

# Инструмент для анализа клиентских отзывов

Проект в рамках всероссийской научно-технологической программы по решению проектных задач в области искусственного интеллекта и смежных дисциплин «Сириус.ИИ»



# Состав проектной команды

1. Шалин Егор Вячеславович, 9 класс, Рязанская область – капитан / кодер
2. Базин Святослав Дмитриевич, 9 класс, Рязанская область – кодер
3. Байбунова Анастасия Сергеевна, 9 класс, Рязанская область – дизайнер презентации / аналитик
4. Шишов Иван Игоревич, 9 класс, Рязанская область – кодер



# Формулировка проблемы, которую необходимо решить

В современном мире мнение клиентов играет ключевую роль в успехе любого бизнеса. Для Тинькофф банка это особенно важно, так как конкуренция на рынке финансовых услуг постоянно усиливается.

Однако, собрать и проанализировать огромное количество отзывов от клиентов вручную оказывается непросто. Отзывы поступают с различных источников: интернет-ресурсы, социальные сети, форумы. Кроме того, тексты отзывов требуют тщательной обработки, чтобы извлечь из них полезную информацию о предпочтениях и потребностях клиентов.



# Формулировка проблемы, которую необходимо решить

Мотивацией для решения этой проблемы является стремление улучшить качество обслуживания клиентов банка, повысить удовлетворенность и лояльность пользователей. Понимание того, что нравится клиентам и что вызывает у них недовольство, позволит банку принимать обоснованные решения по улучшению продуктов и услуг. Адаптация под требования клиентов поможет не только удерживать существующих клиентов, но и привлекать новых, что в конечном итоге способствует росту бизнеса и увеличению прибыли.



# Анализ области

## Виды отзывов клиентов о продуктах Тинькофф:

Клиенты Тинькофф банка оставляют разнообразные отзывы о продуктах и услугах банка. Это могут быть отзывы о работе мобильного приложения, интернет-банке, кредитных продуктах, качестве обслуживания и т.д. Отзывы могут содержать как положительные, так и отрицательные комментарии, а также предложения по улучшению сервиса.



# Анализ области

## Способы сбора отзывов:

- 1. Мониторинг в социальных сетях:** Отслеживание упоминаний банка и его продуктов в социальных сетях. Сбор отзывов по хэштегам и упоминаниям бренда.
- 2. Отзывы на специализированных платформах:** Использование отзывов на сайтах типа Яндекс.Отзывы и других специализированных ресурсах, где пользователи оставляют свои отзывы о банке и его продуктах.
- 3. Обратная связь через мобильное приложение:** Анализ отзывов и оценок, оставленных клиентами напрямую в мобильном приложении банка.



# Анализ области

## Анализ отзывов и выводы:

- 1. Выявление основных тем:** Анализ текстов отзывов с помощью методов обработки естественного языка для выявления основных тем, которые обсуждают клиенты (например, качество обслуживания, удобство интерфейса, проблемы с техническими аспектами и т. д.).
- 2. Определение тональности:** Анализ эмоциональной окраски текстов для определения соотношения положительных и отрицательных отзывов.
- 3. Выявление трендов:** Идентификация повторяющихся проблем или пожеланий клиентов для выявления наиболее актуальных аспектов, требующих внимания и улучшений.



# Анализ области

**Сравнение подходов других компаний:**

## 1. Внутренние разработки компаний-конкурентов:

**Подход:** Компании разрабатывают собственные алгоритмы анализа отзывов с использованием методов машинного обучения, таких как модели Word2Vec и алгоритмы кластеризации текста.

**Преимущества:** Полный контроль над разработкой и настройкой системы под специфику компании.

**Недостатки:** Требуется значительных временных и финансовых ресурсов на разработку и поддержку.

## 2. Использование сторонних аналитических платформ:

**Подход:** Компании обращаются к сторонним сервисам, специализирующимся на анализе отзывов клиентов.

**Преимущества:** Экономически эффективное решение, не требующее больших вложений в разработку собственных алгоритмов.

**Недостатки:** Ограниченная гибкость и возможности адаптации к специфике компании.





# Анализ области

## Выбор и обоснование:

С учетом стремления Тинькофф банка к инновациям и высокой ориентации на потребности клиентов, наиболее перспективным вариантом кажется разработка собственной системы анализа отзывов клиентов. Это позволит банку максимально адаптировать анализ к специфике своего бизнеса и клиентской аудитории, обеспечивая эффективное и точное выявление проблем и потребностей клиентов.



# Описание предлагаемого решения

## **Ввод данных:**

Система будет автоматически собирать данные из различных источников. Данные будут обрабатываться для исключения дубликатов и шума, а также для приведения к единому формату.

## **Анализ данных:**

После предварительной обработки данные будут подвергаться анализу с использованием методов машинного обучения и алгоритмов обработки текста.

## **Создание отчетов и дашбордов:**

На основе результатов анализа будут формироваться отчеты и дашборды, представляющие информацию о наиболее актуальных проблемах, трендах и потребностях клиентов.



# Описание предлагаемого решения

## Преимущества системы:

**Комплексный анализ:** Система позволит проводить комплексный анализ отзывов клиентов, выявляя как позитивные, так и негативные аспекты работы банка.

**Быстрая обработка данных:** Автоматизированный процесс сбора и обработки данных позволит быстро реагировать на изменения в отзывах клиентов и анализировать их в реальном времени.

**Выявление ключевых трендов:** Система поможет выявить ключевые тренды и проблемы, что позволит банку принимать обоснованные решения по их решению.



# Описание предлагаемого решения

## Работа специалистов с системой:

### 1. Мониторинг и анализ результатов:

Специалисты будут регулярно мониторить дашборды и отчеты, создаваемые системой, для получения актуальной информации о текущем состоянии и динамике отзывов клиентов. Они будут анализировать полученные данные, выявлять ключевые тренды, проблемы и паттерны поведения клиентов.

### 2. Разработка стратегий:

На основе результатов анализа специалисты будут разрабатывать рекомендации и стратегии по улучшению сервиса и продуктов банка. Они будут участвовать в процессе принятия решений на основе данных системы, определяя приоритеты и направления деятельности.



# Высокоуровневый план реализации решения

## 1. Сбор данных:

Разработка автоматизированной системы сбора данных из различных источников, таких как социальные сети, форумы, сайты отзывов и официальные каналы коммуникации банка.

Использование API и web-скрапинга для сбора данных в реальном времени.

## 2. Предобработка данных:

Проведение предварительной обработки текстов отзывов, включая удаление стоп-слов, символов пунктуации, а также лемматизацию и стемминг.

Преобразование текстов в векторное представление с использованием моделей Word2Vec для дальнейшего анализа.



# Высокоуровневый план реализации решения

## 3. Анализ и классификация:

Разработка алгоритмов машинного обучения для определения тональности отзывов, выявления основных тем и проблем, а также кластеризации отзывов для выявления общих трендов.

Использование моделей классификации, кластеризации и алгоритмов обработки естественного языка для анализа текстов.

## 4. Визуализация и отчётность:

Разработка интерактивных дашбордов и отчетов для визуализации результатов анализа.

Предоставление интуитивно понятного интерфейса для работы с данными и принятия управленческих решений.



# Описание отдельных шагов плана

## 1. Сбор данных:

**Методы:** Использование API социальных сетей, веб-скрапинг с общедоступных сайтов, сбор отзывов с официальных каналов коммуникации банка.

**Модели:** Не применяются.

**Источники данных:** Социальные сети, форумы, сайты отзывов, официальные сайт и приложение.

**Превращение в систему:** Данные будут собираться с помощью скриптов и программ, автоматизирующих процесс сбора данных из различных источников.



# Описание отдельных шагов плана

## 2. Предобработка данных:

**Методы:** Токенизация, удаление стоп-слов, стемминг, лемматизация.

**Модели:** Не применяются.

**Источники данных:** Данные, собранные на предыдущем этапе.

**Превращение в систему:** Предобработка будет осуществляться с использованием скриптов на языке программирования Python и библиотек `rumorphy3` и `nltk` для обработки текста. (так как они достаточно удобны и быстры)





# Описание отдельных шагов плана

## 3. Анализ данных:

**Методы:** Машинное обучение, алгоритмы обработки естественного языка.

**Модели:** Модель Word2Vec (Word2Vec хорошо справляется с большими объемами данных и может быстро обучаться на больших корпусах текстов) для векторного представления слов, модель Naïve Bayes (она может быть эффективна, когда требуется быстрое решение на больших объемах данных без необходимости в сложных настройках) для классификации тональности отзывов, модель DBSCAN (так как она не требует указания количества кластеров заранее и может обнаруживать кластеры любой формы и размера) для кластеризации.

**Источники данных:** Подготовленные текстовые данные из предыдущего шага.

**Превращение в систему:** Разработка и обучение моделей на основе предоставленных данных, интеграция в систему для автоматического анализа.



# Описание отдельных шагов плана

## 4. Создание отчетов и дашбордов:

**Методы:** Визуализация данных с помощью библиотеки Plotly.

**Модели:** Не применяются.

**Источники данных:** Результаты анализа данных.

**Превращение в систему:** Созданные отчеты и дашборды будут доступны для сотрудников через интерфейс портала, размещённого в интернете.



## 2 этап

**Ссылка на репозиторий github с решением:**

[https://github.com/egorshalin08/siriusii\\_tinkoff\\_Shalin\\_EV](https://github.com/egorshalin08/siriusii_tinkoff_Shalin_EV)

