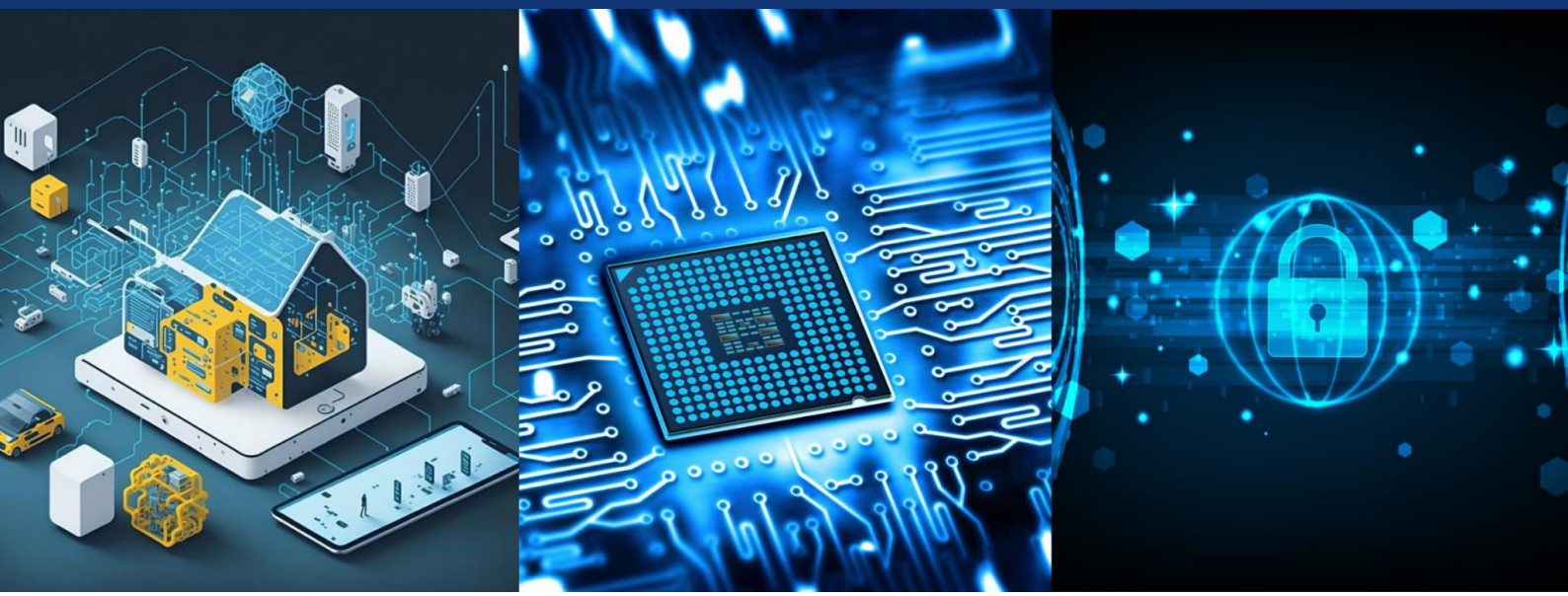
IJIRCCE

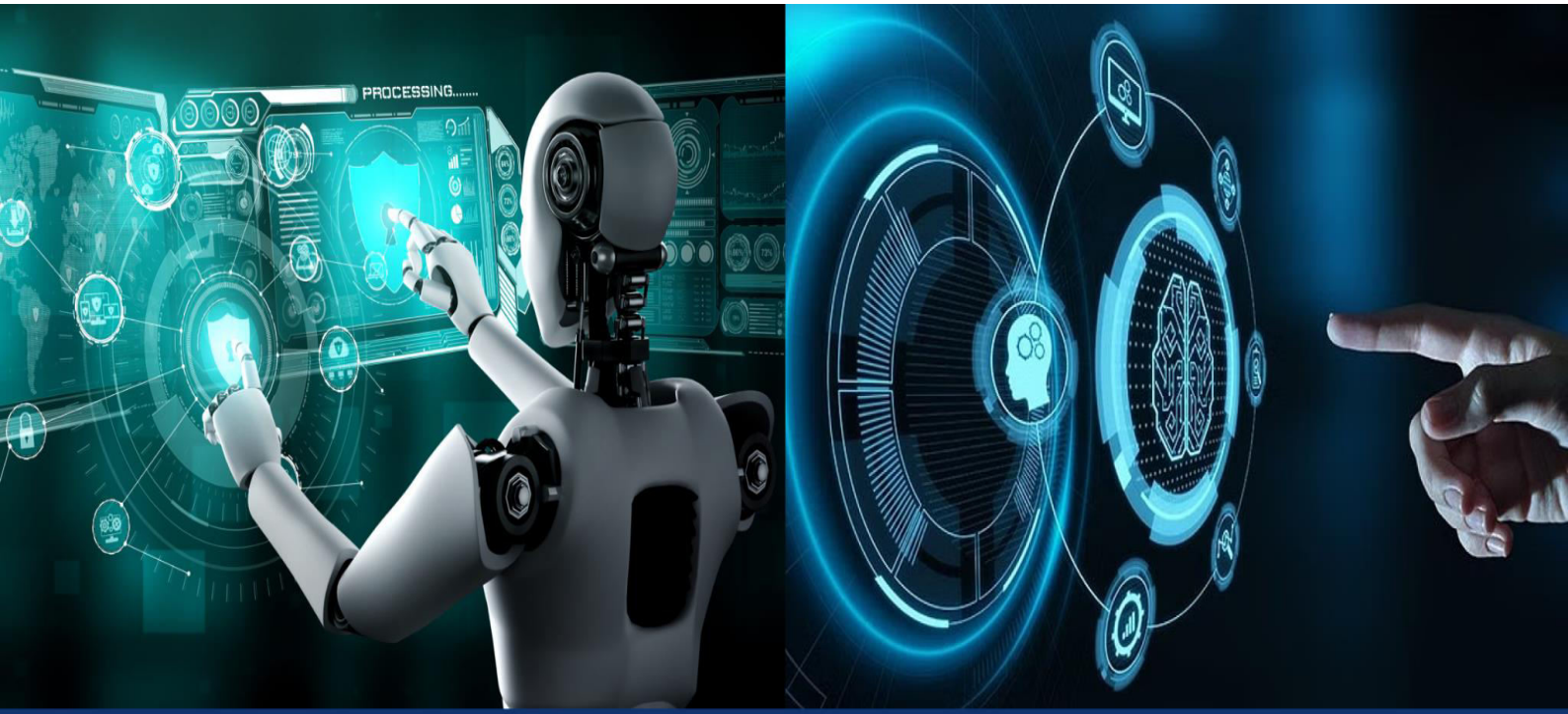
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

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



Impact Factor: 9.927

Volume 14, Issue 5, May 2026

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405051

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

Sales Forecasting using Machine Learning - Linear Regression & Random Forest Regressor

Kavya G D, Dr. puja Shashi

PG Student, Department of MCA, City Engineering College, Bengaluru, India

Professor & HOD, Department of MCA, City Engineering College, Bengaluru, India

ABSTRACT: By projecting future sales using prior data and market patterns, sales forecasting is necessary to corporate planning. Accurate forecasting facilitates better decision-making, resource optimisation, and inventory management for businesses. Large datasets and intricate sales patterns are often challenging for conventional forecasting methods to manage, which results in fewer accurate forecasts. An approach based on machine learning sales forecasting system utilising Random Forest Regressor and Linear Regression methods is displayed in this study. While Random Forest Regressor enhances prediction accuracy by managing non-linear connections and minimising overfitting, Linear Regression is used to examine the connection between the input variables and sales numbers. The system includes feature selection, data collecting, preprocessing, model training, and evaluation. Measures like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score are used to gauge performance. The system determines Which model is more effective for accurate sales prediction by comparing the two methods. The suggested method assists companies in enhancing marketing tactics, financial planning, and inventory control, ultimately boosting productivity and profitability.

KEYWORDS: Demand forecasting, Machine learning, Retail analytics, Sales forecasting, and Inventory management.

I. INTRODUCTION

The technique of projecting future sales using market trends, client demand, past sales data, and other contributing factors is known as sales forecasting. It is crucial in assisting companies in making wise choices about budgeting, production scheduling, inventory control, and marketing tactics. Companies can enhance a company's overall performance and minimise losses from overstocking or understocking products by using accurate sales projections.

When working with huge and complicated datasets, traditional sales forecasting approaches might not produce precise results because they primarily rely on statistical techniques and manual analysis. Machine learning has become a successful method for more accurately forecasting future sales because of the growth of data in contemporary firms. Large data sets can be analysed by machine learning algorithms, which can also find hidden patterns and produce accurate forecasts.

The project's objective is to forecast sales utilising the Random Forest Regressor and Linear Regression methods. While Random Forest Regressor manages intricate patterns and enhances prediction performance, Linear Regression aids in comprehending the relationship between input variables and sales figures. To produce precise sales forecasts, the system incorporates preparing data, training models, and testing, and performance evaluation. This method helps companies increase profitability and make better strategic decisions.

II. LITERATURE SURVEY

1.**Title**: Machine Learning-Based Forecasting Sales with Predictive Analytics in IT Organisations
(2026)
Authors: Ahmed et al
Abstract: Using machine learning algorithms to increase sales prediction accuracy in IT companies is the main topic of this study. It emphasises how businesses may make better judgements by using predictive analytics.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

2.**Title**: A Systematic An analysis of the machine learning literature in Sales Prediction Modelling (2026)
Authors: Saptadi et al many models of machine learning for prediction and discusses their advantages for handling complicated sales data.

3. **Title**: Time series analysis of retail sales and machine learning models(2026)
Authors: Dar et al
Abstract: The study combines traditional time-series methodology with machine learning techniques to increase forecasting efficacy in retail industries.

4. **Title**: Forecasting Retail Sales with Models for Machine Learning (2025)
Authors: Mustapha and Sithole
Abstract: This study contrasts many algorithms for machine learning for retail sales forecasting in an effort to boost company profitability.

5.**Title**: Retail Sales Forecasting using sophisticated machine learning algorithms(2024)
Authors: Malik et al.
Abstract: In order to forecast sales patterns in this study explores advanced algorithms such as XGBoost and ARIMA in retail settings.

III. METHODOLOGY

EXISTING PROBLEM

The majority of traditional methods for sales forecasting rely on human calculations and crude statistical techniques, which can produce inaccurate results. These methods struggle to manage large amounts of sales data from multiple sources. They can't identify complex patterns like seasonal trends and consumer behaviour. Inaccurate forecasting may lead to either overstocking or understocking of products. Poor decision-making can result in monetary losses for businesses.

PROPOSED SOLUTION

To get around the drawbacks of conventional forecasting techniques, this study proposes a machine learning-based sales forecasting system that makes use of Random Forest Regressor and Linear Regression. The system collects historical sales data and uses preprocessing techniques such data purification, feature selection, and handling missing figures. Linear regression is used to find correlations between variables, while Random Forest regression analyses intricate patterns in the information to improve prediction accuracy. Performance indicators are used to evaluate the training models in order to choose the optimal forecasting model. For businesses, This technology makes it possible for more accurate forecasting, improved inventory management, and better decision-making.

IV. SYSTEM DESIGN

The Superstore dataset is used to collect historical sales information as the initial stage in the system design process. Following collection, the data is pre-processed to remove inconsistencies, duplicates, and missing values. Following cleaning, key elements that affect sales prediction are found via feature extraction. Following the division of the dataset into testing and training data, The model is built using machine learning methods. Finally, but just as importantly, the system produces precise sales forecasting predictions that assist companies in making better decisions and strategies.

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405051

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

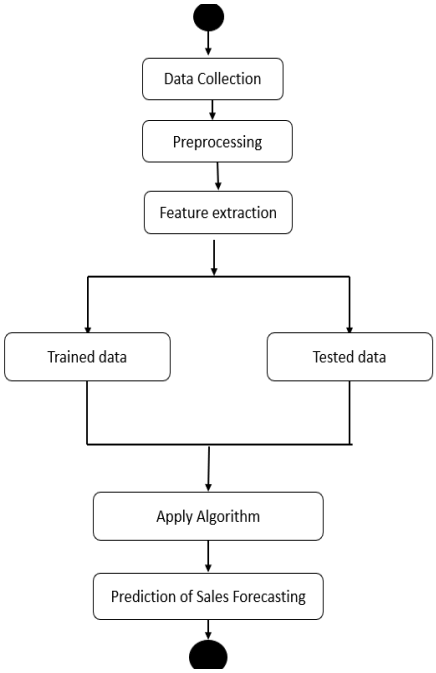


FIG 1: System Design of Sales Forecasting

V. SYSTEM ARCHITECTURE DESIGN

To generate accurate sales forecasts, the several stages of the System Architecture Design for Forecasting Sales with Machine Learning work together. The Superstore sales dataset is collected as input at the start of the process. The data is then preprocessed to remove missing values, duplicates, and inconsistent information. Feature extraction and engineering are then used to determine crucial factors influencing revenue. After the processed data is separated into training and testing datasets, Machine learning techniques are used. In the endThe model is assessed and generates sales forecasting results that help businesses improve inventory control and decision making.

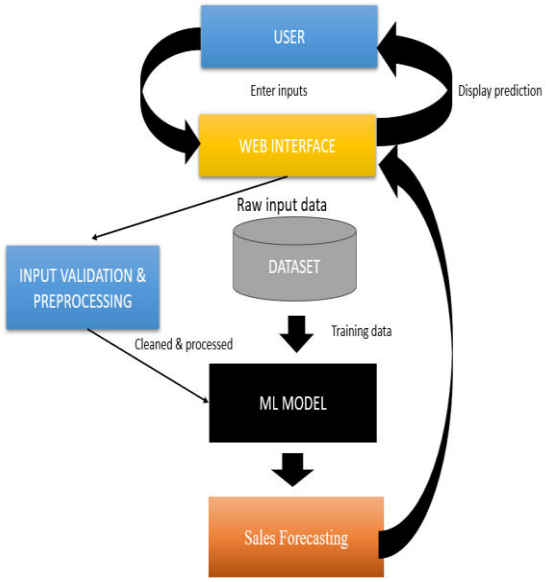


FIG 2 : System Architecture desgin of sales forcasting.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

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

VI. IMPLEMENTATION

Python was utilised in the creation of machine learning-based sales forecasting project and data science technologies including Pandas, NumPy, Matplotlib, and Scikit-learn. The Superstore Sales Dataset was first loaded and pre-processed in order to handle missing values, remove duplicate records, and convert category data into numerical form. Exploratory data analysis, or EDA was used to identify patterns and trends in sales. Feature engineering was hired to produce meaningful features for prediction. Following the division of the dataset, the Random Forest Regressor and Linear Regression models were trained. GridSearchCV was utilised for hyperparameter tuning in order to improve accuracy. Following the models' evaluation using MAE, RMSE, and R^2 score. The optimal model was chosen to forecast sales.

VII. RESULTS AND DISCUSSION

The project's findings demonstrate that both Random Forest Regressor and Linear Regression were effectively trained to forecast sales values using the Superstore Sales Dataset. The models' performance was assessed using the R2 score, RMSE, and MAE. Due of its inability to handle intricate sales patterns, linear regression produced basic forecast findings but demonstrated lesser accuracy.

However, by generating lower error levels and a higher R2 score, the Random Forest Regressor outperformed the others. This suggests that it captured non-linear correlations in the sales data more well. Additionally, characteristics including Profit, Quantity, and Order Month had a significant influence on sales prediction, according to feature importance analysis.

VIII. CONCLUSION

By employing Random Forest Regressor and Linear Regression models on the Superstore Sales Dataset, this research effectively created a machine learning-based system for forecasting sales data data, feature engineering, analysis, and model training preparation, and assessment were all components of the project. Random Forest Regressor outperformed Linear Regression Regarding The precision of predictions between the two models. The findings aid companies in making better-informed pricing decisions, sales strategy, and inventory control. This research illustrates how machine learning may enhance forecasting precision and promote company expansion.

REFERENCES

1. Ahmed, M., "Predictive Analytics for Sales Forecasting in IT Enterprises Using Machine Learning Techniques," pp. 45–52, 2026.
2. "Machine Learning Applied to Sales Prediction Modelling: A Systematic Literature Review," Saptadi, R., Kumar, P., and Lee, J. Journal of Artificial Intelligence Research, vol. 18, no. 1, pp. 78–89, 2026.
3. "Integration of Time Series and Retail Machine Learning Models Sales Forecasting," Journal of Intelligent Systems International, pp. 112–120, 2026, Dar, S., Sharma, R., and Patel, K.
4. " Using Predictive Machine Learning Models Retail Sales," International Journal of Business Analytics, pp. 56–64, 2025, Mustapha, A., and Sithole, T.
5. "Sales Forecasting in Retail Using Advanced Machine Learning Algorithms," Malik, H., Verma, S., and Rao, P., Journal of Computer-Based Intelligence, pp. 34–41, 2024.
6. "Retail Sales Forecasting Using Deep Learning: Systematic Literature Review," Eglite, K., and Birzniece, I. Procedia Computer Science, vol. 200, pp. 150–158, 2022.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details