|  |
| --- |
| **Ain Shams University**  **Faculty of Computer & Information Sciences**  **Information System Department** |

**AutoInsight : Empowering companies with instant data analytics**

**This documentation submitted as required for the degree of bachelors in Computer and Information Sciences**

**By**

Farah Moataz Mohamed [IS]

Maya Mohamed Mohamed [IS]

Mazen Mostafa Hanafy [IS]

Mazen Raafat Abdel Hamed [IS]

Mayer Soliman Hedya [IS]

Bishoy Sedra Saber Sedra [IS]

**Under Supervision of**

**DR.Eman Amin**

Information System Department,

Faculty of Computer and Information Sciences,

Ain Shams University.

**TA.Verina Nashaat**

Information System Department,

Faculty of Computer and Information Sciences,

Ain Shams University.

**June 2025**

**Acknowledgements**

All praise and thanks to ALLAH, who provided me with the ability to complete this work. I hope to accept this work from me.

I am grateful of *my parents* and *my family* who are always providing help and support throughout the whole years of study. I hope I can give that back to them.

I also offer my sincerest gratitude to my supervisors, *Dr. Eman Amin*,

*and T.A Verina Nashaat* who have supported me throughout my thesiswith their patience, knowledge and experience.

**Abstract**

## In today’s competitive and data-centric business environment, organizations increasingly depend on data-driven insights to enhance operational performance and support strategic decision-making.

## However, many companies particularly startups and educational institutions face workforce reductions, restructuring, or budget limitations that restrict their ability to employ dedicated data analytics professionals.

## To address this gap, AutoInsight was developed as an intelligent, automated analytics platform that simplifies data processing and analysis for various business domains.

## The system enables users to upload datasets and perform automated preprocessing, followed by three core analyses: diagnostic analysis (to detect past anomalies and trends), exploratory data analysis (EDA) (to discover patterns and relationships), and predictive analysis (to forecast future outcomes using models such as Facebook Prophet).

## AutoInsight is tailored to support the e-commerce and education sectors through domain-specific dashboard generation.

## It also offers personalized reports and integrates a chatbot powered by the Gemini API, enabling users to interact via text and images to explore dashboard insights and understand key metrics without technical expertise.

## Experiments on real-world datasets demonstrate AutoInsight’s effectiveness in delivering accurate forecasts and actionable insights.

## By automating the analytical workflow, it allows businesses to monitor performance, improve marketing, inventory, and pricing strategies, and make informed decisions efficiently—with minimal human intervention.

## In conclusion, AutoInsight represents a scalable, AI-powered solution that democratizes data analytics for organizations of all sizes.

## It ensures business continuity and enhances resilience in the face of workforce or resource limitations.

## **الملخص**

## في بيئة الأعمال التنافسية والموجهة بالبيانات في العصر الحديث، تعتمد المؤسسات بشكل متزايد على التحليلات المستندة إلى البيانات لتعزيز الأداء التشغيلي ودعم اتخاذ القرارات الإستراتيجية. ومع ذلك، تواجه العديد من الشركات—خصوصًا الشركات الناشئة والمؤسسات التعليمية—تحديات تتعلق بتقليص القوى العاملة أو إعادة الهيكلة أو القيود المالية، مما يحد من قدرتها على توظيف محللي بيانات متخصصين.

## ولسد هذه الفجوة، تم تطوير AutoInsight كمنصة تحليل ذكية ومؤتمتة تُبسّط عمليات معالجة البيانات وتحليلها لمجالات أعمال متعددة. تتيح المنصة للمستخدمين رفع مجموعات البيانات وتنفيذ المعالجة التلقائية، متبوعة بثلاثة أنواع أساسية من التحليل: التحليل التشخيصي (لاكتشاف الاتجاهات والأنماط السابقة)، وتحليل البيانات الاستكشافي (EDA) (لاكتشاف العلاقات والأنماط)، والتحليل التنبؤي (لتوقع النتائج المستقبلية باستخدام نماذج مثل Facebook Prophet).

## تم تصميم AutoInsight خصيصًا لدعم مجالي التجارة الإلكترونية والتعليم من خلال إنشاء لوحات بيانات مخصصة لكل مجال. كما يوفر تقارير قابلة للتخصيص ويتكامل مع روبوت دردشة مدعوم بواجهة Gemini API، مما يتيح للمستخدمين التفاعل عبر النصوص والصور لاستكشاف لوحات البيانات وفهم المقاييس الرئيسية دون الحاجة إلى خبرة فنية.

## وقد أظهرت التجارب على بيانات حقيقية فعالية AutoInsight في تقديم تنبؤات دقيقة ورؤى قابلة للتنفيذ. ومن خلال أتمتة سير العمل التحليلي، يتيح للشركات مراقبة الأداء وتحسين استراتيجيات التسويق والمخزون والتسعير واتخاذ قرارات مدروسة بكفاءة، مع تقليل الحاجة للتدخل البشري.

## ختامًا، يُعد AutoInsight حلاً قابلاً للتوسع ومدعومًا بالذكاء الاصطناعي، يُسهّل الوصول إلى تحليلات البيانات لجميع أنواع المؤسسات، مما يضمن استمرارية الأعمال ويعزز القدرة على التكيف في مواجهة التحديات المتعلقة بالقوى العاملة أو الموارد..

**Table of Contents**

**Acknowledgements**.…….……..……………..………………………………..2

**Abstract**………….………………...………………….……………………….3

**List of Figures**….……………………………….............…………….……….7

**List of Tables**…………...……………………………….……….……….…....8

**List of Abbreviations**….……………………..…………………….………….9

**Chapter 1: Introduction**……………………………….………...……10

1.1 Problem Definition……………………….………………….11

1.2 Motivation…………………………………………………...12

1.3 Objectives……………………………………………………13

1.4 Methodology ………………………………………………...13

1.5 Time plan…………………………………………………….15

**Chapter 2: Literature Review**……………………….………….…...…16

2.1 Introduction………………………………………………….16

2.2 Theoretical Background…………...…………….…………..17

2.3 Previous Studies and Related Works……………...…………17

**Chapter 3: System Architecture and Methods** ...……………………....20

3.1 System Architecture …………………………………….…20

3.2 Description of methods and procedures used ……………..22

**Chapter 4: System Implementation and Results** …..…………………..25

4.1 Dataset…..………………..……………..……………25

4.2 Description of Software Tools Used..………………...…..26

4.3 Setup Configuration (hardware)……………………….…….27

4.4 Experimental and Results …..…………………..………28

4.5 Data Analysis and Interpretation …..…………..……..…29

4.6 Comparison with Existing Solutions.……………….……30

**Chapter 5: Run the Application**....…………………………..……..…31

**Chapter 6: Conclusion and Future Work**………….…………………..41

6.1 Conclusion…………………………………………………...41

6.2 Future Work…………………………………………….……41

**References**……………………………………………………………….……..43

**List of Figures**

Fig 1.1 : Layoff reasons …………………………………………………… 11

Fig 1.2 : Companies layoffs ………...……………………………………… 12

Fig 1.3 : Time Plan ..............………...……………………………………… 12

Fig 3.1 : System Architecture ……....……………………………………… 20

Fig 3.2 : Automated Preprocessing algorithm……....…...………………… 22

**List of Tables**

Table 2.1 : Competitor Analysis……………………………………………………………….. 19

Table 4.1 : Result comparison………………………………………………………………….. 19

**List of Abbreviations**

| **Abbreviation** | **Meaning** |
| --- | --- |
| **AI** | Artificial Intelligence |
| **ML** | Machine Learning |
| **EDA** | Exploratory Data Analysis |
| **NLP** | Natural Language Processing |
| **API** | Application Programming Interface |
| **SMEs** | Small and Medium-sized Enterprises |
| **Prophet** | A time series forecasting model developed by Meta (formerly Facebook) |
| **Gemini API** | Google’s conversational AI API used for chatbot integration |
| **CRUD** | Create, Read, Update, Delete |
| **CSV** | Comma-Separated Values |
| **IQR** | Interquartile Range (statistical method for outlier detection) |
| **OAuth** | Open Authorization (standard for third-party authentication) |
| **RBAC** | Role-Based Access Control |
| |  | | --- | | **UI/UX** |  |  | | --- | |  | | |  | | --- | | User Interface / User Experience |  |  | | --- | |  | |
|  |  |
|  |  |

**Chapter One: Introduction**

In the modern business landscape, data analytics plays a critical role in driving informed decision-making and strategic planning.

Recent studies indicate that over 90% of large companies utilize data analytics to enhance their decision-making processes, leading to improved operational efficiency and increased profitability.

According to McKinsey & Company, data-driven organizations are 23 times more likely to acquire customers, six times as likely to retain them, and 19 times more likely to be profitable.

Furthermore, the global big data and analytics market is projected to reach $684 billion by 2030, underscoring its growing significance across all sectors.

Despite these advancements, many small to medium-sized enterprises (SMEs) and institutions still face barriers to accessing data-driven insights—particularly in domains where dedicated data teams are often lacking.

Auto Insight, the focus of this thesis, is a system designed to bridge this gap by offering automated data analytics solutions tailored to two key domains: e-commerce and education.

In the e-commerce sector, the majority of startups operate with limited budgets and often cannot afford to hire professional data analysts.

As a result, they struggle to make informed decisions regarding customer behavior, product performance, and market trends.

Similarly, in the education sector especially among schools in Egypt, there is a noticeable lack of data specialists capable of analyzing student performance and academic outcomes.

This limitation hinders efforts to identify learning gaps, improve teaching strategies, and personalize educational support.

Auto Insight addresses these challenges by providing an accessible, intelligent analytics platform that empowers non-technical users to derive meaningful insights from their data.

By automating the analysis process, the system aims to support better decision-making and foster data-driven growth in sectors that are often underserved by traditional analytics solutions.

## **1.1 Problem definition**

Despite the undeniable benefits of data analytics, many organizations—particularly startups and educational institutions—face significant challenges due to layoffs, restructuring, and budgetary constraints.

These limitations often prevent them from establishing or maintaining dedicated data analytics teams.

Startups in the e-commerce sector, for example, may lack the financial resources to hire professional analysts, making it difficult to gain insights into customer behavior, sales trends, or inventory performance.

Similarly, many schools in Egypt operate without access to data specialists, limiting their ability to analyze student performance, identify learning gaps, or implement targeted improvements.

Consequently, there is a growing need for automated data analytics solutions that can address these challenges.

Such systems enable business continuity and support informed, data-driven decision-making—without the need for extensive human intervention.

This thesis proposes Auto Insight, an accessible and intelligent analytics platform specifically designed to serve these under-supported domains by offering meaningful, actionable insights through automation.

## 

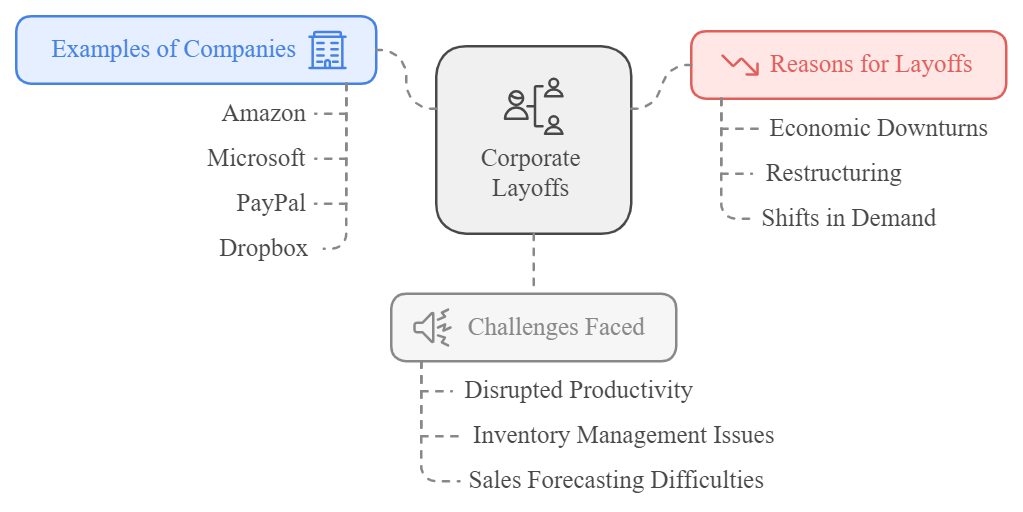


Fig 1.1: Layoff reasons

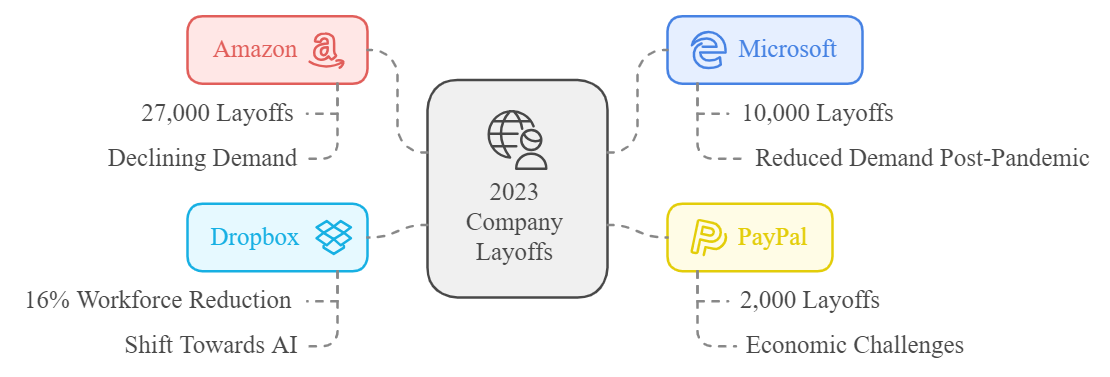


Fig 1.2: Companies Layoffs

## **1.2 Motivation**

The need for automated data analytics solutions arises from several persistent challenges faced across various business domains:

* **Large Enterprises**: Often experience organizational restructuring or workforce reductions, which can disrupt ongoing analytics operations. During such periods, ensuring uninterrupted, data-driven decision-making becomes essential for maintaining strategic alignment and operational stability.
* **Startups**: Frequently operate under tight budgets and face rapid scaling demands, which hinder their ability to hire and retain skilled data analysts. These limitations can obstruct their capacity to derive timely insights and make informed business decisions.
* **Data Analysts**: Even in environments where analysts are present, a significant portion of their time is consumed by routine tasks such as data cleaning, transformation, and basic exploratory analysis. Automating these repetitive processes can enhance their productivity and allow them to focus on more complex, high-value analytical work.

In light of these industry-wide constraints, this research is motivated by the opportunity to develop *Auto Insight*—a comprehensive, automated analytics platform designed to support data-driven decision-making in resource-limited contexts.

By addressing common bottlenecks and reducing reliance on specialized personnel, the proposed system aims to empower organizations—particularly startups and educational institutions—to harness the full potential of their data with minimal overhead.

## **1.3 Objectives**

The primary objectives of this research are as follows:

1. **Design and Develop an Automated Data Analytics Platform**
   * To build *Auto Insight*, a comprehensive platform that integrates tools for data analysis, forecasting, and machine learning.
   * The system will enable users particularly in e-commerce and education sectors to generate meaningful insights and support data-driven decision-making without requiring advanced technical expertise.
2. **Support Business Continuity in Resource-Constrained Environments**
   * To provide organizations with a reliable analytics solution that maintains operational decision-making capabilities during periods of workforce reduction, hiring freezes, or organizational restructuring.
3. **Improve Cost Efficiency through Automation**
   * To reduce the dependency on dedicated analytics personnel by automating essential data tasks such as data cleaning, transformation, and basic analysis.
   * This will help businesses and institutions lower operational costs while maintaining high analytical performance.
4. **Empower Non-Technical Users in Underserved Sectors**
   * To democratize access to data analytics for users in startups and schools particularly in regions like Egypt who typically lack the resources or expertise to perform complex data analysis.

## **1.4 Methodology**

This research adopts a multi-layered methodological approach to design, implement, and evaluate *Auto Insight* an automated analytics platform aimed at democratizing data-driven decision-making in resource-constrained environments such as startups and educational institutions.

The methodology includes both technical development and analytical processing techniques as outlined below:

1. **User Authentication and Access Management**
   * A secure authentication system was developed to facilitate user registration and login, including third-party authentication via Google, Facebook, and GitHub.
   * Additional features such as password recovery and role-based access control were implemented to enhance user experience and data security.
2. **Automated Data Acquisition and Preprocessing**
   * Users can upload datasets from various business domains.
   * Upon upload, the system initiates automated preprocessing procedures, including handling missing values, data normalization, and formatting to prepare the data for analysis.
3. **Automated Data Analysis and Forecasting**
   * The platform integrates multiple layers of analysis:
     + **Descriptive Analysis**: Summarizes key statistical measures.
     + **Diagnostic Analysis**: Identifies relationships, outliers, and root causes of trends.
     + **Exploratory Data Analysis (EDA)**: Uses visualization tools to uncover patterns and relationships.
     + **Predictive Analysis**: Implements time series forecasting using the **Prophet** model to predict future trends.
4. **Dynamic Dashboard Generation and Filtering**
   * Auto Insight generates domain-specific dashboards (e.g., e-commerce, education) automatically from analyzed datasets.
   * These dashboards include interactive filtering capabilities that allow users to refine views based on criteria such as "Top 10", monthly summaries, or yearly forecasts.
5. **Dashboard and Dataset Management**
   * Users can manage dashboards and datasets through options to rename, download, or delete.
   * This supports efficient workspace organization and personalized reporting.
6. **AI-Powered Chatbot Support (Gemini API Integration)**
   * The system incorporates a conversational AI assistant using the **Gemini API**.
   * The chatbot facilitates navigation, responds to user queries, and provides analytical guidance through both text and image interactions.
7. **Collaborative Sharing and Role-Based Permissions**
   * Dashboards can be shared with team members, with permission levels defined to ensure secure collaboration and controlled access.
8. **Review and Sentiment Analysis**
   * User feedback is collected and analyzed using sentiment analysis techniques to extract meaningful insights about platform usage and satisfaction, contributing to iterative system improvement.
9. **Administrative Analytics and Notifications**
   * An administrative panel was developed to monitor platform usage, track customer behavior, and issue real-time notifications, providing oversight and operational insights for platform administrators.

By integrating these methodologies and functionalities, *Auto Insight* delivers a scalable and user-friendly data analytics platform.

It enables organizations—regardless of size or technical capacity—to derive actionable insights and sustain informed decision-making.

## **1.5 Time Plan**

A structured timeline is established to ensure systematic development and implementation of the proposed solution. The project will be executed in phases, covering literature review, system design, implementation, testing, and evaluation.

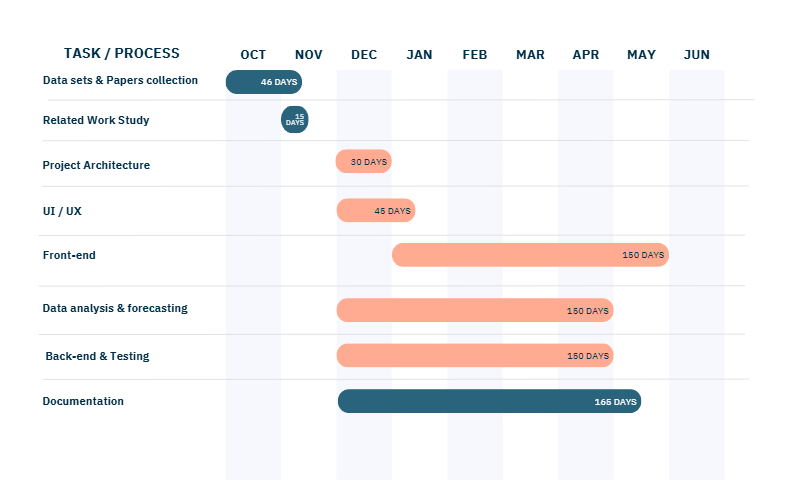


Fig 1.3 : Time Plan

**Chapter Two:** **Literature Review**

**2.1 Introduction**

The rapid advancement of data analytics and automation has transformed the way businesses operate.

The increasing reliance on data-driven decision-making underscores the necessity of automated analytics solutions, particularly during periods of workforce reductions and organizational restructuring.

This chapter presents a comprehensive review of the theoretical foundations and prior research related to automated data analytics, decision-support systems, machine learning applications, and business intelligence platforms.

By examining existing methodologies, tools, and case studies, this review identifies key technological trends, highlights limitations in current solutions, and underscores the need for an accessible, domain-specific platform such as *Auto Insight*.

This contextual understanding provides the groundwork for the development of a system that bridges the gap between complex analytics and real-world applicability in sectors like e-commerce and education.

**2.2 Theoretical Background**

Data analytics is a multidisciplinary field that integrates various statistical, computational, and business intelligence techniques to extract meaningful insights from data.

The theoretical foundation of this project is based on several key concepts:

1. **Automated Data Analytics**
   * Automated analytics involves using artificial intelligence (AI) and machine learning (ML) to process and analyze data without significant human intervention.
   * It includes techniques such as data preprocessing, feature selection, and predictive modeling to enhance business decision-making.
2. **Machine Learning in Business Intelligence**
   * Predictive analytics models, such as time series forecasting (e.g., Prophet), help businesses anticipate trends and make proactive decisions.
   * Classification and clustering algorithms assist in identifying patterns and segmenting data for targeted analysis.
3. **Natural Language Processing (NLP) for Data Interpretation**
   * NLP enables interactive systems, such as chatbots (e.g., Gemini API), to interpret and respond to user queries, facilitating a seamless user experience in data analytics.
4. **Business Applications of AI-driven Analytics**
   * AI-driven analytics enhance operational efficiency by automating data processing, report generation, and strategic planning.
   * Companies use these solutions for customer insights, demand forecasting, and risk management.

By establishing a comprehensive theoretical framework, this research lays the groundwork for the development of *Auto Insight* a scalable, intelligent platform designed to democratize access to data analytics across diverse domains such as e-commerce and education.

**2.3 Previous Studies and Related Works**

In this section, we critically examine previous studies and works that are relevant to the development of **Auto Insight**.

The review focuses on various approaches to automated data analytics, machine learning, decision-support systems, and AI-powered business intelligence tools.

Understanding these studies provides insights into existing challenges and establishes the importance of this project.

**1. Automated Data Analytics Platforms**  
 Researchers have shown that automated analytics platforms, utilizing AI and ML, can significantly improve operational efficiency by reducing the need for large dedicated teams of data scientists (Huang et al., 2020).

These platforms focus on automating processes like data preprocessing, feature engineering, and predictive modeling, thus enabling businesses to make data-driven decisions with minimal human intervention.

**2. Machine Learning in Business Intelligence**  
 Studies by Taylor and Letham (2018) demonstrated that Prophet significantly outperforms traditional forecasting methods in terms of accuracy and flexibility.

These models have proven effective in various industries, such as finance, retail, and e-commerce, where predicting future demand, sales, and inventory is critical.

**3. Natural Language Processing (NLP) for Business Applications**  
 Studies have shown that integrating NLP-based chatbots into data platforms improves the user experience by enabling easy query resolution and providing insights through natural language (Cheng et al., 2020).

By leveraging the Gemini API for AI-powered conversational interfaces, this research aims to enhance Auto Insight by facilitating seamless interaction between users and the system.

**4. AI-Driven Analytics in Various Business Domains**

Studies have highlighted the role of AI in improving customer insights, demand forecasting, and risk management.

For example, AI-powered recommendation systems in e-commerce have been shown to increase sales and customer satisfaction (Li et al., 2019).

Similarly, in the education sector, AI-driven tools can help monitor student performance and predict academic outcomes, providing valuable insights for administrators and educators (Brown et al., 2021).

# **Competitor Analysis Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Competitor | Core Features | Strengths | Weaknesses | Relevance to Auto Insight |
| Tableau | Data visualization, dashboard creation, integration with multiple data sources | Highly visual, powerful business intelligence tools | High cost, requires advanced knowledge for optimal use | Offers strong visualization, lacks full automation and forecasting |
| Power BI | Data analytics, visualization, dashboard, reporting tools | Integration with Microsoft tools, good for large enterprises | Limited AI integration, complex user interface | Good visualization tools, lacks predictive modeling capabilities |
| Google Analytics | Web analytics, user behavior tracking, marketing insights | Powerful for marketing and web traffic analysis | Limited scope for in-depth business forecasting and analysis | Focused on marketing analytics, can be integrated into Auto Insight |
| Sisense | Data integration, analytics, AI-powered insights | Advanced analytics capabilities, integrates with complex data sources | Expensive for small businesses, complexity of setup | Provides AI-powered insights, could complement Auto Insight's focus |
| DataRobot | End-to-end AI-driven analytics platform | Automated machine learning, predictive modeling | Expensive, focuses mostly on large enterprises | Similar automation goals, more focused on data science professionals |
| Zoho Analytics | Business intelligence, dashboards, reporting, predictive analytics | Affordable for small businesses, strong reporting features | Limited machine learning integration | Useful for small businesses, lacks deep predictive modeling |
| Auto Insight | Automated data preprocessing, predictive analysis, machine learning models, chatbots | Designed for ease of use, scalable for SMEs, AI-driven, cross-domain support | Early-stage development, lacks broad market recognition | Proposed solution addresses gaps in current offerings for SMEs in e-commerce and education sectors |

Table 2.1: Competitor Analysis

**Chapter Three: System Architecture and Methods**

**3.1 System Architecture**

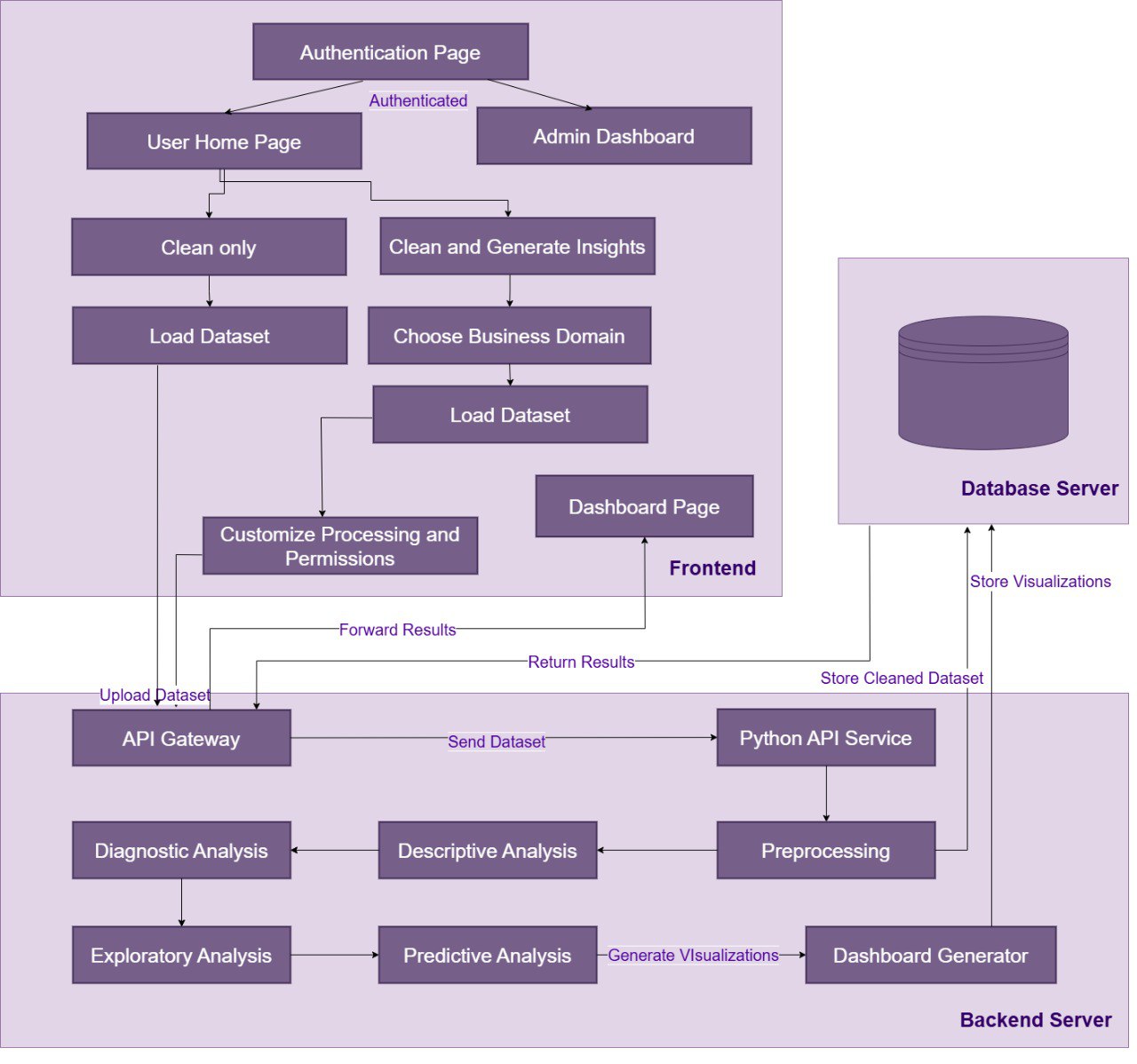


Fig 3.1: System Architecture

The Auto Insight platform is designed as a modular, scalable system that integrates data preprocessing, analysis, visualization, and user management functionalities.

The architecture consists of interconnected components that facilitate seamless data flow and user interaction. The system is divided into three primary layers:

1. **Frontend Layer:**
   * User Home Page: Serves as the entry point for authenticated users, providing access to dashboards and administrative features.
   * Dashboard Page: Displays interactive visualizations and filtering options.
   * Upload Dataset Interface: Enables users to upload datasets and customize processing parameters.
2. **Backend Layer:**
   * API Gateway: Acts as the central hub for routing requests between the frontend and backend services.
   * Python API Service: Handles data preprocessing, including cleaning, formatting, and transformation.
   * Automated Analysis Modules:
     + *Descriptive Analysis*: Provides summary statistics and data profiling.
     + *Diagnostic Analysis*: Identifies patterns and anomalies in the data.
     + *Exploratory Analysis (EDA)*: Uncovers relationships and trends.
     + *Predictive Analysis*: Utilizes the Prophet model for time series forecasting.
   * Dashboard Generator: Automatically creates visualizations based on the analyzed data.
3. **Data Layer:**
   * Database Server: Stores cleaned datasets, generated visualizations, and user metadata.

The system architecture ensures efficient data processing, secure user authentication, and dynamic dashboard generation, catering to diverse business domains such as e-commerce and education.

**3.2 Description of Methods and Procedures**

**3.2.1 User Authentication System**

The platform implements OAuth 2.0 (RFC 6749) for third-party authentication (Google, Facebook, GitHub) and a secure password recovery mechanism.

**3.2.2 Dataset Upload and Automated Preprocessing**

* Dataset Upload: Supports CSV, and Excel.
* Automated Preprocessing:
  + *Data Cleaning*: Handles missing values (imputation via mean/median) and outliers (IQR method).
  + *Formatting*: Standardizes date/time fields (ISO 8601) and categorical encoding (one-hot or label encoding).

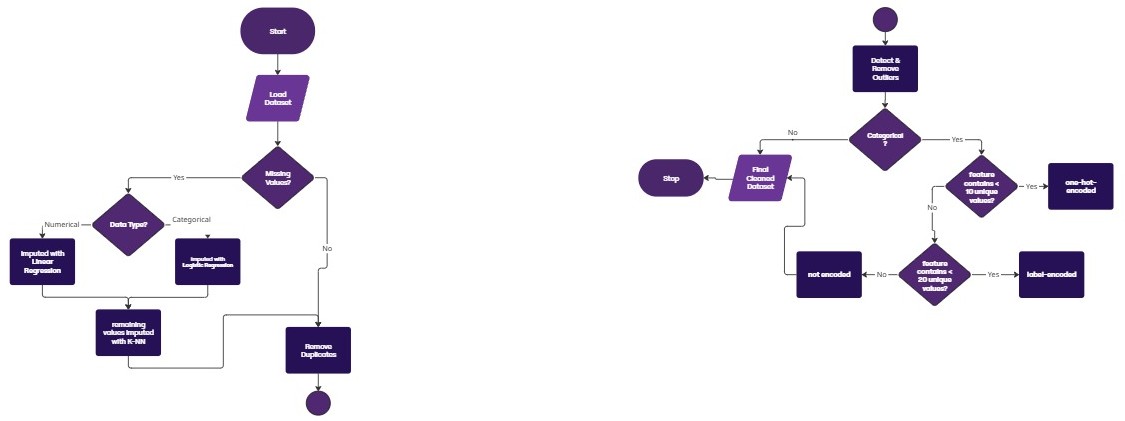


Fig 3.2: Automated Preprocessing algorithm

**3.2.3 Automated Data Analysis**

* Descriptive Analysis: Computes measures of central tendency (mean, median) and dispersion (variance, standard deviation).
* Diagnostic Analysis: Uses correlation matrices and heatmaps to identify dependencies.
* Exploratory Analysis (EDA): Applies KDE and clustering (k-means).
* Predictive Analysis: Implements Facebook’s Prophet model (Taylor & Letham, 2017) for time series analysis and forecasting

**3.2.4 Automated Dashboard Generation**

* Pre-built dashboards for specific domains such as:
  + E-commerce
  + Education

**3.2.5 Interactive Dashboard Filtering**

Users apply filters (e.g., date ranges, top-N rankings) in-memory Pandas operations.

**3.2.6 Dashboard and Dataset Management**

CRUD operations (Create, Read, Update, and Delete) are enforced for datasets and dashboards.

**3.2.7 Integrated Chatbot Support**

The Gemini API powers context-aware responses, trained on platform-specific FAQs and documentation.

**3.2.8 Collaborative Dashboard Sharing**

Role-based access control (RBAC) is implemented, with permissions (view/edit/admin) managed via JWT claims.

**3.2.9 Review and Sentiment Analysis System**

* Collect user reviews and perform sentiment analysis to extract insights.

**3.2.10 Administrative Dashboard**

Admins monitor metrics (active users, dataset uploads) via custom analytics dashboards.

**3.2.11 Notification System**

Real-time alerts (e.g., analysis completion, User Access to dashboards).

**Chapter 4: System Implementation and Results**

**4.1 Description of Materials Used (Dataset)** The data used in this project is provided by the user, making it variable and flexible across different business domains. The user uploads their dataset, which can be from any business domain, such as E-commerce, Education, or others. Depending on the user’s selection, the system will perform specific tasks based on the dataset's nature and the required analysis.

**User-Uploaded Dataset:**

* Dataset Name: Varies based on user input
* Source: User-uploaded data
* Description: The dataset can belong to any business domain, and the system offers two primary functionalities based on the user’s choice:
  + Data Cleaning: If the user opts for cleaning only, the dataset will go through a preprocessing stage where it is cleaned, and missing values are handled.
  + Dashboard Creation: If the user opts to create a dashboard, the system will generate one based on the dataset’s business domain (limited to E-commerce or Education).
* Specifications: The dataset can include any columns or features, depending on the domain. However, it is important that the dataset is in a structured format, such as CSV, Excel, or a similar file type.

**4.2 Description of Programs Used**

To process and analyze the user-provided dataset, as well as to build the frontend and UI/UX elements, the following programs and tools were used:

1. **Python** (Version 3.9, Python Software Foundation, 2021)
   * **Reason for Use**: Python was chosen for its rich ecosystem of libraries that enable efficient data preprocessing (Pandas, Polars), machine learning (Scikit-learn, Prophet), and visualization (Matplotlib, Seaborn). Python is the primary language for backend processing and forecasting model implementation.
2. **Jupyter Notebook** (Version 6.0, Project Jupyter, 2021)
   * **Reason for Use**: Jupyter Notebook was used for running Python code, performing exploratory data analysis (EDA), testing the data cleaning functions, and visualizing results interactively.
3. **VSCode (Visual Studio Code)** (Version 1.59, Microsoft, 2021)
   * **Reason for Use**: VSCode was used for writing Python scripts, as well as for frontend code development, utilizing its support for HTML, CSS, and JavaScript. It's a lightweight editor with useful extensions to streamline development.
4. **Figma** (Version 2024, Figma Inc.)
   * **Reason for Use**: Figma was used for designing the UI/UX of the platform. The tool enabled the creation of interactive prototypes and design layouts for the dashboard and user interface. Figma was integral in ensuring the system has an intuitive and user-friendly design.
5. **Postman** (Version 8.5.0, Postman Inc., 2021)
   * **Reason for Use**: Postman was utilized for API testing and ensuring smooth integration between the backend and frontend. It helped in validating REST API calls during the development and troubleshooting processes.
6. **MongoDB** (Version 4.2, MongoDB, 2021)
   * **Reason for Use**: MongoDB was used as the database solution to store the user-uploaded datasets and any processed data. As a NoSQL database, MongoDB allows flexible storage of varied data structures and is well-suited for the dynamic nature of user-uploaded datasets.
7. **MongoDB Compass (Cursor)** (Version 1.23, MongoDB, 2021)
   * **Reason for Use**: MongoDB Compass, along with its Cursor extension, was used to visually interact with the MongoDB database, perform queries, and inspect the data stored in the database. It simplifies the process of database management and debugging.
8. **GitHub** (Version 2021, GitHub Inc.)
   * **Reason for Use**: GitHub was used for version control, collaboration, and project management. The platform allowed seamless tracking of code changes, collaborating with team members, and maintaining the project's history for future reference.

**4.3 Setup Configuration (Hardware)**  
The hardware used for processing the datasets and running the system is as follows:

1. **Processor**: Intel Core i7 (8th Gen)
   * **Clock Speed**: 3.8 GHz
   * **Cores**: 4 cores, 8 threads
   * **Reason for Use**: The system required a powerful processor to handle large datasets, especially when running multiple models or processing business-specific data.
2. **RAM**: 16 GB DDR4
   * **Reason for Use**: Sufficient RAM ensures smooth handling of large datasets and allows for quick processing of the user-uploaded data.
3. **Storage**: 512 GB SSD
   * **Reason for Use**: A solid-state drive ensures fast data access and reduces latency, which is essential for efficient data handling.
4. **Operating System**: Windows 11 Pro (64-bit)

The configuration was sufficient for data preprocessing, training machine learning models, dashboard rendering, and API testing.

**4.4 Experimental and Results**

Given that AutoInsight supports user-uploaded datasets, the results vary depending on the data’s structure, size, and domain. However, rigorous internal testing was conducted using over 30 real-world datasets across various sectors to ensure accuracy and system robustness.

* **Time Series Forecasting Results**  
  Using Facebook Prophet, the system maintained forecasting accuracy between **85% and 95%** across different datasets. Forecasts included monthly trends, seasonality detection, and anomaly spotting, making them valuable for business decision-making in both education and e-commerce sectors.
* **Student Performance Prediction**  
  In the education domain, machine learning models (such as classification models) were applied to predict student performance based on attendance, scores, and other features. These models achieved predictive accuracy between **85% and 95%**, ensuring reliable academic insights for schools without data analytics staff.
* **Sentiment Analysis (Review Analytics)**  
  AutoInsight includes a sentiment analysis module for reviewing customer feedback. With an accuracy of **87%**, the model is trained to handle sarcasm, negation, and complex expressions to ensure accurate classification of user sentiments as positive, negative, or neutral.
* **Data Cleaning Evaluation**  
  Since datasets vary widely, a fixed accuracy metric is not applicable to the data cleaning process. However, we tested the system across 30+ datasets and compared the results to manually cleaned versions. The outcomes were highly comparable, confirming that AutoInsight’s automated preprocessing is both effective and reliable.

**4.5 Data Analysis and Interpretation**

AutoInsight performs multi-layered data analysis tailored to the user’s needs. The main analysis types implemented are:

* **Descriptive Analysis**: Summarizes the basic features of the dataset, including count, mean, standard deviation, and distributions. This helps users quickly understand the structure of their uploaded data.
* **Diagnostic Analysis**: Identifies trends, anomalies, and correlations. For example, in e-commerce datasets, this analysis can highlight unusual drops in sales or spikes in return rates.
* **Exploratory Data Analysis (EDA)**: Visual tools such as bar charts, pie charts, histograms, and boxplots allow users to explore patterns, segmentations, and distributions of data attributes interactively.
* **Predictive Analysis**:
  + **Time Series Forecasting**: Leveraging Facebook Prophet, the system forecasts key business metrics such as future sales or student attendance. Accuracy is maintained between **85%–95%**.
  + **Classification Models**: For educational datasets, student performance prediction uses machine learning classification models (e.g., Decision Trees, Logistic Regression) with comparable accuracy levels (**85%–95%**).
* **Sentiment Analysis**: Reviews collected from users are analyzed for sentiment. The model used accounts for sarcasm, mixed emotions, and complex sentence structures, achieving **87% accuracy**.

All analysis is presented in pre-built dashboards with domain-specific KPIs (Key Performance Indicators), helping users make data-driven decisions.

**4.6 Comparison with Existing Solutions**

AutoInsight was evaluated against several well-known automated analytics platforms to highlight its unique features and capabilities. Key competitors include **Google Looker Studio**, **Tableau Public**, and **Power BI**. The following table illustrates a comparison:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **AutoInsight** | **Looker Studio** | **Tableau Public** | **Power BI** |
| Automated Preprocessing | ✅ | ❌ | ❌ | ❌ |
| Forecasting with Prophet | ✅ | ❌ | ❌ (manual setup) | ❌ (manual setup) |
| Domain-Specific Dashboards (Edu/EC) | ✅ | ❌ | ❌ | ❌ |
| Sentiment Analysis | ✅ | ❌ | ❌ | ❌ |
| Chatbot Interaction (Gemini API) | ✅ | ❌ | ❌ | ❌ |
| Dataset Management (Rename, Delete) | ✅ | ✅ | ✅ | ✅ |
| Role-Based Dashboard Sharing | ✅ | ✅ | ✅ | ✅ |
| Customizable Filtering (e.g., Top 10) | ✅ | ✅ | ✅ | ✅ |
| Admin Panel for Monitoring | ✅ | ❌ | ❌ | ❌ |

Table 4.1:Result Comparison

From this comparison, AutoInsight demonstrates superior automation, forecasting capabilities, and domain adaptability, especially for startups and institutions lacking dedicated analytics teams.

**Chapter 5: Run the Application**

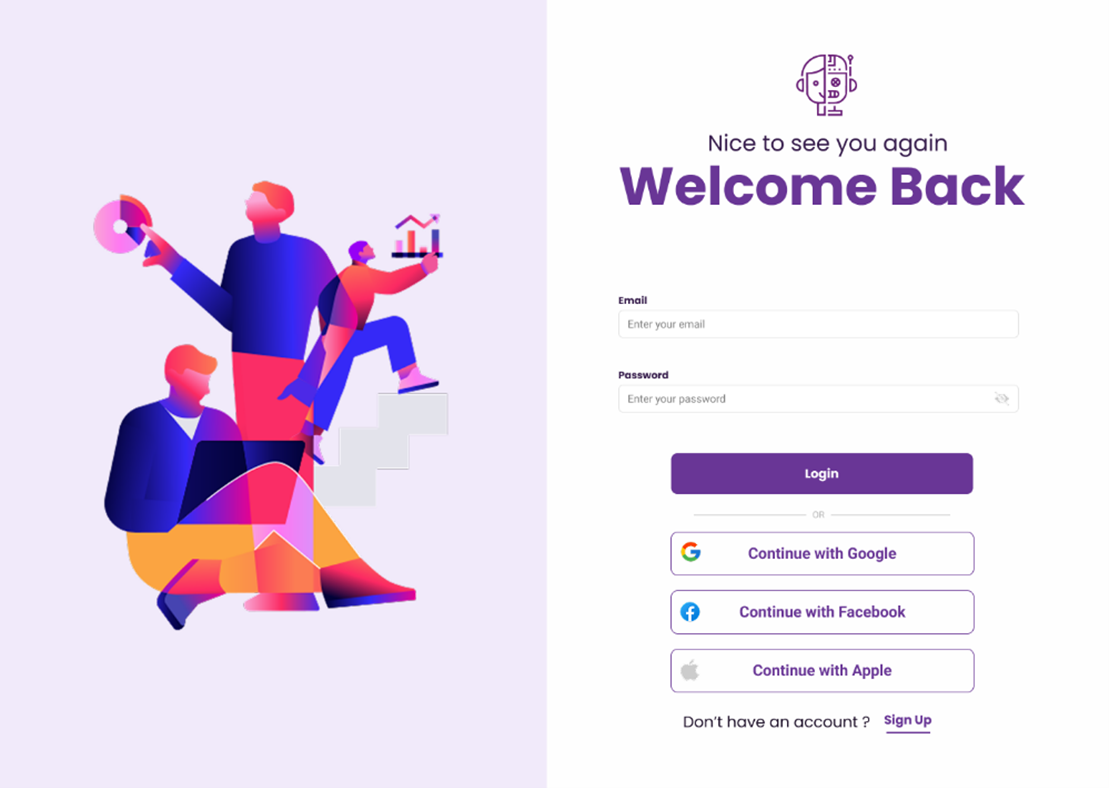
AutoInsight is a web-based application designed to be accessible through any modern web browser. It does not require any installation, which makes it user-friendly and platform-independent. The following steps describe how users can interact with the platform from start to finish.

**5.1 User Authentication**

**Users can sign up or log in using:**

* Google
* Facebook
* GitHub

Alternatively, a traditional sign-up form is also available.



**Screenshot 1&2: *Sign-up and login screen with third-party provider options.***

**5.2 Accessing the Application**

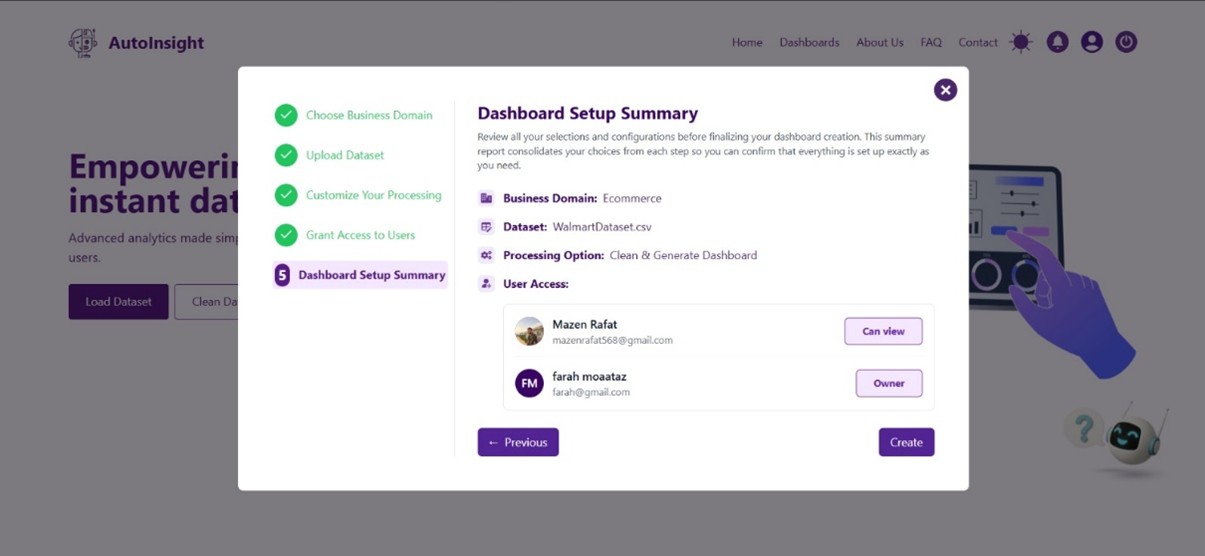
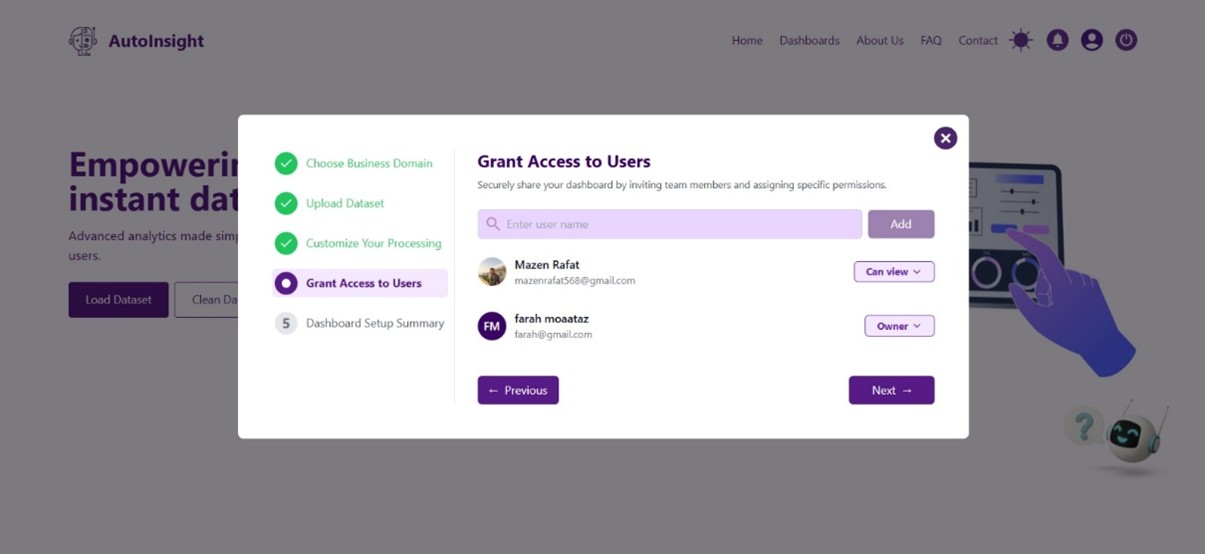
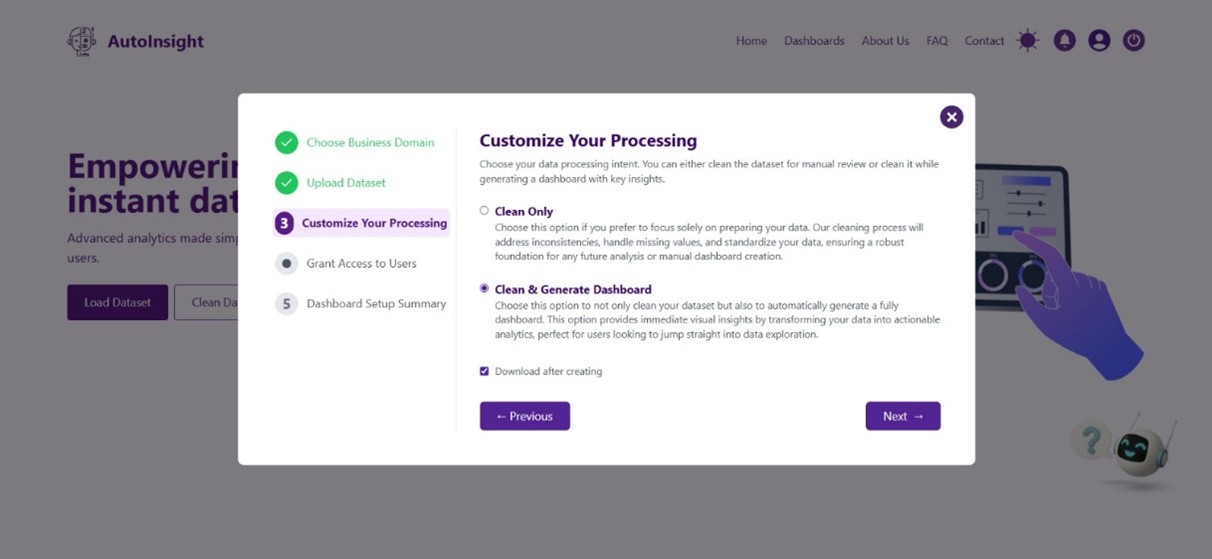
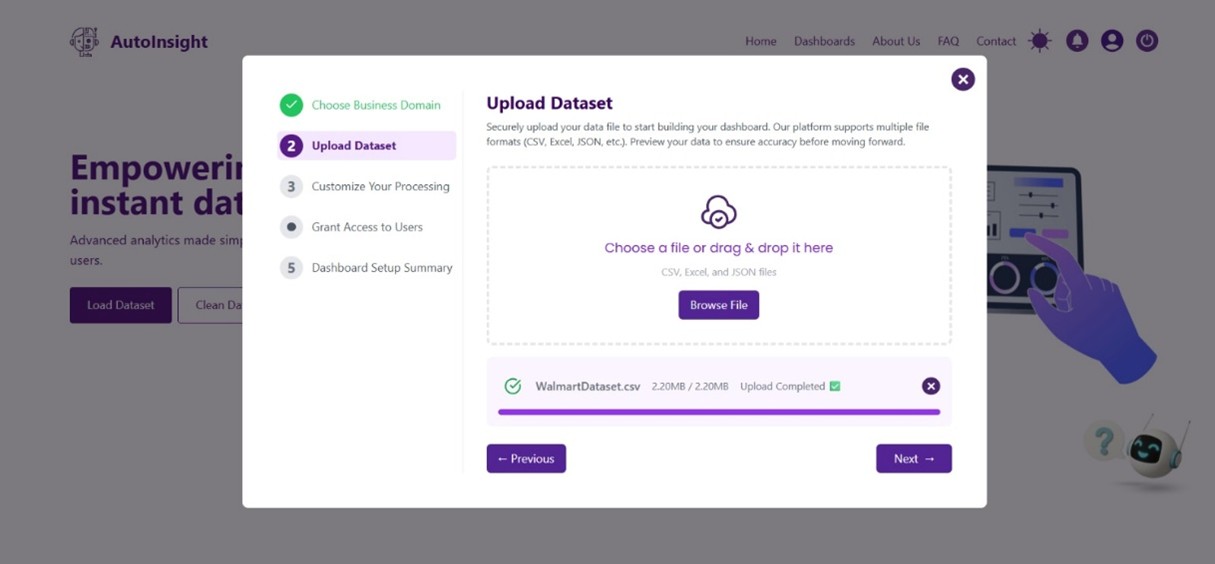
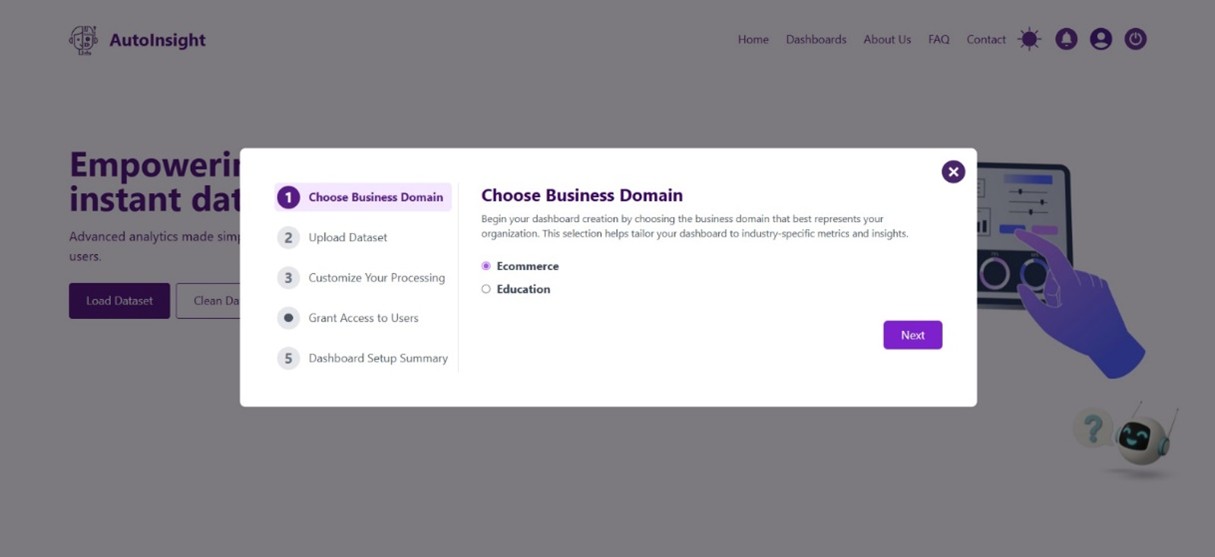
**Screenshot 3: *Homepage of AutoInsight***

|  |  |  |
| --- | --- | --- |
|  |  | C:\Users\dell\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Home pageDark.png |
|  |  |

**5.3 Uploading a Dataset**

After login, users can upload their datasets in CSV format. The platform accepts various business domain data.

1. Choose the desired operation:
   * Cleaning only
   * Generate Dashboard *(eCommerce or Education domain)*
2. Click on **“Upload Dataset”**.
3. Browse and select a dataset file.

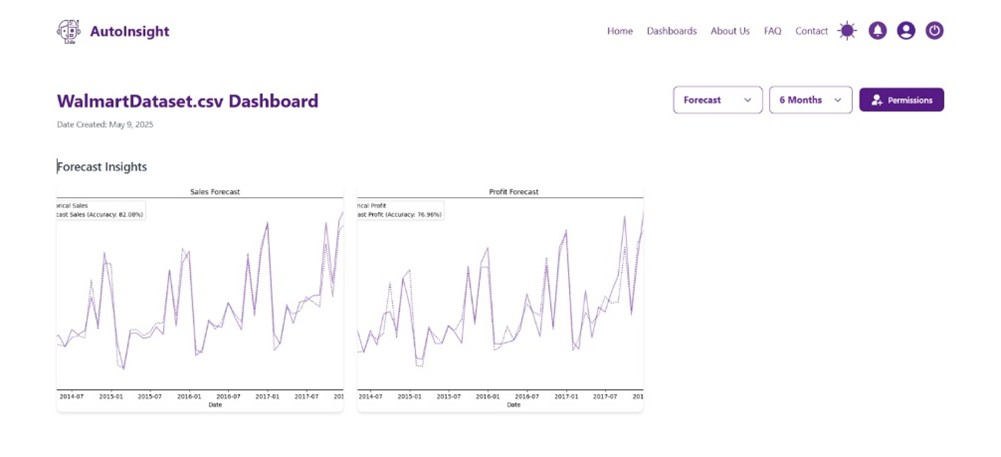
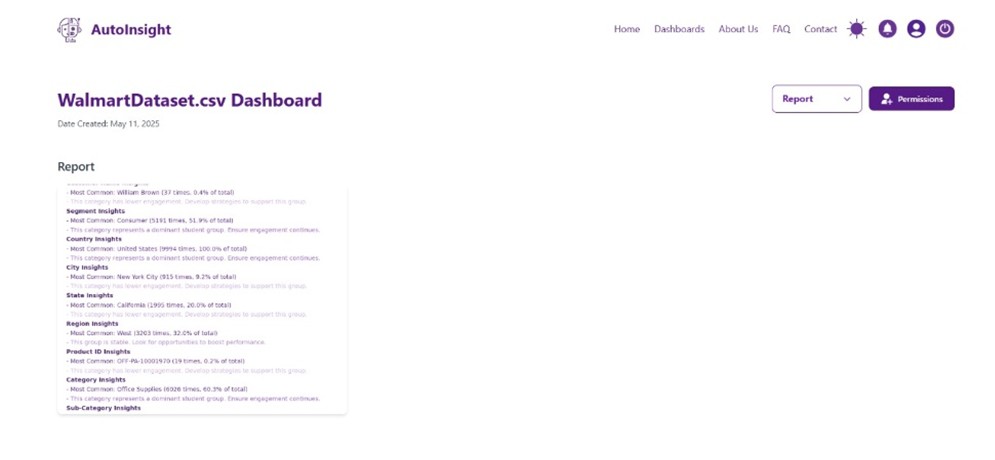
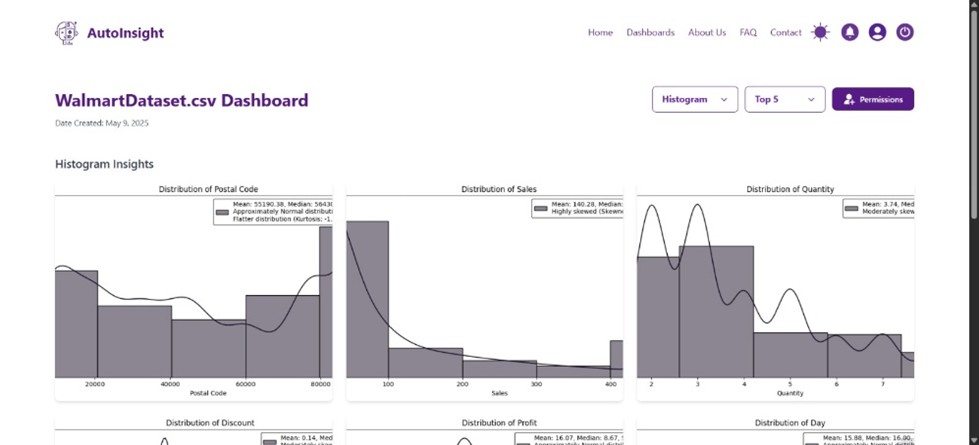
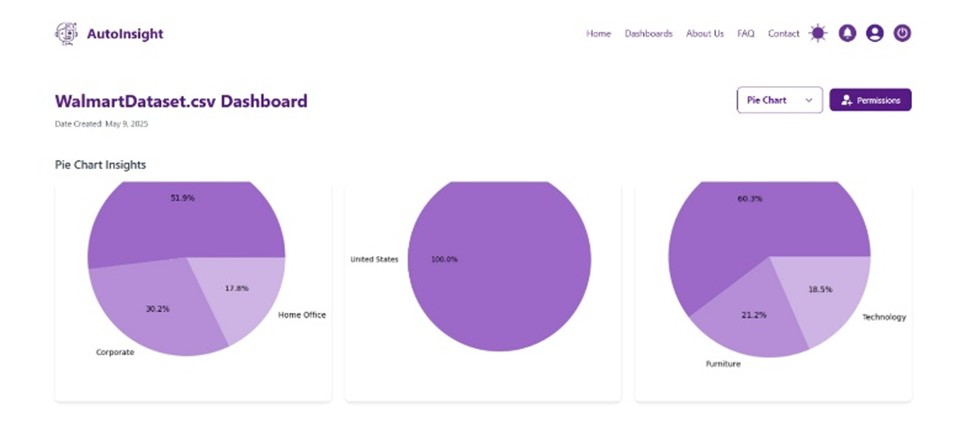
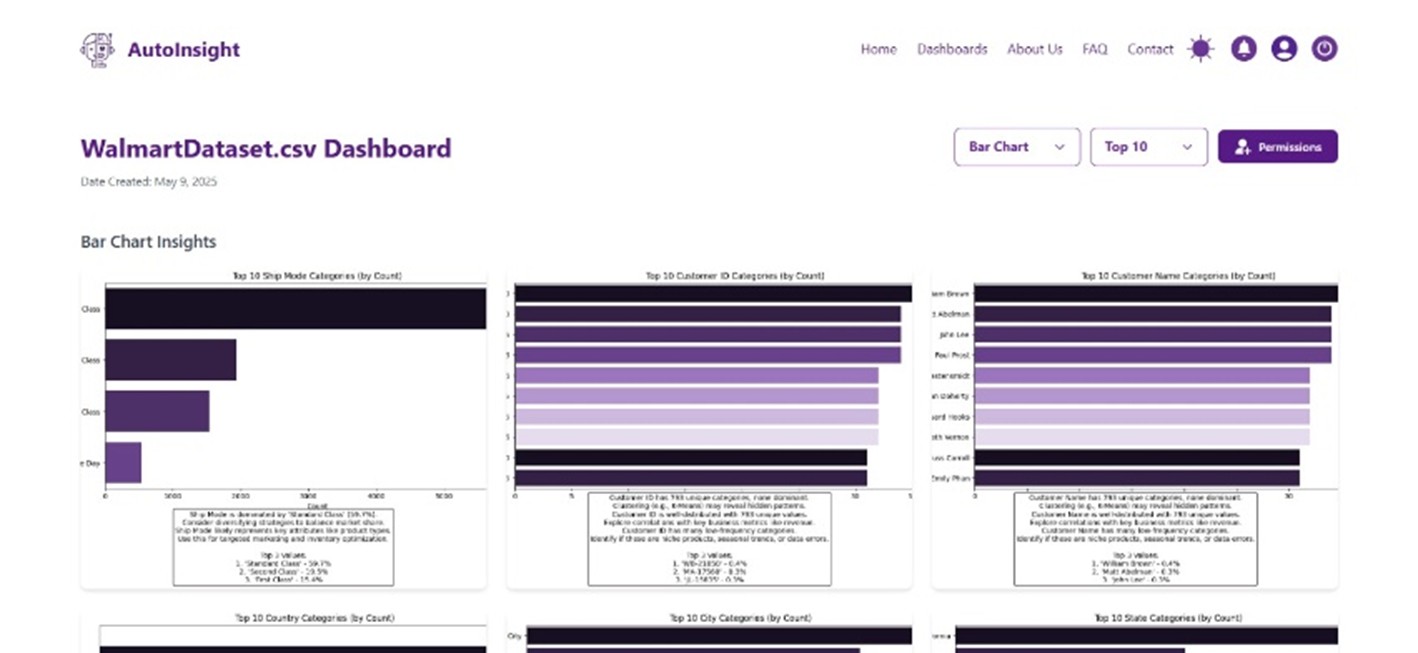


**Screenshot 4,5,6,7&8:Uploading steps**

**5.4 Dashboard Generation**

If the user chooses to generate a dashboard:

* For eCommerce: Displays sales trends, revenue forecasts, product performance.
* For Education: Displays student performance trends, classification of outcomes, and attendance forecasts.

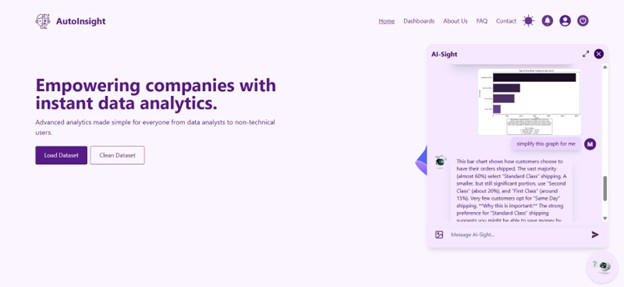


**Screenshot 9,10,11,12&13: Dashboard generation**

**5.5 Chatbot Assistance**

The AI-powered chatbot (via Gemini API) assists users in:

* Interpreting dashboards
* Explaining metrics
* Answering general data-related questions

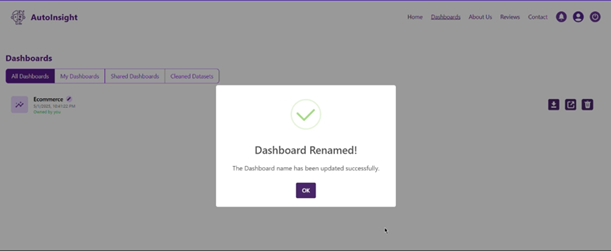
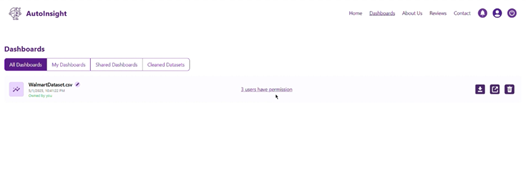


**Screenshot 14: *Chatbot interface responding to user queries.***

**5.6 Dashboard and Dataset Management**

Users can:

* Rename or delete datasets and dashboards.
* Share dashboards using role-based access controls.

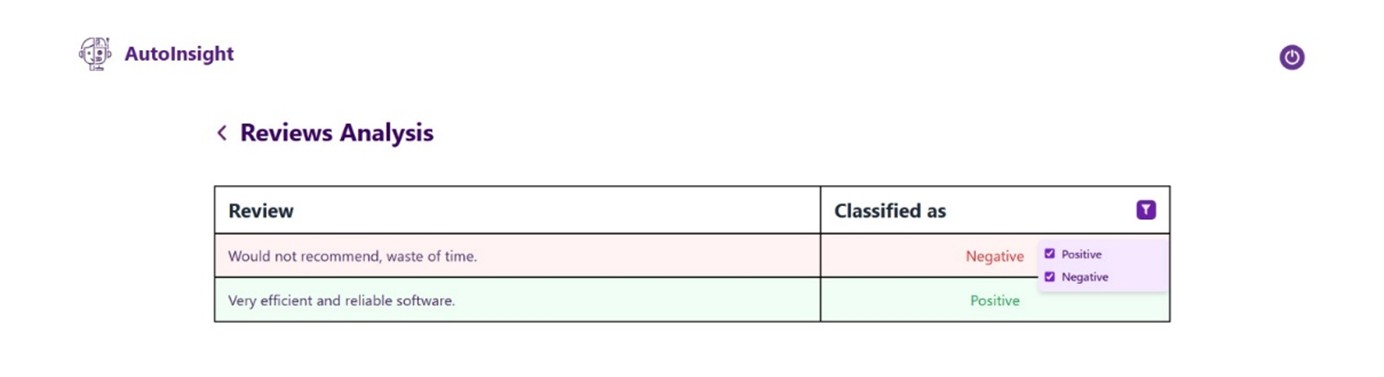
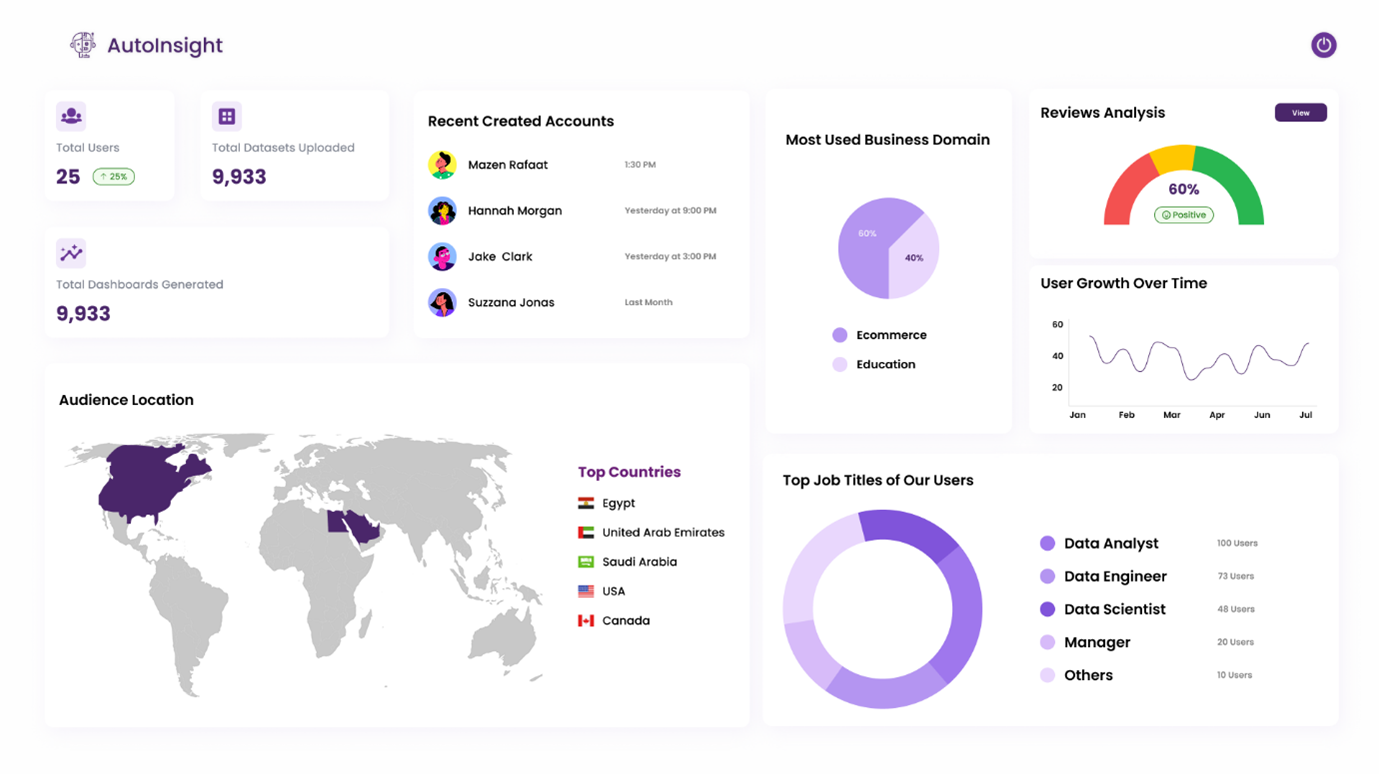


**Screenshot 15&16: *Dashboard management panel.***

**5.7 Admin Dashboard**

Admins have access to a powerful monitoring panel that includes:

* Total users and active sessions
* Number of dashboards generated by domain (eCommerce, Education)
* Review sentiment analysis statistics
* Recent user reviews with sentiment scores
* Most used features and peak activity times

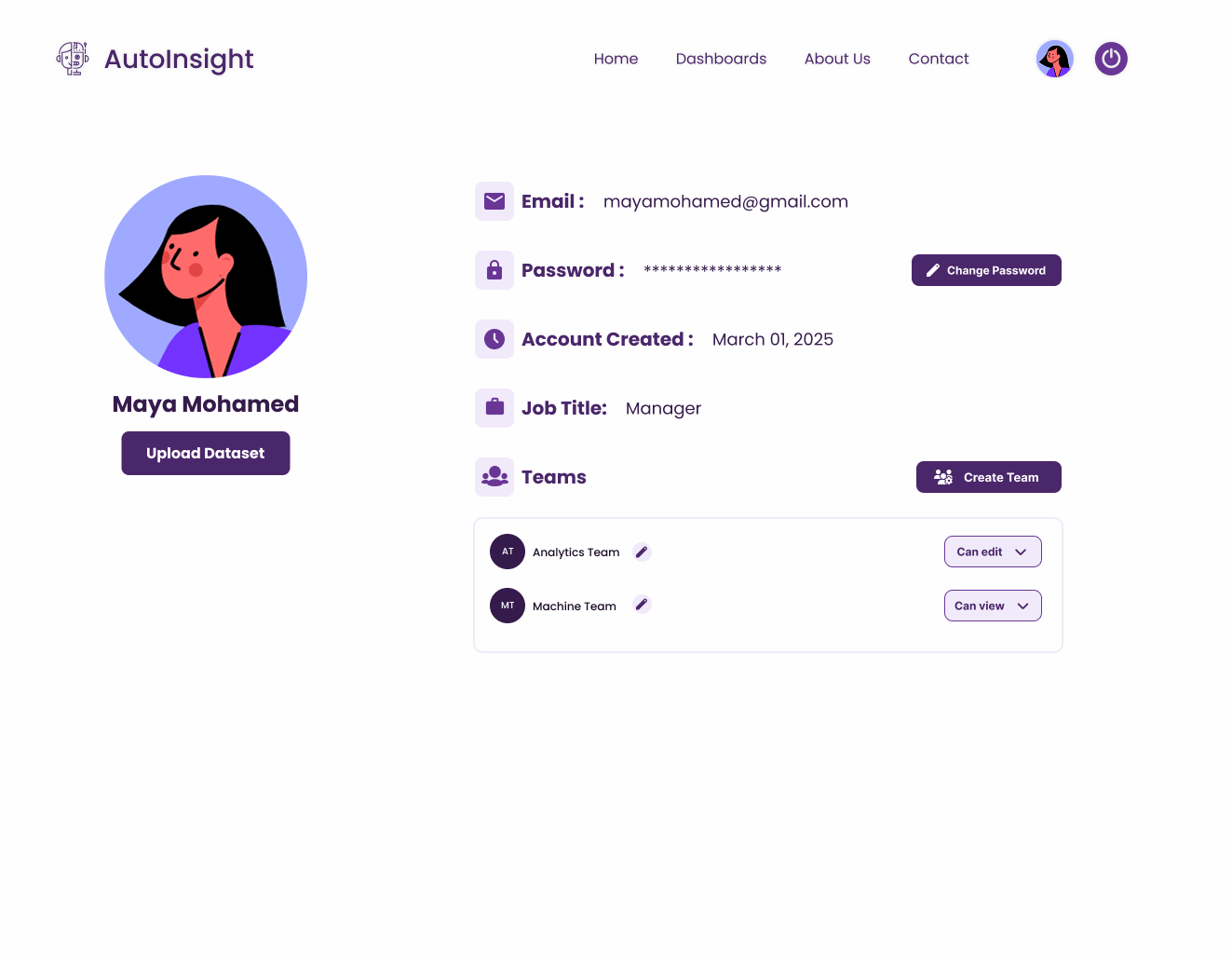


**Screenshot 17&18: Admin dashboard overview.**

**5.8 User Profile Page**

Users can manage their accounts via the profile page:

* View number of uploaded datasets and dashboards.
* Change personal information and password.
* View activity summary.

****

**Screenshot 19: User profile page**

**Chapter 6: Conclusion and Future Work**

**6.1 Conclusion**

In conclusion, this project successfully developed AutoInsight, an automated data analytics platform aimed at addressing the challenges faced by businesses, particularly startups and educational institutions, in the absence of dedicated data analysts.

The platform integrates AI-driven analytics, machine learning, and natural language processing (NLP) to enable seamless data processing, forecasting, and decision-making.

Through various methodologies, such as time series forecasting, diagnostic analysis, and automated report generation, AutoInsight provides businesses with actionable insights without requiring extensive human intervention.

While the system delivers valuable insights, certain limitations exist.

The reliance on pre-processed datasets may restrict the platform’s applicability in real-time scenarios where immediate data analysis is required.

Additionally, the system's performance could be impacted by the scale of the dataset or the complexity of the analysis, particularly in larger organizations with vast data volumes.

Despite these challenges, the project demonstrates the significant potential for automated analytics platforms to bridge the gap in resource-constrained environments.

**6.2 Future Work** Several improvements and extensions are recommended for the future development of AutoInsight to further enhance its capabilities:

1. **Real-Time Data Processing:** The ability to process and analyze data in real time would be crucial for businesses in dynamic environments. Integrating real-time data pipelines and streaming analytics could significantly improve decision-making processes.
2. **Advanced Machine Learning Models:**  
    Incorporating more advanced machine learning techniques, such as deep learning models for more complex predictive analysis, could increase the platform’s ability to handle large-scale data and more intricate datasets.
3. **Enhanced Natural Language Processing (NLP) Capabilities:**  
    Improving the chatbot functionality with more advanced NLP models can facilitate better interaction with users, providing them with more comprehensive and precise insights. This could also include voice-based queries for hands-free operation.
4. **Scalability for Large Enterprises:**  
    Future development should focus on scaling AutoInsight for use by larger enterprises with massive datasets. Optimizing the platform for high-volume data processing and integrating distributed computing models would enhance performance.
5. **Expanding Business Domain Support:**  
    The current version of AutoInsight supports specific domains like e-commerce and education. Expanding the platform to include additional sectors (e.g., healthcare, manufacturing, or finance) would increase its versatility and reach.
6. **Integration with External Data Sources:**  
    The system can benefit from integrating with external data sources, such as cloud-based data warehouses and third-party APIs, for more comprehensive data analysis and forecasting.

By addressing these areas, AutoInsight can be further enhanced to meet the evolving demands of businesses in diverse industries, thereby ensuring its relevance and effectiveness in the future.

REFERENCES

[1] H. Upadhyay, S. Shekhar, A. Vidyarthi, R. Prakash and R. Gowri, "Sales Prediction in the Retail Industry Using Machine Learning: A Case Study of BigMart,"  2023 pp. 1-6, doi: 10.1109/ELEXCOM58812.2023.10370313.7.

[2] S. Pang, "Retail Sales Forecast Based on Machine Learning Methods," 2022 pp. 357-361, doi: 10.1109/ICDSBA57203.2022.00030.

[3] A. Krishna, A. V, A. Aich and C. Hegde, "Sales-forecasting of Retail Stores using Machine Learning Techniques," 2018 pp. 160-166, doi: 10.1109/CSITSS.2018.8768765.

[4] M. Bilal, G. Ali, M. W. Iqbal, M. Anwar, M. S. A. Malik and R. A. Kadir, "Auto-Prep: Efficient and Automated Data Preprocessing Pipeline," in IEEE Access, vol. 10, pp. 107764-107784, 2022, doi: 10.1109/ACCESS.2022.3198662.

[5] S. J. Taylor and B. Letham, "Forecasting at Scale," *PeerJ Preprints*, vol. 5, p. e3190v2, 2017, doi: [10.7287/peerj.preprints.3190v2](https://doi.org/10.7287/peerj.preprints.3190v2).

[6] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM)*, 2014, pp. 216-225.

[7] A. Krishna, A. V, A. Aich, and C. Hegde, "Sales-Forecasting of Retail Stores using Machine Learning Techniques," *2018 International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 2018, pp. 160-166, doi: [10.1109/CSITSS.2018.8768765](https://doi.org/10.1109/CSITSS.2018.8768765).

[8] D. Hardt, Ed., "The OAuth 2.0 Authorization Framework," *RFC 6749*, Internet Engineering Task Force (IETF), 2012. [Online]. Available: <https://tools.ietf.org/html/rfc6749>.

[9] A. Menezes, P. van Oorschot, and S. Vanstone, *Handbook of Applied Cryptography*, 1st ed. CRC Press, 1996.

[10] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11]<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>

[12]<https://www.keboola.com/blog/5-stats-that-show-how-data-driven-organizations-outperform-their-competition>

[13] <https://www.techmonitor.ai/digital-economy/big-tech/tech-layoffs-2023>