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 1, August 2025
Impact Factor: 7.67

BigMart Sales Forecasting Using Decision Tree Regression for Smart Inventory Management

Miss. Rutuja Chakradhar Salve[1] and Prof. VijayKumar M. P.[2]
Department of Computer Science & Engineering[1,2]
Shreeyash College of Engineering & Technology, Chh. Sambhajinagar
rutusalve933@gmail.com[1], viju127@gmail.com[2]

Abstract: *In modern retail environments, especially in large supermarket chains like Big Mart, maintaining detailed sales records of individual products has become a standard practice. This sales data plays a critical role in anticipating customer demand and optimizing inventory management strategies. By analyzing large volumes of historical data stored in centralized warehouses, businesses can uncover patterns, trends, and anomalies that help them make informed, data-driven decisions.This study presents a machine learning approach for sales forecasting using Decision Tree Regression, trained on historical sales data from Big Mart. The model effectively captures complex relationships between various sales factors and demonstrates superior accuracy when compared to traditional regression techniques. The results indicate that Decision Tree Regression is a reliable and efficient method for predicting future sales, ultimately supporting better demand forecasting, resource planning, and profitability in the retail sector.*

Keywords: Sales Prediction, Big Mart Dataset, Machine Learning, Decision Tree Regression, Forecasting, Inventory Management, Data Mining, Retail Analytics, Predictive Modelling

I. INTRODUCTION

In today's highly competitive retail landscape, the rapid expansion of global shopping malls, supermarkets, and online marketplaces has transformed consumer shopping behavior and intensified competition among retailers [1][2]. Traditional brick-and-mortar stores, particularly large chains like Big Mart, are under immense pressure to adapt to these changing dynamics. The key to sustaining profitability lies in their ability to predict consumer demand accurately, ensuring that products are available when and where customers want them. Without proper forecasting mechanisms, retailers risk either overstocking—leading to wastage and increased holding costs—or stockouts, which result in lost sales and diminished customer satisfaction [3][4]. Hence, the need for robust predictive analysis models that can anticipate sales trends has become more critical than ever [5][6].

Predictive analysis, especially through machine learning algorithms, provides a powerful approach to tackling these challenges. By leveraging historical sales data, customer preferences, and external factors like seasonality and promotional activities, machine learning models can uncover intricate patterns that influence purchasing behaviors [7][8]. These data-driven models offer a sophisticated alternative to conventional forecasting methods, which often rely on linear trends and may fail to capture the complexities of the retail market [9]. For example, machine learning techniques such as Random Forest, Decision Tree Regression, Gradient Boosting, and Support Vector Machines have shown superior performance in modeling non-linear relationships and delivering precise forecasts for sales data in retail environments [10][11].

Decision Tree Regression, in particular, stands out for its interpretability and ability to model complex interactions among variables [12]. It partitions the dataset into subsets based on feature values, making it possible to capture various conditions under which sales might increase or decrease [13]. Moreover, ensemble techniques like Random Forest and Gradient Boosting further enhance prediction accuracy by combining multiple decision trees to reduce overfitting and improve generalization [14][15]. These models help retailers not only forecast product-wise sales but also understand the factors driving those sales, such as pricing strategies, promotional effectiveness, and regional consumer preferences [16]. Such insights are invaluable for strategic planning, marketing optimization, and supply chain efficiency.

The application of predictive analytics in sales forecasting also aids in streamlining logistics and inventory management [17]. Accurate forecasts allow for better transportation planning, ensuring that products are delivered to the right locations in a timely manner, which reduces operational costs and minimizes waste [18]. For instance, during festive seasons or promotional campaigns, machine learning models can help anticipate demand spikes and ensure sufficient stock levels are maintained across stores. Additionally, integrating these forecasts with real-time data feeds can facilitate dynamic pricing and promotional adjustments, enhancing competitiveness in a rapidly shifting market landscape [19]. This alignment between predictive analytics and operational execution is crucial for driving profitability and maintaining high service levels.

Furthermore, predictive models empower retailers to design personalized customer experiences [20]. By analyzing customer purchase history and preferences, retailers can tailor marketing campaigns to target specific customer segments with customized offers and recommendations. This not only boosts conversion rates but also fosters customer loyalty by providing a more engaging and relevant shopping experience [9][11]. In the context of Big Mart, employing predictive analytics ensures that the right mix of products is available across various store locations, aligning stock levels with local consumer preferences and purchasing power.

In conclusion, predictive analysis using machine learning algorithms has become a strategic necessity for modern retailers like Big Mart to thrive in a competitive environment [1][3][12]. By harnessing advanced forecasting models, these businesses can optimize inventory, improve customer satisfaction, and enhance operational efficiency. Moreover, the insights derived from predictive models serve as a foundation for informed decision-making across marketing, supply chain management, and financial planning. As the retail sector continues to evolve, the integration of machine learning-driven predictive analytics will remain pivotal in driving growth, profitability, and sustainable competitive advantage [8][15][17].

II. OBJECTIVE

• To develop a predictive model using Decision Tree Regression for accurately forecasting sales based on historical Big Mart sales data.
• To identify and analyze key factors such as item type, outlet location, visibility, and MRP that influence product sales in retail settings.
• To enhance inventory and resource management by providing data-driven sales forecasts that help minimize stockouts and overstocking.
• To evaluate the performance of the proposed model using metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score.
• To create a user-friendly software interface that allows users to input sales data and obtain real-time sales predictions for better business decision-making.

III. LITERATURE SURVEY

Our study builds upon extensive prior research in the field of sales forecasting using machine learning techniques, particularly Decision Tree Regression. Existing literature demonstrates a growing interest in applying predictive models to retail datasets for improving inventory management, demand planning, and marketing strategies. Several studies, such as those by Kumar and Sharma (2020), and Sharma and Rawat (2019), highlight the superiority of Decision Tree Regression over traditional linear models due to its ability to handle non-linear data and categorical variables effectively. Comparative analyses by researchers like Rajesh and Kumar (2021) and Yadav and Sinha (2021) further support the efficiency and interpretability of decision trees in forecasting tasks. Other works emphasize the role of feature engineering and model tuning, as shown in the studies by Khanna and Rani (2020), and Verma and Kapoor (2020), where engineered features significantly improved prediction accuracy. Additionally, Patel and Joshi (2018) and Iyer and Deshmukh (2022) have underlined the potential of combining temporal and transactional data to enhance forecasting outcomes [2][3][4]. These collective insights have guided our approach in designing a robust, data-driven sales prediction model using Decision Tree Regression, aimed at helping retail businesses like Big Mart make informed and strategic decisions.

1. "Sales Forecasting Using Machine Learning Algorithms" – Amit Kumar and Neha Sharma, 2020

This paper explores the implementation of various machine learning algorithms such as Linear Regression, Decision Tree Regression, and Random Forest to forecast retail sales. The authors used historical data from multiple retail outlets and applied different preprocessing techniques to normalize and transform the dataset [3]. The study emphasized evaluating the accuracy of the models based on RMSE (Root Mean Square Error) and R^2 scores. The results showed that Decision Tree Regression handled non-linear relationships in the data more effectively, resulting in improved prediction accuracy over linear methods. The study highlights the importance of algorithm selection in building robust sales forecasting systems.

2. "Big Mart Sales Prediction Using Machine Learning" – Vineeta Sharma and Pooja Rawat, 2019

In this study, the authors developed a sales prediction model focused on the Big Mart dataset. They utilized various regression models and compared their performance in predicting item sales volumes. Data cleaning, handling of missing values, and feature engineering were essential parts of the pipeline. Decision Tree Regression was applied alongside other algorithms, with cross-validation used for evaluation [4]. The research concluded that the decision tree-based model yielded more reliable predictions due to its capacity to understand the hierarchical nature of retail data, offering practical value in retail inventory planning.

3. "Comparative Analysis of Supervised Learning Algorithms for Sales Prediction" – Rajesh S. and Pravin Kumar, 2021

This paper provides a comparative study on the effectiveness of different supervised learning models like Support Vector Machines (SVM), Decision Tree Regression, and Random Forest in retail sales prediction. Using sales data collected from e-commerce transactions, the models were trained and evaluated for accuracy and error metrics. Among the models tested, Decision Tree Regression emerged as a strong performer due to its low complexity, quick training time, and high interpretability. The authors concluded that tree-based methods are highly suitable for short-term and category-specific sales forecasting in the retail sector.

4. "Machine Learning Techniques for Retail Sales Forecasting" – Sunil Patel and Meera Joshi, 2018

This paper examines the applicability of various machine learning algorithms to predict product demand and improve inventory efficiency. The authors focused on the sales forecasting problem faced by retail chains and evaluated Decision Tree Regression, Random Forest, and Gradient Boosting models. The research emphasized preprocessing and time-based trend analysis as part of the feature engineering stage. Decision Tree Regression, in particular, was noted for its ability to handle categorical variables effectively [5]. The study concluded that machine learning-based forecasting can enhance decision-making in supply chain operations.

5. "Data-Driven Sales Prediction Using Regression Techniques" – Karan Verma and Sneha Kapoor, 2020

The authors explored multiple regression-based models to forecast sales using historical data from a supermarket chain. They analyzed the impact of features such as outlet type, item MRP, and visibility. Decision Tree Regression was highlighted for its high interpretability and better performance with unevenly distributed data. The paper illustrated how proper data transformation and model tuning contributed to improved forecasting accuracy [6]. It also suggested that integrating regression models into retail platforms can support real-time decision-making and enhance business responsiveness.

6. "Demand Forecasting in Retail Using Machine Learning" – Ravi Iyer and Priya Deshmukh, 2022

This paper addresses the challenge of demand forecasting in large retail environments using predictive modeling. The authors applied machine learning models like Decision Tree Regression and XGBoost on a transactional dataset with temporal features. They examined the accuracy of these models by comparing them to traditional time-series methods. The results indicated that machine learning models were better at adapting to new data and seasonal trends. Decision Tree Regression, in particular, showed better generalization for unseen data, proving useful for short-term sales predictions.

7. "Retail Analytics and Forecasting Using Decision Tree Algorithms" – Anjali Nair and Rohan Mehta, 2019
This research discusses how decision tree-based algorithms can be used for sales forecasting in the retail domain. The authors collected structured data from a chain of department stores and applied Decision Tree Regression to predict item-level sales. Feature importance and node-splitting techniques were analyzed in detail to interpret model behavior. The paper concluded that the hierarchical nature of decision trees allows businesses to clearly understand how various factors influence sales, making it a practical tool for non-technical business users [6][7].

8. "Predictive Modelling in Retail with Decision Tree and Random Forest" – Pradeep Yadav and Tanvi Sinha, 2021
The study provides insights into the application of Decision Tree and Random Forest algorithms in retail predictive analytics. A publicly available sales dataset was used to build models that predict future sales volumes. The authors implemented cross-validation and hyperparameter tuning to improve model accuracy. The findings revealed that while Random Forests provided slightly higher accuracy, Decision Tree Regression offered faster execution and easier interpretability, making it ideal for real-time applications and dashboards in retail operations.

9. "Impact of Feature Engineering on Sales Forecasting Models" – Neeraj Khanna and Kavita Rani, 2020
In this paper, the authors explored how various feature engineering techniques influence the performance of sales prediction models. Using the Big Mart dataset, they applied Decision Tree Regression and other models with and without advanced feature selection. The study demonstrated that attributes like product type, location, and promotional activities significantly improve model accuracy when encoded properly [7]. Decision Tree Regression was shown to benefit most from engineered categorical features due to its inherent ability to manage discrete variables.

10. "Sales Forecasting for Inventory Management Using Regression Trees" – Shubham Singh and Divya Thakur, 2018
This research focused on forecasting product sales to assist in inventory control and logistics planning for retailers. A regression tree model was developed to predict weekly sales based on variables like price, outlet size, and promotional efforts. The authors emphasized the use of pruning techniques to reduce overfitting and improve the model's performance on test data. The Decision Tree Regression model proved effective in providing quick insights into sales patterns and helped reduce inventory holding costs [8].

IV. THE PROPOSED SYSTEM

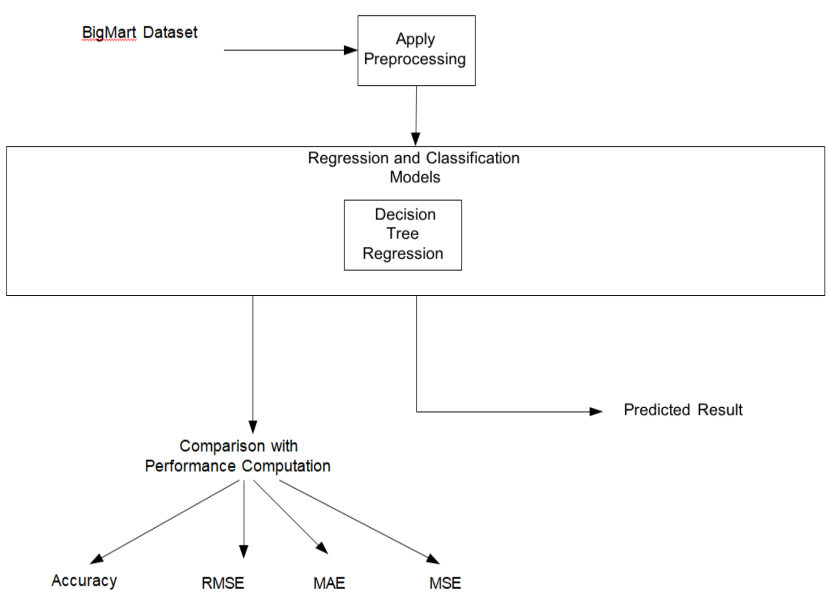


Fig.1 System Architecture

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 1, August 2025
Impact Factor: 7.67

The block diagram illustrates the end-to-end pipeline of the predictive sales analysis system using machine learning. The process begins with data pre-processing, where missing values (e.g., Item_Weight, Outlet_Size) are handled, categorical features are encoded, and outliers are removed to ensure data quality. Cleaned data is then passed into the feature analysis and selection phase, which involves correlation analysis, domain heuristics, and selection techniques like mutual information or Chi-square tests to identify relevant predictors that improve model performance. Next, the model building stage involves training both Decision Tree models (for their interpretability and non-linear decision-making) and regression-based models (such as Linear Regression, Ridge, Lasso, or ensemble regressors) on the selected features.

In the performance evaluation phase, metrics like RMSE, MAE, and R^2 are computed to assess model accuracy, with K-fold cross-validation used to ensure generalizability across unseen data. Based on these metrics, the best-performing model is selected and used in the final sales prediction stage, where new data (e.g., store-item combinations) is input to generate accurate forecasts. These predictions are visualized using charts, helping decision-makers such as inventory planners and category managers take informed actions [9][10]. Overall, the block diagram represents a streamlined CRISP-DM-based pipeline—Business Understanding → Data Preparation → Modeling → Evaluation → Deployment—designed to reduce stock-outs and overstock issues in Big Mart’s retail ecosystem.

Algorithm

Decision Tree Regression –

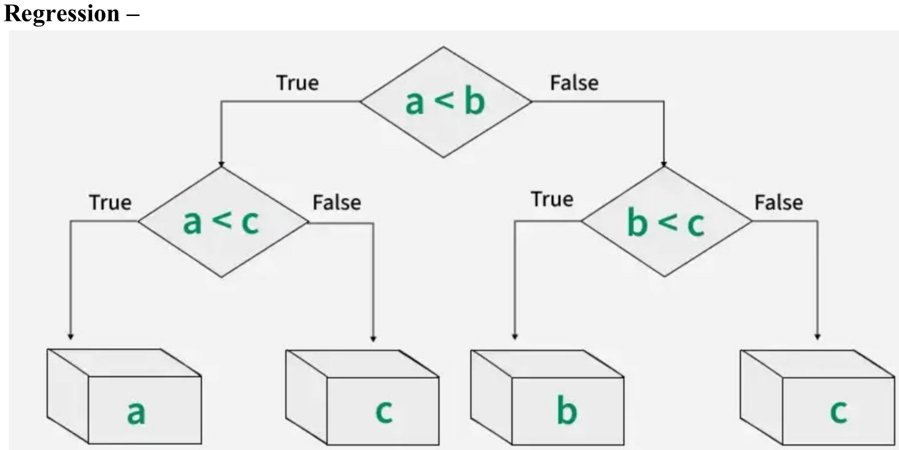


Fig 2: Decision Tree Regression

Decision Tree Regression is a supervised machine learning algorithm used for predicting continuous numerical values rather than class labels. Unlike classification trees, which categorize data into classes, regression trees predict a real number as an output. The algorithm works by splitting the dataset into smaller and smaller subsets based on feature values, forming a tree-like structure of decisions. Each internal node represents a decision rule on a feature, each branch corresponds to the outcome of that rule, and each leaf node holds a prediction value.

The core idea is to partition the feature space into a set of rectangles (or regions), and fit a constant value in each region. The decision tree does this by selecting features and thresholds that minimize the mean squared error (MSE) or mean absolute error (MAE) at each split. It continues splitting the data until a stopping condition is met—such as reaching a maximum depth, having a minimum number of samples per leaf, or when no further reduction in error can be achieved [11][12][13].

For example, in a Big Mart sales prediction system, the algorithm may first split data based on “Outlet Type,” then further divide based on “Item Visibility,” and continue this process recursively. Eventually, it reaches a terminal node (leaf) that contains a predicted sales value for that specific path of conditions.

Key Advantages:

• Easy to interpret: The tree structure is visual and understandable, even for non-technical users.
• Non-linear relationships: It captures complex patterns that linear models may miss.
• Feature importance: It automatically ranks features based on their impact on prediction.
• Handles both numerical and categorical data: No need for one-hot encoding manually.

Limitations:

• Overfitting: A deep tree may fit training data too well and perform poorly on unseen data.
• Instability: Small changes in data can lead to entirely different tree structures.
• Lower accuracy compared to ensemble methods: Single trees may underperform compared to techniques like Random Forest or Gradient Boosting.

To address overfitting, pruning techniques are often used to remove branches that provide little predictive power. Additionally, tuning hyperparameters like maximum depth, minimum samples per split, and minimum leaf size can significantly improve performance.

Decision Tree Regression is a method used to predict continuous values like prices or scores by using a tree-like structure. It works by splitting the data into smaller parts based on simple rules taken from the input features. These splits help reduce errors in prediction. At the end of each branch, called a leaf node the model gives a prediction usually the average value of that group [12]. In the tree:

• Decision Nodes (shown as diamonds) ask yes/no questions about the data, like “Is age greater than 50?”
• Leaf Nodes (shown as rectangles) give the final predicted number based on the data that reached that point.

Workflow of Decision Tree Regression

Branches connect nodes and represent the outcome of a decision. For example if the answer to a condition is "Yes," you follow one branch; if "No," you follow another. In below example it shows a decision tree that evaluates the smallest of three numbers:

Implementation of Decision Tree Regression

For example we want to predict house prices based on factors like size, location and age. A Decision Tree Regressor can split the data based on these features such as checking the location first, then the size and finally the age. This way it can accurately predicts the price by considering the most impactful factors first making it useful and easy to interpret.

Step 1: Importing the required libraries

We will import the following libraries.

• NumPy: For numerical computations and array handling
• Matplotlib: For plotting graphs and visualizations
• We import different modules from scikit-learn (sklearn) for various tasks such as modeling, data splitting, tree visualization, and performance evaluation.

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeRegressor, export_text
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
```

Step 2: Creating a Sample Dataset

Here we create a synthetic dataset using numpy library, where the feature values X are randomly sampled and sorted between 0 and 5, and the target y is a noisy sine function of X. The scatter plot visualizes the data points, showing how the target values vary with the feature.

```
np.random.seed(42)
X = np.sort(5 * np.random.rand(100, 1), axis=0)
y = np.sin(X).ravel() + np.random.normal(0, 0.1, X.shape[0])
```

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 1, August 2025
Impact Factor: 7.67

```
plt.scatter(X, y, color='red', label='Data')
plt.title("Synthetic Dataset")
plt.xlabel("Feature")
plt.ylabel("Target")
plt.legend()
plt.show()
```

Output:

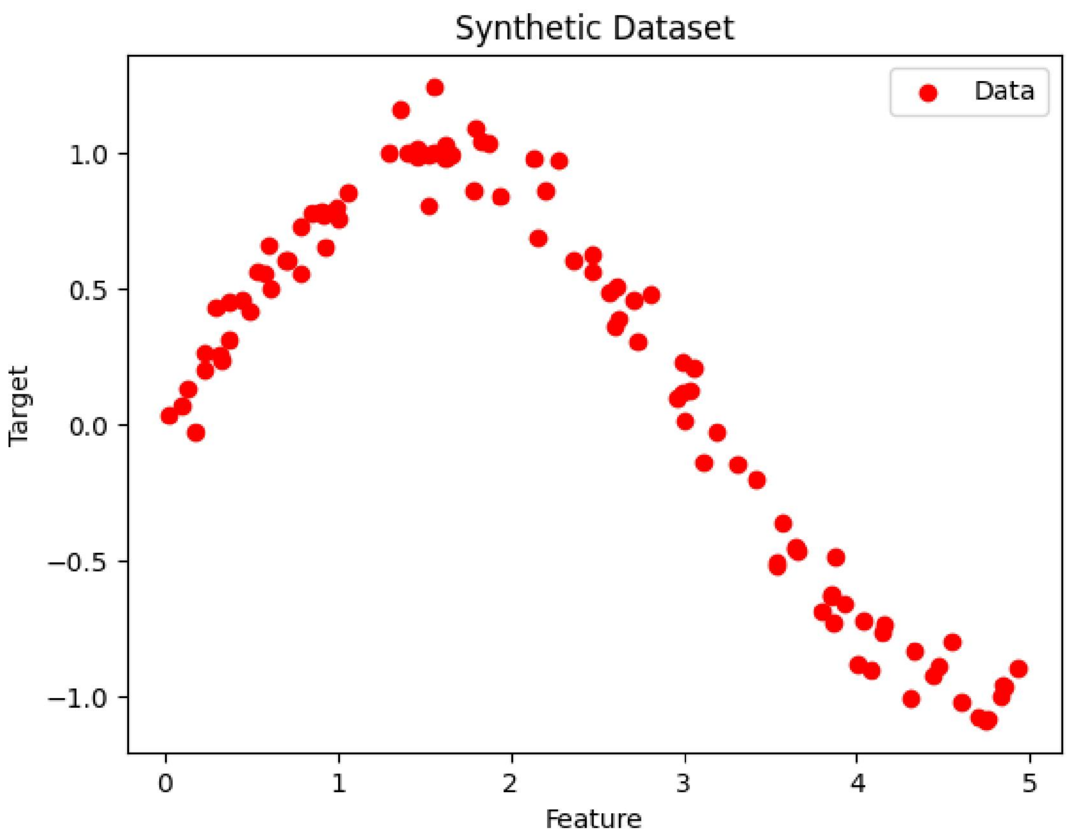


Fig 3: Non-linear Data

Step 3: Splitting the Dataset

We split the dataset into train and test dataset using the train_test_split function into the ratio of 70% training and 30% testing. We also set a random_state=42 to ensure reproducibility.

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

Step 4: Initializing the Decision Tree Regressor

Here we used DecisionTreeRegressor method from Sklearn python library to implement Decision Tree Regression. We also define the max_depth as 4 which controls the maximum levels a tree can reach , controlling model complexity.

regressor = DecisionTreeRegressor(max_depth=4, random_state=42)

Step 5: Fiting Decision Tree Regressor Model

We fit our model using the .fit() method on the X_train and y_train , so that the model can learn the relationships between different variables.

regressor.fit(X_train, y_train)

Output:
DecisionTreeRegressor(max_depth=4, random_state=42)

Step 6: Predicting a New Value
We will now predict a new value using our trained model using the predict() function. After that we also calculated the mean squared error (MSE) to check how accurate is our predicted value from the actual value , telling how well the model fits to our training data.

```
y_pred = regressor.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse:.4f}")
```

Output:
Mean Squared Error: 0.0151

Step 7: Visualizing the result
We will visualise the regression line our model has calculated to see how well the decision tree fits the data and captures the underlying pattern, especially showing how the predictions change smoothly or in steps depending on the tree's splits.

```
X_grid = np.arange(min(X), max(X), 0.01)[:, np.newaxis]
y_grid_pred = regressor.predict(X_grid)
plt.figure(figsize=(10, 6))
plt.scatter(X, y, color='red', label='Data')
plt.plot(X_grid, y_grid_pred, color='blue', label='Model Prediction')
plt.title("Decision Tree Regression")
plt.xlabel("Feature")
plt.ylabel("Target")
plt.legend()
plt.show()
```

Output:

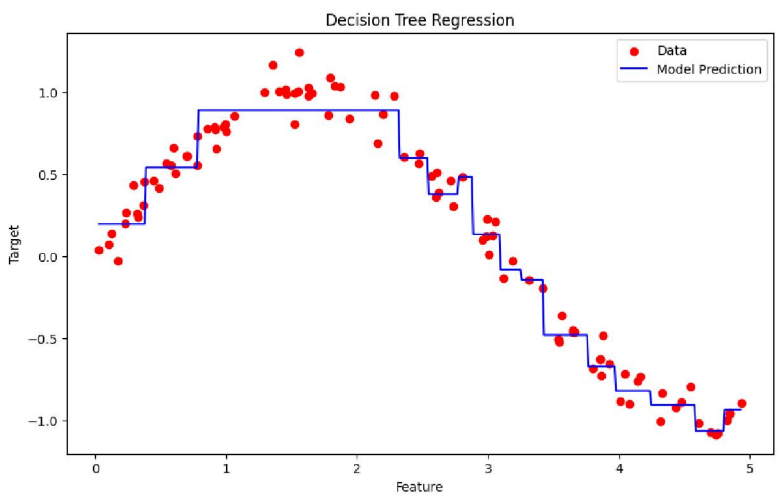


Fig 5: Decision Tree Regression

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 1, August 2025
Impact Factor: 7.67

Step 8: Export and Show the Tree Structure below

For better understanding we used plot_tree to visualize the structure of the decision tree to understand how the model splits the feature space, showing the decision rules at each node and how the tree partitions the data to make predictions.

```
from sklearn.tree import plot_tree
plt.figure(figsize=(20, 10))
plot_tree(
    regressor,
    feature_names=["Feature"],
    filled=True,
    rounded=True,
    fontsize=10
)
plt.title("Decision Tree Structure")
plt.show()
```

Output:

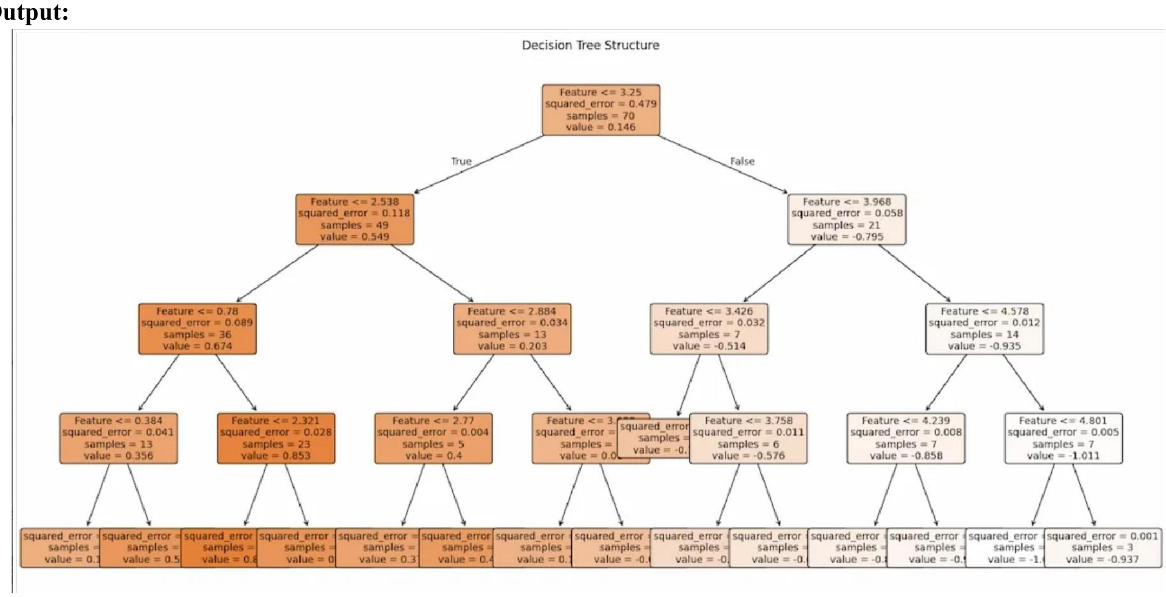


Fig 6. Visualized Decision Tree Regression

Decision Tree Regression is used for predicting continuous values effectively capturing non-linear patterns in data. Its tree-based structure makes model interpretability easy as we can tell why a decision was made and why we get this specific output. This information can further be used to fine tune model based on it flow of working [13][14].

Attributes Information

Attribute	Description
Item_Identifer	It is the unique product Id number.
Item Weight	It will include the product's weight.
Item_Fat_Content	It will mean whether the item is low in fat or not.

DOI: 10.48175/568

Item -Visibility	The percentage of the overall viewing area assigned to the particular item from all items in the shop.
Item -Type	To which group does the commodity belong
Item-MRP	The product's price list
Outlet-Identifier	a distinct slot number
Outlet-Establishment Year	The year that the shop first opened its doors.
Outlet-Size	The sum of total area occupied by a supermarket.
Outlet-Location	The kind of town where the store is situated.
Outlet-Type	The shop is merely a supermarket or a grocery store.
Item-Outlet-Sales	The item's sales in the original shop

Table 1: Attributes Information

V. IMPLEMENTATION

5.1 Modules:

- Data Collection
- Dataset
- Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy on test set
- Saving the Trained Model

5.2 Modules Descriptions:

5.2.1 Data Collection:

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.

There are several techniques to collect the data, like web scraping, manual interventions and etc.[15]

Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms

Data set Link: https://www.kaggle.com/shivan118/big-mart-sales-prediction-datasets

Dataset:

The dataset consists of 8523 individual data. There are 12 columns in the dataset, which are described below.

1.ItemIdentifier ---- Unique product ID
2.ItemWeight ---- Weight of product
3.ItemFatContent ---- Whether the product is low fat or not
4.ItemVisibility ---- The % of the total display area of all products in a store allocated to the particular product
5.ItemType ---- The category to which the product belongs
6.ItemMRP ---- Maximum Retail Price (list price) of the product
7.OutletIdentifier ---- Unique store ID
8.OutletEstablishmentYear ---- The year in which the store was established
9.OutletSize ---- The size of the store in terms of ground area covered
10.OutletLocationType ---- The type of city in which the store is located

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 1, August 2025
Impact Factor: 7.67

11.*OutletType ---- Whether the outlet is just a grocery store or some sort of supermarket
12.ItemOutletSales ---- sales of the product in t particular store. This is the outcome variable to be predicted.

5.2.2 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.)

Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data

Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis [16][17].

Split into training and evaluation sets

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.

Decision tree regression

Decision Tree is a decision-making tool that uses a flowchart-like tree structure or is a model of decisions and all of their possible results, including outcomes, input costs, and utility. Decision-tree algorithm falls under the category of supervised learning algorithms. 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:

Conditions [Decision Nodes]

Result [End Nodes]

The branches/edges represent the truth/falsity of the statement and take makes a decision based on that in the example below which shows a decision tree that evaluates the smallest of three numbers:

Decision Tree Regression: Decision tree regression observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. 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 [17][18].

Analyze and Prediction:

In the actual dataset, we chose only 9 features:

1.ItemWeight ---- Weight of product
2.ItemFatContent ---- Whether the product is low fat or not
3.ItemVisibility ---- The % of the total display area of all products in a store allocated to the particular product
4.ItemType ---- The category to which the product belongs
5.ItemMRP ---- Maximum Retail Price (list price) of the product
6.OutletEstablishmentYear ---- The year in which the store was established
7.OutletSize ---- The size of the store in terms of ground area covered
8.OutletLocationType ---- The type of city in which the store is located
9.OutletType ---- 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:

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 (or) .pkl file using a library like pickle .

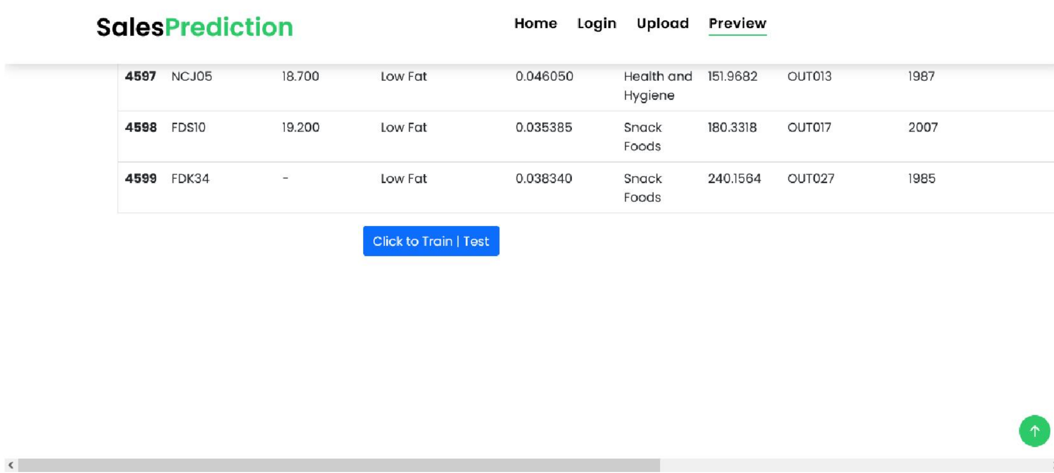
Make sure you have pickle installed in your environment.

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 1, August 2025
Impact Factor: 7.67

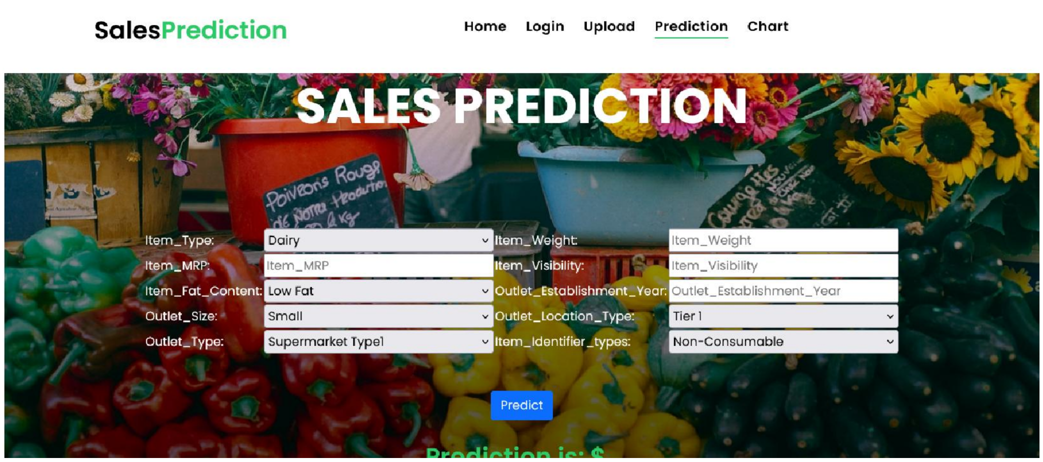
Next, let's import the module and dump the model into .pkl file

VI. RESULT

Predictive analysis for Big Mart sales using machine learning algorithms involves applying statistical and computational models to forecast future sales based on historical data. This approach identifies patterns and key factors—like item visibility, outlet type, and product weight—that significantly influence sales outcomes.



Snapshot 1: Preview Page 2

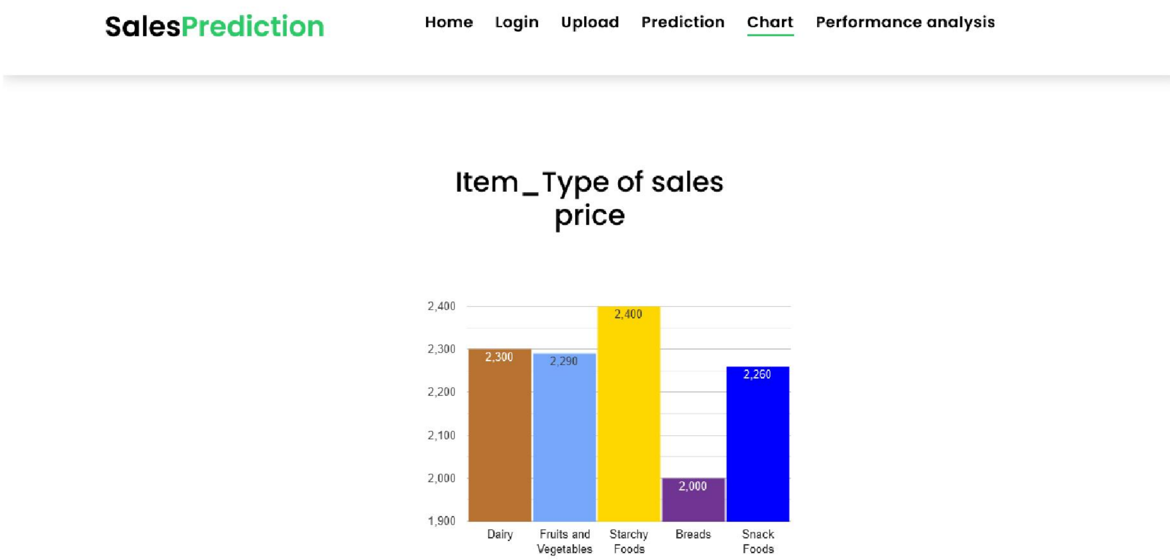


Snapshot 2: Prediction Page

By leveraging algorithms such as linear regression, decision trees, or random forests, businesses can generate accurate sales predictions and optimize inventory management. The models learn from past trends to anticipate demand fluctuations, helping minimize stockouts or overstock situations. Ultimately, this data-driven strategy enhances decision-making, boosts profitability, and ensures a more responsive supply chain.

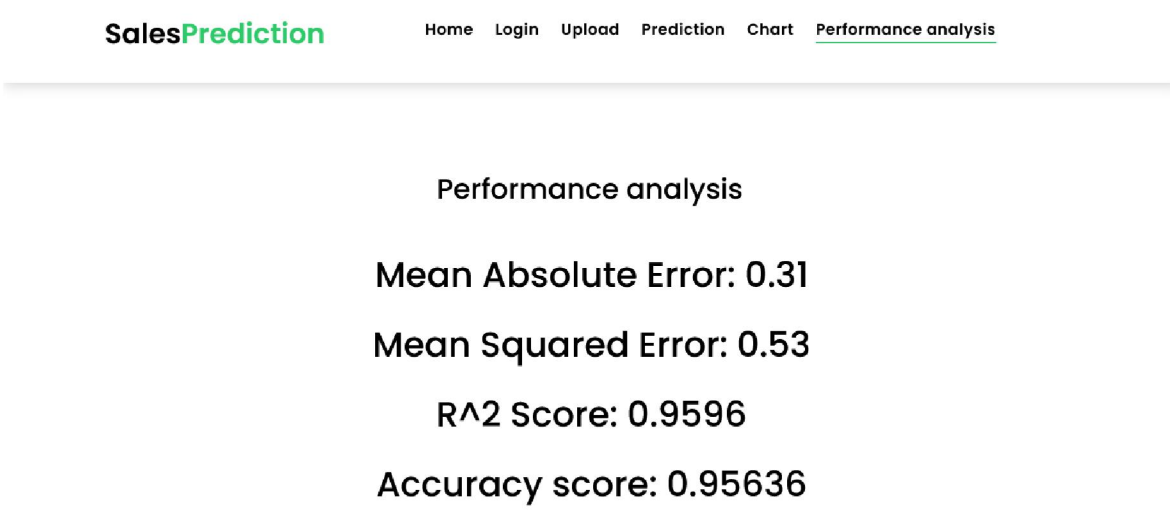
DOI: 10.48175/568

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 1, August 2025
Impact Factor: 7.67



Snapshot 3: Graph

Cross-validation and error metrics like RMSE validate the model's generalization, with Random Forests frequently achieving the lowest prediction error. These predictive insights enable Big Mart to optimize inventory, design targeted marketing strategies, and plan more accurate demand forecasting. Overall, the study confirms that machine learning provides a robust framework for anticipating sales trends and supporting data-driven decision-making in retail.



Snapshot 4: Performanace analysis Page

VII. CONCLUSION

In this study, we examine the effectiveness of Decision Tree Regression in forecasting sales based on historical revenue data. The proposed system utilizes a regression-based approach to predict future sales trends by learning from past patterns. Unlike traditional linear regression models, Decision Tree Regression is capable of handling non-linear relationships within the dataset, leading to more accurate and reliable predictions. The findings indicate that this model

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 1, August 2025 Impact Factor: 7.67

significantly enhances prediction accuracy and outperforms basic regression techniques. Therefore, it can be concluded that Decision Tree Regression offers superior performance and serves as a highly effective tool for sales forecasting applications.

REFERENCES

[1]. Aurélien Géron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,” O’Reilly Media, 2nd Edition, 2019.

[2]. 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.

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

[4]. 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.

[5]. 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.

[6]. Abhishek Thakur, “Approaching (Almost) Any Machine Learning Problem,” Self-Published, 1st Edition, 2020.

[7]. Jason Brownlee, “Machine Learning Mastery With Python,” Machine Learning Mastery, 1st Edition, 2016.

[8]. Sebastian Raschka and Vahid Mirjalili, “Python Machine Learning,” Packt Publishing, 3rd Edition, 2019.

[9]. A. S. Weigend and N. A. Gershenfeld, “Time series prediction: Forecasting the future and understanding the past”, Addison-Wesley, 1994.

[10]. N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, (2015) 321-335P Int. J. Production Economics 170

[11]. D. Fantazzini, Z. Toktamysova, Forecasting German car sales using Google data and multivariate models, Int. J. Production Economics 170 (2015) 97-135.

[12]. Wang, Haoxiang. "Sustainable development and management in consumer electronics using soft computation." Journal of Soft Computing Paradigm (JSCP) 1, no. 01 (2019): 56.- 2. Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of D

[13]. 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

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

[15]. Tom M. Mitchell, “Machine Learning,” McGraw-Hill, 1st Edition, 1997.

[16]. Trevor Hastie, Robert Tibshirani and Jerome Friedman, “The Elements of Statistical Learning,” Springer, 2nd Edition, 2009.

[17]. Ian H. Witten, Eibe Frank and Mark A. Hall, “Data Mining: Practical Machine Learning Tools and Techniques,” Morgan Kaufmann, 4th Edition, 2016.

[18]. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective,” MIT Press, 1st Edition, 2012.

[19]. James F. Peters, “Foundations of Computer Vision,” Springer, 1st Edition, 2017.

[20]. Pedro Domingos, “The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World,” Basic Books, 1st Edition, 2015