

Инструмент для анализа клиентских отзывов

АО «Тинькофф Банк»



01

Команда

Состав проектной
команды



Команда «2k:club»



Красных Кристина
Эдуардовна

- 10 класс
- Кировская область
- Участник
- Аналитик, МI разработчик



Кривцова Екатерина
Дмитриевна

- 10 класс
- Кировская область
- Капитан
- Дизайнер, МI разработчик

I

02

Проблема

Мотивация, задача



Мотивация



Tinkoff
Bank

В наши дни важно не только
получать обратную связь от
клиентов, но и корректно
обрабатывать эту информацию,
оценивать репутацию фирмы,
создавать актуальные доработки



Анализ отзывов и
клиентоориентированность
становятся одними из факторов
конкурентоспособности на рынке

87% покупателей читают онлайн-отзывы,
прежде чем совершить какую-либо покупку

Проблема



Ручная
обработка
ОТЗЫВОВ
неэффективна

- Большое количество времени
- Зависит от качества анализа человека
- Субъективность оценки



СИРИУС.ИИ

Задача

Разработать инструмент, который
анализирует данные
из открытых источников и помогает
сотрудникам
Тинькофф улучшить клиентский опыт

I

03

Анализ

Поиск наиболее
перспективной стратегии



Отзывы



Позитивные

- Удобство использования услуг и качество обслуживания

Некорректные

- Недостоверные
- Оскорбительные
- Спам

Негативные

- Проблемы обслуживания и качества продуктов

Предложения

- Идеи улучшения продукта, его оптимизации

Нейтральные

- Не содержащие ярко выраженных эмоциональных оценок

Платформы сбора отзывов



«Тинькофф Отзывы»

- Сервис компании Тинькофф, разработанный, чтобы избежать сгенерированных отзывов и их накрутки

Выберу.ру

- Сайт со встроенной аналитикой отзывов (ИИ Валера)
- 2700 отзывов о Тинькофф

Отзовик

- Отзывы об организациях в открытом доступе
 - 8200 отзывов о Тинькофф

Data sets

- Яндекс Карты
- 500 тыс. отзывов о российских организациях

И другие...

Разные компании = разный подход



- Яндекс использует свою нейросеть для анализа отзывов на товары в сервисе Яндекс Маркет
- Сначала алгоритм выбирает самые **качественные мнения**: подробные, с описанием опыта использования и **соответствующие множеству других критериев**. Затем нейросеть отмечает особенности товара, о которых пишут чаще всего. Для создания списка плюсов и минусов с помощью YandexGPT нужно **не менее 10 качественных отзывов** на товар
- Модель учитывает накрученные отзывы
- Возможность некорректной оценки

**MPSTATS**

- Это инструмент аналитики и управления продажами на маркетплейсах Wildberries, Ozon и Яндекс Маркет
- MpStats опрашивает каждый товар 1 раз в час от лица пользователя маркетплейса, формируя статистику по товару
- С помощью NLP сервис выделяет самые популярные жалобы и преимущества товара
- Дорогая подписка (около 10.000 руб/мес)
- Анализ только на указанных маркетплейсах



Сравнительная характеристика



Тип анализа	Сложность реализации	Точность	Область применения	Итог (с учетом поставленной задачи)
Ручной анализ	★	★	★ ★ ★	№3
Модели машинного обучения	★ ★	★ ★ ★	★	№1
Нейросеть	★ ★ ★ ★	★ ★	★ ★	№2

I

04

Решение

Описание системы анализа и взаимодействия работников с инструментом



Схема работы инструмента

I

Сбор отзывов

Выделение
ключевых слов

Сортировка по
типам

Отправление
данных
работникам

Для каждого этапа можно
найти подходящую модель
или нейросеть, либо
обучить самостоятельно

Обмен работник – программа, и
обратно, удобно осуществлять
через бота в сети Telegram

Преимущества чат – ботов Телеграм

Этапы передачи и получения готовых данных являются важной частью, поэтому среда обмена этого должна быть комфортной:

- работает 24/7;
- не требует специального ПО;
- высокая скорость передачи данных;
- конфиденциальность;
- простота реализации.



I

05

План

Шаги создания решения: от
сбора данных до создания
аналитической системы



Сбор и фильтрация информации



1. Веб-парсер

- a) Извлечение данных с веб-сайтов
- b) Удаление html разметки
- c) Преобразование отзывов в строки



2. Извлечение ключевых слов

- a) Удаление ненужных символов и служебных полей
- b) Удаление личных данных, приведение слов в начальную форму
- c) Извлечение ключевых слов



3. Группировка похожих отзывов



- а) Переход из словесного вида к векторно-числовому
- а) Разделение отзывов на тематические подгруппы (положительные, отрицательные, некорректные, предложения)

4. Отправка сообщений работникам для решения проблем и внедрения новых функций



- а) Определение паттернов в каждой группе
- а) Передача аналитических данных работникам через чат-бота Telegram

I

06

Отдельные шаги

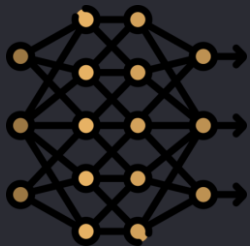
Методы, модели, полноценная
система



Использование ChatGPT и API OpenAI



Для реализации **выделения ключевых слов** потребуется какая-либо нейросеть, лучше всего с текстовыми данными на данный момент справляется **ChatGPT**.



OpenAI API - это мощная облачная платформа, предназначенная для предоставления разработчикам беспрепятственного **доступа** к предварительно **обученным** **моделям** **искусственного** **интеллекта**.



Universal Sentence Embedder



Universal Sentence Encoder (USE) — это инструмент в области обработки естественного языка, разработанный компанией Google.

Его основная функция — преобразование **текстовых данных** в **высокомерные векторы** (в embeddings), которые отражают **семантическое значение** предложений. В отличие от традиционных embeddings, которые представляют отдельные слова, USE генерирует embeddings **для целых предложений или коротких абзацев**.



Обучение модели кластеризации



Разделение отзывов на тематические группы



Обучение модели, используя агломеративную кластеризацию.

Обучать модель можно так же на **Data Set** от Яндекс Карт с отзывами, чтобы она была более приспособлена под конкретную задачу проекта.



Агломеративный метод

Агломеративная кластеризация — метод формирования кластеров, при котором каждый объект сначала находится в **отдельном кластере**, затем объекты группируются в значительно более крупные кластеры.

Полноценная система



01

Интеграция моделей

Интегрировать выбранную модель в рабочий процесс системы, настроить ее для работы в реальном времени или планируемом режиме

02

Тестирование системы

Протестировать систему с моделями на пилотной группе пользователей

03

Исправление ошибок

Выявить и исправить возможные ошибки, работа в цикле 2-3, пока не будут исправлены все возможные неисправности системы

04

Мониторинг и обновление

Следить за работой системы, мониторить показатели качества моделей, обновлять модели при необходимости с учетом новых данных



Доп.
задание



Сбор и обработка данных



ШАГ I

- Платформа: Тинькофф.Отзывы, а конкретно – отзывы на Тинькофф Топливо, т.к. их 300.000+
- Способ: request.post, оформление данных в формат json (текст и рейтинг)



ТИНЬКОФФ

Как результат
получаем «чистые»
данные из отзыва и
их рейтинг

ШАГ II

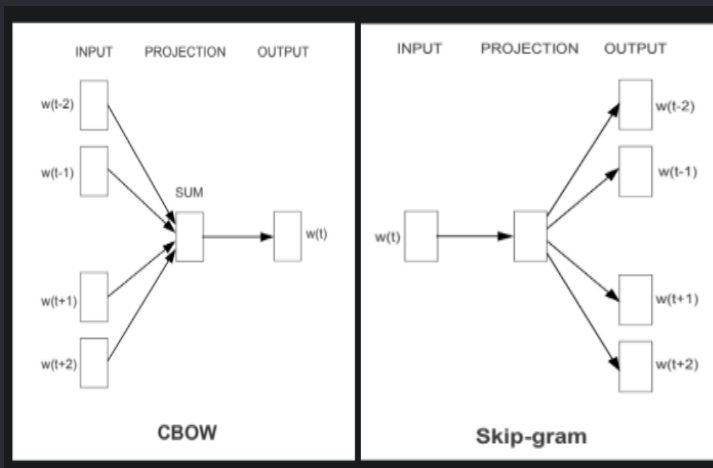
- Токенизация, удаление стоп-слов, стемминг с помощью библиотеки NLTK



Векторизация отзывов

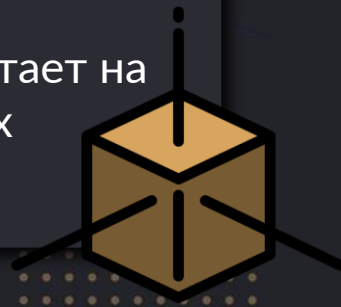


Способ: использование методов NLP, модель Word2Vec



На выходе Word2Vec получают координаты векторов, соответствующих текстовым данным.

Технология работает на базе искусственных нейронных сетей.





- Косинусы полученные векторов можно сравнивать и находить паттерны, тем самым обнаруживая сильные и слабые места компании
- Выявление аномалий в эмоциональной окраске отзыва и его рейтинге (например, «Всё понравилось!» с рейтингом 3)
- Определение зависимости даты публикации отзыва от его эмоциональной окраски или группы
- ... и многое другое

