



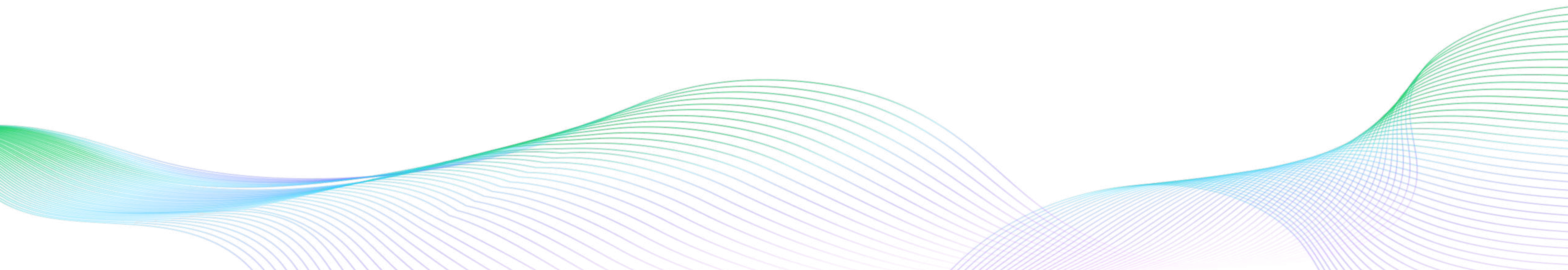
Эффективный скоринг дефолтности физлиц на основе открытых данных

Николай Друкман

SpectrumData

О компании

SpectrumData – IT-компания, предоставляющая решения в области Big Data и Data mining на основе данных о людях, организациях, автомобилях и объектах недвижимости



Наши проекты



Маркетплейс данных
для бизнеса



Проверка истории
автомобиля



Бюро кредитных
историй

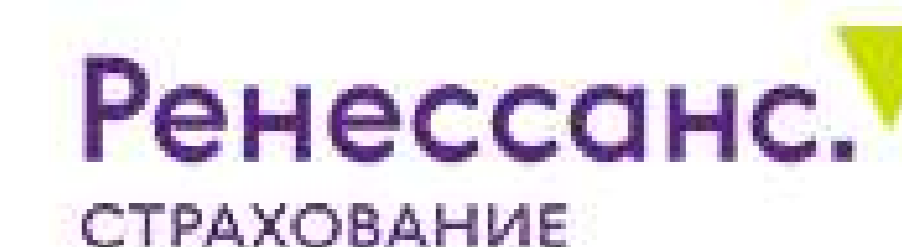


Проверка физических
лиц



Проверка
контрагентов

Нам доверяют



Преимущества SpectrumData

Главные преимущества SpectrumData:

1. Наличие большой накопленной истории:

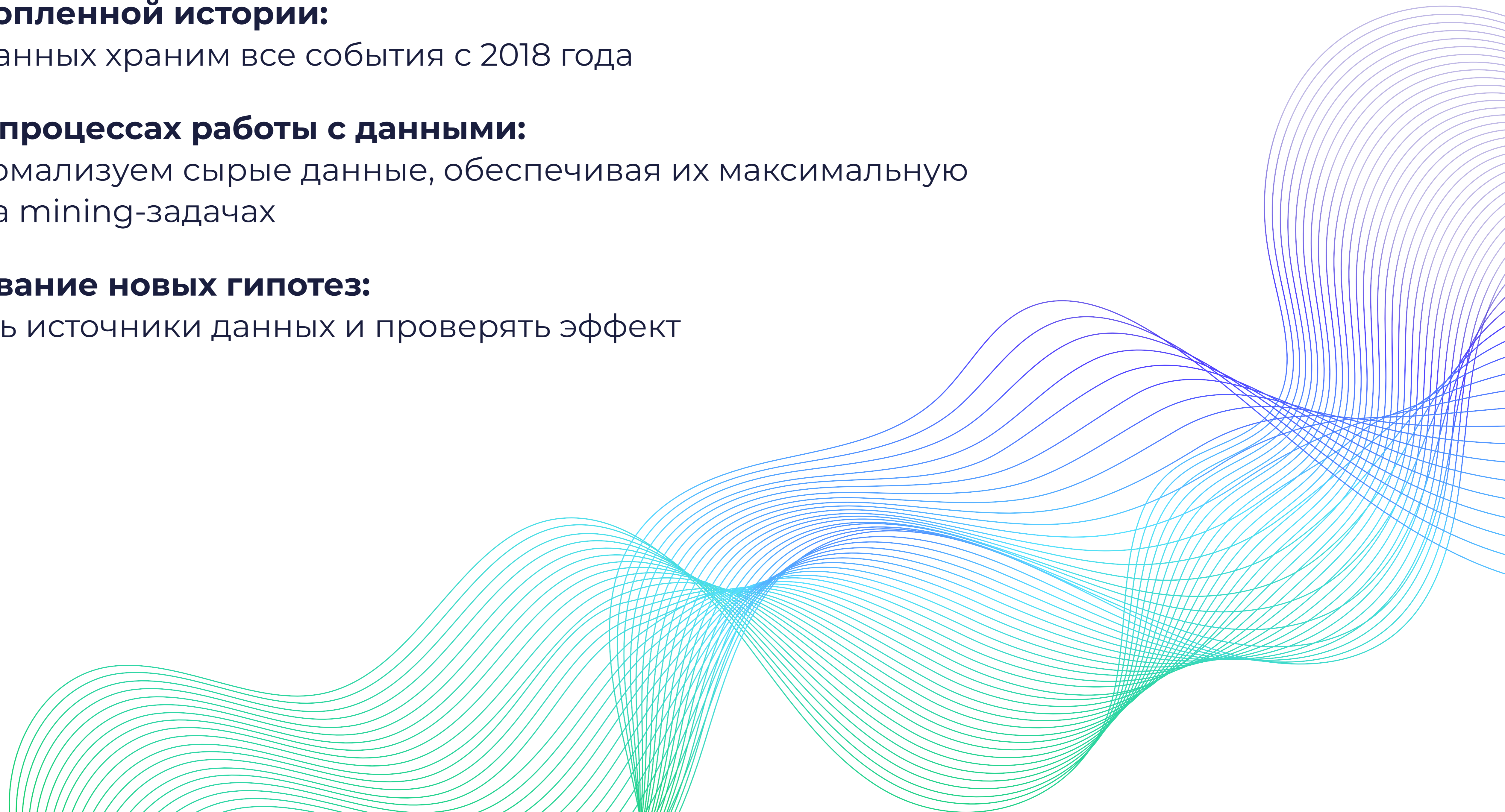
По ключевым блокам данных храним все события с 2018 года

2. Высокая экспертиза в процессах работы с данными:

Аккуратно чистим и нормализуем сырые данные, обеспечивая их максимальную результативность в Data mining-задачах

3. Непрерывное тестирование новых гипотез:

Не перестаем добавлять источники данных и проверять эффект от их использования





**хороший заемщик по мнению
нейросети Kandinsky**

Скоринг дефолтности физлиц

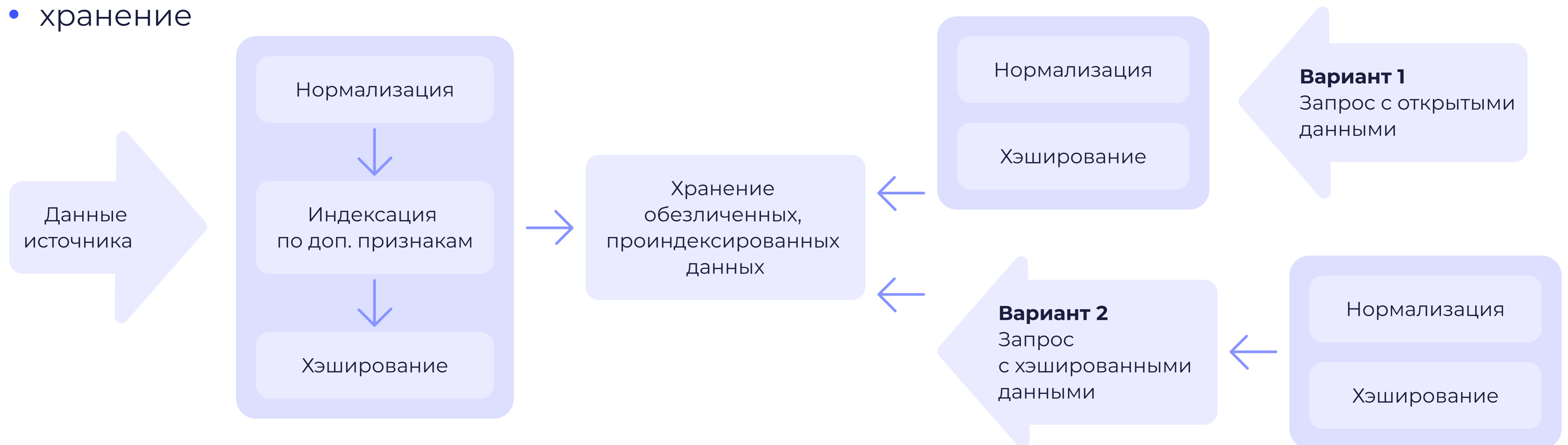


**плохой заемщик по мнению
нейросети Kandinsky**

Как готовятся данные?

Мы обрабатываем данные, в том числе собственные, в несколько этапов:

- предварительная нормализация данных
- хэширование (обезличивание)
- индексация по полу, возрасту, региону
- хранение



Как строится модель?

Модель скоринга построена на 3 группах данных, которые влияют на вероятность дефолта клиента:

1. История владения автомобилями:

- технические характеристики
- история ДТП
- штрафы ГИБДД
- история лишения прав
- актуальная рыночная стоимость ТС
- и др.

2. Правовая благонадежность:

- исполнительные производства ФССП (с историзмом)
- арбитражи
- дела судов общей юрисдикции
- и др.

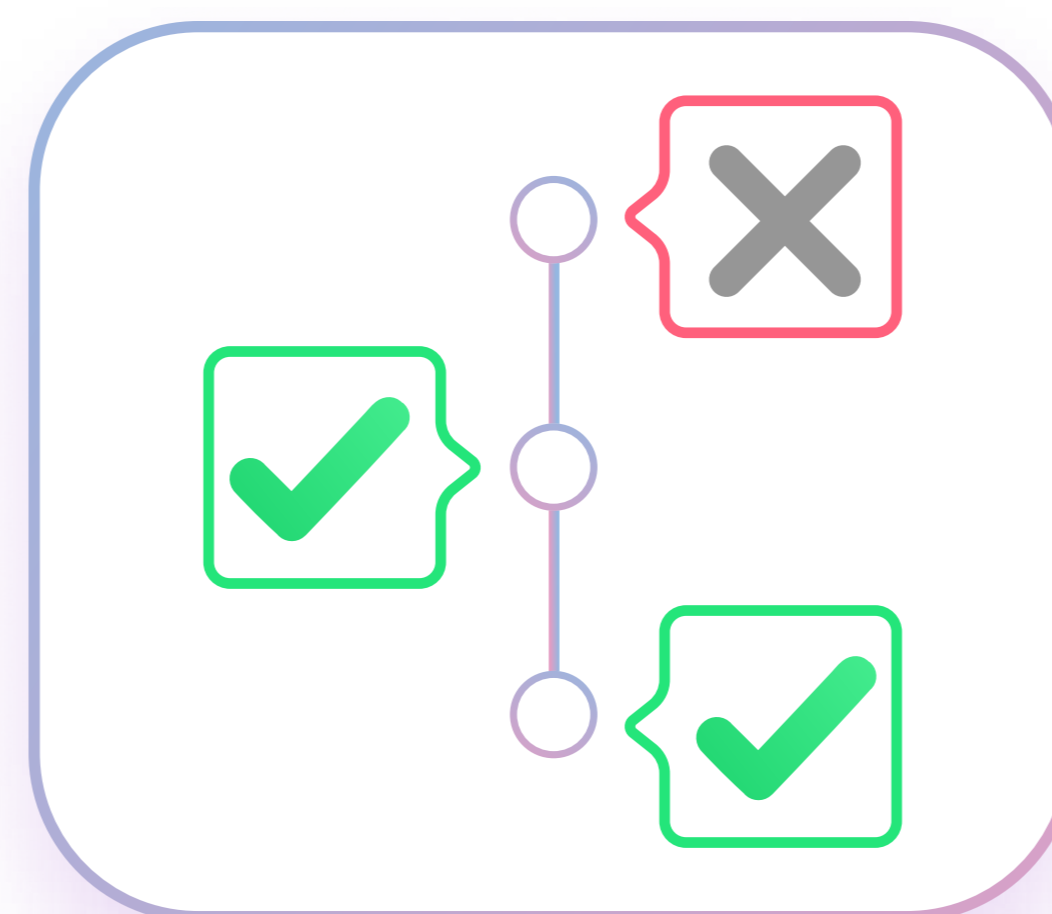
3. Бизнес-активность:

- деловые связи физлица из ЕГРИП/ЕГРЮЛ
- участие в управлении организациями
- и др.

Моделирование скоринга дефолтности



- На основе данных Spectrum бюро собрали выборку исторических кредитов за 2021 год для разработки generic-моделей
- Сегменты клиентов и таргеты выбрали так, чтобы с большей вероятностью показывать приросты для интересующих заказчиков сегментов



- Корректно отбираем ретрособытия, используя «консервативные» отступы от даты заявки
- Не позволяем каких-либо заглядываний в «будущее»



- Используем основанные на градиентом бустинге алгоритмы

Моделирование скоринга дефолтности



Аккуратно настраиваем модели и выбираем значимые переменные под таргеты



- Обеспечиваем и контролируем стабильность переменных и предсказаний моделей во времени
- Гарантируем SLA в части поведения скоринга во времени (в частности, PSI)

Результаты пилотов

>1% Gini

**прирост к скоринговым
системам заказчиков**

- Результат получен на основе текущих пилотов с 3-мя банками из топ-10 РФ
- Generic-модели, построенные на данных Spectrum бюро, дают эффект, однако более значимые приросты получаются от уточненных под таргет заказчика моделей

Промышленное использование



скорость ответа
<1 секунды *



Держим нагрузку
100 rps



Готовы одновременно
предоставлять несколько
моделей под различные
типы продуктов/сегменты

* 99.98% запросов, согласно SLA

Пайплайн тестирования моделей



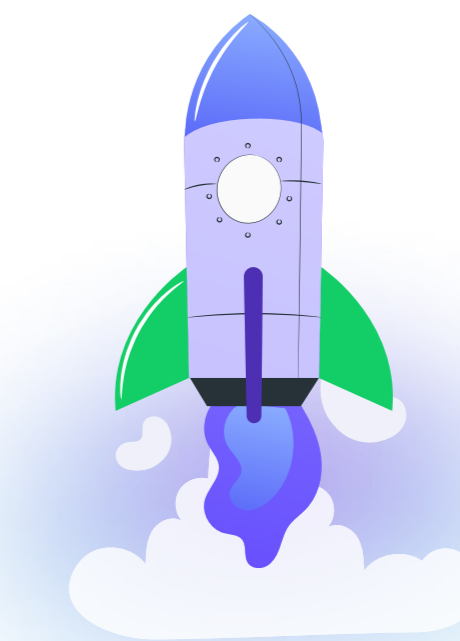
Поиск сигнала

- Финансовая организация передает репрезентативную выборку с полями для мэтчинга и частично заполненными целевыми переменными
- SpectrumData дообучает модели под таргет заказчика и скорит отложенную выборку
- Заказчик оценивает метрику в рамках слепого тестирования и выбирает интересующие для дальнейшего использования модели



Большой тест

- Заказчик предоставляет необходимую для скоринга выборку целиком
- SpectrumData скорит выборку отобранными на предыдущем шаге моделями
- Компании совместно оценивают экономический эффект от использования в процессе моделей SpectrumData



Внедрение

- Согласование коммерческих условий
- IT-интеграция



Данные для скоринга заемщика



Изображения сгенерированы нейросетью Midjourney

Данные для скоринга заемщика

На основе анализа 40 источников мы выделили 8 наборов данных, которые больше всего влияют на вероятность дефолта клиента:

- история владения автомобилями
- актуальная рыночная стоимость ТС
- история ДТП
- штрафы ГИБДД
- исполнительные производства ФССП (с историзмом)
- история лишения водительских прав
- деловые связи физлица из ЕГРИП/ЕГРЮЛ
- участие в управлении организациями.



Преимущества данных об автомобилях

Основные данные об автомобилях, доступные в SpectrumData:

- технические характеристики
- история владения автомобилями
- актуальная рыночная стоимость автомобиля
- история использования автомобиля в такси
- история ДТП
- штрафы ГИБДД.

Преимущества данных об автомобилях, предоставляемых SpectrumData:

1. Высокая корреляция с риском дефолта

Особенно эффективны при скоринге по автокредитам и ипотеке

2. Уникальность

Никакие другие участники рынка не использовали ранее такие данные для скоринга

3. Экспертиза в данных

Работаем с данными с 2010 года.

Работаем со всеми страховыми компаниями по направлению ОСАГО



Преимущества и особенности источника ФССП

Эффект от использования превосходит самостоятельный парсинг:

- нормализация позволяет избежать потери данных
- распознавание смены фамилий с высокой точностью
- не нужно самостоятельно анализировать статьи и суммы ИП

Историзм повышает эффективность моделей по оценке дефолтности на 17%:

- поиск как открытых, так и ранее закрытых ИП
- в 7 раз больше данных для аналитики
- возможность отслеживания смены фамилий

В составе скоринга дефолтности:

- учитываются пол и возраст клиента, а также региональная специфика
- принимается во внимание тяжесть статей и скорость закрытия ИП





Верификация предмета залога



Изображения сгенерированы нейросетью Midjourney

Верификация предмета залога

Предоставление полных данных о транспортном средстве, принимаемом в залог, на этапе заведения заявки на кредит

Возможности применения:

- подтверждение принадлежности ТС заемщику
- оценка рисков, связанных с ликвидностью залогового авто

Доступные данные о транспортном средстве:

- оценка рыночной стоимости с учетом истории эксплуатации: ДТП, количества владельцев, использования в такси
- ограничения: нахождение в залоге, розыске, арбитражные дела и долги собственника
- документы и технические характеристики
- регистрационные действия
- и любые другие данные о ТС

