

АНАЛИЗ ПОЛЬЗОВАТЕЛЬСКИХ ОТЗЫВОВ МЕТОДАМИ NLP КАК ИНСТРУМЕНТ ВЫЯВЛЕНИЯ ФАКТОРОВ ПОТРЕБИТЕЛЬСКОГО ВЫБОРА НА МАРКЕТПЛЕЙСЕ

Васютинская А.А.¹

Научный руководитель – преподаватель Канунникова К.И.¹

¹Университет ИТМО

avasyutinskaya@niuitmo.ru

Введение

В условиях стремительной цифровизации экономики пользовательский контент (UGC), в особенности текстовые отзывы, стал ключевым фактором, формирующим репутацию бренда и влияющим на принятие потребительских решений [1, 3]. Потенциальные покупатели часто ориентируются на мнение других пользователей, что делает отзывы эффективным инструментом маркетинга [1, 3]. Однако для бизнеса, представленного на маркетплейсах, массивы текстовых данных остаются недоиспользованным ресурсом. На практике анализ сводится к отслеживанию рейтинга, который, как показывают исследования [4], зачастую не отражает полной картины восприятия товара, так как сводит многообразный пользовательский опыт к одному числовому значению, не раскрывая, какие именно характеристики продукта вызвали удовлетворение или недовольство.

Цель исследования заключается в проведении анализа пользовательских отзывов, используя методы NLP (natural language processing) для выявления ключевых аспектов восприятия товара, их тональности и влияния на общий рейтинг.

Основная часть

Работа продемонстрирует применимость подхода на примере отзывов о нескольких моделях проводных механических клавиатур средней ценовой категории с Яндекс.Маркета (более 500 отзывов на каждую модель). Выбор маркетплейса обусловлен наличием значительного объема развернутых отзывов, формирующих достаточную основу для проведения тематического и тонального анализа. Такой подход ориентирован на конкретную товарную категорию, что позволяет минимизировать искажающее влияние цены и сосредоточиться на функциональных характеристиках.

Исследование предполагает реализацию анализа в среде Python с использованием библиотек обработки естественного языка и включает следующие этапы:

1. Предварительная обработка данных. Проводится очистка текстов отзывов, токенизация и нормализация. Данный этап является стандартной процедурой подготовки данных к тематическому и тональному анализу [2, 5].

2. Выделение аспектов. Для автоматического выявления ключевых аспектов обсуждения товаров планируется применение модели тематического моделирования, основанной на трансформерных эмбедингах. Подход к извлечению аспектов из пользовательских отзывов опирается на исследования в области аспектно-ориентированного анализа (ABSA) [2, 4].

3. Анализ тональности. Для каждого отзыва и выделенного аспекта определяется эмоциональная окраска. Планируется использование предобученной модели RuBERT, адаптированной для русского языка. Контекстная модель позволяет учитывать смысл высказывания в целом, что особенно важно при анализе коротких и неформальных текстов. Применение методов sentiment-анализа широко рассматривается как инструмент поддержки маркетинговых решений и анализа пользовательского поведения [3].

4. Расчет метрик по аспектам. По каждому аспекту рассчитываются частота его упоминаний, средняя тональность, доля позитивных отзывов и средний рейтинг отзывов, в которых аспект присутствует. Эти метрики позволяют оценить значимость отдельных характеристик товара для потребителей и выявить аспекты, ассоциированные с более высокими или низкими оценками.

5. Анализ связи аспектов и рейтинга. С помощью линейной регрессии оценивается, как аспекты и их тональность статистически связаны с изменением общего рейтинга товара. Подобный подход применялся для анализа влияния содержания отзывов на рыночные показатели и продажи [4]. Предполагается выявить аспекты, оказывающие наибольшее влияние на итоговую оценку товара.

6. Сравнительный анализ моделей. На завершающем этапе проводится сопоставление выбранных моделей клавиатур по структуре выявленных аспектов и их тональной характеристике. Это позволит определить различия внутри одного ценового сегмента и выявить факторы, формирующие преимущества отдельных товаров.

Выводы

Предлагаемое исследование направлено на демонстрацию того, что анализ текстовых отзывов с использованием методов NLP позволяет получить более детализированное представление о восприятии товара по сравнению с использованием только среднего рейтинга.

Ожидается, что структурирование отзывов по аспектам и расчет соответствующих метрик позволит выявить характеристики, наиболее значимые для потребителей, а также определить аспекты, статистически связанные со снижением или повышением оценки.

Практическое значение работы заключается в возможности применения предложенного подхода для поддержки управленческих решений на маркетплейсах: корректировки характеристик продукта, позиционирования, описания товара и коммуникации с потребителями. Таким образом, текстовая аналитика рассматривается как инструмент повышения конкурентоспособности и оптимизации продуктовых стратегий.

Литература

1. Бакеева Г. С. Роль пользовательских отзывов и UGC в формировании репутации бренда // Вестник науки. – 2025. – № 2 (83). – С. 28-38.
2. Al-Ghuribi S. M., Noah S. A. M., Tiun S. Unsupervised Semantic Approach of Aspect-Based Sentiment Analysis for Large-Scale User Reviews // IEEE Access. – 2020. – Vol. 8. – P. 218592–218613.
3. Kauffmann E. et al. A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making // Industrial Marketing Management. – 2020. – Vol. 90. – P. 523-537.
4. Li X., Wu C., Mai F. The effect of online reviews on product sales: A joint sentiment-topic analysis // Information & Management. – 2019. – Vol. 56, No. 2. – P. 172-184.
5. Mao Y., Liu Q., Zhang Y. Sentiment analysis methods, applications, and challenges: A systematic literature review // Journal of King Saud University - Computer and Information Sciences. – 2024. – Vol. 36, No. 4. – P. 1-16.