

Применение машинного обучения в банковском риск-менеджменте



Екатерина
Золотарева
2021

Департамент анализа данных и машинного обучения

Департамент анализа данных и машинного обучения



Мы занимаемся
разработкой
интеллектуальных
сервисов, основанных
на технологиях
машинного обучения,
обработки данных и
интернета вещей, в
интересах реальных
заказчиков

Реализуем проекты ИИ
для реальных заказчиков

Создаем интеллектуальные
системы и сервисы

Встраиваем реальные
проекты в образование

Наши проекты и интересы

Предиктивная
аналитика

Анализ текстов и
мультимедиа данных

Умные
производства

Интеллектуальные
приложения и боты

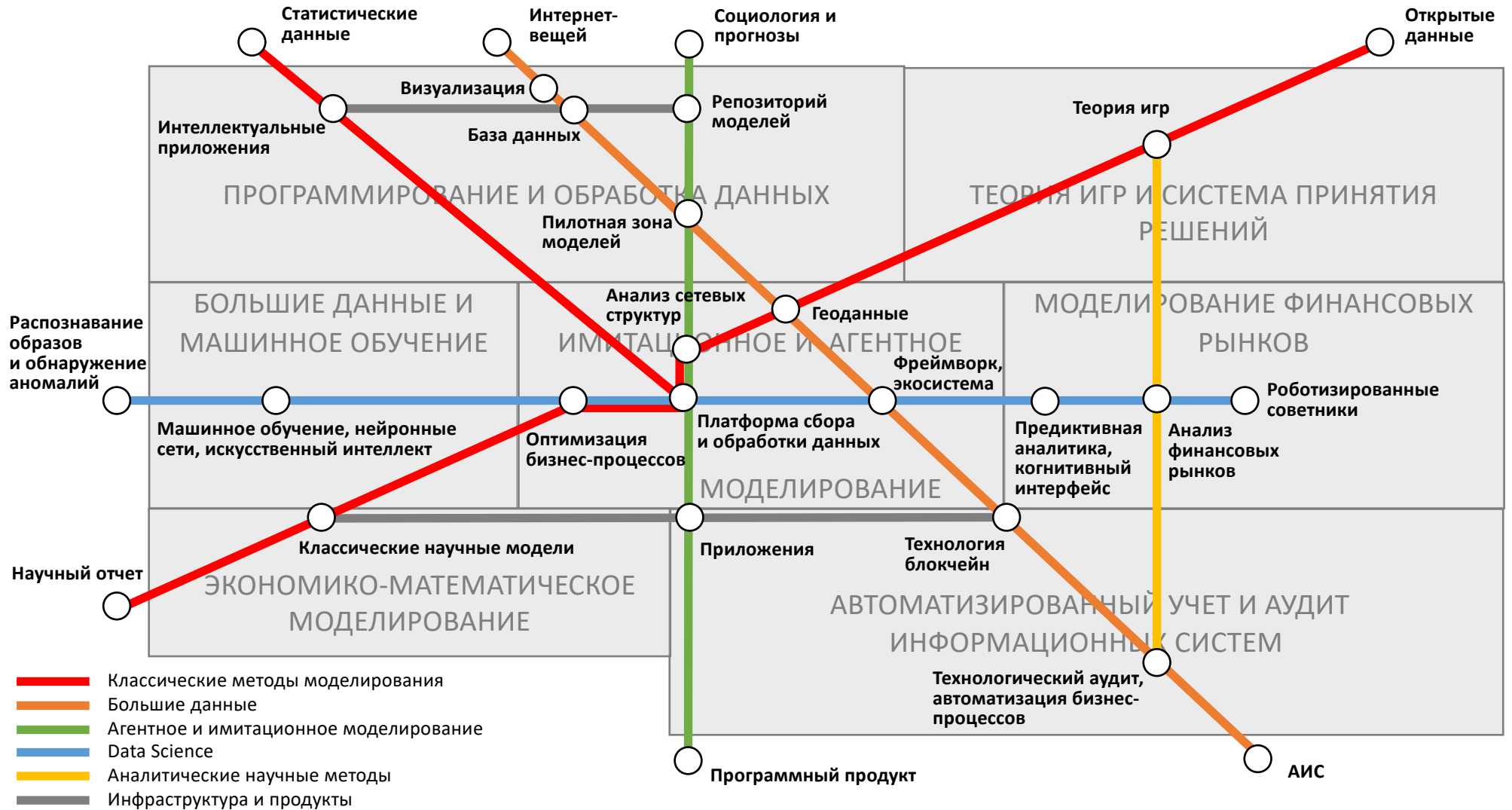
Системы визуализации
данных

Интернет
вещей

Анализ структуры
сложных сетей

Информационная
безопасность

Карта компетенций департамента



Машинное обучение в риск-менеджменте

Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ банков-контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

Выявление аномалий и нестыковок в данных

Распознавание мошенничества и конфликтов

Идентификация пользователей

Машинное обучение в банковском риск-менеджменте

«Банк России привлекает искусственный интеллект и большие данные, чтобы точнее настраивать денежно-кредитную политику и оперативнее отлавливать проблемы на финансовых рынках».



Источник: Роботы, винтажи и big data: Банк России оттачивает модели по ДКП. Интервью с А.Морозовым 08.10.2020 <https://cbr.ru/press/event/?id=8158>

Проекты

- Винтажи макроданных
- Онлайн-инфляция
- Контроль за банками
- Выявление манипулирования



Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ банков-контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

Выявление аномалий и нестыковок в данных

Распознавание мошенничества и конфликтов

Идентификация пользователей

Скоринг клиентов- физических лиц

InClass Prediction Competition

Кредитный скоринг-2

Классификация заемщиков по кредитному статусу

104 teams · 20 years to go

Overview Data Notebooks Discussion Leaderboard Datasets ... My Submissions **Submit Predictions**

Overview Edit

Description

Evaluation

[+ Add Page](#)

Данное соревнование является частью курса "Искусственный интеллект и компьютерные науки" и "Технологии обработки больших данных" в Финансовом университете.

Участники соревнования строят модели кредитного скоринга. Речь идет о том, что когда клиент обращается в банк с заявлением о предоставлении кредита, банк принимает решение о выдаче кредита или об отказе в предоставлении кредита с использованием статистических моделей на основании информации о тех клиентах, которые уже брали кредит (кто-то из них выполнил свои обязательства по кредитному договору, а кто-то не выполнил). На вероятность возврата кредита может влиять много факторов, причем сложным образом, и для прогнозирования результатов по каждому отдельному случаю участникам соревнования необходимо построить модель машинного обучения, которая на основании данных из заявления о выдаче кредита предсказывает, вернет ли заемщик этот кредит.

InClass Prediction Competition

Кредитный скоринг-2

Классификация заемщиков по кредитному статусу

104 teams · 20 years to go

Overview Data Notebooks Discussion **Leaderboard** Datasets ... My Submissions **Submit Predictions**

Public Leaderboard Private Leaderboard

This leaderboard is calculated with all of the test data. [Raw Data](#) [Refresh](#)

#	Team Name	Notebook	Team Members	Score	Entries	Last
1	Chernov Valera			0.81457	11	2mo
2	AnastasiaRazoryonova			0.81362	1	2mo

Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ банков-контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

Выявление аномалий и нестыковок в данных

Распознавание мошенничества и конфликтов

Идентификация пользователей

Анализ банков-контрагентов

Цель

Ранняя идентификация риска неплатежеспособности

Варианты постановки задачи

- Предсказание отзыва лицензии
- Предсказания санации или отзыва лицензии

- Предсказание присвоения дефолтного рейтинга
- Предсказание снижения рейтинга
- Предсказание текущего рейтинга
- Предсказание будущего рейтинга

Только открытые данные

Источник : Банк России
Ограничения: запаздывающие индикаторы

Источник: международные рейтинговые агентства
Ограничения: рейтинги есть не у всех

Метки: рейтинги международных агентств

Исходные данные

Рейтинги S&P и Fitch (Bloomberg), период 2000-2020

~ 850 записей по 116 банкам, включая 18 банков с отозванной лицензией и 23 ликвидированных

Оцифровка рейтинговой шкалы

- Рейтинги ранжированы от 0 до 23, где 0 означает дефолт
- Отсутствие рейтинга (NR) или его отзыв (WD) оцифровывались вручную

	Рейтинг:		Шкалы:
	S&P	Fitch	от 0 до n
"условный" блок AAA	AAA		23
"условный" блок AA	AA+		22
	AA		21
	AA-		20
"условный" блок A	A+		19
	A		18
	A-		17
"условный" блок BBB	BBB+		16
	BBB		15
	BBB-		14
"условный" блок BB	BB+		13
	BB		12
	BB-		11
"условный" блок B	B+		10
	B		9
	B-		8
"условный" блок CCC	CCC+		7
	CCC		6
	CCC-		5
"условный" блок CC	CC		4
"условный" блок C	C		3
"условный" блок R	R		2
"условный" блок перехода из R в D	SD	RD	1
"условный" блок D	D		0
	NR		
	WD		

Признаки: данные отчетности

Исходные данные (только открытые)

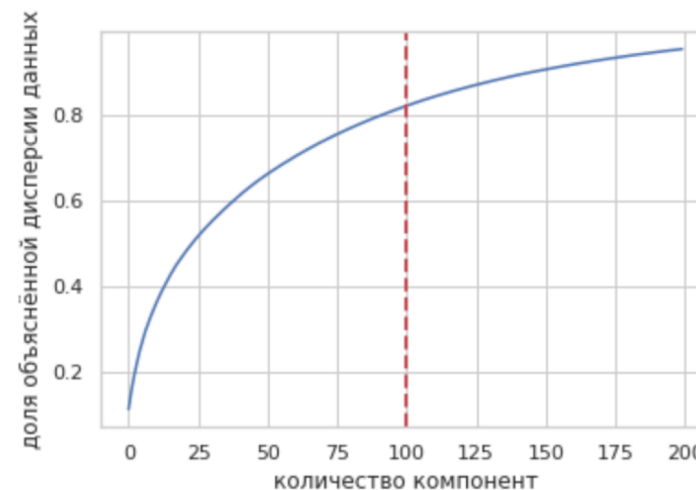
- 101 форма (бухгалтерский баланс)
- 102 форма (отчет о прибылях и убытках)
- 123 форма (капитал)*
- 135 форма (нормативы)*

Преобразования

- Временные срезы (глубина 12 месяцев)
- Масштабирование (по активам)
- Нормализация
- Выделение главных компонент (PCA)
- Группировки и синтетические показатели*

Итоговый датасет

- ~ 12 тыс. записей по 107 банкам на 155 разных дат
- ~ 350 преобразованных признаков
- ~ 70% данных на обучении, 30% на валидации



Общая схема модели машинного обучения

Известные данные
с известными результатами



Разделение выборки

Тестовая выборка (20-40%)

Обучающая выборка (60-80%)

Отделение
истинных
результатов



Оценка качества обученной модели
на контрольной выборке
путем сравнения предсказанных результатов с истинными

Обучение модели



NumPy 

 pandas

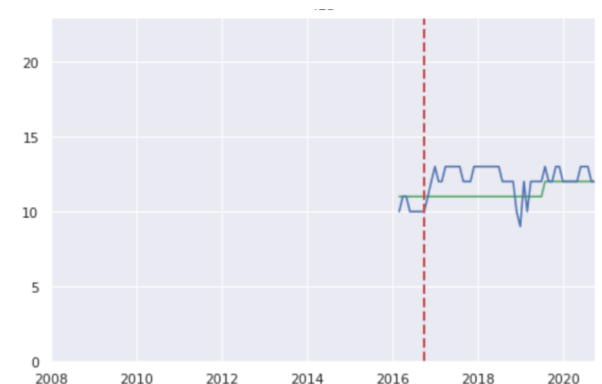
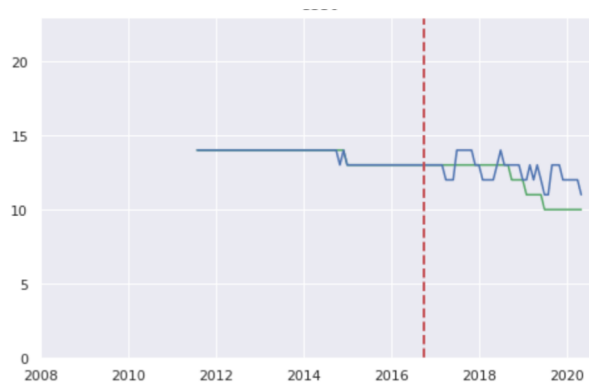
scikit-learn
Machine Learning in Python

Результаты: предсказания значения рейтинга

Временной горизонт:
текущий месяц

Модели:

- XGBoost Regressor

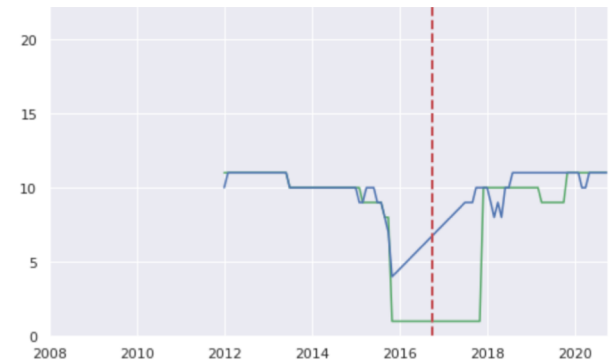


Оценка качества:

- MAE
- RMSE
- R2

Классификатору сложно
идентифицировать редкие события.

Результаты регрессии лучше:
абсолютная ошибка
около 1 пункта

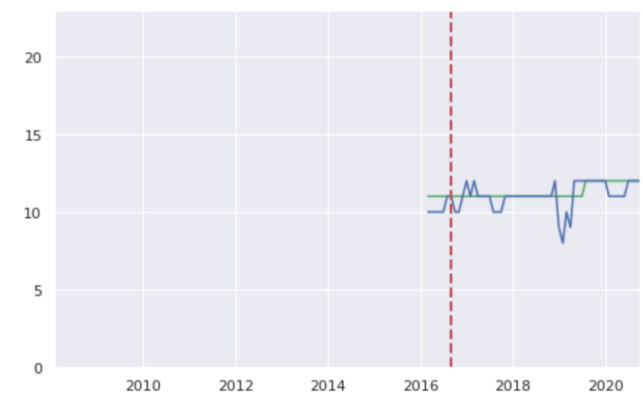
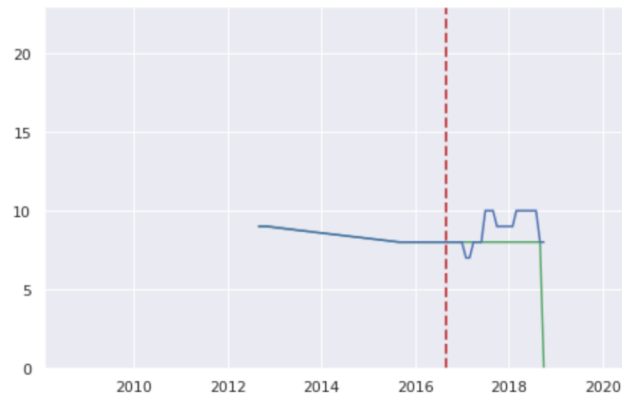


Результаты: предсказания значения рейтинга

Временной горизонт:
следующий месяц

Модели:

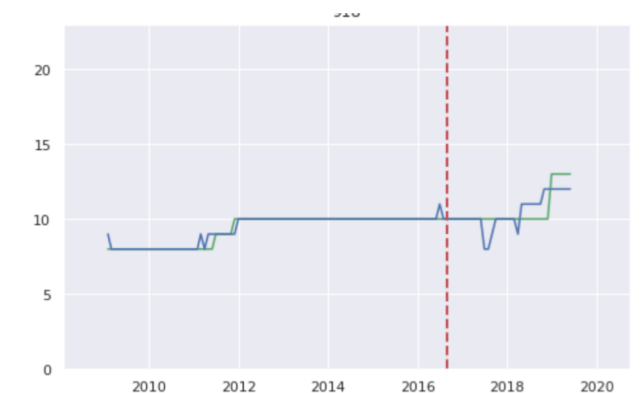
- XGBoost Regressor



Оценка качества:

- MAE
- RMSE
- R2

Точность сопоставима с предыдущей моделью.
Совмещение двух моделей позволит
предсказывать события снижения рейтинга



Перспектива

Модификация признаков

- Синтетические признаки
- Дополнительные отчетные формы
- Другие открытые данные (котировки, новости)
- Закрытые данные (непубликуемая отчетность)

Модификация меток

- Эксперименты с временными горизонтами и рисковыми событиями
- Обогащение закрытыми данными (внутренние рейтинги, профсуждения)

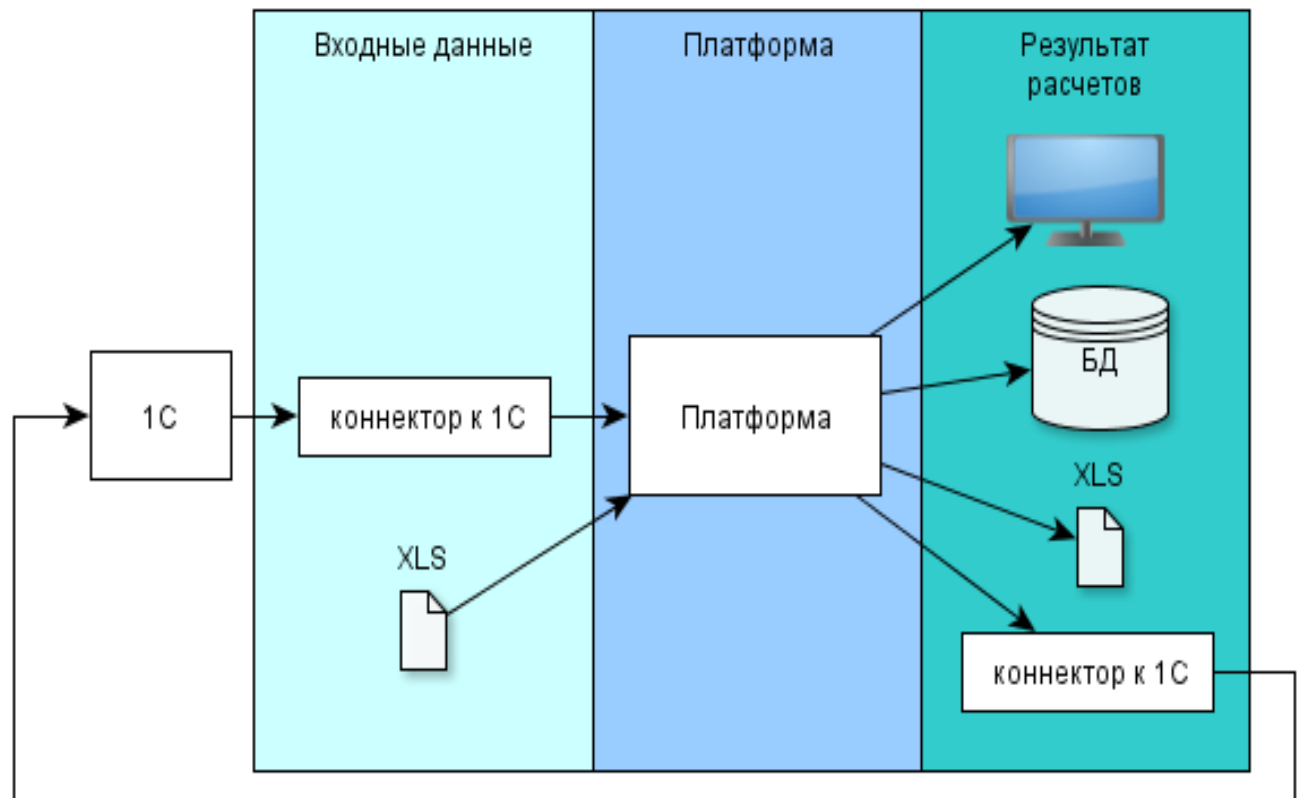
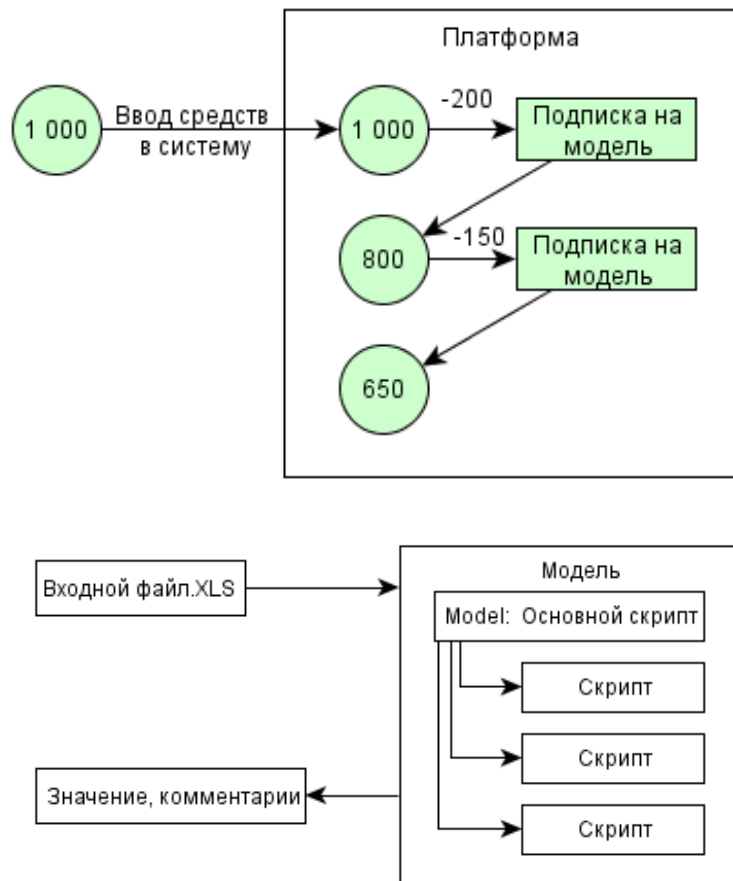
Альтернативное использование

- Скоринг клиентов-юридических лиц
- Оценка рисков страховых компаний

Кому может быть интересно?

- Регулятору
- Банкам, страховым компаниям, фондам
- Вкладчикам
- Экспертным ресурсам

Model Repository и FinDataHub.com



Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

Выявление аномалий и нестыковок в данных

Распознавание мошенничества и конфликтов

Идентификация пользователей

Распознавание состояния рынка

Разработана система машинного обучения, с 80%-ной точностью предсказывающая моменты разворота рыночных трендов

Метрики качества

традиционные		
Accuracy	Precision / Recall / F1	AUC
нетрадиционные		
Freq – частота переключений	Delay – задержка предсказания	TimeInPos – время в позиции
в итоге все проще		
Доходность		

Исходные данные

Open
Close
High
Low
Volume

Что предсказывают?

Цена актива (\$)
Доход или доходность (\$ или %)
Направление изменения цены (↑ или ↓)
Рыночный тренд (↑ или ↓)
на основании данных из прошлого

В итоге: доходность торговой стратегии 30% на DAX



Альтернативное использование:
предсказание дефолта по облигациям

Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ банков-контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

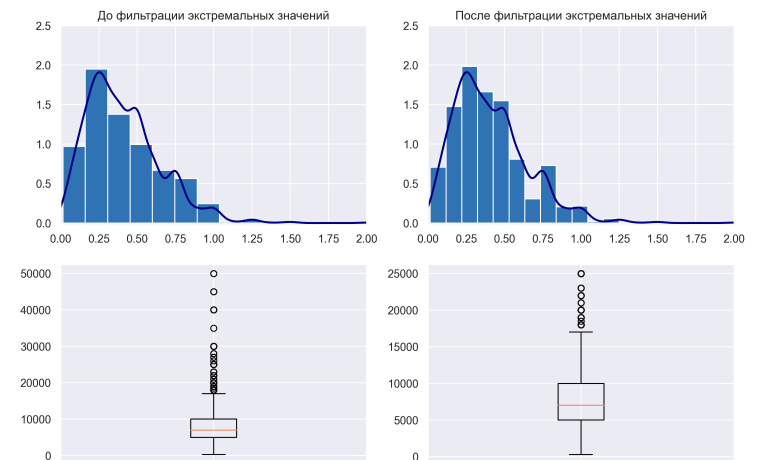
Выявление аномалий и нестыковок в данных

Распознавание мошенничества и конфликтов

Идентификация пользователей

Выявления аномалий и нестыковок

- Методы оценивания функций распределения (например, ядерные оценки плотностей)
- Алгоритмы изолирующего случайного леса
- Одноклассовый метод опорных векторов
- Модификация метода главных компонент
- Комбинация указанных методов



Машинное обучение в банковском риск-менеджменте



Кредитные риски

Скоринг клиентов ФЛ

Скоринг клиентов ЮЛ

Анализ банков-контрагентов

Рыночные риски

Распознавание рыночных трендов

Предсказание дефолтов по облигациям

Операционные риски

Выявление аномалий и нестыковок в данных

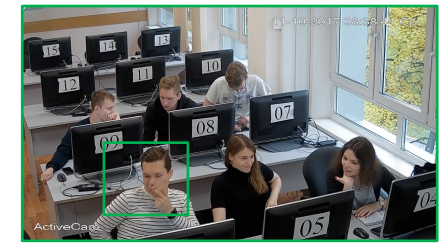
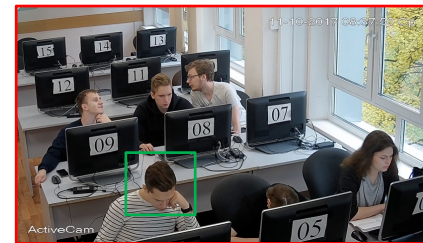
Распознавание мошенничества и конфликтов

Идентификация пользователей

Видео, речевая и текстовая аналитика в операционных рисках

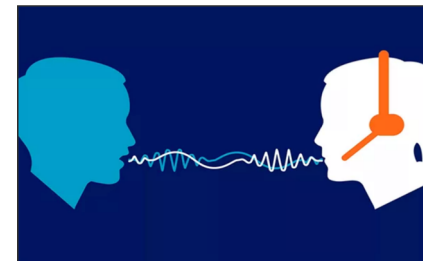
Видеоаналитика

- Детекция
- Идентификация по фото
- Распознавание эмоций по выражению лица



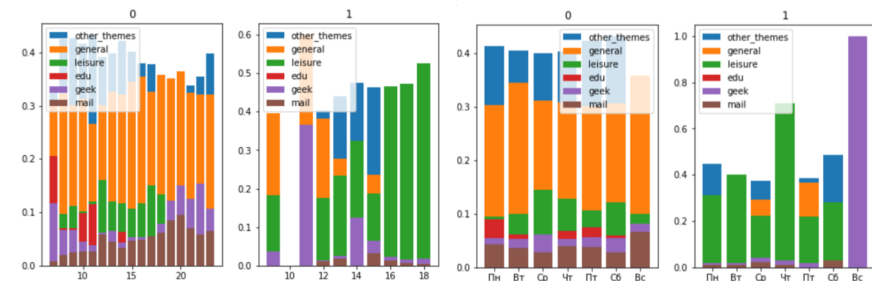
Речевая аналитика

- Идентификация по голосу
- Распознавание эмоций по тембру



Текстовая аналитика

- Идентификация пользователя по истории браузера



Наши заказчики

Заказчики НИОКР



Евразийская
экономическая
комиссия



ФГУП НИИСУ



Российский фонд
фундаментальных
исследований



Альфа-Капитал



НПП «ПУЛЬСАР»



Минэкономразвития
России



Министерство
финансов
Республики Крым



Секретариат Совета
МПА СНГ



ФГУП ГосНИИАС



Всероссийский союз
страховщиков



ПРАВИТЕЛЬСТВО
РОССИЙСКОЙ
ФЕДЕРАЦИИ



Пенсионный фонд
России



Федеральный центр
проектного
финансирования



Всероссийская
академия
внешней торговли



ФГУП ЦНИИ центр



Палата Налоговых
Консультантов



СБЕРБАНК



ГОСУДАРСТВЕННАЯ ДУМА
ФЕДЕРАЛЬНОГО СОБРАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ



Екатерина Золотарева, к.э.н.

Доцент Департамента анализа данных и
машинного обучения

Руководитель Научно-технического центра
банковских информационных технологий

elzolotareva@fa.ru

Департамент анализа данных и машинного обучения

105187, Москва, Щербаковская ул., 38 ☎ +7 (499) 503-47-02, доб. 4713 ✉ bigdata@fa.ru

