



Решение Андеррайтера

**Быстрое одобрение,
сбор данных, которых не было в
витрине**



Цель

- Основная цель: улучшить модель.
- Для этого нужно:
 - - рассмотреть данные, которых нет в витрине
 - - выбрать наиболее эффективные
 - - убедиться на лучше модели, что данные улучшают точность
 - - обновить синтетическую модель с учетом НОВЫХ данных



Входные данные

Заявки в диапазоне дат: 2020-01-01 по 2022-02-13

Витрина MART_NORMA_AUTO: x строк, x столбцов.

Одобрено 30367

- Отмена 54615
- Отмена+Одобрено 84982

Выбор количества прошлых кредитов

Будем рассматривать

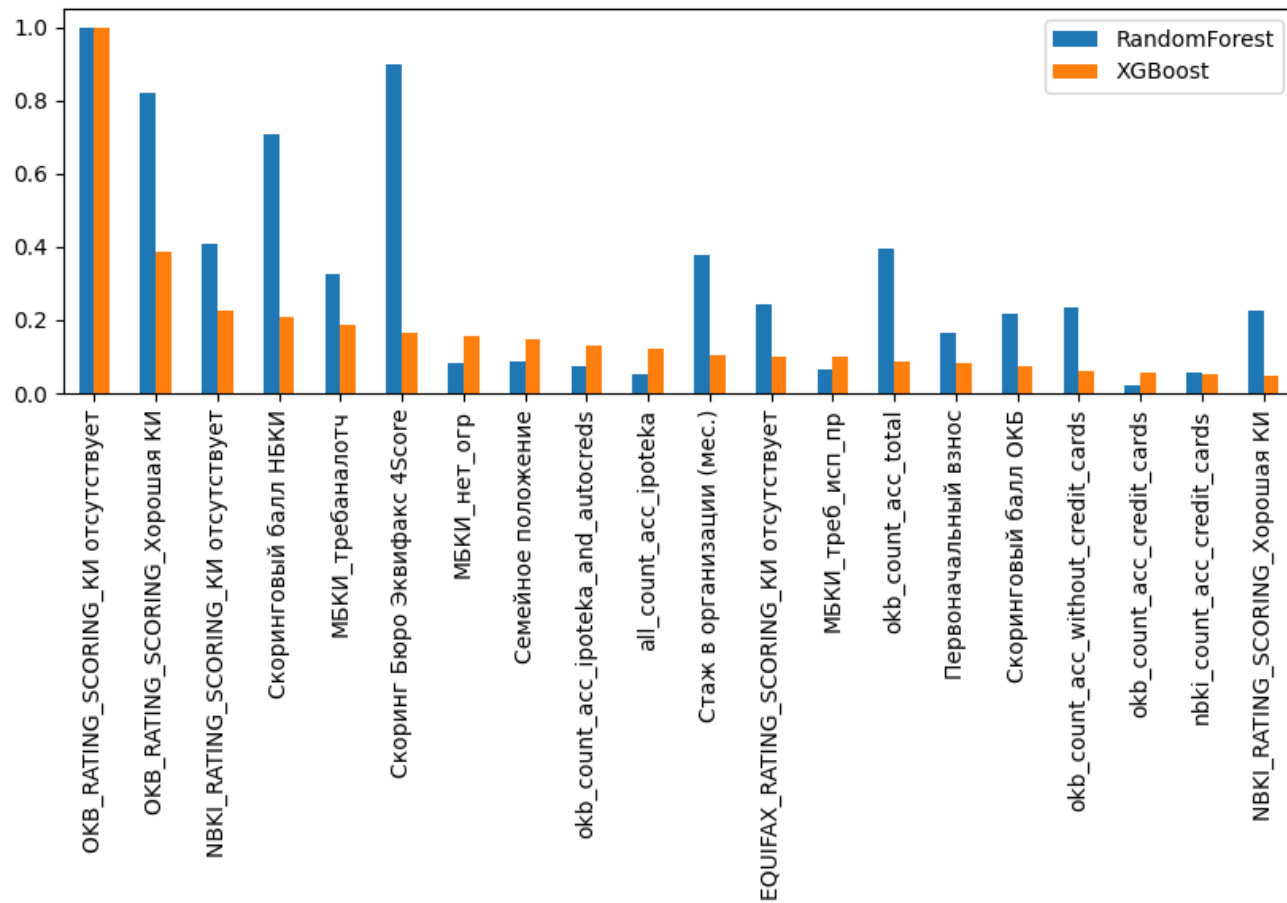
- общее количество крдитов (?_count_acc_total)
- количество кредитов без типа "Кредитная карта" (?_count_acc_without_credit_cards)
- количество кредитов типа "Кредитная карта" (?_count_acc_credit_cards)
- количество кредитов для типа Ипотека и Автокредит (?_count_acc_ipoteka_and_autocreds)
- количество кредитов для типа Ипотека (?_count_acc_ipoteka)
- количество кредитов для типа Автокредит (?_count_acc_autocreds)

Для кредитных историй:

- НБКИ (nbki)
- ОКБ(okb)
- Эквифакс(equifax)
- Консолидированная кредитная история(all)

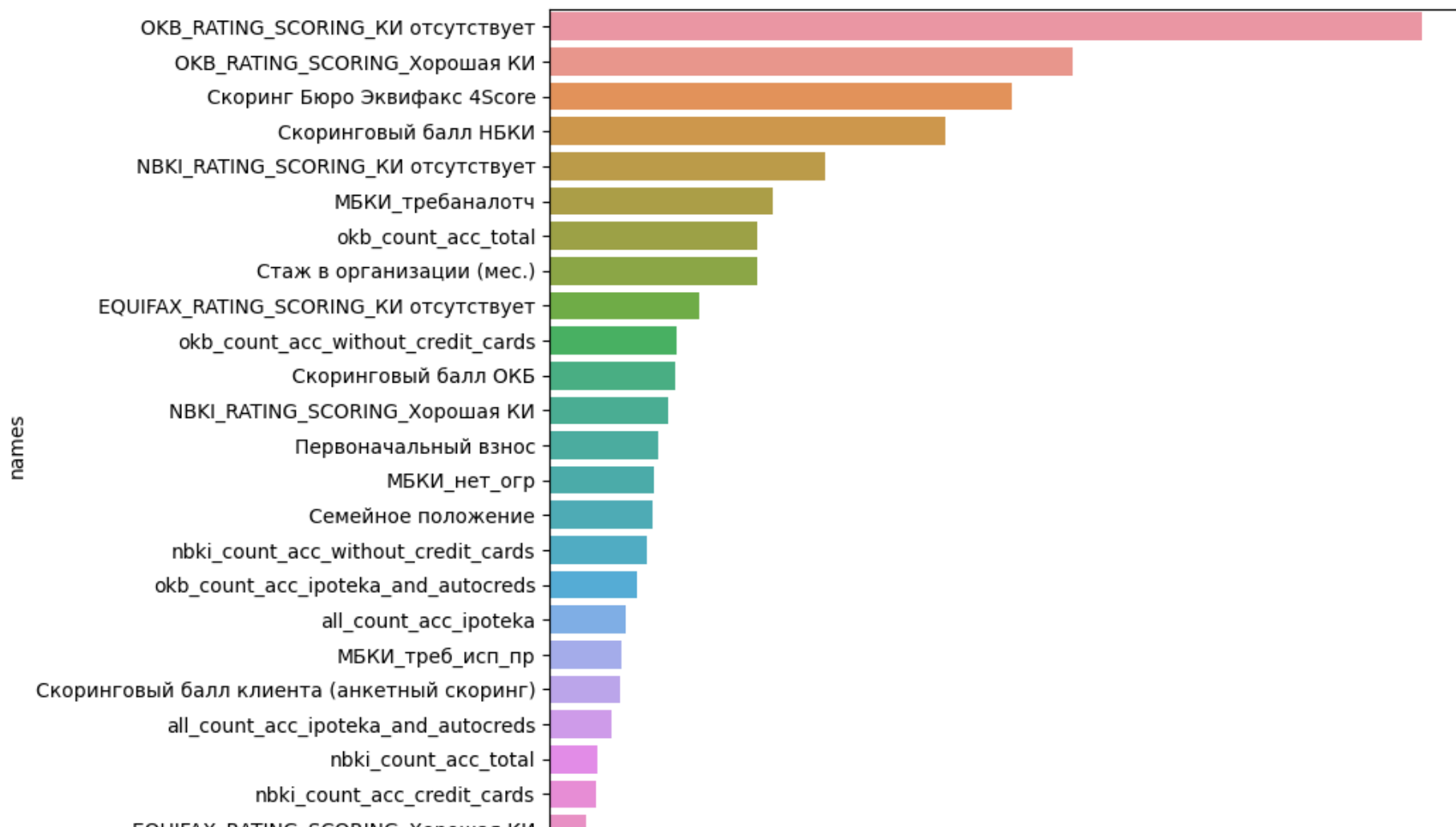
Выбор количества прошлых кредитов

Важность признаков для XGBoost и RandomForest



Выбор количества прошлых кредитов

Важность признаков сумма XGBoost и RandomForest





Выбор количества прошлых кредитов

В итоге оставлены наиболее влиятельные поля

- okb_count_acc_total
- okb_count_acc_without_credit_cards
- nbki_count_acc_without_credit_cards
- all_count_acc_ipoteka

Как мы видим общее количество от ОКБ оказалось наиболее достоверным.

Рассмотрение кредитов без Кредитных карт может быть полезным.

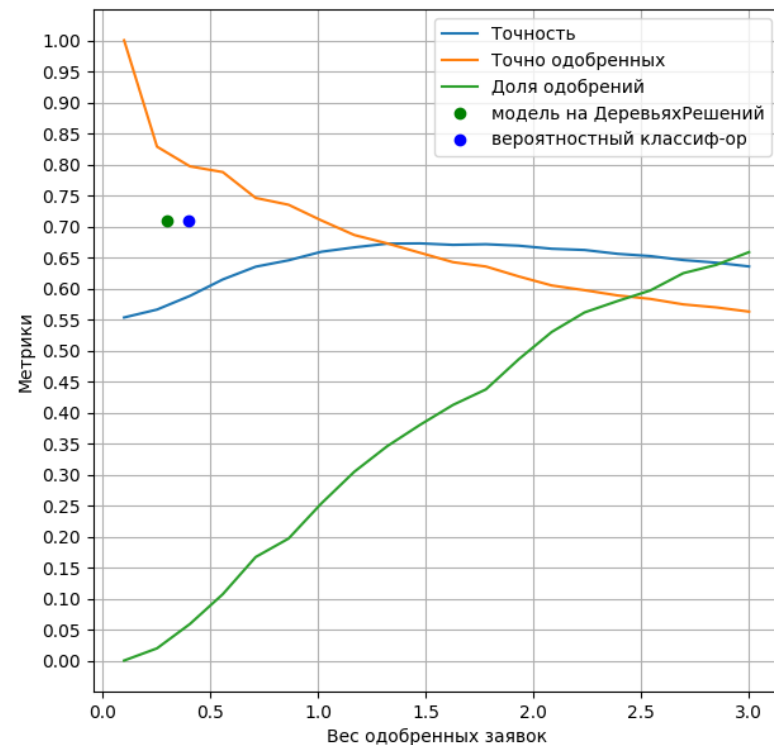
Наличие ипотечных кредитов тоже может быть фактором.

А связка Ипотека+Автокредиты — себя не оправдала для количественной оценки.

Выбор количества прошлых кредитов

Лучшая модель «ДО»

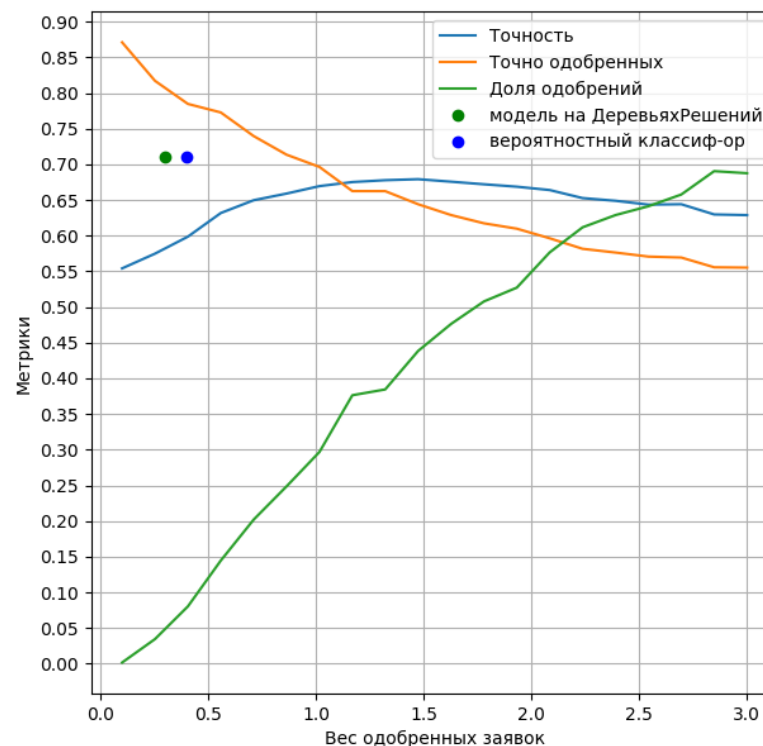
•



Выбор количества прошлых кредитов

Лучшая модель «ПОСЛЕ»

- Под голубой линией увеличилась площадь на 1-2%
- На 1% увеличилось количество одобренных. На пол процента увеличилась точность.



Выбор статистики по просрочкам

Возьмем просрочку для:

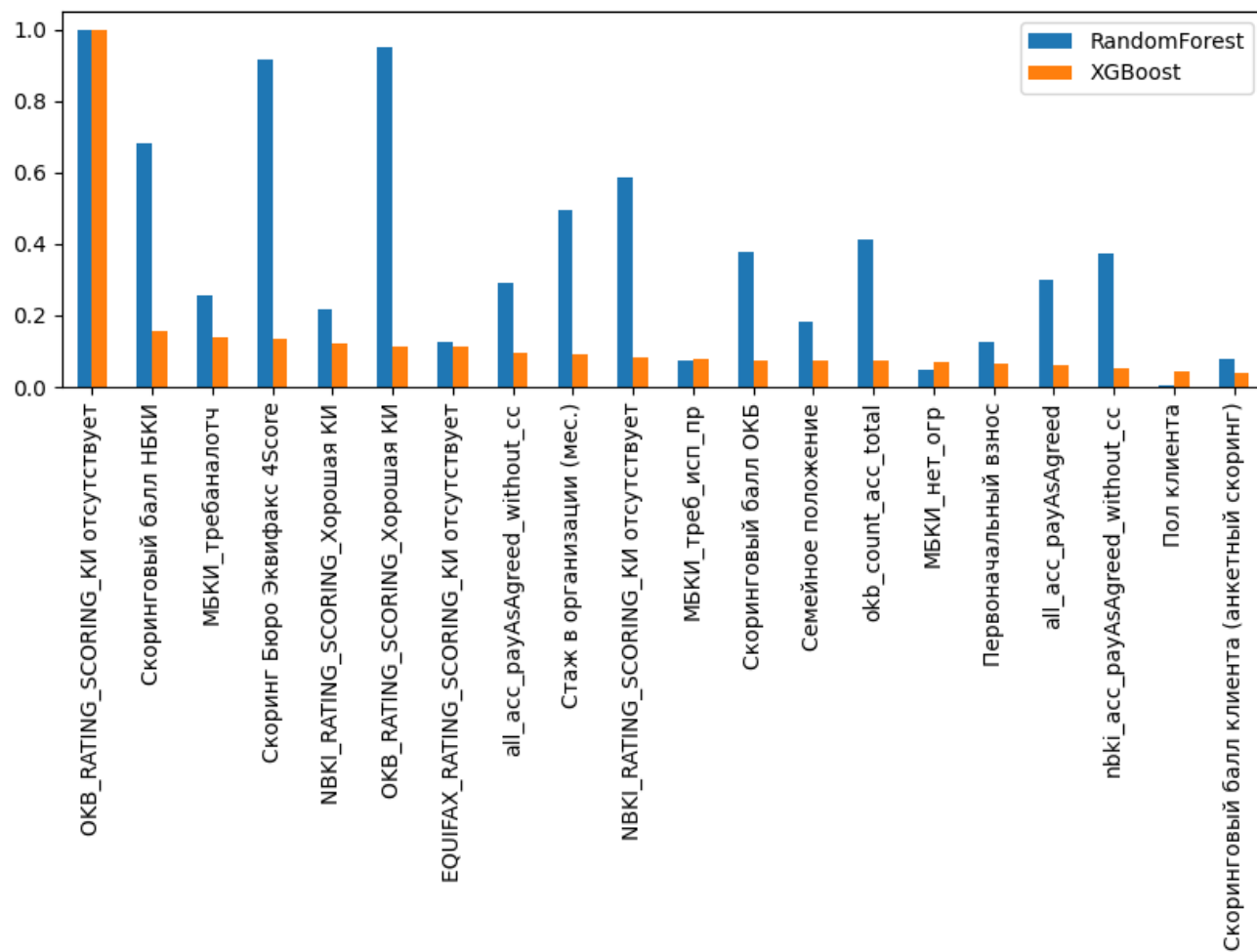
- Оплата без просрочек
- Просрочка от 1 до 7 дней — включает в себя Просрочка от 1 до 5 дней
- Просрочка от 1 до 29 дней
- Просрочка от 8 до 29 дней
- Просрочка от 30 до 59 дней
- Просрочка от 60 до 89 дней
- Просрочка от 90 до 119 дней
- Просрочка более 120 дней

- Проделаем это для
- - кредитов без Кредитных карт и для кредитов по Кредитным картам
- ипотека и автокредиты
- - ипотека
- - автокредиты

У нас получилось 18*4 новых полей для проверки. Значения берутся как отношение просрочек к количеству выплат.

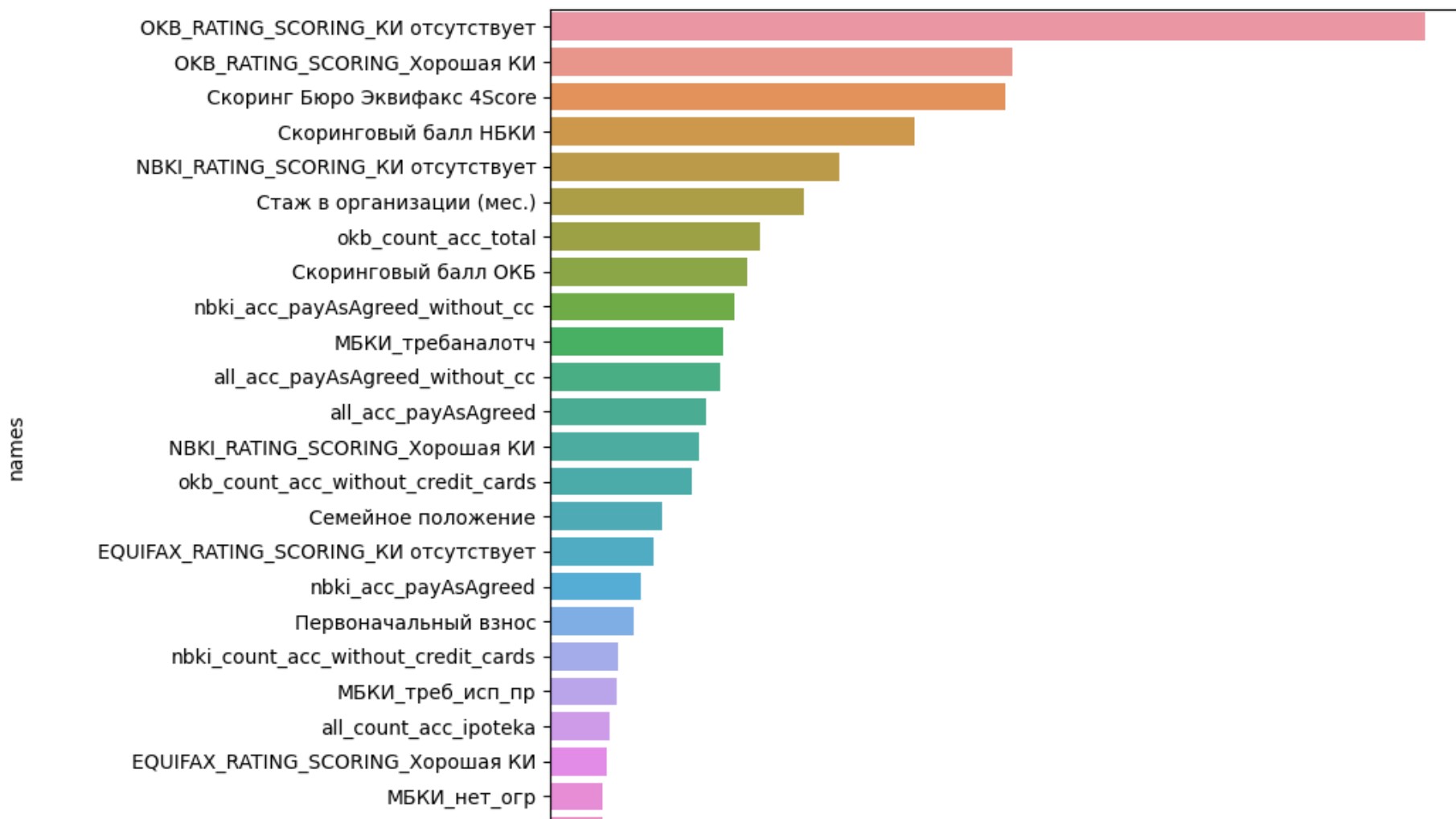
Выбор статистики по просрочкам

Важность признаков для XGBoost и RandomForest



Выбор статистики по просрочкам

Важность признаков сумма XGBoost и RandomForest



Выбор статистики по просрочкам

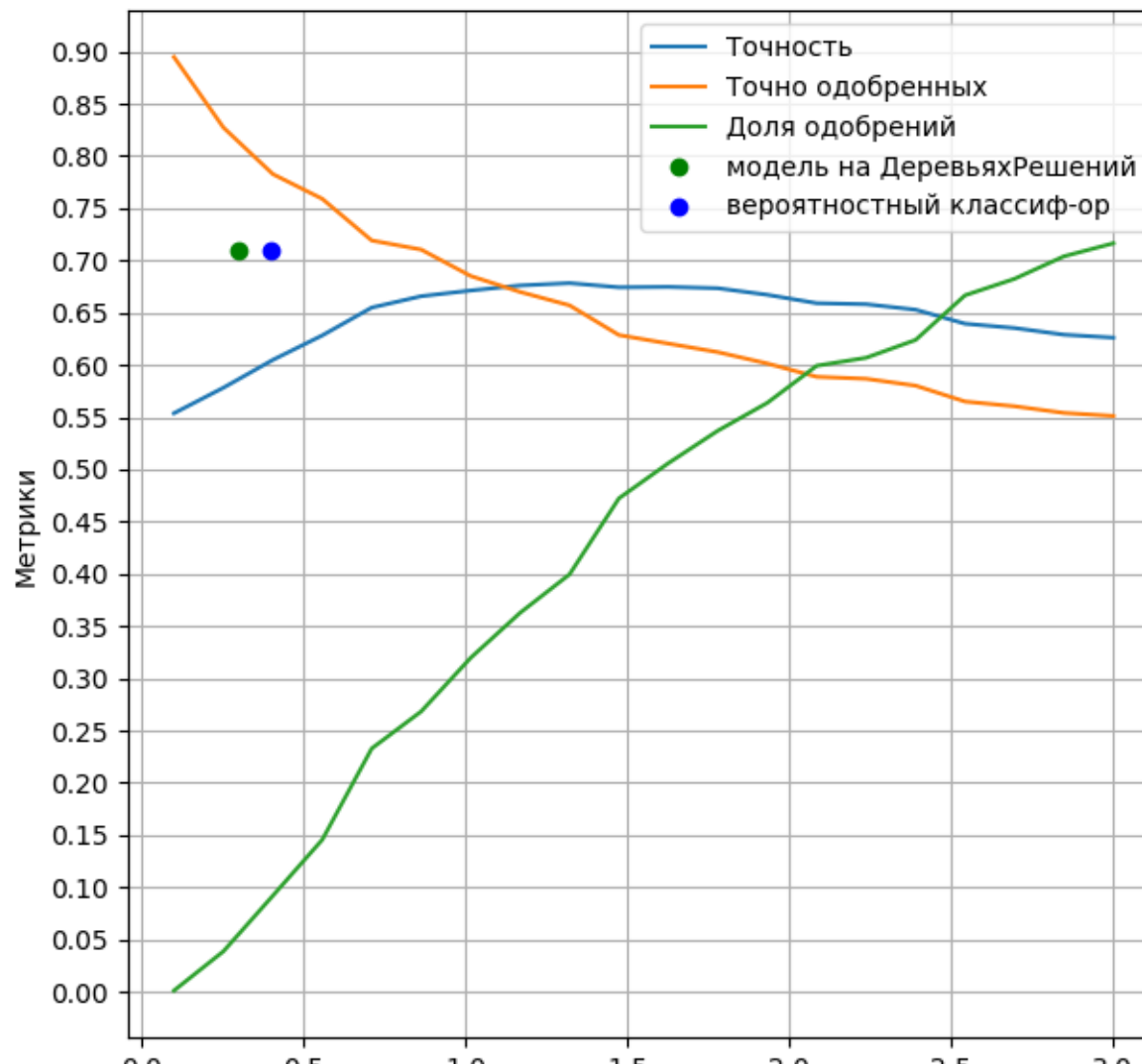
Отмечаем

- - nbki_acc_payAsAgreed_without_cc
- - all_acc_payAsAgreed_without_cc
- - nbki_acc_payAsAgreed
- - all_acc_payAsAgreed

НБКИ показал себя значительно лучше чем ОКБ. Отношение вовремя оплаченны выплат ко всем выплатам ожидаемо показало хороший результат.(Нужно проверить что нет ошибки)

- Количество выплат без Кредитных карт показало себя лучше чем общее количество выплат без просречки.
-

Выбор статистики по просрочкам



Статистика по просрочкам

два

Возьмем

- - количество активный счетов
- - количество активных с суммой больше 50 тыс.
- - количество активных с суммой меньше= 50 тыс.
- - количество закрытых с суммой больше 50 тыс.
- - количество закрытых с суммой меньше= 50 тыс.
- - средний процент просрочек с 1 до 29 дней для закрытых
- - средний процент просрочек от 29 до 120 дней для закрытых

- Проделаем это для
- - НБКИ (nbki)
- - ОКБ(okb)
- - Эквифакс(equifax)
- - Консолидированная кредитная история(all)

А так же, возьмем

- - НБКИ процент PayAsAgreed с суммой больше 50 тыс.
- - НБКИ процент PayAsAgreed с суммой меньше 50 тыс.
- - консолидирования PayAsAgreed с суммой больше 50 тыс.
- - консолидирования PayAsAgreed с суммой меньше= 50 тыс.
- У нас получилось $7*4+4 = 32$ новых поля

Статистика по просрочкам

два

- Используя аналогичные шаги как в пунктах выше, выбираем следующие поля:
- - количество закрытых с суммой больше 50 тыс.
- - количество закрытых с суммой меньше= 50 тыс.
- - средний процент просрочек от 29-до 120 дней для закрытых

Дополнительные поля

Возьмем и распарсим скоринги:

- Проверка организации-работодателя на наличие в списке ликвидированных(0prov):
status_code, status_type, status_groupid, status_isActing, status_groupName, companyType, okfs_code, includeInList, change_history, legalAddresses
 - - Эквивалент 4Score: scor
 - - AFS НБКИ: rules_count
 - - ОКБ National hunter: 'MA_SMT', 'MA_RAD', 'MA_AS', 'LCL_MA', 'MULT_M', 'MA_PAS', 'MA_EMP', 'MA_SPE', 'MA_MS_', 'MA_REF', 'MA_MTE', 'MA_SPA', 'MA_MS', 'MA_SAM', 'MA_AS_', 'MA_SWT'
 - - НБКИ Биометрия: matchImages, matchResults, match_max, match_avg

Рассматривая корреляцию полей и их важность убираем коррелирующие с низкой важностью:

- - 0prov_status_groupid
- - ex4scor_scoring
- - nbki_biom_resp_match_max
- - 0prov_status_isActing
- - 0prov_status_groupName_Действующее
- - nbki_biom_resp_matchResults
- - okbnh_MA_SPE

Дополнительные поля — частотный анализ

- отрицательное влияние:
- - 0prov_okfs_code: при значении - 54 (14,1)
- - okbnh_MA_MTE при значении — 1 (257,44)
- - okbnh_MA_MS — 1 (18,4)
- - nbki_biom_resp_matchImages — больше девяти(~500, ~100)

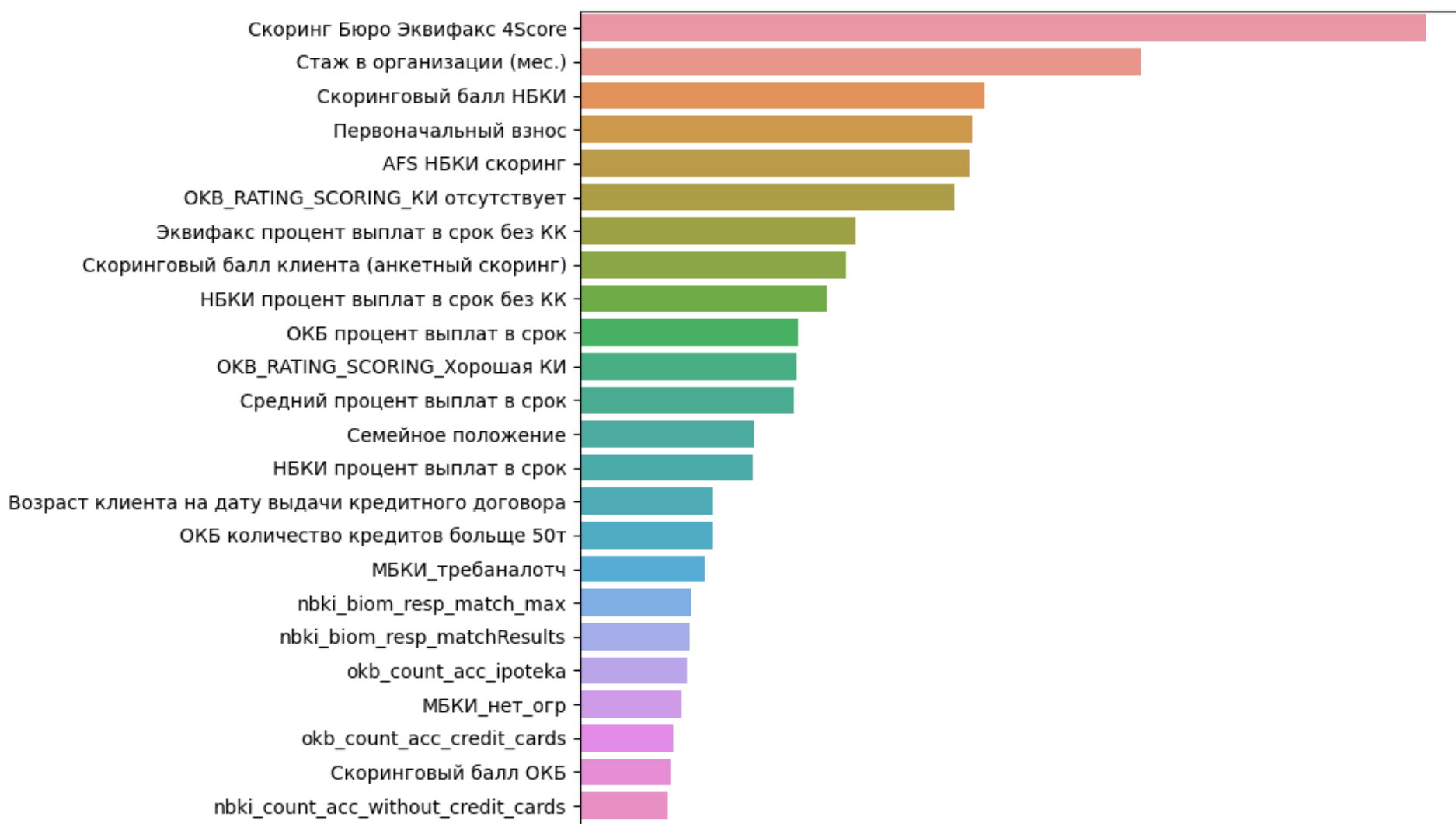
Шаги фильтрации полей — 100->50

Переходим от 100 к 50 признакам:

- 1 Проведен поиск лучшей модели
- 2 Была определена важность
- 3 Затем проведена серия экспериментов по определению оптимального количества столбцов выбирая самые важные
- 4 Удалены 7 самых коррелирующих переменных.
- 5 Проверено, что точность улучшилась
- 6 Снова проведен поиск лучшей модели
- 7 Проверено отсутствие коррелирующих и найдена важность признаков

Важность 50 признаков

Важность признаков сумма XGBoost и RandomForest



Лучшая модель на тестовой выборке

- Accuracy: 0.691234
- AUC: 0.733127
- Precision: 0.674832
- Recall: 0.206910
- Одобрённых:
0.106039
-
-