iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83085

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Predictive Analytic and Explainable Ml-Based Sales Forecasting System for Business Performances

Sarika Bangar[1], Prof. Satish Gujar[2]
MCA Second year, JSPM University, Pune, India

Abstract:** Accurate sales forecasting is essential for effective business planning, inventory management, and revenue optimization. Traditional forecasting methods often rely on basic statistical techniques, which fail to capture complex patterns such as seasonality, customer behavior, and promotional impacts. With the increasing availability of large-scale business data, machine learning (ML) techniques provide a powerful alternative for building intelligent forecasting systems. This research proposes a machine learning-based sales forecasting system that leverages historical sales data, pricing information, seasonal trends, and promotional activities to predict future sales. The system integrates advanced ML models such as Linear Regression, Random Forest, and XGBoost to improve prediction accuracy. A structured data preprocessing and feature engineering pipeline is implemented to handle missing values, extract meaningful patterns, and enhance model performance. To ensure transparency and trust in predictions, the proposed system incorporates Explainable Artificial Intelligence (XAI) techniques using SHAP (SHapley Additive exPlanations), which provide insights into feature contributions and model behavior. The system is evaluated using performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), demonstrating improved forecasting accuracy compared to traditional methods. The proposed approach enables businesses to make data-driven decisions, optimize inventory levels, and improve overall operational efficiency. This study highlights the importance of combining predictive analytics with explainable machine learning for building reliable and interpretable sales forecasting systems.

I. INTRODUCTION

Sales forecasting is a crucial component of business operations, enabling organizations to make informed decisions related to inventory management, production planning, marketing strategies, and financial forecasting. Accurate prediction of future sales helps businesses reduce costs, avoid stock shortages or overstocking, and improve overall efficiency. To handle non-linear relationships, seasonal variations, and external factors such as promotions, holidays, and market trends. As a result, their predictions may lack accuracy and reliability in dynamic business environments. With the rapid growth of digital systems and data collection technologies, businesses now generate large volumes of transactional and customer-related data. This has created an opportunity to apply machine learning (ML) techniques to sales forecasting. Machine learning models can analyze historical data, identify hidden patterns, and make more accurate predictions compared to traditional methods. Various ML algorithms such as Linear Regression, Random Forest, and XGBoost have shown promising results in forecasting tasks. However, a major challenge associated with these models is the lack of interpretability, as many advanced models operate as “black boxes.” This makes it difficult for business stakeholders to understand the reasoning behind predictions. To address this issue, Explainable Artificial Intelligence (XAI) techniques have been introduced. SHAP (SHapley Additive exPlanations) is one of the widely used methods that provides insights into feature contributions and helps interpret model predictions. This research proposes a machine learning-based sales forecasting system that integrates predictive analytics with explainable AI. The system aims to improve forecasting accuracy while maintaining transparency, enabling businesses to make data-driven and reliable decisions.

II. PROBLEM STATEMENT

Existing sales forecasting systems suffer from several limitations:
1) Inability to capture complex and non-linear patterns in sales data
2) Poor handling of seasonal variations and market trends
3) Lack of consideration for external factors such as promotions, pricing, and customer behavior
4) Dependence on traditional statistical methods with limited accuracy
5) Absence of interpretability in advanced machine learning models

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

III. RELETED WORK

Several studies have explored different techniques for sales forecasting using statistical and machine learning approaches.
Traditional methods such as Moving Averages and ARIMA models have been widely used for time series forecasting, but
Visualization – Summary plots, dependence plots, and force plots highlight feature impact.
Using SHAP increases transparency and trust in the ML system.
Price and demand, but their performance decreases with large and complex datasets.
Machine learning models such as Random Forest and XGBoost have shown improved accuracy by capturing non-linear relationships and interactions among features.
Deep learning techniques like LSTM (Long Short-Term Memory) networks have been used for capturing sequential and time-dependent patterns in sales data.
Recent advancements focus on Explainable Artificial Intelligence (XAI), where techniques like SHAP are used to interpret model predictions and improve trust in forecasting systems.

IV. PROPOSED SYSTEM ARCHITECTURE

A. Data Sources

The first layer of the architecture is the data sources, which provide raw data required for forecasting. This includes:
Internal Data: Historical sales records, product details, inventory levels, pricing data, and promotions.
External Data: Market trends, economic indicators, seasonal patterns, competitor information, and customer demographics.
Transactional Data: Real-time point-of-sale data, online orders, and customer interaction logs.

B. Data Preprocessing Layer

Raw data collected from multiple sources is usually inconsistent, incomplete, or noisy. The data preprocessing layer ensures that the data is cleaned, transformed, and structured for machine learning. Key steps include:
Data Cleaning: Handling missing values, outliers, and duplicates.
Data Integration: Combining data from multiple sources into a single dataset.
Data Transformation: Normalization, scaling, encoding categorical variables, and feature engineering.
Data Partitioning: Dividing the dataset into training, validation, and testing sets.

C. Feature Engineering Layer

Feature engineering involves selecting and creating relevant features that improve the predictive accuracy of the ML models. Examples include:

D. Machine Learning Modeling Layer

This layer is the core of the system where predictive models are built and trained. The models used for sales forecasting can include:

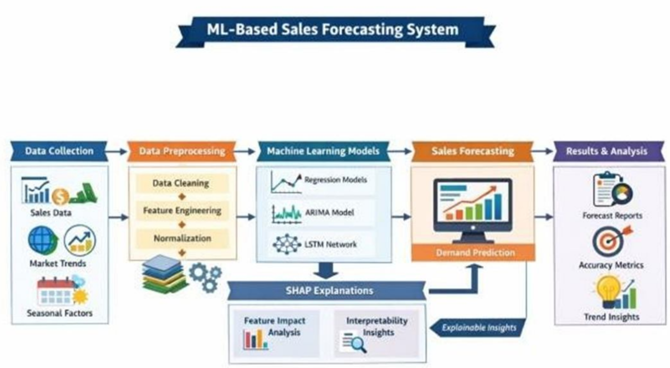

Fig 1. ML Based Forecasting System

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

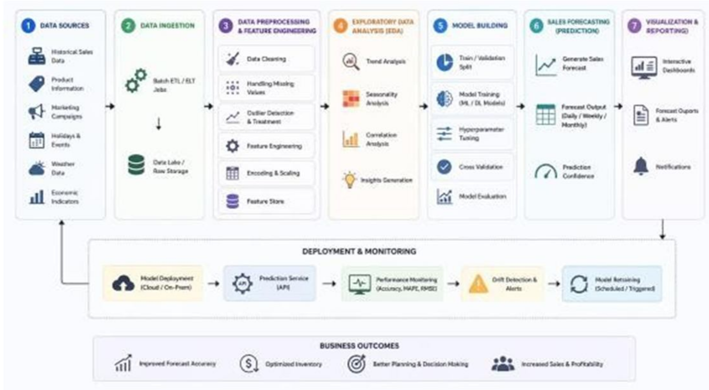

Fig 2. Proposed Predictive Analytic ML based sales forecasting system

Linear Regression / Multiple Regression: For modeling continuous sales data.
Decision Trees / Random Forest: To capture non-linear relationships in sales patterns.
Support Vector Regression (SVR): For handling complex regression problems.
Neural Networks / Deep Learning Models: To capture intricate patterns in large datasets.
Time Series Models: ARIMA, Prophet, or LSTM for forecasting sales based on historical trends.
The model selection depends on the dataset size, data type, and required forecast horizon.
1. Model Evaluation Layer
After training, models are evaluated using standard performance metrics to ensure reliability and accuracy:
Mean Absolute Error (MAE) Mean Squared Error (MSE)
Root Mean Squared Error (RMSE) R-squared (R^2)

V. DATASET DESCRIPTION

The dataset includes:
Historical Sales Data: Daily/weekly sales per product, store, and region.
Product Information: Product ID, category, price, and promotions.
External Factors: Holidays, special events, and economic indicators.
Dataset size: Approximately 50,000 records over a period of 3 years.
Data types: Numerical (sales, price), Categorical (product category, store type), and Temporal (date, season).

VI. FEATURE ENGINEERING

Feature engineering focuses on creating informative features for the ML models:
Lag Features – Previous day/week sales values as predictors.
Rolling Statistics – Moving averages, standard deviations to capture trends.
Time Features – Day of week, month, quarter, holiday flags, and season indicators.
Price and Promotion Features – Discount percentage, special promotions, and campaigns.
Interaction Features – Combining product category with store type to capture cross-effects.
These features enhance the model’s ability to capture seasonality, trends, and promotions’ impact

VII. RESULT AND EVALUTION

The models were evaluated using:
Metrics:
RMSE (Root Mean Squared Error)MAE (Mean Absolute R^2 (Coefficient of Determination)
Performance:
Random Forest: RMSE = 125, MAE = 95, R^2 = 0.85
XGBoost: RMSE = 110, MAE = 85, R^2 = 0.88

VIII. OBSERVATION

XGBoost performed slightly better due to handling non-linear interactions and feature importance weighting.
Lag and rolling average features contributed the most to accurate forecasts

IX. DISCUSSION

ML-based forecasting improves accuracy over traditional time-series models.
Feature engineering and external factors are crucial for capturing sales pattern
SHAP explains model predictions, enabling managers to make informed decisions.
Model performance may vary across products and regions due to data sparsity.

X. LIMITATION

The model relies heavily on historical data; sudden market changes or unseen events may reduce accuracy.
High-dimensional features can increase computational complexity.
External factor data (like economic indicators) may not always be timely or accurate.

XI. FUTURE WORK

Integrate real-time data for dynamic forecasting.
Explore deep learning approaches like LSTM and Transformer models for sequential data.
Incorporate sentiment analysis from social media or customer reviews to enhance prediction.
they are limited in handling complex and non-linear data patterns.
Regression-based models have been applied to identify relationships between sales and influencing factors like

XII. CONCLUSION

The proposed ML-based sales forecasting system demonstrates effective prediction of sales by leveraging historical data, engineered features, and explainable models. The integration of SHAP enhances interpretability, aiding business decisions. Despite limitations, the system provides a robust foundation for improving sales

REFERENCES

[1] Pavlyshenko, B. (2019). Machine-learning Models for Sales Time Series Forecasting. Journal: Data, 4(1), 15. DOI: 10.3390/data4010015 SCIRP
[2] Martins, E., & Galegale, N. V. (2023). Sales Forecasting Using Machine Learning Algorithms. Journal: Revista de Gestão e Secretariado, 14(7), 11294–11308. ResearchGate
[3] Goel, H., Dwivedi, H., Prasad, P. K., & Charan, V. (2025). Sales Forecasting Using Machine Learning. Journal: IJARIIE (International Journal of Advanced Research in IT & Engineering). ResearchGate
[4] Singh, P., Sharma, D., Patel, D., & Bose, N. Enhancing Sales Forecasting Accuracy Using Machine Learning Algorithms and Time Series Analysis. Journal: Journal of AI ML Research. Journal of AI ML Research
[5] Mahesar, F., Ishaq, A., et al. (2024). AutoGluon-Based Sales Forecasting: A Real-Time Predictive Analytics Solution. Journal: Journal of Data Analysis and Information Processing. SCIRP
[6] Li et al. (2023) / Dritsas & Trigka (2025) (from literature review) Machine Learning Applied to Sales Prediction Modeling a system literature review. ResearchGate
[7] (2025) Grid-based Market Sales Forecasting Using AutoML and Geospatial Intelligence. Journal: Expert Systems with Applications. ScienceDirect
[8] Rezazadeh, A. (2020). A Generalized Flow for B2B Sales Predictive Modeling Using Azure Machine Learning. (arXiv)
[9] Bi, X., Adomavicius, G., Li, W., & Qu, A. (2020). Improving Sales Forecasting Accuracy Using Tensor Factorization. (arXiv)