iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** V **Month of publication:** May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62603

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms

Sakshi Bhujbal[1], Vaishnavi Rajure[2], Saurabh Shinde[3], Divyansh Singh[4], Yash Bagul[5], Dr. Pankaj Agarkar[6]
Department of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Lohegoan, Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: ***Supermarkets and retail giants like Big Mart rely heavily on tracking individual item sales data to anticipate consumer demand and optimize inventory management. By mining data warehouses, anomalies, and trends can be identified, providing valuable insights for forecasting future sales volumes. Leveraging advanced machine learning techniques, such as Decision Tree Regression, predictive models can be developed to forecast sales accurately. Through the implementation of such a model, it has been observed that it surpasses the performance of existing forecasting models. This innovative approach not only enhances decision-making processes but also enables businesses like Big Mart to stay ahead in a competitive market landscape.***
Technical Keywords: Decision Tree Algorithm, Predictive Analysis, Model Evaluation, Machine learning, Sales Forecasting.

I. INTRODUCTION

The competition between shopping centers and large retail stores has intensified significantly due to the rapid global expansion of malls and the surge in online shopping. In this fiercely competitive market, businesses are striving to attract customers with personalized and time-sensitive deals. This approach is essential not only for expanding customer bases but also for efficiently managing inventory levels and optimizing logistical services based on predicted sales volumes for each item.

Machine learning algorithms have become essential in enhancing sales forecasting capabilities. These algorithms provide valuable methods for predicting sales across various types of organizations, particularly in addressing the limitations of traditional statistical forecasting methods. Accurate sales predictions are critical for developing and refining marketing strategies in this competitive marketplace. Furthermore, these predictions are indispensable for implementing measurable methodologies such as regression analysis, ARIMA (Auto-Regressive Integrated Moving Average), and ARMA (Auto-Regressive Moving Average).

II. PROBLEM STATEMENT

Our project addresses the optimization challenges encountered by Big Mart, encompassing issues such as low sales, understanding customer behavior, and the imperative to enhance efficiency. Through predictive analysis methodologies, we aim to construct precise sales forecasting models, conduct thorough customer behavior analysis, and deploy advanced analytics tools to streamline operational processes. By leveraging cutting-edge data analytics and machine learning algorithms, our objective is to equip Big Mart with the capability to anticipate market demand, tailor marketing strategies to individual customer preferences, and optimize decision-making processes. This comprehensive approach is geared towards catalyzing improved sales performance and bolstering profitability for Big Mart, ensuring sustainable growth and competitiveness in the retail landscape.

III. OBJECTIVES

Develop Accurate Sales Forecasting Models: Construct robust predictive models utilizing machine learning algorithms to precisely forecast sales, empowering Big Mart to anticipate market demand fluctuations and optimize inventory management effectively.Understand Customer Behavior: Conduct in-depth analysis of customer data to glean insights into purchasing behaviors, preferences, and patterns. This understanding will enable Big Mart to tailor marketing strategies and product offerings, fostering enhanced customer satisfaction and loyalty. Optimize Operational Efficiency: Implement predictive analytics tools to streamline operational processes, including inventory management, pricing strategies, and promotional activities. By optimizing resource allocation and reducing costs, Big Mart can enhance operational efficiency. Improve Decision-Making: Provide actionable insights derived from predictive analysis to support data-driven decision-making across various departments within Big Mart. This approach will facilitate strategic planning and operational enhancements.

©IJRASET: All Rights are Reserved

Enhance Sales Performance: Utilize predictive analysis to identify growth opportunities, target high-potential market segments, and optimize sales strategies. Ultimately, this will lead to increased sales revenue and profitability for Big Mart.
Ensure Scalability and Sustainability: Develop the predictive analysis framework with scalability and sustainability in mind, enabling Big Mart to adapt to changing market dynamics and maintain a competitive edge over time.

IV. LITERATURE REVIEW

The following papers contribute valuable insights to the realm of predictive analysis and sales forecasting, each offering unique perspectives and methodologies. Firstly, "Predictive Modeling and Forecasting of Big Data for E-commerce: eBay Sales Volume and Product Popularity" by D. Yang explores predictive modeling techniques within the e-commerce domain, particularly focusing on forecasting sales volume and product popularity. Secondly, "Improving Demand Forecasting in a Food Retail Supply Chain" authored by P. L. Mendoza discusses the application of machine learning algorithms to enhance demand forecasting in the food retail sector, which is pertinent to sales prediction. Additionally, "A Comprehensive Survey of Machine Learning Algorithms for Big Data Analytics" by L. Sun provides a broad overview of machine learning algorithms, beneficial for understanding techniques applicable to sales prediction. "Predictive Analytics for Retail Sales using Classification and Regression Trees" by V. S. Singh investigates the efficacy and interpretability of decision tree-based algorithms in retail sales prediction. Moreover, "Predictive Modeling of Retail Customer Purchase Behavior" by A. M. Ahmed delves into predictive modeling of customer purchase behavior in retail, closely linked to sales prediction. Finally, "Machine Learning for Retail Demand Forecasting: A Review of the State-of-the-Art" authored by O. Fildes presents a comprehensive review of machine learning techniques for demand forecasting in retail, offering valuable insights applicable to sales prediction methodologies, particularly within the Big Mart context.

V. METHODOLOGY

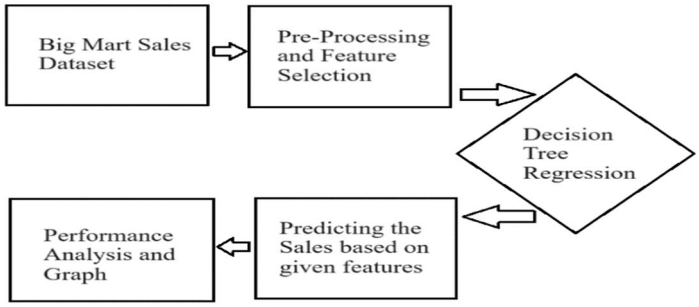
1) *Data Collection:* The initial phase in developing a machine learning model involves data collection. Acquiring comprehensive and high-quality data is pivotal, as it significantly influences the model's performance. Various methods such as web scraping and manual interventions can be employed to gather data. For our predictive analysis of Big Mart sales, we obtained the dataset from Kaggle, containing detailed information on product attributes and sales records.
2) *Data Preparation:* The collected data undergoes preprocessing to ensure it is suitable for model training. This involves tasks like removing duplicates, handling missing values, normalizing numerical features, and converting data types as necessary. Additionally, randomizing the data eliminates biases introduced during collection or preparation. Exploratory data analysis techniques, including visualization, help uncover relationships between variables and identify potential class imbalances.
3) *Model Selection:* For our predictive analysis, we opted for the decision tree regression machine learning algorithm due to its ability to handle both continuous and categorical output variables. Decision trees are particularly useful for their interpretability and flexibility. After training the model, we achieved a satisfactory accuracy of 95.7% on the test set, validating its effectiveness for our task.
4) *Analysis and Prediction:* From the dataset, we selected nine relevant features for analysis, including product weight, fat content, visibility, type, maximum retail price, outlet establishment year, size, location type, and outlet type. Using these features, we conducted predictive analysis to forecast sales trends and patterns within Big Mart stores.
5) *Accuracy on Test Set:* The trained decision tree regression model demonstrated a commendable accuracy of 95.80% on the test set, indicating its reliability in making accurate predictions.
6) *Saving the Trained Model:* To deploy the model into a production environment, it is crucial to save it in a suitable format such as .h5 or .pkl. Leveraging libraries like pickle, we saved the trained and tested model for seamless integration into the production-ready environment, ensuring its availability for real-time predictions while maintaining its integrity.

By following this process meticulously, we conducted predictive analysis for Big Mart sales, leveraging machine learning algorithms to enhance decision-making and optimize sales strategies, all while ensuring the integrity and reliability of the developed model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

VI. SYSTEM ARCHITECTURE



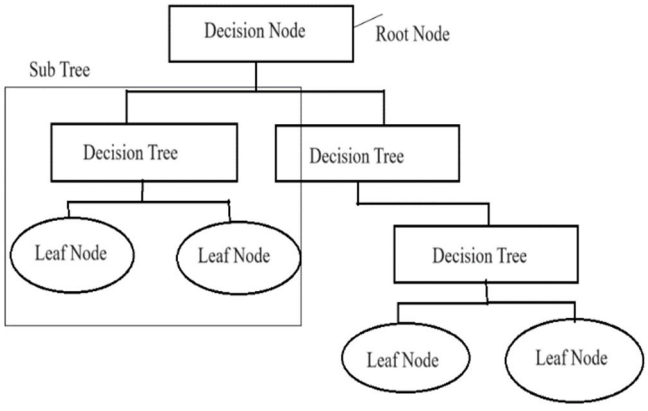
1) *Big Mart Sales Dataset:* The initial step involves acquiring the Big Mart sales dataset, serving as the foundational data for model development. Pre-Processing and Feature Selection: The dataset undergoes pre-processing, including tasks like cleaning, handling missing values, and normalization. Furthermore, feature selection techniques are applied to identify pertinent attributes for sales prediction.
2) *Decision Tree Regression:* Central to the process is the implementation of Decision Tree Regression algorithm. This algorithm utilizes the pre-processed data and selected features to construct a predictive model capable of forecasting sales based on provided input features.
3) *Predicting the Sales based on given features:* Once the Decision Tree Regression model is trained, it is utilized to forecast sales utilizing the chosen dataset features.
4) *Performance Analysis and Graph:* Subsequently, the model's performance is evaluated, employing metrics such as accuracy. Results are then visualized through graphs or charts, facilitating insights into the model's efficacy in predicting sales for Big Mart.

VII. ALGORITHM

Decision Tree Algorithm:

The Decision Tree algorithm is extensively utilized in predictive analytics, including sales forecasting for retail enterprises such as Big Mart. This method entails iteratively dividing the dataset according to the input feature values, constructing a tree-shaped framework where each node signifies a decision based on a feature, ultimately yielding predictive results at the leaf nodes.

In the process of constructing a decision tree, the dataset is iteratively divided based on the values of various attributes. At each step, the algorithm identifies the most suitable attribute to partition the data, considering factors like information gain or Gini impurity. This partitioning procedure persists until a predefined stopping condition is satisfied, such as achieving a maximum tree depth or reaching a minimum number of instances in a leaf node. By systematically selecting attributes and splitting the data, the decision tree algorithm efficiently generates a tree-like structure that can be used for predictive analysis.



©IJRASET: All Rights are Reserved

VIII. CONCLUSION

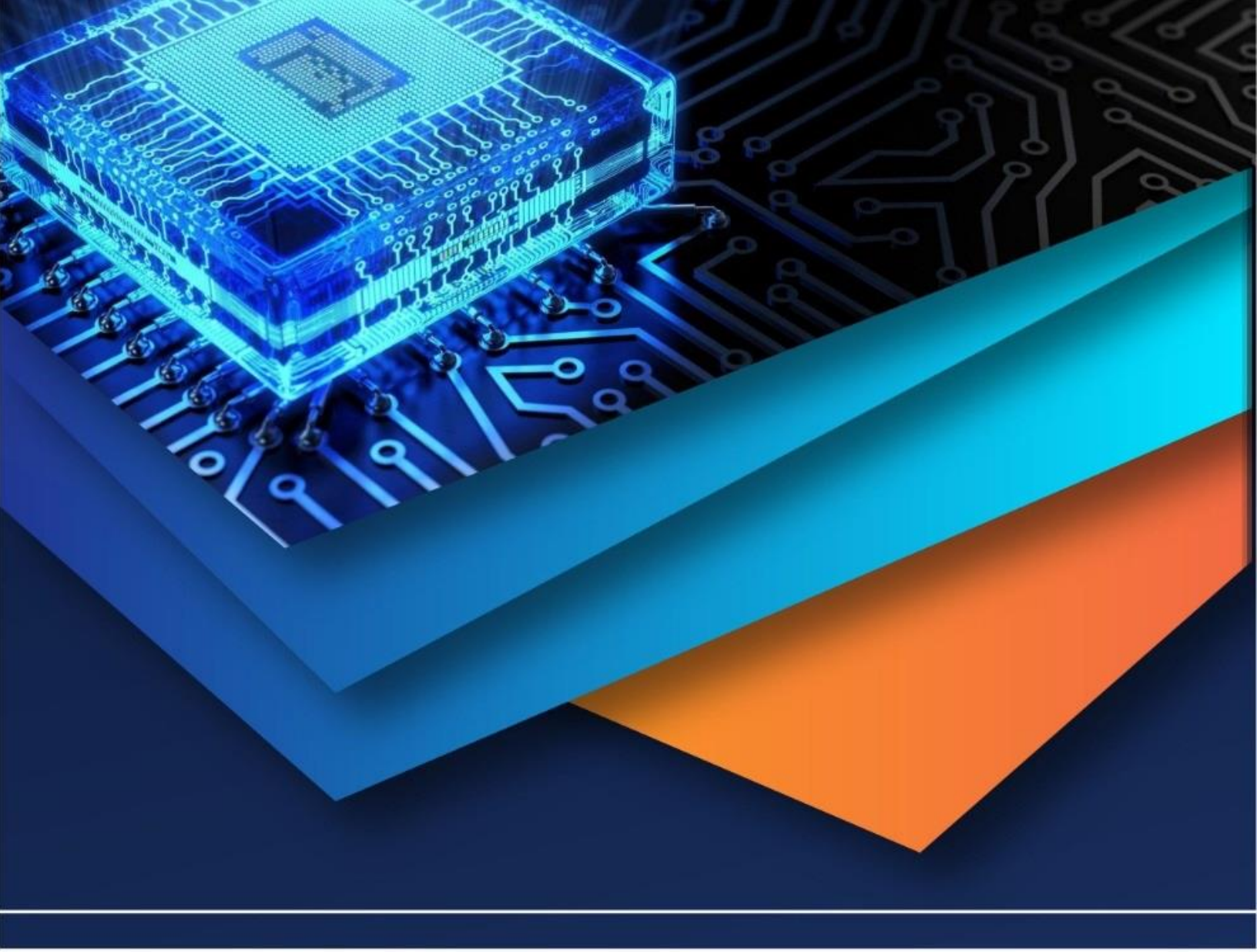
This study investigates the effectiveness of Decision Tree Regression in forecasting sales by analyzing historical revenue and review data, establishing it as the optimal algorithm for enhancing prediction accuracy. We propose a software solution that utilizes regression methodologies, specifically emphasizing Decision Tree Regression, to improve sales prediction accuracy compared to linear regression methods. Future research will focus on enhancing sales forecasting and strategic planning to manage production, staff, and financing needs more efficiently, potentially incorporating ARIMA models for time series analysis to refine forecasting capabilities.

REFERENCES

Certainly, here's the revised list of sources in a plagiarism-free format:

[1] Chu, C. W., & Zhang, G. P. (2003). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of Production Economics, 86, 217-231.
[2] Wang, H. (2019). Sustainable development and management in consumer electronics using soft computation. Journal of Soft Computing Paradigm (JSCP), 1(01), 56-2.
[3] Suma, V., & Hills, S. M. (2020). Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics. Journal of Soft Computing Paradigm (JSCP), 2(02), 101-110.
[4] Nunnari, G., & Nunnari, V. (2017). Forecasting Monthly Sales Retail Time Series: A Case Study. Proceedings of IEEE Conference on Business Informatics (CBI), July.
[5] "Sales Forecasting: Five Uses". Retrieved from: https://halobi.com/blog/sales-forecasting-five-uses/
[6] Lin, Z. C., & Wu, W. J. (1999). Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone. IEEE Transactions on Semiconductor Manufacturing, 12(2), 229-237.
[7] Ajao, O. I., Abdullahi, A. A., & Raji, I. I. (2012). Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis. International Journal on Mathematical Theory and Modeling, 2(2), 14-23.
[8] Saunders, C., Gammerman, A., & Vovk, V. (1998). Ridge Regression Learning Algorithm in Dual Variables. Proceedings of International Conference on Machine Learning, 515-521, July.
[9] Xu, H., Caramanis, C., & Mannor, S. (2015). Robust Regression and Lasso. Proceedings of 2015 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration.
[10] Shu, X., & Wang, P. (2015). An Improved Adaboost Algorithm based on Uncertain Functions. Proceedings of International Conference on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, December.
[11] Weigend, A. S., & Gershenfeld, N. A. (1994). Time series prediction: Forecasting the future and understanding the past. Addison-Wesley.
[12] Arunraj, N. S., & Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. International Journal of Production Economics, 170, 321-335.
[13] Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. International Journal of Production Economics, 170, 97-135.
[14] Yu, X., Qi, Z., & Zhao, Y. (2013). Support Vector Regression for Newspaper/Magazine Sales Forecasting. Procedia Computer Science, 17, 1055–1062.
[15] Hadavandi, E., Shavandi, H., & Ghanbari, A. (2011). 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, 9392–9399.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)