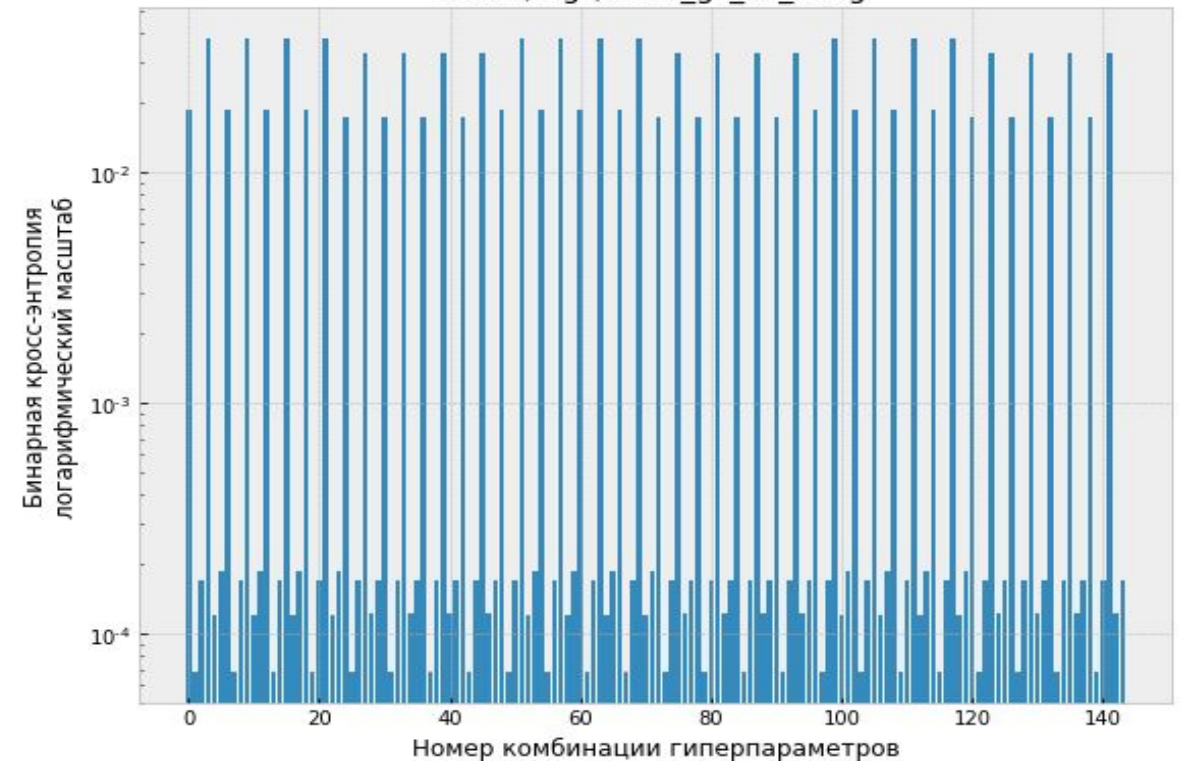
- Далее был проведен анализ значимости признаков для их дальнейшей интерпретации и синтеза новых признаков.

# 4. Результаты построенной классификационной модели

Поиск по сетке (CV на 3 фолда)  
Лог: ./logs/2020\_gs\_cb\_0.log



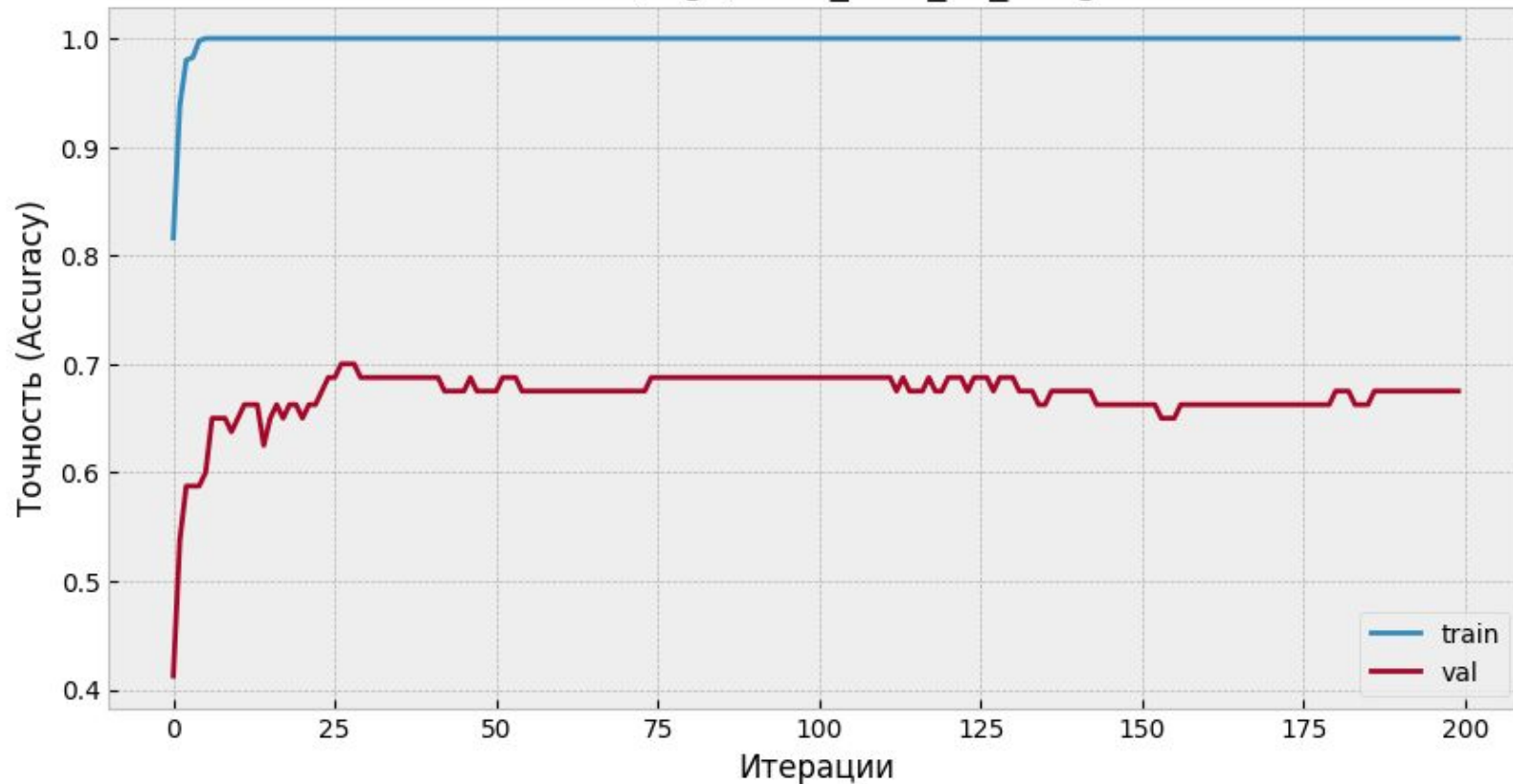
Поиск по сетке (CV на 3 фолда - среднее)  
Лог: ./logs/2020\_gs\_cb\_0.log



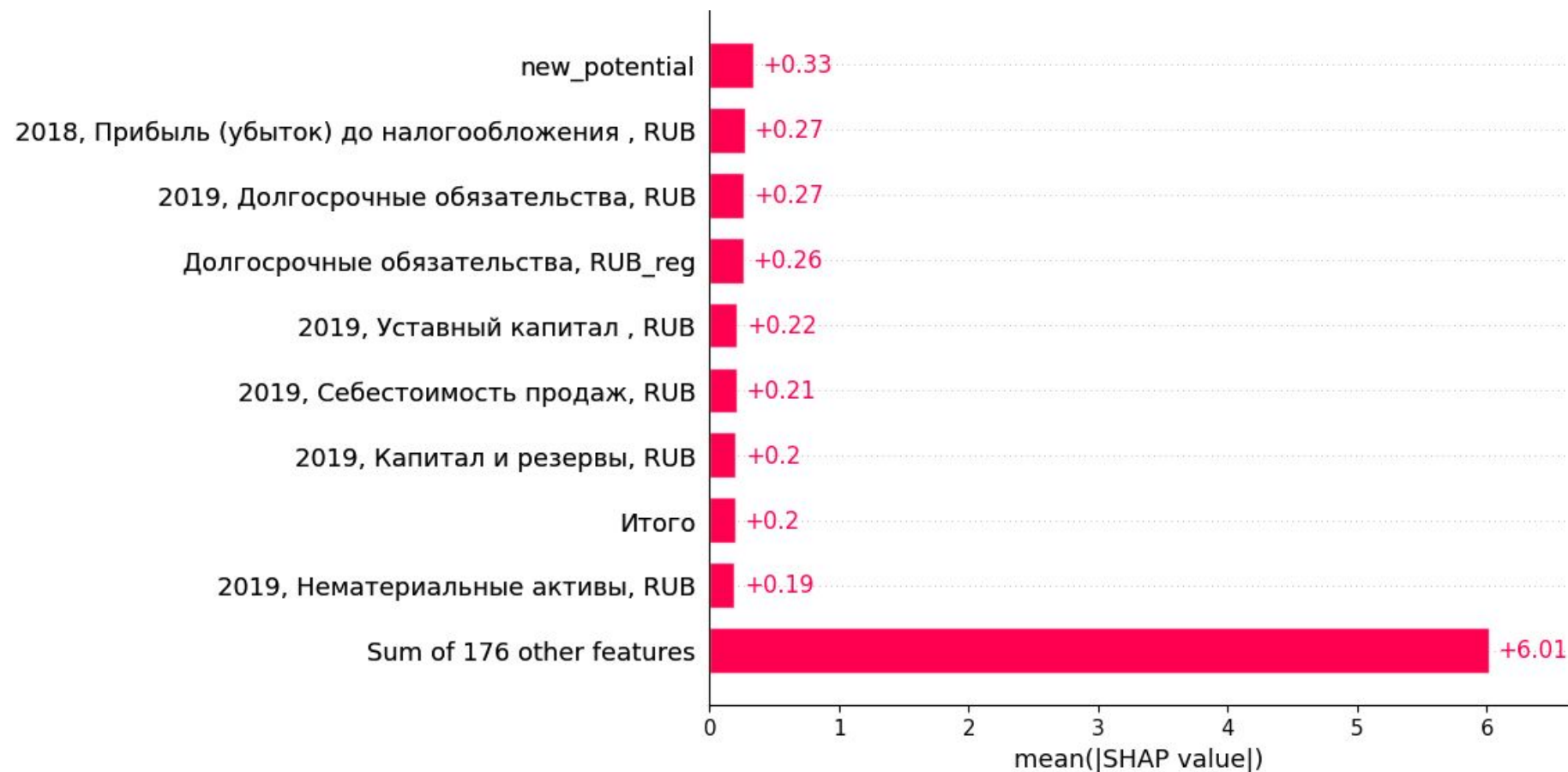
На графиках приведены результаты работы классификатора для порога 0. Ассурасу: 0.70

# 4. Результаты построенной классификационной модели

Обучение лучшего классификатора  
Лог: ./logs/2020\_best\_cb\_0.log



## 4. Обучение классификатора



Коэффициенты считались на всех доступных данных (включая валидационные).  
\* Признаки \_reg – тренд по историческим данным, \_cat – наличие исторических данных, new\_potential – мультипликатор характеризующий финансовый потенциал



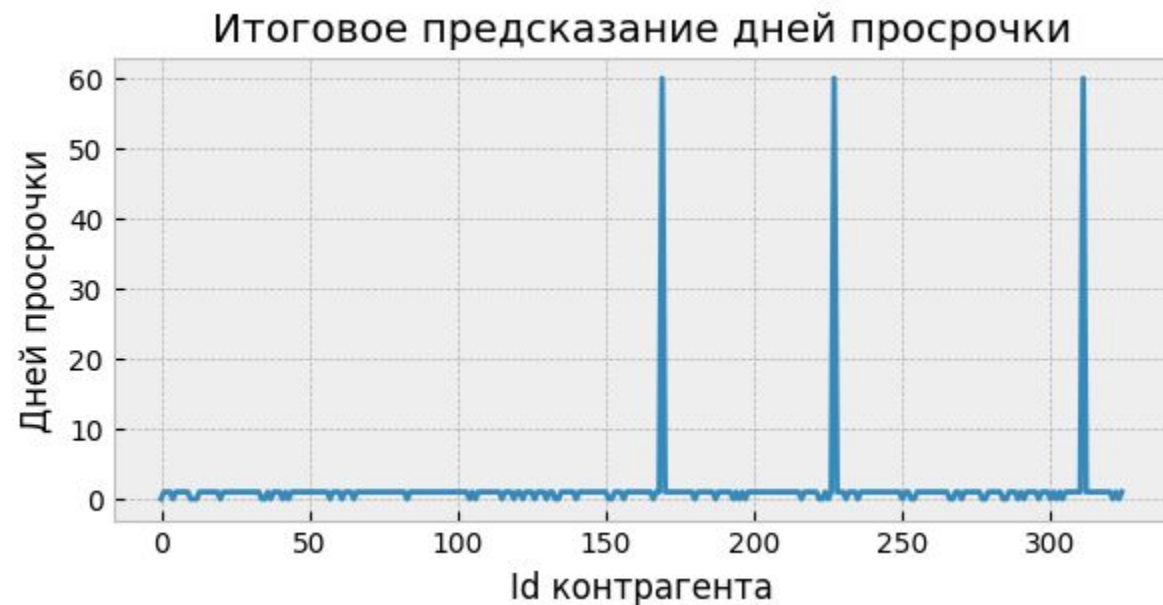
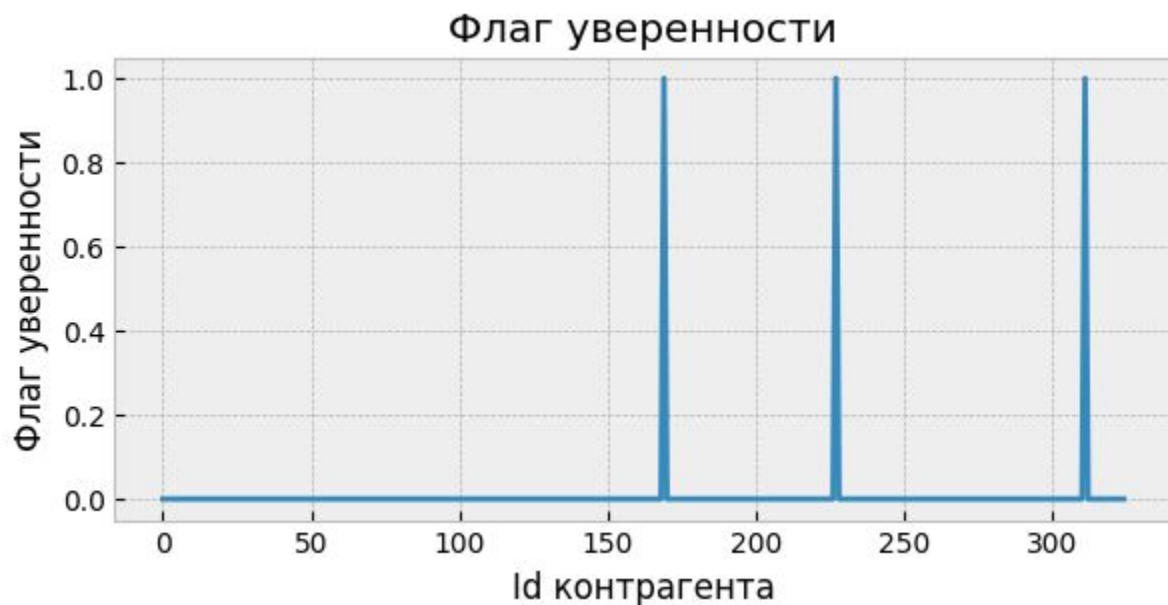
## 5. Получение прогноза для 2021 года

Сведем оценку классификаторов к следующей логике:

- Если классификаторы подряд оценивают класс просрочки как положительный, финальная оценка - больший (по количеству дней) классификатор и его "флаг уверенности" 1.
- Если порядок нарушен (например Просрочка  $> 0 == 0$ , Просрочка  $> 30 == 1$  - оценка больший (по количеству дней) классификатор и его "флаг уверенности" 0.

Таким образом флаг уверенности выступает индикатором согласованности предсказаний классификаторов.

## 5. Получение прогноза для 2021 года



Модель предсказывает достаточно много незначительных просрочек (менее 30 дней), и 3 большие просрочки (более 60 дней) для данных 2021 года.

## 6. Рекомендации по улучшению прогноза

1. Необходимы сведения о каждом займе (дата, сумма), что позволит привязать более адекватные значения макроэкономических показателей.
2. Наличие сведений о принудительно взысканных годах.
3. Вид экономической деятельности.
4. Количество поставок для данного контрагента.
5. Является ли Северсталь главным поставщиком контрагента.

# 4. Рекомендательная система [!версия для 2019 года!]

Любая аналитика необходима в конечном счете для лица, которое на ее основе принимает решение. Таким образом, необходимо сделать понятную неспециалисту систему, которая бы не была перегружена техническими подробностями, а давала понятную информацию.

В нашем случае она объединяет 3 сущности:

1. Предсказание классификаторов – дает в процентах оценку надежности контрагента.
2. Графическое представление наиболее значимых признаков, повлиявших на данную оценку.
3. Сводку по уже имеющимся данным, в разрезе признаков из п. 2.<sup>12</sup>

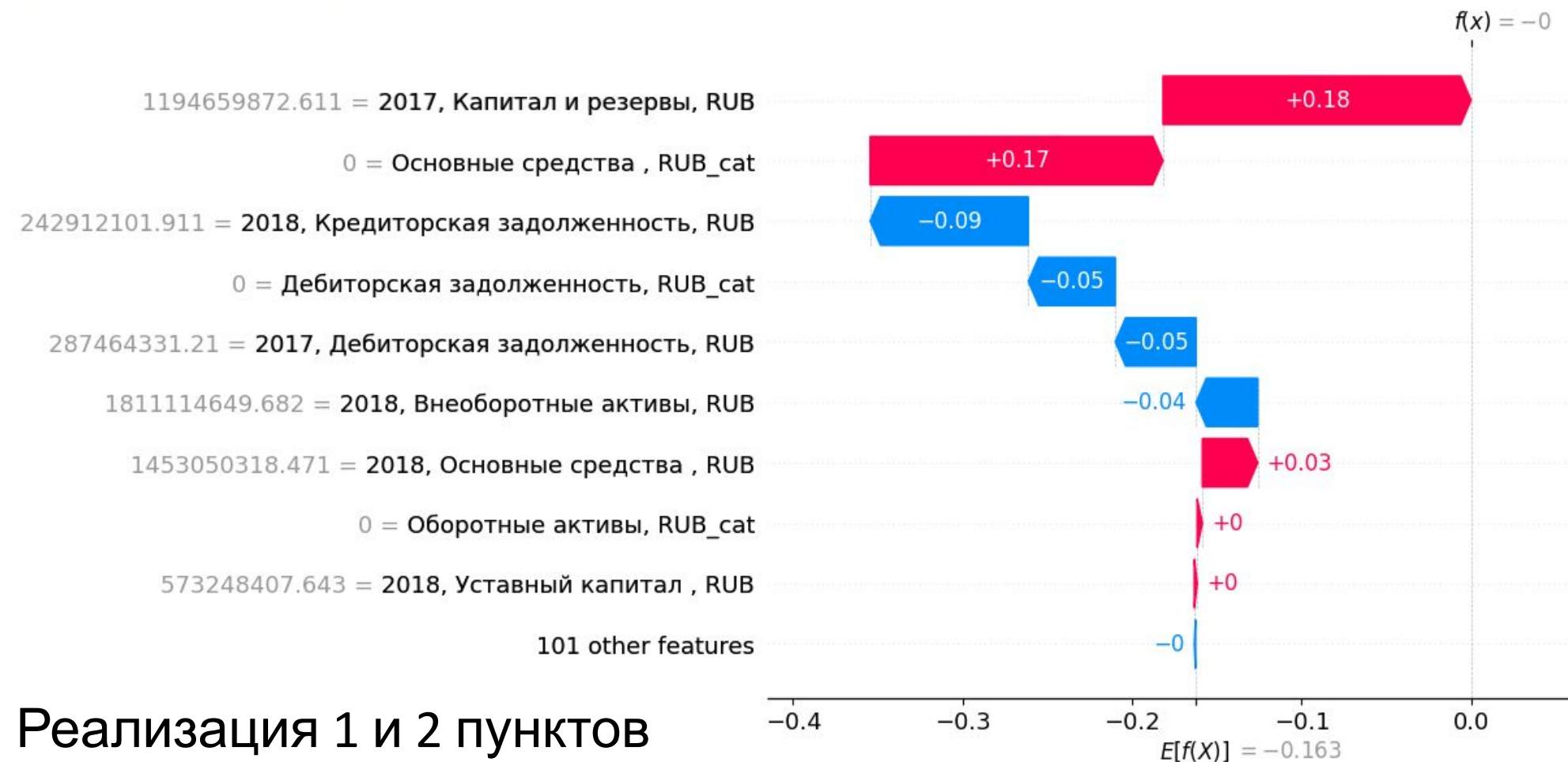


# 4. Рекомендательная система

Введите индекс строки для построения отчета: 500

Контрагент 500 классифицирован как 'контрагент с высоким риском' [84.4%]

Наиболее значимые параметры контрагента, повлиявшие на решение:



Реализация 1 и 2 пунктов

## 4. Рекомендательная система

===== Сводка по наиболее значимым признакам=====

[предполагается наличие исторических данных]

['Аналогичные' контрагенты - процент, допустивших ПДЗ по данному критерию (выше/ниже) в зависимости от 'Вклада']

	Признак	Вклад	Аналогичные
0	2018, Кредиторская задолженность, RUB	-0.093322	21.8
1	Дебиторская задолженность, RUB_cat	-0.051280	1.1
2	2017, Дебиторская задолженность, RUB	-0.047656	16.2
3	2018, Внеоборотные активы, RUB	-0.036394	9.4
4	2018, Оборотные активы, RUB	-0.000654	19.4

	Feature Id	Importances
0	Trend Заёмные средства (краткосрочные), RUB	8.819509
1	2018, Уставный капитал, RUB	7.294156
2	Trend Кредиторская задолженность, RUB	5.944278
3	Trend Заёмные средства (долгосрочные), RUB	5.144275
4	2018, Дебиторская задолженность, RUB	4.760046



### Реализация 3 пункта

В приведенной «тетрадке» данная идея обернута в класс при помощи которого возможно интерактивное взаимодействие.