

### Актуальность темы:

Онлайн-обзоры имеют большое значение для поддержки принятия потребителями решений о покупке. В настоящее время существует огромное количество платформ, которые позволяют пользователям размещать отзывы о различных товарах, услуги или рассказать о своем опыте работы в той или иной компании; эти обзоры отражают потребительский опыт и снижают неопределенность при онлайн-покупках [1]. Однако часто на популярные товары оставляется огромное количество отзывов и появляется задача ранжирования отзывов по полезности [2].

По результатам исследования Local Consumer Review Survey 2022 все больше и больше потребителей читают отзывы. В 2021 году 77% "всегда" или "регулярно" читали их перед покупкой товара (по сравнению с 60% в 2020 году), в 2022 году эта цифра увеличилась до 98%. 57% утверждают, что они были бы не склонны приобретать товары у продавцов, которые не реагируют на отзывы.

Таким образом анализ отзывов может быть полезен не только потенциальным потребителям, но и владельцам компании для контроля качества своей продукции. Особенно это может быть актуально для компаний, специализирующихся на здравоохранении, техническом обслуживании, выпуску уходовой продукции и косметических средств.

Для ручного анализа отзывов и настроений клиента к различным аспектам товара, услуги или компании в целом может потребоваться большое количество времени, в некоторых случаях даже месяцы. При анализе отзывов могут возникать различные проблемы - отзывы могут быть неоднозначными или бесполезными, к такой информации также необходим особый подход.

В области анализа тональности текста сейчас актуально использование трансформеров для подробного анализа отзывов и др. текстов, т. е. не просто предоставление рейтинга всего текста, а полный отчет об отношении к различным объектам, например отношение пользователя к самому фильму, к режиссеру или актерам. Это должно позволить строить более точные рекомендательные системы и облегчить контроль качества товаров.

### Объект исследования:

Методы на основе нейронных сетей для анализа тональности корпуса текстов.

### Предмет исследования:

Применение трансформеров для эффективного анализа эмоционально окраски русскоязычных отзывов с объектно-ориентированным анализом тональности текста.

### Цель исследования:

Разработать сервис для подробного анализа отзывов пользователей с использованием трансформеров и их ранжирование по степени полезности. Повысить точность предсказания при работе с русскоязычными тестами. Создать и разметить датасет, основанный на данных, взятых с таких популярных маркетплейсов, как Ozon, Wildberries и AliExpress, который можно будет использовать для выявления наилучшей модели для аспектного анализа отзывов на русском языке.

Основная проблема — это отсутствие инструментов для качественного анализа русскоязычных отзывов с популярных площадок, нет универсального корпуса текстов. В исследованных статьях были использованы различные подходы — модели (трансформеры, сверточные нейронные сети, рекуррентные нейронные сети, гибридные и т. д.) и методы (использование тональных словарей с применением нечеткой логики). В зависимости от датасета результаты анализа могут быть разными, но их точность в среднем составляет <70%, таким образом необходимо создать единый словарь отзывов и подобрать подходящую модель для достижения точности выше 70%.

Для достижения поставленной цели будут поставлены и решены следующие задачи:

1. Провести аналитический обзор предметной области: анализ существующих исследований и архитектур нейронных сетей для анализа тональности текста.
2. Подготовить наборы данных для тренировки, валидации и тестирования нейронной сети, которые можно использовать для сравнения с созданным датасетом, изучить их характеристики.
3. Произвести парсинг необходимых сайтов, извлечение, обработку и анализ полученных данных, на основе которых будет создаваться датасет.
4. Подход к решению поставленной проблемы - разработать архитектуру нейронной сети и протестировать ее на собранных данных.
5. Проанализировать полученные результаты.

Гипотеза исследования:

С помощью разработанной архитектуры нейронной сети (трансформер) и собранного набора данных, охватывающих большую часть отзывов на товары, которые люди оставляют на маркетплейсах возможно создать удобный сервис для подробного анализа отзывов пользователей как с выбранных площадок, так и с других сайтов и интернет-магазинов. Данная архитектура подходит не только для эффективного анализа эмоционально окраски отзывов и семантического анализа, но и объектно-ориентированного анализа тональности текста, что позволяет извлекать из отзывов не только общее отношение пользователя и судить о положительности или отрицательности отзыва, но и определить, чем конкретно он был доволен или не доволен, какие эмоции испытывал по отношению к различным аспектам товара или услуги (для этого необходимо добиться точности оценки более 70%). Также созданный корпус отзывов может считаться универсальным и подходить для анализа отзывов с различных интернет-магазинов.

В качестве методов исследования используется анализ статей различных авторов и сравнение подходов. Для выявления частных проблем существующих исследований и усовершенствования процесса обработки текста используется метод дедукции – переход от общих концепций к созданию конкретной архитектуры нейронной сети. В процессе разработки трансформера и подбора гиперпараметров применялись экспериментальные методы. Для классификации отзывов использовались статистические методы.