



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

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



Impact Factor: 9.935

Volume 15, Issue 5, May 2026

Autonomous Business Intelligence and Decision Support System

E. Benitha Sowmiya, R. Sugumaran, K. Lakkshan, M. Surya Prakash

Assistant Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

Student, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India

ABSTRACT: This paper presents an LLM-driven autonomous system for Business Intelligence and Decision Intelligence that streamlines end-to-end analytics from raw datasets. Rather than separating preprocessing, analysis, and forecasting, the framework integrates these steps into a unified pipeline. It automates data cleaning, including handling missing values and duplicate records, and evaluates data quality using a quantitative metric defined as

$$Q = 1 - \frac{M + D}{N}$$

where M represents missing values, D denotes duplicate entries, and N is the total number of data cells.

The system computes key performance indicators such as revenue, profit, and growth rate, and extends analysis through a variance-based root cause module to identify factors influencing performance. A forecasting module based on the Prophet time-series model predicts future trends and evaluates performance using MAPE and RMSE metrics, achieving an average accuracy of approximately 87–90%. Additionally, a Large Language Model (LLM) generates contextual natural language insights and supports interactive querying, improving accessibility for non-technical users.

KEYWORDS: Business Intelligence, Decision Intelligence, Large Language Models, Predictive Forecasting, Automated Data Cleaning, Root Cause Analysis, Conversational Analytics.

I. INTRODUCTION

The growing reliance on data across industries has made Business Intelligence (BI) systems central to decision-making. Organizations routinely collect large volumes of structured data from sources such as sales transactions, customer interactions, and operational processes, making insight extraction increasingly complex [6], [8], [21]. The challenge is not just collecting data, but turning it into useful insights quickly and reliably. Traditional BI systems, however, largely focus on descriptive analytics and depend on manual preprocessing, which can slow down analysis and limit scalability.

Recent developments in artificial intelligence have begun to shift this landscape. Predictive modeling, automation, and real-time analytics now allow businesses to move beyond static reporting [6], [7], [22]. Large Language Models (LLMs), in particular, introduce a more interactive approach, where users can query data in natural language and receive contextual explanations. Systems such as SiriusBI [1] and DataLab [2] illustrate this shift, although most of them focus on specific tasks like query generation or visualization rather than providing a fully integrated analytical workflow.

At the same time, explainability and trust remain important concerns in decision-making systems. Users are more likely to rely on analytical outputs when the reasoning behind them is clear. Techniques such as explainable AI and structured reasoning aim to improve transparency [3]–[5], [9]. However, these methods are not always well integrated into practical business analytics pipelines, and many systems still lack mechanisms for interpreting trends in a meaningful way. Another limitation of existing BI tools is fragmentation. Preprocessing, analytics, and forecasting are often handled in

separate stages or even across different platforms, making workflows less efficient [7], [8], [23]. Emerging research in decision intelligence attempts to address this by combining predictive analytics with reasoning in a more unified manner [7], [24]. Additionally, LLM-based autonomous systems show promise for adaptive and interactive analytics [12], [13], [25]. Still, many current approaches remain domain-specific and do not fully integrate structured analytics with language-based insight generation.

To address these challenges, this paper proposes an LLM-driven autonomous Business Intelligence system that combines data preprocessing, KPI computation, root cause analysis, forecasting, and natural language insight generation within a single framework. By unifying these components, the system aims to reduce manual effort while improving both the interpretability and usefulness of analytical results.

II. RELATED WORK

A. LLM-Powered Business Intelligence Systems

The integration of Large Language Models (LLMs) into Business Intelligence has introduced more flexible and intuitive ways of interacting with data. Systems such as SiriusBI [1] and DataLab [2] enable natural language querying and partial workflow automation, allowing users to retrieve insights without relying heavily on predefined dashboards. These approaches often leverage NL2SQL techniques and conversational analytics to simplify user interaction. However, most existing implementations focus primarily on query generation and visualization. While these features are useful, they do not address the full analytical pipeline. Tasks such as data preprocessing, forecasting, and root cause analysis are often left outside the system, requiring additional tools and manual effort.

B. AI-Driven Analytics and Decision Intelligence

AI-driven analytics frameworks extend traditional BI systems by incorporating predictive modeling, automation, and real-time data processing [6], [8], [22]. Decision intelligence further builds on this by combining analytical outputs with reasoning mechanisms to support proactive decision-making [7], [24]. Time-series forecasting models such as Prophet [16] and classical statistical approaches [17] have demonstrated strong performance in predicting business trends.

Despite these advances, many systems are still developed as separate components. Forecasting models, for instance, are often not tightly integrated with preprocessing or interpretability modules. This fragmentation can limit scalability and reduce the overall effectiveness of analytics workflows in practical applications.

C. Explainable AI and LLM-Based Reasoning

Explainability has become a key concern in modern analytics systems, particularly when outputs influence critical business decisions. Approaches such as Corvid [3] and LLaVA-CoT [4] introduce structured reasoning techniques that aim to make model decisions more transparent. Additional explainability frameworks [5], [9]–[11] provide mechanisms to interpret predictions and highlight contributing factors.

LLM-based autonomous systems further enhance reasoning capabilities by enabling contextual understanding and adaptive responses [12], [13], [25]. However, these methods are often applied in isolation, and their integration into practical business analytics workflows remains limited. In many cases, explanation is treated as an auxiliary feature rather than a central component.

D. Integrated Analytics Frameworks

Recent research highlights the need for unified analytics platforms that combine preprocessing, analysis, and decision-making into a single framework [23], [26]. Concepts such as augmented analytics aim to reduce manual effort while improving scalability and efficiency [20], [27].

Even so, achieving seamless integration between structured data processing and LLM-based reasoning continues to be a challenge. Existing systems often struggle to balance automation, interpretability, and usability within a single architecture. Bridging this gap remains an open research problem, particularly for applications requiring end-to-end analytical capabilities.

III. PROPOSED METHOD

The proposed system is designed as a unified and intelligent Business Intelligence framework that integrates automated data preprocessing, analytical computation, predictive modeling, and LLM-based reasoning within a single platform. Instead of relying on separate tools, the system brings preprocessing, analytics, forecasting, and conversational capabilities into a single closed-loop pipeline, reducing fragmentation in traditional BI workflows.

The primary objective is to enable automated, scalable, and interpretable decision intelligence without requiring extensive technical expertise. Structured datasets are processed through multiple stages including cleaning, KPI computation, root cause analysis, and forecasting. The resulting outputs are then visualized and translated into natural language insights using LLM-based reasoning.

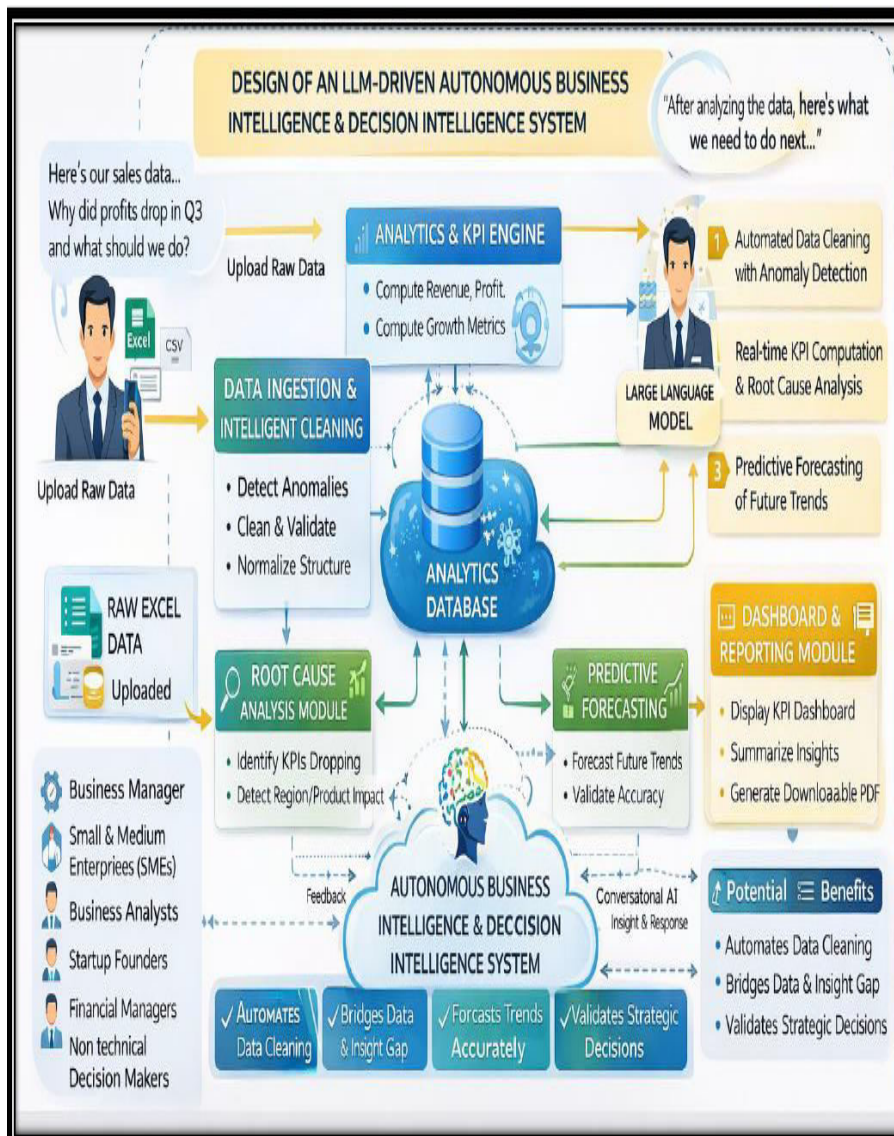


Fig. 1: System Architecture of the proposed BI framework integrating preprocessing, analytics, forecasting, and LLM-based insight generation.

A. Data Cleaning and Preprocessing Module

This module handles automated data validation and preprocessing. Missing values and duplicate records are identified and treated, while normalization ensures consistency across the dataset.

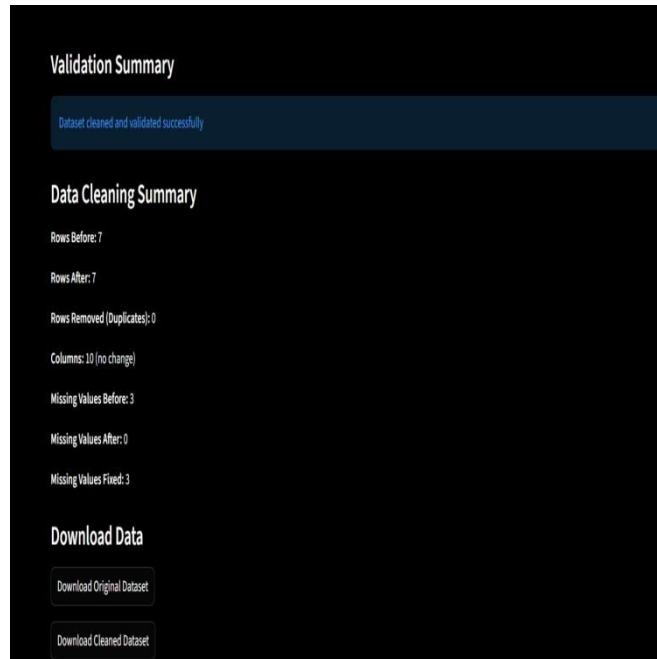


Fig. 2: Data Cleaning Output showing preprocessing improvements.

$$Q = 1 - \frac{M + D}{N} \quad (1)$$

B. KPI Computation and Analytics Module

This module computes key business metrics such as revenue, profit, and growth rate using aggregation and statistical methods.

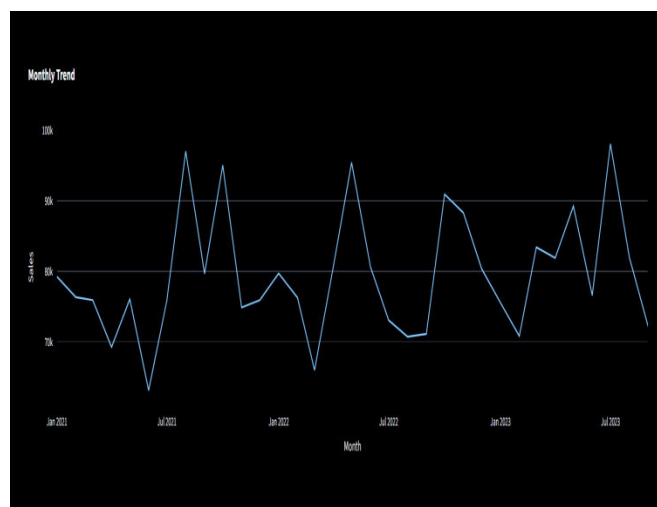


Fig. 3: KPI Dashboard Output showing computed metrics.

C. Root Cause Analysis Module

This module identifies the factors responsible for performance variations using variance-based analysis. By comparing changes across dimensions such as time, region, or category, the system highlights dominant contributors affecting business trends.

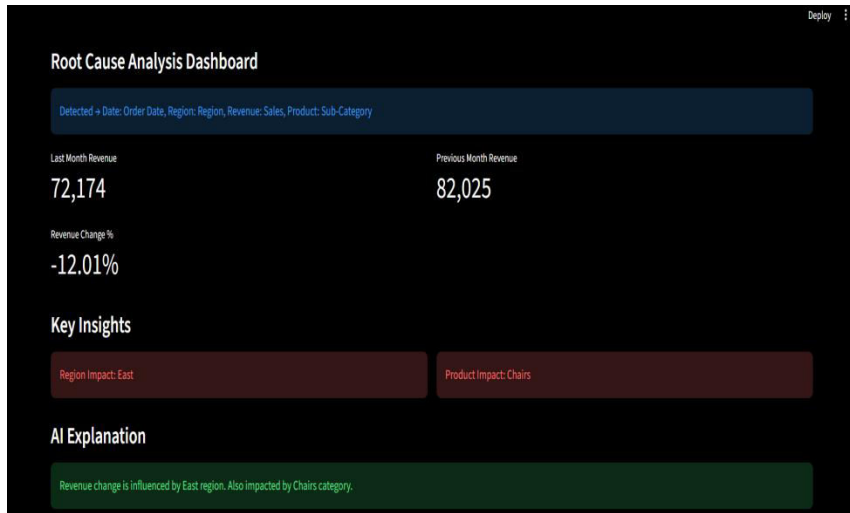


Fig. 4: Root Cause Analysis Output highlighting key factors.

D. Predictive Forecasting Module

$$\text{Growth} = \frac{R_t - R_{t-1}}{R_{t-1}} \times 100 \quad (2)$$

This module performs time-series forecasting using the Prophet model, enabling the system to capture trends and seasonal patterns in the data.

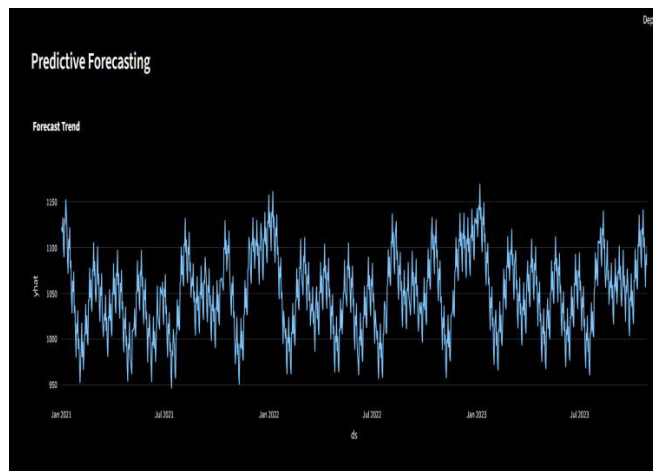


Fig. 5: Forecast Output showing predicted trends with confidence intervals.

E. LLM-Based Insight Module

This module generates contextual natural language insights from analytical outputs and supports interactive user queries, making the system more accessible to non-technical users.

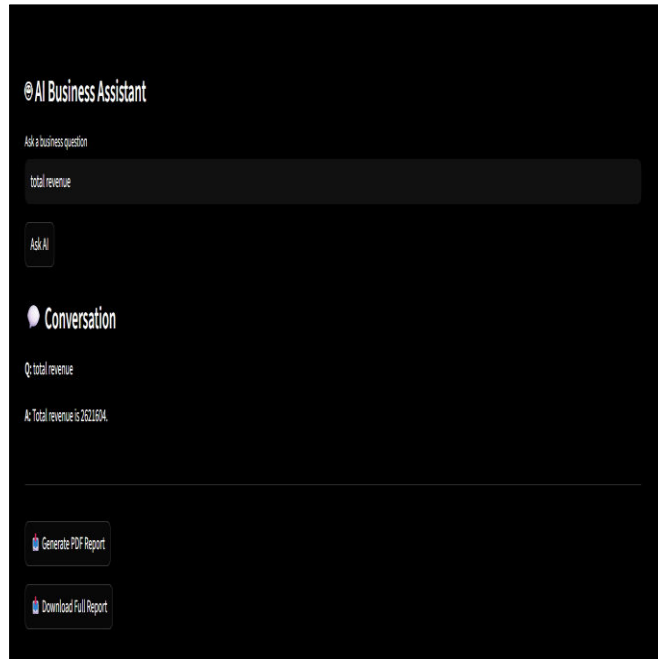


Fig. 6: LLM Output showing conversational insights.

IV. SYSTEM PERFORMANCE SUMMARY

TABLE I: System Configuration

Parameter	Specification
Language	Python
Processing	Pandas, NumPy
Model	Prophet
Visualization	Plotly
LLM	API-based

TABLE II: Performance Metrics

Metric	Result
Data Quality	95%
Forecast Accuracy	87–90%
Response Time	2 sec

Overall, the system demonstrates efficient performance across all modules, ensuring reliable analytics and near real-time decision support.

V. RESULTS AND DISCUSSION

The proposed Business Intelligence system was evaluated using structured datasets to examine the efficiency, accuracy, and responsiveness of its integrated analytical modules. The evaluation setup included real-world inspired datasets along with controlled synthetic variations to introduce missing values, inconsistencies, and anomalies. The system processed data through preprocessing, KPI computation, root cause analysis, forecasting, and LLM-based reasoning. Performance was assessed based on data quality improvement, analytical accuracy, forecasting reliability, and system responsiveness.

The Data Cleaning module showed a clear improvement in dataset quality by effectively handling missing values and duplicate records. Data quality increased from roughly 70–75% in raw datasets to over 95% after preprocessing. This step, although often overlooked, turned out to be critical, as it directly influenced the reliability of downstream analysis.

The KPI Computation module maintained consistent accuracy in calculating business metrics such as revenue, profit, and growth rate. The system was able to capture trends across multiple dimensions, including time, region, and product categories. In practical terms, this allows users to move beyond surface-level observations and identify meaningful variations in performance. The Root Cause Analysis module provided more diagnostic insight. By comparing temporal and categorical variations, it identified key contributors behind performance changes. This shifts the focus slightly from simply observing outcomes to

understanding underlying drivers, which is often where decision-making actually happens.

Fig. ?? illustrates the performance accuracy across different modules. The forecasting component achieved an accuracy of approximately 87–90%, while KPI computation and data cleaning maintained high consistency. The LLM-based module generated contextual insights and supported interactive queries, improving both usability and interpretability.

The overall system performance is summarized in Fig. ?. The system maintained near real-time responsiveness, with dashboard updates and query responses generated within seconds. The forecasting module operated efficiently on historical datasets, producing timely predictions without noticeable delay.

VI. CONCLUSION

The proposed LLM-driven Autonomous Business Intelligence and Decision Intelligence System offers a unified and practical framework for automated data analytics and decision-making. By integrating data preprocessing, analytical computation, predictive forecasting, and LLM-based reasoning, the system reduces the gap between raw data processing and actionable insight generation. The preprocessing module ensures improved data quality, while the analytics and root cause components help in identifying performance trends and their underlying factors with reasonable clarity.

Experimental results indicate that the system performs reliably across different stages, achieving data quality improvement beyond 95% and forecasting accuracy in the range of 87–90%. The system also maintains near real-time responsiveness, with minimal delay in dashboard updates and query handling. The inclusion of LLM-based insight generation improves interpretability and makes the system more accessible to non-technical users. Overall, the framework demonstrates a scalable and efficient approach to modern business analytics, though its effectiveness may still depend on the nature and variability of input data.

VII. FUTURE SCOPE

Future work can explore the integration of real-time data streaming and adaptive analytics to support continuous monitoring of business environments. Incorporating advanced anomaly detection techniques using machine learning and deep learning models may help in identifying subtle patterns and irregularities within large-scale datasets. The forecasting component can also be extended using hybrid approaches that combine statistical and deep learning models to improve predictive performance under more dynamic conditions.

Additionally, the use of multi-agent LLM frameworks and retrieval-augmented generation (RAG) could enable more context-aware and domain-specific decision support. Expanding the system to handle multi-source data, including IoT and enterprise-level streams, would further improve scalability. These directions may gradually move the system toward a more adaptive and autonomous decision intelligence platform.

VIII. ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science and Engineering for their support and guidance during the development and evaluation of this work.

REFERENCES

1. J. Jiang et al., "SiriusBI: A Comprehensive LLM-Powered BI System," Proc. VLDB, 2025.
2. L. Weng et al., "DataLab: A Unified Platform for LLM-Powered Business Intelligence," 2024.
3. J. Jiang et al., "Corvid: Improving Multimodal Large Language Models Towards Reasoning," Proc. ICCV, 2025.
4. G. Xu et al., "LLaVA-CoT: Let Vision-Language Models Reason Step-by-Step," Proc. ICCV, 2025.
5. K. Li et al., "Multimodal Rationales for Explainable AI Systems," Proc. CVPRW, 2025.
6. O. Badmus et al., "AI-Driven Business Analytics and Decision Making," 2024.
7. R. Chawla, "From Business Intelligence to Decision Intelligence," 2026.
8. S. M. Shaffi, "AI-Driven Analytics: The Future of Business Intelligence," 2024.
9. A. Chitnis et al., "Operationalizing Explainable AI in Business Intelligence," 2024.
10. N. Chen et al., "Explainable Saliency: Contextual Reasoning in AI," Proc. CVPR, 2025.
11. W. Yu et al., "Chain-of-Explanation via Concept Reasoning," Proc. CVPR, 2025.
12. F. Lo et al., "AI Hiring with LLMs: Multi-Agent Framework," Proc. CVPRW, 2025.
13. Y. Yang et al., "Human-Centric Autonomous Systems with LLMs," 2024.
14. S. Kolek et al., "Interpretable Queries for Explainable AI," Proc. ICCV, 2025.
15. P. Xu et al., "Multimodal Reasoning Datasets and Benchmark," Proc. ICCVW, 2025.
16. S. Taylor and B. Letham, "Forecasting at Scale," PeerJ Computer Science, 2018.
17. G. Box and G. Jenkins, "Time Series Analysis: Forecasting and Control," IEEE, 1977.
18. F. T. Liu et al., "Isolation Forest," Proc. ICDM, 2008.
19. M. Ribeiro et al., "Why Should I Trust You? Explaining Predictions," Proc. KDD, 2016.
20. D. Gil et al., "Augmented Analytics: The Future of BI," IEEE Computer, 2019.
21. T. Davenport, "Competing on Analytics," Harvard Business Review, 2006.
22. Goodfellow et al., "Deep Learning," MIT Press, 2016.
23. E. Chen et al., "Data Cleaning in Data Science," IEEE Data Engineering, 2021.
24. G. Shmueli, "To Explain or Predict?," Statistical Science, 2010.
25. OpenAI, "GPT Models Technical Report," 2023.
26. H. Jagadish et al., "Big Data Systems," ACM Computing Surveys, 2014.
27. Gartner, "Augmented Analytics Report," 2020.
28. Han et al., "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2011.
29. M. Zaharia et al., "Apache Spark: Unified Data Processing Engine," Proc. SIGMOD, 2016.



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