

Enhancing Mart Sales Forecasting Using Machine Learning Techniques

MR.M.AJAY KUMAR ,VITHANALA JAHNAVI,RAYANABOINA MANISHA,SAMA USHA REDDY

Assistant Professor, Guru Nanak Institute of Technology, CSE Department, Hyderabad
Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology
Hyderabad,Telangana.

ABSTRACT:

Currently, supermarket run-centers, Big Marts keep track of each individual item's sales data in order to anticipate potential consumer demand and update inventory management. Anomalies and general trends are often discovered by mining the data warehouse's data store. For retailers like Big Mart, the resulting data can be used to forecast future sales volume using various machine learning techniques like big mart. A predictive model was developed using Decision Tree Regression for forecasting the sales of a business such as Big -Mart, and it was discovered that the model outperforms existing models.

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

II. LITERATURE REVIEW

The given literature survey discusses different machine learning, deep learning, soft computing, and data mining techniques used for object detection, sales forecasting, sustainable electronics management, and demand prediction. Wu et al. (2022) proposed a Deep Interactive U-Net (DI-U-Net) model for accurate infrared small object detection using high-resolution feature learning and dense feature interaction. Ching Wu Chu and Guoqiang Peter Zhang (2023) compared linear and nonlinear forecasting models and found neural network models more effective for retail sales prediction. Wang Haoxiang (2022) used ANFIS-based soft computing techniques for sustainable consumer electronics management and green supply chain optimization. Suma and Shavige Malleshwara Hills (2021) applied data mining methods to predict demand for refurbished electronics in the Indian market using real-world e-commerce data. Giuseppe Nunnari and Valeria Nunnari (2021) used Neuro-Fuzzy and Neural Network approaches for retail sales time series forecasting and achieved better performance than traditional models. Zone-Ching Lin and Wen-Jang Wu (2021) demonstrated the effectiveness of Multiple Linear Regression in solving nonlinear overlay accuracy problems in imaging systems. Overall, these studies show the importance of intelligent computational methods in prediction, forecasting, sustainability, and detection applications.

III. METHODOLOGY

The project consists of several modules including Data Collection, Dataset, Data Preparation, Model Selection, Analyze and Prediction, Accuracy on Test Set, and Saving the Trained Model. In the Data Collection stage, data is gathered using methods like web scraping and manual collection, which is important for improving model performance. The dataset contains 8523 records and 12 attributes related to product and store information, with ItemOutletSales as the target

variable. During Data Preparation, the data is cleaned, normalized, randomized, visualized, and divided into training and testing sets. For Model Selection, the Decision Tree Regression algorithm was chosen because it achieved high prediction performance. The model uses 9 important features such as ItemWeight, ItemFatContent, ItemVisibility, ItemType, ItemMRP, OutletEstablishmentYear, OutletSize, OutletLocationType, and OutletType for analysis and prediction of sales. The model achieved an accuracy of 95.80% on the test dataset. Finally, the trained model is saved into a .h5 or .pkl file using the pickle library for future use in production environments.

Existing System Disadvantages

- a. Existing system models prediction analysis which gives less accuracy.
- b. Forecasting methods and applications contains Lack of Data and short life cycles.

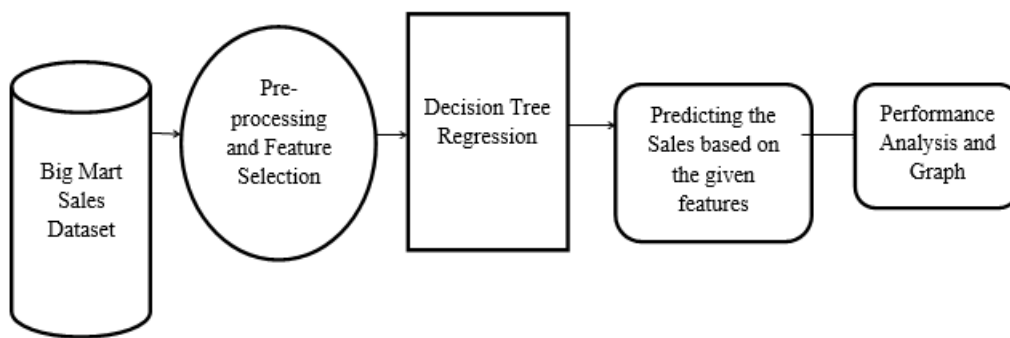
PROPOSED SYSTEM:

The main objective of the proposed system is to predict future sales using previous year sales data with the help of the Decision Tree Regression algorithm. The system aims to identify the most efficient and accurate prediction model while analyzing important factors that influence product and store sales. It also helps business organizations manage large amounts of sales data effectively through smart forecasting techniques. The predicted results support Big Mart executives in improving business strategies, increasing profits, and making better decisions regarding new store locations and sales management.

Proposed System Advantages

1. Machine learning helps businesses analyze sales data quickly and accurately for better decision-making.
2. The system helps identify customer needs and improve marketing strategies effectively.

SYSTEM ARCHITECTURE



MODULES:

1.Data Collection

Data collection is the first step in developing a machine learning model. Better quality and larger amounts of data improve model performance. Data can be collected using methods like web scraping and manual collection.

2.Dataset

The dataset contains 8523 records and 12 attributes related to products and stores, such as Item Weight, Item Type, Item MRP, Outlet Size, and Item Outlet Sales. The target variable is ItemOutletSales, which is predicted by the model.

3.Data Preparation

In this stage, the data is cleaned by removing duplicates, handling missing values, correcting errors, and converting data types. The data is also normalized, visualized, randomized, and divided into training and testing sets.

4. Model Selection

The Decision Tree Regression algorithm was selected because it achieved high performance with 95.7% accuracy. It is a supervised learning algorithm that predicts continuous output values using a tree-like structure.

5. Analyze and Prediction

For prediction, 9 important features were selected from the dataset, including Item Weight, Item Fat Content, Item Visibility, Item Type, Item MRP, Outlet Establishment Year, Outlet Size, Outlet Location Type, and Outlet Type.

IV. IMPLEMENTATION

The implementation of enhancing mart sales forecasting using machine learning techniques begins with collecting sales and product-related data from Big Mart stores. The data is then preprocessed by handling missing values, removing duplicates, and converting categorical data into numerical form... After training, the model predicts future sales using test data and achieved an accuracy of about 95.8%. Finally, the trained model is saved using the pickle library for future business analysis and sales forecasting applications.

V. EXPERIMENTAL RESULTS



Fig: Home Page

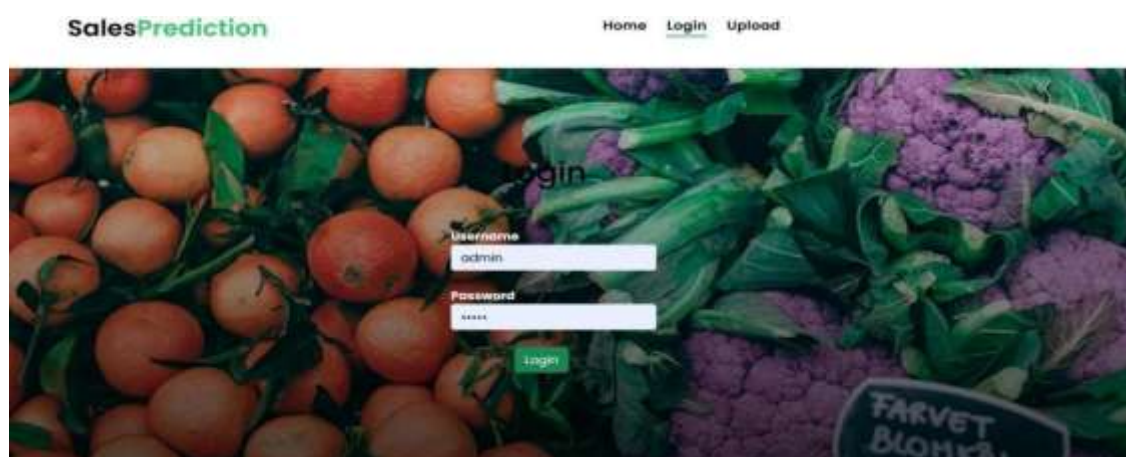
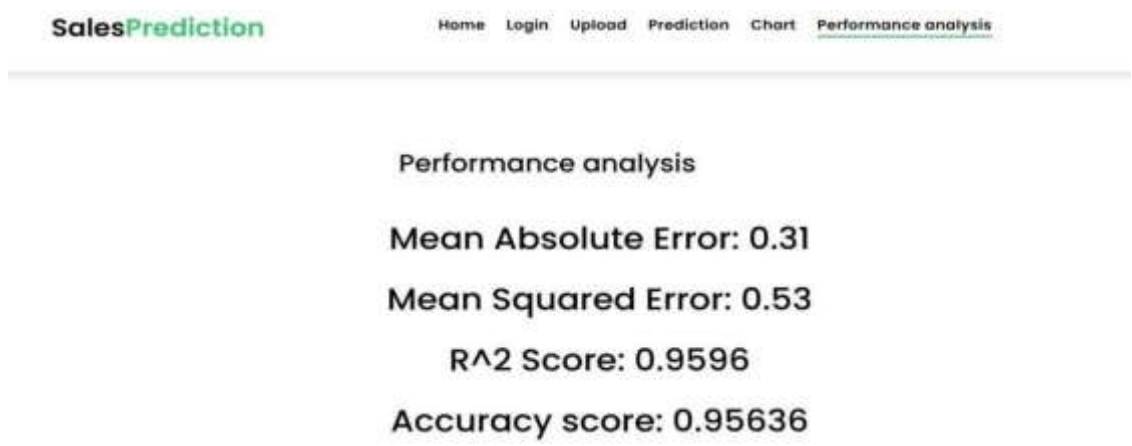


Fig: Login Page



Item ID	Product Name	Category	Weight (kg)	Unit	Price (USD)	Expiry Date	Supplier	Quality Score	Storage Temp (°C)	Shelf Life (Days)	Notes
4994	FDK33	6.480	Low Fat	0.0003000	Dairy	1/5/2028	OUT027	1802	Medium	Tier 2	Supermarket - 3174.6088 Type3
4996	FDL58	5.780	Regular	0.074648	Snack Foods	262.2568	OUT046	1997	Small	Tier 1	Supermarket - 5273.093 Type1
4997	HCJ05	18.700	Low Fat	0.049350	Health and Hygiene	151.8862	OUT013	1887	High	Tier 3	Supermarket - 2354.3004 Type1
4998	FDSP0	18.200	Low Fat	0.025385	Snack Foods	180.3378	OUT017	2007	-	Tier 2	Supermarket - 3428.2542 Type1
4999	FDK34	-	Low Fat	0.038340	Snack Foods	340.1564	OUT027	1883	Medium	Tier 3	Supermarket - 7193.8830 Type3

[Click to Train / Test](#)

VI. CONCLUSION

The Enhancing Mart Sales Forecasting using Machine Learning Techniques project successfully predicts future sales by analyzing historical sales data and important product and store features. The Decision Tree Regression algorithm provided high prediction accuracy and helped identify factors affecting sales performance. The system improves business decision-making by supporting better inventory management, marketing strategies, and profit planning. The obtained results show that machine learning techniques can effectively enhance sales forecasting and help retail businesses like Big Mart achieve better growth and operational efficiency.

VII. FUTURE ENHANCEMENT

In future, the sales forecasting system can be improved by using advanced machine learning and deep learning algorithms such as Random Forest, XGBoost, and Neural Networks to achieve higher prediction accuracy. Real-time sales data and customer feedback can also be integrated for better forecasting performance. The system can be enhanced with graphical dashboards and cloud-based storage for easier business analysis and accessibility. Additional factors such as seasonal trends, customer behavior, online sales, and market conditions can be included to make the predictions more accurate and useful for business growth and decision-making.

REFERENCES

1. Machine Learning techniques for Big Mart sales prediction using Decision Tree Regression and data mining approaches.
2. Suma, V., and Shavige Malleshwara Hills, "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics," 2021.
3. Giuseppe Nunnari and Valeria Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study," 2021.
4. Ching Wu Chu and Guoqiang Peter Zhang, "Sustainable Development and Management in Consumer Electronics Using Soft Computation," 2023.
5. Wang Haoxiang, "Sustainable Development and Management in Consumer Electronics Using Soft Computation," 2022.
6. Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning, Springer Publications.
7. Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press.
8. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media.