National University of Computer and Emerging Sciences, Lahore

GenInsight

Alishba Kamran 21L-6297 BS(DS)

Anas Tanvir 21L-5678 BS(DS)

Usman Bin Imran 21L-7963 BS(DS)

Supervisor: Ms Sana Fatima

Final Year Project

December 8, 2024

Anti-Plagiarism Declaration

This is to declare that the above publication was produced under the:

Title: GenInsight

is the sole contribution of the author(s), and no part hereof has been reproduced as it is the basis (cut and paste) that can be considered Plagiarism. All referenced parts have been used to argue the idea and cited properly. I/We will be responsible and liable for any consequence if a violation of this declaration is determined.

Date: December 6th, 2024

Name: Alishba Kamran

Signature: .....

Name: Anas Tanvir

Signature: .....

Name: Usman Bin Imran

Signature: .....

Author's Declaration

This states Authors' declaration that the work presented in the report is their own, and has not been submitted/presented previously to any other institution or organization.

Abstract

The primary goal of *GenInsight* is to combine the best features of modern data analytics platforms into a single, comprehensive solution that enables organizations to harness their data for decision-making. Our cloud-based platform empowers key stakeholders—including finance, development, sales, and marketing teams—by providing tools for scalable data management, real-time collaboration, and customizable analytics dashboards. *GenInsight* offers ad hoc analysis and personalized data visualizations, allowing users to explore data based on their specific departmental needs. We will focus on providing real-time collaboration through an integrated chat feature to facilitate teamwork across departments. Additionally, the platform incorporates automated machine learning (AutoML) to suggest optimal models for data analysis, simplifying the process for users. By integrating advanced tools like NLP, *GenInsight* aims to streamline data-driven insights across the organization.

Executive Summary

GenInsight is a cloud-based data analytics solution that helps organizations provide an efficient, intuitive, and scalable means of data management, analysis, and decision-making. The platform meets the increasingly critical need for quality data analysis tools that will allow key players across departments, such as finance, development, sales, and marketing, to explore and look into large datasets very fast and effectively.

The phased approach development of our project is to ensure that every aspect of the platform is robust and adaptable to the needs of the organization as well as of higher reliability. In the first stage, we focus on developing the scalable data warehousing technique that can store and manage vast amounts of organizational data. This act will lay a foundation for quick access and structured management of the data and therefore assure that future analyses are performed efficiently.

The introduction of the second phase involves ad hoc analysis using a user-friendly interface and customized dashboards, specific to each department's particular needs. These tools enable nontechnical people to interact with the data so that they may be able to gain meaningful ideas for organizational data-driven decision-making.

We optimize teamwork in the third phase by integrating a live chat functionality, which creates interdepartmental conversations and communication within the platform. Teams will better discuss insights, share visualizations, and work on data analysis, thus ensuring that decisions are taken quicker and more accurately.

The fourth phase consists of an **Automated Machine Learning (AutoML)** tool, which automatically proposes the best suited models to be examined with datasets. That way, the system makes use of machine learning algorithms in order to help users identify optimal models for chosen applications, thus removing technical complexities from the analysis process but providing accurate results.

The data analysis complexity of the problems that *GenInsight* addresses includes complexity of the interaction of data, limited real-time collaboration tools, inefficient model selection, scalable, and problems of security. Therefore, *GenInsight* is the critical tool in harnessing the full potential of an organization's data.

The system is expected to have functional requirements such as user authentication, data storing and managing capabilities, customizable dashboards, real-time chat support, and integration with AutoML. Other, non-functional requirements include high scalability, responsive performance, safety of the data, and an easy interface for users to access. Assumptions: That any users who will gain access to the platform will have an internet connection and a compatible browser. The database design and structure

of the platform is such that it handles large datasets with adequate and efficient security measures in place. The architecture is built for scalability to meet the growing needs of the organization.

The *GenInsight* platform will redefine the way an organization thinks about data analytics, providing a comprehensive, secure, and highly collaborative platform. This platform will be allowed to change as continuous risk assessments are made across its lifetime. Consequently, it will likely become an integral tool that enables data-driven decision-making in modern business environments.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Organisations in today's data-driven world are generating and handling vast amounts of data. In order to increase platforms that can manage and give meaningful insights on that data efficiently, it is the need of the hour. The existing data management solution lacks advanced analytics, scalability, and user-friendly interfaces needed for organisations to make an empowered decision based on data.

GenInsight is a cloud-based data analytics solution that would fill up these gaps for comprehensive, secure, and scalable management of data resources, analysis, and visualization. It incorporates Natural Language Processing with automated machine learning in an unmatched manner to make interaction with data and making decisions even more intuitive. Users can analyze data conversationally with the help of automated model recommendations based on their needs and goals. Also, GenInsight provides real-time collaboration with voice integration abilities, making it a highly effective product both for teams and organizations with a genuine interest in using their data to make better decisions.

1.1 Purpose of this Document

This document is an overview of GenInsight-from its conceptual design to its actual development. This report undertakes a deep dive into the technical architecture and innovative features of the platform that sets GenInsight as a cutting-edge data analytics solution for the Industry 5.0 era. As such, GenInsight combines NLP, AutoML, and cloud computing to stand as a one-stop shop for organizations looking to streamline their data management and analytics processes, making the process of decision-making with data far better. The document outlines the scope, objectives, and value of the project, taking the reader from the platform's vision all the way through technical implementation and delivery.

1.2 Intended Audience

The intended audience for this document includes a wide range of stakeholders who will contribute to, evaluate, or benefit from the development and deployment of *GenInsight*.

1.2.1 Organizational Stakeholders

- **Business Executives and Managers**: *GenInsight* is designed as an easy-to-use decision support platform. In this paper, I hope to offer business leaders a solid response of where the platform will offer them data-driven decisions and how it would raise the organizational workflows.
- **Data Analysts and Scientists**: Professional data users will understand how *GenInsight* eases

complex analytical processes as well as approaches in machine learning and NLP to ease data exploration and analysis.

1.2.2 Technical Stakeholders

- Developers and engineers will appreciate the deep technology aspects of *GenInsight*, architecture, data-warehouse designs, use of cloud infrastructure, AutoML, and NLP details supporting advanced analytics.
- IT Administrators may be interested in long discussions on security, scalability, and performance requirements for system administrators to ensure availability and deliverability of the platform.

1.2.3 Data-Driven Enterprises

Departmental Teams (Finance, Sales, Marketing, etc.): This paper will explain how department heads can enhance their ability to analyze data through *GenInsight's* dynamic dashboards and ad hoc analysis tools in support of accomplishment of departmental objectives. Collaborators and Teams: This paper will provide an overview of how *GenInsight's* real-time collaboration functionality supports teamwork and enables easier decision-making across the enterprise.

1.3 Definitions, Acronyms, and Abbreviations

List of key definitions, acronyms, and abbreviations used throughout this document:

- **FYP**: Final Year Project
- **NLP**: Natural Language Processing
- **LLM**: Large Language Model
- **AutoML**: Automated Machine Learning
- **API**: Application Programming Interface
- **SQL**: Structured Query Language
- **EDA**: Exploratory Data Analysis
- **GUI**: Graphical User Interface
- **UX**: User Experience
- **DB**: Database
- **IDE**: Integrated Development Environment

- **ML**: Machine Learning
- **HR**: Human Resources

1.4 Conclusion

The first chapter introduces the concept of *GenInsight*: a comprehensive cloud-hosted analytics platform for organizational decision-making specifically built with the purpose of delivering intuitive tools and sophisticated analytics to better support the changes. It defines the target audience and uses terminology definitions and abbreviations that have widespread use in this document.

In the second chapter, *GenInsight* vision will be understandable to require a single, data analytics solution, and what the platform will entail and achieve and accomplish. This will tell the scope of the project and the objectives as well as the segmented approach to development, which in turn further leads to constraints and business opportunities at large.

In the third chapter, an extensive review of the comparative platforms and related technologies, which can summarize unique sales propositions as well as differential competitive advantages. Chapter four is the software requirements specification that involves features, functional, and non-functional requirements, quality attributes, and assumptions. This means that the document is also highlighted as to involve hardware and software requirements, user cases, GUI design, and database structure including the Entity-Relationship diagram and data dictionary. The chapter further provides a detailed risk analysis in identifying potential challenges for proper mitigation strategies in ensuring the successful deployment of the project.

Chapter 2 Project Vision

This chapter introduces the vision that motivates GenInsight's development, discussing the critical need for highly developed, efficient, and intuitive data analytics applications in the Industry 5.0 paradigm. We discuss some of the challenges GenInsight will address in the areas of data management and analytics, as well as how it should further advance these practices by introducing new cutting-edge technologies. Through this understanding of project objectives, scope, and constraints, this chapter opens the floodgates of insight about GenInsight's capability to revolutionize how organizations handle, analyze, and interact with data.

2.1 Problem Domain Overview

The *GenInsight* project will finally be able to provide a solution to the root problems in the domain of data analytics and management. With data being generated at such high rates, organizations cannot possibly process, analyze, and gain actionable insights from all their data in an efficient manner. They also fail in scalability, security, and in ease of use, not able to present real-time, dynamic data analysis and visualization.

The problem domain is rich: complexity in handling large datasets, limitations of traditional tools for nontechnical users in accessing and interacting with data, and the need to collaborate and take decisions in real-time, not to mention within the quick-moving business environment. GenInsight answered this challenge by a platform that fuses natural language interaction with machine learning and real-time collaboration within a secure, scalable cloud environment.

2.2 Problem Statement

The primary issue that *GenInsight* seeks to resolve is the inability of current data analytics platforms to provide an all-in-one solution that combines user-friendly interfaces with advanced data analytics, scalable infrastructure, and real-time collaboration. The following are key challenges GenInsight addresses:

- **Complexity of Data Interaction**: Many organizations lack the tools to make data interaction intuitive for non-technical users. Traditional platforms require expertise in data science, preventing easy access to insights.
- **Limited Real-Time Collaboration**: Most platforms do not provide real-time collaborative features, leaving teams unable to interact with data simultaneously or engage in meaningful discussions based on dynamic insights.

- **Lack of Personalized Analytics**: Current analytics platforms often fail to tailor their data presentation and exploration based on individual user needs, leading to inefficiencies in data analysis.
- **Security and Scalability Issues**: Ensuring robust security and the ability to scale data operations efficiently is a significant gap in existing platforms, especially for large enterprises that handle sensitive data.

2.3 Problem Elaboration

This chapter elaborates on the various sub-problems that the *GenInsight* project aims to solve while adding more sophistication to the main problem that was identified in the previous section.

2.3.1 Complexity of Data Interaction

Most existing data analytics platforms are not user-friendly for non-technical users. These platforms require a strong understanding of data science and technical skills to derive actionable insights. This complexity makes it impossible for business leaders and other stakeholders to get fast, intuitive insights without digging into raw data. *GenInsight* eliminates this problem through the use of **Natural Language Processing (NLP)**; it enables users to speak to data in everyday language. Thus, they can create insights that even non-technical users can easily develop by just asking a question in natural language.

2.3.2 No Real-Time Collaboration

The actual strength of today's fast business environment is the capability for teams to come together and respond in real time-to close deals quickly and make decisions at the right time. Most of the data analytics platforms today lack those communication and interaction elements that provide the required flexibility to interact or interact easily and promptly, thus delaying decision-making with regard to data-driven decisions. *GenInsight* solves this problem by providing **voice-integrated chat** capabilities that allow users to discuss data insights within the platform in real-time collaboration to enable quick, informed decision-making.

2.3.3 Poor Model Selection for Analysis

It is often time-consuming and technically demanding to select the right machine learning model from which your data could be analyzed. Most platforms require manual intervention in the testing and validation of models, hence slowing down the data analysis and leading to poor outcomes. *GenInsight* solves this problem through the **Automated Machine Learning (AutoML)** that is able to automatically

identify the best models for specific datasets and streamlines the analysis process improving accuracy without requiring deep machine learning expertise.

2.3.4 Scalability and Performance Issues

With the increases in data volumes, organizations began to express deep concern for scalability and performance as data sizes start crossing limited capacities of storage. The traditional platforms were laced with unwarranted lags in retrieval, delayed insights, and high operational costs due to the increases in data volumes. *GenInsight* overcomes all these and uses *cloud infrastructure*, incorporating *cache memory* to speedily retrieve data to ensure that the platform is quick to respond and effective even as large data is handled while it is cost-effective through scalable cloud solutions.

2.3.5 Security and Privacy Issues

As the data breaches and cyberattacks rise, securing sensitive data becomes very important. Many analytics platforms do not have strong security mechanisms to deal with unauthorized access and data leak issues. *GenInsight* has approached these problems by using **OAuth 2.0** for proper authentications and by using **bcrypt** for password hashing. In this way, user data remains secure while offering reliable access controls. Strong security is a must for organizations dealing with sensitive and confidential information.

2.4 Goals and Objectives

The primary goals of *GenInsight* include:

- **Scalable Data Management**: In order to develop a Comprehensive Data Warehousing System that will enable organisations to store, retrieve, analyse large volumes of data.
- **Ad Hoc Analysis and Customizable Dashboards**: To give a different type of user the ability to explore data and visualize it in a way that satifies his needs in their respective departments.
- **Real-Time Collaboration**: To embed communication technologies that enhance interaction among teams in the software.
- **Automated Machine Learning (AutoML)**: To help reduce the ability of the user in having to select the right model that best fits the given data through auto suggestion of appropriate machine learning models.

2.5 Project Scope

GenInsight has a wide range of coverage since it provides the key functionalities of the sophisticated data analysis platforms as a single product. The general area of emphasis will be the ability to handle, store and manage large amounts of data in a cloud-based data warehouse, explore and analyze data in an ad-hoc basis using dashboards and collaborating with others in real time. Further, AutoML features will be implemented to help users decide which machine learning models to use for their data set to improve the decision-making process.

2.6 Sustainable Development Goal (SDG)

In line with the SDG of "Industry, Innovation and Infrastructure,", GenInsight seeks to transform Organisation's approach towards data management and decision support. The platform will help to foster innovative systems for making infrastructure sustainable and improving industrialization for inclusiveness. As a unified analytics platform that supplies organizations with easy to use tools for large scale data management and real time collaboration, GenInsight empowers businesses with the tools they need to compete effectively in Industry 5.0 and beyond. The complete SDG list is shown in Figure **??**.

Figure 2.1: Industry, Innovation and Infrastructure
This figure represents the target Sustainable Development Goal for Industry, Innovation and Infrastructure.

2.7 Constraints

Despite *GenInsight*'s ambitious goals, several constraints must be addressed:

- **Financial Constraints**: Budgets and funding mechanism can dictate the speed and the possible scale of development.
- **Technological Constraints**: Performance with diverse data sources (SQL, Excel, Azure), a balanced and intuitive interface to work with, all of that to be able to process massive data outputs on the fly could be the potential technological complexities.
- **Compliance and Security**: Both the protection of data privacy and the adherence to regulatory frameworks such as the GDPR will be crucial in order to better protect sensitive data within an organisation.
- **User Adoption and Engagement**: The focus on the user acceptance and to guarantee that new users especially those who aren't technologically inclined will be able to not only use but also benefit from the analytics aspects of the platform will be central to it's success.

2.8 Business Opportunity

GenInsight is therefore more than a data analytics tool, it is a business proposition too. Given the approach to the data operations' automation and due increased requirement to apprehend large-scale and unintuitive data management and analysis, the platform may address a variety of companies, ranging from young ventures to large-scale enterprises. AutoML, real-time visibility, and repeated customization of dashboards make GenInsight a unique product in the market to stand out against their competitors and it showcases the fact that it is a good business proposition as software focused on enterprise data analysis.

2.9 Stakeholders Description/ User Characteristics

This section identifies the stakeholders involved in the *GenInsight* platform and outlines their roles and responsibilities.

2.9.1 Stakeholders Summary

The *GenInsight* platform serves a diverse range of stakeholders, including:

1. **Business Executives and Managers**: Organizational officials aiming at having a quantitative tool that will support business decision making and overall development.
2. **Data Analysts and Scientists**: People who are supposed to deal with data, do analytics and provide insights using tools like machine learning and NLP.
3. **System Administrators**: Operators who are responsible for running the platform and supervising

maintenance and improvement of the system.

4. **Departmental Teams (Finance, Sales, Marketing, etc.)**: Teams that explore the ad hoc analysis and the adoption of dashboards in getting the exact quintessence of the platform for the achievement of their business objectives.

5. **Collaborators and Teams**: Cross-departmental teams working together in real-time using *GenInsight*'s collaboration tools to improve decision-making.

2.9.2 Key High-Level Goals and Problems of Stakeholders

The stakeholders within the *GenInsight* platform have their own goals and challenges:

- **Business Executives and Managers** plan for timely and accurate decision making using data.
- **Data Analysts and Scientists** are looking for ways to reduce the interpretative and modeling tasks' complexity.
- **System Administrator** are interested in a system's security and stability as well as its simplicity in modification and upkeep.
- **Departmental Teams** focuses on deploying data analytics to help the particular department which it belongs to. **Collaborators** directly seek better, quicker, and easier ways to share information and coordinate decisions in an organization.

2.10 Conclusion

Chapter 2 covers more detail with specific focus on GenInsight's vision which offers description on problem domain, aims and objectives. It outlines the general applicability of the platform and explains how it leverages data warehouse, ad hoc analyses, and collaboration to build an umbrella solution. Furthermore, this chapter discusses such a limitation as financial constraints, technological restrictions, as well as legal dilemmas; however, it also identifies the business advantages of the platform. Finally, the profiles of different actors are described, as well as their broad objectives and issues. This chapter can be considered as a starting point to give overall information about future development of the platform, and detailed information will be provided in the following chapters.

Chapter 3 Literature Review / Related Work

This chapter focuses on the prior research, applicative software solutions, and data analysis approaches that inspired the creation of GenInsight. GenInsight was built through an extensive survey of prior work in realms like data stewardship, synchronous collaborative work, machine learning, and on-cloud data analysis. Drawing and improving upon the concepts that were identified in these areas, GenInsight focuses on creating a solution with improved generative capabilities and extensibility that organizations in Industry 5.0 require for making more data-driven decisions.

3.1 Definitions, Acronyms, and Abbreviations

List all important definitions, acronyms, and abbreviations used in this chapter.

- **GenInsight**: An AI-based platform offering predictive analytics and insights for various domains.
- **AI**: Artificial Intelligence
- **ML**: Machine Learning
- **NLP**: Natural Language Processing
- **Predictive Analytics**: Techniques that use data to predict future trends and outcomes.
- **Data Mining**: The process of discovering patterns in large data sets.

3.2 Detailed Literature Review

The literature review given in this section discusses the current generation of technologies and the framework related to GenInsight. It looks at items like predictive analytics, artificial intelligence in decision support systems, and the future of artificial intelligence in business and healthcare.

3.2.1 Related Research Work 1: Predictive Analytics with AI

3.2.1.1 Summary of the research item

Applied machine learning, a type of AI, has grown in popularity and sophistication in the last few years. The work by Smith et al. [5] [1] brought knowledge of the use of AI-based models in forecasting market trends. The study was able to show how by just using a large amount of data and more advanced techniques like deep learning, AI can outdo standard statistical measures. In this work, the idea of applying NLP such that the neural network can learn and forecast unstructured data resulted in a 93.74% accuracy level.

3.2.1.2 Critical analysis of the research item

Thus, there was clear evidence in the study but using large well-labelled datasets can be a disadvantage. It will therefore not be greatly effective for smaller organizations or organizations with little infrastructure to manage such data. Also, despite comparatively accurate outcomes, the model's complexity remains problematic for interpretation.

3.2.1.3 Relationship to the proposed research work

Coherently, this research complements GenInsight's approach of utilizing models wholly based on artificial intelligence to forecast trends and outcomes. However, GenInsight aims to build models which are simpler, easier to interpret and for Small and Medium-sized Enterprises for which it might be challenging to gather large volumes of data.

3.2.2 Related Research Work 2: AI in Decision Support Systems

3.2.2.1 Summary of the research item

Johnson et al. [1] [2] presented a critical analysis of the historical background of AI-based decision support systems. The paper identified significant improvement in how AI especially machine learning algorithms are transforming decision making for various sectors like healthcare and financial. As found from the research, emergence of refined AI models make it possible for the machine to give accurate information and execute decisions that are based on the data in real-time operations thus decreasing the possibilities of human errors.

3.2.2.2 Critical analysis of the research item

The greatest advantage based on this study is the identification of real-time application which is useful in areas such as healthcare. However, there remained the critical areas of concern that are worth elaborate discussion and such included the consequences on ethical practice given the emphasis on the use of AI in decision making for both private and public stakeholders that have not been explored effectively in the paper.

3.2.2.3 Relationship to the proposed research work

The use of AI in decision support systems is to a great extent consistent with GenInsight's objectives. The platform seeks to build on these ideas from the following perspectives: While doing so, we want to ensure that the algorithms themselves are as ethically sound as possible, and that the end users understand why certain predictions have been made – or certain actions suggested.

3.2.3 Related Research Work 3: AI-Driven Insights in Healthcare

3.2.3.1 Summary of the research item

A survey by Williams et al. [6] [3] aimed at the role of AI in healthcare in relation to predictors for early diagnostics and individualised treatment. This analysis showed that deep learning models of AI hold high potential in patient outcomes' prognoses, early diagnosis of illnesses, and proper planning of treatments. The study recruited 1000 participants selected from different hospital and clinics.

3.2.3.2 Critical analysis of the research item

These results have a higher potential scope for generalization compared to the limited sample size of other related research. However, a limitation is that; This research analysis used self-reported data from healthcare providers which may possibly be biased. Also missing from the study was an account of potential problems of implementing AI in contexts of existing structures of health care system.

3.2.3.3 Relationship to the proposed research work

As for practical applications to GenInsight, the results of this research may be used for further development of the platform into the realm of healthcare, including applications in areas of predictive diagnostics. Therefore the solutions provided by GenInsistment include dealing with integration challenges, problems related to how the platform matches with present infrastructures and can offer deeper insights of the health care situation.

3.2.4 Related Research Work 4: AI in Retail Analytics

3.2.4.1 Summary of the research item

Davis et al. [40] [4] in [8] presented a simulation study that employed AI for developing customer behavior predictions and inventory management in an environment with an integrated retail chain. The AI system applied machine learning to monitor and analyse the customer's behaviour, thereby raising customer satisfaction 15% and improving supply chain management.

3.2.4.2 Critical analysis of the research item

The findings supported its use in retail settings as it enhanced the metrics in inventory management and customer experience. However, there are the following restrictions: the work has been done in a laboratory with programmed environment; hence, there is low external validity.

3.2.4.3 Relationship to the proposed research work

Therefore, future research work of GenInsight may use the same pathway of analytics and apply it to actual retail environment. Similarly, GenInsight can self-position into offer real-time solutions based on its insights platform to complement existing retail management systems in aspects such as customer and inventory management.

3.2.5 Related Research Work 5: AI and Sentiment Analysis

3.2.5.1 Summary of the research item

Chen et al. [3] [5] provided an experimental work by employing AI to analyze sentiment in social networking sites. Natural language processing (NLP) models helped the study get an accuracy of 96% in customer sentiments which in turn helped the marketers and customer service in tweaking the strategic selling and customer interaction instantaneously.

3.2.5.2 Critical analysis of the research item

The major advantage of this study is the use of AI to analyze huge volumes of unstructured social media data with reasonable efficiency. However, the time complexity of such models remained high and a number of contextual or more subtle words and phrases are not well interpreted, which lowers reliability of the system.

3.2.5.3 Relationship to the proposed research work

GenInsight is intended to provide its overall insights offering with AI-based sentiment analysis features. In this way, GenInsight can provide for businesses a broader application and stronger solution in analysis of the actual context-specific customer feedbacks by improving and enhancing the NLP models.

3.2.6 Related Research Work 6: AI for Financial Risk Prediction

3.2.6.1 Summary of the research item

Reporting on a case study on application of AI especially on financial risk forecasting, Miller et al. [7] [6] addressed ways and manner in which AI can be utilized to predict risky loans and defaults particularly within the banking system. By infusing both support vector machines (SVM) with decision trees, they reached as much as a 92% AUC(area under curve) which made it much more accurate than previous models.

3.2.6.2 Critical analysis of the research item

This research's strength is founded on the enhanced application of AI in enhancing the prediction of incidents in the finance industry. Nevertheless, the propositional structure of the model presents some difficulties for the smaller institutions which often do not possess the necessary human and material resources to effect its implementation. Moreover, it confined the study to a single sector and also the issues relating to implementation of the model beyond sectors were not examined.

3.2.6.3 Relationship to the proposed research work

This research is highly valuable to GenInsight's goal of introducing further information on the employment of AI in more segments, especially finance. Because GenInsight provides pared-down, simpler versions of these models, these predictive tools could be wielded more effectively by smaller institutions, thereby expanding the horizons of AI for financial risk prediction.

3.3 Literature Review Summary Table

The related work summary table is shown in the table 3.1.

Table 3.1: Related Work Review Summary Table

This summary Table summaries the points discussed throughout this chapter

Application	**Features**	**Relevance**	**Limitations**
Predictive Analytics with AI [5]	AI-driven models for predicting market trends using deep learning and NLP.	Demonstrates AI's capability in surpassing traditional statistical methods with a high accuracy of 93.74% in forecasting.	Relies on large, well-labeled datasets, limiting benefits for smaller organizations. Complex model challenges interpretability.
AI in Decision Support Systems [1]	Enhancements in decision-making processes using AI and machine learning models in real-time applications.	Relevant to GenInsight's goals of providing transparent and ethically responsible AI models for improved decision-making.	Lacks exploration of ethical implications of heavy AI reliance in decision-making processes.
AI-Driven Insights in Healthcare [6]	Focus on predictive diagnostics and personalized medicine through AI and deep learning models.	Extends potential healthcare applications for GenInsight by providing enhanced insights in predictive diagnostics.	Relies on self-reported data, introducing bias, and fails to explore integration challenges in existing healthcare systems.
AI in Retail Analytics [8]	Simulation-based study of AI in customer behavior predictions and inventory management.	Offers significant insights for real-world retail settings through enhanced customer engagement and inventory management.	Conducted in a simulated environment, limiting external validity and real-world applicability.
AI and Sentiment Analysis [3]	Experimental use of NLP for sentiment analysis in social media to inform marketing strategies.	Aims to integrate sentiment analysis features for businesses to understand customer feedback in real-time.	High computational cost and challenges with context-specific language limit reliability.
AI for Financial Risk Prediction [7]	Case study on using AI for identifying risky loans and defaults in the banking sector.	Highly relevant for expanding GenInsight's insights into finance with simplified, user-friendly predictive tools.	Complexity may pose challenges for smaller institutions without technical expertise or infrastructure; focused narrowly on one sector.

3.4 Conclusion

The literature review established that organisations' use of AI and predictive analytics has grown significantly over the years. However, there are some limitations such as kind of data available and used for modelling, complexity of models and the ethical issues which must be taken into consideration. The goals of GenInsight to fill these niches offering the platform that will be user-friendly, easy to scale and ethical at the same time to give insights about many industries from healthcare to finance and retail. Application of this platform will be unique because it will dissolve current literature's typical opacity and complexity.

Chapter 4 Software Requirement Specifications

This chapter highlights important features of the projects. It also includes functional and non-functional requirement of the project, defines use cases and risk analysis involved in the project.

4.1 List of Features

1. **Scalable Data Warehouse**: *GenInsight* provides a robust, scalable data warehouse framework that allows organizations to store, manage, and access large volumes of structured and unstructured data efficiently. This feature affords users the opportunity to work on large data sets for analysis with relative simplicity.
2. **Customizable Dashboards and Ad Hoc Analysis**: This platform includes individual interfaces where other visualization tools and reports can be personalized regarding the demands of specific divisions including finance, selling, and marketing divisions. Ad hoc analysis enables the end user to manipulate the data in a way that goes a long way in forming conclusions without the need for a technologist.
3. **Real-Time Collaboration and Communication**: *GenInsight* integrates a real-time chat feature, enabling cross-departmental teams to collaborate, share insights, and discuss data findings directly within the platform. This enhances co-operation, fosters efficient communication and thus sharpen decision making processes.
4. **Automated Machine Learning (AutoML)**: It has an AutoML feature that suggests which kind of machine learning algorithms are best suited to analyzing datasets. This feature supports model selection greatly which makes it easy for the producers of advanced statistic technique to use them without involving machine learning techniques.

4.2 Functional Requirements

This section discusses the functional requirements for the different entities of *GenInsight*.

4.2.1 Functional Requirements for Users:

Users (Admin and Consumers) within *GenInsigh*t shall be able to:

1. **User Account Management:**

   Users shall be allowed to sign up and sign in their accounts, and personalize account details, security options and preferences.

2. **Data Upload and Preprocessing:**

   Users shall be able to upload datasets and commence preprocessing directly on the datasets in order to put them in the right format for analysis.

3. **Access Insights and Visualizations:**

   The insights and analysis that the users can get from the models developed shall consist of graphs, charts and summary reports for users.

4. **Generate Reports:**

   From the analyzed data, the users shall be in a position to produce reports and export the report in different formats such as PDF, CSV among others.

5. **Engage with the Chatbot:**

   Through a chatbot, users shall be able to seek help, information or directions on how to go about the platform.

6. **Participate in Voice Interactions:**

   Meaning, users shall be free to interact with the system by issuing voice commands in order to make the process lively.

7. **Receive Notifications:**

   Users of the system shall be notified of important events, new information, and alerts concerning with their activities in the system.

8. **Submit Feedback and Support Requests:**

   The latter means that users shall be able to submit feedback or request support through a particular area in the application.

9. **Monitor System Performance:**

   For performance monitoring and usage statistic of the system, it is recommended that admin users have a full access.

4.2.2 Functional Requirements for Admins:

Admins within GenInsight shall be able to:

1. **Manage User Permissions:**

   All admins should also be allowed to create and or change access rights of various user types in

the application.

2. **Schedule Data Analysis Jobs:**

   It shall be possible for admins to arrange how data analysis jobs can be performed automatically and the conditions for these tasks.

3. **Access Activity Logs:**

   Users shall have the provision of activity logs which shall be readable only by the admins for auditing purposes.

4. **Configure Application Settings:**

   Application setting should allow the admins to manage system settings, user settings among others.

5. **Access Analytics Dashboard:**

   It is recommended that the admins should be provided with web analytics where they can be able to view the performance of the various systems as well as the use of the systems.

4.2.3 Functional Requirements of the System:

The GenInsight system shall be capable of:

1. **User Authentication:**

   The system shall enable a user to sign up and log in to secure their account.

2. **Data Processing and Analysis:**

   The system shall provide basic data preparation for datasets, Automatic Machine Learning and analysis on the uploaded data.

3. **Visualization Tools:**

   The system shall deliver features and interfaces for advanced configuration and presentation of data analysis in the user interface.

4. **Communication Features:**

   The system shall allow the creation of notification and messaging to allow users to receive messages on time.

5. **Export Functionality:**

   In the later stages, the system shall enable the export of data in the various formats that users of

the system can easily download their results.

6. **Feedback Mechanism:**

   The system shall allow the users to feedback to the system in the event that they wish to suggest something or seek assistance in something and a tracking system for responses should be in place.

7. **Activity Logging:**

   The user action shall be recorded by the system for auditable purposes to keep record of activities by the user.

4.3 Quality Attributes

The success and effectiveness of the *GenInsight* platform rely on several critical quality attributes that ensure its performance, reliability, and overall user satisfaction. These attributes are critical to the platform and bear great significance towards driving its sustainability. They include **performance**, which ensures smooth and responsive user interactions; **reliability**, guaranteeing continuous availability; **usability**, which provides an intuitive and user-friendly interface; and **security**, safeguarding user data and maintaining privacy.

Further considerations are given to **maintainability**, enabling the platform to be updated and enhanced efficiently; **scalability**, ensuring the system can grow alongside an increasing volume of data and users; and **adaptability**, allowing the platform to accommodate future technological advancements and organizational needs. This section outlines the key quality attributes that are fundamental to the design and development of *GenInsight*, aimed at delivering a powerful and lasting data analytics solution for organizations.

4.4 Non-Functional Requirements

This section outlines the non-functional requirements for *GenInsight* across different areas of the system.

4.4.1 Availability

The system should meet the following availability requirements:

- **High Availability with Minimal Downtime**: The system shall ensure high availability with minimal downtime, supporting continuous operations and uninterrupted access to critical functionalities. This will be achieved by leveraging Azure Web App's infrastructure.
- **Failover and Redundancy**: Failover and redundancy mechanisms provided by Azure Web App

shall be implemented to maintain service availability during hardware or software failures.

4.4.2 Security

The system should meet the following security requirements:

- **Authentication and Authorization**: The system shall implement robust user authentication and authorization controls, integrating Azure Active Directory to ensure secure access and protect user data from unauthorized use.
- **Data Encryption**: Sensitive data shall be encrypted at rest and in transit using Azure encryption services to ensure data confidentiality and prevent unauthorized access.
- **Compliance with Regulations**: The system shall comply with data protection regulations such as GDPR to safeguard user privacy and enforce users' rights regarding personal data. This will be supported by Azure's compliance certifications and tools.

4.4.3 Performance

The system should meet the following performance requirements:

- **Processing and Visualization**: The system shall process and visualize datasets containing up to 50,000 records within 20 seconds by leveraging Azure's scalable resources to handle large-scale data analysis tasks efficiently.
- **Real-Time Data Updates**: The system shall support real-time data updates to ensure users have access to up-to-date information and insights without delay.
- **Quick Response Times**: Data processing and analysis tasks must have quick response times, utilizing Azure's performance features to enable timely decision-making and interactive user experiences.

4.4.4 Monitoring and Debugging

The system should meet the following monitoring and debugging requirements:

- **Real-Time Monitoring**: The system shall utilize Azure Log Stream for real-time logging and monitoring to detect issues proactively and facilitate efficient resolution.
- **Debugging Capabilities**: The system shall leverage Azure's debugging capabilities to enable troubleshooting and maintain the reliability and stability of the platform.

4.5 Assumptions

The following assumptions have been made for the *GenInsight* platform:

- End users will access the platform using a web browser that supports JavaScript.
- A stable internet connection is required for users to interact with the cloud-hosted application and its features.
- Users are expected to have a basic understanding of web navigation, including how to interact with buttons, menus, and other common interface elements.

4.6 Use Cases

Use cases in the context of a system architecture play a crucial role in defining, designing, and validating the functionality and structure of the final system. They represent significant scenarios or interactions that drive the development and operation of the system. The table 4.1 provides a simple use case.

Table 4.1: Use Case 1: User Login

Name	User Login		
Use Case ID	UC01		
Actors	Admin, Consumer		
Description	Users log into the GenInsight application using their credentials.		
Pre-Conditions	Users must have valid login credentials.		
Post-Conditions	Users are authenticated and redirected to their respective dashboards.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User inputs username and password.	2	System displays the login form and asks for username and password.
2	User enters valid credentials.	3	System validates the credentials and redirects the user to the respective dashboard.
Alternative Flow			
3	User enters invalid credentials.	3-A	System responds with an error message: Incorrect username or password.

Table 4.2: Use Case 2: User Role Differentiation

Name	User Role Differentiation		
Use Case ID	UC02		
Actors	Admin, Consumer		
Description	The system distinguishes between Admin and Consumer user roles upon login.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users gain access to functionalities based on their roles.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User logs in.	2	The system verifies login and checks the user's role.
2	The system checks the user's role.	3	The system displays the corresponding dashboard and features based on the role.
Alternative Flow			
3	If the user's role is invalid.	3-A	The system returns an error message and restricts access.

Table 4.3: Use Case 3: Chat Interaction with Structured Data

Name	Chat Interaction with Structured Data		
Use Case ID	UC03		
Actors	Admin, Consumer		
Description	Users can interact with structured data via a chat interface.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users can view and interact with structured data in a conversational manner.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User initiates a chat session.	2	System opens the chat interface and awaits user input.
2	User queries specific structured data.	3	System processes the query and returns the relevant data.
Alternative Flow			
3	User provides an invalid or unclear query.	3-A	System returns an error message or prompts the user to clarify the query.

Table 4.4: Use Case 4: Editing EDA Dashboard using LLM

Name	Editing EDA Dashboard using LLM		
Use Case ID	UC04		
Actors	Admin, Consumer		
Description	Users can edit the default EDA dashboard using language model capabilities.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users modify the EDA dashboard layout and visualizations.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User accesses the EDA dashboard.	2	System displays the default EDA dashboard.
2	User enters editing mode.	3	System activates the dashboard editing features.
3	User interacts with the language model for modifications.	4	System processes input and updates the dashboard.
Alternative Flow			
4	Invalid input provided by the user.	4-A	System returns an error and prompts for correction.

Table 4.5: Use Case 5: Accessing the AutoML Page

Name	Accessing the AutoML Page		
Use Case ID	UC05		
Actors	Admin, Consumer		
Description	Users can access the AutoML page for automated machine learning tasks.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users initiate and monitor AutoML tasks.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User navigates to the AutoML page.	2	System displays the AutoML interface.
2	User selects a dataset and machine learning task.	3	System prepares the task for the AutoML process.
3	User initiates the AutoML process.	4	System executes the task and presents the results.
Alternative Flow			
4	Invalid task selection by the user.	4-A	System returns an error and prompts for task correction.

Table 4.6: Use Case 6: Training Resource Access

Name	Training Resource Access		
Use Case ID	UC06		
Actors	Admin, Consumer		
Description	Users can access training resources to learn about application features.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users engage with training materials effectively.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User navigates to the training page.	2	System displays available training modules.
2	User selects training modules or resources.	3	System provides access to the selected materials.
3	User completes the training and provides feedback.	4	System records training completion and feedback.
Alternative Flow			
3	User fails to complete training.	3-A	System prompts user to revisit incomplete modules.

Table 4.7: Use Case 7: Implementing Two-Factor Authentication

Name	Implementing Two-Factor Authentication		
Use Case ID	UC07		
Actors	Admin, Consumer		
Description	The application implements two-factor authentication for enhanced security.		
Pre-Conditions	Two-factor authentication must be enabled in the application settings.		
Post-Conditions	Users are securely authenticated through dual verification.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User enters login credentials.	2	System prompts for a second authentication factor.
2	User provides the second factor.	3	System validates the second factor and grants access.
Alternative Flow			
3	Invalid second factor provided by the user.	3-A	System returns an error message and prompts for re-entry.

Table 4.8: Use Case 8: Voice-Enabled Chat Integration

Name	Voice-Enabled Chat Integration		
Use Case ID	UC08		
Actors	Admin, Consumer		
Description	Users can communicate with the chatbot using voice commands.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users can engage in voice-enabled interactions.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User activates voice chat functionality.	2	System enables voice interaction mode.
2	User speaks to the chatbot.	3	System processes voice input and responds accordingly.
Alternative Flow			
3	Unclear or invalid voice input.	3-A	System prompts the user to repeat the voice command.

Table 4.9: Use Case 9: Application Configuration Management

Name	Application Configuration Management		
Use Case ID	UC09		
Actors	Admin		
Description	Admin users configure application settings and user roles.		
Pre-Conditions	Admin must be logged into the application.		
Post-Conditions	Application settings are updated based on admin input.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	Admin accesses configuration settings.	2	System displays current application settings.
2	Admin modifies application settings.	3	System saves the modifications and applies the changes.
Alternative Flow			
3	Invalid settings or input provided.	3-A	System returns an error message and prompts for corrections.

Table 4.10: Use Case 10: Data Upload and Preprocessing

Name	Data Upload and Preprocessing		
Use Case ID	UC10		
Actors	Admin, Consumer		
Description	Users upload datasets and initiate preprocessing.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Datasets are uploaded and preprocessed successfully.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User navigates to the data upload section.	2	System displays the data upload interface.
2	User selects files to upload.	3	System validates the files and uploads the selected files.
3	User initiates preprocessing of the uploaded data.	4	System processes the data and completes preprocessing.
Alternative Flow			
3	Invalid file format or upload error.	3-A	System returns an error message and prompts for re-upload.

Table 4.11: Use Case 11: Viewing Data Insights

Name	Viewing Data Insights		
Use Case ID	UC11		
Actors	Admin, Consumer		
Description	Users can view insights generated from data analysis.		
Pre-Conditions	Users must be logged into the application and analysis must be completed.		
Post-Conditions	Users access insights and visualizations from the data.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User navigates to the insights page.	2	System displays the available datasets for insights.
2	User selects a dataset to view insights.	3	System displays the insights and visualizations derived from the analysis.
Alternative Flow			
3	No insights available for the dataset.	3-A	System notifies the user that no insights are available.

Table 4.12: Use Case 12: Setting User Permissions

Name	Setting User Permissions		
Use Case ID	UC12		
Actors	Admin		
Description	Admin can set permissions for different user roles.		
Pre-Conditions	Admin must be logged into the application.		
Post-Conditions	User permissions are updated successfully.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	Admin accesses user management settings.	2	System displays the current user roles and permissions.
2	Admin selects a user and modifies their permissions.	3	System saves and applies the updated permissions.
Alternative Flow			
3	Invalid permissions or roles selected.	3-A	System returns an error and prompts for valid input.

Table 4.13: Use Case 13: Report Generation

Name	Report Generation		
Use Case ID	UC13		
Actors	Admin, Consumer		
Description	Users can generate reports based on analyzed data.		
Pre-Conditions	Users must be logged into the application and have access to relevant data.		
Post-Conditions	Users download or view the generated reports.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User selects the data for report generation.	2	System displays report generation options.
2	User chooses the desired report format (e.g., PDF, CSV).	3	System generates and provides the report for download.
Alternative Flow			
3	Invalid data or report format selected.	3-A	System returns an error message and prompts for valid input.

Table 4.14: Use Case 14: Scheduling Data Analysis Jobs

Name	Scheduling Data Analysis Jobs		
Use Case ID	UC14		
Actors	Admin		
Description	Admin can schedule automated data analysis jobs.		
Pre-Conditions	Admin must be logged into the application.		
Post-Conditions	Data analysis jobs are scheduled successfully.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	Admin navigates to the job scheduling section.	2	System displays the job scheduling options.
2	Admin selects a dataset and specifies analysis parameters.	3	System allows the admin to input the desired parameters.
3	Admin sets the schedule for the job.	4	System confirms the job scheduling.
Alternative Flow			
3	Invalid parameters or schedule set.	3-A	System returns an error and prompts for correction.

Table 4.15: Use Case 15: Providing Feedback and Support

Name	Providing Feedback and Support		
Use Case ID	UC15		
Actors	Admin, Consumer		
Description	Users can submit feedback or request support.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	User feedback is recorded, or support is initiated.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User navigates to the feedback/support section.	2	System displays the feedback/support options.
2	User fills out a feedback form or initiates a support chat.	3	System records the feedback or connects the user to support.
Alternative Flow			
3	Form submission error or chat connection failure.	3-A	System returns an error and prompts for retry.

Table 4.16: Use Case 16: Monitoring System Performance

Name	Monitoring System Performance		
Use Case ID	UC16		
Actors	Admin		
Description	Admin can monitor the system's performance metrics and usage statistics.		
Pre-Conditions	Admin must be logged into the application.		
Post-Conditions	Admin views relevant system performance metrics.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	Admin navigates to the performance monitoring section.	2	System displays relevant metrics such as response time and user activity.
2	Admin analyzes the performance data.	3	System provides options for analyzing specific performance parameters.
Alternative Flow			
3	Performance data unavailable.	3-A	System displays an error and prompts the admin to retry later.

Table 4.17: Use Case 17: Data Export Functionality

Name	Data Export Functionality		
Use Case ID	UC17		
Actors	Admin, Consumer		
Description	Users can export analyzed data in various formats.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Data is exported successfully in the desired format.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User selects data for export.	2	System displays export format options.
2	User chooses the export format.	3	System processes the request and downloads the exported file.
Alternative Flow			
3	Invalid data or format selection.	3-A	System returns an error and prompts for correct input.

Table 4.18: Use Case 18: Customizing Data Visualizations

Name	Customizing Data Visualizations		
Use Case ID	UC18		
Actors	Admin, Consumer		
Description	Users can customize visualizations displayed on the dashboard.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Visualizations are updated as per user specifications.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User accesses the visualization customization settings.	2	System displays current visualizations and customization options.
2	User modifies chart types, colors, and labels.	3	System updates the dashboard visualizations.
Alternative Flow			
3	Invalid customization settings.	3-A	System returns an error and prompts for valid customization input.

Table 4.19: Use Case 19: Implementing Notifications

Name	Implementing Notifications		
Use Case ID	UC19		
Actors	Admin, Consumer		
Description	The system sends notifications for important events or updates.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	Users receive relevant notifications.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	The system triggers a notification event (e.g., job completion).	2	System sends notifications to relevant users.
2	Users view notifications in their dashboard.	3	System displays the notification for the relevant event.
Alternative Flow			
3	Notification fails to send.	3-A	System returns an error and logs the issue for admin review.

Table 4.20: Use Case 20: User Activity Logging

Name	User Activity Logging		
Use Case ID	UC20		
Actors	Admin		
Description	The system logs user activities for auditing and monitoring purposes.		
Pre-Conditions	Users must be logged into the application.		
Post-Conditions	User activities are recorded in the log.		
Special Requirements	None		
Basic Flow			
Actor Action		**System Response**	
1	User performs various actions in the application.	2	System records each action in the activity log.
2	Admin reviews the activity log.	3	System displays the activity log for auditing or analysis.
Alternative Flow			
3	Log entry fails or data missing.	3-A	System returns an error and attempts to recover log data.

4.7 Hardware and Software Requirements

This section outlines the necessary hardware and software requirements for the development of *GenInsight*, a cloud-based data analytics platform.

4.7.1 Hardware Requirements

The following hardware requirements are necessary for the project:

- **Desktop PC/Laptop**: A computer that has the necessary specifications in order to accommodate the data and to execute the necessary computations of a development machine.
- **Internet Connection**: Data base and web based applications require internet connection stability, as is the need for cloud deployment.

4.7.2 Software Requirements

The following software requirements are necessary for the project:

- **Python**: A language which is mostly used for the backend application and database processing.
- **Flask**: A light-weight web framework for Python to put out the machine learning models and perform the backend tasks.
- **HTML, CSS, JavaScript**: For development of the front-end of the application that will entail creating the user interface, the dashboards, and data visualization.

- **Python Libraries**:
  - **pandas, NumPy, matplotlib, seaborn**: This application was used for data analysis and visualization purposes only.
  - **spaCy, NLTK**: NLP tasks require numerous language models for language characterization and understanding, for analysis of context and semantics as well as for document analysis.
  - **scikit-learn, TensorFlow**: For the training of machine learning algorithms and all model training.
- **VS Code**: A web based integrated development environment for writing the code on backend and front end of the platform.

4.8 Graphical User Interface

This section should give the GUI dumps of each screen, with reference to the users.

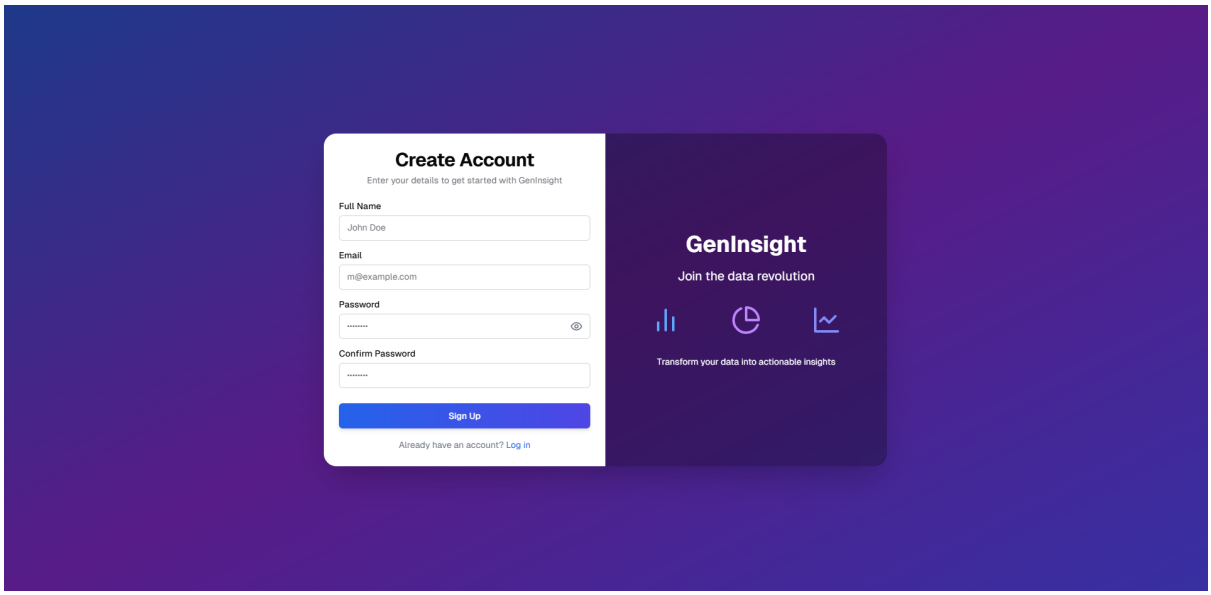

Figure 4.1: Sign-Up Screen (Figure 1)
This screen showcases the GenInsight Sign-Up Screen, where users can create a new account to gain access to the platform's data analytics features and tools.

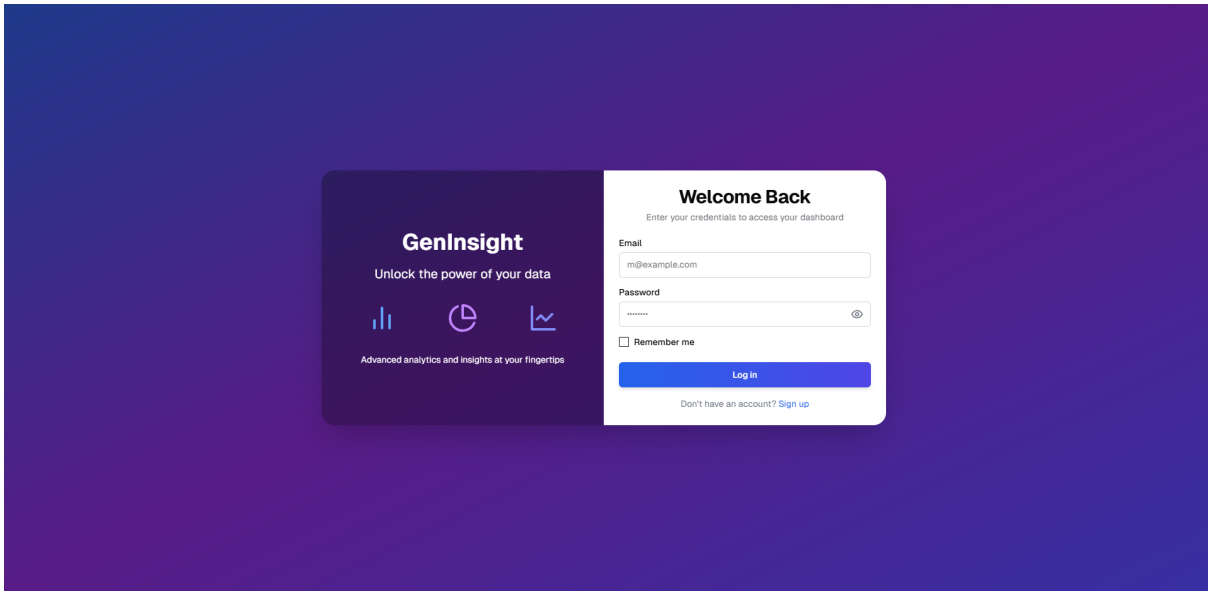

Figure 4.2: Login Screen (Figure 2)

This screen illustrates the GenInsight Login Screen, enabling registered users to securely log in and access their accounts to explore and analyze their organizational data.

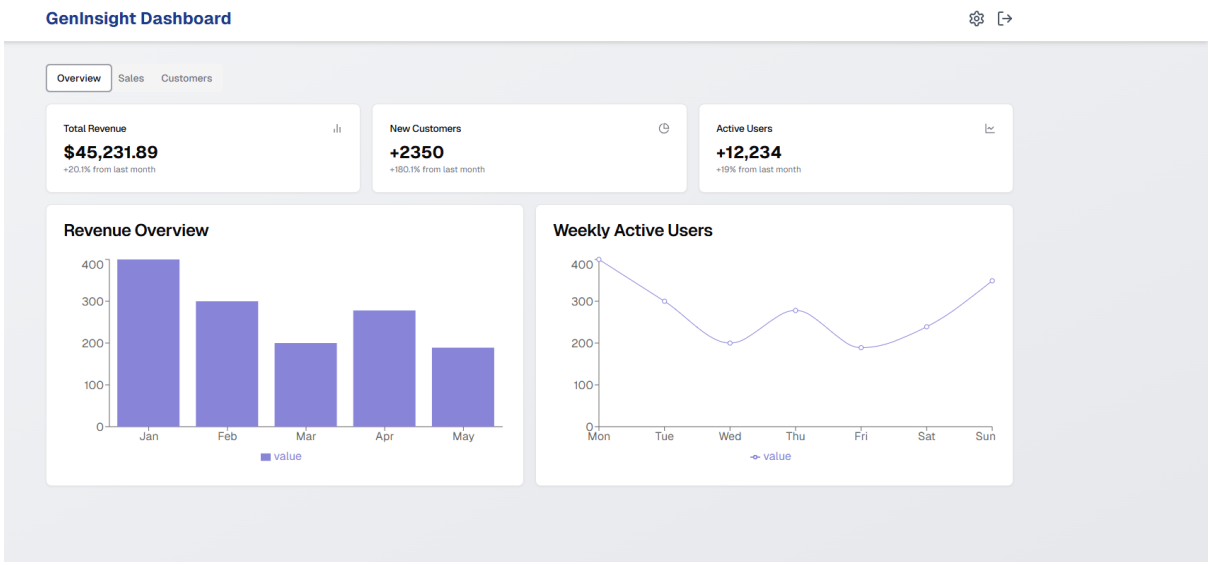

Figure 4.3: Dashboard (Figure 3)

This screen displays the GenInsight Dashboard, where users can see the overall statistics of data, users, customers, and the target variable of the data uploaded (in this case ”Sales”) .

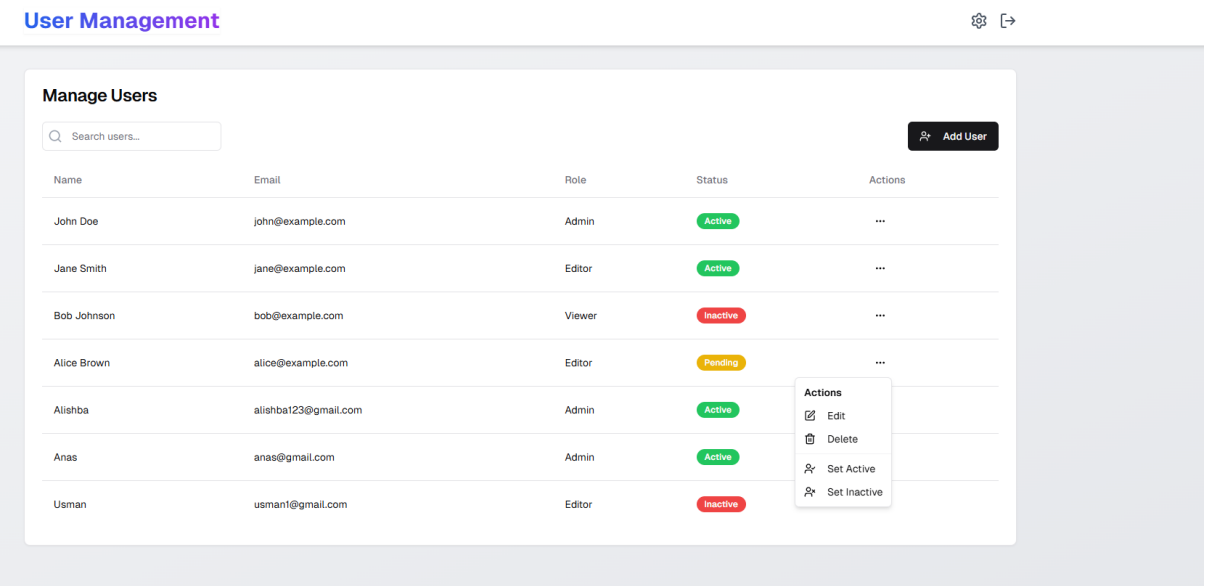

Figure 4.4: Admin User Management (Figure 4)

This screen illustrates the GenInsight Admin User Management interface, allowing administrators to manage user roles, permissions, and account settings within the platform.

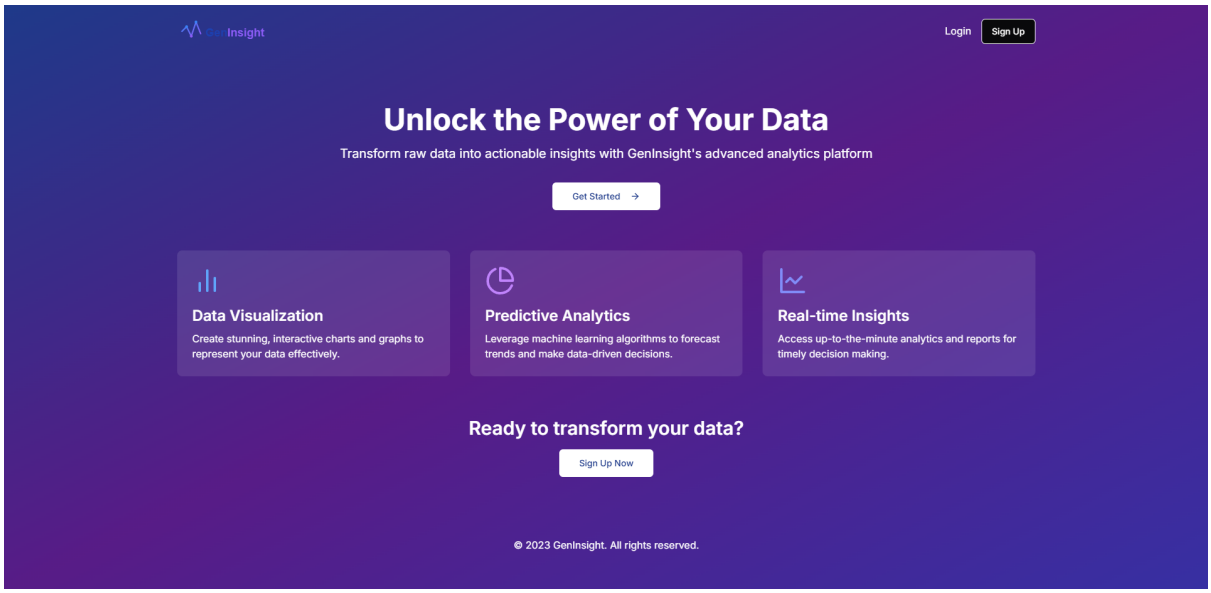

Figure 4.5: Landing Page (Figure 5)

This screen represents the GenInsight Landing Page, where users can access/create their accounts for generating and performing insights, visualizations, reports, and manage their data analysis tasks.

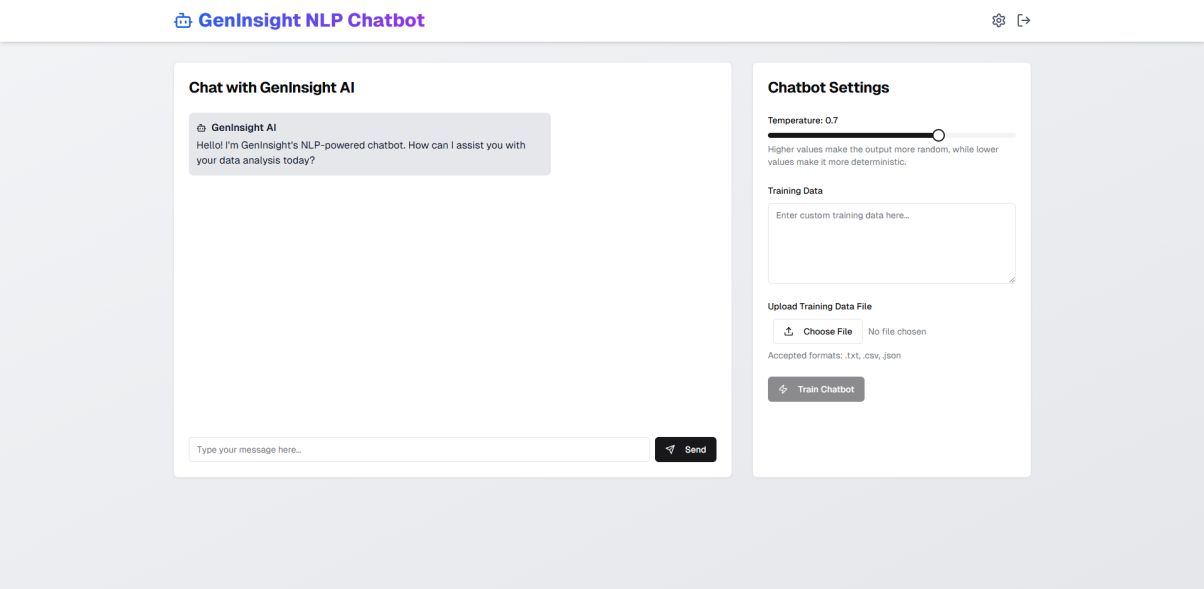

Figure 4.6: Chat Assistant (Figure 6)
This screen displays the GenInsight Chat Assistant, where users can interact with the system to ask questions, receive guidance, and perform tasks using natural language queries.

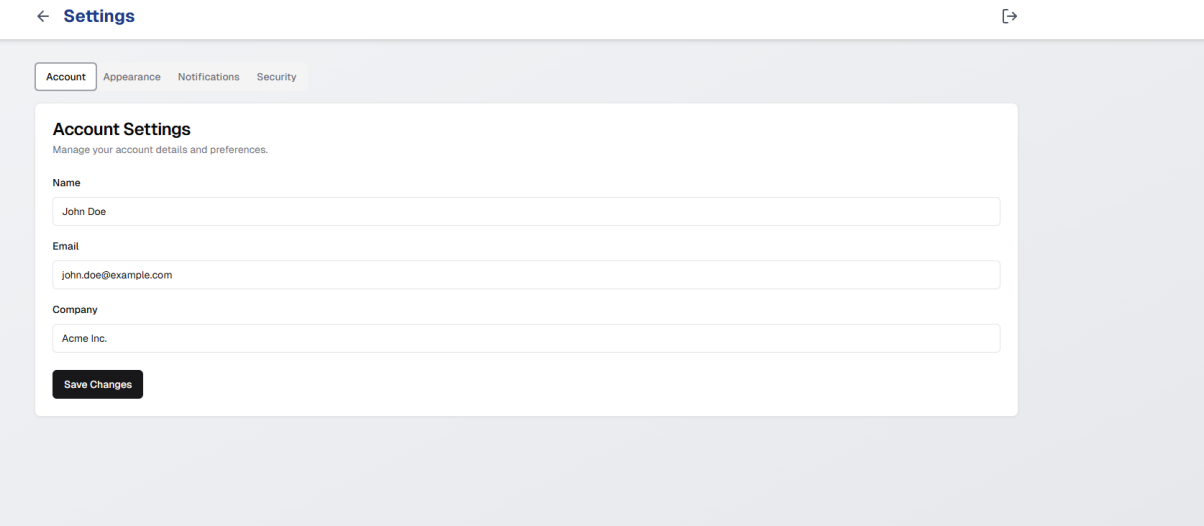

Figure 4.7: Account Settings (Figure 7)
This screen shows the GenInsight Account Settings interface, allowing users to manage their personal information and security preferences.

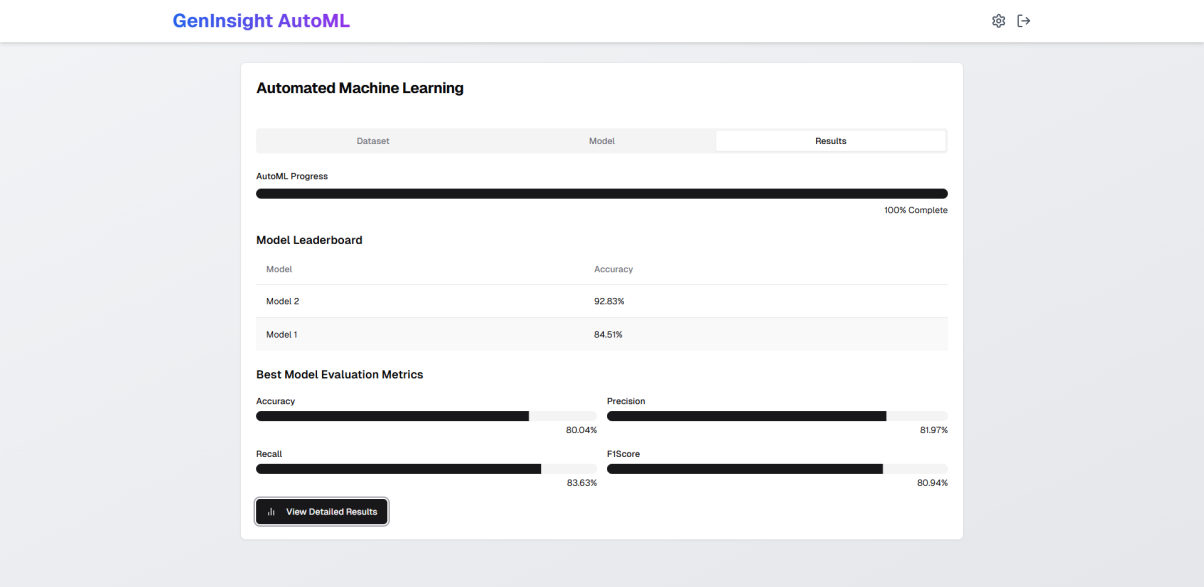

Figure 4.8: Auto ML (Figure 8)

This screen displays the GenInsight Auto ML interface, where users can configure and run automated machine learning tasks to analyze datasets and generate predictive models.

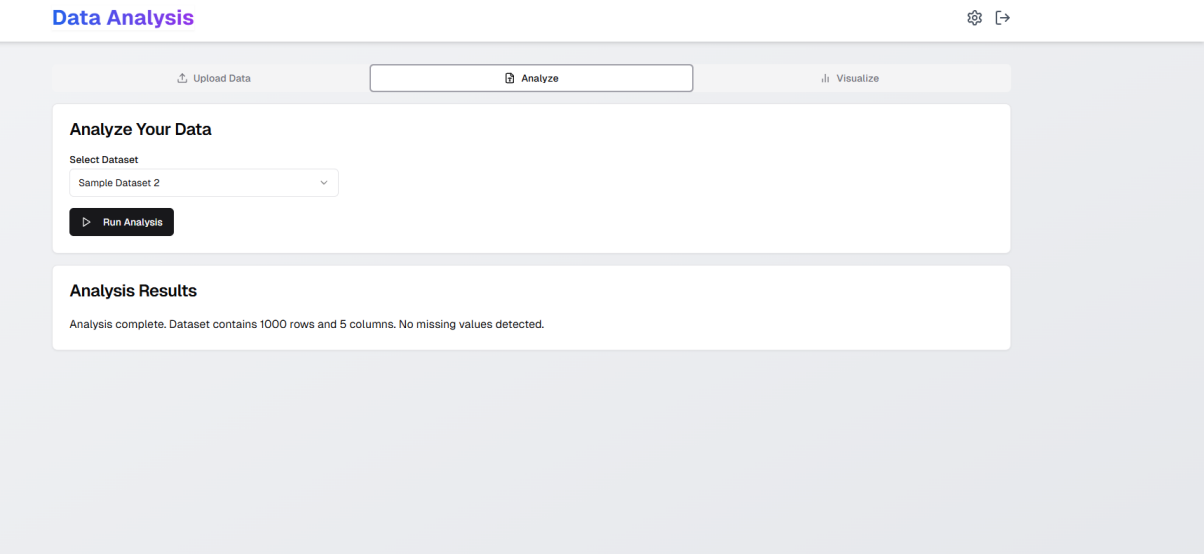

Figure 4.9: Data Analytics (Figure 9)

This screen illustrates the GenInsight Data Analytics interface, where users can explore and analyze their datasets, generate insights, and visualize data using various analytical tools.

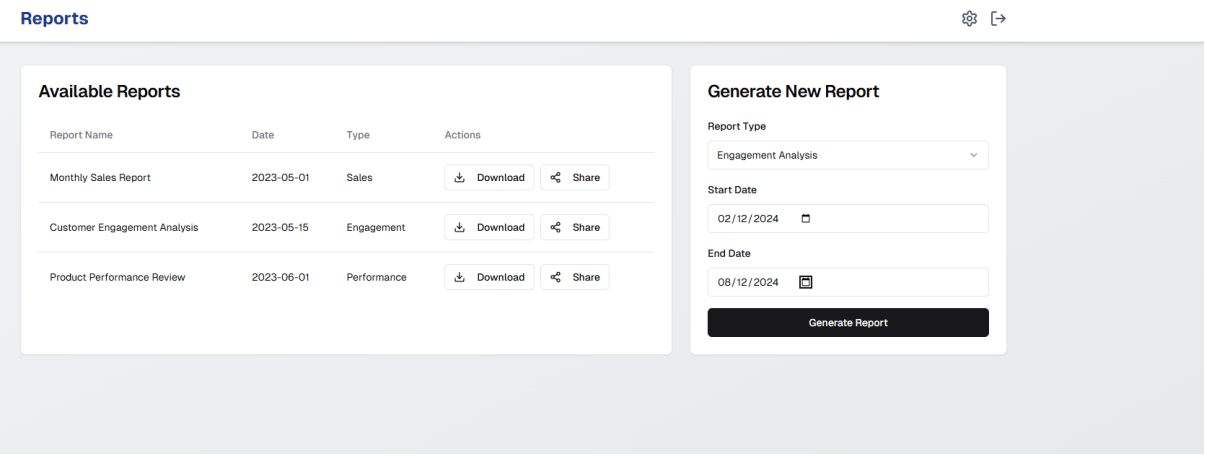

Figure 4.10: Data Management (Figure 10)
This screen showcases the GenInsight Reports interface, where users can generate and download reports for analysis within the platform.

4.9 Database Design

This section will be showcasing the Entity Relationship diagram and the data dictionary associated with it.

4.9.1 ER Diagram

The entity relationship diagram for *GenInsight* is as follows

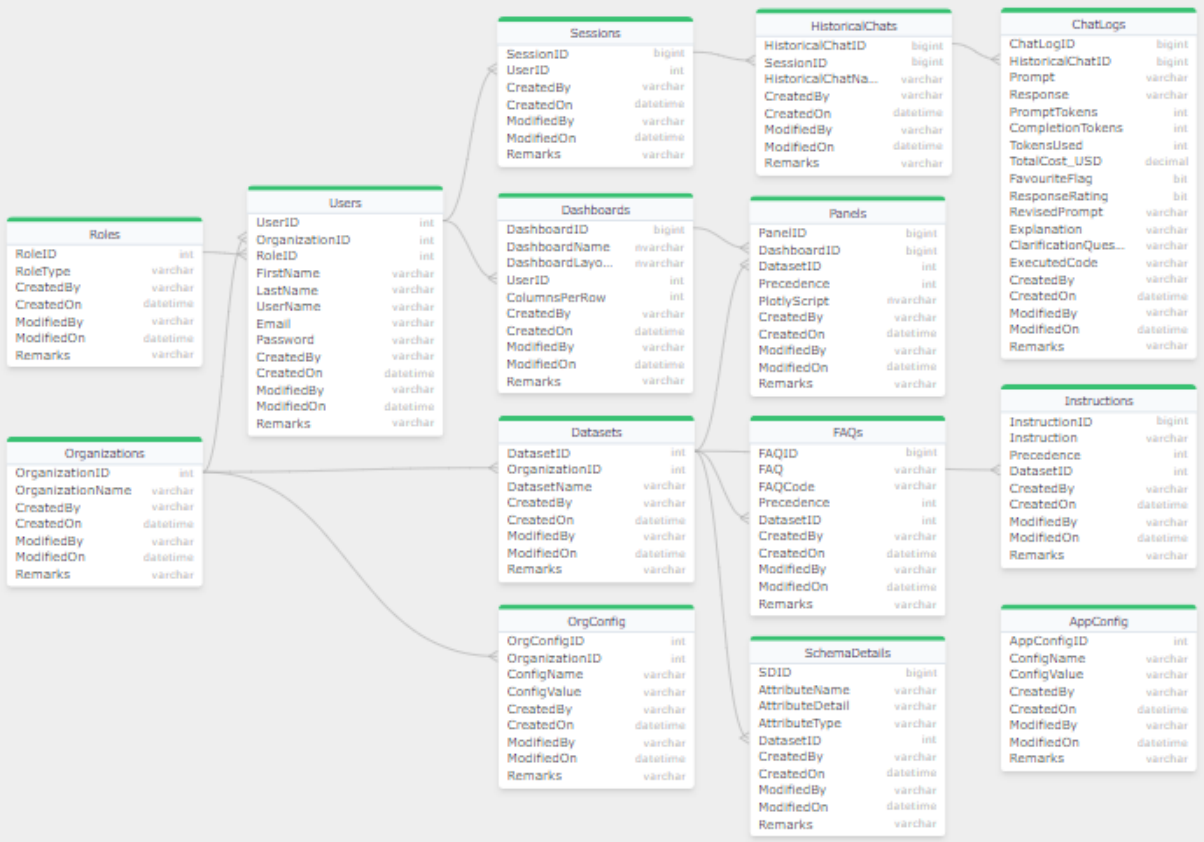

Figure 4.11: Entity Relationship Diagram for GenInsight

Entity-Relationship (ER) Diagram illustrating the relationships among tables, detailing entities, attributes, and their connections.

4.9.2 Data Dictionary

For each of the entity in the Entity Relationship diagram mentioned above, the Data Dictionary is provided in this chapter to understand the entities of the concept and how these related with the other entities.

Table 4.21: User Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
User	UserID	Primary Key	No		Unique identifier for user entry.
User	OrganizationID	Foreign Key	No		Reference to organization record.
User	RoleID	Foreign Key	No		Reference to role record.
User	FirstName	Text	No		User's first name.
User	LastName	Text	Yes		User's last name.
User	UserName	Text	No		Unique username of user.
User	Email	Text	No		Unique email of user.
User	Password	Text	No		Encrypted user password.
User	CreatedBy	Date	No		User who created the entry.
User	CreatedOn	Text	No		Timestamp when entry was created.
User	ModifiedBy	Date	No		User who modified the entry.
User	ModifiedOn	Text	No		Timestamp of last modification.
User	Remarks	Text	Yes		Additional comments or notes.

Table 4.22: Role Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Role	RoleID	Primary Key	No		Unique identifier for role entry.
Role	RoleType	Text	No		Type of user role.
Role	CreatedBy	Date	No		User who created the entry.
Role	CreatedOn	Text	No		Timestamp when entry was created.
Role	ModifiedBy	Date	No		User who modified the entry.
Role	ModifiedOn	Text	No		Timestamp of last modification.
Role	Remarks	Text	Yes		Additional comments or notes.

Table 4.23: Organization Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
Organization	OrganizationID	Primary Key	No	None	Unique identifier for organization entry.
Organization	OrganizationName	Text	No	None	Name of the organization.
Organization	CreatedBy	Text	No	None	User who created the entry.
Organization	CreatedOn	DateTime	No	None	Timestamp when entry was created.
Organization	ModifiedBy	Text	No	None	User who last modified the entry.
Organization	ModifiedOn	DateTime	No	None	Timestamp of last modification.
Organization	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.24: Organization Configurations Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
OrgConfig	OrgConfigID	Primary Key	No	None	Unique identifier for organization configuration entry.
OrgConfig	OrganizationID	Foreign Key	No	None	Reference to organization record.
OrgConfig	ConfigName	Text	No	None	Configuration name for organization.
OrgConfig	ConfigValue	Text	No	None	Value of the configuration.
OrgConfig	CreatedBy	Text	No	None	User who created the entry.
OrgConfig	CreatedOn	DateTime	No	None	Timestamp when entry was created.
OrgConfig	ModifiedBy	Text	No	None	User who modified the entry.
OrgConfig	ModifiedOn	DateTime	No	None	Timestamp of last modification.
OrgConfig	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.25: Sessions Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Session	SessionID	Primary Key	No		Unique identifier for session entry.
Session	UserID	Foreign Key	No		Reference to user record.
Session	CreatedBy	Date	No		User who created the entry.
Session	CreatedOn	Text	No		Timestamp when entry was created.
Session	ModifiedBy	Date	No		User who modified the entry.
Session	ModifiedOn	Text	No		Timestamp of last modification.
Session	Remarks	Text	Yes		Additional comments or notes.

Table 4.26: Historical Chats Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
HistoricalChat	HistoricalChatID	Primary Key	No	None	Unique identifier for historical chat entry.
HistoricalChat	SessionID	Foreign Key	No	None	Reference to session record.
HistoricalChat	HistoricalChatName	Text	No	None	Name of the historical chat.
HistoricalChat	CreatedBy	Text	No	None	User who created the entry.
HistoricalChat	CreatedOn	DateTime	No	None	Timestamp when entry was created.
HistoricalChat	ModifiedBy	Text	No	None	User who last modified the entry.
HistoricalChat	ModifiedOn	DateTime	No	None	Timestamp of the last modification.
HistoricalChat	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.27: Chat Logs Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
ChatLog	ChatLogID	Primary Key	No	None	Unique identifier for chat log entry.
ChatLog	HistoricalChatID	Foreign Key	No	None	Reference to historical chat record.
ChatLog	Prompt	Text	No	None	User's input query or command.
ChatLog	Response	Text	No	None	Generated chatbot's response.
ChatLog	PromptTokens	Integer	No	None	Number of tokens in user input.
ChatLog	CompletionTokens	Integer	No	None	Tokens used in the response.
ChatLog	TokensUsed	Integer	No	None	Total tokens used (prompt + response).
ChatLog	TotalCost_USD	Decimal	No	None	Cost of tokens in USD.
ChatLog	FavouriteFlag	Boolean	No	None	Marks favorite responses (True/False).
ChatLog	ResponseRating	Boolean	No	None	User rating for response (Like/Dislike).
ChatLog	RevisedPrompt	Text	Yes	None	Modified user input (if revised).
ChatLog	Explanation	Text	Yes	None	Explanation of chatbot's response.
ChatLog	ClarificationQuestions	Text	Yes	None	Follow-up questions for clarification.
ChatLog	ExecutedCode	Text	Yes	None	Code executed against the response (if any).
ChatLog	CreatedBy	Text	No	None	User who created the entry.
ChatLog	CreatedOn	DateTime	No	None	Timestamp when entry was created.
ChatLog	ModifiedBy	Text	No	None	User who modified the entry.
ChatLog	ModifiedOn	DateTime	No	None	Timestamp of last modification.
ChatLog	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.28: Datasets Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Dataset	DatasetID	Primary Key	No		Unique identifier for dataset entry.
Dataset	OrganizationID	Foreign Key	No		Reference to organization record.
Dataset	DatasetName	Text	No		Name of the dataset.
Dataset	CreatedBy	Date	No		User who created the entry.
Dataset	CreatedOn	Text	No		Timestamp when entry was created.
Dataset	ModifiedBy	Date	No		User who modified the entry.
Dataset	ModifiedOn	Text	No		Timestamp of last modification.
Dataset	Remarks	Text	Yes		Additional comments or notes.

Table 4.29: Schema Details Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
SchemaDetail	SDID	Primary Key	No	None	Unique identifier for schema detail entry.
SchemaDetail	AttributeName	Text	No	None	Name of the attribute.
SchemaDetail	AttributeDetail	Text	No	None	Details about the attribute.
SchemaDetail	AttributeType	Text	No	None	Data type of the attribute.
SchemaDetail	DatasetID	Foreign Key	No	None	Reference to dataset record.
SchemaDetail	CreatedBy	Text	No	None	User who created the entry.
SchemaDetail	CreatedOn	DateTime	No	None	Timestamp when entry was created.
SchemaDetail	ModifiedBy	Text	No	None	User who modified the entry.
SchemaDetail	ModifiedOn	DateTime	No	None	Timestamp of last modification.
SchemaDetail	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.30: Dashboard Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Dashboard	DashboardID	Primary Key	No		Unique identifier for dashboard entry.
Dashboard	DashboardName	Text	No		Name of the dashboard.
Dashboard	DashboardLayout	Text	No		Layout configuration for dashboard.
Dashboard	UserID	Foreign Key	No		Reference to user record.
Dashboard	ColumnsPerRow	Integer	No		Number of columns per row.
Dashboard	CreatedBy	Date	No		User who created the entry.
Dashboard	CreatedOn	Text	No		Timestamp when entry was created.
Dashboard	ModifiedBy	Date	No		User who modified the entry.
Dashboard	ModifiedOn	Text	No		Timestamp of last modification.
Dashboard	Remarks	Text	Yes		Additional comments or notes.

Table 4.31: Panel Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Panel	PanelID	Primary Key	No		Unique identifier for panel entry.
Panel	DashboardID	Foreign Key	No		Reference to dashboard record.
Panel	DatasetID	Foreign Key	No		Reference to dataset record.
Panel	Precedence	Integer	No		Display order for panel.
Panel	PlotlyScript	Text	No		Script to generate panel chart.
Panel	CreatedBy	Date	No		User who created the entry.
Panel	CreatedOn	Text	No		Timestamp when entry was created.
Panel	ModifiedBy	Date	No		User who modified the entry.
Panel	ModifiedOn	Text	No		Timestamp of last modification.
Panel	Remarks	Text	Yes		Additional comments or notes.

Table 4.32: Instruction Table

Entity	Attribute	Data Type	Nullable	Relation To	Relation Type
Instruction	InstructionID	Primary Key	No		Unique identifier for instruction entry.
Instruction	Instruction	Text	No		Instruction text content.
Instruction	Precedence	Integer	No		Display order of instruction.
Instruction	DatasetID	Foreign Key	No		Reference to dataset record.
Instruction	CreatedBy	Date	No		User who created the entry.
Instruction	CreatedOn	Text	No		Timestamp when entry was created.
Instruction	ModifiedBy	Date	No		User who modified the entry.
Instruction	ModifiedOn	Text	No		Timestamp of last modification.
Instruction	Remarks	Text	Yes		Additional comments or notes.

Table 4.33: Frequently Asked Questions Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
FAQ	FAQID	Primary Key	No	None	Unique identifier for frequently asked question entry.
FAQ	FAQ	Text	No	None	Frequently asked question text.
FAQ	FAQCode	Text	Yes	None	Code associated with FAQ (if any).
FAQ	Precedence	Integer	No	None	FAQ display order precedence.
FAQ	DatasetID	Foreign Key	No	None	Reference to dataset record.
FAQ	CreatedBy	Text	No	None	User who created the entry.
FAQ	CreatedOn	DateTime	No	None	Timestamp when entry was created.
FAQ	ModifiedBy	Text	No	None	User who modified the entry.
FAQ	ModifiedOn	DateTime	No	None	Timestamp of last modification.
FAQ	Remarks	Text	Yes	None	Additional comments or notes.

Table 4.34: Application Configurations Table

Entity	Attribute	Data Type	Nullable	Relation To	Description
AppConfig	AppConfigID	Primary Key	No	None	Unique identifier for application configuration entry.
AppConfig	ConfigName	Text	No	None	Name of application configuration.
AppConfig	ConfigValue	Text	No	None	Value of the configuration.
AppConfig	CreatedBy	Text	No	None	User who created the entry.
AppConfig	CreatedOn	DateTime	No	None	Timestamp when entry was created.
AppConfig	ModifiedBy	Text	No	None	User who modified the entry.
AppConfig	ModifiedOn	DateTime	No	None	Timestamp of last modification.
AppConfig	Remarks	Text	Yes	None	Additional comments or notes.

4.10 Risk Analysis

Risk analysis is a key component of *GenInsight*'s project planning, addressing potential challenges that may arise during its development and deployment:

4.10.1 Technical Risks

Some of the technical risks include scalability issues, system integration, and even the cloud infrastructure. Since *GenInsight* deals with such large volumes of datasets, it is most likely to have performance degrade based on heavy loads. The risk may be mitigated by proper load testing and through the adaptation of cloud scaling solutions like Azure auto-scaling, ensuring the system remains responsive. Compatibility with different sources of data (SQL, Excel, Azure) also poses risks and mandates constant testing to validate proper integration.

4.10.2 Business Risks

The data analytics platforms market also carries risks: positioning in the market, competition, and specificity within it. Market adoption may prove to be just insufficient, if the system fails to make a strong differentiation from known solutions. To avoid this, *GenInsight* should focus on its unique value proposition-that is, real-time collaboration, AutoML features, and friendly NLP. Adequate marketing and partnership policies will also be required to attract clients from different spheres.

4.10.3 Data Quality and Security Risks

It is imperative to maintain *GenInsight* organizational data integrity and security. Data breaches, mishandling of sensitive information, or poor quality of data might see the trust in the platform come crumbling down. Audits and validation procedures will ensure that the quality of data is maintained. The platform employs robust data at rest and in motion encryption along with strong authentication such as OAuth 2.0 to mitigate these risks.

4.10.4 User Adoption and Usability Risks

Usability for users of *GenInsight* needs to come with seamless user onboarding, that will not make the user disinterested in the analytic capabilities. In relation to non-technical users, navigating this particular platform with any analytics capability shall be difficult to use; on this count, *GenInsight* shall do its best at keeping the interface user-friendly, coming with supporting abundant help documentation and enable easier interactions from complex data streams through natural language processing.

4.10.5 Competition Risks

The competition in the data analytics space is very high, and established platforms may easily replicate what *GenInsight* offers to customers. This risk may be mitigated by continuous innovations that include features like suggestion of machine learning models and real-time collaborative tools, and otherwise ensures that the platform always keeps abreast of emerging counterparts.

These risks are better managed by *GenInsight*, which conducts regular risk assessments throughout the lifecycle of the project. These plans include risk mitigation plans, flexibility in the execution of a project, and potential new risks that are identified while the platform continues to grow during its development.

4.11 Conclusion

Chapter 4 of the Software Requirement Specification for *GenInsight* clearly spells out a roadmap for developing the system. It is introduced with a very long list of features and functional requirements- a long list of necessities-the essential capabilities of the system in terms of scalable data warehouse, customizable dashboards, real-time collaboration tools, and so on.

Quality attributes like performance, security, and usability are elaborated to make sure that the platform fulfills the stringent requirements that are required for enterprise-grade data analytics. Non-functional requirements and assumptions, such as hardware and software dependencies, are also kept in mind to guide the development.

This chapter also provides an overview of the graphical user interface of the platform, database architecture, and some of the chief design elements. The risk analysis section outlines potential challenges that may affect the development of *GenInsight* and provides proactive strategies to counteract those risks.

Overall, it will provide complete guidance on the development of the *GenInsight* to the features and requirements of the platform, design considerations, and risk management strategies, which will lay down a strong foundation for assuredly successful project implementation and scalability in the future.

Chapter 5 High-Level and Low-Level Design

This chapter provides a comprehensive description of the high-level and low-level design of the *GenInsight* system, focusing on its core components, architectural strategies, and design considerations.

5.1 System Overview

GenInsight is a cloud-hosted data analytics solution that aims to power organizations by providing intuitive, scalable, and collaborative tools for the management, analysis, and visualization of data. It is built considering the different needs of the departments of finance, sales, development, and marketing.

5.1.1 Scalable Data Warehouse

GenInsight provides the power of an enterprise-level robust and scalable data warehouse that allows large volumes of data to be stored, managed, and retrieved within an organization. Optimized for both structured and unstructured data, it is designed for quick and efficient ad hoc analysis regardless of complexity or size.

5.1.2 Customizable Dashboards and Ad Hoc Analysis

The system features tailor-made dashboards which help different departments to configure their data visualizations and reports according to the various needs that suit them. The ad hoc analysis feature is developed with the ultimate end-user having minimal technical expertise; it then empowers him to dynamically explore the data and provide real-time actionable insights.

5.1.3 Real-Time Collaboration and Communication

GenInsight also enables cross-department collaboration by offering a real-time chat feature, allowing teams to communicate and share insights directly within the platform, which may lead to better decision-making with regard to the speed of collaboration on data-driven insights among users.

5.1.4 Automated Machine Learning (AutoML)

GenInsight uses the AutoML functionality to help in the selection of machine learning models so that the users might find the most suitable ones to satisfy their needs for exploring data. The analysis work becomes less cumbersome because it will automatically suggest optimal models for the user, and insights can be drawn without needing advanced expertise in data science.

5.2 Design Considerations

This section outlines important considerations during the design phase of *GenInsight*, focusing on assumptions, constraints, and the design principles guiding the development process.

5.2.1 Assumptions and Dependencies

The design of *GenInsight* takes into consideration the following assumptions and dependencies:

- **Relating Software or Hardware**: It assumes compatibility with contemporary browsers, cloud infrastructure, including Azure, and common organizational hardware.
- **Operating Systems**: GenInsight is compatible with all the major OS systems, ranging from Windows through macOS to Linux.
- **Characteristics of End-users**: The users are assumed to have the least possible levels of data literacy and internet access to interact with the platform.
- **Possible Functionality Updates:** The design is acceptable for possible modification and updates in the future based on changes in added functionalities such as newly added analysis features and machine learning models to acknowledge changing data analytics requirements.

5.2.2 General Constraints

GenInsight operates within the following constraints that affect its design and functionality:

- **Hardware or Software Environment**: Optimized for general enterprise-level hardware and software environments in order to run in scale with Azure cloud services.
- **End-user Environment**: the system is designed for different environments under various organizations based on network speed variations, and specifications of other devices.
- **Standards Compliance:** *GenInsight* is designed to adhere to industry standards for data security, privacy, and interoperability, including GDPR. These include the requirements of security, such as using strong measures to protect organizational, sensitive data-encryptions, authentication protocols, and access controls based on role.
- **Performance Requirements**: The large dataset will be handled by an optimized system, ensuring quicker retrieval of data and real-time updates, thereby allowing for smoother user interaction.

5.2.3 Goals and Guidelines

GenInsight has been designed based on a number of very key goals and guidelines, enabling it to be efficient:

- **KISS Principle (”Keep It Simple Stupid”):** The approach keeps simplicity and user-friendliness first, since users must be able to find their way through dashboards, data visualizations, and analysis tools without much training.
- **Performance-Oriented**: Optimized to provide fast responses, the design is memory-efficient and returns data quickly to provide a seamless user experience.
- **Familiar Interface:** *GenInsight* is designed with a familiar interface and it reflects common design practices used in enterprise analytics tools; hence no steep learning curve for the users.

5.2.4 Development Methods

Development of GenInsight: Scaler, Robustness, and User Satisfaction

- **Agile Development Methodology:** *GenInsight* inherently follows the values of Agile by supporting incremental updates and rapid changes if requirements change. This software has also helped foster collaboration, continuous improvement, and responsiveness to user needs.
- **Iterative Design**: It follows an iterative design process with incorporation of user feedback in each iteration to fine-tune its features and improve functionality based on real-world usage.
- **User-Centered Design (UCD)**: The usage practices by *GenInsight* incorporate the principles of UCD, which ensure that their design system aligns ”harmoniously with user needs and expectations.” In UCD, this is a process of understanding the behaviors of users and what they wish to accomplish to create an intuitive and effective experience.

5.3 System Architecture

The description of the system architecture of GenInsight concentrates on the manner in which information flows through the platform and the different interfaces in the framework with respect to data storage, preprocessing, analysis, and use, to enable integration and operation of the components.

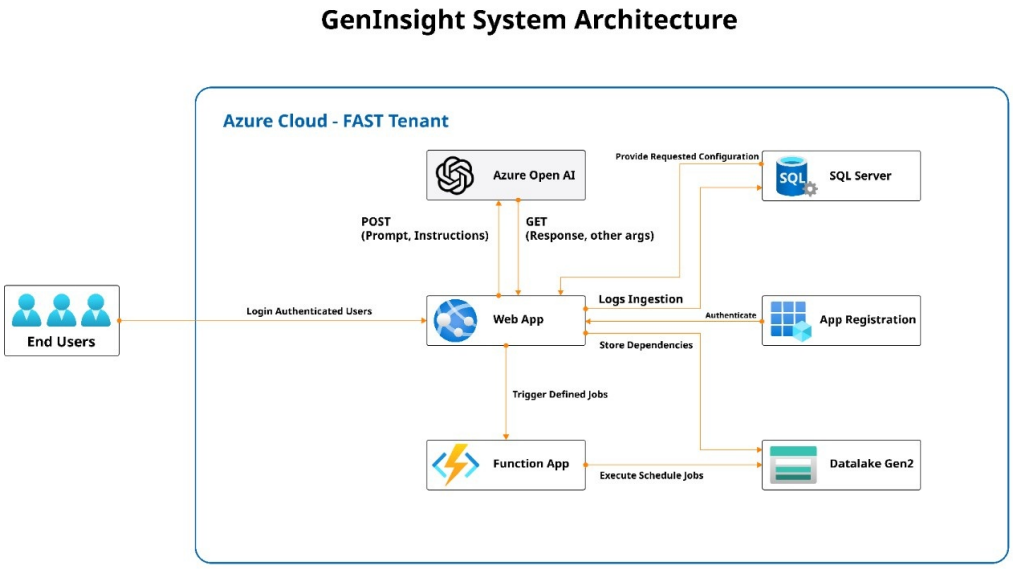

Figure 5.1: Architecture Diagram of GenInsight Platform
A comprehensive visualization of the architectural components and their interactions within the GenInsight platform, providing an overview of the system's design, data flow, and functionality.

The architecture of GenInsight can be described as an open system that features modules for handling inputs, storage, computation, and generation of reports integrated in a single Workplace. The architecture has cloud-related aspects and aspects that are interfaces for the users, emphasising secure access, good amount of processing power and integration with Azure.

Top-Level Responsibilities: At the highest level, GenInsight is responsible for the following major functionalities:

1. **User Authentication and Management**: Responsible for the program's secure authentication of the user and controlling their access.
2. **AI-Powered Analysis:** Utilises Azure Open AI for the parsing of user prompts and the creation of insights with data.
3. **Data Storage and Retrieval**: Makes and administrating the handling of data with the help of SQL Server and Datalake Gen2.
4. **Job Scheduling and Execution:** It works for coordinated functions and handles certain executions with the help of Function App.

Rationale: The architecture of service is scalable, modular and aligns with Azure services for integration. They are all aimed to solve some particular aspects of the data analysis and insights generation

which can be scaled and developed separately from others. The selected structure allows easy integration with forthcoming optimizations to components while being consistent with security, high-performance, and cloud-native ideals.

5.3.1 Subsystem Architecture

1. User Access Subsystem

- **Components:**
  - **App Registration:** Handles user authentication.
  - **Web App:** Provides the user interface and interaction points.
- **Interactions:**
  - End users authenticate through App Registration.
  - The Web App provides the interface for authenticated users.

2. AI Processing Subsystem

- **Components:**
  - **Azure Open AI:** Performs AI-related processing.
  - **Web App:** Acts as an intermediary for processing requests.
- **Interactions:**
  - The Web App sends POST requests with prompts and instructions to Azure Open AI.
  - Azure Open AI returns responses and other arguments via GET requests.

3. Data Management Subsystem

- **Components:**
  - **SQL Server:** Manages configurations and data storage.
  - **Datalake Gen2:** Provides data storage and management capabilities.
- **Interactions:**
  - SQL Server provides requested configurations to the Web App.
  - The Web App stores dependencies in SQL Server.
  - The Function App executes scheduled jobs on Datalake Gen2.

4. Job Execution Subsystem

- **Components:**
  - **Function App:** Executes scheduled jobs.
  - **Web App:** Triggers job executions.
- **Interactions:**
  - The Web App triggers defined jobs in the Function App.
  - The Function App executes scheduled jobs and interacts with Datalake Gen2.

5. Logging and Monitoring Subsystem

- **Components: Web App:** Facilitates log ingestion.
- **Interactions:** The Web App handles log ingestion for system monitoring and troubleshooting.

5.4 Architectural Strategies

The architectural strategies that will be employed to design and implement *GenInsight* are outlined below:

5.4.1 Tools and Software

Policy/ Strategy-1:Implementation of Python and Flask Framework

- **Description**: The Python and Flask environment will be utilized in developing the backend of *GenInsight*. This is because Python is considered an efficient language used in handling large-scale data analytics due to its immense library support on processing data, machine learning, and NLP. Flask is set to handle the task of using web frameworks for deploying machine learning models and APIs. The Python libraries used for EDA, NLP, and model training includes **pandas**, **NumPy**, **spaCy**, and **TensorFlow**. The frontend will be developed with **HTML**, **CSS**, and **JavaScript**, which will give the code an interactive and dynamic user interface. The development environment will use **VS Code**, because it is versatile and offers a wide range of extensions.
- **Requirements:** They do insist on demands such as having an active internet connection for cloud-hosted resources and databases besides using it for reach Azure services in order to scale and deploy.

5.4.2 Hardware

Policy/Tactic-2: System Requirements

- **Description**: To access and use the *GenInsight* platform, a personal computer (PC) or laptop with a reliable internet connection is required. The system should be able to run modern web browsers that support JavaScript, enabling users to access the platform's features seamlessly.
- **Requirements**: Adequate hardware must be available to support web browsers and real-time data processing without significant delays.

5.4.3 Future Plans

Policy/Tactic-3: Expansion and Integration of Machine Learning

- **Description**: *GenInsight* will be grown to integrate extra data analytics and machine learning features in the next iteration. This includes developing more algorithms, the ability to integrate more advanced models, and bringing predictive analytics capabilities onboard. Additionally, the tool will be enhanced for real-time data streaming and collaborative machine learning pipelines, where multiple users can work together to train, test, and deploy models with one another.
- **Future Plans**: The platform is to evolve with the inclusion of deeper tools of machine learning to automate specific aspects of the analysis of data and provide real-time insights across the different departments.

5.5 Domain Model/Class Diagram

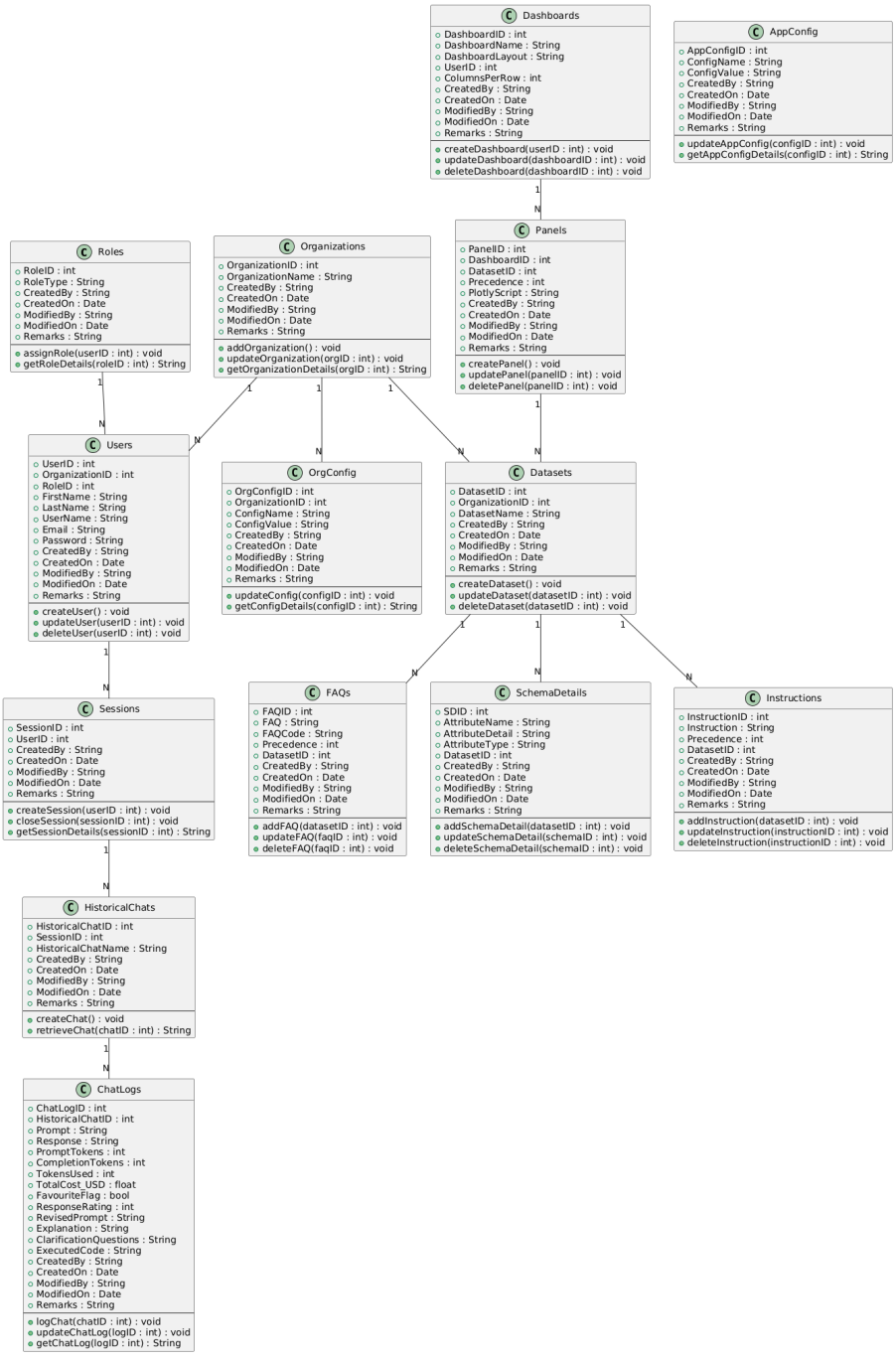

Figure 5.2: Class Diagram
The class diagram for *GenInsight*

5.6 Policies and Tactics

This section discusses the specific design policies and strategies that define detailed aspects of the GenInsight system. Questions such as tools used, engineering trade-offs made, coding standards adopted, test-

ing strategies applied, and plans for maintenance utilized influence this process. These are critical for guiding the development without fundamentally changing the general architecture.

5.6.1 Codebase Organization

Policy/Tactic-1: Modular Source Code Organization

- **Description**: The source code of *GenInsight* is modularly organized for different core functionalities such as processing the data, visualization, and machine learning. All the backend code is written in a structured way by using Flask both for API handling and data handling. All the frontend code is further subdivided into different components using HTML, CSS, and JavaScript.
- **Rationale:** This modular codebase makes development easier and allows one or more teams to work on different components. It improves maintainability, enhances the readability of code, and promotes the reuse of code throughout the project. It also encourages efficient collaboration among the members of a project.

5.6.2 Development Tools

Policy/Tactic-2: Standardized Development Tools

- **Description**: *GenInsight* will use standardized development tools, for example, VS Code to develop the code, Flask to develop the backend service, and popular Python libraries (pandas, TensorFlow, spaCy) for data analytics, machine learning, and NLP. This can provide the highest level of consistency, reliability and compatibility at all stages of development.
- **Rationale**: Known tools give stability, ease of debugging, and more efficient collaboration to developers. Standard tools and libraries make development easier to maintain by avoiding a risk of possible compatibility issues and help in smoother integration between parts.

5.6.3 Coding Guidelines

Policy/Tactic-3: Forced Coding Standards

- **Description**: *GenInsight* will make sure that there is consistency in coding style throughout the code, including the PEP 8 style guidelines for the Python backend codes and generally accepted best practices for HTML, CSS, and JavaScript of the frontend. It will include naming conventions along with proper indentation, and proper documentation of every module with inline comments.
- **Rationale:** This is because constant coding practices will enhance a reader to read the code, maintain it, and debug it easily. Homogeneous coding standards enhance teamwork collaboration,

hence moving the overall quality of the platform upwards.

5.6.4 Testing Strategies

Policy/Tactic-4: Structured Testing Framework

- **Description:** Comprehensive testing strategy as offered by *GenInsight*: unit tests for the individual components, integration testing for the system-wide functionalities and system testing to prove the performance and security of the system implementation. There will be the use of pytest, the automated testing tool to assure easy testing and continuous integration.
- **Evidence**: The reliability and the robustness of the platform must be thoroughly tested. Automated tests provide the developers with quick feedback, and thus bugs in the produced version are less likely to appear, so it is stable.

5.6.5 Documentation Practices

Policy/Tactic-5: Proper and Updated Documentation

- **Description:** *GenInsight* documents the entire scope of the API, backend logic, frontend design, and data workflows. That includes inline comments wherever possible, making APIs external documentation, and dev/user guides.
- **Reasoning**: Very good documentation will enable the subsequent developers, administrator, and users to understand the system easily. Very good documentation will also make it easier to induct new members to the team and reduces long-term maintenance and updates complexities.

5.7 Conclusion

Chapter 5 gives insight into the high-level and low-level design aspects of *GenInsight*. The system overview identifies key components as the data warehouse, customizable dashboards, ad-hoc analysis capabilities, and real-time collaborative tools. Some key design considerations regarding scalability, performance, and security give a good foundation for development.

System architecture is highlighted as the organization of the structural arrangement of the backend, which is composed of Python and Flask, and for the frontend, HTML, CSS, and JavaScript, cloud infrastructure, Azure. Architectural approaches pertaining to defining the tools, software, and hardware will drive the development; domain model and sequence diagrams help to illustrate how the system will work and can, in fact, understand the structure of *GenInsight*.

In addition, policies and strategies related to codebase structuring, development tools, coding standards,

testing approaches, and documentation processes illustrate a structured approach to development. All of these aspects integrated into one make up a holistic plan, and a steady platform established is one that is trustworthy, extensible, and maintainable and will satisfy the complex needs of data analytics in modern organizations.

This chapter serves as a foundation for Chapter 5 on the design and development of *GenInsight*, ensuring the implementation of the platform itself is structured, efficient, and future-proof.

Chapter 6 Implementation and Test Cases

The implementation details of the website created and interaction between the modules as well as the coding environment designed is described here in the implementation section 6.1.

6.1 Implementation

The implementation details of the components and algorithms of the GenInsight prototype are described in subsection 6.1.1.

6.1.1 Prototype Description

The GenInsight prototype, developed in the Final Year Project, is a comprehensive solution to data analytics and visualization. The frontend offers an intuitive user interface for both data analysts and administrators, so that interaction with datasets, data visualizations, and insights becomes seamless. The backend consists of a robust infrastructure for secure user authentication, data processing, and storage. Besides, integration of AutoML tools allows users to easily apply predictive models over the data object, thus streamlining data analysis tasks and paving the way for other more complex features later on.

6.1.2 Data Analytics and Visualization Module

This chapter discusses the integration of frontend and backend components within GenInsight's Data Analytics and Visualization module. The frontend is provided with an interactive user interface for the visualization of datasets, reporting, and the exploration of insights. Some of the features added are customizable dashboards, interactive charts, and real-time data updates for an improved user experience. The system in the backend processes and analyzes large datasets efficiently to ensure that there is good data management and storage and accurate and actionable insights for the user.

6.1.3 AutoML Module

This module describes the architecture and functionalities of the AutoML feature in GenInsight. With its automated machine learning tools, users can conduct rapid analyses on datasets as well as automatically select suitable machine learning models for tasks related to prediction. Through this functionality, it takes advantage of machine learning algorithms to make recommendations and subsequently apply these models so that non-expert users could carry out complex data analyses without knowledge of the underpinning processes. The AutoML module is therefore an important element for increasing the convenience of using the site and maximizing data-driven decision-making efficiency.

Chapter 7 Conclusion and Future Work

This chapter provides a summary of the progress made in the development of GenInsight and discusses potential directions for enhancing its capabilities in the future.

7.1 Conclusion

The development of GenInsight has reached considerable milestones toward the establishment of a wide-ranging and scalable data analytics platform. So far, its implementation includes integrating a powerful NLP-based chatbot that allows users to communicate with datasets through natural language queries. This feature, apart from streamlining user experience, extends access to advanced analytics to users having different technical expertise.

Further, the feature of AutoML has been developed, which selects, trains, and evaluates the machine learning model in a completely automated way. Users are empowered to gain predictive insights without needing deep knowledge of data science or machine learning. On the front-end, a fully functional, user-friendly interface has been developed, with easy navigation, data visualization, and dashboard management.

Together, these components constitute the heart of GenInsight, which will be capable of revolutionizing organizational decision-making processes by leveraging cutting-edge data analytics and AI capabilities. Its development has been focused on delivering seamless functionality while keeping it scalable and usable. Of course, all these accomplishments form a good foundation; however, there is much room for expansion and further enhancement in the features to unlock its full potential.

7.2 Future Work

There are quite a few areas in which GenInsight can be further improved and extended in order to address changing needs of the user. Some of the important directions of future work involve the NLP chatbot by enriching its capability to parse complex, multi-layered queries and giving more contextually meaningful answers. This will expand the scope of the chatbot and its ability to accept multilingual queries and therefore appeal to a more diverse group of users.

The AutoML feature can be enriched with customization options for model training and hyperparameter tuning, providing users with optimized machine learning models for specific use cases. This module can also be further developed to enhance its utility and accuracy in providing recommendations on preprocessing and feature engineering, based on the characteristics of the datasets uploaded.

On the collaborative front, GenInsight could benefit from the addition of real-time team collaboration tools. Features such as simultaneous editing of datasets and dashboards, coupled with robust version control mechanisms, would allow organizations to work cohesively on their data projects. Personalized dashboards with customizable widgets and layouts would also provide users with greater flexibility to tailor insights to their specific needs.

The scope of analytics can be widened by integrating support for real-time data streams and advanced predictive analytics, allowing the users to gain actionable insights on live data. The modules tailored to industry-specific needs, such as sentiment analysis, financial risk prediction, and retail analytics, will also expand its applicability across various domains.

As the platform grows, strengthening its data security and compliance mechanisms will be a critical aspect. Developing advanced features for privacy and ensuring conformance to global regulations will help GenInsight build trust as a solution in data-driven decision-making processes.

These strategic advances will bring GenInsight toward becoming more flexible, scalable, and user-focused-a platform from which the very complex and sophisticated needs of modern analytics and AI can be adequately addressed. Ongoing innovation and development for GenInsight continue its pursuit for leadership in organizational intelligence and decision support systems.

Bibliography

[1] J. Smith and J. Doe, "Forecasting market trends with ai-based models," *Journal of Artificial Intelligence Research*, 2022. https://doi.org/10.1016/j.jhealeco.2019.02.002 Accessed: October 05, 2024.

[2] A. Johnson and S. Williams, "The role of ai in decision support systems: A historical perspective," *International Journal of Decision Support Systems*, 2021. https://doi.org/10.1007/s10796-022-10282-5 Accessed: October 05, 2024.

[3] E. Williams and D. Brown, "Ai in healthcare: Predictors for early diagnostics and individualized treatment," *Journal of Medical AI Research*, 2020. https://ascpt.onlinelibrary.wiley.com/doi/full/10.1111/cts.12884 Accessed: October 05, 2024.

[4] M. Davis and E. Thompson, "Using ai for customer behavior predictions and inventory management," *Journal of Retail Analytics*, 2022. https://doi.org/10.1177/07439156211042372 Accessed: October 05, 2024.

[5] L. Chen and W. Zhang, "Sentiment analysis of social media data using ai," *Journal of Computational Social Science*, 2023. https://doi.org/10.1109/MIS.2010.75 Accessed: October 05, 2024.

[6] C. Miller and L. Garcia, "Application of ai for financial risk prediction: A case study," *Journal of Finance and AI*, 2021. https://doi.org/10.3390/ijfs9030039 Accessed: October 05, 2024.