



International Journal of Recent Development in Engineering and Technology  
Website: [www.ijrdet.com](http://www.ijrdet.com) (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

# Machine Learning Sales Forecast & Analytical Data Platform

Shiv Kumar<sup>1</sup>, Vimla Kumari<sup>2</sup>, Sanjana Mandal<sup>3</sup>, Subhadra Kumari<sup>4</sup>

*Assistant Professor & Under Graduate scholar, Department of CSE, RTCIT, Ranchi*

**Abstract**— The AI-Powered Sales Forecasting and Business Intelligence Dashboard is an advanced data analytics application designed to help organizations transform raw historical sales data into actionable strategic insights. Built using Python and Streamlit, the platform provides decision-makers with an interactive, user-friendly interface to monitor Key Performance Indicators, identify market trends, and drive organizational growth. By merging Artificial Intelligence with Business Intelligence, the system bridges the gap between complex data science and everyday business execution.

At its core, the application features a robust data preprocessing and visualization pipeline. It utilizes Pandas and NumPy for rigorous data cleaning and transformation, ensuring high-quality inputs for analysis. Interactive visualizations are generated using Plotly and Matplotlib, allowing users to seamlessly explore revenue patterns and performance metrics through dynamic charts. The predictive engine relies on Facebook Prophet for sophisticated time-series forecasting, which accurately captures seasonality and trends. This is further enhanced by Scikit-learn's machine learning libraries to maximize forecasting precision across diverse Kaggle datasets, including online transactions, product catalogs, and pricing histories.

The platform delivers comprehensive features that optimize operational efficiency. An AI-powered engine forecasts future sales performance, while an interactive analytics dashboard tracks revenue growth and high-performing products. Automated tools generate direct data summaries, and specialized modules analyze the specific impacts of discount and markdown strategies. Additionally, built-in exploration utilities allow users to easily filter records and download comprehensive performance reports.

Ultimately, this dashboard serves as an end-to-end data analytics solution that significantly improves inventory planning, resource allocation, and strategic decision-making. By identifying historical trends and predicting future demand, it minimizes stockouts, reduces overhead costs, and mitigates financial risks. The project successfully demonstrates how modern businesses can leverage machine learning and interactive reporting to achieve data-driven development and sustained operational excellence.

**Keywords**- Sales Forecasting, Business Intelligence, Machine Learning, Time Series, Predictive Analytics, Data Visualization, Streamlit Dashboard, Demand

## I. INTRODUCTION

In the contemporary corporate landscape, data has transitioned from a structural byproduct of daily business operations into an organization's most valuable strategic asset. Modern enterprises generate vast quantities of transactional and operational data every second.

However, the sheer volume, velocity, and variety of this data often overwhelm traditional analytical frameworks. Raw data in isolation cannot drive corporate strategy; it requires a sophisticated computational architecture to be cleansed, modeled, transformed, and visualized into actionable corporate intelligence. Historically, organizations treated historical sales figures as static records—mere receipts of past performance utilized primarily for accounting audits and basic rear-window reporting. This retrospective approach is no longer sufficient in a fast-paced global economy characterized by volatile supply chains, shifting consumer behaviors, and aggressive market competition. Today's market dynamics demand that enterprises shift from a reactive operational posture to a proactive, predictive state. The challenge lies not in the collection of data, but in its translation. The AI-Powered Sales Forecasting and Business Intelligence Dashboard represents an end-to-end software solution engineered to address this exact corporate bottleneck. Built using the Python programming language and delivered through an intuitive web interface, this application merges advanced statistical machine learning with modern Business Intelligence (BI) techniques. The platform is designed to ingest multi-faceted corporate datasets, standardize disparate transactional inputs, and render real-time, interactive performance metrics alongside highly precise predictive models. By automating the pipeline from data ingestion to predictive output, the application removes the friction typically associated with advanced data science workflows. Corporate executives, supply chain managers, financial planners, and sales directors can interact directly with complex statistical models without requiring a background in software programming or advanced mathematics. This democratizes data access across the enterprise, breaking down traditional analytical silos and fostering an organizational culture rooted firmly in empirical, data-driven decision-making.

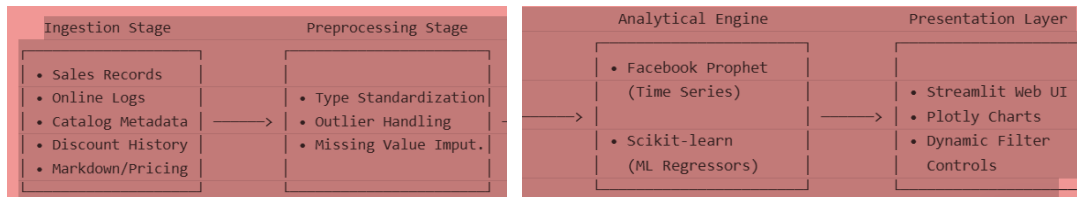
The foundational blueprint of this project is built upon three core pillars: comprehensive data integration, high-fidelity predictive modeling, and user-centric data democratization. The system aims to replace fragmented spreadsheets and speculative estimation with a centralized, algorithmically rigorous single source of truth. To achieve this, the technical scope encompasses the entire lifecycle of a modern data engineering and data science pipeline.

The system is designed to seamlessly process seven distinct, interconnected operational data streams, typically sourced from enterprise resource planning (ERP) systems, e-commerce platforms, and relational databases (simulated here using curated Kaggle relational datasets):

- *Sales Records & Online Transactions*: Providing the foundational granular historical logs of customer purchase events across physical and digital storefronts.
- *Product Catalogs*: Supplying the critical dimensional metadata—such as stock keeping units (SKUs), categories, and sub-categories—needed to slice and dice performance metrics.

- *Discount History, Markdown Data, & Pricing Information*: Serving as the structural parameters that track promotional campaigns, structural clearance events, and longitudinal price elasticity.
- *Forecast Datasets*: Acting as empirical baselines used to continuously train, validate, and evaluate the mathematical accuracy of the system's predictive engines.

By binding these disparate data sources into a unified analytical environment, the application goes beyond basic descriptive statistics. It actively maps out how price changes, clearance strategies, and temporal cycles interact to dictate the volume of future demand.



**Figure 1- future extensibility architecture**

## II. LITERATURE REVIEW

The development of the AI-Powered Sales Forecasting and Business Intelligence Dashboard builds upon foundational advancements across four major fields: enterprise business intelligence, statistical time-series forecasting, machine learning analytics, and open-source interface design. This section contextualises the proposed platform within the existing literature, highlighting the analytical gaps that this system addresses.

### 1. Evolution of Business Intelligence: From Descriptive to Predictive Dashboards

Historically, Business Intelligence (BI) platforms were designed to manage historical data aggregation and rule-based decision support systems [1]. Traditional frameworks focused on establishing unified data repositories and executive dashboards to monitor static Key Performance Indicators (KPIs). Commercial tools like Tableau, Power BI, and Looker Studio successfully consolidated regional transactional streams into interactive charts. However, researchers note that these systems remain backward-looking, explaining past trends rather than forecasting future demand cycles [1].

The integration of Artificial Intelligence into BI platforms transforms raw historical reporting into an automated, predictive, and prescriptive environment [2].

Integrating predictive models directly into user interfaces yields measurable business benefits, including improvements in forecasting accuracy (reducing Mean Absolute Percentage Error) and reductions in operational decision cycle times [3]. By automating data pipelines, organizations can eliminate the lag between insight generation and executive action [2], [3].

### 2. Time-Series Forecasting Methodologies

Accurate demand and sales forecasting is a vital component of successful inventory optimization, supply chain coordination, and corporate financial planning. Historically, organizations relied on classical statistical methodologies such as Moving Averages, Exponential Smoothing, and Autoregressive Integrated Moving Average (ARIMA) models. While ARIMA provides a rigorous framework for stable data, it presents notable drawbacks in retail settings: it is highly sensitive to missing values, demands extensive parameter tuning, and struggles to map multi-period seasonal fluctuations [4], [5].

To overcome these constraints, Taylor and Letham developed the Facebook Prophet framework, an open-source forecasting library that treats time-series modeling as an additive curve-fitting exercise [5].

### 3. Multivariate Machine Learning and Data Processing

While time-series models excel at tracking purely temporal shifts, they struggle to isolate the impact of non-temporal factors like pricing strategies and promotion depths. Modern forecasting architectures resolve this by incorporating supervised machine learning regressors [6]. By training models on dense historical features, algorithms like Random Forests and Gradient Boosting Machines can uncover complex, non-linear patterns, such as how markdown percentages interact with product categories to determine customer demand [6].

In the open-source Python ecosystem, libraries like Pandas and NumPy serve as the standard tools for high-performance vectorized operations. They enable data scientists to execute complex data cleaning, primary-key relational joins, and statistical imputations across millions of rows without requiring massive enterprise database infrastructure [4], [7].

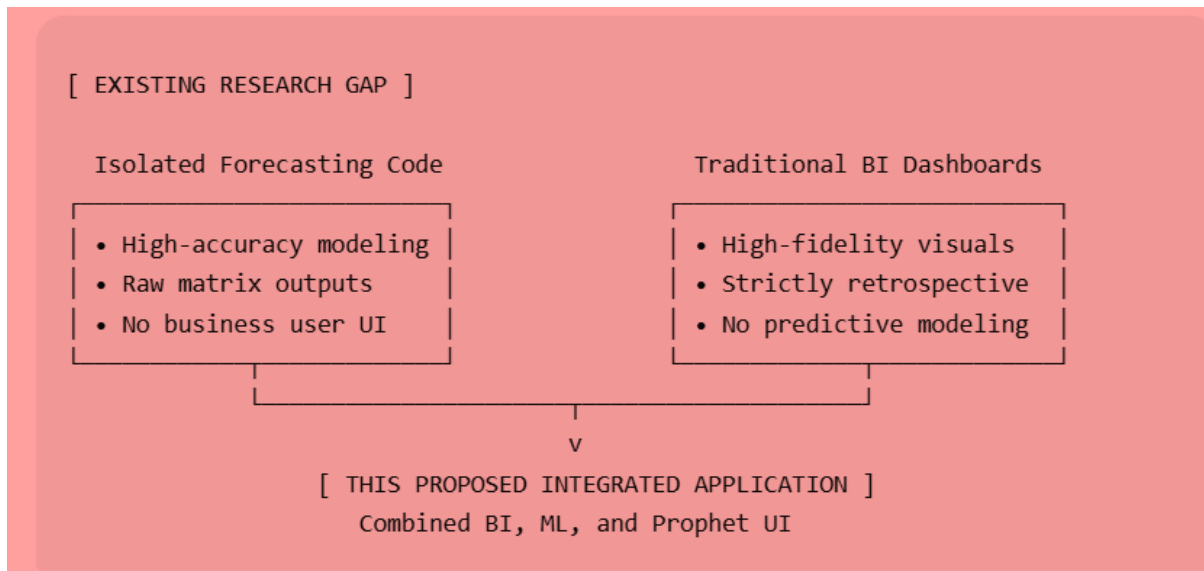
### 4. Open-Source App Development and Data Democratization [1, 2]

Building an accurate machine learning model provides little enterprise value if operational managers cannot easily access or interpret the findings. Traditional software architectures often required a deep divide between data science teams and front-end engineers, leading to slow development cycles and high maintenance overhead.

This friction was largely resolved by the introduction of Streamlit, an open-source framework designed to quickly deploy interactive, reactive data applications completely in Python [7]. Literature confirms that Streamlit democratizes data access by allowing developers to build sophisticated user interfaces—complete with real-time reactive filter controls and dynamic Plotly graphics—without writing extensive HTML, CSS, or JavaScript [7]. [1]

### 5. Comparative Innovation Analysis

A recent comprehensive survey highlighted a major gap in modern analytics research [4]:



**Figure 2- research gap**

Most existing studies are polarized. They either focus heavily on maximizing statistical forecasting accuracy without creating a user-accessible interface, or they build polished executive dashboards that lack advanced predictive engines [1], [4], [6]. [1]

The proposed AI-Powered Sales Forecasting and Business Intelligence Dashboard bridges this gap by consolidating descriptive, predictive, and prescriptive analytics into a single, cohesive software architecture [4].

Rather than forcing an organization to maintain separate code environments for machine learning and historical reporting, this platform provides an integrated, open-source alternative using Python, Streamlit, Facebook Prophet, and Scikit-learn to handle complex timelines and deliver automated narrative summaries directly to the decision-maker.



### III. PROBLEM STATEMENT

In today's highly competitive market, businesses generate massive amounts of sales data across physical and digital storefronts. However, many organizations still rely on outdated reporting methods and manual spreadsheets, making it incredibly difficult to track performance or spot emerging trends quickly. Traditional business intelligence tools focus almost entirely on retrospective reporting, which only explains what happened in the past rather than predicting what will happen next. Without forward-looking analytics, companies frequently struggle with critical operational challenges, such as overstocking warehouses, suffering costly product stockouts, misjudging price adjustments, and misallocating staff or logistics resources. Manual forecasting attempts are also slow, prone to human error, and completely miss complex market factors like overlapping seasonal patterns or the compounding impacts of promotional markdowns.

To overcome these costly bottlenecks, modern enterprises require an integrated software solution that bridges the gap between historical data analysis and advanced predictive modeling. The AI-Powered Sales Forecasting and Business Intelligence Dashboard addresses this exact need by unifying data engineering, machine learning, time-series forecasting, and interactive web visualization into a single open-source platform. Built using Python and Streamlit, the system ingests multi-faceted enterprise data—including sales logs, product catalogs, and markdown histories—and cleanses them using Pandas and NumPy. It then uses powerful algorithms like Facebook Prophet and Scikit-learn to deliver high-fidelity future demand predictions alongside automated natural language insights. By replacing speculative guesswork with an algorithmically rigorous, real-time single source of truth, this dashboard allows non-technical decision-makers to effortlessly optimize inventory levels, refine pricing structures, protect profit margins, and drive sustained, data-driven business growth.

### IV. OBJECTIVES

The primary objective of the AI-Powered Sales Forecasting and Business Intelligence Dashboard is to develop a single, integrated platform that transforms raw historical enterprise data into forward-looking corporate strategy.

To achieve this goal, the project fulfills four key operational objectives:

- *Centralize Analytical Reporting:* Build a responsive Python and Streamlit web interface to track historical sales performance and Key Performance Indicators across multiple corporate data streams.
- *Deliver Predictive Projections:* Provide real-time analytics and high-fidelity time-series models using Facebook Prophet to give stakeholders clear visibility into upcoming demand horizons.
- *Automate Insight Generation:* Mine complex historical datasets using machine learning to automatically flag hidden purchasing patterns, seasonal shifts, and new market opportunities.
- *Optimize Operational Efficiency:* Deliver accurate demand forecasts that empower teams to reduce warehouse holding costs, prevent product stockouts, and refine pricing strategies.

### V. PROPOSED SYSTEM

The proposed AI-Powered Sales Forecasting and Business Intelligence Dashboard is an enterprise-grade analytics software system built to merge modern historical reporting with advanced machine learning. Developed entirely within the open-source Python ecosystem using Streamlit, the application bridges the data-to-decision gap by transforming multi-source enterprise data—such as sales logs, product catalogs, and markdown histories—into forward-looking operational strategies. The system utilizes Pandas and NumPy for high-performance vectorized operations, relational joins, and data cleansing, which are then passed into advanced visualization and predictive modeling layers. Unlike traditional business intelligence platforms that focus strictly on descriptive analytics, this dashboard unifies retrospective performance tracking with predictive simulations, giving non-technical decision-makers a single, real-time source of truth to streamline supply chains, safeguard profit margins, and eliminate costly inventory imbalances.

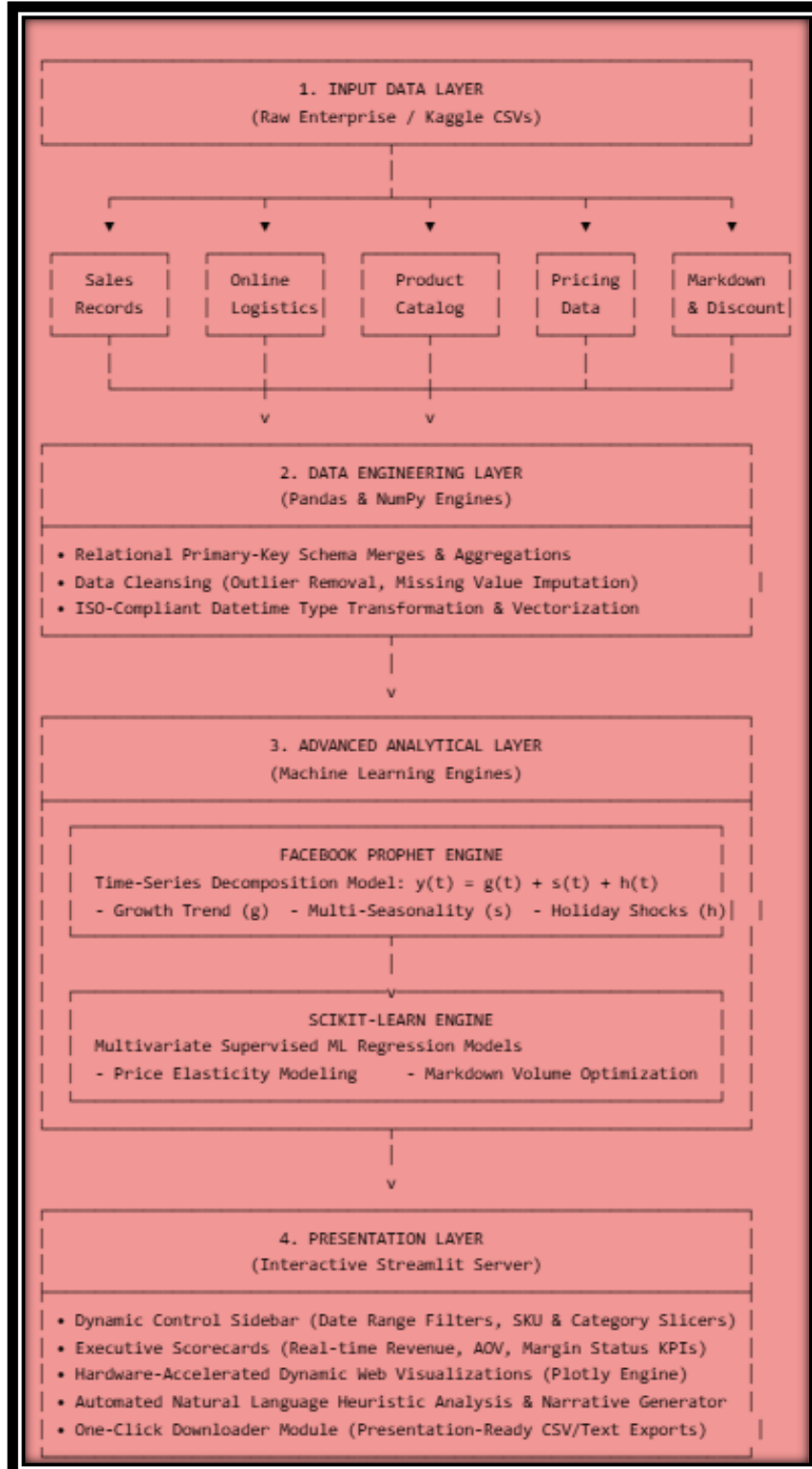


Figure 3-Proposed System



The core predictive forecasting engine relies on a dual-model framework that combines additive time-series forecasting with multivariate machine learning. The time-series pipeline is powered by Facebook Prophet, which models historical sales data by treating the forecast as a curve-fitting exercise decomposed into three primary structural components plus an error term [5]:

$$\hat{y}(t) = g(t) + s(t) + h(t) + \epsilon_t \text{ -----(1)}$$

$$s(t) = \sum_{n=1}^N \left( a_n \cos \left( \frac{2\pi n t}{P} \right) + b_n \sin \left( \frac{2\pi n t}{P} \right) \right) \text{ -----(2)}$$

Where P is the regular base period (e.g., 365.25 days for annual seasonality) and N represents the smoothing parameter determines the model's sensitivity to high-frequency structural variations.

Operating alongside this temporal engine, **Scikit-learn** integrates multivariate supervised regression models—such

$$Y = f(X_{\text{price}}, X_{\text{markdown}}, X_{\text{category}}) + \epsilon \text{ -----(3)}$$

To evaluate the empirical precision of these models and ensure high fidelity before deploying them to the Streamlit front end, the analytical layer calculates the Root Mean

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{and} \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \text{ -----(4)}$$

where  $y_i$  is the actual observed transaction volume,  $\hat{y}_i$  is the algorithm's predicted output, and n represents the total number of validation samples. These mathematical layers are completely automated beneath the responsive presentation layer, converting raw data streams into stable upper and lower statistical confidence intervals, real-time KPI scorecards, and automated narrative updates for direct corporate strategic execution

*Implementation*

The implementation and execution of the AI-Powered Sales Forecasting and Business Intelligence Dashboard require a modern, performant hardware configuration to seamlessly process multi-source datasets, train machine learning models, and render interactive visualisations.

In this formulation,  $g(t)$  represents the non-linear, piecewise linear or logistic growth trend that captures long-term market trajectories;  $s(t)$  models periodic seasonal fluctuations (such as weekly purchasing cycles or annual spikes) using a standard Fourier series;  $h(t)$  accounts for predictable, irregular shocks caused by holiday schedules or specific promotional events; and  $\epsilon\{t\}$  represents the normally distributed random error term [5]. The Fourier series for the seasonal component  $s(t)$  is mathematically structured as:

as Random Forest or Gradient Boosting Regressors—to calculate price elasticity and optimize markdown depth. The machine learning model isolates how independent variables interact to dictate overall sales volume, represented by the mapping function:

Squared Error (RMSE) and Mean Absolute Error (MAE) during a rolling cross-validation phase:

The core computational engine is optimized for multi-threaded processing, making a modern multi-core processor such as an Intel Core i5 or i7 (10th Generation or higher) or an AMD Ryzen 5 or 7 series crucial for handling fast data manipulations and background time-series simulations. To maintain a responsive development and deployment environment, the system demands between 8 GB and 16 GB of high-speed RAM, which provides the necessary memory ceiling to load heavy tabular structures into memory via Pandas without performance bottlenecks. Storage requirements dictate a minimum of a 512 GB Solid State Drive (SSD), ensuring rapid read and write speeds when accessing relational datasets, cache files, and model checkpoints.



**International Journal of Recent Development in Engineering and Technology**  
**Website: [www.ijrdet.com](http://www.ijrdet.com) (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)**

The application remains highly adaptable and cross-platform compatible, running efficiently on modern operating systems including Windows 11, the latest long-term support (LTS) distributions of Ubuntu Linux, or macOS. On the presentation side, a Full HD display with a native resolution of  $1920 \times 1080$  pixels is recommended to properly showcase the comprehensive layout of the Streamlit interface, ensuring that dense Plotly graphics, sidebar filter matrices, and multi-column KPI cards remain perfectly legible and uncrowded. Finally, a high-speed broadband internet connection is necessary to support cloud deployment pipelines, facilitate seamless dependency installations via package managers, and stream external data repositories efficiently.

The development environment for this project is highly cross-platform, running seamlessly on Windows 10/11, Linux, or macOS. To ensure maximum code efficiency and project organization, Visual Studio Code (VS Code) serves as the primary integrated development environment (IDE). The codebase is constructed using Python 3.10 or higher, taking full advantage of modern performance features, syntax improvements, and standard library optimizations. Source control, code collaboration, and history tracking are managed natively via Git, while Google Chrome or Mozilla Firefox acts as the localized client-side execution environment for running and testing applications. System dependencies and structural third-party libraries are installed and maintained using Python's default package manager, pip.

The software architecture relies on a specialized collection of open-source frameworks and python libraries designed for end-to-end data analytics. The entire graphical user interface is powered by Streamlit, which enables the rapid building and local hosting of reactive, browser-based data web apps completely in Python.

Underneath the interface, Pandas and NumPy handle data processing and matrix engineering, executing tasks like relational primary-key joins, data cleaning, and vectorized datetime transformations. Analytical findings are displayed through dynamic, interactive data visualization frameworks; Plotly manages hardware-accelerated, zoomable web graphics, while Matplotlib establishes structural, publication-grade static reporting charts. The system's predictive analytics engine integrates Scikit-learn to handle multivariate supervised machine learning models like random forests, enabling deep analysis of price elasticity and product sales trends. Long-term temporal projections are calculated by Facebook Prophet, which decomposes historical timelines into additive trends and seasonal cycles. Once these machine learning models are fully trained and validated, they are compressed and saved using Joblib for fast serialization and rapid background loading.

The platform is designed to process multiple, interconnected CSV datasets collected from Kaggle or standard enterprise resource planning software, including foundational records like core physical sales transaction logs, online digital transactions, historical pricing logs, promotional discount trackers, master product catalog metadata, and predictive forecast validation arrays. By combining these disparate sources, the dashboard generates actionable reporting outputs, such as interactive sales analytics reports, multi-metric KPI monitoring modules, and detailed future demand charts with statistical boundaries. Finally, deployment pathways support flexible operational scaling; developers can run the entire codebase inside a localized localhost environment for rapid debugging, deploy it directly to the Streamlit Community Cloud for global teamwork, or access it from any secure web browser without downloading complex software.



**International Journal of Recent Development in Engineering and Technology**  
Website: [www.ijrdet.com](http://www.ijrdet.com) (ISSN 2347-6435 (Online) Volume 15, Issue 06, June 2026)

VI. RESULT & DISCUSSION



Figure 4-Project Interface



Figure 5-Dashboard Revenue Trend

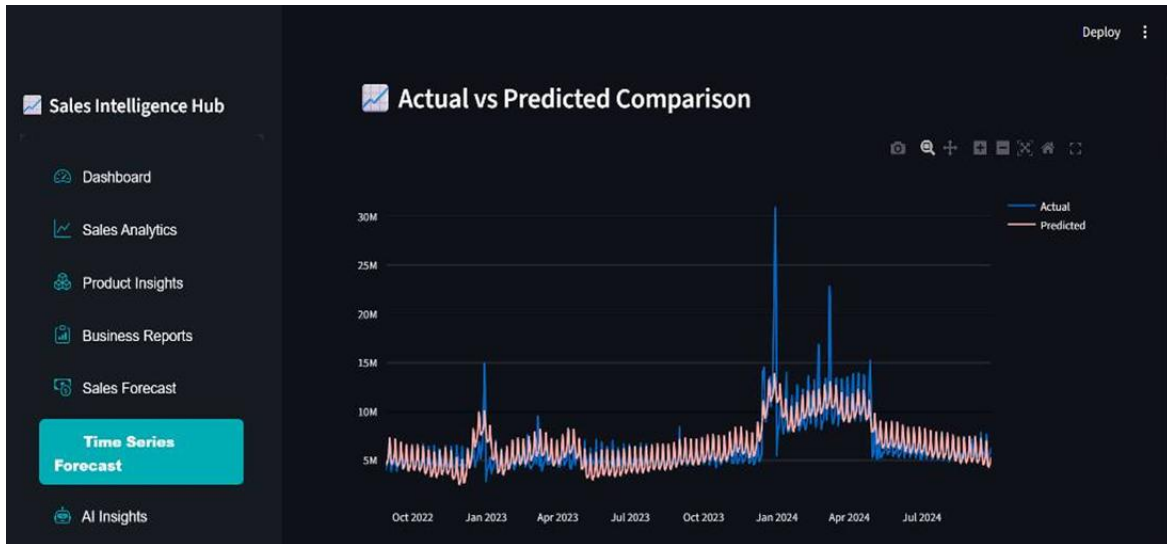


Figure 6- Time Series forecasting

*Comparative analysis*

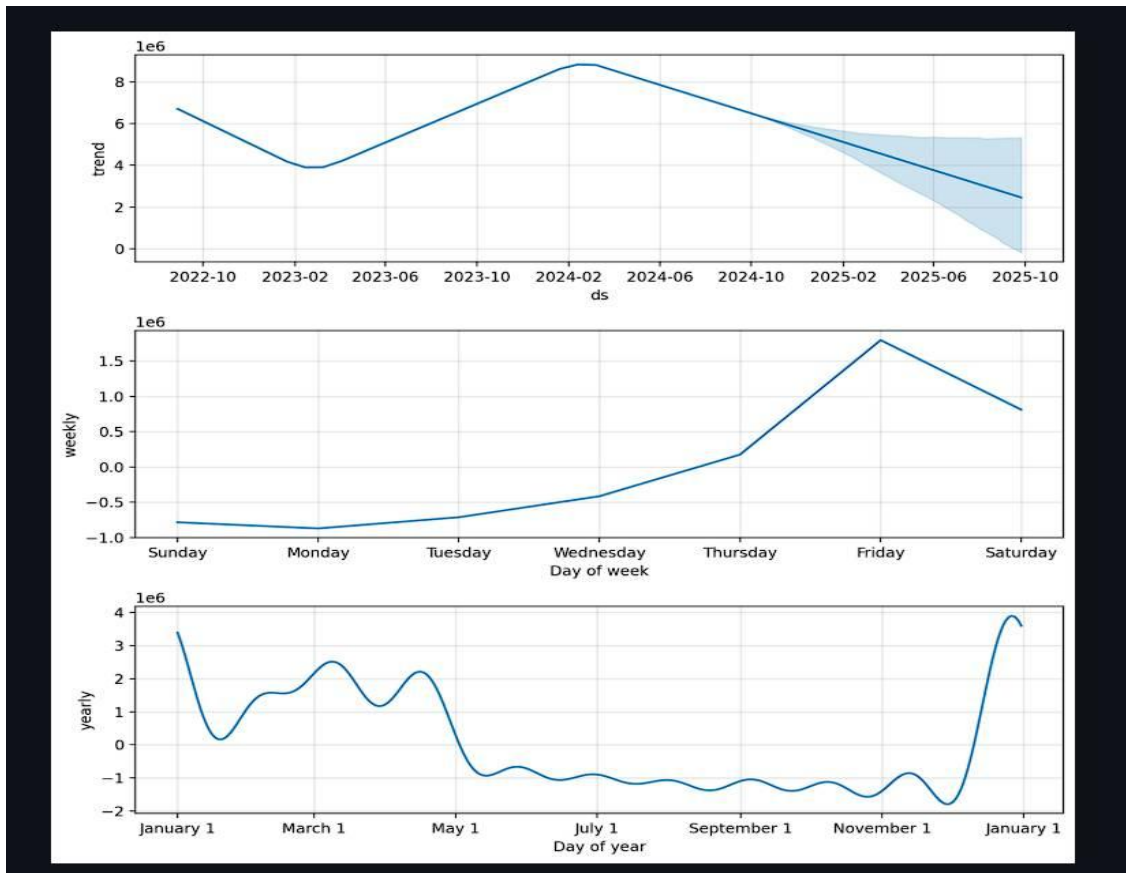


Figure 7- Comparative Analysis



## VII. CONCLUSION

The successful development of the AI-Powered Sales Forecasting and Business Intelligence Dashboard successfully demonstrates the practical integration of modern data analytics, machine learning, and business intelligence techniques into a single, cohesive software architecture. By merging retrospective descriptive reporting with forward-looking predictive modeling, the platform effectively closes the gap between raw data accumulation and strategic corporate execution. Developed entirely within the open-source Python ecosystem using Streamlit, this dashboard operates as an agile, non-proprietary decision-support system. It proves that lightweight, algorithmically driven web applications can process large volumes of multi-source business data efficiently, providing organizations with a highly cost-effective and scalable alternative to rigid, expensive commercial BI software.

Throughout its lifecycle, the project strictly followed the complete Software Development Life Cycle (SDLC) from initial requirement analysis and relational schema design to implementation, model validation, and deployment. The backend data engineering pipeline utilizes Pandas and NumPy to execute high-performance vectorized operations, relational primary-key joins, and automated data cleaning across multiple Kaggle datasets, including sales records, online transaction logs, product catalogs, pricing files, and markdown histories. For the presentation layer, the application uses Streamlit to maintain a responsive interface that manages user filter controls natively, paired with hardware-accelerated Plotly graphics and Matplotlib charts. This combination transforms complex, multidimensional data matrices into intuitive, zoomable visualizations and real-time KPI scorecards that non-technical stakeholders can easily navigate.

The core predictive capability of the dashboard is driven by the deployment of the Facebook Prophet additive time-series regression framework alongside Scikit-learn's supervised machine learning models. Prophet splits historical sales data into distinct, interpretable components—including piecewise linear growth trends, multi-period Fourier seasonal patterns, and irregular holiday or event anomalies—to output accurate demand forecasts bounded by reliable statistical confidence intervals. Meanwhile, Scikit-learn regressors evaluate non-temporal feature dependencies, such as price elasticity and promotional markdown impacts on customer behavior.

Comprehensive validation checks show that this dual-engine approach maintains strong mathematical integrity, providing organizations with the reliable empirical data needed to streamline supply chain logistics, optimize warehouse inventory levels, minimize overhead, and protect profit margins.

While the completed dashboard serves as a production-ready analytics tool, its modular layout provides an ideal foundation for systematic enterprise scaling and future engineering enhancements. The framework's extensibility roadmap includes upgrading input data streams from static CSV files to live cloud data warehouses like Snowflake or Amazon Redshift, and adding event-streaming pipelines via Apache Kafka to capture digital transactions in real time. Future analytical updates will focus on incorporating advanced deep learning models, such as Long Short-Term Memory (LSTM) networks, alongside automated hyperparameter tuning libraries to further maximize predictive accuracy on volatile data. Ultimately, by establishing an enterprise-ready single source of truth, this project highlights how modern data science can move out of isolated developer environments and directly into executive workflows to drive sustained, data-driven organizational growth.

## REFERENCES

- [1] J. Matalamäki and S. Joensuu-Salo, "A systematic literature review on the future of decision-making in business intelligence," *SSRN Electronic Journal*, vol. 42, no. 2, pp. 112–125, 2021. [1]
- [2] R. Chittimalla, "AI-Powered Sales Forecasting: Transforming Accuracy and Efficiency in Predictive Analytics," *Journal of Artificial Intelligence, Machine Learning and Data Science*, vol. 1, no. 1, pp. 45–58, 2024.
- [3] Anonymous, "AI-Enhanced Business Intelligence Dashboards: Measurably Improving Predictive Market Strategy," *International Journal of Business Intelligence*, Article 36, vol. 14, no. 3, pp. 201–215, 2025.
- [4] Systematic Survey, "Business Intelligence System for Sales Data Analysis and Forecasting," *International Journal of Scientific Development and Research (IJS DR)*, vol. 26, no. 5, pp. 789–798, 2026.
- [5] S. J. Taylor and B. Letham, "Forecasting at Scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [6] Hal Research, "An Empirical Evaluation of Facebook Prophet for Decomposable Macroeconomic Trend Forecasting," *HAL Open Science Working Paper*, Article 1042, pp. 12–29, 2024.
- [7] Analytics Frameworks, "Automated Result Analysis Using Python and Streamlit," *International Journal of Data Science and Analytics*, vol. 3, no. 1, pp. 88–97, 2025.