IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 6, June 2025

Enhancing Mart Sales Using ML Techniques

M. Ajay Kumar[1], A. Rutwik[2], B. Laxman[3], B. Dhashvanth Sai[4]

Assistant Professor, Department of Computer Science & Engineering[1]
Students, Department of Computer Science & Engineering[2,3,4]
Guru Nanak Institute of Technology, Hyderabad

Abstract*: Accurate sales forecasting is pivotal for optimizing inventory management and securing business profitability in the highly competitive retail sector. This study aims to enhance sales prediction for Big Mart by leveraging Decision Tree Regression, a machine learning approach adept at capturing complex, non-linear relationships within diverse datasets. The proposed methodology analyses extensive historical sales records in tandem with critical features—such as product categories, store attributes, and promotional strategies—to unearth underlying patterns and trends that drive sales performance. By rigorously preprocessing and engineering these features, the model is calibrated to provide precise future sales predictions, which in turn facilitate impactful insights for inventory optimization. These insights enable retailers to minimize inventory wastage, avoid stockouts, and calibrate pricing and promotional strategies, thereby supporting informed, data-driven decision-making. Moreover, the integrated approach underscores the importance of building robust predictive models that not only adapt to historical trends but also dynamically respond to changing market conditions, ensuring improved operational efficiency and a sustained competitive advantage in a dynamic retail market.*

Keywords: inventory management

I. INTRODUCTION

In the highly competitive retail industry, accurate sales forecasting is essential for optimizing inventory management, reducing wastage, and ensuring product availability. Effective prediction models enable businesses to make informed decisions that enhance operational efficiency and maximize profitability. Traditional statistical methods often struggle to capture complex, non-linear relationships in sales data, making machine learning approaches increasingly popular for predictive analytics. This study explores the application of Decision Tree Regression, a robust tree-based algorithm, to enhance sales forecasting for Big Mart. By analysing historical sales data and key influencing factors—such as product categories, store attributes, and promotional strategies—the model uncovers patterns that drive accurate future predictions. Decision Tree Regression provides interpretable outcomes, making it an ideal choice for data-driven decision-making. The insights generated from this approach help in reducing stockouts, minimizing excess inventory, and refining strategic planning, ultimately supporting Big Mart's ability to maintain competitiveness in the dynamic retail landscape.

II. LITERATURE REVIEW

Jil P. Bhatti, Janak H. Maru, & Devangi R. Paneri (2024) : The paper titled ″A Review of Machine Learning Techniques for Retail Sales Prediction: Deep Learning vs. Traditional Methods″ analyses ten different machine learning approaches for sales forecasting, comparing traditional statistical models with deep learning techniques. The authors emphasize that deep learning models such as CNN and LSTM outperform conventional methods in terms of accuracy and computational efficiency. Their study underscores the importance of leveraging AI-driven forecasting techniques to improve business decision-making. The paper discusses how CNNs can extract spatial patterns in sales data, while LSTMs are particularly useful for capturing temporal dependencies, making them ideal for time-series forecasting. The authors also highlight the challenges of implementing deep learning models, such as the need for large datasets and computational resources, and propose strategies to mitigate these limitations.

ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Impact Factor: 5.731
Volume 5, Issue 6, June 2025

Rashmi Amardeep et al. (2024) : The paper titled ″Integrating LSTM, ARIMA, and XGBoost for Sales Forecasting and Inventory Optimization″ focuses on machine learning-based sales forecasting and inventory optimization. The study integrates LSTM, ARIMA, and XGBoost to identify patterns and correlations within historical sales data. The authors highlight how these models help businesses optimize inventory management, devise targeted marketing strategies, and allocate resources efficiently. Their findings suggest that combining multiple machine learning techniques enhances forecasting accuracy. The study provides a comparative analysis of different forecasting models, demonstrating that LSTM excels in capturing long-term dependencies, ARIMA is effective for short-term trend analysis, and XGBoost offers high predictive power by leveraging ensemble learning. The authors also discuss practical applications of their approach, such as demand forecasting for seasonal products and dynamic pricing strategies.

Hitesh S.M, Yukthi A, & Prof. Ramya B.N (2024) : The paper titled ″Sales Prediction Using Machine Learning Techniques″ examines how machine learning models enhance sales forecasting accuracy by leveraging advanced data mining techniques. The study identifies the Gradient Boost Algorithm as the most effective model for predicting sales trends, outperforming traditional statistical methods in handling complex, non-linear relationships in sales data. The authors emphasize the importance of predictive analytics in optimizing business operations, particularly for E-commerce industries, which face challenges due to increasing data volume and evolving consumer behavior. By integrating decision analysis and predictive modeling, the research demonstrates how machine learning-driven forecasting improves revenue generation and operational efficiency. The study also proposes a performance evaluation framework to assess different models, ensuring businesses adopt the most suitable predictive approach for inventory management, pricing strategies, and market trend analysis. Additionally, the adaptability of these models for real-time monitoring applications highlights their relevance in dynamic retail environments, helping businesses prevent stockouts, reduce surplus inventory, and enhance overall profitability through data-driven decision-making.

Riddhi J. Kotak & Rajnish Rakholia (2024) : The paper titled ″A Comprehensive Study on Different Machine Learning Approaches for Retail Sales Forecasting: Methods, Procedures, Obstacles, and Prospects″ provides an extensive analysis of various machine learning techniques applied to retail sales forecasting. The study emphasizes the importance of accurate sales predictions for inventory optimization, resource allocation, and operational efficiency in retail businesses. Traditional forecasting methods often struggle to capture the complex and dynamic patterns in retail sales data, leading to inefficiencies in stock management and revenue planning. The authors investigate a diverse range of machine learning models, including statistical techniques, univariate and multivariate time-series forecasting models, deep learning algorithms, regression methods, and neural networks. Special attention is given to Decision Tree Regression, Light GBM, and XGBoost, which have demonstrated superior performance in handling non-linear relationships and large-scale retail datasets. The study also discusses the computational challenges associated with machine learning models, such as memory consumption and processing efficiency, and proposes strategies to optimize forecasting accuracy while minimizing resource usage.

S. Ravi Teja Reddy & P. Malathi (2022) : The paper titled ″Comparative Analysis of Decision Tree Regression and Linear Regression for Sales Forecasting″ compares Decision Tree Regression with Linear Regression for sales forecasting using Big Mart sales data. The researchers conducted experiments to determine the best algorithm for predicting future sales trends. Their findings revealed that Decision Tree Regression achieved 97.5% accuracy, significantly outperforming Linear Regression, which had an accuracy of 87.6%. The study highlights the advantages of tree-based models in handling complex, non-linear relationships in sales data, making them a preferred choice for retail forecasting. The authors emphasize that Decision Tree Regression is particularly effective in capturing interactions between multiple variables, such as product type, store attributes, and promotional efforts, leading to more precise predictions. Additionally, the study discusses how businesses can leverage these insights to optimize inventory management, reduce wastage, and improve profitability.

Existing System

The Retail Real-Time Sales Prediction System Using LSTM and XGBoost is an advanced machine learning framework designed to enhance sales forecasting accuracy in dynamic retail environments. By integrating Long Short-Term Memory (LSTM) networks for time-series forecasting and XGBoost for predictive analytics, the system effectively captures both temporal dependencies and non-linear relationships within sales data. Unlike traditional forecasting models, which rely solely on historical trends, this system leverages real-time data ingestion through seamless Point-of-Sale (POS) system integration, ensuring adaptability to market fluctuations. A Flask-based backend enables efficient data processing, while an interactive React dashboard allows intuitive visualization of predicted trends. The predictive models analyse key factors such as seasonal variations, promotions, inventory levels, and consumer behaviour, facilitating more informed decision-making. By enhancing forecasting reliability, this system empowers retailers to minimize stockouts, optimize inventory, reduce financial risks, and improve customer satisfaction, ultimately driving higher profitability and operational efficiency. Future enhancements may include incorporating deep reinforcement learning techniques and automated hyperparameter tuning to further refine prediction accuracy, making the system a cornerstone for data-driven retail strategies.

Existing System Disadvantages:

- High Computational Cost
- Complex Implementation
- Data Dependency
- Limited Interpretability
- Scalability Issues

Proposed System

The proposed system leverages Decision Tree Regression to enhance sales forecasting for retail environments such as Big Mart by utilizing historical sales data and key features—including product type, outlet attributes, promotional efforts, and seasonal variations—to accurately predict future sales trends. This system integrates modules for data collection, preprocessing, and feature engineering to ensure the quality and relevance of the inputs, followed by a machine learning model that effectively captures complex, non-linear relationships inherent in sales patterns. By automatically ranking feature importance, the model aids in identifying the most critical factors influencing sales, while its inherent interpretability allows business stakeholders to easily understand and trust the resultant predictions. Additionally, the system is designed for scalability and fast training, enabling real-time updates and adjustments in dynamic retail scenarios. Ultimately, this Decision Tree Regression-based approach not only enhances prediction accuracy but also supports data-driven decision-making for optimized inventory management, strategic pricing, and improved overall profitability in competitive retail markets.

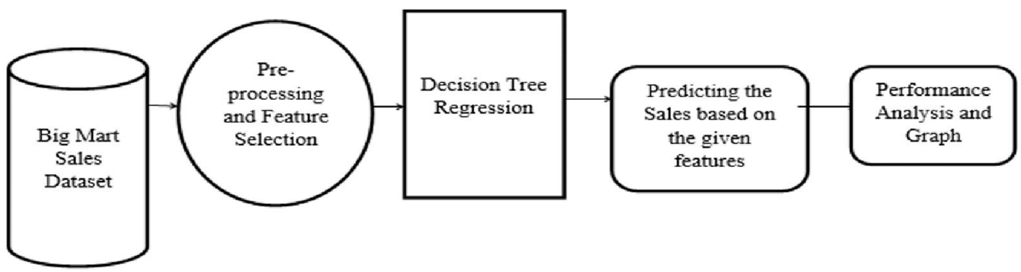
Proposed System Advantages:

- Ability to Capture Non-linear Relationships
- Interpretability and Transparency
- Scalability and Efficient Training
- Real-time Data Adaptability
- Enhanced Inventory and Decision Support
- High Prediction Accuracy

IV. SYSTEM ARCHITECTURE

The system architecture for a Decision Tree Regression-based sales forecasting system is designed as a modular and layered framework that efficiently integrates data ingestion, preprocessing, model training, prediction, and evaluation.

At the foundation, raw sales data—collected from historical records, inventory logs, POS systems, and promotional datasets—is fed into a data ingestion layer where it undergoes thorough cleaning and standardization. This preprocessing stage is critical; it handles missing values, normalizes features, and applies feature selection techniques to extract the most relevant predictors. The refined data is then passed to the core predictive engine, where a Decision Tree Regression model is trained. This model incrementally partitions the dataset by selecting optimal split points based on minimizing error metrics such as Mean Squared Error, effectively capturing non-linear relationships inherent in the sales trends. The trained model is subsequently deployed within a prediction module that processes new data inputs to generate real-time sales forecasts. Additionally, an evaluation layer monitors model performance using statistical metrics, ensuring the predictions maintain high accuracy and reliability over time. This layered architectural design not only enhances computational efficiency and scalability—vital for handling large volumes of retail data—but also offers an interpretable structure that informs stakeholders of the influential factors driving sales, thereby supporting robust data-driven decision-making in inventory management and strategic planning.



V. METHODOLOGY

MODULES:

Data Collection:

The system begins by gathering raw data from diverse sources—historical sales records, point-of-sale (POS) transactions, inventory logs, and promotional details. This comprehensive data acquisition ensures that all facets of retail operations are captured, providing a rich foundation to recognize seasonal trends, regional variances, and the impacts of promotional events.

Preprocessing:

Once collected, the raw data undergoes rigorous cleaning and standardization. In this phase, missing information is addressed, inconsistencies are corrected, and all variables are normalized. This preparation is critical to minimize noise and ensure that the subsequent analysis and predictions are built on a reliable, uniform dataset.

Feature Extraction:

The system then focuses on extracting and selecting key predictors that directly influence sales outcomes. Important variables—such as product types, outlet characteristics, promotional activities, and seasonal indicators—are identified and engineered. By isolating these critical factors, the predictive power of the sales forecasting approach is significantly enhanced.

Model Training and Learning :

With the refined dataset in place, the system applies a Decision Tree Regression algorithm to learn from historical trends. During training, the algorithm iteratively splits the data based on optimal decision thresholds, capturing both non-linear relationships and intricate interactions among features. This learning process enables the system to build a robust predictive model that mirrors the complexities of real-world sales dynamics.

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 5, Issue 6, June 2025
Impact Factor: 5.731

Forecast Generation :

Once the model structure is ready, it is compiled by defining the optimizer, loss function, and evaluation metrics. The Adam optimizer is chosen for its efficiency and adaptability. Since the task involves classifying data into multiple categories, categorical cross-entropy is used as the loss function.

Performance Evaluation :

Training the model involves feeding the preprocessed and feature-extracted data into the model and allowing it to learn from the labeled examples. The model will adjust its internal weights using backpropagation to minimize the loss function. The dataset is typically divided into training and validation sets to assess the model's performance on unseen data. Monitoring the loss and accuracy during training helps ensure the model is learning effectively and not overfitting.

VI. RESULT

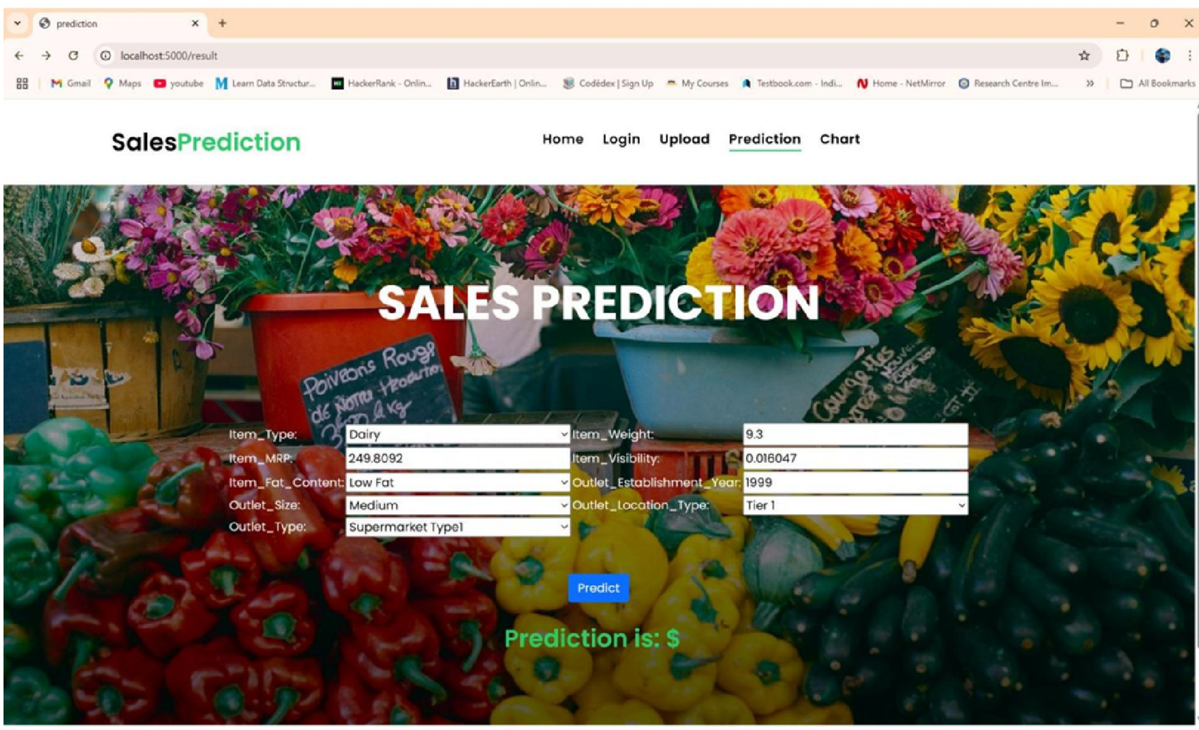


Figure 1.1: Enter the required inputs needed for the prediction of sales.

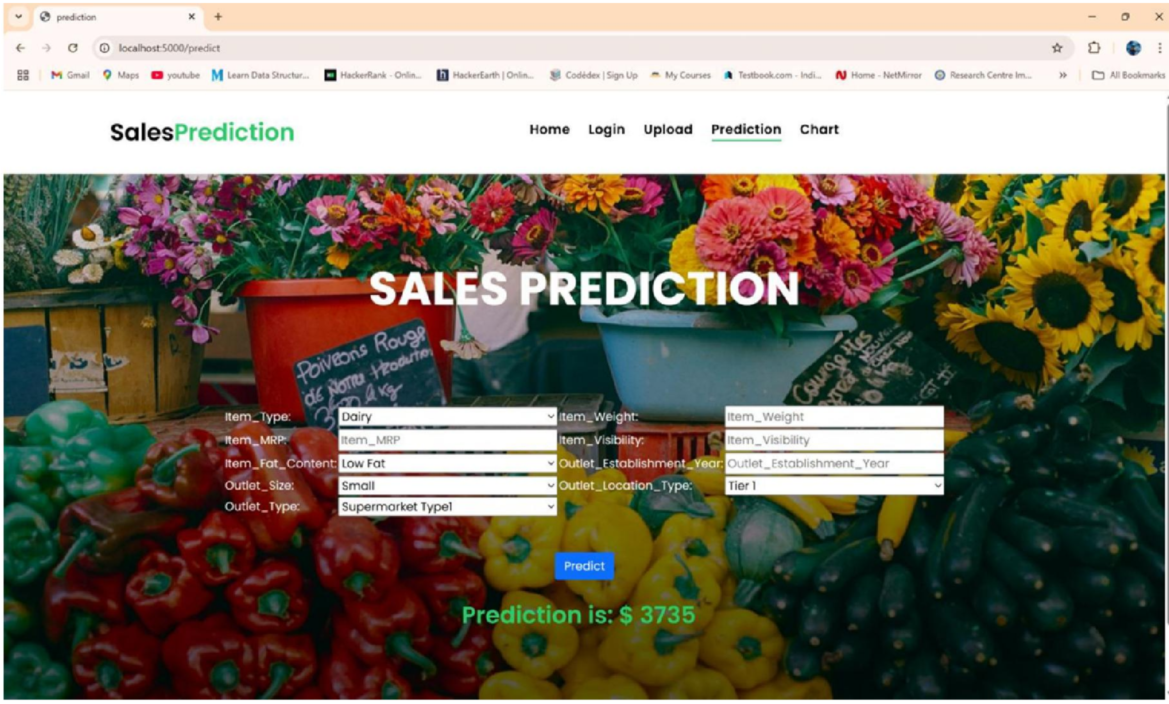


Figure 1.2: The sales for the product is predicted

VI. CONCLUSION

Retail businesses face increasing pressure from dynamic market conditions and growing competition, making advanced forecasting methods essential for success. This study illustrates how Decision Tree Regression markedly improves mart sales forecasting, offering an interpretable and efficient predictive tool for retail operations. By harnessing machine learning techniques, the proposed approach not only facilitates data-driven decision-making but also enhances profitability through optimized inventory management. The model's transparent structure enables stakeholders to identify the most influential sales drivers, thereby empowering strategic business planning. For future research, integrating ensemble methods such as Random Forests and XGBoost, incorporating real-time data streams, and refining predictive models will be crucial for improving adaptability and robustness in rapidly changing retail environments.

REFERENCES

[1] Ching Wu Chu and Guoqiang Peter Zhang, “A comparative study of linear and nonlinear models for aggregate retails sales forecasting”, Int. Journal Production Economics, vol. 86, pp. 217- 231, 2003.

[2] Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.- 2.

[3] Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." Journal of Soft Computing Paradigm (JSCP) 2, no. 02 (2020): 101- 110

[4] Giuseppe Nunnari, Valeria Nunnari, “Forecasting Monthly Sales Retail Time Series: A Case Study”, Proc. of IEEE Conf. on Business Informatics (CBI), July 2017.

[5] https://halobi.com/blog/sales-forecasting-five-uses/.

[6] Zone-Ching Lin, Wen-Jang Wu, “Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone”, IEEE Trans. On Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[7] O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, “Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis”, Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

[8] C. Saunders, A. Gammerman and V. Vovk, "Ridge Regression Learning Algorithm in Dual Variables", Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.
[9] "Robust Regression and Lasso". Huan Xu, Constantine Caramanis, Member, IEEE, and Shie Mannor, Senior Member, IEEE. 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration."An improved Adaboost algorithm based on uncertain functions".Shu Xinqing School of Automation Wuhan University of Technology.Wuhan, China Wang Pan School of the Automation Wuhan University of Technology Wuhan, China.
[10] Xinqing Shu, Pan Wang, "An Improved Adaboost Algorithm based on Uncertain Functions", Proc. of Int. Conf. on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.
[11] A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding the past", Addison-Wesley, 1994.
[12] N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321-335P
[13] D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.
[14] X. Yua, Z. Qi, Y. Zhao, Support Vector Regression for Newspaper/Magazine Sales Forecasting, Procedia Computer Science 17 (2013) 1055–1062.
[15] E. Hadavandi, H. Shavandi, A. Ghanbari, An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: a Case study of the printed circuit board, Expert Systems with Applications 38 (2011) 9392–9399