

УДК 004.89:339.13

141. ПРОГРАММНОЕ СРЕДСТВО РЕАЛИЗАЦИИ РЕШЕНИЯ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА КОНКУРЕНТНОЙ СРЕДЫ НА РЫНКЕ БЫТОВОЙ ТЕХНИКИ И ЭЛЕКТРОНИКИ

Янковская Я.А., студент гр. 272302

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

*Сторожев Д.А. – магистр экон. наук, старший преподаватель
кафедры ЭИ*

Аннотация. В данной статье рассматривается разработка веб-ориентированного программного средства для автоматизированного анализа конкурентной среды на рынке бытовой техники и электроники. Описаны ключевые этапы работы системы, включая интеллектуальный сбор данных, оценку тональности клиентских отзывов с помощью нейросетевых архитектур семейства Transformer, плотностную кластеризацию текстов и прогнозирование динамики цен. Представлена распределенная микросервисная архитектура приложения, реализованного с использованием фреймворков Django и FastAPI, а также даталогическая модель базы данных PostgreSQL. Подробно рассмотрены функциональные возможности системы и логика взаимодействия её компонентов для поддержки принятия управленческих решений.

Ключевые слова. Анализ конкурентной среды, машинное обучение, парсинг данных, обработка естественного языка, прогнозирование, микросервисная архитектура, база данных, Python, Django, FastAPI, PostgreSQL.

В условиях стремительного развития цифровой экономики и возрастающей конкуренции компании розничной торговли сталкиваются с необходимостью оперативно реагировать на изменения внешней среды. Объем информации, доступной о конкурентах, постоянно растет: новостные публикации, отзывы пользователей, частое обновление цен, изменения продуктовых линеек и маркетинговая активность формируют сложную, динамичную и слабо структурированную информационную среду.

Традиционно для оценки конкурентной среды применялись экспертные методы и ручной мониторинг, включающие периодический обход сайтов конкурентов, ведение сводных электронных таблиц и классический SWOT-анализ на основе агрегированных статистических данных [1]. Оценка репутации и потребительской лояльности базировалась на выборочных социологических опросах, фокус-группах или ручном чтении отзывов аналитиками. Однако специфика предметной области – рынка бытовой техники и электроники – существенно усугубляет проблему информационной перегруженности и делает подобные подходы нежизнеспособными.

В условиях гиперконкуренции цены на продукцию на транснациональных маркетплейсах подвергаются корректировкам алгоритмами динамического ценообразования по несколько раз в сутки. Одновременно с этим, текстовые отзывы на площадках омниканальных сетей генерируются в огромных объемах, содержа тысячи скрытых инсайтов о слабых местах конкурентов. Ручной сбор данных характеризуется высокой ресурсоемкостью, низкой оперативностью и сильным влиянием человеческого фактора (ошибки ввода, субъективность в оценке тональности текста).

Актуальность темы обусловлена необходимостью повышения качества и скорости принятия решений путем перехода от устаревших экспертных оценок к автоматизированным data-driven подходам. Применение современных алгоритмов машинного обучения, в частности определение тональности публикаций с помощью архитектуры трансформеров [2], векторизация текстов [3], автоматическое выявление тематических кластеров [4], а также оценка динамики цен и построение прогнозов [5], позволяет бизнесу своевременно реагировать на рыночные угрозы. Интеграция ML-моделей с веб-архитектурой обеспечивает масштабируемость и оптимизацию рутинных аналитических процессов.

Объектом исследования являются информационные и бизнес-процессы сбора, структурирования и анализа данных о конкурентной среде на рынке бытовой техники и электроники. Предметом исследования выступают методы, алгоритмы и программные средства машинного обучения, применяемые для парсинга данных, анализа естественного языка, кластеризации текстов и прогнозирования рыночных показателей.

Таким образом, целью работы является разработка веб-ориентированного программного средства для совершенствования процесса анализа конкурентной среды с использованием методов машинного обучения, обеспечивающего эффективный сбор, обработку, кластеризацию и прогнозирование рыночных данных.

Для достижения поставленной цели в работе решаются следующие задачи: анализ предметной области и существующих подходов к оценке конкурентной среды на рынке бытовой техники и электроники, разработка архитектура веб-ориентированного программного средства с использованием микросервисного подхода, реализация методов автоматизированного сбора и предварительной обработки данных о конкурентах, включая парсинг веб-ресурсов, разработка и интеграция алгоритмов машинного обучения для анализа текстовой информации (оценка тональности, кластеризация) и прогнозирования динамики цен, проектирование реляционной базы данных для хранения и обработки структурированных и неструктурированных данных, интеграция компонентов системы в единый программный комплекс для поддержки принятия управленческих решений, проведение проверки работоспособности и оценка применимости программного средства в задачах анализа конкурентной среды.

Научная новизна работы заключается в интеграции Transformer-моделей, кластеризации и прогнозирования в единую систему, разработке микросервисной архитектуры для конкурентного анализа, формализации процесса анализа конкурентной среды как ML-задачи.

Практическая значимость проекта заключается в создании готового инструментального комплекса, который может применяться предприятиями электронной коммерции для регулярного мониторинга рынка. Использование данной системы обеспечивает формирование аналитических дашбордов, ранжирование угроз на основе тональности отзывов и прогнозирование ключевых метрик, что минимизирует влияние человеческого фактора и сокращает время на анализ конкурентной среды.

Разрабатываемое программное средство представляет собой комплексную веб-ориентированную систему анализа конкурентной среды. Назначение системы заключается в повышении качества и эффективности процесса оценки рынка с применением алгоритмов машинного обучения для автоматического сбора, классификации, кластеризации и прогнозирования параметров рыночной активности.

Для реализации математического аппарата и алгоритмов интеллектуальной обработки применяются передовые инструменты экосистемы Python:

- scikit-learn – для реализации классических моделей машинного обучения и базовой кластеризации;
- PyTorch – в качестве основного фреймворка для построения и инференса глубоких нейронных сетей и архитектур семейства Transformer;
- pandas и numpy – для векторизации, быстрой предобработки и матричных вычислений при анализе больших массивов данных.

В качестве основного веб-каркаса системы используется фреймворк Django, обеспечивающий маршрутизацию и контроль доступа, в то время как ресурсоемкие задачи скрапинга и ML-вычислений инкапсулированы в независимые асинхронные микросервисы на базе FastAPI.

Для формализации функциональных требований и описания логики взаимодействия конечных пользователей с программным комплексом была спроектирована диаграмма вариантов использования, представленная на рисунке 1. Данная диаграмма отражает внешнее поведение системы и фиксирует зоны ответственности каждой категории пользователей.

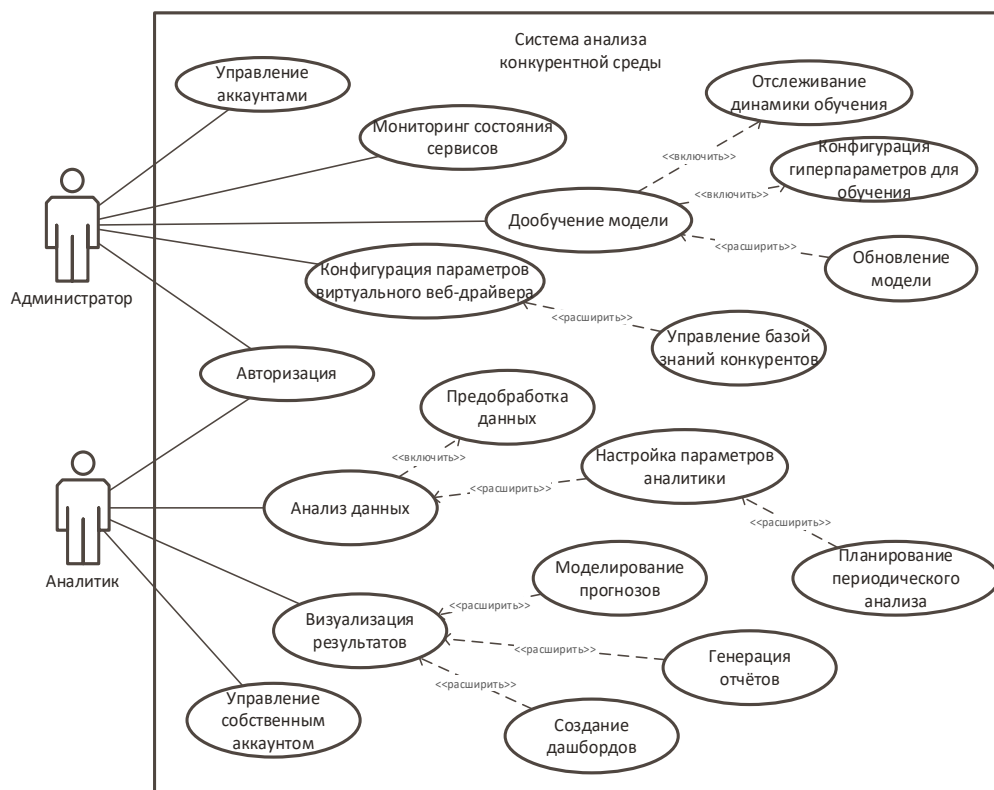


Рисунок 1 – Диаграмма вариантов использования программного средства

В системе реализована ролевая модель управления доступом, предусматривающая два типа пользователей: администратора и аналитика, работающих в защищенном контуре с обязательной авторизацией.

Администратор обеспечивает техническое сопровождение системы, включая управление учетными записями, мониторинг состояния сервисов, настройку параметров сбора данных и управление контуром машинного обучения (дообучение и обновление моделей).

Аналитик является основным пользователем системы и выполняет задачи анализа данных: инициирует обработку информации, настраивает параметры аналитики и осуществляет визуализацию результатов в виде отчетов и интерактивных дашбордов для поддержки принятия решений.

Таким образом, определенные функциональные требования и сценарии использования образуют надежный фундамент для последующей программной реализации, обеспечивая разделение бизнес-логики и технических настроек платформы.

В качестве системы управления базами данных выбрана PostgreSQL, обеспечивающая надежное хранение данных за счет поддержки ACID-транзакций, ссылочной целостности и масштабируемости. Данные, включая результаты парсинга, интеллектуальной обработки и системные логи, организованы в реляционной структуре, нормализованной до третьей нормальной формы.

Разработана даталогическая модель (рисунок 2), включающая таблицы с определением ключей и связей. Центральным элементом является таблица компаний-конкурентов, связанная с каталогом товаров и историей изменения цен (отношения «один-ко-многим»), что позволяет формировать временные ряды.

Текстовые данные хранятся в таблицах отзывов и упоминаний, а результаты машинного обучения – в специализированных таблицах прогнозов и аналитики. Административный контур объединяет пользователей, настройки и отчеты.

Для предотвращения дублирования реализованы ограничения уникальности, обеспечивающие корректное обновление данных при автоматическом сборе. Разработанная структура является масштабируемой основой для интеграции с модулями анализа и прогнозирования.

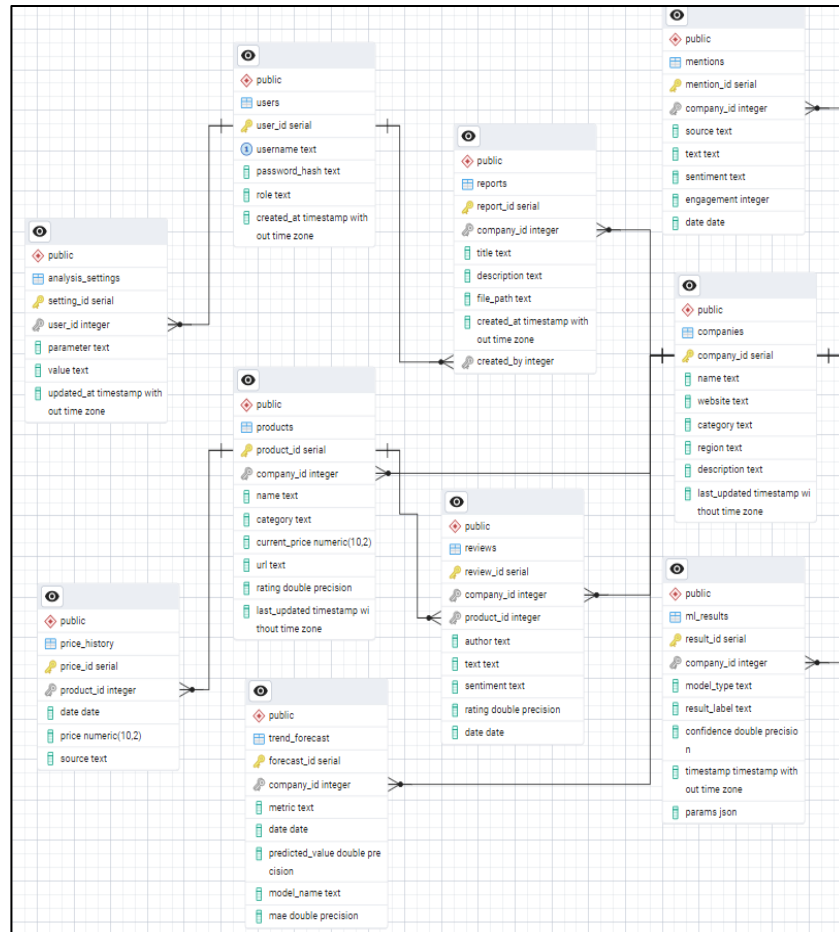


Рисунок 2 – Даталогическая модель базы данных

Разработанная система имеет распределенную многокомпонентную архитектуру, физическая структура которой представлена на диаграмме развертывания (рисунок 3).

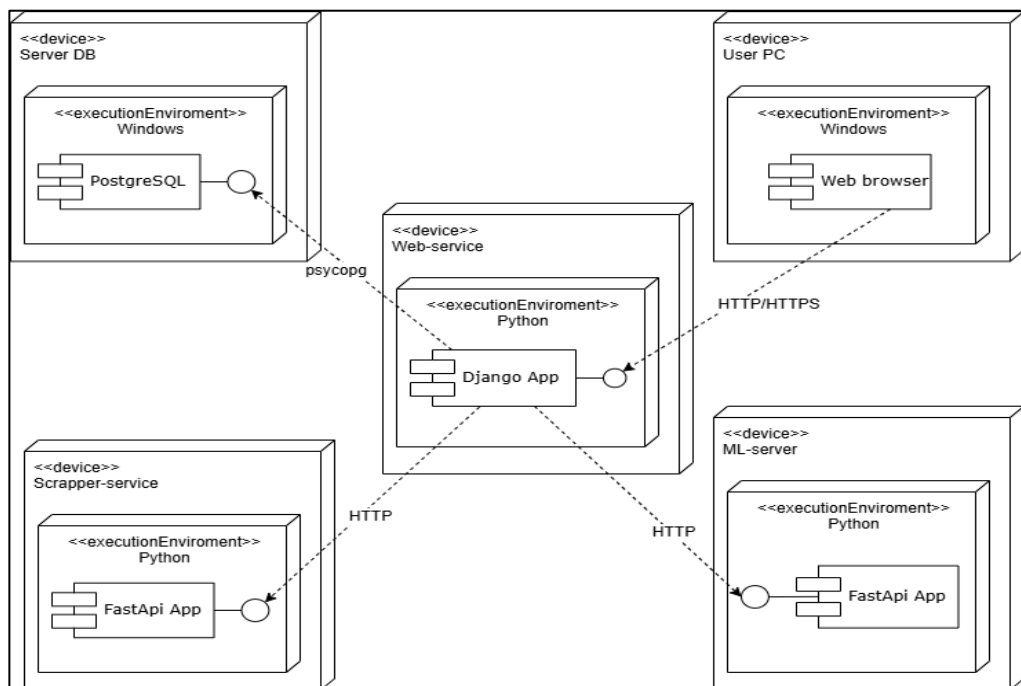


Рисунок 3 – Диаграмма развёртывания

В модели выделяются независимые узлы, включающие клиентский компьютер пользователя, сервер веб-приложения, сервер базы данных, специализированный ML-сервер и сервис сбора данных. Каждое устройство содержит соответствующую среду выполнения для программных артефактов. Обмен данными между узлами осуществляется посредством сетевых протоколов передачи гипертекста, а взаимодействие с базой данных обеспечивается выделенным драйвером. Подобная модель демонстрирует распределенность архитектуры и независимость отдельных подсистем, каждая из которых разворачивается и работает автономно.

Центральным связующим звеном выступает основной веб-сервис, к которому подключаются модули дообучения, машинного обучения, парсинга, скрапинга и подсистема базы данных. Архитектура отличается слабой связанностью подсистем, где взаимодействие осуществляется через явные программные интерфейсы. Подобное архитектурное решение упрощает замену или доработку отдельных модулей без внесения изменений в работу остальной системы.

Динамические аспекты функционирования проектируемой системы, а также изменение состояний ключевых объектов в процессе выполнения анализа конкурентной среды формализованы с использованием диаграммы последовательности (рисунок 4). Данная модель отражает взаимодействие между основными компонентами системы и позволяет проследить полный цикл обработки пользовательского запроса, начиная с его инициации и заканчивая представлением результатов анализа.

Процесс анализа конкурентной среды реализуется в виде последовательности взаимосвязанных этапов. На первом этапе аналитик формирует запрос через пользовательский интерфейс веб-приложения, задавая параметры анализа, такие как перечень конкурентов, временной интервал и тип исследуемых данных. После получения запроса основной веб-сервис инициирует процесс сбора информации, передавая задачу специализированному модулю скрапинга, функционирующему как независимый микросервис.

Модуль сбора данных осуществляет автоматизированный парсинг внешних источников, включая веб-сайты конкурентов и платформы с пользовательскими отзывами, после чего агрегированные данные передаются в контур интеллектуальной обработки. На следующем этапе ML-сервис выполняет комплексную предобработку данных, включающую очистку, нормализацию и векторизацию текстовой информации. Далее применяются алгоритмы машинного обучения для решения аналитических задач: оценка тональности отзывов, кластеризация текстов с целью выявления тематических групп, а также прогнозирование динамики цен и других рыночных показателей.

Полученные результаты передаются обратно в веб-сервис, где осуществляется их сохранение в базе данных с обеспечением целостности и структурированности. На заключительном этапе обработанная информация визуализируется в пользовательском интерфейсе в виде интерактивных дашбордов, графиков и отчетов, обеспечивая аналитика инструментами для интерпретации данных и принятия обоснованных управленческих решений.

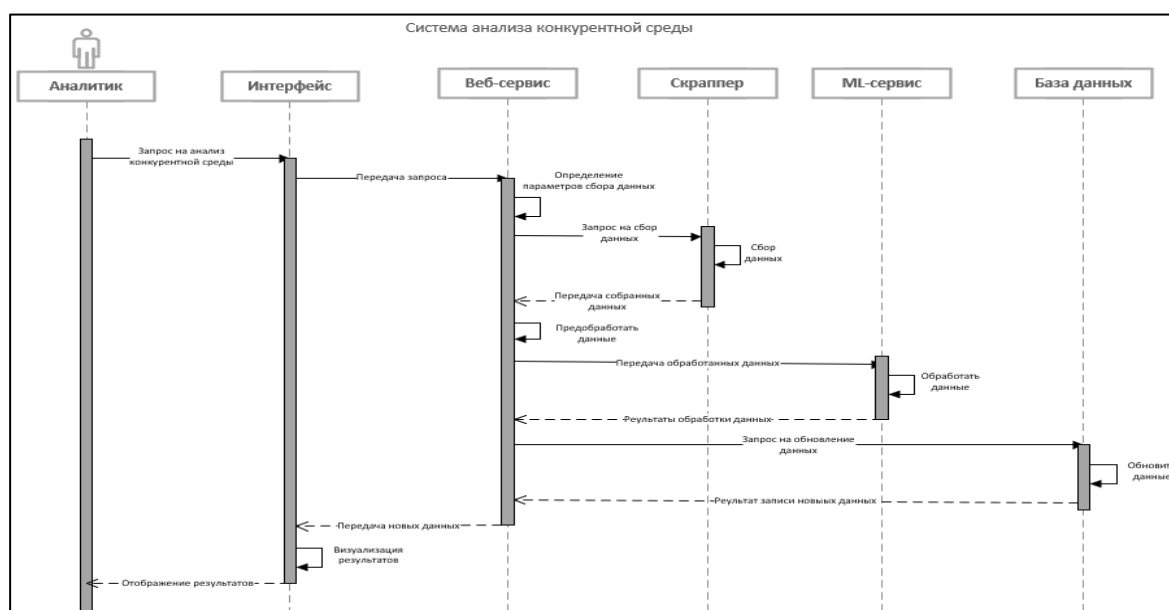


Рисунок 4 – Диаграмма последовательности

Таким образом, представленная диаграмма последовательности демонстрирует согласованное взаимодействие распределенных компонентов системы и подтверждает эффективность выбранной архитектуры для реализации задач анализа конкурентной среды в условиях обработки больших объемов данных.

В результате, было разработано веб-ориентированное программное средство для совершенствования процесса анализа конкурентной среды, направленное на повышение качества и оперативности принятия управленческих решений. В процессе работы была создана устойчивая распределенная микросервисная архитектура, спроектирована и реализована нормализованная реляционная база данных, а также сформирована надежная инфраструктура взаимодействия между модулями сбора информации, основным веб-сервисом и контуром предиктивной аналитики. Интеграция современных алгоритмов машинного обучения позволила успешно автоматизировать ресурсоемкие задачи оценки тональности клиентских отзывов, кластеризации слабоструктурированных текстов и прогнозирования динамики цен. Разработанная система показала стабильную работу при обработке больших массивов данных, что подтверждает её готовность к практическому применению предприятиями электронной коммерции для комплексного мониторинга рынка.

Список использованных источников:

1. Методы конкурентного анализа в розничной торговле [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://www.cfin.ru/management/strategy/competit/analysis.shtml>. – Дата доступа: 25.03.2026.
2. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment>. – Дата доступа: 25.03.2026.
3. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://arxiv.org/abs/1908.10084>. – Дата доступа: 26.03.2026.
4. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://umap-learn.readthedocs.io/en/latest/>. – Дата доступа: 27.03.2026.
5. Forecasting at Scale with Prophet [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://facebook.github.io/prophet/>. – Дата доступа: 28.03.2026.

UDC 004.89:339.13

SOFTWARE TOOL FOR IMPLEMENTING A SOLUTION BASED ON MACHINE LEARNING METHODS FOR COMPETITIVE ANALYSIS IN THE HOUSEHOLD APPLIANCES AND ELECTRONICS MARKET

Yankouskaya Y.A., student of group 272302

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

D.A. Storozhev, Master of Economic Sciences

Annotation. This article discusses the development of a web-oriented software tool for automated competitive analysis in the household appliances and electronics market. The key stages of the system's operation are described, including intelligent data collection, sentiment analysis of customer reviews using Transformer neural network architectures, density-based text clustering, and price dynamics forecasting. The distributed microservice architecture of the application, implemented using Django and FastAPI frameworks, as well as the PostgreSQL database datalogical model, are presented. The functionality of the system and the interaction logic of its components to support managerial decision-making are considered in detail.

Keywords. Competitive analysis, machine learning, data parsing, natural language processing, forecasting, microservice architecture, database, Python, Django, FastAPI, PostgreSQL.