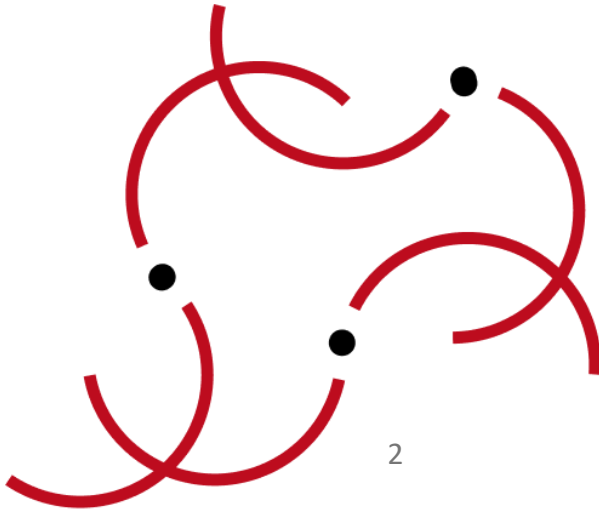
ИИ для кредитного скоринга

Digital Skills Academy

Введение: Риск как центральный вопрос

- Идеальные условия для ИИ: большой объём исторических данных, чёткая целевая переменная, сотни параметров, высокая цена ошибки
- Серьёзные опасности: модель влияет на реальные деньги, доступ к финансированию, справедливость
- Финансовые убытки, регуляторные штрафы, этические вопросы



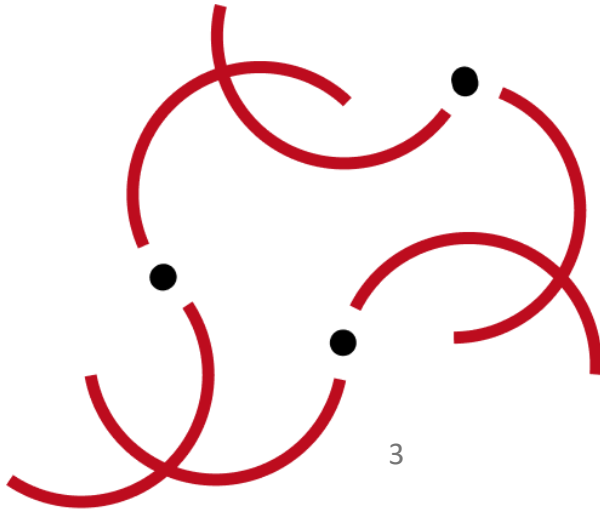
Digital Skills Academy

Основная задача скоринга: классификация и вероятность

Задача: выдать число (вероятность дефолта) на основе данных о клиенте

- Раньше: ручное построение экспертами (медленно, с ограничениями)
- Теперь: нейросети и алгоритмы машинного обучения берут все доступные данные и ищут закономерности

Казалось бы, идеально, но есть проблемы



Первый миф: «Больше данных = точнее модель»

Ошибка выжившего и смещения в данных

- Историческая выборка — люди, которым уже дали кредиты, не вся совокупность
- Модель хороша на знакомых сегментах, ошибается на новых

Отсутствие данных о некоторых событиях

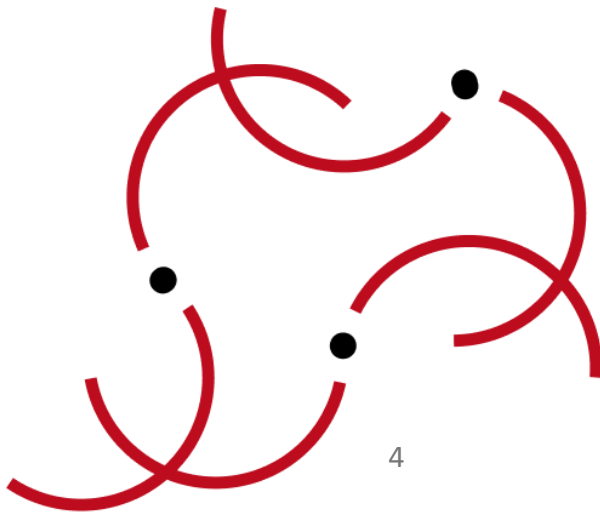
- Обучена на спокойной экономике, не справляется во время кризиса

Скрытые переменные

- Модель ловит корреляцию, а не причину
- При применении на новый регион ошибается

Практический пример: COVID и недостаточность данных 2016-2019

Вывод: нужны правильные данные, покрывающие разные экономические состояния



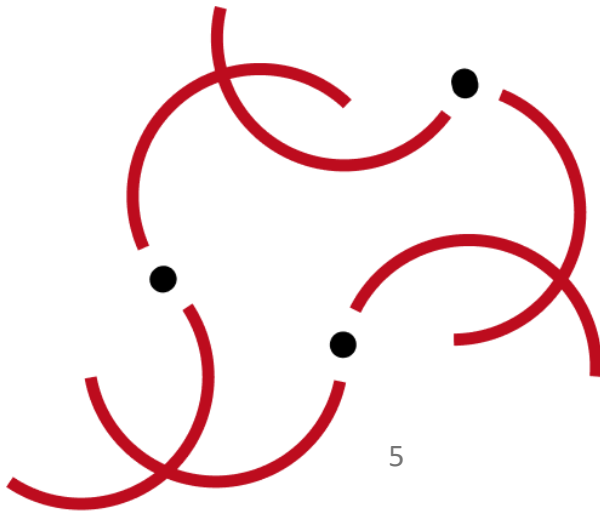
Digital Skills Academy

Второй миф: «Модель ловит истинную причину дефолта»

- Модель находит корреляции, не понимает причинность

Пример: пол клиента может коррелировать с дефолтом, но это не причина

- Косвенная дискриминация через переменные, которые коррелируют с защищёнными признаками



Digital Skills Academy

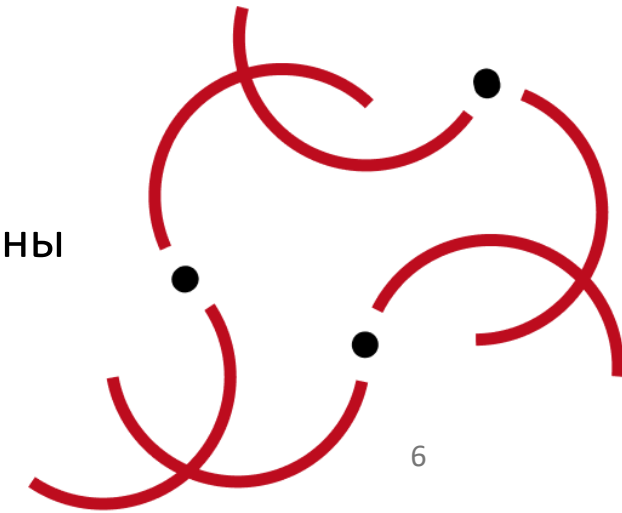
Регуляторные требования: fairness, explainability, недискриминация

Главный вопрос: почему алгоритм отказал клиенту в кредите?

Парадокс: сложные модели точнее, но менее объяснимы

Три ключевых требования:

1. **Fairness (справедливость)**
   Нет дискриминации по защищённым признакам (явной и скрытой)
2. **Explainability (объяснимость)**
   Возможность объяснить, почему алгоритм принял конкретное решение
3. **Auditability (проверяемость)**
   Модель протестирована, валидирована, результаты документированы



Digital Skills Academy

ИИ в скоринге: где он действительно работает

Первая зона: обнаружение скрытых паттернов в данных

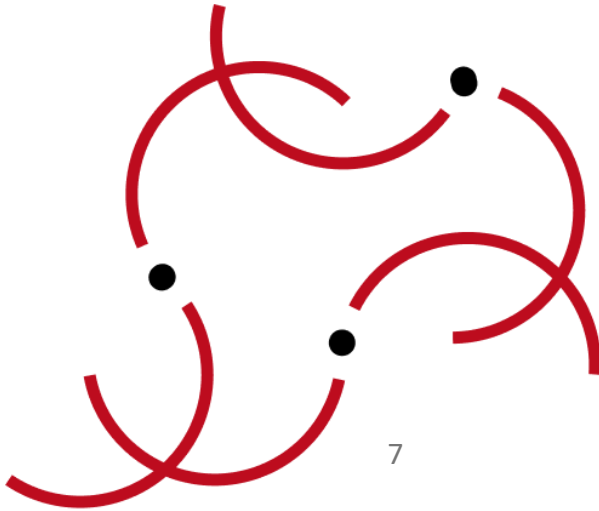
- Нелинейности и взаимодействия между переменными
- Повышение точности на 5-15 процентных пункта

Вторая зона: детекция смещений и аномалий в данных

- Автоматическое отслеживание систематических ошибок по сегментам
- Выявление красных флагов для справедливости

Третья зона: валидация и регулярная переподготовка

- Автоматический процесс переподготовки
- Регулярная проверка на изменения в рынке и поведении



Digital Skills Academy

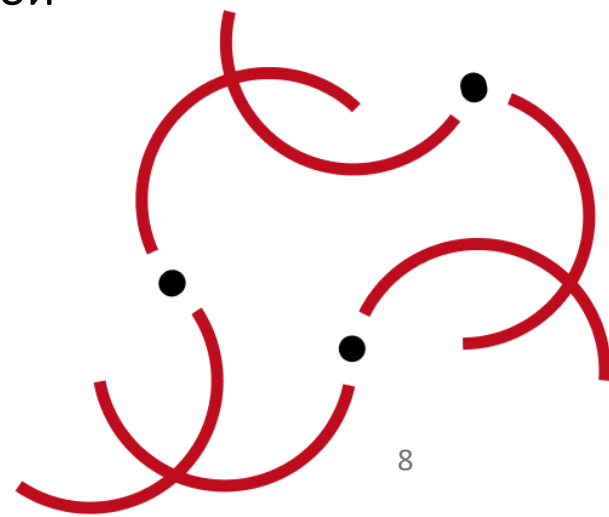
Пример: ИИ находит скрытый риск

Сценарий: 5000 заемщиков, 4% дефолт, стандартная модель 80% точности

Модель машинного обучения с дополнительными переменными: 87% точности

Обнаружена закономерность: молодые люди с большим кредитом в первый раз — высокий риск

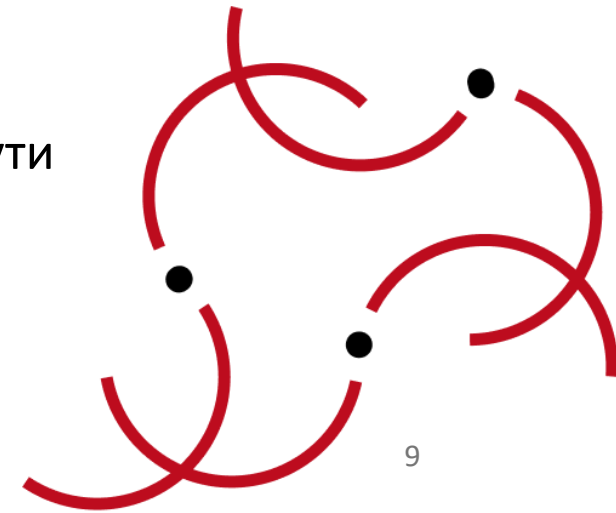
- Хороший доход и стаж маскируют отсутствие опыта управления большой задолженностью
- Взаимодействие: большой кредит + нет истории больших кредитов = высокий риск



Digital Skills Academy

Ключевой риск: модель ловит не причину, а артефакт

- Точность улучшилась, но возникла проблема: непропорционально большой процент отказов женщинам
- Причина: молодые женщины в исторических данных брали меньшие кредиты
- Модель через взаимодействие переменных косвенно ловит дискриминацию
- Регуляторная проблема: не использует пол явно, но дискриминирует по сути



Digital Skills Academy

ИИ как помощник в валидации моделей: three-line defence

First line: разработчик сам проверяет качество

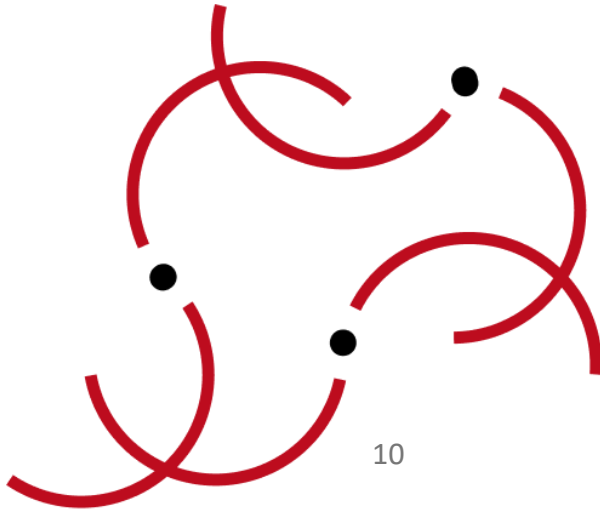
- Точность, распределение оценок, систематические ошибки

Second line: независимая команда риск-менеджмента

- Fairness, explainability, стабильность
- Критический взгляд

Third line: внутренний аудит

- Документация, compliance, регуляторные требования



Digital Skills Academy

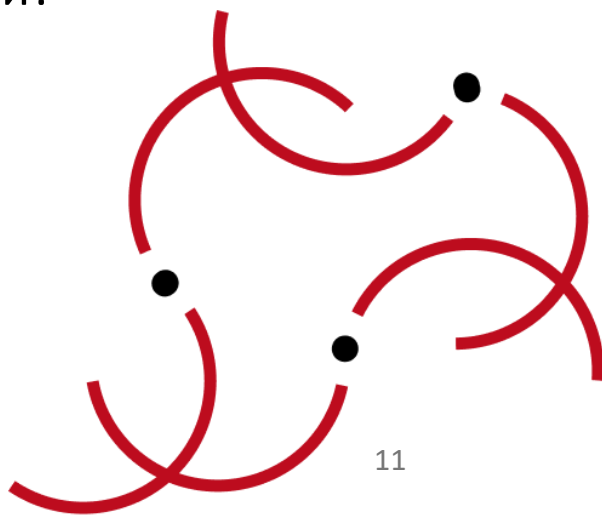
ИИ в первой линии: автоматическая диагностика модели

Описываете модель ИИ: переменные, параметры, как она работает

ИИ проводит диагностику: ищет проблемы в точности по сегментам

Вопросы, которые ИИ может выявить:

- Почему точность на одних сегментах выше, чем на других?
- Где систематические переоценки или недооценки риска?
- На каких сегментах модель может сломаться при изменении экономики?
- Какие переменные косвенно ловят дискриминацию?



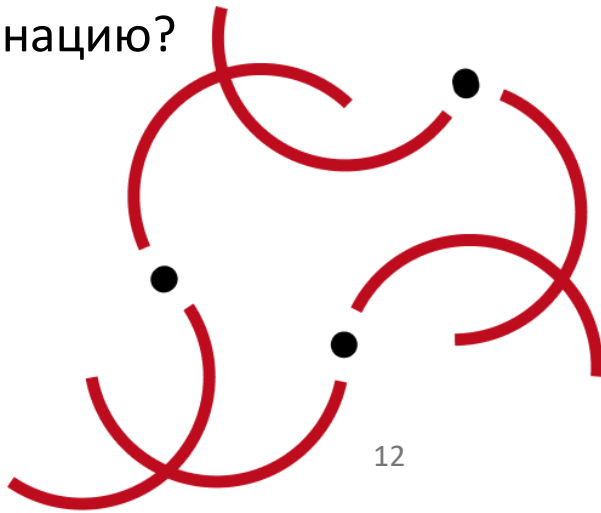
Digital Skills Academy

Промпт: диагностика скоринг-модели с ИИ

Ты — опытный аналитик рисков. Провести критическую диагностику модели

Твоя задача:

- **Выяви паттерны:** на какие сегменты модель выглядит менее надёжной?
- **Проверь справедливость:** есть ли скрытая дискриминация?
- **Выяви риски смещений:** на каких сегментах модель может подвести в будущем?
- **Определи слепые пятна:** какие переменные косвенно ловят дискриминацию?
- **Дай рекомендации:** что сделать перед запуском в работу?
- **Структура:** Проблема — Почему это проблема — Рекомендация

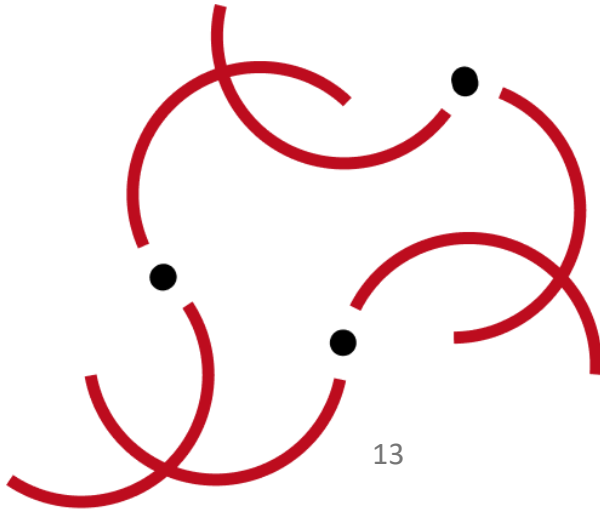


Вторая и третья линии защиты: ИИ для независимой проверки

Вторая линия: ИИ автоматизирует проверку моделей

- Система критериев для проверки fairness, explainability

Третья линия: ИИ готовит отчёты о соответствии регуляторным требованиям



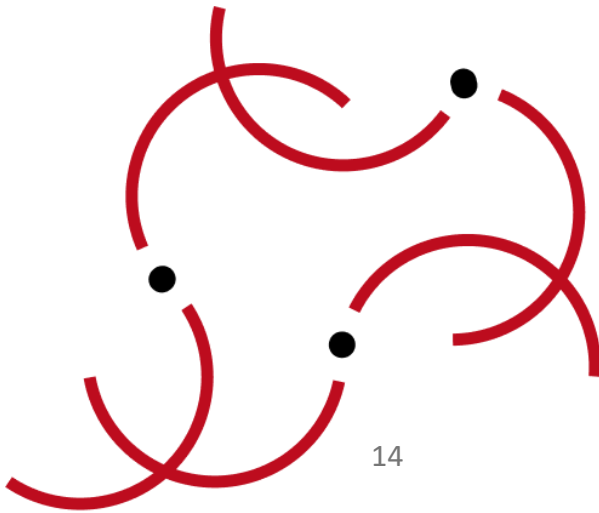
Digital Skills Academy

Практикум:
Анализ скоринг-портфеля и поиск проблем

Сценарий: портфель 50 000 активных клиентов, дефолт вырос с 2,8% до 4,2%

Контекст:

- Модель запущена январь 2022 на данных 2021
- Произошли экономические потрясения (СВО, санкции, инфляция, рост ставок)
- Потери банка: ≈14 млн рублей вместо ожидаемых 8,4 млн
- Дефолт в новых картах 2023: 5,5%



Digital Skills Academy

Практикум: Этап 1 — Диагностика

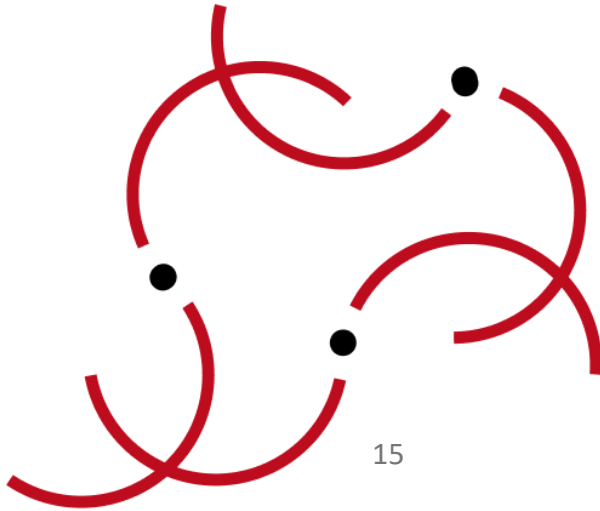
Исторический контекст:

- Январь-март 2022: СВО, санкции, валютный шок
- Апрель-декабрь 2022: инфляция, ставки вросли 8%-20%
- 2023: ставки остаются высокими

Поведение клиентов изменилось: доходы в стрессе, миграция между регионами

Модель недооценивает риск по всем сегментам, особенно:

- Молодые клиенты
- Регионы
- Розничная торговля
- Возможные причины: переобучение на экономике 2021, структурные изменения в базе, изменение поведения

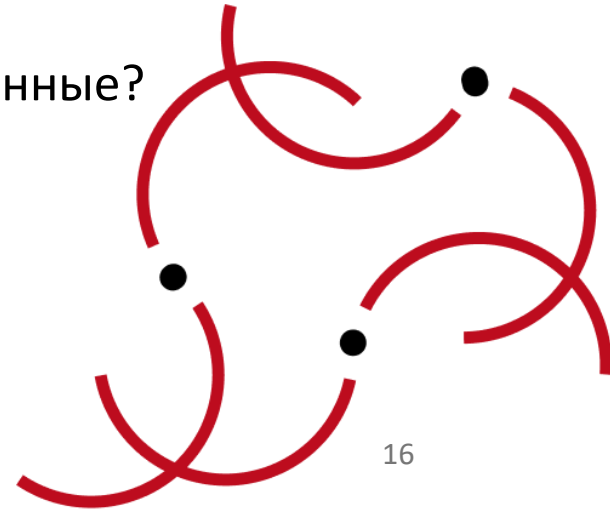


Digital Skills Academy

Практикум: Этап 2 — Анализ с ИИ

Твоя задача как аналитика риска:

1. **Диагностика:** на какие сегменты модель смотрит хуже всего?
2. **Скрытый драйвер:** какая переменная коррелирует с ростом дефолтов?
3. **Механизм ошибки:** почему модель не поймала это? Что произошло в экономике?
4. **Что делать:** пересчитать, добавить консервативность, новые переменные?
5. **Как избежать в будущем:** какой процесс валидации установить?



Практикум: Этап 3 — Валидация на новых данных

Неправильный подход: обучить на 2023, тестировать на 2023 — данные уже содержат ответ

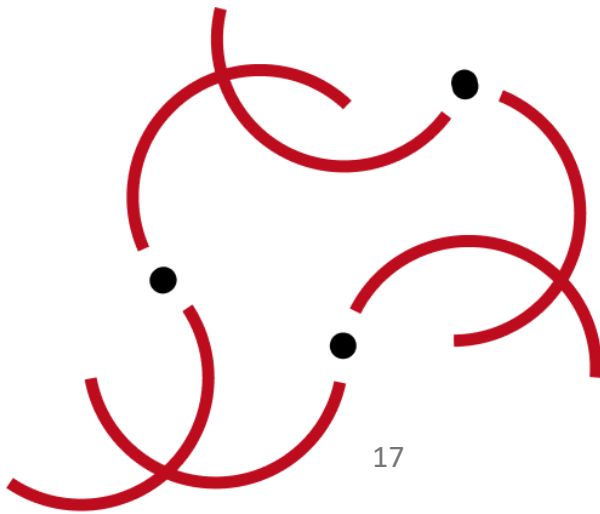
Правильный подход: walk-forward validation

Процесс:

- Window 1: обучаю 2021, тестирую H1 2022
- Window 2: обучаю 2021+H1 2022, тестирую H2 2022
- Window 3: обучаю 2021+2022, тестирую H1 2023
- Window 4: обучаю 2021+2022+H1 2023, тестирую H2 2023

Видим, как модель теряет в точности при изменении экономики

Определяем: когда произошли структурные изменения, на каких сегментах падение больше всего



Digital Skills Academy

Практикум:
Этап 4 — Новая модель и её ограничения

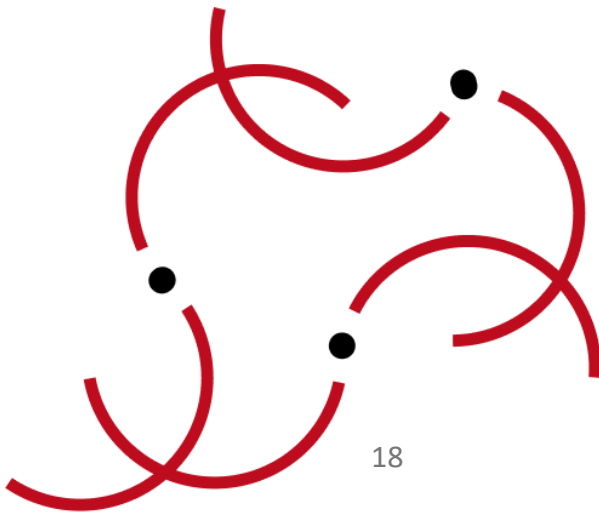
Новая модель (с новыми переменными и взаимодействиями): 83% точность

Правильный подход: признать неопределённость

- Данные охватывают спокойную экономику и кризис, но будущее неизвестно
- Установить консервативные требования

Мониторинг:

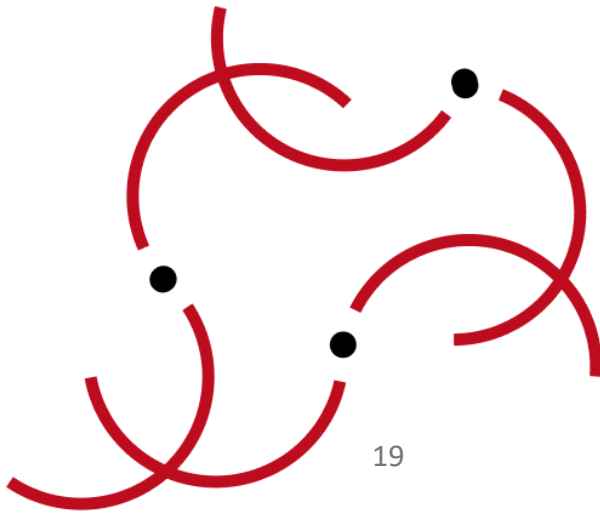
- Каждые три месяца проверять дефолты в портфеле
- Если дефолт вырос >30%, переработать модель
- Отслеживать работу модели по сегментам



Digital Skills Academy

Финальная идея:
Модель — это не истина, это живой инструмент

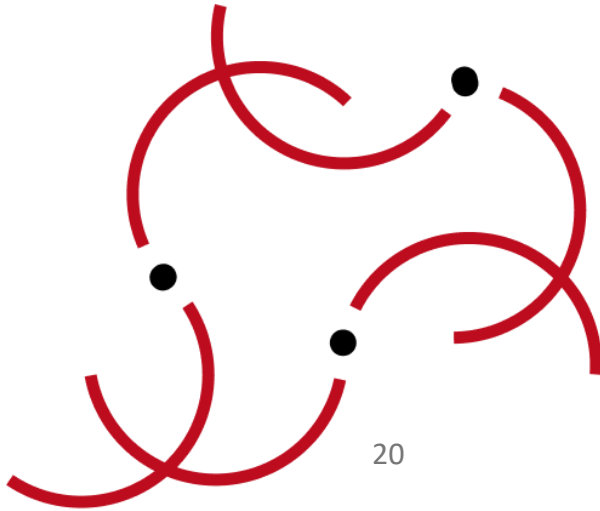
- Скоринг-модель работает в определённых условиях
- Как только условия меняются, модель может перестать работать
- ИИ помогает не верить модели слепо
- Постоянно задавать вопросы: работает ли сейчас, где ошибается, что изменилось?
- Здоровый скептицизм к модели



Digital Skills Academy

Выводы

- Кредитный скоринг — одна из зрелых областей применения ИИ в финансах
- Но здесь возникают серьёзные опасности: скрытая дискриминация, переобучение, прокси-переменные
- ИИ контролирует риски как инструмент валидации и диагностики
- Регуляторные требования жесткие: fairness, explainability, auditability — не рекомендации, а требования
- Скоринг-модель — живой инструмент, нужна постоянная переподготовка и мониторинг



Digital Skills Academy

Домашнее задание

Вариант А (с доступом к данным):

- Разработайте скоринг-модель на реальном портфеле
- Проведите диагностику по сегментам
- Проверьте на fairness и скрытую дискриминацию
- Спроектируйте валидацию и мониторинг

Вариант В (теоретический):

- Напишите 2-3 страницы: как разработывать, валидировать, мониторить скоринг-модели?
- Какие требования, какие проверки?

Для обоих: Используйте ИИ критически, опишите, какие выводы были полезны, что отклонили и почему

