Министерство образования Новосибирской области

ГБПОУ НСО «Новосибирский авиационный технический колледж имени Б.С. Галущака»

разработка бизнес-аналитической платформЫ

Системный анализ и проектирование

Разработал:

Бушин А.Н.

2024

**Содержание**

[1 Описание предметной области 3](#_Toc163748706)

[2 Проектирование ИС 6](#_Toc163748707)

[3 Создание приложения 8](#_Toc163748708)

[4 Аналоги и прототипы 12](#_Toc163748709)

[5 Подготовка тестовых сценариев 13](#_Toc163748710)

[6 Вывод 14](#_Toc163748711)

# **1 Описание предметной области**

Область бизнес-аналитики включает в себя комплексный процесс, начиная с сбора данных и заканчивая принятием обоснованных бизнес-решений. На этапе сбора данных аналитики определяют необходимые источники данных и разрабатывают стратегию их сбора, включая внутренние и внешние источники, такие как базы данных компании, открытые данные, социальные медиа и веб-аналитика. Этот процесс также включает очистку и структуризацию данных для последующего анализа. На этапе анализа данных аналитики применяют различные методы анализа, такие как статистический анализ, машинное обучение, алгоритмы кластеризации и прогнозирования, чтобы выявить закономерности и взаимосвязи между переменными. Цель анализа данных - получить понимание текущего состояния бизнеса, выявить возможности для роста и оптимизации процессов.

Полученные результаты анализа данных должны быть интерпретированы с учетом бизнес-контекста. Аналитики оценивают важность найденных паттернов и тенденций для бизнеса, а также определяют возможные последствия и риски, что позволяет принять обоснованные решения на следующем этапе. Наконец, на основе результатов анализа данных и их интерпретации бизнес-аналитики помогают руководству и другим заинтересованным сторонам принимать обоснованные решения. Это могут быть стратегические решения, тактические решения по управлению процессами и операциями, а также оперативные решения для реагирования на изменения в окружающей среде. Кроме того, бизнес-аналитики предоставляют рекомендации по оптимизации ресурсов и улучшению эффективности бизнес-процессов.

Но в тоже время существуют проблемы в области бизнес-аналитики, которые могут оказывать существенное влияние на эффективность работы компании и качество принимаемых решений. Недостаточная компетентность аналитиков в программировании, особенно у тех, кто имеет экономическое образование, может привести к тому, что они не могут использовать современные инструменты и методы анализа данных для выявления ключевых трендов и паттернов в информации. Это, в свою очередь, может замедлить процесс принятия решений и снизить конкурентоспособность компании на рынке.

Низкая скорость выполнения запросов в базах данных из-за большого объема данных также может стать проблемой. Большие объемы данных и множество баз данных могут вызвать задержки при получении необходимой информации для анализа, что снижает оперативность принятия решений и способствует потере возможностей.

Различный инструментарий у каждого аналитика также может создать препятствия для эффективной работы команды. Отсутствие стандартизации в использовании инструментов и технологий для анализа данных может затруднить совместную работу, обмен знаниями и обучение новых сотрудников, что в конечном итоге снизит производительность и качество работы.

Сложность выводов из-за объема информации также является серьезной проблемой. При анализе больших объемов данных аналитики могут столкнуться с трудностями в интерпретации результатов из-за избытка информации. Это может привести к неполноте анализа, упущению важных тенденций и неправильному принятию решений, что в конечном итоге может нанести ущерб бизнесу.

Для решения проблем в области бизнес-аналитики можно использовать комплексный подход.

Во-первых, создание синтетической копии продуктовой базы данных поможет ускорить выполнение запросов и сократить время на отладку. Эта копия будет содержать данные, аналогичные по свойствам основной базе данных, но с уменьшенным объемом, что обеспечит быстрый доступ к данным для аналитических целей.

Во-вторых, разработка системы запросов на русском языке позволит аналитикам без навыков программирования легко и эффективно работать с данными, что уменьшит порог вхождения и повысит производительность.

Третий шаг заключается в создании инструментов визуализации данных. Это поможет аналитикам наглядно представлять результаты своих запросов, улучшая понимание данных и выявляя скрытые паттерны и взаимосвязи.

Наконец, внедрение системы искусственного интеллекта для анализа визуализаций запросов и вывода ключевых выводов и рекомендаций поможет аналитикам быстрее и точнее интерпретировать данные, принимать обоснованные решения и оптимизировать бизнес-процессы.

Такой комплексный подход к решению проблем в бизнес-аналитике позволит повысить эффективность работы, улучшить качество принимаемых решений и обеспечить конкурентоспособность компании на рынке.

# **2 Проектирование ИС**

Клиент-серверные приложения в области бизнес-аналитики могут выполнять различные функции, помогающие аналитикам и бизнес-пользователям в анализе данных и принятии решений. Вот несколько возможных функций такого приложения:

Предоставление возможности визуализации данных в форме графиков, диаграмм, дашбордов и т. д., что помогает аналитикам быстро понять ключевые тренды и паттерны.

Автоматическое формирование отчетов на основе анализа данных с возможностью настройки параметров отчетов пользователем.

Создание интерактивных дашбордов, которые позволяют пользователям исследовать данные и проводить анализ в реальном времени, а также устанавливать фильтры и параметры для получения более детальной информации.

Интеграция с различными источниками данных, такими как базы данных, веб-сервисы, файлы Excel и т. д., для получения комплексной информации.

Возможность интеграции с различными внешними сервисами и API для получения дополнительной информации и расширения функциональности приложения.

На основе вышеперечисленных функций разработана диаграмма прецедентов, которая показана на рисунке 1.

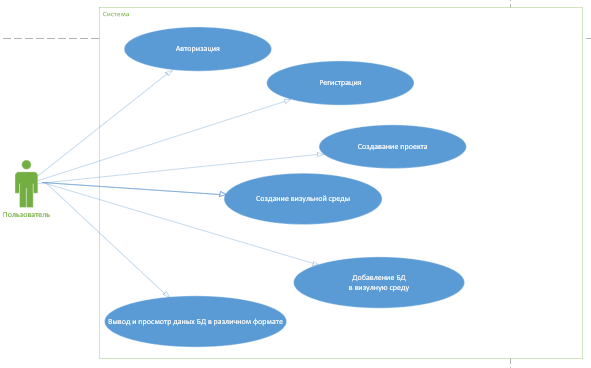


Рисунок 1 – диаграмма прецедентов

# **3 Создание приложения**

На основе требований к программному продукту разработано веб-приложение, которое позволяет отправлять запросы к базе данных Trino и получать ответы на основе обработки текста с помощью искусственного интеллекта.

Приложение принимает два параметра запроса: "trino\_id" и "question". Затем оно получает учетные данные для подключения к базе данных Trino из сервиса Vault, после чего подключается к базе данных и выполняет SQL-запрос, который возвращает схему таблиц.

Загружая данные из двух текстовых файлов, оно формирует вопрос, включая схему таблиц и вопрос, указанный в параметрах запроса. Полученный вопрос обрабатывается с использованием модели обработки текста ChatOpenAI, которая возвращает ответ.

Основная логика программы находится в файле main.py, код которого описан в листинге 1.

Листинг 1 – основной модуль

**import** **httpx**

**import** **hvac**

**from** **fastapi** **import** FastAPI, Response

**from** **httpx\_socks** **import** SyncProxyTransport

**from** **langchain\_core.prompts** **import** ChatPromptTemplate

**from** **langchain\_openai** **import** ChatOpenAI

**from** **trino.dbapi** **import** connect

**from** **src.conf** **import** VAULT\_URL, VAULT\_KEY, OPENAI\_KEY, PROXY\_URL

app = FastAPI()

**@app.get**("/")

async **def** **root**():

**return** {"message": "Hello World"}

**@app.get**('/sql')

async **def** **sql**(trino\_id: str, question: str):

client = hvac.Client(

url=VAULT\_URL,

token=VAULT\_KEY,

)

trino\_creds = client.secrets.kv.read\_secret\_version(path=trino\_id)['data']['data']

conn = connect(host=trino\_creds['host'],

port=trino\_creds['port'],

user=trino\_creds['user'],

)

cur = conn.cursor()

cur.execute('SHOW schemas')

cur.execute(f'SHOW TABLES in {trino\_creds["catalog"]}.public')

tables = cur.fetchall()

schema = ''

**for** table **in** tables:

schema += cur.execute(f'SHOW CREATE TABLE {trino\_creds["catalog"]}.public.{table[0]}').fetchall()[**0**][**0**] + '**\n**'

cur.close()

conn.close()

**with** open("static/prompt/system.txt") **as** f:

system = f.read()

**with** open("static/prompt/prompt.txt") **as** f:

user = f.read()

user = user.format(tables=schema, question=question)

llm = ChatOpenAI(openai\_api\_key=OPENAI\_KEY,

http\_client=httpx.Client(transport=SyncProxyTransport.from\_url(PROXY\_URL)))

prompt = ChatPromptTemplate.from\_messages([

("system", system),

("user", "{input}")

])

chain = prompt | llm

**return** {'sql': chain.invoke({"input": user}).content, }

**return** Response(content=chain.invoke({"input": user}).content, media\_type="application/xml")

При отправке запроса к этому веб-приложению, указывается два параметра: "trino\_id" и "question". Веб-приложение использует библиотеку hvac для доступа к сервису Vault. Оно запрашивает учетные данные для подключения к базе данных Trino из хранилища секретов Vault. После получения учетных данных приложение подключается к базе данных Trino и выполняет SQL-запрос, который возвращает схему таблиц. Далее приложение загружает два текстовых файла, содержащих информацию о системе и промпт для пользователя. Используя данные из этих файлов, оно формирует вопрос, включая схему таблиц и вопрос, который вы указали в параметрах запроса. Сформированный вопрос отправляется в модель обработки текста ChatOpenAI, которая обрабатывает его и возвращает ответ. Результат работы модели возвращается в качестве ответа на ваш исходный запрос.

В файле conf.py, код которого описан на листинге 2 находится переменные, которые используются для хранения конфиденциальной информации или настроек, таких как URL-адреса API, токены аутентификации и другие конфигурационные параметры. Использование переменных среды позволяет изолировать конфиденциальные данные от кода и сохранить их в безопасности.

Листинг 2 – модуль хранения конфиденциальный данных

**import** **os**

**from** **dotenv** **import** load\_dotenv

load\_dotenv()

VAULT\_URL = os.environ.get("VAULT\_URL")

VAULT\_KEY = os.environ.get("VAULT\_TOKEN")

OPENAI\_KEY = os.environ.get("OPENAI\_KEY")

PROXY\_URL = os.environ.get("PROXY\_URL")

На листинге 3 описан код dockerfile. Dockerfile используется для создания Docker-образа, который содержит окружение Python с установленными зависимостями для запуска приложения. Docker-образы представляют собой контейнеры, которые упаковывают в себе все необходимое для работы приложения, включая код, зависимости и окружение выполнения.

Листинг 3 – dockerfile

FROM python:**3.11**-slim

RUN pip install poetry

WORKDIR /app

RUN pip install httpx hvac fastapi httpx-socks langchain-core langchain-openai trino

COPY pyproject.toml /app/

RUN test -f pyproject.lock && COPY pyproject.lock /app/ || true

RUN poetry install --no-dev

COPY . /app/

RUN adduser -u **5678** --disabled-password --gecos "" appuser && chown -R appuser /app

USER appuser

CMD ["poetry", "run", "python", "src/main.py"]

В данном dockerfile происходит установка инструментов управления зависимостями Poetry, который помогает управлять зависимостями Python-проекта и создавать виртуальные среды, а также устанавлива зависимости проекта, указанные в файле pyproject.toml. Далее копируются файлы проекта внутрь контейнера. Затем создается пользователь appuser, чтобы запускать приложение от имени этого пользователя после чего запускается команда по умолчанию для выполнения приложения.

Таким образом, это веб-приложение позволяет вам легко и удобно выполнять SQL-запросы к базе данных Trino и получать ответы на ваши вопросы с помощью искусственного интеллекта.

# **4 Аналоги и прототипы**

На основе выделенного функционала были выделены и рассмотрены 2 основных аналога: Tableau и Power BI.

Tableau — это одна из ведущих платформ визуализации данных, предназначенная для создания интерактивных дашбордов, отчетов и аналитических приложений. Она позволяет пользователям легко анализировать данные и делиться результатами исследований с коллегами.

Tableau предоставляет широкий набор инструментов для визуализации данных, включая различные типы графиков, таблиц, карт и других элементов. Пользователи могут легко создавать интерактивные дашборды, перетаскивая и отпуская данные для быстрого анализа. Tableau интегрируется с различными источниками данных, такими как базы данных, облака, файлы Excel и многие другие. Это позволяет пользователям легко анализировать данные из различных источников в единой среде.

Power BI — это, в свою, очередь бизнес-аналитическая платформа, разработанная Microsoft, которая позволяет пользователям создавать отчеты, дашборды и интерактивные приложения для анализа данных. Power BI является частью семейства продуктов Microsoft Office и обеспечивает глубокую интеграцию с другими сервисами компании.

Power BI предоставляет пользователю мощные инструменты для визуализации данных и анализа трендов. С его помощью можно создавать интерактивные дашборды, делиться отчетами и проводить анализ данных на основе множества источников. Power BI также предоставляет возможности аналитики на основе ИИ для автоматического выявления важных трендов и прогнозирования. Power BI интегрируется с другими продуктами и сервисами Microsoft, такими как Microsoft Excel, Azure и Dynamics 365. Это позволяет пользователям легко анализировать и объединять данные из различных источников, создавать мощные аналитические приложения и делиться результатами с коллегами.

5 Подготовка тестовых сценариев

В процессе разработки бизнес-аналитической платформы необходимо провести ряд тестовых сценариев, чтобы убедиться в корректной работе функциональности и соответствии требованиям проекта. Ниже, в таблице 1, приведены примеры тестовых сценариев для различных компонентов приложения:

Таблица 1 – Тестовые сценарии

|  |  |  |
| --- | --- | --- |
| **Ввод** | **Ожидаемый результат** | **Фактический результат** |
| Анализ данных из базы данных | Данные успешно загружаются и отображаются на графиках и диаграммах | Данные успешно загружаются и отображаются в графическом виде |
| Создание отчета о продажах | Отчет содержит информацию о продажах за выбранный период с детализацией по различным параметрам, таким как продукты, регионы и каналы продаж | Отчет успешно создается и содержит необходимую информацию о продажах |
| Прогнозирование спроса | Модель успешно анализирует предыдущие данные и создает прогноз спроса на будущий период с учетом выбранных параметров | Модель успешно создает прогноз спроса на основе предыдущих данных и выбранных параметров |
| Анализ эффективности маркетинговых кампаний | Аналитический отчет содержит информацию о результативности различных маркетинговых кампаний, включая ROI, конверсию и другие метрики | Отчет успешно анализирует данные о маркетинговых кампаниях и предоставляет информацию о их эффективности |

6 Вывод

Бизнес-аналитическая платформа является важным инструментом для организаций любого масштаба, помогая им принимать обоснованные решения на основе данных.

Современные организации все больше осознают важность данных для принятия стратегических решений. Бизнес-аналитическая платформа предоставляет им необходимые инструменты для сбора, анализа и визуализации данных.

Платформа должна быть гибкой и способной адаптироваться к различным потребностям бизнеса. Это включает в себя возможность интеграции с различными источниками данных, создание разнообразных отчетов и аналитических дашбордов, а также наличие инструментов для проведения прогнозирования и моделирования.

Важным аспектом платформы является обеспечение безопасности данных. Организации должны иметь возможность контролировать доступ к данным и обеспечивать их защиту от несанкционированного доступа и утечек.

В целом, бизнес-аналитическая платформа играет важную роль в современном бизнесе, обеспечивая организациям доступ к данным и аналитическим инструментам для принятия стратегических решений