

20.11.2021



ПРЕДСКАЗАНИЕ ДЕФОЛТА КОМПАНИИ

ХАКАТОН - 2021

ReLU_people



ПОВЕСТКА

ключевые темы для обсуждения

01

Анализ
данных

02

Генерация
признаков

03

Модель

04

Оценка
финальной
модели

05

Insights

06

Возможности
дальнейшего
использования

20.11.2021

ReLU_people



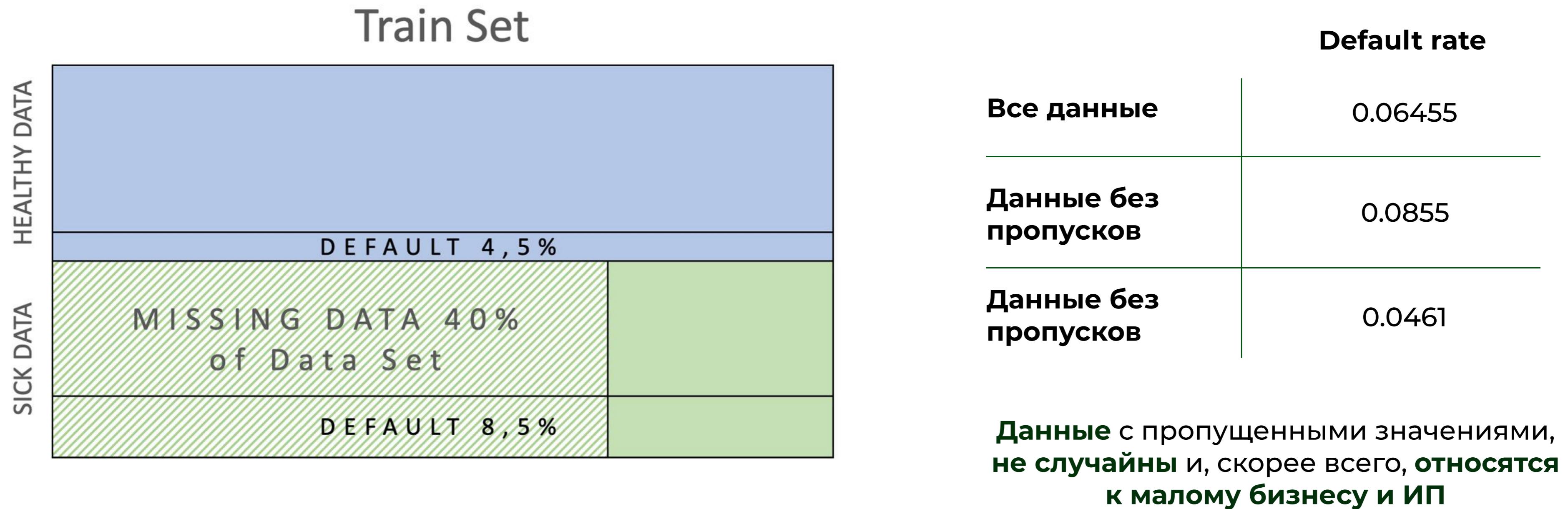
АНАЛИЗ ДАННЫХ

ХАКАТОН - 2021

3



АНАЛИЗ ПРОПУЩЕННЫХ ДАННЫХ



20.11.2021

ReLU_people



ГЕНЕРАЦИЯ ПРИЗНАКОВ

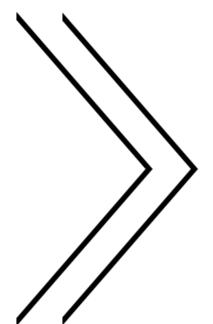


СОЗДАНИЕ НОВЫХ ПРИЗНАКОВ

150+

новых признаков:

- активы
- пассивы
- признаки-отношения показателей
- показатели рентабельности
- индикаторы пропусков
- индикаторы “возраста” компании
- etc



20.11.2021

ПРЕОБРАЗОВАНИЯ:

Финансовые метрики

- Прибыльность
- Ликвидность
- Кислотность
- Cash gap days

Log() трансформации

- $\log(\text{Revenue}+1)$



20.11.2021

ReLU_people



МОДЕЛЬ



WOE-BINNING + ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

МОДЕЛЬ:

Логистическая
регрессия

ПРОСТРАНСТВО ПРИЗНАКОВ:

Исходные и сгенерированные
признаки, подвергнутые WOE-
трансформации

ОБОСНОВАНИЕ ВЫБОРА:

- WOE-трансформация позволяет **восстановить монотонную зависимость между независимыми переменными и таргетом**
- **Более интерпретируемое заполнение пропущенных значений** при их большом количестве
- **Уменьшается чувствительность модели к выбросам**
- **Упрощается обработка выбросов**

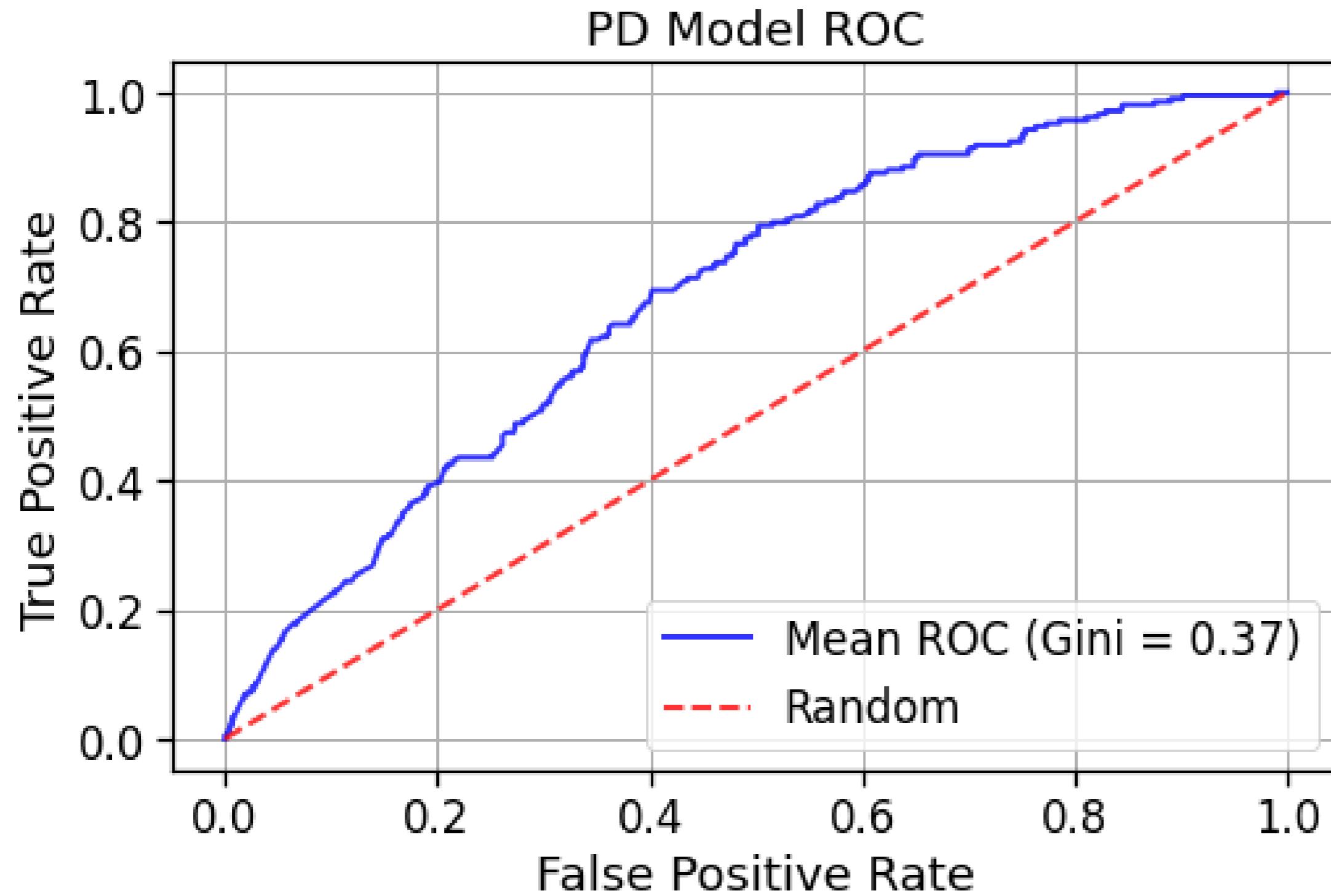
20.11.2021

ReLU_people



ОЦЕНКА ФИНАЛЬНОЙ МОДЕЛИ

ОЦЕНКА WOE МОДЕЛИ



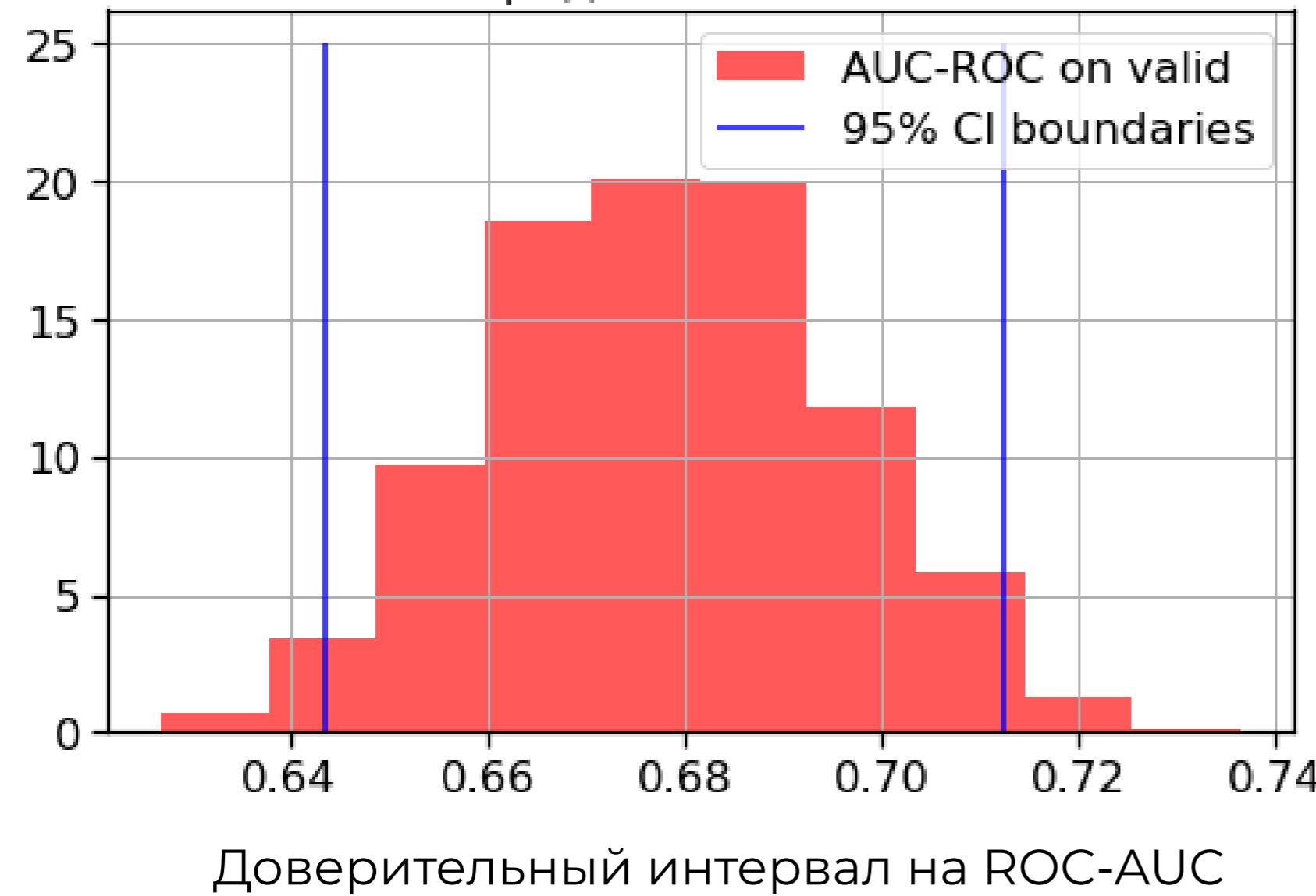
ReLU_people

ROC - AUC score = **0.6798**
Accuracy score = **0.5982**

ОТБОР ПРИЗНАКОВ



Распределение AUC-ROC



отбор по IV

фильтрация незначимых признаков по Gini (<5%)

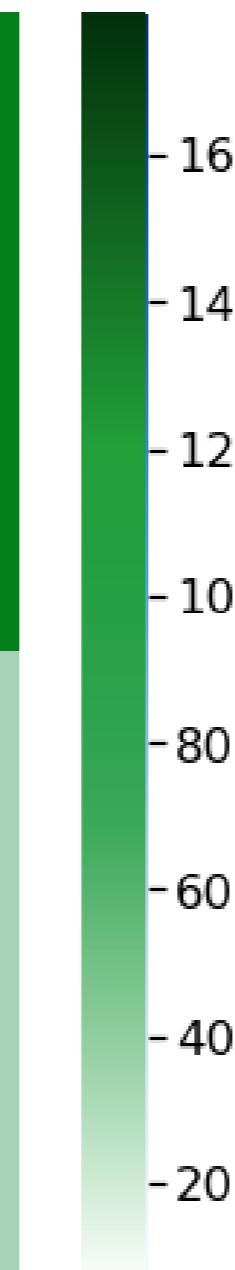
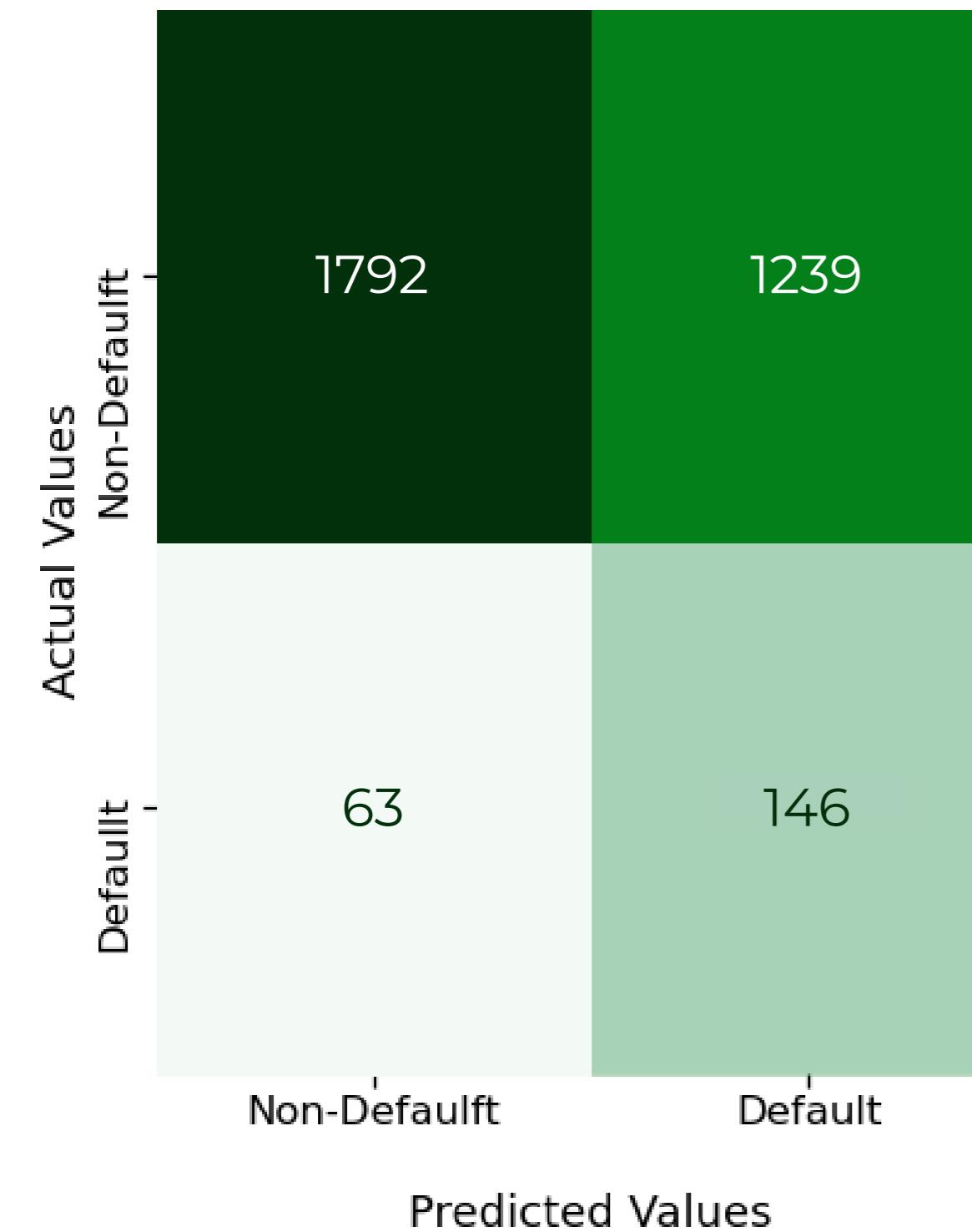
L1/L2- регуляризация

ReLU_people

ЭКОНОМИЧЕСКАЯ ИНТЕРПРЕТАЦИЯ РЕЗУЛЬТАТОВ



PD Classification Confusion Matrix



Предположим, что банк рассчитывает **МАХ сумму кредита** следующим образом:

$$\mathbf{MAX} = \mathbf{NP} + 0.5 * \mathbf{A} - \mathbf{L}$$

NP - чистая прибыль
A - текущие активы
L - заемный капитал

Тогда, при условии, что все компании^{*} взяли максимально возможную сумму кредита под **8,14%**["] на **3 года**

Недополученная прибыль - **1 679 245 ₽**

Убытки - **51 837 457 ₽**

ReLU people

20.11.2021

ReLU_people



INSIGHTS



НА ОСНОВЕ ДАННЫХ

Налоговая нагрузка **сильнее всего** влияет на риск банкротства компании

Отношение управленческих расходов к прибыли компании **практически не оказывает** влияния на риск возникновения дефолта

Большая часть пропущенных данных относится к малому бизнесу и ИП

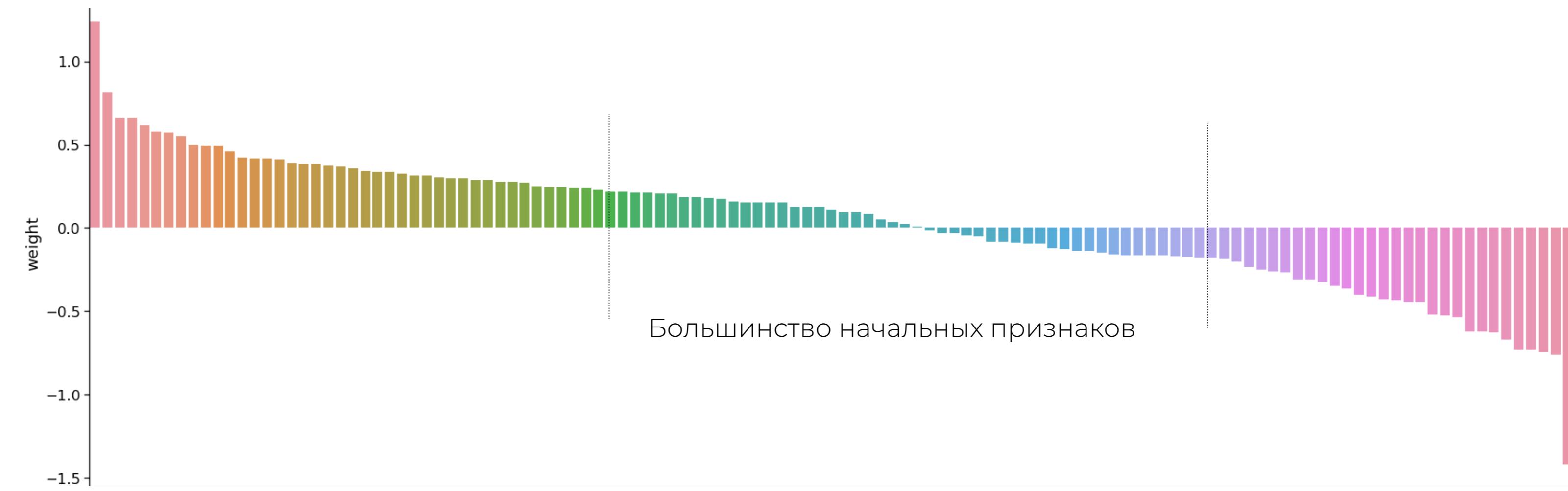
Уровень риска среди компаний данного типа - в **2 раза** выше

Для более точной оценки подобных компаний необходимо получать дополнительные данные об их финансовой деятельности.

НА ОСНОВЕ ДАННЫХ



Наблюдается сильное влияние признаков, сгенерированных как WoE методом, так и преобразованиями исходных данных



Генерация новых признаков позволила выделить заметно более сильные зависимости.



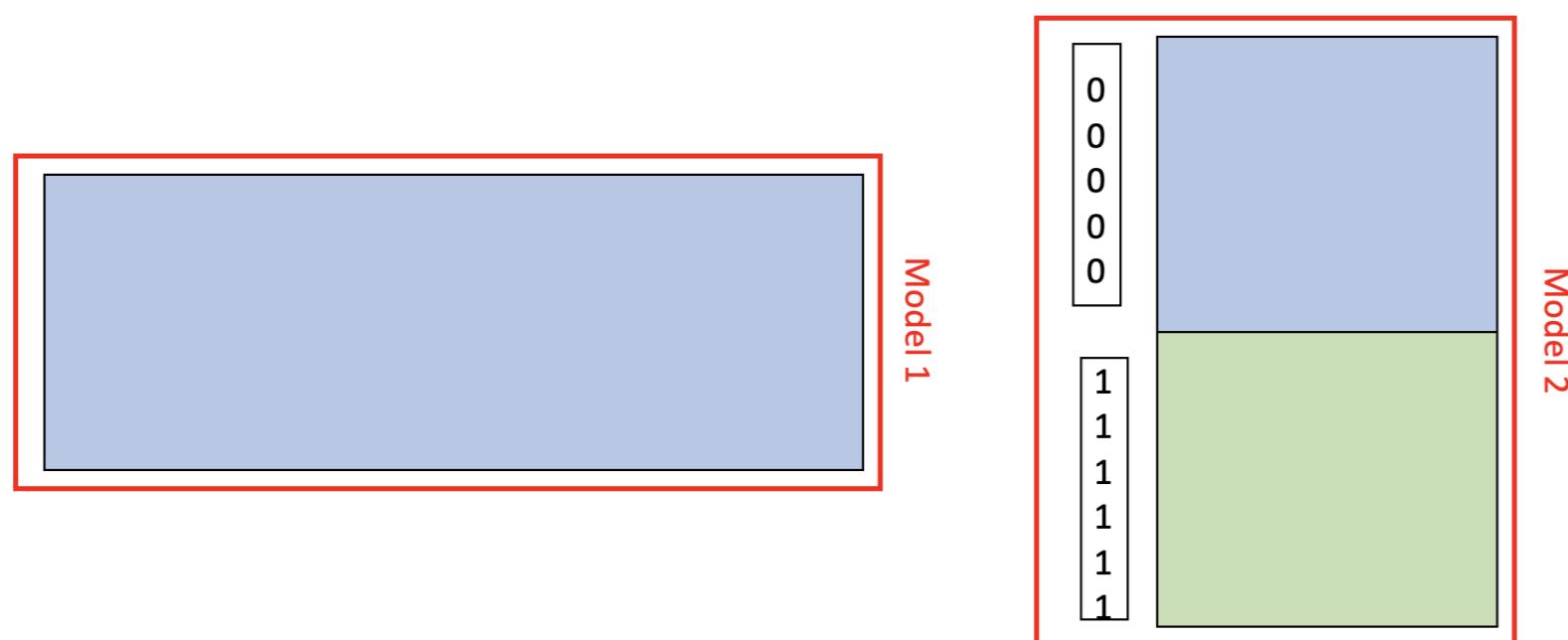
ВОЗМОЖНОСТИ ДАЛЬНЕЙШЕГО РАЗВИТИЯ МОДЕЛИ

ВОЗМОЖНЫЕ ВАРИАЦИИ МОДЕЛЕЙ



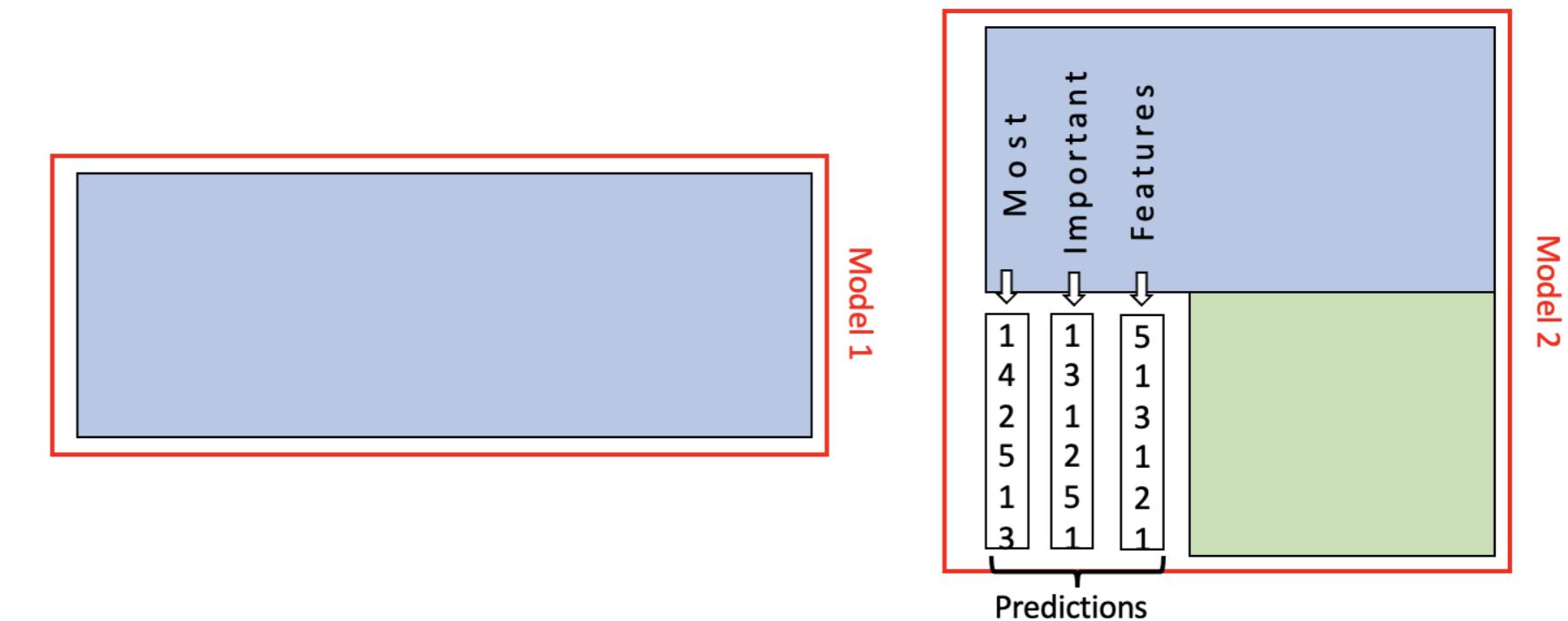
1

Мы предлагаем дальше исследовать результаты двух отдельных моделей в новых комбинациях



2

Другой вариант - предсказание самых важных признаков.



20.11.2021



eganz@edu.hse.ru

kmscheblanov@edu.hse.ru

kpkovalev@edu.hse.ru

narysin@edu.hse.ru

СПАСИБО!

Вопросы?

ХАКАТОН - 2021

ReLU_people

20.11.2021

ReLU_people

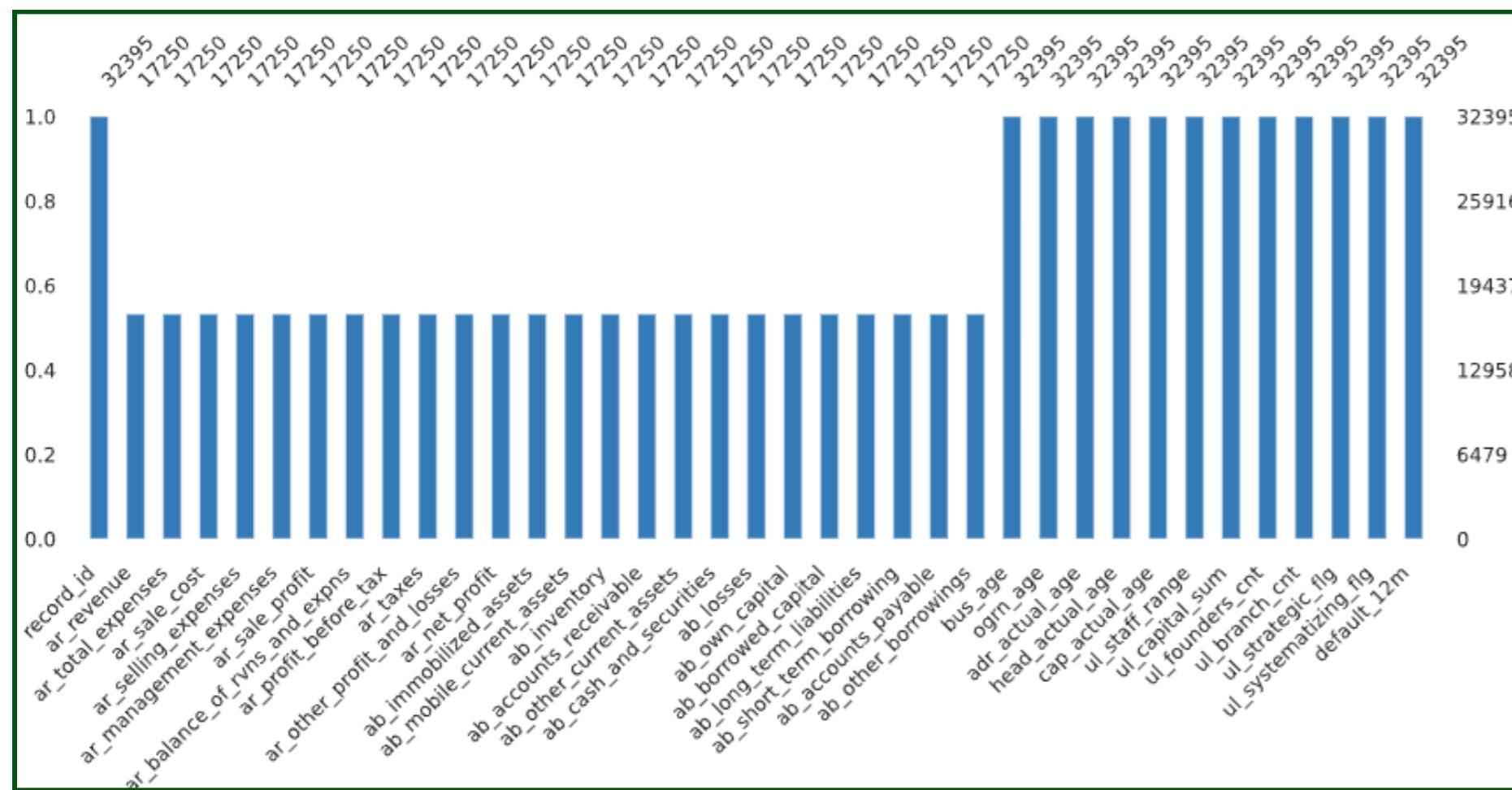


ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

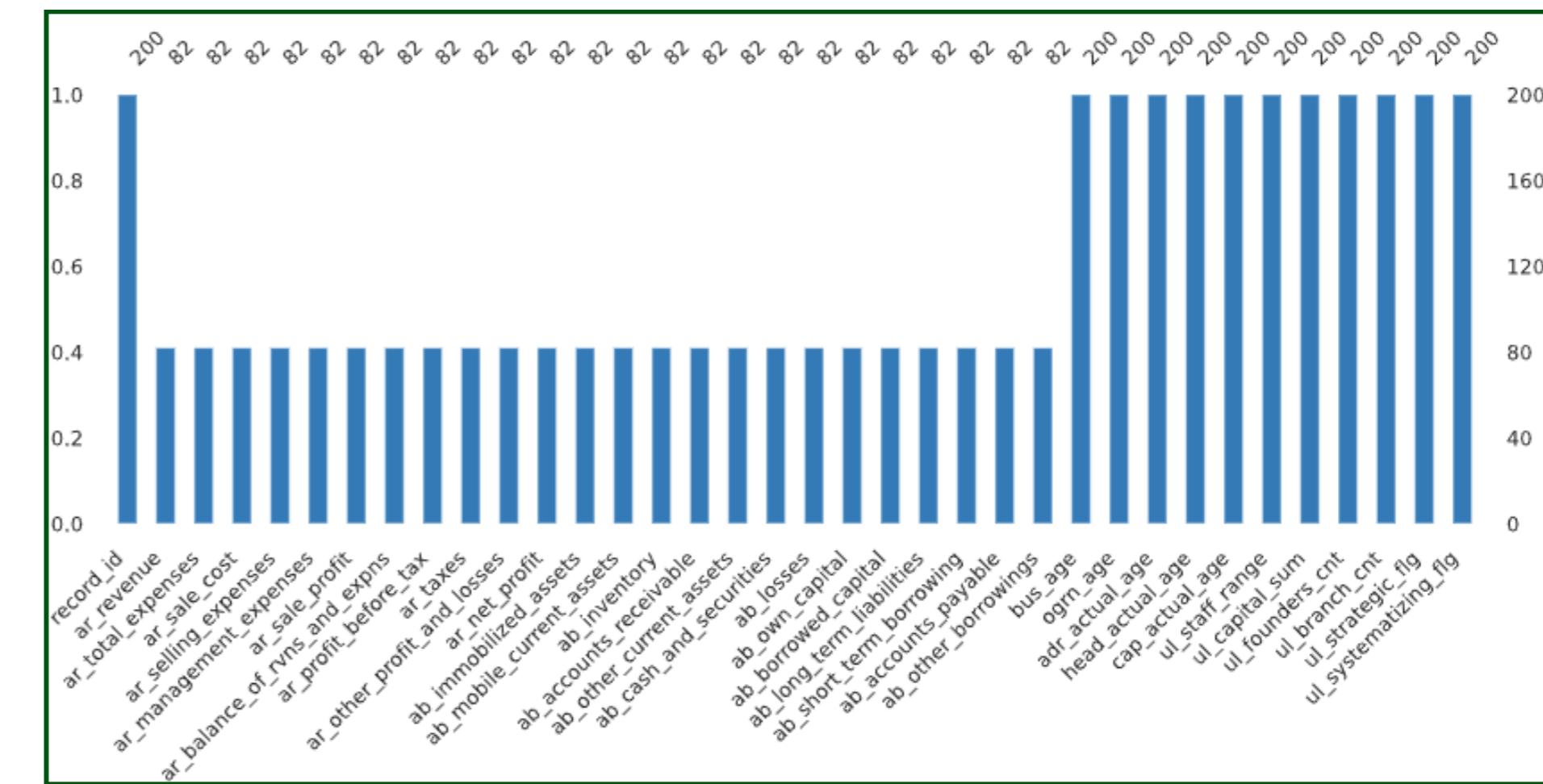
АНАЛИЗ ПРОПУЩЕННЫХ ДАННЫХ



Гистограмма пропущенных значений в
обучающей выборке



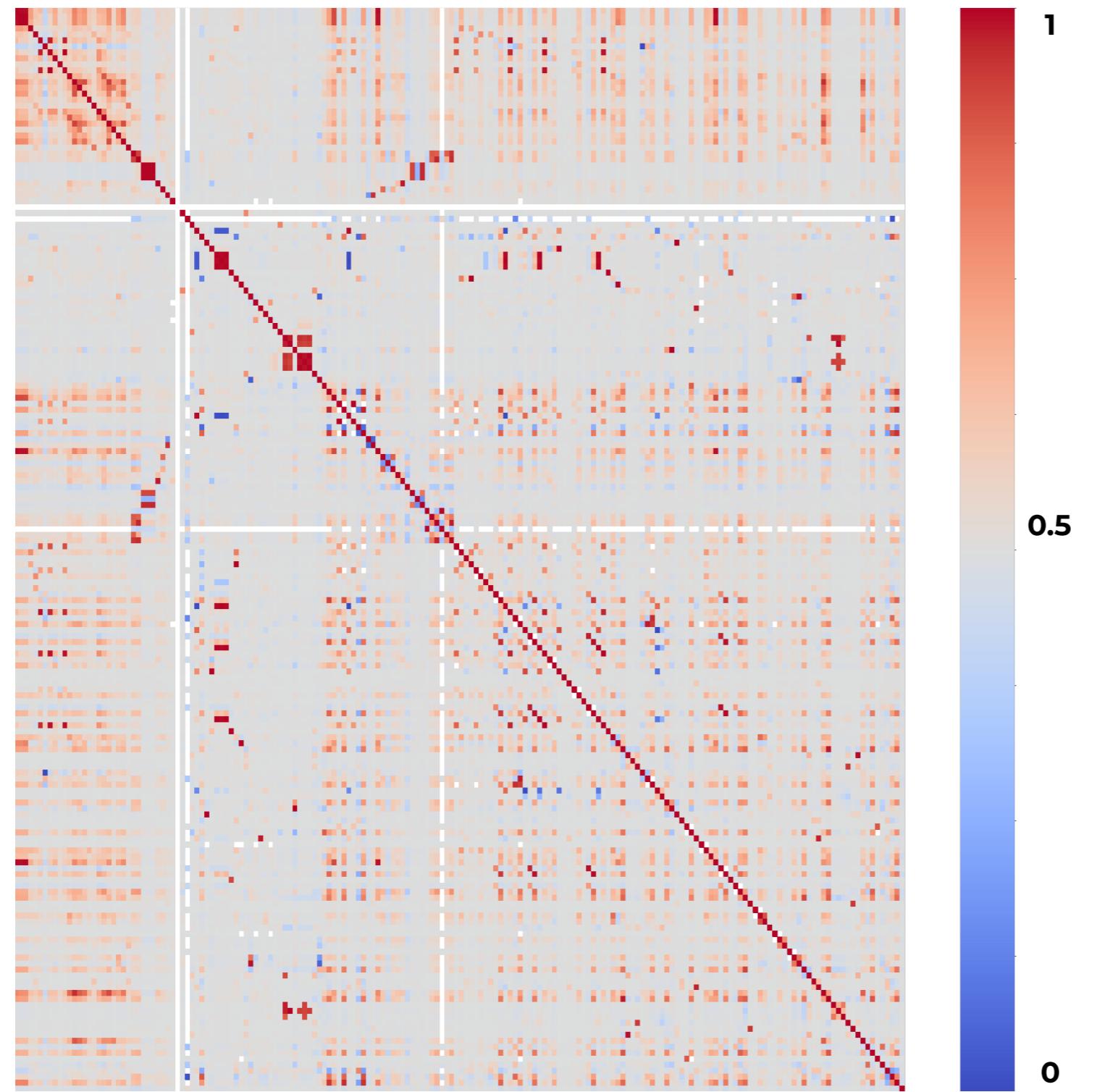
Гистограмма пропущенных значений в
тестовой выборке



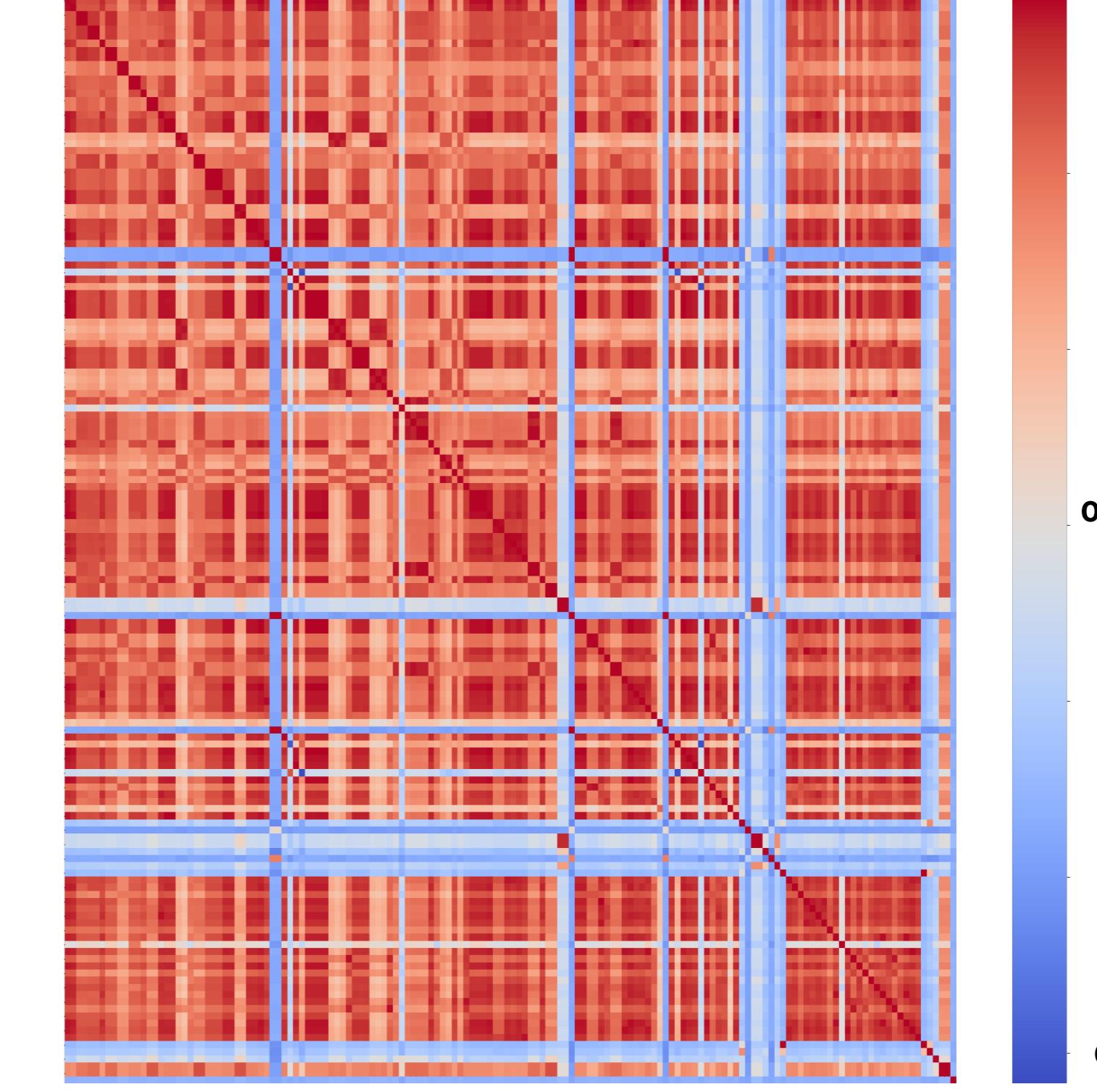
ИСХОДНЫЕ VS WOE-ТРАНСФОРМИРОВАННЫЕ ПРИЗНАКИ



20.11.2021



Матрица корреляций для исходных признаков

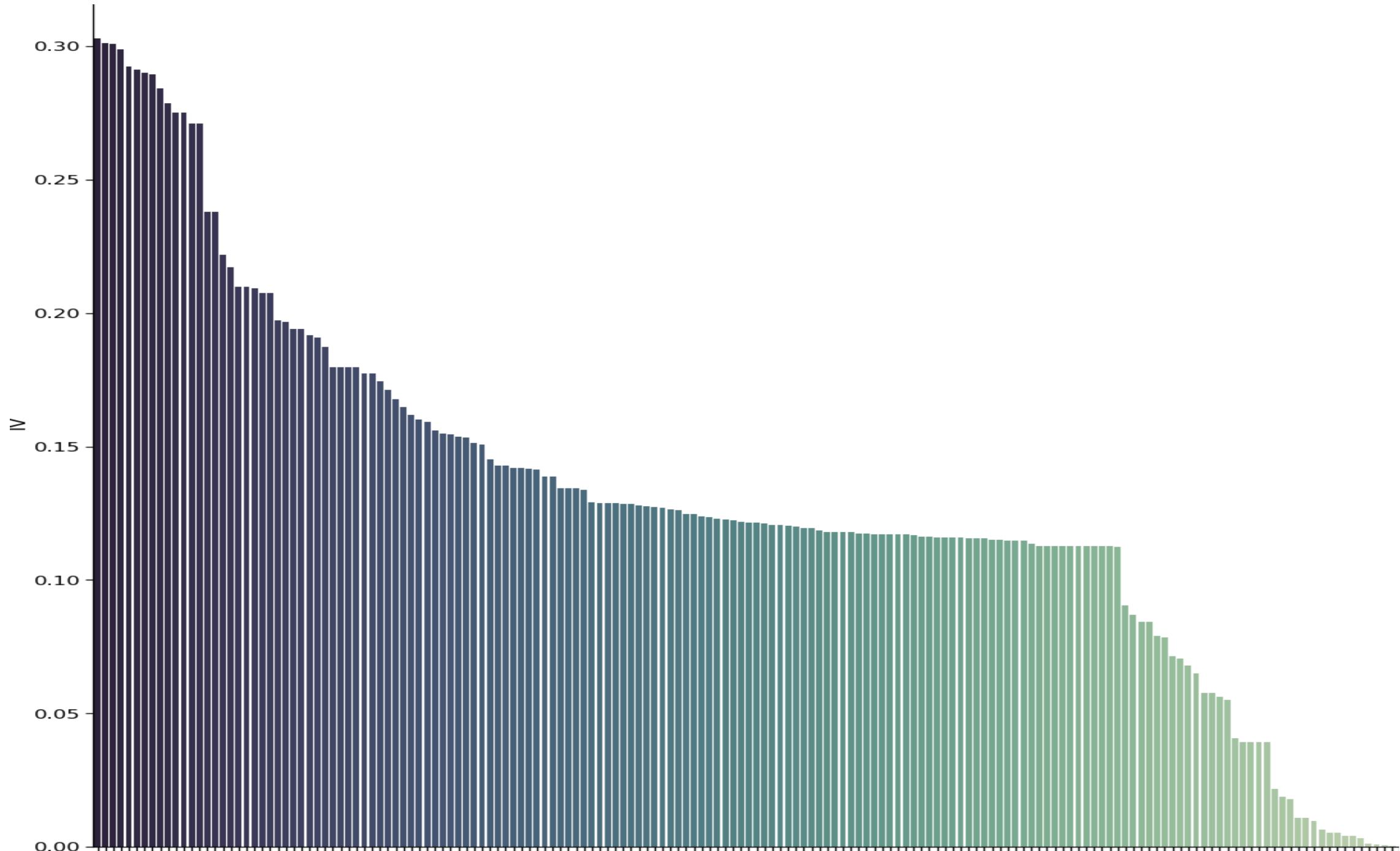


Матрица корреляций для WOE-трансформированных признаков

ReLU_people

21

ГИСТОГРАММА ВАЖНОСТИ ПРИЗНАКОВ ПО IV



Наиболее информативные признаки:

1. Налоговая нагрузка
2. Отношение прибыли к оборотным активам
3. Нераспределенная прибыль (на конец последнего года)
4. Отношение текущих денег к доле займов
5. Флаг новой компании
6. Отношение коммерческих расходов к выручке
7. Выручка
8. ROI
9. Отношение нераспределенной прибыли к чистой прибыли
10. Ликвидность
11. Длительность с даты образования ОРГН
12. Debt-to-Equity (D/E) Ratio
13. Доля оборудования в собств капитале
14. Доля долгосрочных обязательств в структуре займов
15. Средняя прибыль от филиала

ReLU_people