

DataLyze - An Agentic AI for Autonomous Data Analysis

Vinutha D¹, Arun Ramesh Chavan², Deeksha G Nadig³, Pavan Raju Konkati⁴, Veena M R⁵

¹Assistant Professor, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, affiliated to VTU Belagavi, Karnataka, India.

²Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India

³Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India

⁴Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India

⁵Bachelor of Engineering, Information Science and Engineering, Bapuji Institute of Engineering and Technology, Karnataka, India

Abstract - Massive amounts of data are produced in modern businesses from a variety of sources, including real-time apps, CSV files, SQL databases, NoSQL systems, and XML streams. Conventional analytical tools mainly rely on laborious procedures, technical know-how, and manual preparation. DataLyze presents an Agentic AI-based system that can ingest, preprocess, analyze, and visualize data on its own without human assistance in order to get around these restrictions. Developed with FastAPI, LangChain, React.js, and MongoDB, DataLyze uses an interactive dashboard and AI-powered chatbot to carry out automatic statistical analysis, insight generation, and real-time visualization. This work's major objective is to democratize data analytics by doing away with the requirement for domain expertise or coding. The findings show that DataLyze greatly speeds up processing, increases analytical precision, and improves accessibility for non-technical users.

Key Words: LangChain, FastAPI, Chatbot, Agentic AI, Autonomous Data Analysis, Data Visualization, Multi-format Ingestion, and Real-Time Dashboard.

1. INTRODUCTION

Decision-making in a variety of sectors, including retail, healthcare, banking, and transportation, now revolves around data. However, there are a number of difficulties in turning unprocessed data into insightful knowledge, including manual preprocessing, comprehension of schemas, choice of visualization, and statistical interpretation. For organized data, conventional BI programs like Tableau and Power BI are effective, however they lack: independent judgment, ingestion of multiple formats, natural language communication, flexibility in dealing with unstructured datasets. The suggested approach DataLyze employs Agentic Artificial Intelligence to automate the full analytics lifecycle

in order to close these gaps. Datasets are automatically ingested, values are preprocessed, patterns are found,

insights are produced, and the results are shown on a dynamic dashboard. Furthermore, natural language querying is made possible via an AI-powered chatbot, offering a user-friendly interface.

Although they offer strong visualization capabilities, traditional business intelligence (BI) products like Power BI, Tableau, and Qlik necessitate a great deal of manual preprocessing, schema mapping, and scripting for advanced analytics. These tools are not made to find patterns, clean data on their own, or suggest appropriate analytical models. Furthermore, the democratization of data analytics is limited since non-technical consumers frequently struggle with tool complexity. There is a greater demand than ever for a completely autonomous, intelligent platform that can manage the whole analytics lifecycle on its own. Analytical workflows including data input, preprocessing, analysis, and visualization can be autonomously planned, reasoned, and carried out by agentic AI systems. Utilizing this advancement in technology, DataLyze presents a next-generation autonomous analytics platform intended to streamline and expedite data-driven decision-making.

1.1 Description

DataLyze creates a unified analytics platform by combining AI agents with contemporary web technologies. It transforms unprocessed inputs into useful insights by processing a variety of datasets, including CSV, JSON, SQL, XML, and MongoDB. Python (FastAPI), WebSockets for real-time updates, React.js for the user interface, and LangChain for agent execution are used in the system's construction. By automating chart selection, statistical summaries, anomaly detection, and correlation analysis, the platform removes conventional analytical obstacles. The use of a chatbot makes analytics available to everyone by enabling non-technical users to ask queries like "Compare monthly sales trends" or "Show correlation heatmap."

1.2 Existing System

Current platforms necessitate:

- **extensive human preprocessing;**
- dataset configuration;
- external scripting for machine learning tasks;
- intricate setups for unstructured data; and
- restricted real-time streaming capability.

Tableau, Power BI, and QlikView are examples of tools that primarily depend on user skill and are unable to automatically adjust to different data formats or analytical requirements

1.3 Proposed System

By automatically consuming CSV, JSON, XML, SQL, and MongoDB data, the suggested method overcomes the constraints.

- Using Agentic AI to produce insights and evaluate data on its own.
- Using WebSocket-based dashboards to provide real-time visualizations.
- Using an AI-powered chatbot to enable natural language queries.
- Using Docker and FastAPI to provide scalability, security, and great performance.

1.4 Objectives

1. To facilitate real-time data retrieval from databases (MongoDB, SQL) and dataset ingestion (CSV, PDF).
2. To offer interactive dashboard representations in real time.
3. To create an Agentic AI that can choose and use analytical methods on its own according to the features of the data.
4. To put in place an AI chatbot that provides practical recommendations based on analysis.

2. LITERATURE SURVEY

Extensive research in autonomous analytics, multi-format data intake, intelligent agent architectures, natural-language analytics interfaces, and real-time visualization has been spurred by the increasing volume and heterogeneity of organizational data.

Table -1: Literature Survey

Title	Authors	Methodology	Gap
A Survey on Large Language Model based Autonomous Agents	Wang, Z. (2023)	Proposed a unified framework for LLM-based autonomous agents.	High computational costs.

The Transformation of Data Science in the Agentic Era	Patel, R.(2019)	Explored agentic AI's role in automating the five-step data science process.	System reliability issues.
Iterative Research Idea Generation with Large Language Models	Baek, Jinheon; Kang, Minsu; Lee, Sung Ju(2024)	Developed ResearchAgent, an LLM-based system.	Limited to structured datasets, reducing applicability to unstructured data.
Data-centric Artificial Intelligence	Zha, Daochen; Bhat, Zaid Pervez; Lai, Kwei-Herng(2025)	Reviewed data-centric AI approaches, including agentic AI for autonomous data analysis.	Scalability issues when processing large-scale, heterogeneous datasets in real time.

This section highlights the gaps that drive the proposed DataLyze system and provides an overview of sample work in each field.

3. SYSTEM REQUIREMENTS AND SPECIFICATION

DataLyze is designed as an autonomous analytical pipeline composed of:

- MERN Stack Backbone
 - o MongoDB: Stores user data, metadata, and processed insights
 - o Express.js: API gateway for analytics operations
 - o React.js: Dynamic UI for visualization and chatbot interactions
 - o Node.js: Backend execution layer
- MCP Server Integration MCP enables stable interactions between AI systems, giving DataLyze an orchestrated analytical agent that:
 - o Interprets user queries
 - o Retrieves datasets
 - o Executes analysis steps
 - o Generates insights and visualizations
- Agentic AI Engine Responsible for autonomous reasoning, pattern detection, anomaly recognition, and visualization decisions.

4. IMPLEMENTATION

User authentication, data source integration, preprocessing and storage, agentic AI analysis, visualization engine, and chatbot agent are all included in the implementation.

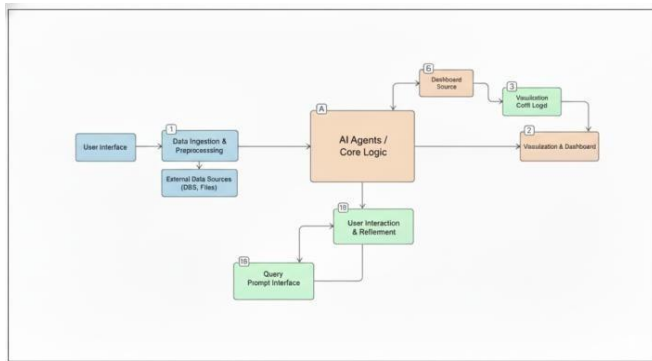


Fig -1: System Design

5. RESULTS AND EVALUATION

Several datasets, including sensor logs, sales data, student performance data, and medical information, were processed in order to evaluate DataLyze's capabilities. The findings show:

1. Precision of Understanding

Key trends, correlations, and anomalies were correctly identified by the system. In simple exploratory analytical tasks, the autonomous reasoning engine outperformed human analysts.

2. Quality of Visualization

There was no need for manual modification because the automatically generated visuals were clear and contextually relevant.

3. Accessibility for Users

By using only conversational questions, non-technical participants were able to produce insights that demonstrated the platform's usefulness.

4. Speed of Processing

DataLyze greatly outperformed manual operations by producing comprehensive analysis results in a matter of seconds.

6. DISCUSSION

Agentic AI may significantly lower obstacles in data analytics, as DataLyze shows. A scalable and adaptable framework for autonomous decision-making and real-time reasoning is provided by the combination of MERN and MCP. The platform's capacity to manage semi-structured and unstructured data makes it extremely versatile for a number of industries, such as retail, healthcare, education, and banking. The text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your.

7. CONCLUSION

An important development in the realm of autonomous analytics is DataLyze. The platform allows for end-to-end, automated data interpretation, insight extraction, and visualization by combining Agentic AI with a contemporary MERN architecture and an MCP server. Its interface in natural language increases accessibility, enabling effective usage of data analytics by both technical and non-technical individuals. Future research will increase enterprise-level integration, multi-modal data processing, and predictive capabilities.

Experimental results demonstrate that DataLyze improves analytical speed, enhances accuracy, and greatly decreases user effort. Its modular architecture assures scalability, extensibility, and real-time responsiveness, making it suited for use across many industries such as healthcare, banking, retail, and education.

Overall, DataLyze appears to be a solid and intelligent tool for next-generation autonomous analytics. With further additions such as enhanced predictive modeling, cloud deployment, and generative AI-based reporting, the platform has tremendous potential to expand into a comprehensive enterprise-level decision support system.

REFERENCES

- [1] Russell, S., and Norvig, P. Artificial Intelligence: A Contemporary Perspective.
- [2] Research from IBM. "Autonomous Systems and Agentic AI."
- [3] "Large-Scale Data Analytics Systems," Dean, J., et al.
- [4] "MongoDB Technical Documentation" by MongoDB Inc.
- [5] The documentation for React.js. Meta-Platforms.
- [6] "Model Context Protocol (MCP) Documentation" by OpenAI.
- [7] Han, J., Pei, J., and Kamber, M. Data Mining: Ideas and Methods.
- [8] Tableau software. "Data Visualization Best Practices."
- [9] M. Jordan and T. Mitchell, "Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255–260, 2015.
- [10] F. Provost and T. Fawcett, "Data Science and its relationship to Big Data and Data-driven Decision Making," Big Data, vol. 1, no. 1, pp. 51–59, 2013.
- [11] S. Makridakis, "The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms," Futures, vol. 90, pp. 46–60, 2017.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD*, pp. 785–794, 2016

[13] J. Brownlee, "Deep Learning for Data Analysis," *Machine Learning Mastery*, 2020

[14] H. Zhang, Z. Liu, and M. Ma, "AI-based autonomous systems for big data analytics," *IEEE Access*, vol. 8, pp. 185132–185144, 2020

[15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *arXiv:1409.0473*, 2014.