

AI-Powered Business Analyst Using Natural Language Processing and Retrieval-Augmented Generation

Dr. M. Kayalvizhi

Assistant Professor,

Department of Information Technology,
Sri Krishna Adithya College of Arts and
Science.

kayalvizhim@skacas.ac.in

SanjeevKailash.D.G

Junior Researcher,

Department of Information Technology,
Sri Krishna Adithya College of Arts and
Science.

23bsit254sanjeevkailashdg@skacas.ac.in

ABSTRACT

In the digital business ecosystem of today, organizations generate huge amounts of data from structured and unstructured sources through sales transactions, customer interactions, marketing campaigns, and operational processes. However, due to a lack of analytical skillsets, time constraints, and fragmented data sources, SMEs often fail to extract actionable insights from the said data. The proposed system leverages Large Language Models, LangChain orchestration, SQL-based databases, and Retrieval-Augmented Generation to provide business insights in real time without requiring technical knowledge: Natural language queries are translated into structured SQL queries, or document searches, the proposed solution improves decision.

The three major improvements are: speed, accuracy, and accessibility. Experimental use cases in sales analysis, marketing optimization, customer segmentation, and financial forecasting demonstrate the system's effectiveness in facilitating data-driven decisions while ensuring data privacy through local model deployment.

Keywords - *Business Analytics, Artificial Intelligence, Natural Language Processing, Retrieval-Augmented Generation, Large Language Models, Decision Support Systems.*

I. INTRODUCTION

The rapid growth of digital technologies is generating massive volumes of business data every day. Business analytics is essential for converting these data elements into meaningful insights that help organizations make strategic decisions. Traditional tools for analytics involve technical skills, report generation through manual effort, and predefined dashboards, which further limit flexibility and usability for non-technical users. As a result, many organizations, especially SMEs, tend to rely on intuition-based decisions rather than fact-based insights.

Moreover, many analytics systems presently lack real-time insight, which leads to delays in business responses to changes in markets. The lack of integration of structured databases with

unstructured documents further reduces the effectiveness of conventional systems. These limitations pinpoint the need for intelligent, adaptive, and user-friendly analytics platforms.

Artificial Intelligence and NLP have become the most transformative technologies in business analytics. Recent breakthroughs with Large Language Models enable machines to understand human language and generate conversational responses, allowing more intuitive interfaces for analytics through conversations. AI-powered chatbots reduce the learning curve associated with analytics tools and empower decision-makers at all organizational levels.

The paper presents an AI-Powered Business Analyst system that works much like a virtual business analyst. It will allow users to ask questions

like "What was the most profitable product?" or "How did the sales trend in the last month?" and get responses instantaneously with high accuracy. By fusing SQL databases with document-based knowledge using RAG, this system provides all-encompassing insights while maintaining data privacy and scalability.

II. RELATED WORK

Various studies have discussed how the usage of AI and machine learning in business analytics enhances decision-making. Traditional Business Intelligence tools greatly rely on static dashboards and predefined reports; however, there is a limitation in terms of flexibility and user interaction. These systems often require skilled analysts to interpret data, thus driving up operational costs for SMEs.

Recent research underlines the role of machine learning in predictive and prescriptive analytics to enable the identification of trends and suggestions on the best course of action. Recently, conversational analytics systems that use NLP have gained significant attention because they can simplify data access through natural language queries. Many such solutions remain cloud-dependent, bringing up issues related to data privacy, latency, and its associated costs.

Retrieval-Augmented Generation has been considered an effective method to combine structured databases with unstructured textual data. RAG-based systems eliminate hallucination issues typical in stand-alone LLMs by enhancing contextual understanding and response accuracy. Despite this, very little research is focused on privacy-preserving, locally deployable AI analytics solutions for SMEs, which this paper aims to address.

III. SYSTEM ARCHITECTURE & DESIGN

The AI Business Analyst will process natural language queries to retrieve insights efficiently via a modular, scalable architecture. Its architecture is designed to provide flexibility, extensibility, and seamless integrations of multiple data sources.

A. System Overview

The system comprises six major components:

1. UpdateUserInterface
2. Natural Language Processing Module
3. Query Interpretation Engine

4. Structured Data Source (MySQL)

5. Unstructured Data Source: ChromaDB

6. Response Generation Module

Each of the modules can function independently, which allows easier maintenance and enhancements in the future. Separation of concerns also enhances system reliability and scaling.

B. Query Processing Flow

1. User submits query in natural language.
2. The query is processed through the LangChain and forwarded to LLaMA.
3. The query is categorized as structured-which are SQL-based, or unstructured-which are document-based.
4. SQL queries run on MySQL databases, while document searches are done through ChromaDB.
5. Results retrieved are validated, summarized, and formatted.
6. The system provides the user with clear and actionable insights.

This pipeline guarantees minimum response time with high accuracy, keeping the response relevant to context.

C. Technology Stack

Programming Language: Python

LLaMA (via Ollama for local deployment)

Framework: LangChain

Database: MySQL

Vector Database: ChromaDB

Performance, privacy, and open-source flexibility are prioritized in the chosen stack.

IV. IMPLEMENTATION METHODOLOGY

The system implementation focuses on usability, accuracy, and data privacy. LLMs tokenize and semantically analyze the natural language queries. Intent detection and entity extraction can help translate user queries into valid actions with high accuracy.

LangChain serves as the orchestration layer, handling interactions between the language model, SQL database, and vector store. Query validation mechanisms are implemented to avoid incorrect or harmful database operations.

The Retrieval-Augmented Generation approach increases the response accuracy by anchoring model outputs in real data. Deploying the model locally with Ollama reduces reliance on the cloud services, minimizes latency, and makes sure

sensitive business data stays within the organization's infrastructure.

V. USE CASES & APPLICATIONS

The Business Analyst, powered with AI, supports a wide range of real-world business scenarios across departments to drive data-informed decisions.

A. Sales Performance Analysis

The system identifies top-performing products, regions, and sales periods. This helps the business discover seasonal trends in products, track sales growth, and optimize pricing strategies accordingly.

B. Marketing Optimization

To measure engagement, conversion rates, and return on investment, data from the marketing campaign is analyzed. The chatbot allows for quick comparisons across multiple channels on the performance of campaigns.

C. Customer Segmentation

Customer information is processed to segment users by demographics, purchase history, and behavior. It allows making personalized marketing efforts and enhancing customer retention.

D. Financial Forecasting

Revenue, expenses, and profit conditions are forecasted using historical financial data analysis. Consequently, these insights will help management in budgeting, assessing risks, and conducting strategic planning activities.

VI. ADVANTAGES AND EVALUATION

The proposed system has several advantages compared to other analytics platforms:

- * Less dependence on skilled data analysts Real-time insights through conversational interaction
- * Seamless integration of structured and unstructured data
- * Improved decision accuracy and speed Improved data privacy by deploying locally

Initial testing with sample business datasets reveals a considerable reduction in query response time, with increased relevance of insights provided. In general, feedback from users suggests greater satisfaction because of both ease of use and clarity in responses.

VII. ADDITIONAL FEATURES OF THE PROPOSED SYSTEM

There are a number of Role-Based Access Control (RBAC)

It can be interfaced with additional functionalities of role-based access, allowing only selected users, such as managers or administrators, to have access to strategic data like financial performance.

• Explainable AI (XAI)

For purposes of increasing trust and transparency, the system should have the capability to explain why it has arrived at certain insights by pointing out the data that it used.

• Automated Report Generation

The AI Business Analyst is also capable of auto-generating business reports either weekly or monthly in PDF format or in the form of dashboards.

• Anomaly Detection

The system is able to pick up any peculiar patterns, for instance, sudden reductions in sales, peculiar expenses, and unexpected customer churn.

VIII. EVALUATION METRICS FOR PERFORMANCE

Some metrics that could be used to assess system performance are:

- Response Time for the Query – This is the time taken by the system for the generation of insights based on the user query.
- Accuracy of SQL translation – This assesses the ability of the conversion from natural language to SQL.
- Relevance Score - Helps determine if the answer is relevant to the user's intent
- User Satisfaction Score - Obtained from feedback surveys.
- System Scalability: The capacity to process growing amounts of data and simultaneous users.

IX. SECURITY AND PRIVACY ISSUES

- "The Local Model Deployment is designed to ensure that business-sensitive information is maintained within a company's infrastructure."
- Prevention of SQL Injection Attacks by validating and sanitizing queries.
- Secured database connections for the transfer of data.
- Audit Logs – compliance tracking via analysis of user queries and system responses.

These features make it an ideal system for privacy-mainly SMEs.

X. SHORTCOMINGS OF THE SYSTEM

The discussed Despite its merits, the proposed system has some limitations:

- Depends upon the quality and organization of the underlying data.
- Analytical queries may need the skills of a good prompt mastery.
- Setup involves domain-specific schema definition.
- Forecasting accuracy may be affected for highly volatile market data These limitations point out areas where there can be growth. XII. Future Enhancements (Extended) Apart from the aforementioned work plans in the future, the system can be improved using:
 - Integration with real-time data streams (IoT, sales feeds) Speech-based answer to the query using the speech to text algorithm.
 - Auto-learning optimization of the query using reinforcement learning.
 - Compatibility with BI visualization software like Power BI & Tableau.
 - Flexible across multiple industry sectors like Healthcare, Retail & Logistics.

CONCLUSION AND FUTURE WORK

This paper presented the AI-Powered Business Analyst that simplifies business analytics through natural language interactions and Retrieval-Augmented Generation. By integrating LLMs, SQL databases, and document retrieval systems, the solution empowers an organization to make informed decisions efficiently and securely.

It shows how this design could help bridge the gap between complex data infrastructures and non-technical users by using conversational AI. In particular, SMEs will benefit from its modular design and privacy-preserving deployment.

Future work will involve the integration of machine learning predictive models, the enhancement of data visualization dashboards, support for multilingual queries, and large-scale performance evaluations to further validate system effectiveness.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Dr.M.Kayalvizhi, Assistant Professor, Department of Information Technology, Sri Krishna Adithya College of Arts and Science, Coimbatore, India, for her valuable guidance, continuous encouragement, and constructive suggestions throughout the course of this research work. The authors also acknowledge the dedicated efforts and contribution of Mr.SanjeevKailash.D.G, whose consistent involvement and commitment were instrumental in the successful preparation and completion of this study.

REFERENCES

- [1] Y. K. Dwivedi et al., "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities and agenda for research, practice and policy," *International Journal of Information Management*, vol. 57, pp. 1-22, 2021.
- [2] I. H. Sarker, "Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective," *SN Computer Science*, vol. 2, no. 5, pp. 1–22, 2021.
- [3] S. Bharadiya, "Machine learning and AI in business intelligence: Trends and opportunities", *International Journal of Computer (IJC)*, Vol. 48, No. 1, pp. 123-134, 2023.
- [4] C. Janiesch, P. Zschech and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021.
- [5] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, 2020, pp. 1-12.
- [6] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, "Prescriptive analytics: Literature review and research challenges," *International Journal of Information Management*, vol. 50, pp. 57-70, 2020.
- [7] J. Wang, J. Tang, and H. Liu, "Natural language processing for intelligent business analytics," *IEEE Access*, vol. 9, pp. 125345–125358, 2021.
- [8] C. Molnar, *Interpretable Machine Learning*, 2nd ed. Cham, Switzerland: Springer, 2022.
- [9] J. Dean and L. A. Barroso, "The tail at scale," *Communications of the ACM*, vol. 56, no. 2, pp. 74–80, 2019.

[10] T. Mikolov et al., "Efficient estimation of word representations in vector space," in *Proc. International Conference on Learning Representations (ICLR)*, 2019.

[11] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for Industry 4.0-based manufacturing systems," *Manufacturing Letters*, vol. 3, pp. 18-23, 2020. [12] IEEE Editorial Board, *IEEE Author Center: Preparing Your Journal Article for Submission*. IEEE, 2024.