Практика реализации подхода «Клиент 360˚» в банке

Андрей Кислый

Лидер практики

Евгений Балюк

Big Data разработчик

СОДЕРЖАНИЕ

- Подход «клиент 360˚»
- Поведенческий анализ
- Архитектура решения
- Опыт реализации

Подход «Клиент 360°»

Личные характеристики

- контакты
- интересы
- социальный статус

Жизненные события

- свадьба
- рождение ребенка
- приобретение жилья

Интересы

- хобби
- поведенческие характеристики

Обогащенный профиль клиента

Товары

- История покупок
- Факты лояльности
- Инциденты
- Отзывы
- Рекомендации

Временные состояния

- Текущее положение
- Намерение совершить покупку

Отношения

- личные
- деловые

Выявление признаков → Поведенческие данные → Статистические модели → Бизнес-логика

Эволюция клиентских данных и социальные медиа

Стадия	Задачи	Требования
Лидо-генерация	- Сегментировать - Оценить потенциал - Дополнительный канал коммуникации	- Прогноз поведенческого профиля - Оценка привлекательности - Прогноз потенциальных продуктов - Оценка вероятности сделки
Кредитная анкета	- Обогатить информацию для скоринга - Борьба с мошенничеством	- Идентификация клиента - Прогноз поведенческого профиля - Обогащение существующей модели скоринга
Развитие клиента	- Единый профиль, максимум информации - Подтвердить/опровергнуть оценку скоринга, прогноз - Понять интересы клиента	- Идентификация дубликатов - Формирование/обогащение поведенческого профиля - Оценка вероятности дефолта
Удержание	- Анализ доходности - До-продажа / Кросс-продажа	- Прогноз интересных продуктов - Оценка вероятности сделки
Урегулирование	- Сбор данных - Подтвердить/опровергнуть данные клиента	- Максимальная детализация данных по клиенту - Оценка поведения при различных сценариях коммуникации

На разных этапах взаимодействия клиента с банком полнота данных о нем меняется и значимость открытых (внешних) данных также разная

Сфера применения

Продажи и Маркетинг

Увеличение числа розничных клиентов, рост объема продаж на одного клиента - за счет ориентации на индивидуальные особенности и предпочтения клиента. Повышение отдачи на проводимые маркетинговые кампании за счет более качественной сегментации.

Управление рисками и скоринг

Использование для оценки риска невозврата всей информации о клиенте, находящейся в открытом доступе. Рост клиентской базы за счет снижения ошибки второго рода.

Безопасность

Противодействие инсайдерской деятельности, мошенничеству и AML

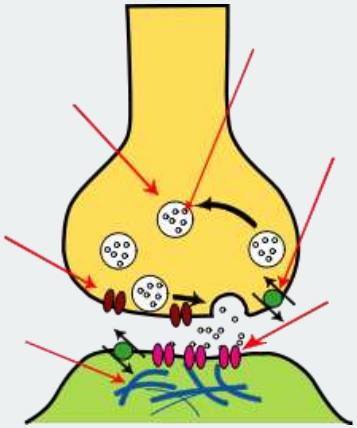
Это не все

Оптимизация скорости обслуживания, данные о доступности каналов обслуживания (банкоматы, терминалы и т.п.), оперативный контроль за качеством обслуживания

Компоненты решения

Техническая составляющая	железо, софт, разворачивание кластера, тюнинг, ETL, организация вычислений, вопросы интеграции
Социальные сети	доступ к данным, мечинг, загрузка, хранение
Обработка данных	технологии анализа текста, графики, категоризация данных
Поведенческая аналитика	поведенческая модель: система метрик, признаков и интерпретаций, технологии их расчета в кластере
Data Science и бизнес аналитика	корреляционный анализ, бизнес-анализ, поиск пользы в данных, бизнес ценности

Поведенческий анализ – Что это



Особенности нашего поведения и примем ли мы то или иное решение напрямую зависят от нашего профиля нейрогуморальной регуляции и накладывают отпечаток на всю нашу жизнь.

В целом можно выделить несколько базовых гормонов, нейромедиаторов и структурных особенностей нашего мозга - которые оказывают наибольшее влияние на специфику нашего поведения.

Базовые гормоны и нейромедиаторы: Адреналин и Норадреналин, Тестостерон, Допамин, Окситоцин и Вазопрессин, Серотонин, Кортизол.

Структурные особенности: развитость лобных долей, качество проводящих путей между различными участками головного мозга и тп – ведут к большей или меньшей выраженности самоконтроля, качеству реализации исполнительных функций мозга (внимание, оперативная память, целенаправленность, долгосрочное планирование…)

Кроме того, на наше поведение также влияет уровень доступности ресурсов и их изменчивости.

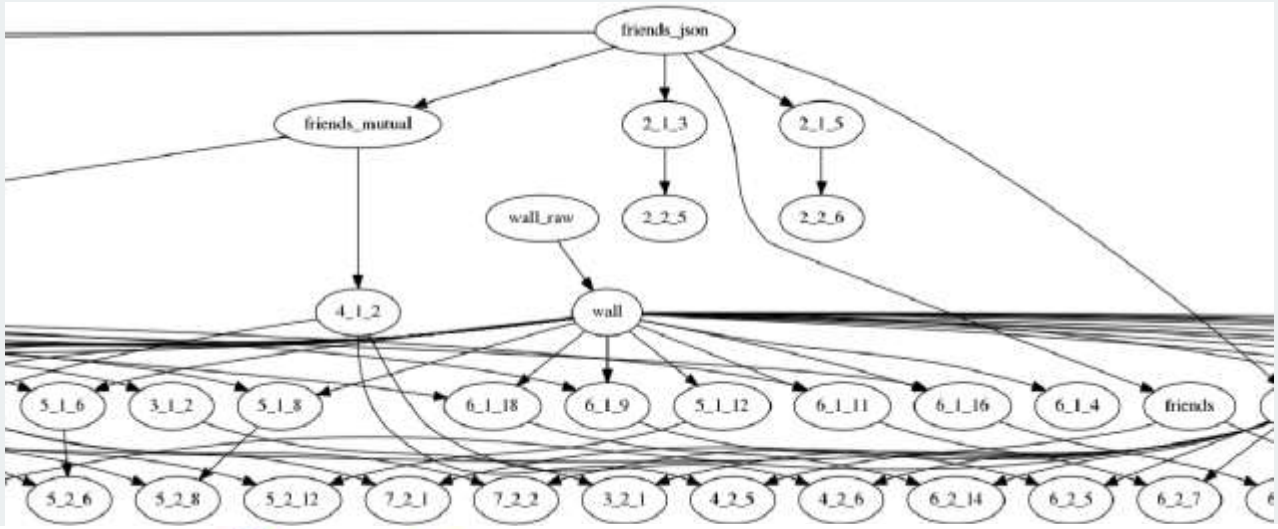
В результате именно доступность и изменчивость ресурсов формирует особенности наших поведенческих стратегий и каждого из нас можно описать как совокупный профиль этих особенностей и свойственных нам стратегий принятия решений и поведения.

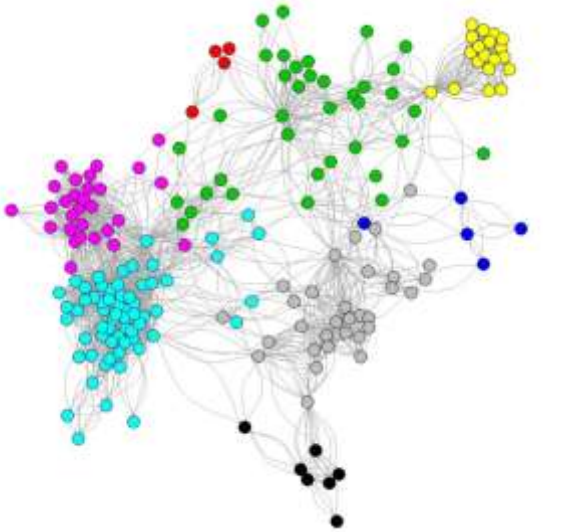
Поведенческий анализ – Рабочий пример

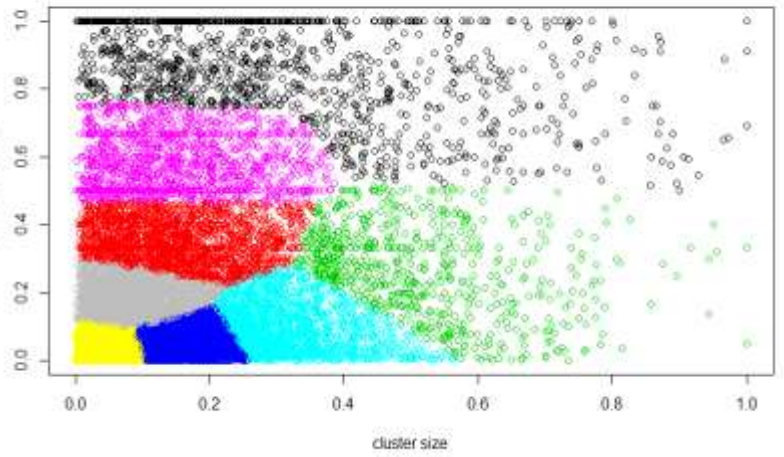
	Возраст	Пол	Регион
Значение	21	Мужской	город < 50 тыс
Вектор психографи-ческого типа	Демонстративный, зависимый	Демонстративный, зависимый	Примитив
Тип риска	Гипероптимизм (узкий фокус внимания, низкий уровень горизонта планирования), импульсивность, уязвимость к социальным провокациям		
Сценарий развития	Взятие завышенной суммы, нецелевое использование, отсутствие планирования, прокрастинация в процессе обслуживания долга, недостаточность обеспечения, потенциальное развитие ситуативно-средового типа		

Поведенческий анализ - Интерпретации

Схема трансформации социального профиля (комплекс метрик, показателей, интерпретаций) для определения типа личности (согласно модели)







Поведенческий анализ - Архитектура сервисов

Обработка открытых данных

Постановка задачи

- Количество цифровых данных растет экспоненциально
- Ручная обработка возможна только в режиме расследования – единичный анализ, человек учитывает контекст данных
- Для качественной автоматизации обработки такого массива данных необходимы инструменты с применением машинного обучения

Выбор архитектуры

Hadoop:

- Распределённое хранение
- Распределённая обработка большого количества данных
- Затраты на масштабирование – минимальны

Hive:

- Удобный SQL - подобный язык запросов
- Распределённые вычисления
- Возможность построчной обработки данных

Выбор архитектуры

Подводные камни:

- скорость обработки когда данных мало
- сложность взаимодействия потоков между собой

Варианты решения:

- отказаться от реализации Hadoop
- вынести часть решений за кластер (для некоторых задач лучше подходит локальная обработка, чем обработка на кластере)

Социальный анализ – получение данных

Сбор данных

- Доступ через REST API
- При ограниченном доступе - 3 запроса в секунду

Формат данных

Права доступа:

- Открытые и закрытые методы
- Ограничения для закрытых методов
- Почему иногда лучше выбрать закрытые методы

Социальный анализ – получение данных

Подводные камни:

- Время на один запрос
- Специфическая работа некоторых методов

Варианты решения:

Нужно больше токенов:

- Просим у сотрудников и должников
- Регистрируем новые аккаунты

Социальный анализ – используемые методы

Кластеризация друзей

- выделение групп друзей, наиболее близких по интересам

Тематический классификатор

- определение интересов пользователя
- поиск целевых категорий

Анализ тональности

- отношения пользователя к тематике

Используемые методы – Кластеризация

Социальные связи наиболее ценная информация в сети.

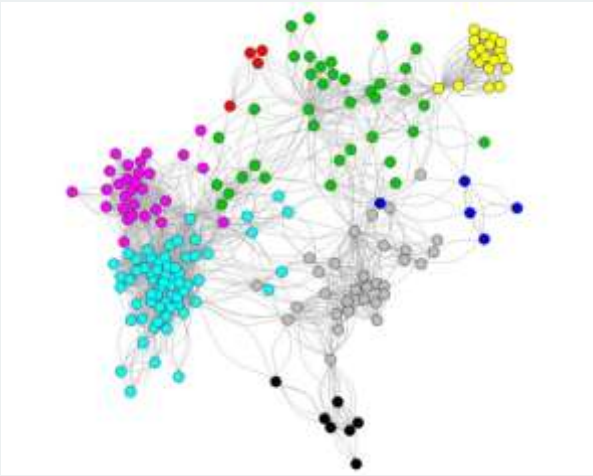
Анализ списка друзей позволяет

- определить круг общения человека
- его интересы
- заполнить отсутствующие поля профиля

Кластеризация проводилась при помощи статистического языка программирования **R** и пакета **igraph**

Алгоритмы:

1. walktrap
2. edge betweenness
3. multilevel
4. label propagation
5. leading eigenvector
6. spinglass



Используемые методы – Кластеризация

Пример анализа (группы сотрудников)

Лучшие кластера:

- Коллеги по работе
- Школа
- Университет
- 1-2 группы по интересам

Минусы:

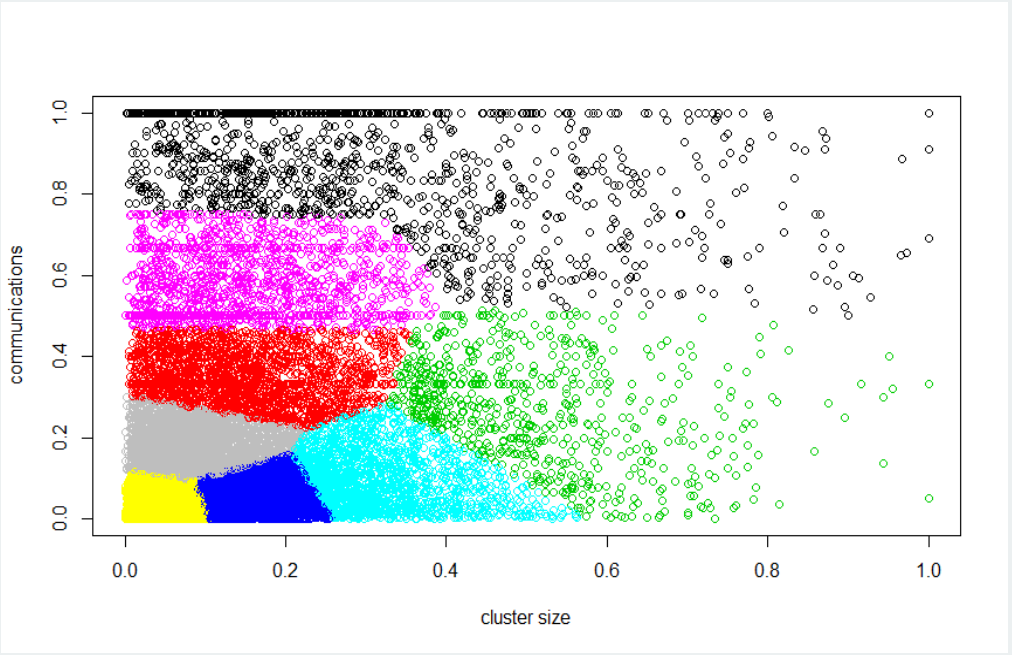
- Иногда один крупный логический кластер разбит на несколько частей
- Нетривиальная автоматизация выявления сути кластера

Используемые методы – Кластеризация

Анализ коммуникаций

Позволяет определить:

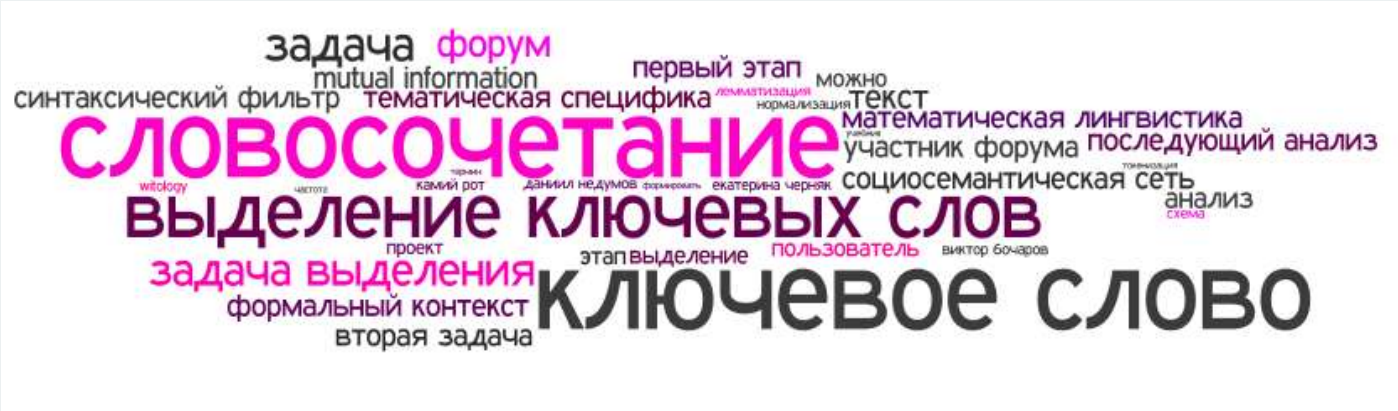
- Как много внимания человек уделяет каждому кластеру/интересу
- Является ли человек неотъемлемой частью кластера, или только пытается стать ею
- Вектор распределения интересов пользователя



Используемые методы – Классификатор

Назначение:

- Определение тематики высказывания
- Для соц-сети
  - выявление интересов пользователя
  - определение сути кластера, если кластер тематический (не вуз/работа)



Используемые методы – Классификатор

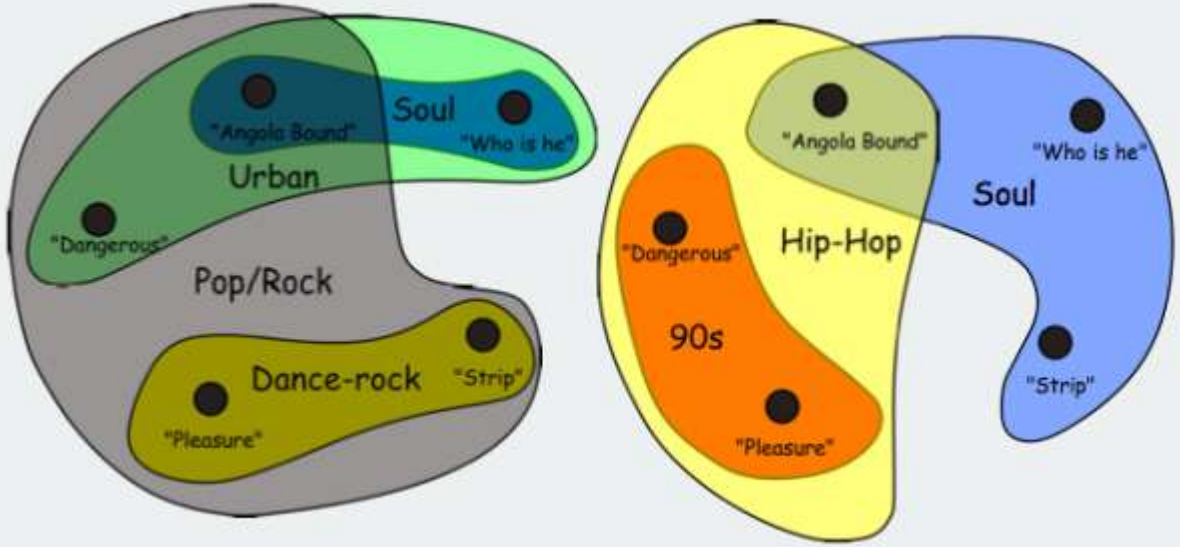
Подготовка выборки

Данные были взяты из википедии по различным категориям:

- Бизнес
- Наука
- Обучение
- История
- Сельское хозяйство
- ...

Используемые методы – Классификатор

Обучение модели
Multilabel



Используемые методы – Классификатор

Результаты

До двух меток на один текст

Точность достаточно высокая. Но пока только для больших текстов

Следующий шаг - тематическая классификация небольших текстов

Используемые методы – Анализ тональности

Постановка задачи

Получить следующую информацию о компании/продукте:

- популярность за время
- популярность в регионе
- количество клиентов в регионе
- отношение и тональность

Корреляционный анализ с имеющейся информацией:

- финансовая отчетность банка
- результаты маркетинговых исследований (опросов)
- прочее

Используемые методы – Анализ тональности

Сбор данных:

1. Определение
2. Сохранение html кода страниц в сыром виде
3. Парсинг кода, получение сообщений и ссылок на профиля пользователей
4. Сбор профилей пользователей

Язык программирования - python

Обработка данных

1. Поиск упоминаний банка в сообщении
2. Анализ тональности высказывания

Параллельная обработка в кластере

Используемые методы – Анализ тональности

Тренировочный сет

Достаточное количество форумов уже имеет отметки пользователей о тональности, это и служило тренировочными данными

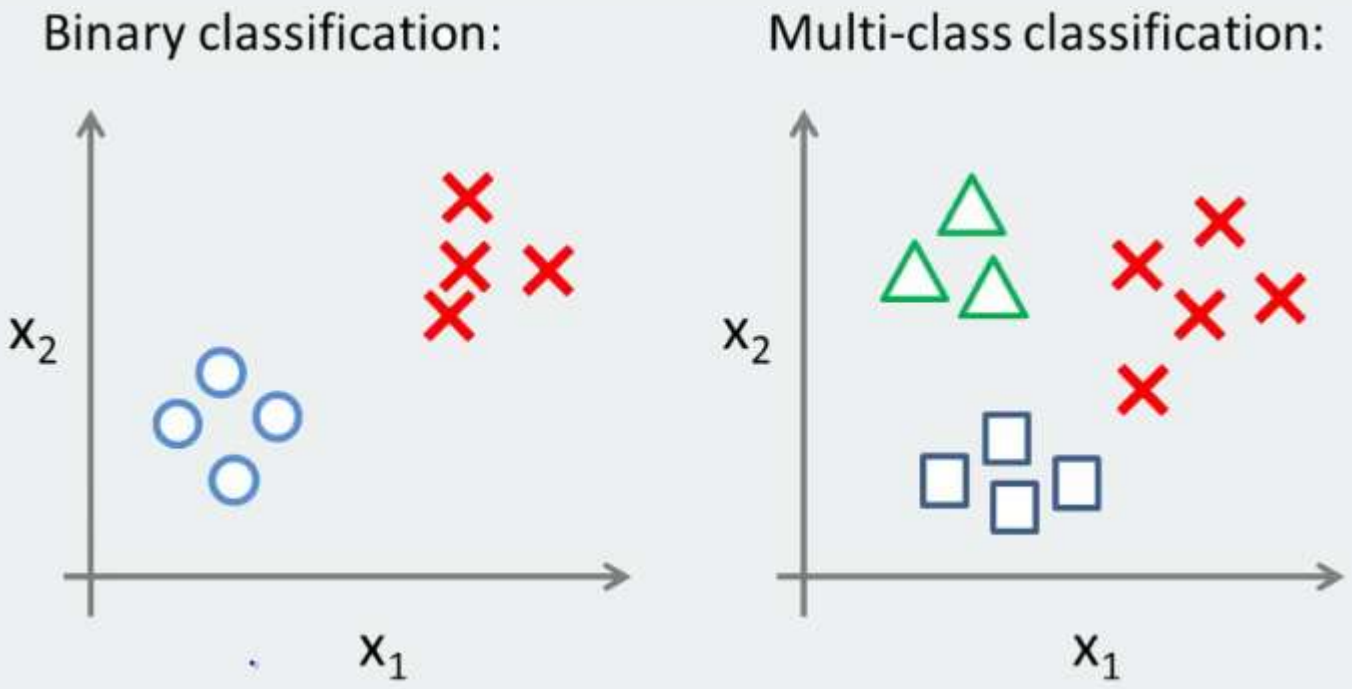
Отзывы о банках были взяты с форумов:

- banker.ua
- banki.ru
- banki.ua
- bankinfo.ua
- infobank.by
- minfin.com.ua

Данные собраны за 2007 - 2014 (июнь)

Используемые методы – Анализ тональности

Для определения трех меток использовалась Multiclass классификация вместе с методом опорных векторов



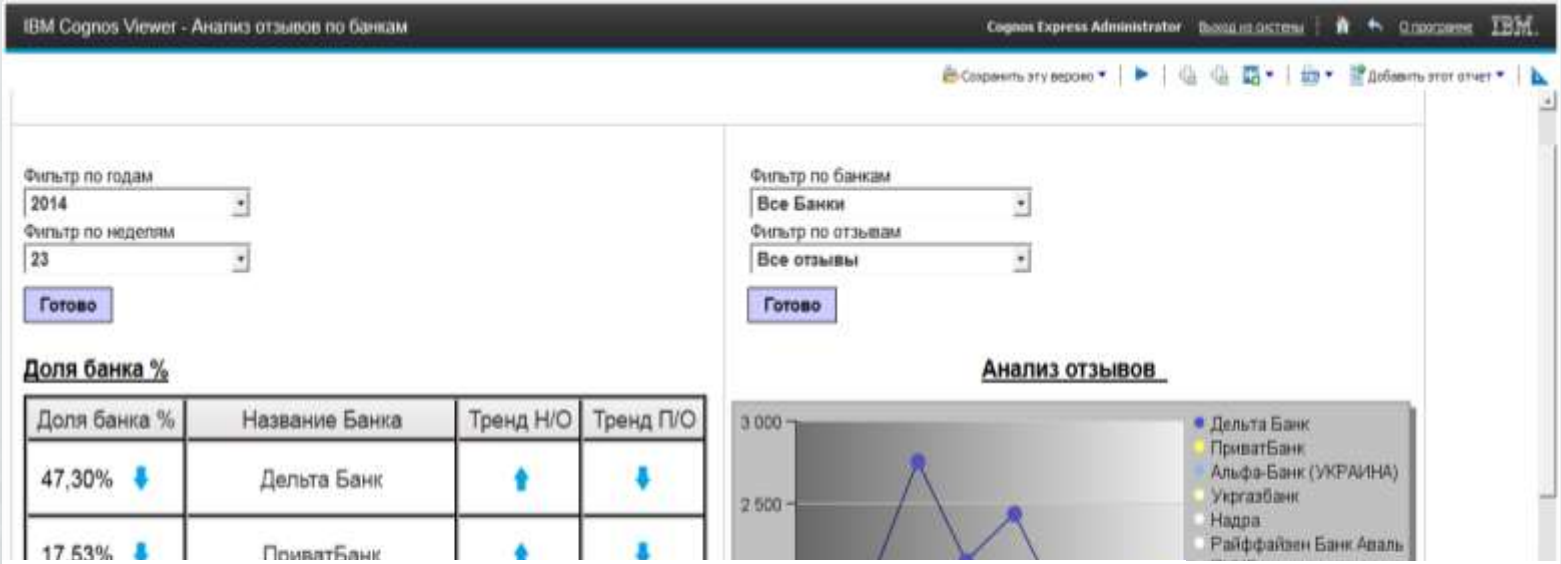
Используемые методы – Анализ тональности

Результаты

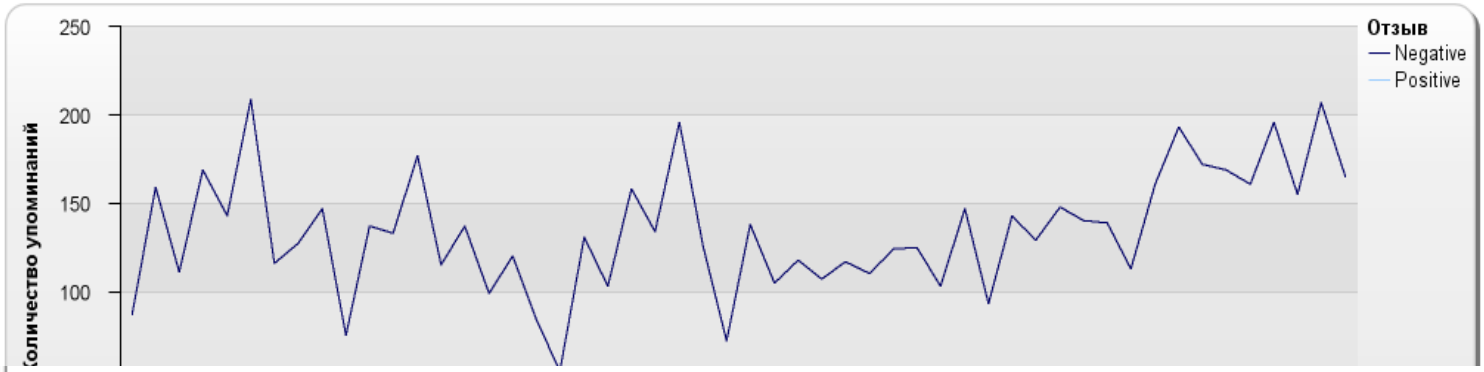
Точность сегодняшней модели составляет 80%

Но не стоит останавливаться пока точность ниже 99%

Пример работы



Динамика обсуждений



Общая статистика за 2014 год

Количество Банков	Количество ресурсов	Количество Регионов	Количество отзывов	Количество позитивных	Количество негативных	Количество нейтральных
13	7	35	38 692	5 644	28 673	4 375

СПАСИБО ЗА ВНИМАНИЕ!

Андрей Кислый

Лидер практики

+38 067 505 99 51

info@ibdi.pro

www.ibdi.pro

Евгений Балюк

Big Data разработчик