



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

Voice-Enabled Intelligent E-Commerce Data Analysis System

PL Subramanian, C Arjun ,KK Harikes

Assistant Professor of the Dept. of Information Technology, K L N College of Engineering, Madurai,
Tamil Nadu, India

B.Tech Student, Dept. of Information Technology, K L N College of Engineering, Madurai, Tamil Nadu, India

B.Tech Student, Dept. of Information Technology, K L N College of Engineering, Madurai, Tamil Nadu, India

ABSTRACT: The rapid growth of e-commerce platforms has led to the generation of large volumes of sales, product, and customer data, commonly stored in formats such as CSV files, PDF reports, invoices, and images. Extracting meaningful insights from these documents manually is time-consuming, inefficient, and often requires technical expertise. This creates challenges, especially for small and medium-scale business owners, in quickly interpreting data for effective decision-making. Therefore, there is a need for an intelligent system that simplifies data analysis and provides clear, actionable insights in an accessible and user-friendly manner.

This project introduces the AI Sales Agent, a voice-enabled intelligent system designed to analyze e-commerce data and deliver accurate, conversational responses to user queries. The system supports multiple data formats and integrates technologies such as natural language processing, machine learning, semantic search, and large language models to provide context-aware insights. By enabling interaction through both voice and text, it enhances usability and reduces manual effort. The solution improves business efficiency and decision-making while contributing to sustainable development goals by promoting innovation, productivity, and efficient resource utilization.

I. INTRODUCTION

The rapid expansion of e-commerce platforms has significantly transformed business operations, leading to the generation of vast amounts of transactional, product, and customer data on a daily basis. This data is commonly stored in formats such as CSV files, PDF reports, invoices, and even image-based documents. Although this information contains valuable insights related to sales trends, product performance, and customer behavior, extracting meaningful information manually is time-consuming and complex.

With advancements in Artificial Intelligence, Natural Language Processing, and Machine Learning, it is now possible to automate data analysis and make it more efficient and user-friendly. The proposed AI Sales Agent is a voice-enabled intelligent system that analyzes e-commerce data and provides instant, conversational responses. By integrating data processing, semantic search, and speech technologies, the system helps business users gain insights quickly and make better decisions with minimal effort.

II. LITERATURE SURVEY

[1] Business Intelligence Tools for Data Analytics (2019)

Researchers studied tools like Tableau and Power BI for analyzing business data and generating dashboards. These tools helped organizations visualize sales trends, customer behavior, and performance metrics effectively. Limitation: Requires technical knowledge and manual effort to interpret data, making it less accessible for non-technical users.

[2] Natural Language Processing for Data Query Systems (2020)

This research focused on using NLP techniques to allow users to query databases using natural language instead of structured queries. It improved accessibility and user interaction with data systems. Limitation: Limited understanding of complex queries and lacks integration with voice-based systems.

[3] Conversational AI for Business Analytics (2021)

Researchers developed AI-powered chatbots to provide business insights through conversational interfaces. These systems used machine learning models to interpret queries and generate responses. Limitation: Accuracy depends on training data and lacks deep integration with real-time data analytics.

[4] Voice-Enabled Intelligent Systems Using Speech Recognition (2022)

This study explored the use of speech recognition and text-to-speech technologies to build voice-enabled systems for user interaction. It improved accessibility and hands-free operation. Limitation: Not fully integrated with data analytics systems and may face challenges with speech accuracy in noisy environments.

III. METHODOLOGY**3.1 Requirement Analysis**

System analysis is the process of studying and understanding the project objectives, user requirements, and technical constraints to design an efficient and practical solution.

The proposed system — A Voice-Enabled Intelligent System for E-Commerce Data Analysis — aims to assist business owners, analysts, and decision-makers by providing an automated platform to analyze sales data, product performance, and revenue trends. The system uses Artificial Intelligence, Natural Language Processing, and voice technologies to enable users to interact with data through text or speech and obtain meaningful insights easily.

3.2 System Design

System design defines the architecture, components, data flow, and interactions between different modules of the system. The design aims to transform functional requirements into a structured framework that guides system development.

For this project, the system integrates data preprocessing, AI-based query processing, and a Streamlit-based user interface. It allows users to upload datasets, interact using natural language or voice, and receive insights through visualizations and responses. The design ensures smooth interaction between input, processing, analytics, and output modules.

3.3 Development Tools and Technologies

The system's web-based interface is implemented using Streamlit, which enables seamless interaction between users and the backend processing modules. Pandas and NumPy are used for data preprocessing, cleaning, and handling structured datasets efficiently. Artificial Intelligence functionalities are powered by LLM-based models and Natural Language Processing techniques, enabling the system to understand user queries and generate relevant responses. Plotly is used for creating interactive visualizations, while speech recognition and text-to-speech libraries are used to support voice-based interaction. Additionally, ChromaDB and TF-IDF techniques are used for semantic search and efficient data retrieval.

3.4 Implementation

The implementation phase involves translating the system design into a functional application by integrating all modules. It includes dataset handling, preprocessing, AI query processing, and user interface development.

The system is implemented using a modular approach where users can upload datasets, input queries via text or voice, and receive results in real time. The AI module processes queries, extracts relevant information, and generates insights, while the visualization module presents the results in graphical formats. The overall system ensures smooth workflow from data input to output generation.

3.5 Testing and Deployment

Software testing is an essential stage in the system development life cycle (SDLC) to ensure that the application functions correctly and produces reliable outputs. In this project, testing is performed to validate data processing accuracy, AI response quality, and system performance.

Testing focuses on validating user inputs, ensuring correct handling of datasets, and verifying the accuracy of generated insights and visualizations. The Streamlit interface is tested for usability and responsiveness to ensure a smooth user experience. The deployment phase involves making the system accessible to users through a web-based platform. The application can be deployed locally or on cloud platforms, allowing business users to upload data, interact with the system, and obtain insights efficiently. The deployment ensures scalability, reliability, and ease of access for real-world usage.

3.6 SYSTEM ARCHITECTURE

The system architecture defines the overall structure and workflow of the Agentic E-Commerce Analytics System, illustrating how different components interact to process user queries and generate intelligent insights. It ensures a smooth flow of data from input to output while maintaining modularity, scalability, and efficiency. The architecture begins with the user interaction layer, where users provide input in the form of text, voice, or uploaded files such as CSV, PDF, or images. This input is then processed through specialized modules including OCR for image text extraction, PDF parsing for document analysis, and data preprocessing tools that clean and structure the data into meaningful formats. The chunking module converts processed data into smaller, searchable text segments, which are stored in a vector database for efficient retrieval.

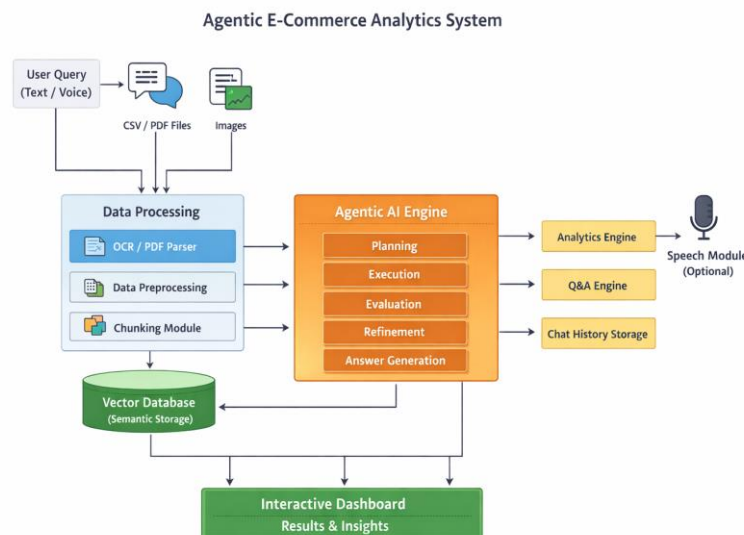


Figure -1: Architecture of the System

The core of the system is the Agentic Engine, which follows a multi-phase workflow consisting of planning, execution, evaluation, refinement, and answer generation. It intelligently analyzes user queries, retrieves relevant data using semantic search, evaluates the quality of retrieved information, and refines the search when necessary to improve accuracy. Supporting components such as the analytics engine compute key metrics like revenue trends and product performance, while the QA engine generates conversational responses. The system also incorporates chat history storage for contextual continuity and optional speech modules for voice interaction. Finally, the processed insights are delivered through an interactive interface, enabling users to easily understand business performance and make data-driven decisions.

IV. EXISTING SYSTEM

The existing e-commerce data analysis process is largely manual and relies on traditional tools such as spreadsheets and static dashboards. Business owners and analysts typically depend on predefined reports and basic data visualization techniques to interpret sales data.

However, these methods are time-consuming, require technical expertise, and often fail to provide real-time insights. They also lack the ability to process unstructured data such as PDFs and images and do not support natural language interaction. As a result, decision-making becomes inefficient, delayed, and less effective in identifying key business trends and opportunities.

V. PROPOSED SYSTEM

The proposed system introduces an AI-powered E-Commerce Analytics Agent that leverages machine learning and natural language processing techniques to provide intelligent insights from sales data. Unlike traditional analysis methods, this system supports multiple data formats, including CSV files, PDFs, and images, enabling comprehensive data processing.

The core of the system is an agentic AI engine that follows a structured workflow consisting of planning, execution, evaluation, refinement, and answer generation. It utilizes semantic search and vector storage to retrieve relevant information and generate context-aware responses using advanced language models. Users can interact with the system through a Streamlit-based web interface by entering text queries or using voice commands. The system provides real-time analytics such as revenue trends, product performance, and sales summaries, making it an efficient and user-friendly solution for data-driven decision-making.

VI. RESULT

The implemented AI Sales Agent successfully analyzes e-commerce data and provides accurate, meaningful insights based on user queries. The system effectively processes structured and unstructured data and delivers relevant responses using AI-driven techniques. It demonstrates reliable performance in identifying key metrics such as total revenue, average order value, and top-performing products.

The Streamlit-based interface enables real-time interaction, allowing users to upload data and receive instant insights through text or voice queries. Sample use cases show that the system can quickly highlight high-performing products, detect revenue trends, and assist in business decision-making. Overall, the results indicate that the system significantly improves the efficiency of data analysis, reduces manual effort, and enhances the ability of businesses to make informed, data-driven decisions.

Voice-Enabled Intelligent E-Commerce Data Analysis System

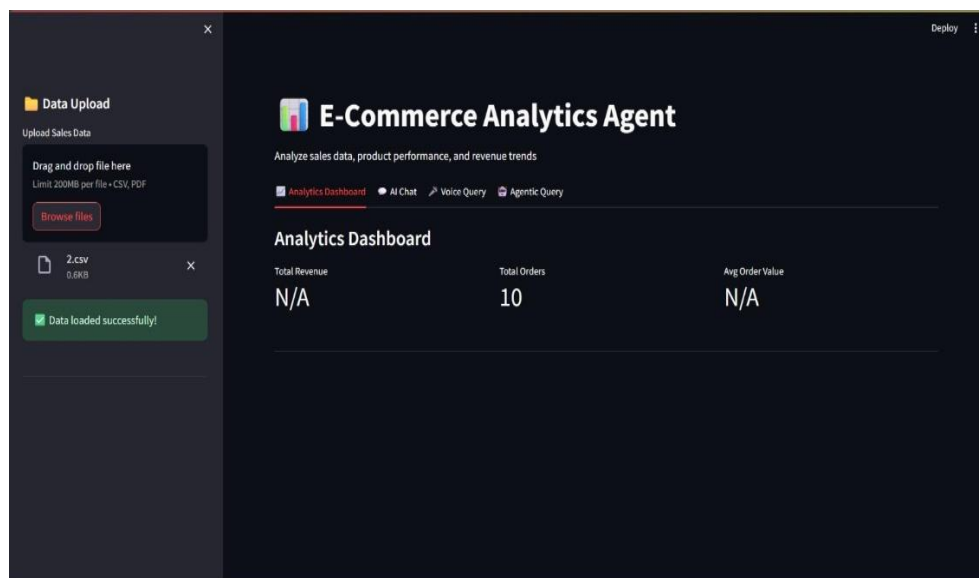


Figure -1: inputs

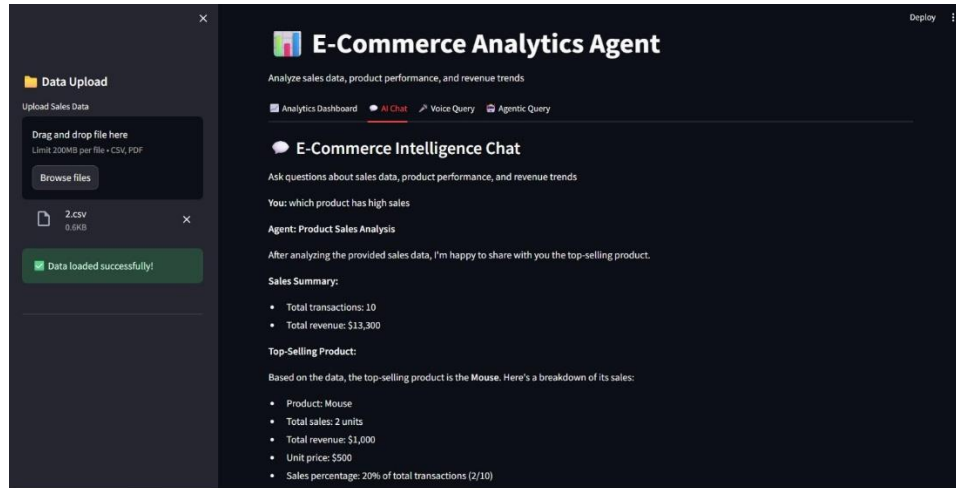


Figure -2: outputs

VII. FUTURE ENHANCEMENT

The current system provides intelligent insights into e-commerce data through automated analysis and AI-driven responses, but several enhancements can further improve its capabilities and usability. Future improvements include integrating real-time data from e-commerce platforms such as Shopify and Amazon APIs to enable live analytics and continuous monitoring of business performance. Incorporating advanced visualization techniques and interactive dashboards can help users better understand trends and patterns in sales and customer behavior. The system can also be enhanced using advanced machine learning and deep learning models for predictive analytics, such as sales forecasting and customer segmentation.

Additionally, expanding support for multiple languages in both text and voice interaction will make the system more accessible to a wider audience. Deploying the application on cloud platforms and developing a mobile version can improve scalability and accessibility for users across different devices. Enhancing security features such as authentication and data encryption will ensure safe handling of sensitive business data. These improvements aim to transform the system into a comprehensive AI-powered business intelligence assistant for modern e-commerce applications.

VIII. CONCLUSION

The project aimed to address the growing challenge of analyzing large volumes of e-commerce data by introducing an intelligent and automated solution. Traditional data analysis methods are often manual, time-consuming, and require technical expertise, making it difficult for businesses to extract meaningful insights efficiently. In contrast, the proposed system provides a smart, user-friendly platform that enables real-time data analysis and interactive querying using natural language and voice inputs.

Through the integration of technologies such as data processing, semantic search, and AI-driven response generation, the system effectively delivers accurate insights into sales performance, product trends, and revenue patterns. The Streamlit-based interface enhances usability, allowing even non-technical users such as business owners and analysts to interact with the system effortlessly and make informed, data-driven decisions.

REFERENCES

- [1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson Education, 2022.
- [2] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 4th ed. Elsevier, 2021.

- [3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” arXiv preprint arXiv:1301.3781, 2013.
- [5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” NAACL-HLT, pp. 4171–4186, 2019.
- [6] A. Radford et al., “Language Models are Few-Shot Learners,” OpenAI, 2020.
- [7] G. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” EMNLP, 2019.
- [8] Streamlit Inc., “Streamlit: The fastest way to build data apps,” Available: <https://streamlit.io>, 2023.
- [9] ChromaDB, “Chroma: The AI-native open-source embedding database,” Available: <https://www.trychroma.com>, 2023.
- [10] Groq Inc., “Groq API for Fast LLM Inference,” Available: <https://console.groq.com>, 2024



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details