В реалиях непрекращающейся **цифровой трансформации** в геометрической прогрессии возрастают требования к качеству клиентского обслуживания [1]. Так анализ тональности текстов позволяет оперативно оценивать эмоциональное состояние клиента, понимать, доволен ли он услугой или сталкивается с проблемой [4]. Персонализированный подход через анализ эмоционального состояния дает возможность предложить более подходящее решение, предотвращая негативные последствия и укрепляя долгосрочные отношения с клиентом, способствуя предотвращению или сокращению оттока пользователей [3].

Также применение **анализа тональности речи** открывает новые возможности для оптимизации маркетинговых стратегий [4]. К примеру, мониторинг оставленных клиентами отзывов позволяет сформировать целостное представление о товаре, что дает возможность подбирать поставщиков на основе рейтинга их продуктов. Оптимизируя расходы компании, появляется возможность уделять больше времени и ресурсов на решение сложных задач, требующих личного контроля и вмешательства в бизнес-процессы.

**Актуальность** темы исследования обусловлена растущей значимостью пользовательских отзывов в цифровой экономике. Современные потребители все чаще принимают решения о покупке, ориентируясь на отзывы, а их **длина** и **тональность** могут существенно влиять на восприятие бренда и уровень доверия к нему. Анализ этого взаимодействия помогает глубже понять особенности поведения покупателей.

**Предметом** настоящего исследования выступает влияние длины отзыва на его эмоциональную окраску и характер оценочных суждений покупателей. **Целью** предлагаемой работы ​​является анализ и установка взаимосвязи объема пользовательского отзыва и его эмоциональной окраски при помощи методов машинного обучения. **Материалом исследования** являются пользовательские отзывы, оставленные покупателями в мобильном приложении и на сайте сети супермаркетов ВкусВилл в категории “Фрукты: лимоны”.

 Исследование основывается на **теоретических положениях** о взаимосвязи длины текста и характера его тональности, а также на методах автоматизированного анализа пользовательских отзывов. В области анализа тональности текстов (Sentiment Analysis) **настоящая работа выдвигает гипотезу** о том, что длина отзыва может быть индикатором его эмоциональной окраски.

В контексте потребительского поведения существует предположение, что короткие отзывы нередко являются реакцией на сильные эмоции, что делает их чаще негативными, так как пользователи быстрее оставляют комментарии в случае неудовлетворенности товаром или сервисом [2]. В то же время длинные отзывы, как правило, требуют от автора большего вовлечения, что может свидетельствовать о его положительном опыте, побуждающем подробно изложить преимущества продукта. Однако противоположный сценарий также возможен: пользователи могут писать развернутые негативные отзывы, если столкнулись с серьезными проблемами, требующими детального объяснения.

Современные методы машинного обучения и обработки естественного языка (NLP - Natural Language Processing) позволяют анализировать закономерности между длиной текста и его тональностью [5]. Автоматизированный анализ отзывов предполагает выявление закономерностей, определяющих, в какую сторону смещается эмоциональная окраска в зависимости от объема пользовательского комментария.

 В данной работе было проведено **эмпирическое исследование** с использованием как качественных, так и количественных методологических подходов. На первом этапе был произведен сбор данных. Были выгружены пользовальские отзывы с сайта сети супермаркетов ВкусВилл в категории “Фрукты: лимоны” в количестве 5000 единиц. Далее библиотека *pandas* используется для формирования *DataFrame* со столбцом review. С помощью модуля *re* (регулярные выражения) выполняется предобработка: текст переводится в нижний регистр, удаляются не-буквенные символы и лишние пробелы. Затем при помощи *nltk* мы загружаем токенизатор для русского языка, а библиотека *transformers* через функцию pipeline позволяет обратиться к модели *blanchefort/rubert-base-cased-sentiment*, которая автоматически определяет тональность (positive, negative или neutral). Результаты классификации записываются в столбец sentiment. Далее рассчитывается длина каждого отзыва (в словах) при использовании метода *split()*. После этого при помощи *pandas* вычисляются средние длины положительных и отрицательных отзывов и итоги визуализируются с помощью *matplotlib*.

Таким образом, полученные **статистически значимые результаты** (проверка на двухвыборочном *t-тесте* из пакета *scipy.stats*) подтвердили гипотезу о том, что длина отзыва связана с его тональностью. Так средняя длина положительных отзывов равна 6.63 в периоде, в то время как средняя длина отрицательных отзывов получилась равной 10.25. Это открывает возможности для дальнейшего изучения данного феномена и его применения в управлении репутацией брендов. Если пользователи сталкиваются с серьезными проблемами, они могут оставить развернутый негативный отзыв, детально описав суть возникших сложностей.

 Для ритейлеров, таких как ВкусВилл, изучение тональности отзывов является важным инструментом для **оценки уровня удовлетворенности клиентов** и выявления потенциальных проблем. Понимание того, как длина отзыва связана с его эмоциональной окраской, позволяет разрабатывать более эффективные стратегии управления репутацией и повышать лояльность потребителей.

*Литература*

*Алиева А. Х.* Особенности использования персонализированного маркетинга в условиях цифровизации // Журнал прикладных исследований. – 2021. – Т. 8. – № 6. – С. 736–741.

*Двойникова А. А., Карпов А. А.* Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // Информационно-управляющие системы. – 2020. – №4 (107). – С. 20–30.

*Karagthala J. J., Shah V.* Analyzing the recent advancements for Speech Recognition using Machine Learning: A Systematic Literature Analysis // Journal of Electrical Systems. – 2024. – Т. 20. – №10s. – С. 1425–1447.

*Kumar V. S. et al.* Multimodal sentiment analysis using speech signals with machine learning techniques // 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). IEEE, 2022. С. 1–8.

*Kumari R., Srivastava S. K.* Machine learning: A review on binary classification // International Journal of Computer Applications. – 2017. – Т. 160. – №7.