



Data Science для Банка Кейс АО «Сбербанк».

06 декабря, 2017

DS, ML и люди, которые этим занимаются. Взгляд Сбербанка



ИССЛЕДОВАТЕЛЬ ДАННЫХ + аналитический инструмент



Разработка
модели
t = 3-4 недели



Результаты
модели



ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ



Разработка
модели
t = 3-4 часа



Результаты
модели



НАПРАВЛЕНИЕ ИИ

процесс принятия решений

ЦЕЛЬ

автоматизация процесса актуализации
и разработки моделей кредитного скоринга

МЕТОДЫ

деревья решений, полносвязные нейросети,
бинарная классификация

ОБЛАСТЬ ПРИМЕНЕНИЯ

кредитный скоринг, оптимизация кредитного
процесса, снижение уровня просроченной
задолженности

Качество моделей аналогично



Ручная
работа
эксперта

- Проверка документов
- Проверка стоп факторов
- Проверка в других источниках
- Проверка работодателя
- Верификация по телефону



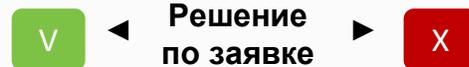
Модель принятия
решений



Андеррайтер

Издержки

Качество



Модель

Издержки

Качество

НАПРАВЛЕНИЕ ИИ

процесс принятия решений, работа с естественным языком/текстом

ЦЕЛЬ

автоматизация процесса принятия решения по кредитной заявке, эмуляция решений андеррайтера

МЕТОДЫ

деревья решений, полносвязные нейросети, бинарные модели

ОБЛАСТЬ ПРИМЕНЕНИЯ

кредитный скоринг, оптимизация кредитного процесса

Кейсы применения Machine Learning от Сбербанка

Кейс: Анализ карточных транзакций клиентов

Паттерн покупки авто

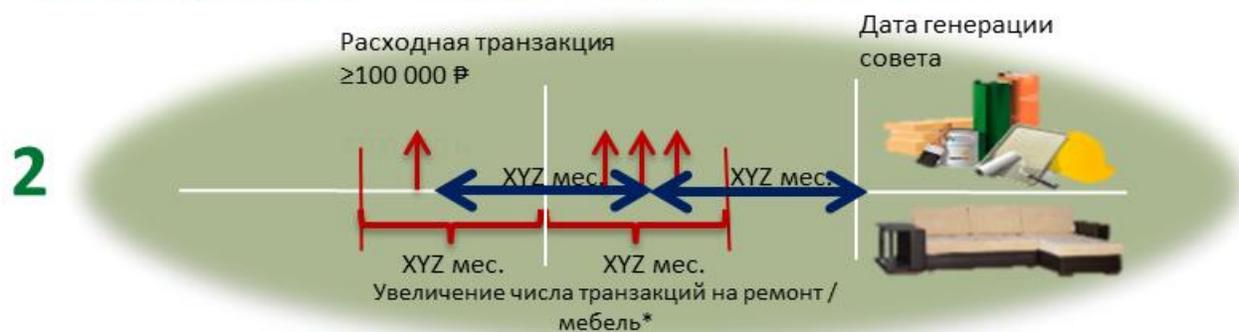


За X месяцев до крупной транзакции не было покупок, связанных с авто, а после появились

Поздравляем с покупкой авто! Полезная информация:

- Тех.сервис
- Мойка

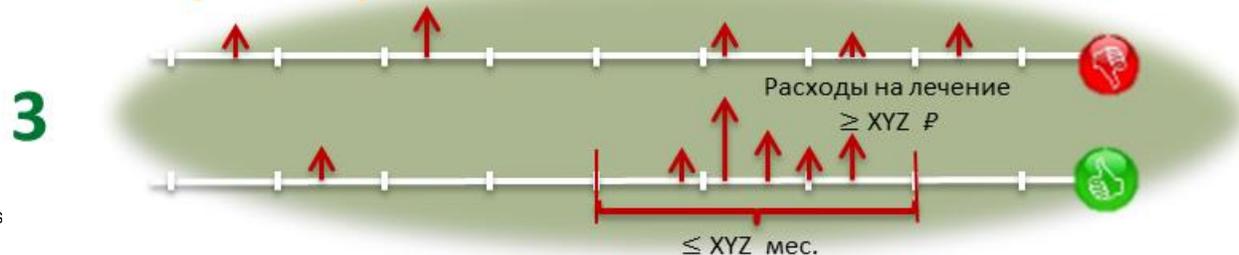
Паттерн ремонта/покупки мебели



При сравнении месяца крупной покупки и месяца после наблюдается рост числа транзакций на ремонт, мебель

Поможем завершить ремонт! Скидки, акции на покупку мебели и строительных материалов

Паттерн затрат на лечение



По истории транзакций находим период в который происходит активная оплата лечения в РФ ($\geq K \text{ ₺}$)

Вы знаете, что можете оформить налоговый вычет?

<https://vozvratnalogov.online/>

Кейсы применения Machine Learning от Сбербанка

Кейс : Оптимизация воронки продаж

ШАГ I

Анализ данных,
в т.ч. транзакционных
Way4, ЦОД, кред. фабрика

ШАГ II

Выявление паттернов и
сегментация клиентов
по характеристикам

ШАГ III

Формирование
продуктовых
предложений на базе
характеристик клиента

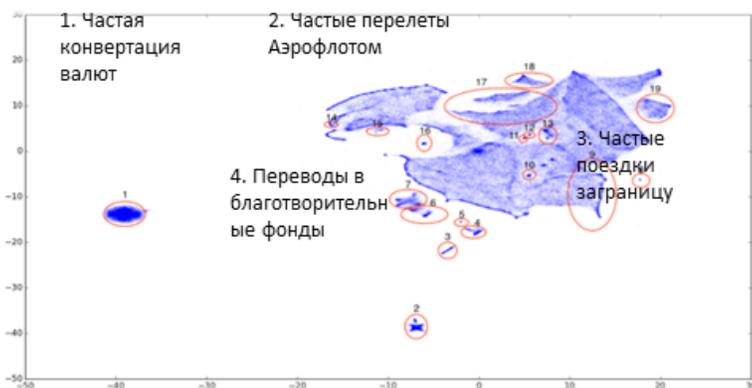


ЭКОНОМИЧЕСКИЙ ЭФФЕКТ

- ✓ Рост эффективности воронки продаж
- ✓ Рост лояльности клиентов

МЕТОДЫ алгоритмы кластеризации, визуализация данных большой размерности с использованием LargeVis

КЛАСТЕРИЗАЦИЯ КЛИЕНТОВ ПО ХАРАКТЕРУ ТРАНЗАКЦИЙ



В ЗАВИСИМОСТИ ОТ КЛАСТЕРА КЛИЕНТА ПРЕДЛОЖИТЬ РЕЛЕВАНТНЫЙ ПРОДУКТ



Паттерн	Продукт
1. Частая конвертация валют	Мультивалютный счет
2. Частые перелеты Аэрофлотом	Карта «Аэрофлот Бонус»
3. Частые поездки за границу	Страховка для выезжающих за рубеж
4. Переводы в благотворительные фонды	Карта «Подари жизнь»

ВЕБ-СЕРВИС ПОИСКА КАНДИДАТОВ

Описание вакансии Заместитель руководителя Администрации Губернатора Московской области (гос. служба):

Функция: Достижение высокого качества государственной гражданской и муниципальной службы. Соблюдение законодательства о государственной гражданской и муниципальной службе Московской области

Основные задачи: Разработка основных направлений кадровой политики и развития гражданской и муниципальной службы. Реализация законодательства по вопросам гражданской службы. Организация профессиональной переподготовки, повышения квалификации гражданских служащих. Координация деятельности государственных органов при решении вопросов поступления, прохождения и прекращения гражданской службы. Формирование и использование кадрового резерва Московской области. Подготовка управленческих кадров

Оценочные критерии:

- Основное образование
- Доп. образование
- Опыт работы
- Нежелательный опыт работы
- Стаж гос. службы
- Общий стаж
- Достижения

ФИО	Баллы	Сбербанк Рейтинг
Березин Илья Сергеевич	100.00	😊
Голубев Андрей Алексеевич	90.50	😊
Григорьев Алексей Владимирович	74.25	😊

Настройки поиска кандидатов:

Образование [0;5]: 2

Опыт работы по специальности [0;5]: 2

Нежелательный опыт работы [-5;0]: -1

Удельный вес общего стажа [0;0.005]: 0.001

Удельный вес стажа гос. службы [0;0.05]: 0.01

Статистика поиска:

- 827 - резюме в базе
- 437 - резюме попало в результаты
- 2 - резюме входит в топ 25%
- 8 - резюме входит в топ 50%

НАПРАВЛЕНИЕ ИИ

работы с естественным языком/текстом

ЦЕЛЬ

разработка алгоритма для автоматического сопоставления резюме с требованиями к кандидатам

МЕТОДЫ

семантический анализ текста, рекуррентные нейронные сети

ОБЛАСТЬ ПРИМЕНЕНИЯ

формирование кадрового резерва, поиск кандидатов

Модели анализа кандидатов и эффективности действующих сотрудников

данные соц.
сетей



выполнение
планов



оценка
эффективности

НАПРАВЛЕНИЕ ИИ

процесс принятия решений, работа с естественным языком/текстом

ЦЕЛЬ

предсказать надежность, эффективность, вероятность оттока работников массовых специальностей

МЕТОДЫ

градиентный бустинг, рекуррентные нейросети, байесовские методы

ОБЛАСТЬ ПРИМЕНЕНИЯ

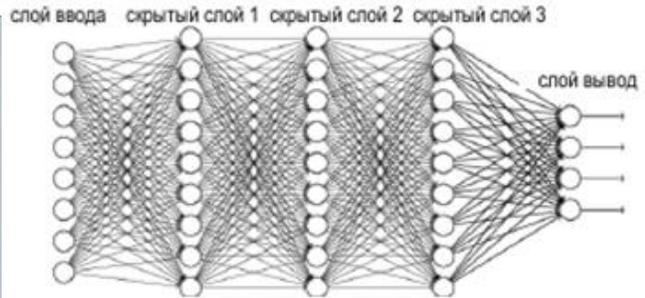
анализ кандидатов, система мотивации, оценка эффективности сотрудников, прогноз оттока

Кейс Михаила Косински

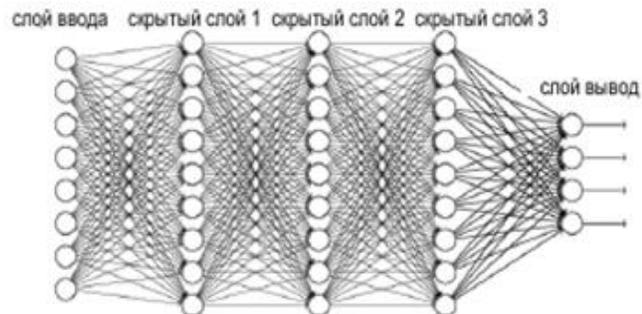
психометрический анализ личности по поведению в соцсетях



1. Фото ➤ 2. Глубокая нейронная сеть ➤ 3. Профиль



Интроверт



Экстраверт

НАПРАВЛЕНИЕ ИИ

компьютерное зрение, работа с естественным языком/текстом

ЦЕЛЬ

прогнозирование типов личности на основе альтернативных источников информации (поведение в соц. сетях, фото и пр.)

МЕТОДЫ

кластеризация, деревья решений, сверточные нейросети, OCEAN

ОБЛАСТЬ ПРИМЕНЕНИЯ

психометрический анализ, исследование социальных групп

Кейсы применения Machine Learning от Сбербанка

Кейс. Моделирование вероятности дефолта для малого бизнеса в режиме реального времени

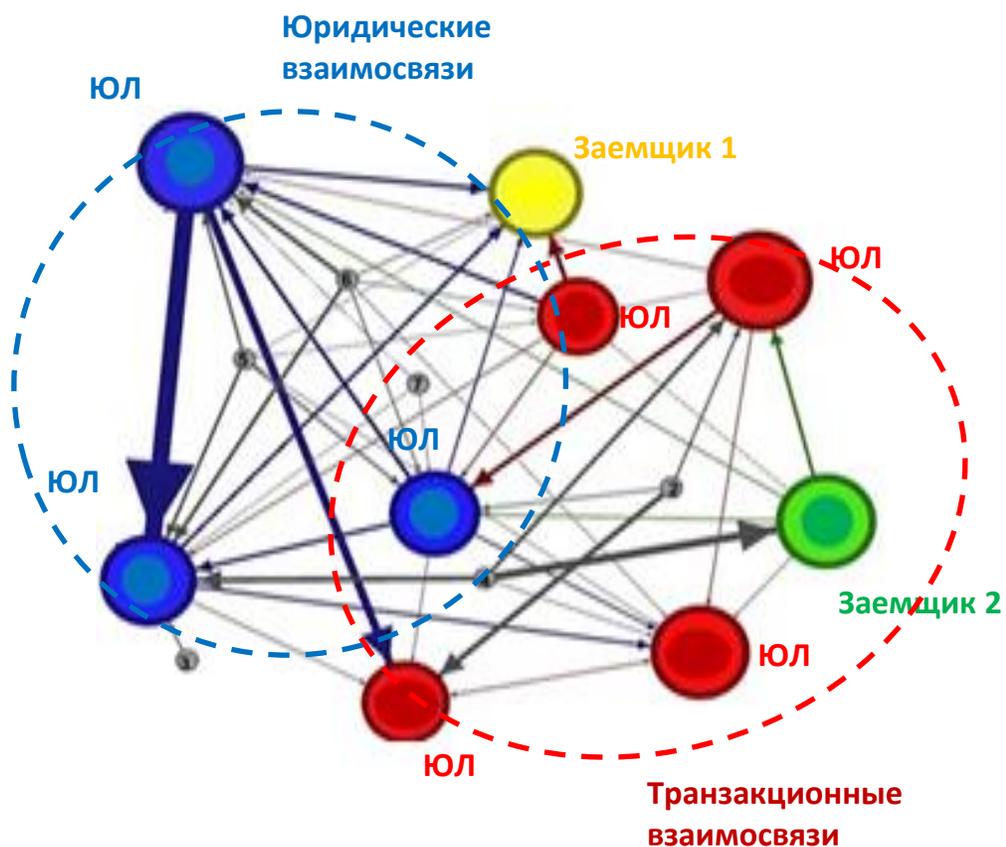
AS IS: ВНЕШНИЕ ИСТОЧНИКИ
оценка рисков за 1-2 дня

TO BE: ТРАНЗАКЦИОННЫЕ ДАННЫЕ
real-time оценка рисков

НЕДОСТАТКИ

- Трудовые и временные затраты
- Недостоверность отчетности
- Субъективность оценки качеств. факторов
- Смещенность оценки, если нет кред. истории
- Сбор полного комплекта затруднителен





НАПРАВЛЕНИЕ ИИ

процесс принятия решений, работа с естественным языком/текстом, компьютерное зрение

ЦЕЛЬ

выявить взаимосвязи между клиентами на основе транзакций и данных из разных источников

МЕТОДЫ

графовая аналитика, деревья решений, семантический анализ текста, рекуррентные нейросети

ОБЛАСТЬ ПРИМЕНЕНИЯ

определение сомнительных операций, в том числе не связанных с экономической деятельностью, для целей комплаенс и риск-менеджмента

Кредитный процесс за 7 минут (1/4)

ГЛАЗАМИ КЛИЕНТА

1

2

Предложение банка,
выбор продукта,
кастомизация кредита

3

Подписание
документации

4

Получение денег
на счет

ПРОЦЕСС: 7 МИНУТ

ГЛАЗАМИ БАНКА

внутри нейросети: **AI-Decision Making** – CNN и **AI-Credit Structuring** – CNN, по мере накопления обратной связи клиентов – с элементами RL/Deep RL на основе решений клиентов

МОДЕЛИ

Предрасчет в облаке

Сегментация клиентов
Расчет риск-метрик
на данных T-1
Расчет подходящих
клиентам продуктовых
предложений



Decision Making
+
Credit Structuring

SAS ОКР

Получение и обработка
ответа БКИ
Корректировка риск-
метрик на данные БКИ
Генерация финальных
предложений клиенту

Online оценка залогов

РАБОТА С КЛИЕНТОМ И СДЕЛКОЙ

Web-интерфейс

Отображение клиенту
автоматически подобранных
продуктов
Интерактивная настройка
продукта под клиента
Подписание кредитной
документации

Back Office

Открытие счета
перечисление денег клиенту

Кейсы применения Machine Learning от Сбербанка

Кейс. Определение B2B- и B2C-цепочек

БИЗНЕС ЗАДАЧА выявить оптимальные B2B и B2C бизнес «цепочки» на основе анализа экономической активности компаний различных отраслей и регионов

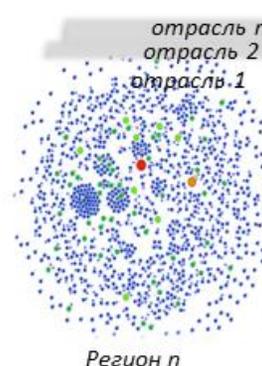
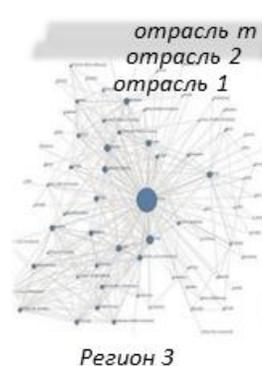
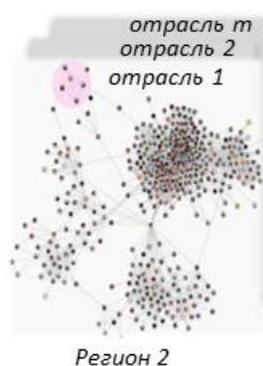
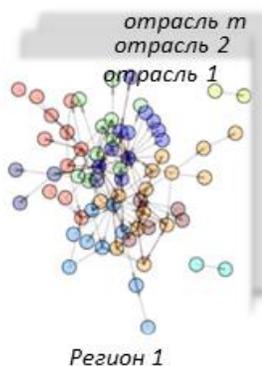
ПРИМЕНЕНИЕ по запросу клиента предоставлять рекомендаций по оптимальным бизнес партнерам, например, при выходе клиента на новый для него регион, рынок и пр.



ЭКОНОМИЧЕСКИЙ ЭФФЕКТ

- ✓ Банк единый marketplace
- ✓ Возможности для новых услуг
- ✓ Рост лояльности

В ОСНОВЕ РАЗРАБОТАННОЙ МОДЕЛИ АНАЛИЗ ГРАФОВ



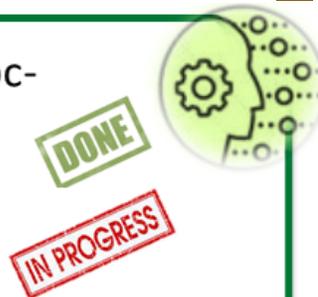
ML АЛГОРИМЫ ДЛЯ ПОИСКА ОПТИМАЛЬНЫХ СВЯЗЕЙ



Кейсы применения Machine Learning от Сбербанка

Кейс. Алгоритмы для чатбота @SberbankML_Bot

- ☑ Разработка алгоритма выявления цепочек «Вопрос-Ответ» (conversational agents)
- ☐ Следующий шаг – наполнение бота содержанием и создание «базы знаний»



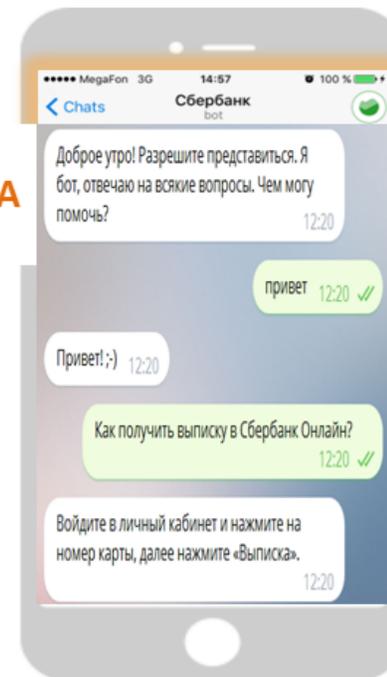
ML ИНСТРУМЕНТЫ

- ✓ Rule-based approach
- ✓ Named-entity recognition
- ✓ Deep Learning
- ✓ Sequence to sequence learning

Требования к библиотеке и данным



В РЕЗУЛЬТАТЕ
ИНТЕРАКТИВНАЯ БЕСЕДА
КЛИЕНТА И SBERBANKBOT



Data Scientist - профессия настоящего

- ▶ **Data Business people** (предприниматели) – группа специалистов, ориентированных на организационные вопросы и получение прибыли от проектов, связанных с данными. Как правило, это люди с инженерным или техническим образованием и степенью MBA.
- ▶ **Data Creatives** (приблизительно можно перевести как "свободные художники" данных) – специалисты, которые творчески подходят к анализу. Иногда это даже хакеры. Как правило, это люди из научной среды с академическим опытом, экономисты и статистики по образованию.
- ▶ **Data Developer** (девелоперы данных) - специалисты, сосредоточенные на технических проблемах управления данными: как получать, хранить данные и учиться на них. Эта группа изо дня в день пишет код, около половины имеют образование в сфере компьютерных наук, у них больше остальных групп прокачаны скиллы Big Data и машинного обучения.
- ▶ **Data Researchers** (исследователи данных) – учёные и исследователи данных, которые практикуют научную работу, публикуются в профильных изданиях, имеют научные степени. В этой группе наименьшее число людей, связанных с бизнесом.

rahmat