

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ПРИКЛАДНОЙ МАТЕМАТИКИ И ИНФОРМАТИКИ
Кафедра информационных систем управления

ЛАПТЁНОК Егор Дмитриевич

**ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ КОРПОРАТИВНОЙ
АНАЛИТИКИ**

Дипломная работа

Научный руководитель:
старший преподаватель
В.В. Конах

Допущена к защите

«___»_____20__ г.

Заведующий кафедрой ИСУ

доктор технических наук, доцент А.М. Недзьведь

Минск, 2025

ОГЛАВЛЕНИЕ

Введение.....	8
Глава 1. Проблемы принятия решений в корпоративной аналитике	10
1.1 Анализ литературы	10
1.2 Виды корпоративной аналитики	12
1.3 Использование моделей искусственного интеллекта.....	13
1.4 Обзор платформ корпоративной аналитики.....	14
1.4.1 Microsoft Power BI.....	14
1.4.2 Tableau	15
1.4.3 Qlik Sense	16
1.4.4 SAP Analytics Cloud	16
1.4.5 Oracle Analytics.....	17
1.4.6 Анализ платформ	17
1.5 Постановка задачи.....	18
1.6 Выводы	19
Глава 2. Алгоритмы и средства для реализации программного обеспечения 20	
2.1 Обзор больших языковых моделей	20
2.1.1 OpenAI	21
2.1.2 Claude.....	21
2.1.3 DeepSeek.....	22
2.1.4 Llama.....	22
2.1.5 Mistral	23
2.1.6 Сравнение моделей	24
2.2 Интеграция искусственного интеллекта.....	25
2.3 Анализ прогнозирующих моделей временных рядов	27
2.3.1 Prophet	28
2.3.2 ARIMA.....	28
2.3.3 Оценка внедрения моделей	29
2.4 Модели	31
2.5 Алгоритмы	32
2.6 Выводы	33
Глава 3. Разработка программного обеспечения	35
3.1 Определение инструментов для разработки	35
3.1.1 Язык программирования	35
3.1.2 Библиотеки и фреймворки	35
3.1.3 Механизмы безопасности.....	36
3.1.4 Среда разработки.....	37

3.1.5 Дополнительные технологии	37
3.2 Архитектура программного обеспечения.....	38
3.3 Реализация программного обеспечения	39
3.3.1 Взаимодействие с большой языковой моделью	40
3.3.2 Разработка серверной части.....	41
3.3.3 Создание интерфейса.....	42
3.3.4 Тестирование и оценка качества программного обеспечения	44
3.4 Выводы.....	45
Заключение	47
Список использованных источников	49

ПЕРЕЧЕНЬ УСЛОВНЫХ ОБОЗНАЧЕНИЙ, СИМВОЛОВ И ТЕРМИНОВ

API — application programming interface (программный интерфейс)

LLM — large language model (большая языковая модель)

MLLM — multimodal large language model (мультимодальная большая языковая модель)

БД — база данных

ИИ — искусственный интеллект

ЛПР — лицо, принимающее решение

ПО — программное обеспечение

СППР – системы поддержки принятия решений

СУБД — система управления базами данных

РЕФЕРАТ

Дипломная работа: 49 страниц, 6 рисунков, 1 таблица, 1 формула, 12 источников.

Ключевые слова: КОРПОРАТИВНАЯ АНАЛИТИКА, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, ИНТЕГРАЦИЯ, АВТОМАТИЗАЦИЯ, ПРИНЯТИЕ РЕШЕНИЙ, PYTHON.

Объект исследования – поддержка принятия решений на основе корпоративной аналитики.

Предмет исследования – интеграция больших языковых моделей искусственного интеллекта в процессы корпоративной аналитики.

Цель исследования – обозначить актуальность программ поддержки принятия решений на основе корпоративной аналитики, разработать новые подходы к применению корпоративной аналитики при помощи искусственного интеллекта.

Методы исследования – системный подход, изучение соответствующей литературы по теме поддержки принятия решений на основе корпоративной аналитики и программная реализация на языке программирования Python.

Полученные результаты и их новизна – на основе поставленной задачи составлены необходимые требования к проектируемой системе, разработана программная реализация данной системы и приведена методика её применения. Новизна заключается в разработке подходов по эффективной интеграции больших языковых моделей в процессы корпоративной аналитики.

Достоверность материалов и результатов дипломной работы – все приведенные аналитические и расчетные материалы правильно и объективно отражают состояние исследуемого вопроса, а все заимствованные из литературных и других источников теоретические и методологические положения и концепции сопровождаются соответствующими ссылками.

Область возможного практического применения – разработанная система может быть использована на предприятиях для внедрения и автоматизации аналитической деятельности.

РЭФЕРАТ

Дыпломная праца: 49 старонак, 6 малюнкаў, 1 табліца, 1 формула, 12 крыніц.

Ключавыя словы: КАРПАРАТЫЎНАЯ АНАЛІТЫКА, ШТУЧНЫ ІНТЭЛЕКТ, ІНТЭГРАЦЫЯ, АЎТАМАТЫЗАЦЫЯ, ПРЫНЯЦЦЕ РАШЭННЯЎ, PYTHON.

Аб'ект даследавання – падтрымка прыняцця рашэнняў на аснове карпаратыўнай аналітыкі.

Прадмет даследавання – інтэграцыя вялікіх моўных мадэляў штучнага інтэлекту ў працэсы карпаратыўнай аналітыкі.

Мэта даследавання – пазначыць актуальнасць праграм падтрымкі прыняцця рашэнняў на аснове карпаратыўнай аналітыкі, распрацаваць новыя падыходы да ўжывання карпаратыўнай аналітыкі пры дапамозе штучнага інтэлекту.

Метады даследавання – сістэмны падыход, вывучэнне адпаведнай літаратуры па тэме падтрымкі прыняцця рашэнняў на аснове карпаратыўнай аналітыкі і праграмная рэалізацыя на мове праграмавання Python.

Атрыманыя вынікі і іх навізна – на аснове пастаўленай задачы складзеныя неабходныя патрабаванні да праектаванай сістэмы, распрацавана праграмная рэалізацыя дадзенай сістэмы і прыведзена метадыка яе прымянення. Навізна заключаецца ў распрацоўцы падыходаў па эфектыўнай інтэграцыі вялікіх моўных мадэляў у працэсы карпаратыўнай аналітыкі.

Дакладнасць матэрыялаў і вынікаў дыпломнай працы – усе прыведзеныя аналітычныя і разліковыя матэрыялы правільна і аб'ектыўна адлюстроўваюць стан доследнага пытання, а ўсе запазычаныя з літаратурных і іншых крыніц тэрэтычныя і метадалагічныя палажэнні і канцэпцыі суправаджаюцца адпаведнымі спасылкамі..

Вобласць магчымага практычнага прымянення – распрацаваная сістэма можа быць выкарыстана на прадпрыемствах для ўкаранення і аўтаматызацыі аналітычнай дзейнасці.

SUMMARY

Diploma work: 49 pages, 6 figures, 1 table, 1 formula, 12 references.

Keywords: CORPORATE ANALYTICS, ARTIFICIAL INTELLIGENCE, INTEGRATION, AUTOMATION, DECISION MAKING, PYTHON.

The object of the research – decision-making support based on corporate analytics.

The subject of the research – integration of large language models of artificial intelligence into the processes of corporate analytics.

The aim of the research – outline the relevance of decision support programs on the basis of corporate analytics, develop new approaches to the application of corporate analytics using artificial intelligence.

Research methods – system approach, study of relevant literature on the topic of decision support based on enterprise analytics and program implementation in the Python programming language.

The results of the work and their novelty – on the basis of the set task the necessary requirements to the designed system are made, the program realization of this system is developed and the methodology of its application is given. The novelty lies in the development of approaches to effectively integrate large language models into enterprise analytics processes.

Authenticity of the materials and results of the diploma work – all given analytical and calculation materials correctly and objectively reflect the state of the issue under study, and all borrowed from literature and other sources theoretical and methodological provisions and concepts are accompanied by appropriate references.

Recommendations on the usage – developed system can be used at enterprises for implementation and automation of analytical activity.

ВВЕДЕНИЕ

Объемы данных на мировом рынке монотонно увеличиваются с невероятной скоростью, а их аналитика стала необходимой частью принятий управленческих решений. Организации нуждаются не только в сборе данных, но и в методах их эффективного анализа, чтобы повысить свою конкурентоспособность и вовремя предотвратить риски. Данные могут представлять собой различные виды и форматы, но, особенно в финансовой сфере, выделяется актуальность работы с временными рядами, которые имеют очень широкую область применения, так как позволяют выявлять тенденции на рынке, обнаруживать скрытые закономерности, улучшить эффективность процессов и составлять прогнозы на будущее.

Современной тенденцией стало использование моделей искусственного интеллекта в качестве инструментов для аналитики. Их способность к обучению и анализу больших объемов данных открывает новые возможности для организаций, что позволит улучшить эффективность методов принятия решений. Но их внедрение в анализ корпоративных данных требует тщательного контроля, включая обеспечение их обучения для достижения достаточно высокой точности результатов, а также отслеживание доступа к данным и моделям.

Цель работы: изучение проблем корпоративной аналитики и разработка легко интегрируемого программного обеспечения, в котором в качестве основного компонента анализа данных используется модель искусственного интеллекта.

Для достижения цели ставятся следующие задачи:

- анализ проблем корпоративной аналитики;
- исследование существующих решений;
- составление требований к API;
- разработка простого и интуитивно понятного интерфейса;
- разработка и реализация подходов для анализа данных;
- интеграция модели искусственного интеллекта в процесс анализа;
- тестирование и оценка качества работы системы.

Актуальность темы заключается в автоматизации процессов корпоративного анализа, повышении скорости и точности прогнозирования для улучшения эффективности внутренних процессов и поддержания конкурентоспособности на рынке. Интеграция искусственного интеллекта в процессы корпоративной аналитики всё более широко распространяется на рынке, но его потенциал не всегда используется в полной мере.

В главе 1 рассматриваются основные понятия данной темы, анализируются современные системы и решения.

В главе 2 представлены разработанные модели и алгоритмы, сравнение и анализ технологий для разработки программного обеспечения.

В главе 3 описана реализация программного обеспечения, включая определение средств и проектирование архитектуры.

ГЛАВА 1

ПРОБЛЕМЫ ПРИНЯТИЯ РЕШЕНИЙ В КОРПОРАТИВНОЙ АНАЛИТИКЕ

1.1 Анализ литературы

Первым делом требуется изучение основного понятийного каркаса, который обеспечит базу для анализа темы, написания алгоритмов и реализации ПО:

- Аналитика — процесс сбора, обработки, анализа и интерпретации данных для получения значимой информации и поддержки принятия решений [4];
- Корпоративная аналитика — процесс использования данных, которые собираются внутри компании, а также актуальных рыночных данных, для принятия управленческих решений и получения информации;
- Временной ряд — данные, которые упорядочены относительно некоторого временного промежутка;
- Искусственный интеллект — технология, позволяющая компьютерам имитировать человеческие возможности к обучению, решению проблем и принятию решений [11];
- Большая языковая модель (LLM) — вид модели глубокого обучения, которая решает задачи по распознаванию и созданию текста на основе поставленной задачи;
- Мультиязычная большая языковая модель (MLLM) — большая языковая модель, способная работать с различными видами данных;
- Система поддержки принятия решений — аппаратно-программный комплекс для автоматизации процесса принятия решений на основе данных, полученных из естественных или искусственных источников [1].

В распоряжении предприятий имеется множество ценных данных, но возникает необходимость их обработки различными методами для получения значимых результатов. Принятие решений на основе данных поможет предприятиям оптимизировать внутренние операции, увеличить производительность и прибыль, а значит, повысить свою конкурентоспособность.

Получение достаточного количества данных уже не является проблемой для большинства организаций. Внутренние процессы, клиенты, датчики и различные устройства — все это приносит множество данных. Поэтому возникает необходимость в выделении и анализе именно полезных и актуальных данных.

В аналитику корпоративных данных также входит изучение данных для выявления результативных идей. Охватывается множество методов, включая бизнес-аналитику, интеллектуальный анализ данных и отчетность [12]. Система корпоративной аналитики может выявить современные тенденции, выбросы и закономерности благодаря объединению и анализу данных из различных источников. Интерактивные панели мониторинга вместе с инструментами бизнес-анализа предоставляют возможность визуализации ключевых показателей эффективности, исторических тенденций и операционных показателей для принятия обоснованных решений. Как пример, можно улучшить маркетинговые стратегии и клиентский опыт проанализировав информацию о поведении клиентов.

Эффективная система корпоративной аналитики должна предоставить именно корректные данные для лица, принимающего решения. Эти данные должны быть предоставлены своевременно, ведь условия на современном рынке очень быстро меняются, поэтому устаревшая информация может исказить текущую ситуацию и поспособствовать принятию неэффективного управленческого решения, а своевременные данные повысят гибкость и улучшат взаимодействие между различными отделами организации.

Организации, которые используют корпоративную аналитику в своих процессах, могут повысить операционную эффективность, увеличить рентабельность, ускорить процесс принятия решений и увеличить шансы нахождения новых путей развития бизнеса. Учёт работы сотрудников, распределение внутренних ресурсов и эффективность процессов становятся более прозрачными благодаря аналитике, что дает компаниям возможность получать данные в реальном времени и быть устойчивыми к вероятным будущим изменениям.

Важный аспект аналитики направлен на изучение взаимодействия с клиентами и улучшение их уровня обслуживания. Информация собирается из непосредственного взаимодействия с клиентами, получения отзывов о работе, анализе конкурентов и других источников. Это улучшит представление о клиентах, их потребностях и предпочтениях, поможет улучшить сервис обслуживания и уровень удержания, что существенно увеличит продажи продукции и услуг.

У многих компаний есть возможность сбора огромного объема информации из множества источников, но часть из них больше

концентрируется именно на сборе, а не на хранении и разработке стратегий их использования. Чем больше количество неструктурированных и частично структурированных данных, скорость их сбора и объем, тем сложнее и медленнее выявление и анализ релевантных данных. Чтобы избежать подобных проблем нужно разработать эффективную стратегию управления данными для получения необходимой информации. Иначе большая часть времени будет уходить на поиск и управление, а не на анализ.

В качестве дополнительной проблемы можно выделить хранение данных. Большие объемы данных собираются на периферии, ещё больше хранится в облачных хранилищах. Такая тенденция ведёт к усложнению систем хранения данных, что также усложняет доступ к информации платформам и сотрудникам предприятий. Некоторые предприятия хранят данные по своим собственным фирменным стандартам, что вызывает множество проблем со совместимостью и возможным пропуском критически важной информации.

1.2 Виды корпоративной аналитики

В большинстве случаев платформы для корпоративной аналитики имеют инструменты для реализации трёх типов аналитики: описательной, предиктивной и предписывающей.

Описательная аналитика больше предназначена для отслеживания ключевых показателей эффективности, которые отображают текущее состояние предприятия. Данный вид аналитики работает с данными для получения картины прошлых результатов, что включает в себя создание отчетов и использование панелей мониторинга вместе с визуализацией для удобного предоставления информации [6].

Описательная аналитика позволяет более подробно изучить имеющиеся данные, выявить места, которые требуют более подробного изучения, и наметить пути для последующего улучшения. Подробное и глубокое изучение данных выявляет причины происхождения рассматриваемых событий или тенденций, что даёт материал для дальнейшей работы в этом направлении.

Предиктивная аналитика использует данные для получения иных результатов. В этом случае процесс выявления тенденций в данных направлен на прогнозирование потенциальных результатов в будущем. Тут строятся предсказательные модели на основе исторических данных, что реализуется при помощи машинного обучения и статистических методов. Главная роль отводится обучению модели и подбору релевантных данных, так как именно

от этого зависит точность прогнозирования модели. С её помощью можно предсказывать поведение клиентов, уровень спроса на определённую продукцию, финансовые тенденции и другие показатели.

Предписывающая аналитика похожа на предиктивную, но рассматривает анализируемые данные с целью составления определённой последовательности действий, если подобная ситуация возникнет в будущем. Данный тип аналитики не останавливается на предсказании будущих результатов, а дополнительно предоставляет различные рекомендации по достижению намеченных результатов. Тут используются методы оптимизации и моделирования, чтобы предоставить эффективные советы для принятия решений.

Перечисленные типы аналитики используют разные инструменты и подходы для анализа, предназначены для получения разных результатов из данных, а поэтому многие платформы предназначены для определенного типа анализа. Работа с несколькими типами аналитики является трудной задачей, эффективное решение которой может быть достигнуто с помощью использования моделей искусственного интеллекта.

1.3 Использование моделей искусственного интеллекта

Платформы корпоративной аналитики всё чаще и чаще обращаются к использованию возможностей ИИ и машинного обучения для оптимизации своих процессов анализа и принятия решений.

Эти технологии позволят гораздо быстрее анализировать огромные объёмы данных, использовать подходы к получению информации, которые недоступны для традиционных способов, что может также улучшить результаты принятия решений [11]. Модели ИИ позволят автоматизировать процесс анализа данных предприятия, что может существенно снизить затраты, улучшить производительность и эффективность. Анализируя данные клиентов для выявления закономерностей и предпочтений, предприятия смогут адаптировать свою продукцию и услуги для лучшего удовлетворения потребностей клиентов. Кроме того, машинное обучение и модели ИИ также могут прогнозировать будущие тенденции и результаты для принятия упреждающих решений и более эффективного распределения ресурсов [8].

Большинство инструментов и подходов в аналитике имеют аналог в виде соответствующей модели ИИ или возможность реализации соответствующей модели. Но интеграция ИИ в процесс корпоративной аналитики требует тщательного внимания ко множеству деталей.

Одним из важнейших аспектов является выбор наиболее подходящей модели, которая будет способна эффективно решать соответствующий спектр задач, подходить требованиям предприятия, демонстрировать высокую адаптивность в условиях меняющегося окружения. Немаловажными требованиями являются ресурсы, которые требуются на интеграцию и развертывание модели, время, данные на её обучение и многие другие факторы.

Однако такие модели будут довольно узкоспециализированы, так как направлены на замену функционала конкретного инструмента. Поэтому одной из основных идей данной работы является интеграция большой языковой модели вместо разработки собственных алгоритмов машинного обучения с нуля и исследование подходов улучшения точности её результатов.

1.4 Обзор платформ корпоративной аналитики

В сфере корпоративной аналитики преобладает использование нескольких известных платформ, которые предлагают полный цикл аналитики — от сбора и трансформации данных до визуализации и генерации отчётов. Среди наиболее известных — Microsoft Power BI, Tableau, Qlik Sense, SAP Analytics Cloud, Oracle Analytics. Они предназначены для поддержки принятия управленческих решений и широко распространены на крупных и средних предприятиях [6].

1.4.1 Microsoft Power BI

Microsoft Power BI представляет собой платформу, которая тесно связана с экосистемой Microsoft, поэтому может собирать информацию из множества источников, поддерживает работу с большими массивами данных, что делает её оптимальным вариантом как для крупных предприятий, так и для малого бизнеса.

Она ориентирована как на ИТ-специалистов, так и на обычных пользователей предприятий. Платформа поддерживает создание дашбордов, отчётов и графиков, а также анализ данных в реальном времени. Power BI имеет широкие возможности для визуализации. Пользователи могут настраивать графики, применять фильтры, использовать географические карты и другие элементы. Однако кастомизация визуальных компонентов

ограничена рамками встроенного редактора и требует знаний языка DAX для реализации более сложной логики.

Платформа поддерживает подключение дополнительного ПО через API-интерфейсы, и сама легко интегрируется в системы предприятий.

Пользователи могут получить доступ к инструментам аналитики на основе ИИ и создавать модели машинного обучения, используя автоматизированные возможности и интеграцию с Azure Machine Learning. Платформа имеет функции, которые позволяют пользователям задавать вопросы и получать ответы о данных с помощью обработки естественного языка. Кроме того, активно продвигается интеграция Copilot для облегчения работы пользователям с данными.

Но данная платформа может оказаться слишком ресурсоёмкой для некоторых предприятий, а интерфейс хоть и является довольно дружелюбным, но в то же время довольно нагромождён, поэтому пользователь может легко потеряться среди множества функций.

1.4.2 Tableau

Tableau является мощной платформой для аналитики, которая ориентирована на интерактивную работу с данными. В ней можно объединять и управлять данными из различных источников, включая данные Salesforce.

Основной акцент делается на интуитивной визуализации, что делает её популярной среди бизнес-аналитиков. Платформа предоставляет гибкие инструменты построения диаграмм, графиков, карт и панелей управления. Пользователь может работать с данными в режиме drag-and-drop без необходимости программирования, но для сложных расчётов используется язык Tableau Calculated Fields.

ИИ в Tableau построен на основе Einstein Trust Layer, а представлен в виде функций для объяснения данных и упрощения взаимодействия пользователя с платформой. В качестве преимуществ ИИ данной платформы можно также выделить персонализированную работу с каждым пользователем и большое внимание к обеспечению безопасности и конфиденциальности данных.

Внедрение Tableau требует развёртывания Tableau Server или Tableau Cloud, имеются клиентские приложения Tableau Desktop и Mobile, а также модуль Prep для обработки данных. Стоимость лицензии платформы довольно высока, а полноценное использование всех возможностей будет доступно лишь с самым высоким уровнем лицензии.

1.4.3 Qlik Sense

Qlik Sense базируется на ассоциативной модели данных, что позволяет пользователям интуитивно исследовать взаимосвязи между показателями. Это делает платформу особенно эффективной при поиске скрытых паттернов и аномалий.

Платформа делает акцент на скорости отклика и обработке больших объемов данных. Большой вниманием уделено широким возможностям для визуализации и интерпретации данных, интерфейс предлагает множество функций, но сделан дружелюбным для пользователя. Qlik Sense учитывает потребность предприятий в работе с данным в реальном времени и реализует технологии для объединения с устройствами и интеграции в системы предприятий.

ИИ в Qlik Sense также в большем реализуется для упрощения пользовательского опыта, работы с естественным языком, создания и подготовки данных. Но ещё имеются инструменты для автоматизированной обработки данных и осуществления предиктивной аналитики при помощи машинного обучения.

Сложность внедрения Qlik Sense умеренная, однако обучение персонала занимает больше времени из-за уникальной логики работы платформы.

1.4.4 SAP Analytics Cloud

SAP Analytics Cloud (SAC) является полностью облачным решением, которое ориентировано на корпоративный сегмент и тесно интегрировано с другими решениями SAP, например, ERP, HANA и другими. Она имеет инструменты для отчётности, планирования и предиктивной аналитики, реализован широкий набор возможностей для визуализации данных, но пользователи должны быть подготовлены и иметь навыки работы с SAPUI.

В SAC внедрено множество функций ИИ для улучшения процесса анализа, выявления и автоматической визуализации данных для упрощения и ускорения работы пользователей системы. Также ведется активная работа над интеграцией генеративной модели Joule в рабочие процессы, которая откроет новые возможности для пользовательского опыта в системе.

Внедрение SAC относительно сложное, особенно при интеграции с локальными источниками данных и построении сложных сценариев

планирования, а также требует серьёзных затрат и нуждается в сертифицированных специалистах SAP.

1.4.5 Oracle Analytics

Oracle Analytics представляет собой комплексную платформу, которая подойдёт для любого аналитика. Можно использовать облачный сервис Oracle Analytics Cloud или локальное решение Oracle Analytics Server — оба решения реализуют современные подходы к обеспечению безопасности и простоты управления. Платформа ориентирована на крупные предприятия, предлагает масштабируемость и высокую степень автоматизации.

Визуализация реализована на высоком уровне с поддержкой прогнозирования, выявления аномалий и трендов. Панели управления можно адаптировать под пользователя, настраивать фильтры, делать вложенные отчёты. Платформа имеет дружелюбный, но местами перегруженный интерфейс.

Oracle Cloud предоставляет множество ИИ-сервисов, из которых AI Assistant, AI Vision, AI Language и AI Document Understanding уже интегрированы в Oracle Analytics. Активно используется автоматическая кластеризация, построение прогнозов, интерпретация результатов и другие возможности машинного обучения. В будущем планируется интеграция LLM для обработки данных и визуализации.

Интеграция платформы с инфраструктурой Oracle происходит довольно просто, но для работы с другой инфраструктурой могут возникнуть дополнительные сложности внедрения. Также платформа имеет высокую стоимость.

1.4.6 Анализ платформ

Обзор рассматриваемых платформ корпоративной аналитики позволил выделить ключевые аспекты, которые необходимо учитывать при реализации подобных решений.

Выделяется сильная привязанность платформ к собственным экосистемам, что существенно затрудняет их интеграцию с внешними системами и снижает гибкость внедрения. Кроме того, платформа должна предусматривать механизмы как для лёгкой интеграции в существующие

инфраструктуры, так и для подключения дополнительных сервисов и модулей, что сделает её гибкой и масштабируемой.

Также важно, чтобы аналитическая платформа была простой в использовании и имела интуитивно понятный интерфейс, что сделает её доступной для широкого круга пользователей и сократит их время на адаптацию и обучение. Но в то же время она должна иметь достаточно надёжные механизмы безопасности для обеспечения конфиденциальности данных.

Выявлен современный тренд на внедрение инструментов машинного обучения и ИИ для анализа, визуализации и интерпретации данных, но в большинстве случаев такие решения реализуются в виде вспомогательных модулей для пользовательских сценариев.

В отличие от подобных решений, основная идея разрабатываемого ПО в ходе дипломной работы заключается в использовании LLM в качестве основного ядра для анализа, визуализации и интерпретации данных, а также взаимодействия с пользователем при помощи естественного языка.

Многие ведущие компании активно работают над внедрением LLM в свои продукты, что подтверждает актуальность и значимость выбранного подхода.

1.5 Постановка задачи

Система должна обеспечить автоматизацию корпоративной аналитики, простоту интеграции и развертывания ПО в существующие системы предприятий, предоставить простой и интуитивно понятный интерфейс с возможностью использования различных видов аналитики благодаря возможностям современного ИИ.

Для достижения этого необходимо:

- спроектировать модели компонентов системы;
- разработать алгоритмы для взаимодействия с LLM и использования ей дополнительных инструментов анализа;
- разработать методы графического представления информации;
- реализовать удобный интерфейс для взаимодействия пользователя с системой и визуализации информации;
- реализовать возможность использования интерфейса через API.

1.6 Выводы

В данной главе рассмотрены ключевые особенности проблем принятия решений в области корпоративной аналитики, выделены ключевые тенденции и требования, которые предъявляются к современным аналитическим системам.

Можно выделить следующие проблемы корпоративной аналитики:

- интеграция платформ аналитики в другие системы;
- обработка различных типов источников данных;
- обеспечение конфиденциальности данных;
- качество и актуальность анализируемых данных;
- масштабируемость и гибкость системы;
- сложность использования системы персоналом.

Особое значение приобретает интеграция технологий машинного обучения и ИИ, особенно актуальной является задача внедрения LLM в платформы для корпоративной аналитики, которые откроют новые возможности в автоматизации процессов анализа и взаимодействия системы с пользователем.

Анализ литературы и обзор существующих платформ подтвердили актуальность и обоснованность выбранного направления в разработке легко масштабируемого и гибко интегрируемого решения, которое использует LLM в качестве основного ядра системы, что представляет собой не просто альтернативу существующим аналитическим платформам, а новый шаг в развитии инструментов для корпоративной аналитики.

ГЛАВА 2

АЛГОРИТМЫ И СРЕДСТВА ДЛЯ РЕАЛИЗАЦИИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Исходя из поставленных задач, разрабатываемое ПО для корпоративной аналитики требует тщательного подбора инструментов и технологий. Наибольшее внимание необходимо уделить выбору LLM, которая является основным ядром системы, проанализировать самые эффективные решения на рынке и выбрать наиболее оптимальное.

Но простая интеграция LLM в процесс аналитики не даст удовлетворительного результата для обеспечения конкурентоспособности платформы. Тут возникает необходимость в разработке моделей и алгоритмов, которые позволят использовать максимум из предоставляемых возможностей ИИ.

Также нужно исследовать точность более узкоспециализированных инструментов для анализа возможности их эффективного внедрения в систему под управлением LLM.

2.1 Обзор больших языковых моделей

LLM являются фундаментом современной интеллектуальной обработки данных. Они позволяют обрабатывать, анализировать и интерпретировать большие объёмы данных с помощью генерации текста на естественном языке. Современная тенденция в сфере ИИ направлена на расширение возможностей языковых моделей путём внедрения им поддержки мультимодальности, то есть способности модели воспринимать и обрабатывать сразу несколько типов входных данных, таких как текст, таблицы, графики, изображения, звук и видео. Такие модели называются мультимодальными большими языковыми моделями (MLLM).

В корпоративной аналитике использование MLLM особенно перспективно, так как позволяет:

- анализировать и интерпретировать документы с таблицами и графиками;
- работать с PDF-документами и смешанными типами данных;

– обрабатывать изображения и предоставлять дополнительные возможности для визуализации и интерпретации.

При выборе модели стоит также учитывать её стоимость и условия доступа, возможность локального развёртывания для обеспечения более надёжной безопасности данных, наличие мультимодальных возможностей.

Далее рассмотрены некоторые наиболее известные и активно используемые LLM.

2.1.1 OpenAI

Компания OpenAI является одним из лидеров в области LLM, в частности, благодаря своей серии Generative Pre-trained Transformer (GPT).

Модель ChatGPT, основанная на архитектуре GPT-4, является одной из наиболее популярных в мире. Она обеспечивает высокую точность интерпретации текстов, уверенно справляется с различными типами задач, включая анализ данных, генерацию отчётов и обработку естественного языка, является одной из лучших моделей по созданию креативных текстов.

Также выделяется поддержкой широкого контекста, устойчивой работой с длинными документами, генерацией связного текста и поддержкой мультимодального ввода, демонстрирует высокую точность в логических, языковых и аналитических задачах. Модель показывает особенно высокие результаты в задачах описательной аналитики, генерации отчётов и текстовой интерпретации структурированных данных [5].

Одним из её преимуществ является развитый API-интерфейс, хорошо документированный и легко интегрируемый в приложения. Однако доступ к GPT-4 возможен только по платной подписке, что может ограничить использование модели в бюджетных или локальных корпоративных решениях. Кроме того, модель остаётся закрытой — невозможно локальное развёртывание, что важно при работе с конфиденциальными данными.

2.1.2 Claude

Claude — LLM компании Anthropic, которая названа в честь американского учёного и инженера Клода Шеннона, в которой большое значение отдается безопасности, интерпретируемости, надёжности и соответствию человеческим ценностям.

Данные модели отличаются высокой производительностью в задачах логических рассуждений, генерации отчётов и обработке длинных документов. Claude также хорошо справляется с суммаризацией и извлечением фактов [5].

Модель является закрытой, поэтому при интеграции через API имеет соответствующие сложности и коммерчески доступна по специальной подписке в рамках платформы Anthropic. По этой же причине она не может быть развёрнута локально, что снижает её гибкость для интеграции в корпоративные среды, а также она в меньшей степени заточена под числовую и аналитическую обработку данных.

2.1.3 DeepSeek

DeepSeek — серия китайских LLM, которые ориентированы на открытость и производительность. Её последние версии показывают высокую эффективность в задачах программирования, анализа числовых данных и структурированных массивов. Они показывают хорошие результаты в генерации программного кода, технических текстов и инструкций, активно развивается в направлении инженерных и табличных задач.

DeepSeek активно развивается как open-source проект, что делает его подходящим для локального развёртывания и использования в условиях ограниченного доступа к облачным API.

Несмотря на ограниченное распространение на рынке, модель показывает конкурентоспособные результаты в задачах описательной и предиктивной аналитики. Однако, DeepSeek пока уступает по полноте документации и стабильности API таким крупным компаниям, как OpenAI и Anthropic, а значит требует более тщательной интеграции и тестирования.

2.1.4 Llama

Llama — открытая архитектура моделей от компании Meta, которые являются одними из наиболее известных и широко используемых в open-source среде, на их архитектуре создаются многие независимые модели. Они предоставляют высокое качество генерации текста при возможности полного локального развёртывания. Llama 2 и Llama 3 являются трансформерными моделями с улучшенными механизмами внимания и обучения, рассчитанными

на масштабируемость. Данные модели отлично подходят для исследовательских проектов и академического использования, хорошо справляются с различными языковыми задачами, не опираясь при этом на собственные источники данных.

Особенность Llama заключается в ориентации на исследовательское и корпоративное сообщество с открытой лицензией (для Llama 2 и частично — для Llama 3). Модели активно используются как основа для создания других моделей и агентов, включая мультимодальные решения (Llava, Llama Adapter и др.). Это делает Llama особенно привлекательной для систем, где важны контроль, безопасность и кастомизация.

Но данные модели требуют самостоятельной настройки, развертывания и не имеют готового публичного API, что осложняет их интеграцию в многие платформы корпоративной аналитики.

2.1.5 Mistral

Mistral — семейство моделей с открытым исходным кодом, разработанное компанией Mistral AI. Они доступны как в виде локальных сборок, так и через публичные API.

Mistral ориентирована на высокую производительность при минимальной ресурсоемкости. Архитектурно она базируется на трансформере с блоками типа sliding window attention, что позволяет эффективно обрабатывать длинные контексты [10].

Главным преимуществом Mistral является открытость и простота интеграции: доступны как исходный код модели, так и разнообразные API через сторонние хостинги, такие как Hugging Face, Fireworks.ai, OpenRouter. Это делает Mistral подходящей для прототипирования и разработки решений, где важны скорость, гибкость и бюджет.

Хотя Mistral уступает GPT-4 и Claude в плане мультимодальности, её открытая архитектура позволяет легко внедрять модель в любые корпоративные системы, включая закрытые среды. Кроме того, для описательной и частично предиктивной аналитики Mistral даёт качественные и воспроизводимые результаты при значительно меньших затратах.

Данные модели лучше всего подходят для систем, которые требуют эффективности и скорости получения результата. Они отлично справляются с программированием и математическими задачами и показывают высокую производительность в популярных бенчмарках.

2.1.6 Сравнение моделей

Для более подробно анализа проведено сравнение технических характеристик последних моделей (см. таблица 2).

Таблица 2 – Сравнение LLM [9]

Comparison criteria	o4-mini (high)	GPT-4.1	Claude 3.7 Sonnet Thinking	DeepSeek R1	Llama 4 Scout	Mistral Medium 3
Context Window	200k	1m	200k	128k	10m	128k
Artificial Analysis Intelligence Index	70	53	57	60	43	49
MMLU-Pro (Reasoning & Knowledge)	83%	81%	84%	84%	75%	76%
GPQA Diamond (Scientific Reasoning)	78%	67%	77%	71%	59%	58%
MATH-500 (Quantitative Reasoning)	99%	91%	95%	97%	84%	91%
AIME 2024 (Competition Math)	94%	44%	49%	68%	28%	44%
Blended usd / 1m tokens	1.93\$	3.50\$	6.00\$	0.96\$	0.28\$	1\$
Median Tokens / s	129.7	123.3	88.9	25.1	124.0	~90
Median First Chunk (s)	52.47	0.57	1.55	3.73	0.48	~0.4

Для определения стоимости и скорости используются характеристики по токенам — минимальным единицам текста, на которые разбиваются данные, контекстным окнам — максимальное количество токенов, которые модель может обработать за один раз, задержке — времени между отправкой запроса и получением ответа, а также показателям точности в некоторых сферах применения.

На основе данного сравнения можно сделать вывод, что лучше всех себя показывает именно модель от OpenAI o4-mini(high). У неё лучшие средние

результаты интеллекта, лучшая точность в бенчмарках в сфере науки и математики, а в AIME 2024 она имеет точность в 94% значительно опережая всех своих соперников. Но вместе с тем она имеет приличную цену для использования, а её огромная задержка может серьёзно повлиять на процесс анализа, что требует дополнительной доработки.

Модели GPT-4.1 и Claude 3.7 Sonnet Thinking имеют чуть более худшие результаты во многих сравнениях, но при этом они имеют гораздо меньшую задержку, что, несомненно, является большим преимуществом по сравнению с o4-mini(high), а модель GPT-4.1 имеет еще и огромное контекстное окно, которое позволит лучше запоминать и адаптироваться к окружению. Но их цена использования крайне высока и может быть неприемлемой для множества предприятий.

Лёгкая модель DeepSeek R1 с маленьким контекстным окном превосходит почти все модели по точности и занимает второе место в данном сравнении. При этом она имеет среднюю стоимость, что делает её одной из самых перспективных в этом сравнении. Однако её задержка выше нормы, а скорость генерации текста невероятно низкая, что приведёт к таким же проблемам, что и модель o4-mini(high).

Модели Llama 4 Scout и Mistral Medium 3 показали результаты, которые оказались ниже, чем у предыдущих моделей, но при этом имеют хорошую скорость и небольшую цену использования. Отставание не столь критично, но модель Llama показала себя немного хуже в бенчмарках научного направления, что потребует её дополнительной доработки при внедрении в процессы аналитики. Но зато Llama имеет самое большое контекстное окно среди всех, что открывает множество возможностей для её широкого применения, а также имеет очень маленькую цену, что несомненно выделяет её на фоне остальных моделей.

2.2 Интеграция искусственного интеллекта

В процессе проектирования интеллектуальной системы корпоративной аналитики особое внимание было уделено вопросу архитектурной интеграции языковой модели. На этапе предварительного анализа были рассмотрены два возможных подхода:

- интеграция нескольких языковых моделей с возможностью их выбора пользователем;
- интеграция одной основной модели, подобранной для широкого круга аналитических задач.

Первая стратегия, которая основана на модели с несколькими LLM, предполагает возможность выбора наиболее подходящего инструмента для конкретной задачи. Такой подход позволяет использовать сильные стороны каждой модели и минимизировать влияние их недостатков на процесс аналитики. В теории такой подход обеспечивает максимальную гибкость и качество обработки за счёт оптимального подбора модели под каждый тип аналитики.

Также имеется возможность одновременного использования моделей используя ансамблевые стратегии при проектировании платформы, что может существенно повысить точность получаемых результатов и минимизировать некоторые недостатки моделей. Но скорость работы такой системы будет неприемлемой, особенно для предприятий, которым нужна отчётность в реальном времени.

Далеко не все пользователи владеют знаниями о различиях между LLM, особенностях архитектуры, скорости, контекстном окне, лучших направлениях их использования и других критериев. В результате, наличие выбора для неподготовленного пользователя будет приводить к неэффективным настройкам, что существенно ухудшит качество анализа данных.

Для поддержки нескольких моделей требуется доступ по API к множеству платформ, что является крайне невыгодным в финансовом плане. А при локальном развёртывании каждую модель придётся обслуживать и поддерживать, для чего нужна мощная вычислительная инфраструктура.

Вторая стратегия основана на использовании одной выбранной языковой модели, адаптированной под специфику корпоративной аналитики. Такой подход иногда может быть менее гибким, но имеет ряд преимуществ:

- такая платформа требует гораздо меньше времени и ресурсов на поддержку и развёртывание, нет необходимости синхронизировать поведение и параметры нескольких LLM;
- сокращается время ответа, минимизируются ошибки, связанные с неправильным выбором модели;
- правильная архитектура системы позволит легко заменить или обновить LLM на более подходящую для предприятия;
- модель можно дообучить на корпоративных данных для улучшения точности её прогнозов для предприятия;
- для платформы, которая использует одну LLM в качестве основного аналитического ядра, гораздо проще подключить дополнительные инструменты для повышения точности решения узкоспециализированных задач.

В результате анализа LLM и стратегии интеграции в качестве основной модели выбрана Mistral Medium 3. Она уступает по многим бенчмаркам лучшим решениям от компаний OpenAI и Anthropic, но всё равно показывает хорошие результаты и является одной из ведущих LLM в мире. Модели Mistral легко интегрируются через API, имеют возможность локального развёртывания и дообучения, показывают очень высокую производительность при генерации отчётов. Mistral позволяет использовать свои модели бесплатно, что существенно упростит исследование возможностей данной LLM в рамках дипломной работы.

Medium 3 является самой актуальной моделью компании на данный момент и во многих бенчмарках превосходит даже модели серии Large.

Одно из преимуществ разрабатываемого ПО заключается в том, что при его интеграции в процессы корпоративной аналитики реального предприятия модель легко может быть заменена на любую другую, которая будет соответствовать возможностям и требованиям предприятия.

2.3 Анализ прогнозирующих моделей временных рядов

Так как результаты LLM могут быть менее точны, по сравнению с результатами инструментов, которые созданы для решения конкретных типов задач, требуется рассмотреть возможности улучшения её точности прогнозирования.

В рамках данной работы будут исследованы подходы для улучшения точности работы с временными рядами, ведь они являются одними из наиболее часто встречающихся типов данных в корпоративной аналитике. Они являются важной областью предиктивной аналитики для прогнозирования прибыли, продаж, спроса и других параметров бизнеса.

LLM хорошо проводят предварительный анализ временных рядов и их интерпретацию, но их предсказательная точность может быть весьма ограничена. Они обучены на текстовых данных и не оптимизированы для решения задач временного моделирования, особенно в условиях сезонности, трендов, аномалий и неполных данных.

В связи с этим в рамках разрабатываемого ПО было предложено комбинировать LLM с узкоспециализированными моделями временного анализа. Такой гибридный подход позволяет использовать сильные стороны обеих технологий. LLM будет использована для анализа задачи, автоматической настройки модели, интерпретации и визуализации результата, а специализированная модель улучшит точность прогноза.

Для реализации данного подхода рассмотрены наиболее популярные и широко применимые прогнозирующие модели временных рядов ARIMA и Prophet.

2.3.1 Prophet

Модель Prophet разработана компанией Meta и представляет собой алгоритм прогнозирования временных рядов, ориентированный на удобство использования и устойчивость к типичным особенностям бизнес-данных. Изначально была создана для специалистов, которые не обладают глубокими знаниями в области статистики, что делает его подходящим вариантом для автоматизированной настройки при помощи LLM.

Prophet ориентирована на практическое использование при наличии сложной сезонности и отсутствия строгих требований к предварительной обработке данных. Она имеет встроенную поддержку нелинейных трендов, включая точки смены тренда, поддержку сезонности, возможности задания особых событий, регуляризации и кросс-валидации [2]. Может автоматически обрабатывать пропущенные значения и аномалии, имеет быструю сходимость и устойчивость к выбросам.

Ещё Prophet выделяется простотой своего интерфейса. Достаточно передать таблицу с временными метками и значениями, а остальная настройка может быть проведена автоматически или через несложные параметры. Это делает Prophet особенно удобной для интеграции в систему, которой управляет LLM.

2.3.2 ARIMA

Модель ARIMA (Autoregressive Integrated Moving Average) является классической статистической моделью для анализа и прогнозирования стационарных временных рядов, которая демонстрирует высокую интерпретируемость и устойчивость на ограниченных и высокоструктурированных данных.

Она основывается на анализе автокорреляции и использует в качестве параметров авторегрессию, интеграцию и скользящее среднее. Эти компоненты отвечают за учёт зависимости текущего значения от предыдущих,

преобразование нестационарного ряда в стационарный и учёт ошибок прошлых предсказаний.

ARIMA основана глубоким статистическом анализе, а потому требует преобразования данных к стационарному виду. Из её особенностей можно выделить поддержку автоматического выбора параметров, возможность расширения в модели SARIMA и SARIMAX для учёта сезонности и внешних переменных.

Несмотря на точность и обоснованность, ARIMA требует более тщательной подготовки данных, ручной настройки параметров, а также знания статистики. Автоматизация работы с ARIMA затруднена, особенно если её настройку выполняет LLM, так как она может ошибаться в определении порядка интеграции или переоценивать автокорреляции, что приведёт к снижению точности.

2.3.3 Оценка внедрения моделей

С целью более тщательной оценки данного подхода улучшения качества прогнозов LLM проведено тестирование её взаимодействия с рассматриваемыми прогнозирующими моделями временных рядов.

Для имитации корпоративных временных рядов была сгенерирована синтетическая выборка длиной 200 дней с выраженной трендовой и сезонной составляющими. Это позволило воспроизвести характерные особенности реальных бизнес-данных. Данные отправляются на анализ LLM для генерации настроек прогнозирующих моделей, которым задаются эти параметры и производится обучение на данных. Для наглядного сравнения результаты визуализированы на графике (см. рисунок 2.1).

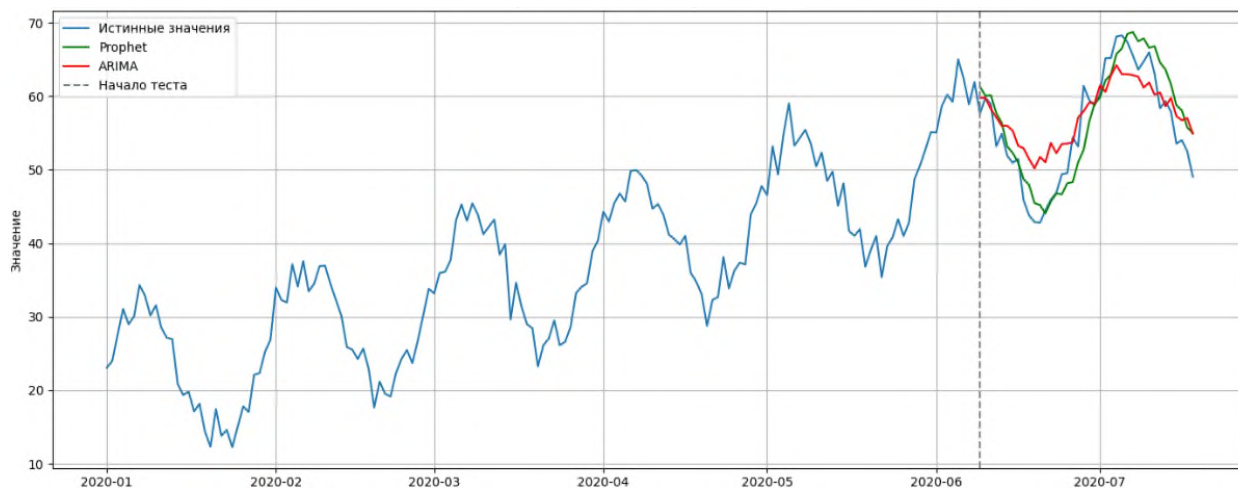


Рисунок 2.1 – Сравнение моделей на данных с сезонностью

Для оценки качества прогнозов использовались две метрики: MAE (Mean Absolute Error) — средняя абсолютная ошибка и RMSE (Root Mean Squared Error) — среднеквадратичная ошибка. Результаты:

- Prophet: MAE – 2.74, RSME – 3.35;
- ARIMA: MAE – 3.58, RSME – 4.23.

Как видно из сравнения, модель Prophet показала лучшие результаты по обоим метрикам. Это объясняется её способностью эффективно учитывать как сезонность, так и общий тренд, а также гибко адаптироваться к изменениям структуры ряда.

Для подтверждения сделанных выводов повторим сравнение на новых данных, в которых будет выражена более сложная сезонность. График сравнения моделей представлен ниже (см. рисунок 2.2).



Рисунок 2.2 – Сравнение моделей на данных со сложной сезонностью

На графиках хорошо видно, что Prophet демонстрирует более плавное следование истинным данным на тестовом отрезке, тогда как ARIMA в ряде точек отклоняется от реальных значений, особенно при усилении тренда.

Результаты показали, что Prophet лучше справляется с данными, в которых присутствует как сезонность, так и тренд. ARIMA позиционируется как более точная модель, но проигрывает модели Prophet из-за более тонкой и сложной настройки. Также скорость работы ARIMA сильно медленнее, чем у Prophet, что также делает её менее перспективной для дальнейшей интеграции с LLM.

Данный эксперимент позволяет выделить модель Prophet как более приемлемую для прогнозирования в системах корпоративной аналитики, а также создает основу для интеграции более сложных моделей, включая ансамбли или гибриды Prophet и LLM.

2.4 Модели

Для представления основных компонентов системы, которые будут обеспечивать анализ корпоративных данных через взаимодействие с LLM, требуется разработать модель системы для обеспечения максимальной эффективности использования намеченного подхода проектирования системы корпоративной аналитики. Исходя из поставленных задач и требований, а также на основе проведенного анализа подходов и инструментов, составлена ключевая модель системы:

$$S = (id, model, type, input, output, com) \quad (2.1)$$

Формула (2.1) представляет собой модель системы, которая соединяет пользователя и LLM, она должна обеспечить максимально эффективное и удобное человеко-машинное взаимодействие.

Основные компоненты модели:

- `id` – уникальный идентификатор задачи для анализа, который позволит системе различать и структурировать процессы взаимодействия;
- `model` – используемая LLM, основное ядро анализа в системе;
- `type` – тип поставленной аналитической задачи;
- `input` – входные данные;
- `output` – результат анализа, включая графическое представление данных;
- `com` – коммуникации между компонентами системы.

Каждый из этих компонентов обеспечивает целостность системы и играет важную роль в обеспечении взаимодействия с пользователем.

Система поддерживает работу с основными типами аналитики, включая описательную, предиктивную и предписывающую, а также имеет возможность предоставить ответ на другие запросы пользователей из данной сферы.

Входные данные формируются из поставленной на анализ задачи, в которой также указываются необходимые ограничения и правила как самого анализа, так и формата ответа, а также самих данных для анализа. Входные данные принимаются в текстовом формате без строго паттерна их записи и имеют ограничения по размеру, которые определяются выбранной моделью LLM. Высокий уровень обработки естественного языка основной модели ИИ позволит эффективно работать с большинством типов данных.

Система должна самостоятельно обеспечить обработку входных данных пользователей, применить необходимые настройки для LLM и передать

данные для анализа. По завершению процесса система аналогичным образом должна обработать полученный результат и передать его пользователю в удобном формате через API или отобразить при помощи интерфейса.

Результаты анализа состоят из двух частей: текстовой и графической. Текстовый ответ состоит из краткого содержательного результата и подробного анализа поставленной задачи. Первая часть ответа может быть использована для последующего более тщательного анализа и визуализации результатов, а вторая часть проходит обработку и выдается пользователю как конечный результат работы системы. Графическая часть реализуется на входных данных и результате, который был предоставлен моделью в текстовой части ответа. По результатам анализа всегда будет представлен подробный отчёт в виде текста, а графическая часть может отсутствовать, что зависит от настроек пользователя.

2.5 Алгоритмы

Алгоритмы — это основа работы любой системы. Они определяют, как данные обрабатываются, как задачи передаются LLM и в каком формате, пользователь получит результаты.

Для корректной и эффективной работы системы алгоритмы системы должны включать в себя следующие этапы:

- подготовку данных в формате, который будет подходить для передачи модели через API или GUI интерфейс;
- использование алгоритма настройки LLM;
- отправку данных LLM и получение результата анализа;
- обработку и передачу полученного результата.

Процесс обработки задачи начинается с определения настроек, которые задал пользователь. Он имеет возможность сделать это через внедренный в систему предприятия API или непосредственно через GUI интерфейс. Очень важно грамотно сформулировать саму задачу, ведь именно она содержит самую важную часть для LLM.

Внутри платформы также реализованы скрытые от пользователя настройки и промпты, которые будут автоматически подобраны алгоритмами исходя из данных пользователя.

В случае некорректной настройки анализа пользователем, например, когда параметры анализа не соответствуют задаче, система автоматически распознает эту ошибку и отредактирует параметры, отдавая при этом приоритет именно поставленной задаче.

При обработке полученного результата идёт разделение данных на те, которые будут использоваться для отображения пользователю, и те, которые могут быть использованы для дальнейшей, более глубокой обработки. Данные для отображения преобразуются в удобный формат для чтения и выводятся через интерфейс системы или возвращаются через API.

Если пользователь системы дополнительно выбрал графическую визуализацию, то алгоритм взаимодействия с LLM повторяется заново, где в качестве задачи уже будет требоваться получить инструкции для отображения графика, а в качестве данных будут передаваться исходные данные пользователя и данные, которые были получены на предыдущем этапе. Затем полученные инструкции преобразуются и выполняются в системе для получения графика, после чего тот будет преобразован в специальный формат для вывода через интерфейс или вернется через API.

Графическая визуализация возможна не для всех видов запросов, поэтому при некорректных запросах пользователю будет выведено сообщение с объяснением ошибки.

Дополнительные действия анализа используются не только для визуализации, но и для использования более узкоспециализированных прогнозирующих моделей временных рядов, что будет осуществляться автоматически при помощи настроек, которые заданы внутри системы.

Такой подход имеет большой потенциал, так как его можно расширить дополнительными инструментами, что улучшит показатели системы в других областях анализа.

2.6 Выводы

В главе проведено подробное исследование популярных LLM, выявлены их сильные и слабые стороны, сделано их тщательное сравнение, по результатам которого выбрана для использования модель Mistral Medium 3.

В главе также рассмотрены и описаны основные модели и алгоритмы, лежащие в основе разработки системы корпоративной аналитики.

Модели представляют собой абстрактные представления компонентов системы, которые необходимы для ее корректного функционирования, они позволяют описать их характеристики, функциональность и взаимодействие с другими компонентами системы. А алгоритмы обеспечивают функциональность системы, включая обработку данных, взаимодействие с LLM и визуализацию результатов.

Также разработаны подходы по улучшению качества работы LLM путём внедрения дополнительных инструментов, которые расширят её функционал, а в качестве практической части исследования подхода намечено внедрение инструментов для прогнозирования временных рядов.

Эффективная реализация описанных подходов и алгоритмов создаст основу для построения автоматизированной системы анализа, способной решать сложные задачи в области корпоративной аналитики.

ГЛАВА 3

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

3.1 Определение инструментов для разработки

Для разработки платформы корпоративной аналитики требуется использование надёжных, гибких и современных инструментов, которые будут обеспечивать высокую производительность, масштабируемость и безопасность работы системы.

3.1.1 Язык программирования

В качестве основного языка для разработки был выбран Python. Он является высокоуровневым интерпретируемым языком, который очень популярен как у начинающих разработчиков, так и активно используется в крупных компаниях и проектах, а потому имеет огромное сообщество пользователей и поддержку.

Python является очень гибким языком, может использоваться для функционального, процедурного и объектно-ориентированного подхода программирования, предоставляет возможность к разделению ПО на модули и пакеты. Позволяет использовать множество современных подходов и инструментов, но главным его достоинством является наличие огромного множества библиотек с решениями, которые можно использовать практически в любом направлении разработки ПО. Поэтому он позволяет реализовать все процессы от обычной обработки данных до полноценного развёртывания системы.

3.1.2 Библиотеки и фреймворки

Python имеет множество библиотек для работы с данными (Pandas, NumPy), их визуализации (Matplotlib, Seaborn), взаимодействия с LLM (OpenAI, MistralAI), создания и развёртывания полноценных приложений.

Для реализации серверной части выбран фреймворк FastAPI, который позволяет создать высокопроизводительный REST API с использованием асинхронной модели обработки запросов, а также легко интегрируется с различными механизмами и внешними сервисами.

Графический интерфейс реализован при помощи фреймворка Gradio, который предназначен для создания и развёртывания дружелюбных интерфейсов с поддержкой работы со множеством видов данных, а также ориентирован непосредственно для взаимодействия с моделями ИИ.

3.1.3 Механизмы безопасности

В корпоративных аналитических системах особое значение приобретает обеспечение информационной безопасности, особенно в контексте обработки конфиденциальных данных и взаимодействия между системами по сети. В условиях использования облачных решений, распределённых архитектур и открытых интерфейсов REST API требования к безопасности выходят на первый план.

Для предотвращения угроз несанкционированного доступа, утечки данных, подделки запросов и других проблем, необходимо реализовать надёжную схему аутентификации и авторизации, обеспечивающую проверку подлинности каждого запроса. Для решения данных задач подходят технологии OAuth 2.0 и JSON Web Token (JWT).

OAuth 2.0 — это современный открытый протокол авторизации, предназначенный для безопасной делегации доступа к защищённым ресурсам между клиентами и серверами. Он стал стандартом и широко используется в корпоративных и облачных решениях. В отличие от традиционной аутентификации, OAuth 2.0 предоставляет механизм делегирования полномочий, при котором клиент может получить доступ к ресурсам пользователя без передачи его учётных данных (логина и пароля) напрямую.

JSON Web Token — это компактный, URL-безопасный формат передачи информации между двумя сторонами. Он часто используется в связке с OAuth 2.0 в качестве формата для токена доступа. Структура JWT состоит из трёх частей, которая содержит информацию об алгоритме подписи и типе токена, описания идентификатора пользователя, времени истечения срока действия, роли, контексте запроса, а также подписи, которая создаётся путём хеширования заголовка и полезной нагрузки с использованием секретного ключа.

Эти технологии применяются для создания пользователям токенов доступа при аутентификации, после чего он используется для выполнения защищенных операций. Токены имеют ограниченный срок действия и могут быть отозваны, что обеспечивает управляемый и безопасный доступ.

3.1.4 Среда разработки

Для использования всех описанных инструментов и технологий отлично подходит PyCharm IDE, среда разработки для Python от компании JetBrains.

Среда ориентирована на Python-разработку, поэтому она имеет инструменты для автоматического дополнения и рефакторинга кода, предлагает решения для инспекции кода и анализа зависимостей, встроенные инструменты для тестирования и отладки. PyCharm имеет удобный и функциональный интерфейс, предоставляет возможности для удобной работы с виртуальными окружениями и пакетными менеджерами, а также включает глубокую интеграцию с системами контроля версий.

PyCharm обладает возможностью интегрирования плагинов для поддержки развёртывания серверной и клиентской частей системы, работы с базами данных и инструментами контейнеризации.

3.1.5 Дополнительные технологии

Платформе корпоративной аналитики для реализации механизмов безопасности с целью обеспечения надёжности и конфиденциальности данных требуется использование базы данных.

Для хранения данных выбрана реляционная БД PostgreSQL, так как она является одним из самых надёжных решений в промышленной разработке и активно используется в системах корпоративной аналитики. На фоне остальных реляционных БД она выделяется повышенной надёжностью и стабильностью, расширенными возможностями работы с форматом JSON и популярностью использования во многих средах для развёртывания.

В реализации ПО для корпоративной аналитики развёртывание БД осуществляется в контейнере при помощи ПО для автоматизации развёртывания и контейнеризации Docker, что обеспечит лёгкую переносимость БД на другие платформы.

3.2 Архитектура программного обеспечения

На начальном этапе разработки ПО проведено проектирование его архитектуры, которая задаёт структурную основу и определяет взаимодействие между основными компонентами. В данной работе разработана архитектура, обеспечивающая эффективную обработку данных и взаимодействие с LLM Mistral.

Архитектура разрабатываемой системы построена с учётом требований к гибкости, масштабируемости и возможности интеграции с внешними платформами. Она состоит из трёх основных слоёв: пользовательского интерфейса, серверной части и слоя искусственного интеллекта (см. рисунок 3.1).

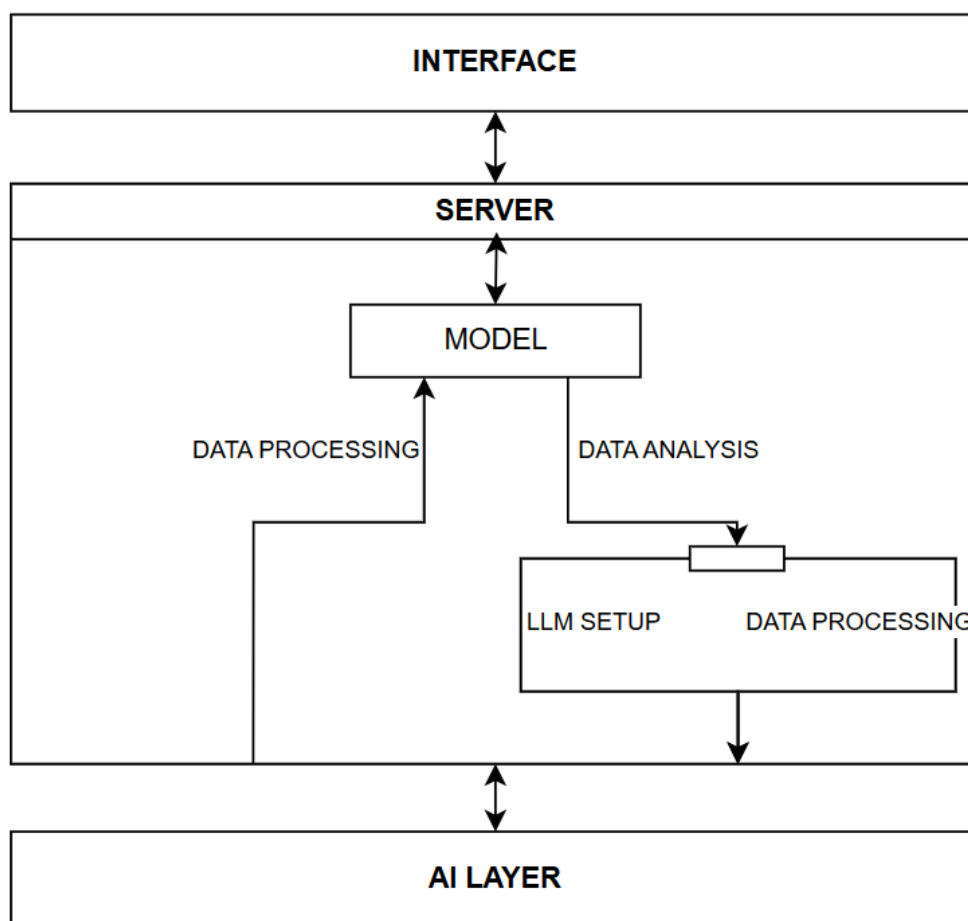


Рисунок 3.1 – Архитектура системы

Пользовательский интерфейс реализован с использованием фреймворка Gradio, который обеспечивает удобный графический интерфейс для взаимодействия пользователя с системой, состоит из элементов для ввода

данных, выбора типа анализа, настройки параметров, создания графиков и отображения результатов.

Серверная часть отвечает за обработку запросов от пользователя, выполнение логики бизнес-процессов и взаимодействие с моделью Mistral через API. Этот слой также реализует механизмы обеспечения безопасности и конфиденциальности данных, а также интеграцию дополнительных инструментов для улучшения качества работы LLM. Серверная часть реализована на языке Python с использованием библиотек FastAPI для создания RESTful-эндпоинтов и обеспечения высокой производительности.

Слой искусственного интеллекта состоит из модели Mistral, которая выполняет аналитическую обработку данных. Этот слой обрабатывает входные данные, предоставляет текстовый анализ и генерирует графики (при необходимости), возвращая результаты в удобном формате.

Исходя из разработанной архитектуры составлена основная схема взаимодействия слоёв системы:

- пользователь вводит данные через интерфейс Gradio;
- данные передаются на серверную часть, где они валидируются, обрабатываются, задаются настройки и формируется запрос для AI Layer;
- сервер отправляет запрос к модели Mistral, получает результаты анализа и передаёт их обратно на пользовательский интерфейс;
- результаты отображаются пользователю в текстовом виде или в виде графика.

Взаимодействие между слоями происходит через REST API для обеспечения гибкости в разработке и возможности интеграции с другими системами. Такой подход к проектированию позволяет масштабировать решение, добавлять новые функции и возможности без кардинальной переработки существующей структуры.

Архитектура системы поддерживает возможность модульного расширения, что позволяет интегрировать дополнительные алгоритмы или модели ИИ при необходимости.

3.3 Реализация программного обеспечения

Разработка ПО для системы корпоративной аналитики происходила в несколько этапов. Сначала были определены задача и требования к функционалу системы, создана архитектурная схема системы, выбраны технологии и инструменты для разработки ПО.

После выполнения всех подготовительных действий началась основная разработка системы.

3.3.1 Взаимодействие с большой языковой моделью

На начальном этапе разработан и реализован модуль взаимодействия с LLM, включающий преобразование входных данных в формат, поддерживаемый Mistral, и обработку её выходных данных. Ниже представлен код основного алгоритма, в котором данные сначала обрабатываются через вызов функции для получения подробного анализа. В функции также применяются дополнительные настройки для модели, которые скрыты от пользователя.

Затем данные разделяются на части для отображения пользователю и для дальнейшей обработки. При необходимости создать график запрашиваются дополнительные инструкции от LLM, затем идёт их выполнение в дополнительном методе, после чего график конвертируется в удобный формат. В конце результат будет возвращен в место вызова функции.

```
import os
from mistralai import Mistral
from ai_process import get_req_prompt_arr, get_plot_prompt_arr,
get_graph

def get_req_prompt_arr(type):
    prompt = analytics_prompts[analytics_types.index(type)]
    return [get_msg(prompt), get_msg(res_prompt)]

def get_plot_prompt_arr():
    return [get_msg(plot_prompt)]

def get_msg(prompt):
    return {"role": "system", "content": prompt}

with open('api_key.txt', 'r') as file:
    api_key = file.read().strip()
model = "mistral-medium-latest"
client = Mistral(api_key=api_key)
```

```

def predict(task, data, type, plot):
    response_text = get_analysis(task, data, type)
    data_only_response, detailed_response =
response_text.split(SEPARATOR)
    detailed_response = detailed_response.split("\n", 1)[1]
    graph = ""
    if plot == "True":
        code = get_plot_code(data)
        graph = get_graph(code)
    return detailed_response, graph

```

Дополнительно реализованы промпты для более точной настройки модели, которые выбираются и используются главным алгоритмом модуля исходя из входных данных и возможности использования различных типов аналитики. В этом же модуле реализованы промпты для задания правил LLM по использованию прогнозирующей модели Prophet и генерации настроек для неё.

3.3.2 Разработка серверной части

На втором этапе реализована серверная часть ПО, а именно создан API-интерфейс при помощи фреймворка FastAPI, настроено его взаимодействие с модулем LLM, реализована обработка входных и выходных данных.

Для обеспечения защиты корпоративных данных и контроля доступа была реализована система аутентификации и авторизации на основе OAuth 2.0 и JWT, которая позволяет ограничивать доступ к функциональности API только авторизованным пользователям и обеспечивает безопасную передачу учетных данных.

Для хранения учетных записей пользователей была использована реляционная модель PostgreSQL. Пользовательская модель реализована через SQLAlchemy:

```

class User(Base):
    __tablename__ = "users"
    id = Column(Integer, primary_key=True, index=True)
    username = Column(String(100), unique=True, index=True)
    hashed_password = Column(String(255))

```

Аутентификация производится через OAuth2PasswordRequestForm и сопровождается генерацией JWT. Создание и проверка JWT реализуются с использованием библиотеки python-jose. При генерации указывается срок действия токена, после чего он возвращается клиенту:

```
token = create_access_token(data={"sub": user.username})  
return {"access_token": token, "token_type": "bearer"}
```

```
def create_access_token(data: dict, expires_delta: timedelta =  
None):  
    to_encode = data.copy()  
    expire = datetime.utcnow() + (expires_delta or  
timedelta(minutes=15))  
    to_encode.update({"exp": expire})  
    return jwt.encode(to_encode, SECRET_KEY,  
algorithm=ALGORITHM)
```

Для валидации токена используется зависимость `oauth2_scheme`, которая автоматически извлекает токен из заголовка `Authorization` и проверяет его подлинность. Главный эндпоинт приложения, отвечающий за выполнение интеллектуального анализа данных, был защищён с использованием зависимости `Depends(get_current_user)`, поэтому запрос к этому эндпоинту невозможен без предварительного получения и передачи JWT-токена в заголовке запроса.

3.3.3 Создание интерфейса

Третий этап включает в себя разработку пользовательского интерфейса приложения при помощи библиотеки `Gradio` и реализации клиентской части ПО. Сначала разработан интерфейс главного окна приложения, код основных компонентов которого представлен ниже.

```
with gr.Tab("Analytics"):  
    gr.Interface(  
        fn=chat_call,  
        inputs=[token_state,  
                gr.components.Textbox(label="Task and Data  
description")],
```

```

        gr.components.Textbox(label="Data", lines=10),
        gr.components.Radio(analytics_types,
label="Analytics Type", value=analytics_types[-1]),
        gr.components.checkbox.Checkbox(label="Create
Plot", value=False)],
        outputs=[token_state,
        gr.components.Textbox(label="Analysis",
lines=10),
        gr.components.Image(label="Image",
visible=False)],
        title="Corporate Analytics AI",
        description="Interact with the AI. Enter a message and
get a response.",
        flagging_mode="never",
        theme=gr.themes.Citrus()
)

```

Графический интерфейс спроектирован с ориентиром на конечного пользователя и предоставляет интуитивно понятный доступ к основным функциям системы (см. рисунок 3.2). Затем для использования реализованных на серверной части механизмов безопасности было разработано стандартное окно с авторизацией пользователя.

Рисунок 3.2 – Графический интерфейс приложения

После разработки интерфейса была реализована логика работы клиентской части приложения в виде методов обработки входных и выходных данных, а также взаимодействие с серверной частью приложения. Также были реализованы примеры запросов для интерфейса, чтобы предоставить пользователю примерный формат использования системы и ускорить его знакомство с работой приложения.

Затем клиентская часть приложения была дополнена функцией авторизации, где перед вызовом основного API клиент сначала получает токен по адресу “/token”, а затем использует его при обращении к основному адресу “/predict”.

3.3.4 Тестирование и оценка качества программного обеспечения

На последнем этапе разработки проведено тестирование ПО и исправление ошибок. В процессе тестирования особое внимание уделялось работе механизмов безопасности и качеству результатов работы системы.

Разработанное ПО показало хорошую производительность и высокое качество как текстовых, так и графических результатов анализа (см. рисунок 3.3).

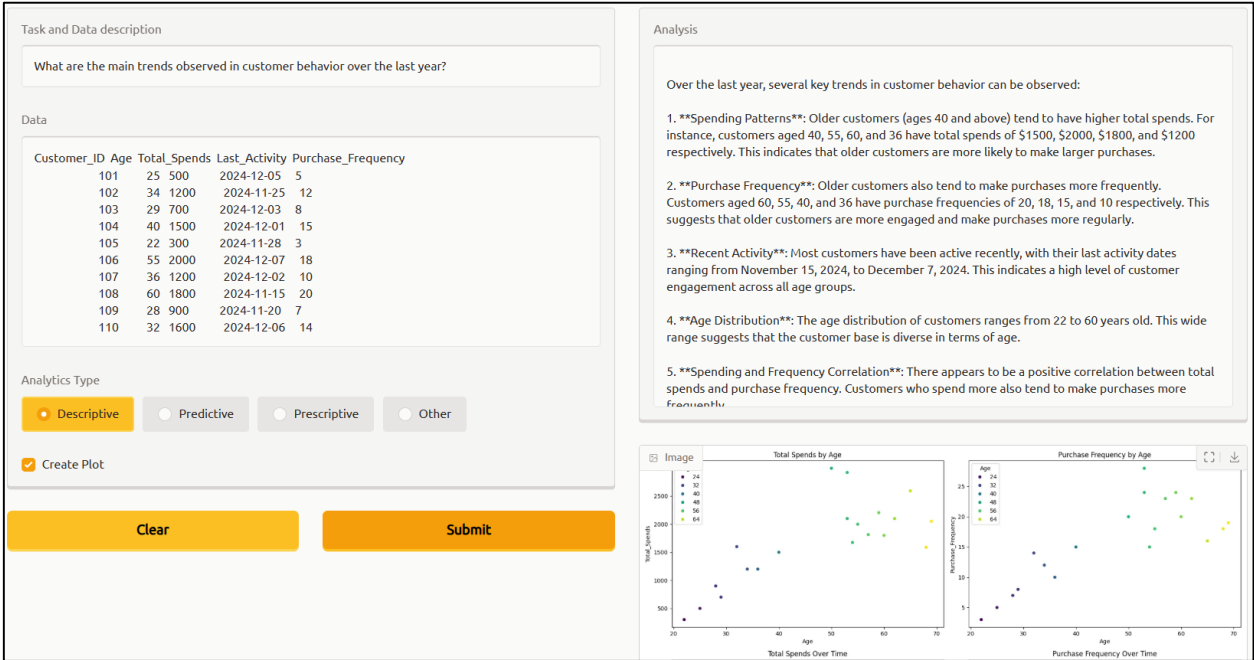


Рисунок 3.3 – Результат анализа

В соответствии с поставленными задачами и целью разработанная система корпоративной аналитики выдаёт подробный текстовый результат и

выводит его в специальное поле. Дополнительно выводится изображение под текстовым результатом, которое содержит исходные данные и, при наличии, результаты проведенного анализа, что значительно улучшает процесс восприятия и интерпретации сложной аналитической информации.

При получении входных данных в виде временных рядов алгоритмы применяли для LLM дополнительные настройки и промпты, в соответствии с которыми она дополнительно использовала модель Prophet для получения результатов (см. рисунок 3.4).

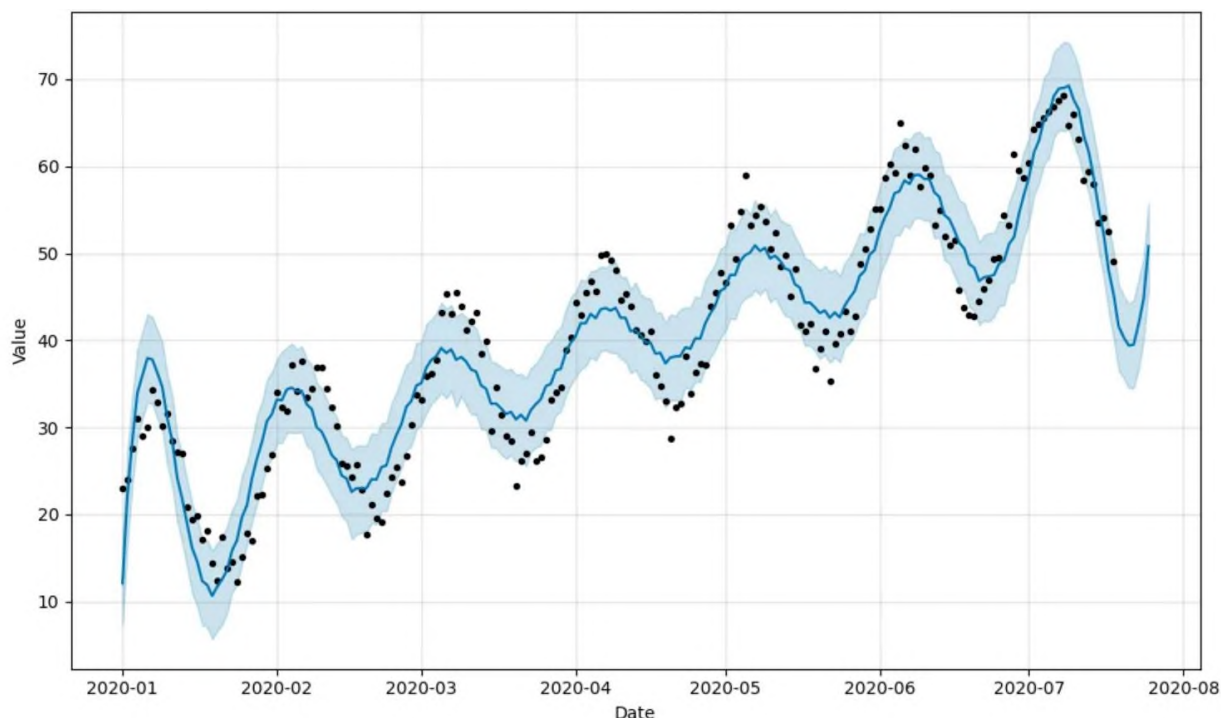


Рисунок 3.4 – Результат использования модели Prophet

Исходя из полученных результатов интеграция прогнозирующей модели временных рядов прошла довольно успешно. Хотя такой подход работает медленнее обычного анализа, видно, что LLM достаточно качественно анализировала данные для подбора настроек модели, что показало хорошие результаты в прогнозировании временных рядов.

3.4 Выводы

Демонстрация работы системы через графический интерфейс подтвердила интуитивность и удобство взаимодействия пользователя с системой. Реализация методов обработки данных, встроенные настройки и визуализация результатов значительно улучшили и упростили процесс

анализа и восприятие информации даже для пользователей без глубоких технических знаний.

Главное преимущество разработанного решения – его универсальность, автоматизация процессов анализа и визуализации данных, возможность интеграции через API, удобный графический интерфейс – всё это делает систему подходящей для широкого спектра задач и различных типов предприятий.

ЗАКЛЮЧЕНИЕ

В рамках выполнения данной дипломной работы было разработано приложение, обеспечивающее автоматизированный анализ данных, включая временные ряды, с использованием LLM как основного инструмента для проведения анализа и автоматизации системы.

Проведён ряд исследований существующих инструментов и разработаны подходы не только к интеграции ИИ в процессы корпоративного анализа, но и к улучшению качества работы LLM и автоматизации системы.

Созданная система позволяет решать широкий спектр аналитических задач, предоставляя результаты в виде прогнозов и рекомендаций. Работа включала в себя этапы от исследования предметной области до практической реализации и апробации разработанного программного обеспечения.

В процессе работы были решены следующие задачи:

- проанализированы современные подходы к построению корпоративных аналитических систем;
- разработаны алгоритмы для взаимодействия с LLM и работы с данными;
- произведена интеграция LLM для обеспечения более глубокого и точного анализа;
- создан графический интерфейс, ориентированный на удобство пользователя, а также обеспечена возможность взаимодействия с системой через API;
- проведено тестирование разработанного ПО, включая оценку его производительности, точности прогнозов и удобства работы.

Также был разработан подход к интеграции систем корпоративной аналитики, который обеспечивает гибкость подключения и возможность автоматической обработки данных.

Разработанное ПО может быть использовано в корпоративной аналитике, включая финансовую, производственную и маркетинговую сферы. Это позволяет компаниям автоматизировать анализ корпоративных данных, прогнозировать изменения и оперативно реагировать на возможные риски. Система также может использоваться в обучении специалистов по работе с данными и искусственным интеллектом, предоставляя реальный инструмент для изучения аналитических процессов.

Дальнейшие перспективы развития модели:

- расширение функционала системы для работы с большими объемами данных и более сложными типами анализа;

- улучшение и дообучение LLM, включая адаптивное обучение на новых данных;
- внедрение дополнительных модулей визуализации для более глубокого анализа и принятия решений;
- интеграция дополнительных инструментов анализа под управлением LLM.

Результаты работы предназначены для компаний, стремящихся к автоматизации аналитических процессов и повышению точности прогнозирования. Они могут быть успешно применены в областях, которые требуют обработки больших объемов данных и выработки рекомендаций в условиях высокой неопределенности. Таким образом, предложенное решение не только решает актуальные задачи корпоративной аналитики, но и закладывает основу для дальнейших исследований в области интеграции искусственного интеллекта в системы управления данными.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Виссия Х. Принятие решений в информационном обществе / Х. Виссия, В.В. Краснопрошин, А.Н. Вальвачев – СПб: ЛАНЬ, 2019. – 227 с.
2. Документация «Prophet Docs» / Facebook [Электронный ресурс]. – Режим доступа: <https://facebook.github.io/prophet/docs>. – Дата доступа: 20.05.2025.
3. Прогнозирование: принципы и практика / Р. Хайндман, Дж. Атанасулопулос – OTexts, 2021. – 450с.
4. Что такое аналитика / SAP [Электронный ресурс]. – Режим доступа: <https://www.sap.com/products/data-cloud/cloud-analytics/what-is-analytics.html>. – Дата доступа: 20.05.2025.
5. A Comprehensive Comparison of All LLMs / AI-PRO [Electronic resource]. – Mode of access: [Electronic resource]. – Mode of access: <https://www.heavy.ai/learn/big-data-analytics/enterprise-analytics>. – Date of access: 20.05.2025.
6. Enterprise Analytics Definition / Heavy.AI [Electronic resource]. – Mode of access: <https://www.heavy.ai/learn/big-data-analytics/enterprise-analytics>. – Date of access: 20.05.2025.
7. Hemachandran K. Predictive Analytics and Generative AI for Data-Driven Marketing Strategies / K. Hemachandran, Debductta Choudhury, Raul Villamarin Rodriguez, 2024 – 290 p.
8. Kumar B.M. The Role of Artificial Intelligence in Decision-Making Processes / B.M. Kumar, Bathina Shireesha, 2024.
9. LLM Leaderboard – Comparison / Artificial Analysis [Electronic resource]. – Mode of access: <https://www.heavy.ai/learn/big-data-analytics/enterprise-analytics>. – Date of access: 20.05.2025.
10. Mistral Docs / Mistral AI [Electronic resource]. – Mode of access: <https://docs.mistral.ai>. – Date of access: 20.05.2025.
11. What is AI? / IBM [Electronic resource]. – Mode of access: [Electronic resource]. – Mode of access: <https://artificialanalysis.ai/leaderboards/models/prompt-options/single/medium>. – Date of access: 20.05.2025.
12. What is enterprise analytics and why are companies using it? / ThoughtSpot [Electronic resource]. – Mode of access: <https://www.thoughtspot.com/data-trends/analytics/enterprise-analytics>. – Date of access: 20.05.2025.