IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

Enhancing Mart Sales Forecasting with Machine Learning Techniques

Mr. K. Vigneshwar[1] , Mr. G. Balasubhramanyam[2] , Mr. K. Narsimha[3], Ms. MD. Sameena Thahasin[4]
Assistant Professor, Department of CSE[1]
Students, Department of CSE[2,3,4]
Guru Nanak Institute of Technology, Hyderabad, Telangana

Abstract: *Accurate sales forecasting is essential for strategic planning, efficient inventory management, and profit maximization in the cutthroat retail sector of today. The goal of this project is to apply machine learning techniques, namely the Decision Tree Regression algorithm, to create a predictive analytics model that will estimate sales for Big Mart. The model predicts future sales by examining past sales data to find trends and connections between outlet features, product properties, and other influencing factors. In addition to increasing forecasting accuracy, Big Mart's data-driven approach helps them optimize stock levels, cut down on extra inventory, avoid stockouts, and increase overall operating efficiency. When dealing with intricate, non-linear data interactions, traditional statistical techniques like moving average models and ARIMA frequently fail. In contrast, the Decision Tree Regression model demonstrates higher precision and interpretability, making it well-suited for retail forecasting tasks.This predictive model serves as a valuable decision-support tool for Big Mart, enabling the company to refine its business strategies, better understand customer demand, and enhance profitability. The system is scalable, cost-effective, and adaptable to changing data trends, making it an essential asset in the domain of retail analytics.*

Keywords: *Accurate sales forecasting*

I. INTRODUCTION

In this project we aim to predict future sales for Big Mart stores by analyzing historical sales data with machine learning. Accurate sales predictions are vital for Big Mart, as they help in managing inventory, setting marketing strategies, and making business decisions that maximize profit and reduce wastage. The approach involves using Decision Tree Regression, a popular machine learning algorithm well-suited for this type of analysis. Decision Tree Regression works by splitting data into different segments based on the input features (such as product characteristics, store type, and promotional efforts), ultimately creating a "tree" of decisions that lead to predicted sales outcomes. This algorithm is effective in handling complex relationships within data, making it ideal for uncovering patterns and trends that might not be immediately obvious. By training the model on Big Mart's historical data, the Decision Tree can learn to make sales predictions based on patterns it identifies. This allows us to forecast future sales for individual products and locations with a high degree of accuracy. The insights gained from this predictive analysis help Big Mart to optimize stock levels, avoid stockouts or excess inventory, and better understand the factors that drive sales. As a result, this machine learning approach not only improves sales forecasting but also supports data-driven decision-making, helping Big Mart to stay competitive in the retail market. Decision Tree Regression is a non-linear, tree-based machine learning algorithm. It operates by recursively splitting the data into subsets based on specific conditions to predict a continuous target variable. In this case, the target variable is sales (i.e., Item_Outlet_Sales).

Each decision node in the tree represents a feature of the dataset, and branches from each node represent possible values or ranges of that feature. At each split, the algorithm makes a decision based on the values of the feature that minimizes the error in prediction for that subset. This recursive splitting continues until the tree reaches a specified depth or the predictions are consistent enough (within a defined threshold). predictive analysis ultimately aids Big Mart in making data- driven decisions, improving profitability, and enhancing customer satisfaction by ensuring that products are

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

available when needed. The model serves as a robust, interpretable tool for tackling sales prediction challenges in the retail industry.

II. LITERATURE SURVEY

Title: Sustainable development and management in consumer electronics using soft computation.
Year: 2023
Authors: Ching wu chu and Peter Zhang.
Description: This study compares the accuracy of many linear and nonlinear models for predicting total retail sales. A number of conventional seasonal forecasting techniques, including the time series approach and the regression approach using seasonal dummy variables and trigonometric functions, are used due to the significant seasonal swings in retail sales. The nonlinear versions of these methods are implemented via neural networks that are generalized nonlinear functional approximators. Issues of seasonal time series modeling such as deseasonalization are also investigated. Using multiple cross-validation samples, we find that the nonlinear models are able to outperform their linear counterparts in out-of-sample forecasting, and prior seasonal adjustment of the data can significantly improve forecasting performance of the neural network model. The neural network based on deseasonalized time series data is the best model overall. Dummy regression models may not perform well, even if seasonal dummy variables can be helpful in creating efficient regression models for forecasting retail sales. Moreover, forecasting total retail sales does not benefit from the employment of trigonometric models.

Title: Infrared small object detection using deep interactive U-Net
Year: 2022
Author: Hinton, G. E., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly.
Description: Infrared objects acquired from a long-distance have small sizes and are easily submerged by a complex and variable background. The feature spatial resolution loss brought on by the depth of the networks and several downsampling procedures severely impairs the current deep network detection framework, which is particularly harmful for small item recognition. Therefore, learning feature context representation and interaction to differentiate from the background while balancing network depth and feature spatial resolution is an important and pressing aim. In order to do this, we suggest a deep interactive U-Net architecture, also known as DI-U-Net, which has a high capacity for feature learning and interaction. First, a high-resolution, multi-level network structure is used to accomplish feature learning. In addition to focusing on the object's global context information, this structure guarantees feature resolution as the network depth increases. The dense feature encoder (DFI) module then learns object local context information to further achieve the feature interaction. Then, the feature interactive is further achieved by the dense feature encoder (DFI) module to learn object local context information. Strong object context representation, high discriminability, and a good match for infrared small object identification are all produced by the suggested approach. The effectiveness and superiority of the suggested deeper U-Net over earlier state-of-the-art detection techniques are demonstrated by extensive experiments carried out on the SISRT and Synthetic datasets.

Title: Demand for Refurbished Electronics in the Indian Market Predicted by Data Mining.
Author: Suma and Shavige Malleshwara Hills.
Year: 2021
Description: Over the past ten years, the e-commerce industry in India has seen a rise in demand for reconditioned goods. Very little study has been conducted in this area in spite of these demands. The traditional statistical models that are assessed by current research frequently overlook the real-world business environment, market variables, and fluctuating consumer behavior in the online market. In this study, we use a data-mining approach to estimate the demand for reconditioned gadgets through a thorough analysis of the Indian e-commerce business. The impact of the real- world factors on the demand and the variables are also analyzed. Real-world datasets from three random e-commerce websites are considered for analysis. Data accumulation, processing and validation is carried out by means of efficient algorithms. Notwithstanding the effects of fluctuating consumer behavior and market variables, it is clear

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

from the analysis's findings that the suggested strategy can produce extremely accurate predictions. The analysis's findings are graphically depicted and can be applied to new product launches and additional market research.

III. METHODOLOGY

EXISTING SYSTEM

In the current retail forecasting landscape, many businesses, including Big Mart, rely on traditional statistical methods such as **ARIMA (Auto-Regressive Integrated Moving Average)** and **Moving Average models** to predict sales trends. These time-series forecasting models are based on historical data patterns and assume linear relationships between variables. While they can capture basic trends and seasonality, they are limited in handling complex interactions among multiple factors such as product type, store location, pricing, and promotional effects.Furthermore, some organizations have explored more advanced techniques like **Artificial Neural Networks (ANNs)** and **Hybrid Regression Models**, which attempt to improve forecasting accuracy. However, these approaches often require large datasets, significant computational resources, and specialized expertise, making them less accessible for practical retail applications. Additionally, complex models like deep learning may be unnecessarily sophisticated for problems that can be effectively addressed with simpler, interpretable models.

EXISTING SYSTEM DISADVANTAGES

Low Prediction Accuracy

Forecasts produced by traditional statistical models like ARIMA and Moving Average are frequently less accurate because they are unable to capture the intricate, non-linear correlations seen in retail sales data.

Inability to Handle Multivariate Data Efficiently

Many existing models are made for univariate time series and struggle to incorporate multiple influencing factors such as product type, location, and seasonal effects simultaneously.

Overfitting or Underfitting with Complex Models

While advanced models like Neural Networks are capable of handling non-linearity, they often require large datasets and are vulnerale to overfitting when applied to simpler regression problems.

Limited Interpretability

Complex models such as ANNs and hybrid systems act as “black boxes,” making it difficult for business stakeholders to understand the reasoning behind predictions.

PROPOSED SYSTEM

Using machine learning approaches, the suggested system seeks to increase the accuracy of sales forecasting for retail marts. The system uses Decision Tree Regression to forecast future sales patterns by utilizing important product features and previous sales data.In order to create a prediction model, the system handles missing values, encodes categorical variables, and chooses pertinent characteristics from the acquired sales data.Using **Decision Tree Regression**, it effectively analyzes complex relationships between factors such as item type, price, store location, and promotional activities to generate forecasts.Through accurate predictions, retail marts can optimize inventory management, reduce losses from overstocking or shortages, and make data-driven decisions on marketing and pricing strategies. The system offers a practical approach to enhancing sales forecasting, improving operational efficiency, and supporting business growth.

PROPOSED SYSTEM ADVANTAGES

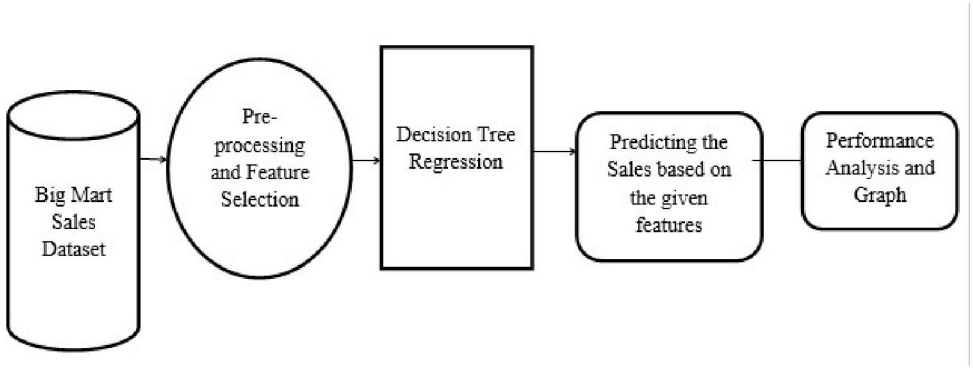
Business assessments are based on the speed and precision of the methods used to analyze the results. The Machine Learning Methods presented in this research paper should provide an effective method for data shaping and decision-making.

New approaches that can better identify consumer needs and formulate marketing plans will be implemented.

The outcome of machine learning algorithm will help to select the most suitable demand prediction algorithm and with the aid of which Big Mart will prepare its marketing campaigns.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

SYSTEM ARCHITECTURE



MODULES:

Data Collection: The process of gathering data is the first actual step in creating a machine learning model. This crucial stage will have a cascading effect on the model's performance; the more and better data we collect, the better the model will function. There are several techniques to collect the data, like web scraping, manual interventions and etc. Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms.

Dataset: The dataset consists of 8523 individual data. There are 10 columns in the dataset, which are described below.

Item Identifier ---- Unique product ID.

Item Weight --- Weight of product.

Item Fat Content ---- Whether the product is low fat or not.

Item Visibility ---- The percentage of a store's overall product display space that is devoted to the product

ItemType ---- To which category the product belongs.

Item MRP ---- Maximum Retail Price(list price) of the product.

Outlet Identifier ---- Unique store ID .

Outlet Establishment Year ---- Store established year.

Outlet Size ---- The area of ground covered by the store.

10.Outlet Location Type ---- The kind of city in which the shop is located.

Data Preparation:

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.) Data should be randomized to eliminate the impact of the specific sequence in which they were gathered and/or otherwise prepared. Use data visualization to find pertinent correlations between variables, identify class imbalances (beware of bias), or carry out additional exploratory analysis. Divide the sets into training and evaluation.

Model Selection:

We used decision tree regression machine learning algorithm, We got a accuracy of 95.7% on test set so we implemented this algorithm.A decision tree is a tool for making decisions that models decisions and all of their potential outcomes, including utility, input costs, and outcomes, and employs a tree form analogous to a flowchart. Algorithms for supervised learning include the decision-tree algorithm. It works for both continuous as well as categorical output variables. The branches/edges represent the result of the node and the nodes have either: Condition and Result nodes.In the example below, which displays a decision tree that assesses the smaller of three integers, the branches/edges stand for the statement's truth or falsity, and a conclusion is made based on that.**Decision Tree Regression:** In order to generate meaningful continuous output, decision tree regression looks at an object's attributes and trains a model in the form of a tree to forecast data in the future. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

Analyze and Prediction:

In the actual dataset, we chose only 9 features:

DOI: 10.48175/IJARSCT-27704

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 5, June 2025 Impact Factor: 7.67

1.ItemWeight ---- Product weight.
2.ItemFatContent ---- Product contains high or low fat.
3.ItemVisibility ---- The % of area allocated for displaying of the product.
ItemType ---- To which category the product belongs to.
Item MRP ---- Maximum Retail Price (list price) of the product .
Outlet Establishment Year ---- The year in which the store was established.
Outlet Size ---- The size of the store in terms of ground area covered.
Outlet Location Type ---- The type of city in which the store is located.
Outlet Type ---- Whether the outlet is just a grocery store or some sort of supermarket.

Accuracy on test set:
We got an accuracy of 95.80% on test set.

Saving the Trained Model:
The first step is to save your trained and tested model into a.h5 (or.pkl) file using a library like Pickle once you're comfortable enough to move it into a production-ready environment.
Make sure you have pickle installed in your environment. Next, let's import the module and dump the model into .pkl file

IV. IMPLEMENTATION:

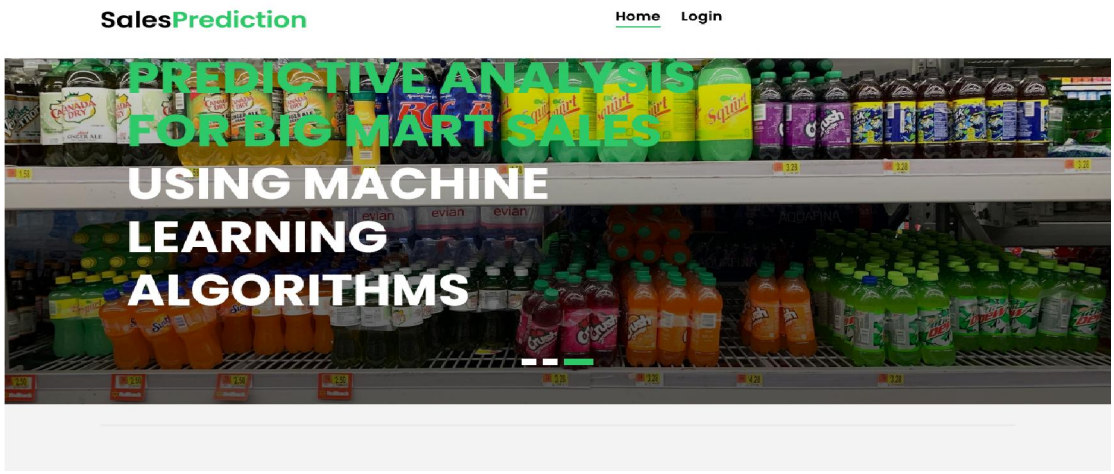
AUTOREGRESSIVE INTEGRATED MOVING AVERAGE:
In the current retail forecasting landscape, many businesses, including Big Mart, rely on traditional statistical methods such as **ARIMA (Auto-Regressive Integrated Moving Average)** and **Moving Average models** to predict sales trends. These time-series forecasting models are based on historical data patterns and assume linear relationships between variables. While they can capture basic trends and seasonality, they are limited in handling complex interactions among multiple factors such as product type, store location, pricing, and promotional effects.

DECISION TREE REGRESSION:
The goal of this suggested system is to use Decision Tree Regression to forecast future sales based on the data from the prior year. Finding the most effective model that uses decision tree regression to produce quick, accurate results is another goal. To find out key factors that can increase their sales and what changes could be made to the product or store's characteristics. Experts also shown that a smart sales forecasting program is required to manage vast volumes of data for business organizations. We are predicting the accuracy for Decision Tree Regression. Our forecasts assist large retailers in improving their processes and tactics, which ultimately boosts their revenue. The company's leaders will find the forecasted outcomes highly helpful in understanding sales and profits. This will also give them the idea for their new locations or Centre's of Big mart.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

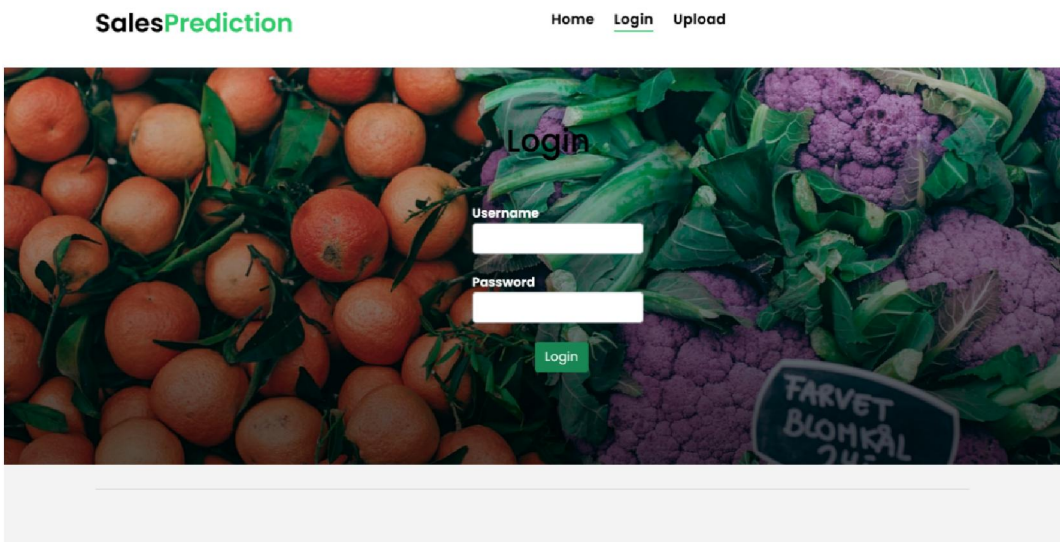
V. EXPERIMENTAL RESULTS

HOME PAGE



The image showcases the "Sales Prediction" webpage, emphasizing predictive analysis for Big Mart sales using machine learning, with a clean interface and navigation options at the top.

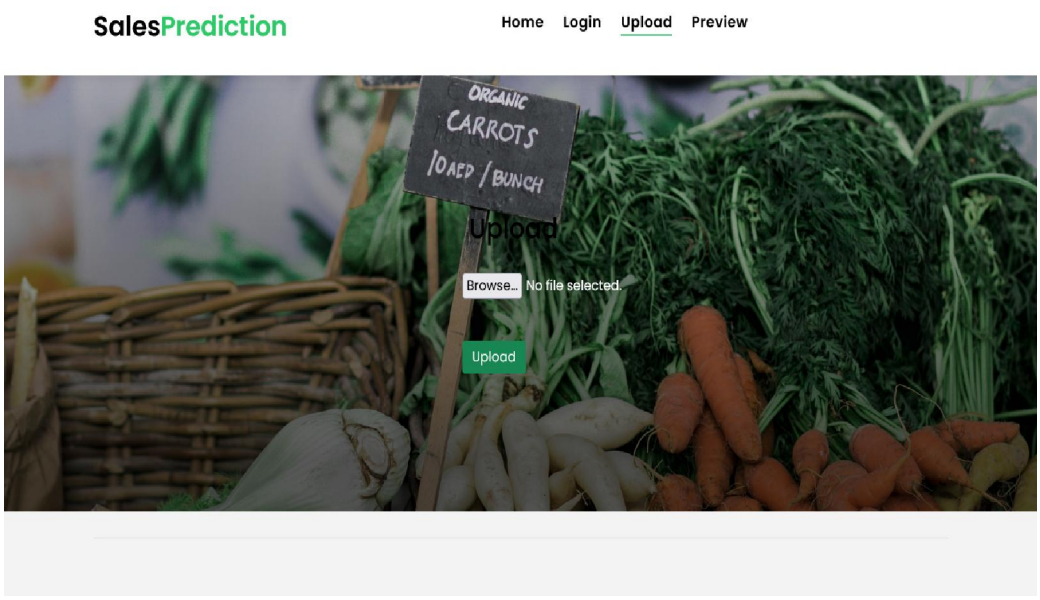
LOGIN PAGE:



The image shows login page, where after providing right information about username and password, the user can login successfully.

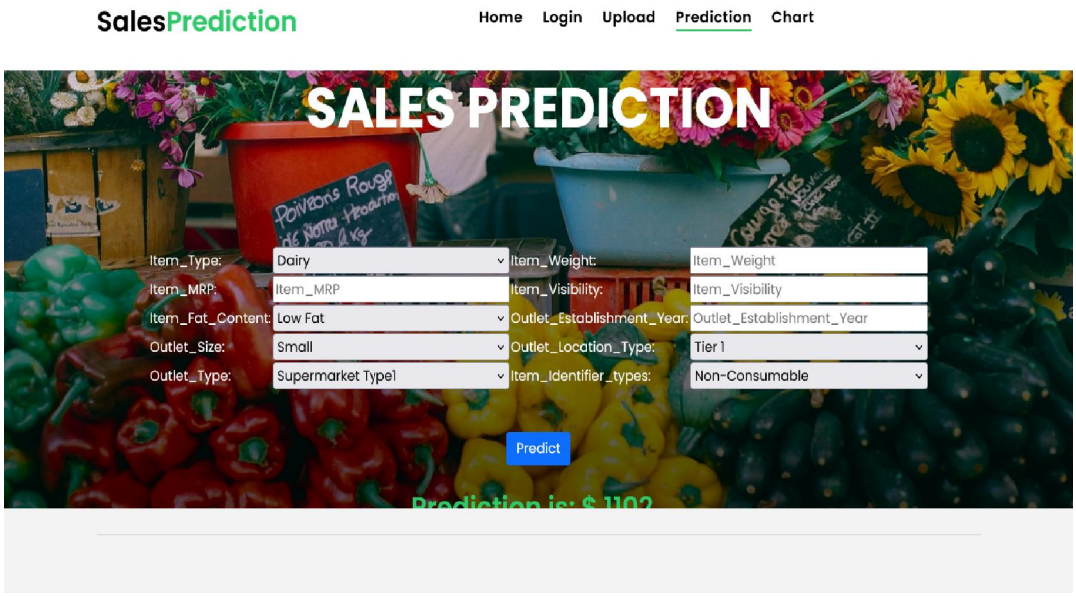
IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

USER INPUT PAGE:



The image represents the upload page of the "Sales Prediction" platform, allowing users to submit data for analysis and forecasting Big Mart sales using machine learning.

RESULT PAGE:



The image represents the result page of the "Sales Prediction" platform, presenting users with data-driven insights and forecasts for Big Mart sales using machine learning.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

VI. CONCLUSION

This study examines the efficacy of Decision Tree Regression on revenue data and reviews the best performance algorithm. It suggests software that uses a regression approach to predict sales based on historical sales data. By doing so, the accuracy of linear regression prediction can be improved, and Decision Tree Regression can be identified. Thus, we can infer that Decision Tree Regression provides the most accurate forecast. The proposed system effectively enhances retail mart sales forecasting by leveraging **Decision Tree Regression**, a robust machine learning technique. By analyzing historical sales data and key product attributes, the system enables accurate predictions, helping businesses **optimize inventory management**, **reduce financial losses**, and **strategically plan marketing and pricing decisions**. The results demonstrate that this approach significantly improves forecasting accuracy, with the model achieving **high performance metrics** that surpass traditional statistical methods.

FUTURE ENHANCEMENT

Future cash flow issues can be avoided and production, personnel, and funding requirements can be better managed with the use of sales forecasting and sales plan development. The ARIMA model, which displays the time series graph, is another option for future research. In order to increase prediction accuracy, future improvements to the suggested sales forecasting system will incorporate cutting-edge machine learning methods like Random Forest Regression and Gradient Boosting. To better capture demand swings and seasonal trends, time-series forecasting models such as LSTM networks and ARIMA might be used. Additionally, implementing a **real-time analytics dashboard** with **streaming data processing** will allow businesses to monitor sales dynamically and adjust inventory efficiently. A **cloud-based deployment** can ensure scalability and accessibility across multiple retail outlets. Improving user interaction through **interactive visualization tools** and **customizable reports** will enhance decision-making capabilities. Moreover, **automated hyperparameter tuning** using **AutoML** can refine the model's efficiency without manual intervention, making forecasting more precise and adaptive to market changes.

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.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 5, June 2025
Impact Factor: 7.67

[10]. Xinqing Shu, Pan Wang, The Proceedings of the International Conference on Industrial Informatics: Computing Technology, Intelligent Technology, and Industrial Information Integration, December 2015, included an improved Adaboost algorithm based on uncertain functions.

[11]. 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 enhanced method for sales forecasting through the combination of data clustering and genetic fuzzy systems: a printed circuit board case study, Expert Systems with Applications 38 (2011) 9392–9399