Текстовая аналитика для анализа большого объема неструктурированных данных

Илья Вигер
Генеральный директор

Разработано при поддержке

Yandex
Data Factory

www.vesolv.com

XXI век: новые подходы к использованию данных

Информация становится одним из основных активов компаний в 21 веке, умение извлекать полезные знания и затем их монетизировать становится ключевым фактором успеха на рынке.

Компании получили доступ к терабайтам, петабайтам, эксабайтам информации. Источники информации разделены на 3 типа: структурированные, неструктурированные и частично структурированные

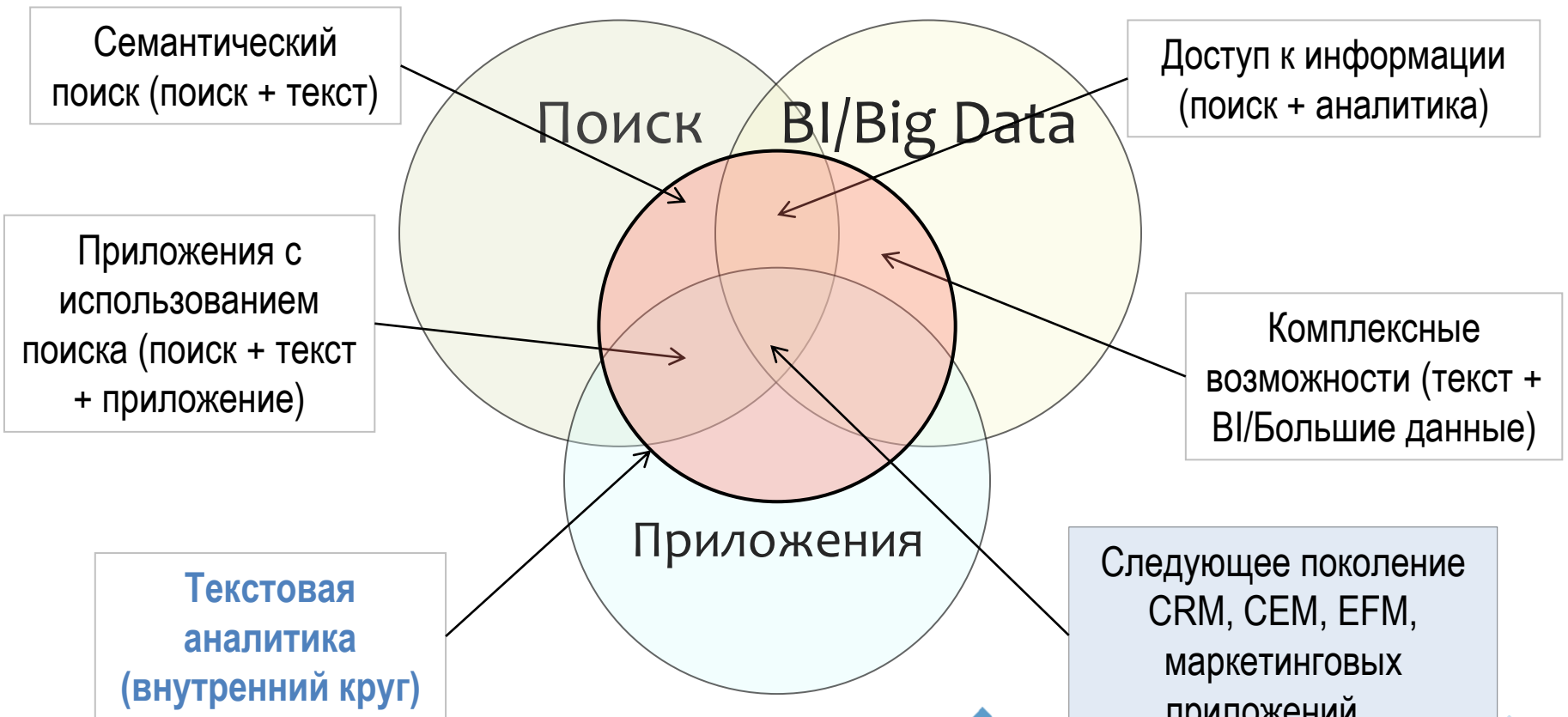
Что можно делать с данными доступными в Интернет, в соц. сетях, или внутри предприятий?

1. Отправлять, пересылать, публиковать, редактировать, архивировать.
2. Искать и индексировать
3. Категорировать и классифицировать в соответствии с задачами предприятия
4. Извлекать информацию и анализировать
5. Проверять гипотезы и моделировать

Человечеству потребовалось 300 тысяч лет, чтобы создать первые 12 эксабайт информации, зато вторые 12 эксабайт были созданы всего за два года

КАК? С помощью инструментов текстовой аналитики!

Что такое «текстовая аналитика»?

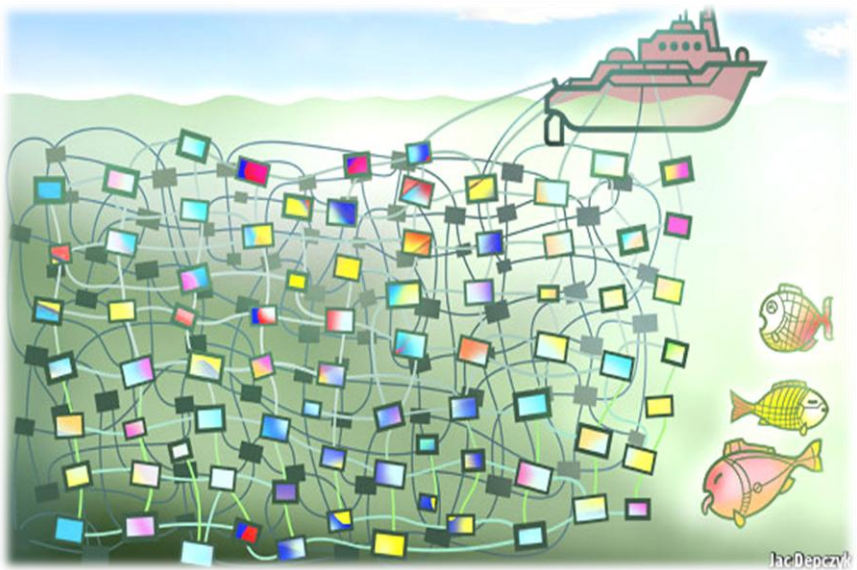


ТА неотъемлемая часть решения ADVANCED BI

	Business Intelligence	**Advanced Analytics**
Вопросы и ответы	**Что происходит?** Когда? Кто? Как? Сколько?	**Почему это происходит?** Произойдет ли это снова? Что если ..? О чем мы не думали, но могли бы спросить/подумать?
Основные функции	Отчеты (KPIs, метрики, дэшборды) Создание запросов на лету Анализ данных (кубы данных, срезы данных, навигация по данным) BI реального времени Автоматический мониторинг и оповещение	Статистически и количественный анализ **Извлечение неструктурированных данных** Предсказательные модели и дополнительные аналитические функции Анализ BIG DATA **Текстовая Аналитика** Сценарное моделирование

Основные возможности ТА: Добывание данных

- Извлечение данных из множества различных источников. Смещение акцента от поиска к обработке извлеченных данных
- Поиск и извлечение информации в режиме реального времени
- Адаптация коннектора к источнику данных при изменении структуры данных. Автоматический разбор неструктурированных документов, выделение заголовка, основной части документа, анализ структуры документа и ссылок
- Определение типов данных. Выбор обработчика в зависимости от типа данных. Обработка аудио, и видео информации
- Широкие возможности по извлечению данных из соц. сетей, публичных источников данных, наиболее посещаемых сайтов и специализированных источников

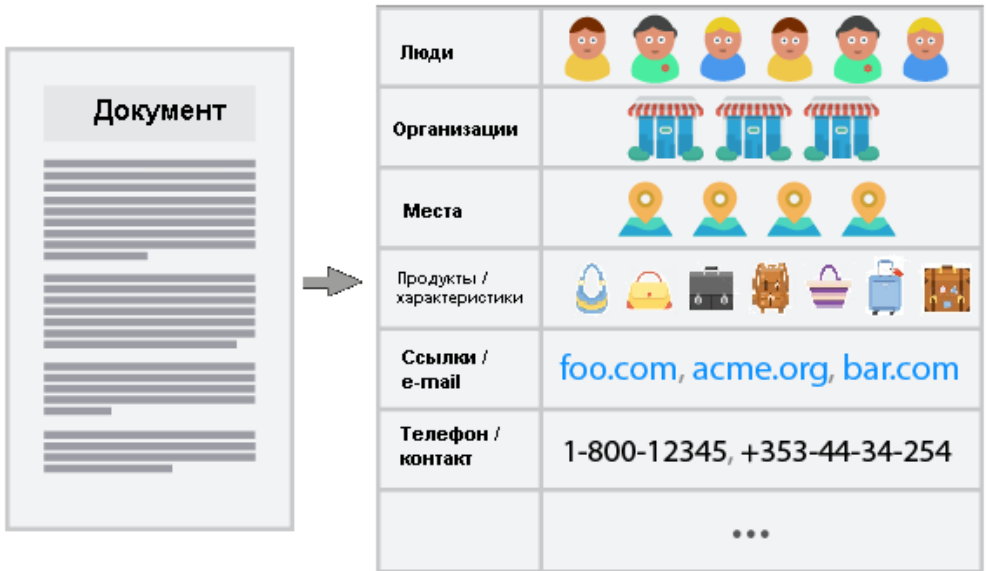


Основные возможности ТА: Категоризация

- Кластеризация документов, создание и наполнение категорий в документах. Выделение параграфов и разделов документов с использованием алгоритмов кластеризации
- Поддержка Pull (авторубрикация) и Push (предопределенные рубрики) сценариев для категорий
- Группировка и категоризация документов и частей документов в соответствии с заданными сценариями
- Использование комплексных алгоритмов для категоризации. Использование самообучаемых (статистических) алгоритмов и алгоритмов обучения с учителем
- Ведение [разных] категорий для разных ролей пользователей
- Использование категорий для последующего извлечения информации из неструктурированного текста

Основные возможности ТА: Извлечение информации

- Извлечение значимой информации из текста
- Выделение собственных имен, людей, организаций, географических мест, наименований, ссылок на сайты, e-mail, сообщения на форумах, сообщения в соц. сетях и пр.
- Выделение продуктов/услуг и их характеристик, в т.ч. характеристик связанных с обслуживанием клиентов
- Наполнение данными CRM-систем. Поиск и наполнение данными о профилях клиентов, отношениях между клиентами, выявление родственных связей, выявление степени влияния между клиентами
- Извлечение прочей информации релевантной для маркетинга (брендтрекинг, позиционирование по сравнению с конкурентами и пр.)

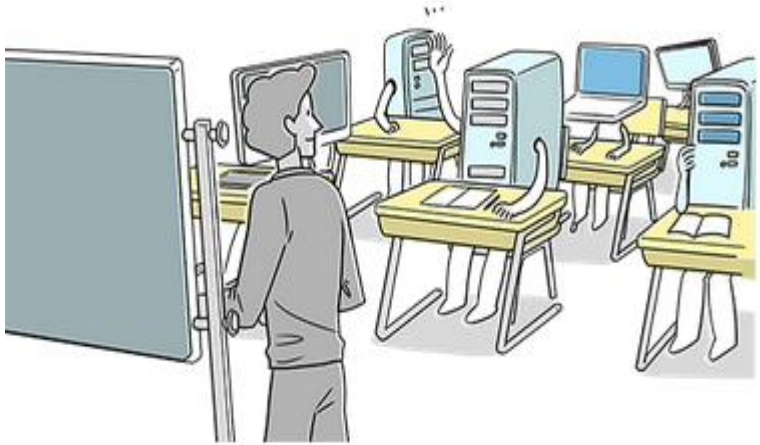


Основные возможности ТА: Анализ мнений

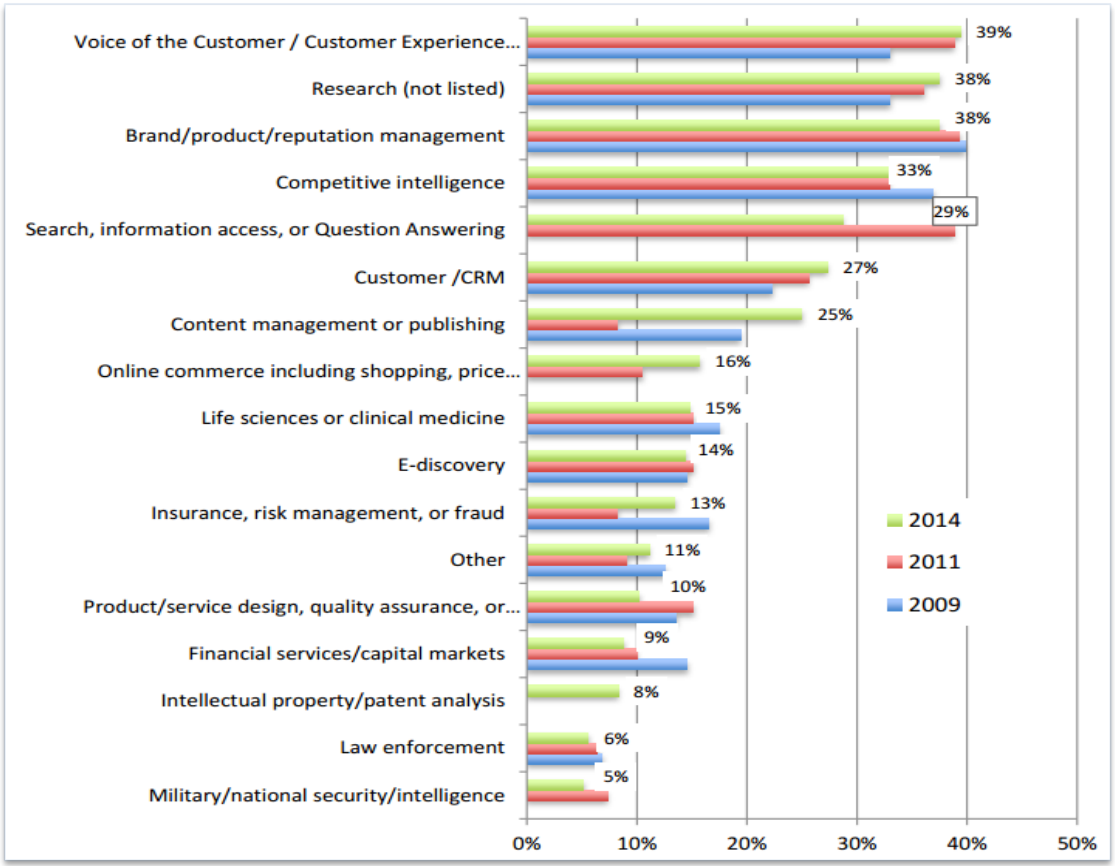
- Анализ мнений используется для дополнительного извлечения скрытой информации, которую сложно извлечь используя классические методы
- Анализ мнений включает выявление тональности сообщений, определение эмоциональной составляющей высказывания, выделение суждений, мечтаний и намерений клиента
- Использование тональности для сценариев активного снятия негатива со стороны клиента
- Разделение фактов и суждений
- Анализ желаний и мечтаний клиентов для разработки новых продуктов и новых маркетинговых стратегий

Основные возможности ТА: Машинное обучение

- Использование математических алгоритмов для обучения компьютера анализу текстовой информации
- Использование комплексных методов обучения: статистических методов, нейронных сетей, метода энтропии, дерева решений и пр.
- Использование машинного обучения вместе с ручным разбором для повышения качества работы ТА
- Разработка новых алгоритмов машинного обучения направленных на «подражание» человеку для автоматизации «диалога с клиентом»



Тенденции



Исследование компании Alta Plana, 2014

Повышение интереса в части

- Голоса клиента, Customer Experience
- Исследованиях разного рода
- Информации о клиентах, CRM
- Управления контентом

Снижение интереса в части

- Управления брендом / продуктом / репутацией
- Анализа конкурентных преимуществ
- Поиска

Применение ТА для «Голоса Клиента»

Области применения голосов

Требования бизнеса

Собранные голоса

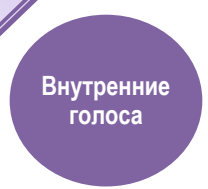
Каналы сбора голоса клиента

- Индекс лояльности (NPS)
- Брендтрекинг
- Раздражители

- Офисы обслуживания
- Call-center
- Точки продаж (POS)
- Зарплатные проекты
- Удаленные каналы (личный кабинет, моб. приложения, банкоматы)
- Телемаркетинг
- Collection

- Новые внутренние проекты
- Совместные проекты с партнерами

- Претензионная база
- Книга ОиП
- Бланки в офисах обслуживания
- Канцелярия
- E-mail компании
- Сайт компании
- Соц.сети (офиц. группы и вне офиц. групп)
- Специализированные сайты
- Моб. Приложения
- Call center



- Опросы
- Анкетирование
- Интервью
- Внутренние порталы

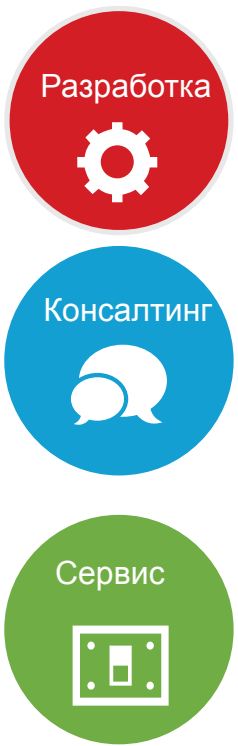
Область применения

Ускоренное удовлетворение потребностей, повышение лояльности клиентов	• Срез клиентских впечатлений, позволяющий определить потребности клиентов и возможности повышения удовлетворенности клиентов • Определение ключевых позитивных и негативных драйверов • Активный контроль отношений с клиентом при взаимодействии с компанией
Повышение эффективности управления брендом и репутацией	• Оперативное выявление случаев негативных для компании ситуаций и реакция на них, прежде чем клиенты поделятся опытом со своими друзьями или информация попадет в масс-медиа • Реакция на негатив со стороны клиента в тот момент когда все еще можно исправить
Современный подход к управлению жизненным циклом продуктов	• Сбор информации о предпочтениях клиентов в отношении продуктов и их преимуществ • Использование сильных и слабых особенностей конкурирующих продуктов • Сбор и углубленный анализ зарождающихся тенденций, их выявление и использование • Увеличение жизненного цикла продуктов за счет выявления возможностей использования продуктов «no name», сегментов клиентов и линеек продуктов • Выявление реакции клиентов на выход продукта в реальном времени

Область применения

Повышение эффективности продаж и маркетинга	• Выявление возможностей up sale и cross sale • Персонализированное предложение для клиентов, использующих Интернет канал • Измерение влияния от акций и кампаний • Измерение эффекта от изменения цены • Выявление наиболее резонансных трендов
Существенное улучшение клиентского сервиса	• Сокращение «посредников» слушающих клиента, сокращение искажений, снижение оттока • Сокращение объема внутренних коммуникаций, автоматизация и повышение эффективности обработки клиентских запросов • Раннее выявление повторяющихся запросов и выработки приемлемого решения • Повышение качества базы знаний для самостоятельного решения проблем клиентом
Повышение эффективности планирования и проектирования	• Автоматизация задач сбора, категоризации и отчетности по обратной связи • Сокращение неизбежных ошибок от ручной категоризации и обработки обратной связи • Снижение стоимости обработки обратной связи • Больше времени и усилий тратиться на развитие бизнеса и увеличение выручки, а не на обработку обратной связи

Платформа VoC от VESOLV и Yandex Data Factory



Голос клиента (Voice of the Customer, VoC)

Решение для автоматизации сбора, и анализа отзывов клиентов из различных источников информации с использованием алгоритмов текстовой, аудио- и видео- аналитики.

Информация извлекается из внутренних информационных систем компании, а также и соц. сетей и открытых источников.

Производится автоматизированный разбор сообщений (рубрикация, определение тональности, выделение фактов и пр.).

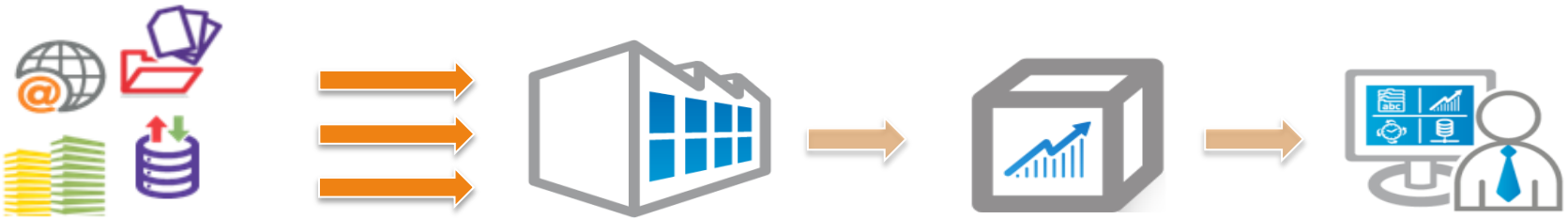
Формируется отчетность для принятия решений.

Больше полезной информации о клиентах, - дешевле

- Сбор информации из разных источников
- Лингвистический, семантический онтологический анализ
- Учет особенностей языка, сленга
- Распознавание текста, аудио/видео. Выделение сути, окраса, суждений и фактов
- Самообучение и обучение с учителем
- Набор преднастроенных отчетов
- Открытый код, возможность доработки внутренними ресурсами
- Разработано при поддержке Яндекс

VESOLV

Как это работает?

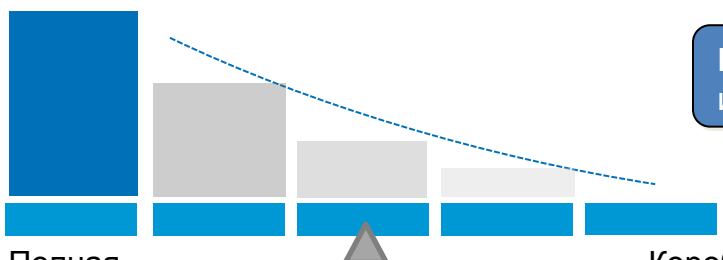


Подключение к источникам	Обработка данных	Создание индексов	Создание отчетов
• Подключение к источникам данных в реальном времени • Получение информации в зависимости от политик безопасности • Использование внутренних и внешних источников данных	• Преобразование форматов данных (HTML, SQL, MS Office, XML, audio…) • Распознавание языка, исправление опечаток • Выделение сущностей, категоризация и классификация • Выделение связей между частями текста, поиск смыслового объекта	• Масштабируемый, высокопроизводительный репозитарий • Полнотекстовый поиск и интерфейс настройки • Создание разных представлений для одного и того же текста • Самообучающийся алгоритм	• Конструктор отчетов с возможностью детализации • Преднастроенные отчеты и дэшборды • Сценарии Real-time • Возможность экспорта и интеграции с другими приложениями

Преднастроенные коннекторы, отчеты, алгоритмы

Кластерная отказоустойчивая архитектура с возможностью масштабирования

Открытая архитектура, возможность доработок



Полная кастомизация

Коробочное решение

Цель: сокращение стоимости извлечения полезной информации

VESOLV

Примеры использования VoC (финансовый сектор)

1. **Снижение оттока клиентов**

VoC позволит сократить негатив клиентов от обслуживания. Клиент приходит в офис компании и получает обслуживание. В течение обслуживания клиента осуществляется анализ разговора для оценки эмоционального состояния клиента. При необходимости осуществляется SMS уведомление менеджера отделения. По завершению обслуживания клиенту отправляется запрос об оценке удовлетворенности, это делается в течение 5-10 минут после завершения обслуживания. Основная идея «поймать» клиента, прежде чем он не начал распространять негативную информацию где-то еще.

2. **Сокращение времени от Идеи до вывода Продукта в продажу**

VoC позволит упростить и ускорить тестирование Идей и рекламных акций. Например для уточнения параметров кривой спроса и чувствительности клиентов к ставке кредита можно запустить «Новый кредитный продукт» и мгновенно начать собирать отзывы и клиентские впечатления о продукте. При этом будет возможность непрерывно сравнивать кол-во клиентов использующих классические продукты и «Новый кредитный продукт», анализировать как характеристики продукта влияют на выбор и использование продукта. Это можно будет делать on-line.

3. **Разработка продуктовой матрицы финансовой организации**

Обычный цикл обновления продуктов и их характеристик составляет несколько месяцев. Основную роль при разработке нового продукта играют подразделения маркетинга и продаж, а также, как правило, внешнее маркетинговое агентство. **VoC** предоставляет участникам процесса предоставить значительно больший объем полезной информации «глазами клиента», а также «глазами сотрудника», чем в обычном сценарии. **VoC** позволит проанализировать

- как чаще всего называют продукты банка клиенты (общепринятое название, а не юридически отточенная формулировка)
- на какие характеристики больше всего обращают внимание, а на какие характеристики не обращают внимание
- на основании каких суждений осуществляют выбор продукта или характеристики
- как позиционируют продукт и его характеристики по сравнению с конкурентами и брендами

Примеры использования VoC (госорганы)

1. **Повышение эффективности работы госорганов**

Взаимодействие органов власти и граждан

- Сбора обращений граждан с использованием разных каналов (горячая линия, сайты к, приемная и пр.)
- Взаимодействие напрямую с гражданами
- Взаимодействия со СМИ

Использование ТА позволяет во-первых, категоризировать обращения граждан по темам обращения, социальному статусу гражданина, и прочим критериям. Во-вторых, сгруппировать обращения граждан по значимости проблемы и резонансности. Наконец, собрать предложения граждан по разрешению проблем и улучшению условий. Поскольку ТА работает в режиме реального времени и может обрабатывать большие объемы обращений ежедневно, то ТА позволяет оперативно отслеживать влияние публикаций в СМИ и прочих соц. медиа на отзывы граждан. Кроме того, ТА позволяет оперативно выявлять зарождающиеся тенденции, находящие отражение в обращениях граждан, принимать необходимые меры, и оценивать эффект от принятых мер.

Компания VESOLV

О КОМПАНИИ

Группа компаний VESOLV– эксперт в области:

- Обработки неструктурированной информации
- Разработки интерфейсов нового поколения
- Business Intelligence и Information intelligence

Наша экспертиза

- Собственные алгоритмы обработки большого объема неструктурированных данных
- Системы поддержки принятия решений
- Внедрение Advanced BI
- Консалтинговые услуги

VESOLV

Технологии Yandex Data Factory для анализа отзывов

MatrixNet

Алгоритм машинного обучения, созданный Яндекс:

- Позволяет настроить модель в соответствии с целями вашего бизнеса
- Позволяет учитывать тысячи факторов
- Устойчив к переобучению

Математические модели

Аппаратно-программное решение на основе алгоритма машинного обучения

- Обучается на малом количестве данных (достаточно 1,5 тысяч отзывов)
- Качество модели увеличивается с ростом количества отзывов и при дополнительном обучении модели

Компьютерное зрение и распознавание изображений

- Оптическое распознавание текстов
- Извлечение фактов – названий банковских продуктов, адресов, дат, персоналий
- Последующий анализ и классификация

Поисковые технологии и обработка естественного языка

- Анализ тональности текста
- Распознавание эмоций
- Группировка похожих публикаций и отзывов
- Мониторинг изменений в поведении и отношении пользователя

Контакты

.

Илья Вигер, CEO

ilya.wieger@vesolv.com

+7 (926) 531-67-71

VESOLV