

## BIG MART PRICE ANALYSIS AND PREDICTION

Usman Jillani<sup>\*1</sup>, Dr Jawad Ahmed<sup>2</sup>, Asst.Prof.Dr. Fawad Naseem<sup>3</sup><sup>\*1,2,3</sup> Faculty of Computer Science & IT Superior University Gold Campus, Lahore, PakistanDOI:<https://doi.org/10.5281/zenodo.17139886>**Keywords**

Analysis Data” “Machine Learning” “Prediction” “GB Regressor” “KNN Model

**Article History**

Received: 21 June 2025

Accepted: 31 August 2025

Published: 17 September 2025

Copyright @Author

Corresponding Author: \*

Usman Jillani

**Abstract**

Nowadays, every company seeks to increase sales and maximize income, driven by meeting customer demand. Using historical data, machine learning models can help predict future outcomes. This research paper examines how machine learning can be utilized to enhance store sales. The models were applied to predict future prices, supporting inventory, and management decisions such as buying and selling stock. In addition, these models contribute to better business policies and help address real-world business challenges. The Big Mart dataset, collected in 2013, contains information on 1,559 products from 10 stores across different cities. Three regression models were applied, with the Gradient Boosting Regressor achieving the highest R-square value of 74%, demonstrating strong performance in price prediction. The KNN Classifier was also tested: on the non-null dataset, it achieved 85% accuracy, while on the dataset with missing values filled using the mode, it achieved 99% accuracy. However, this study has some limitations. The dataset does not include complete customer information, such as reviews, seasonal effects, or other behavioral factors. As a result, the findings from Big Mart stores may not be fully generalizable to other retail environments. Moreover, incomplete or poor-quality datasets can negatively affect prediction accuracy. For future research, collecting more customer-related information—such as shopping behavior, weather conditions (sunny, fair, rainy), reviews, visits during promotions versus non-promotions, and whether customers shop alone or with family—could improve both the accuracy and reliability of predictions.

**INTRODUCTION**

Nowadays, the world's population is increasing rapidly, which creates greater demand for goods and services. As a result, sellers and retailers are seeking ways to expand their businesses and maximize profits [1]. Large shopping malls, such as Big Mart, play a vital role in meeting the needs of ordinary people who prefer convenience. Instead of visiting multiple stores, customers can purchase all their household items in one place. These superstores save time and money for customers by offering discounts and promotions [2]. Maintaining strong customer relationships is crucial for improving store performance, as satisfied customers return and remain loyal [3]. Big Mart uses advanced computer software to track its inventory and

manage supply and demand efficiently [4]. Sales prediction depends on several factors such as historical data, regular sales, holiday seasons, fuel prices, special events, the buyer price index, and even the unemployment rate in the store's surrounding area [5]. Location also plays a major role, as the population density and whether the store is situated in an urban or rural area directly impact sales [6,7]. Data availability is central for prediction—both historical and real-time data enhance forecasting accuracy [8–10]. Data mining techniques allow stores to refine features, manage data, and create effective algorithmic models for forecasting [11]. A key challenge for big stores is the wide variety of items

with different prices and quantities. Developing a machine learning model that ensures accurate predictions is complex, but high-accuracy models help sellers plan [12,13]. Seasonal shortages, such as cold drinks in summer or dry fruits in winter, highlight the importance of forecasting. Sellers can prepare for these shortages in advance, avoiding losses and ensuring customer satisfaction [14,15]. Customer behavior, shaped by both dependent and independent factors, also influences sales predictions [16]. By analyzing time-series data with machine learning, stores can maximize sales and reduce losses [17,18]. Artificial intelligence (AI) plays a central role in modern prediction systems. Models based on regression, supervised learning, and unsupervised learning help identify relationships in data and improve accuracy [19-21]. Today, machine learning models have become highly advanced, solving complex problems in superstores [22]. While some businesses still rely on simple tools like Microsoft Excel, predictive models offer far greater benefits [23]. These tools streamline the movement of products from factory to customer, optimize inventory, and reduce waste. They also provide insights into customer preferences, seasonality, and product performance [33]. Finally, demand forecasting, inventory optimization, and targeted marketing all contribute to better planning, efficient budget management, and improved business growth [34,35]

#### RELATED WORK

The Big Mart Shopping Centre dataset, collected in 2013, has been widely studied by researchers to explore sales and price prediction through machine learning techniques. Various studies have applied different algorithms and evaluation metrics, each highlighting unique insights into predictive modeling for retail data. Elcio Tarallo (2019) examined multiple algorithms such as XGBoost, Linear Regression, Random Forest, and K-Nearest Neighbors (KNN). His findings showed that ensemble models like XGBoost and Random Forest performed strongly, while traditional models such as Linear Regression also provided reliable results. The study emphasized that algorithm selection and tuning are critical for model efficiency. Similarly, M. L. G. Beher and N. Nain (2020) employed Ridge Regression, Decision Tree,

Linear Regression, and XGBoost. Their results demonstrated a relatively low RMSE of 1052, highlighting the importance of regularization methods like Ridge Regression to avoid overfitting in high-dimensional datasets. Assistant Professor Jyothi (2021) compared AdaBoost Regression and Gradient Boosting Regression. While AdaBoost produced a high RMSE, Gradient Boosting performed significantly better, proving effective in capturing non-linearity and complexity in sales data. Around the same time, Naveenraj R (2021) highlighted the efficiency of Decision Tree, Gradient Boosted Trees, and Random Forest, noting that these ensemble tree-based methods consistently achieved lower RMSE values. His findings supported the superiority of tree-based models in retail prediction. Raizada (2021) focused on KNN Regression, noting its simplicity and competitive performance when applied to clean data. Despite being less complex, KNN proved effective in handling both linear and nonlinear relationships. Vidya Chitre (2022) compared Random Forest, Decision Tree, XGBoost, and Linear Regression, with XGBoost achieving the best R-Square value of 61.4%. She also discussed practical applications of these models in markets, outlets, and grocery inventory. Similarly, Shantanu Choudhary (2022) reported that Random Forest produced an R-Square of 61%, outperforming Linear Regression, and highlighted the influence of store locations and demographics on prediction accuracy. Dr. V. Naga Gopiraju (2023) emphasized that hyperparameter tuning significantly enhanced prediction accuracy, especially across varied outlets. Sugiyanti (2023) further showed that outlier detection and tuning using XGBoost improved results, achieving an R-Square of 57.6%. More recently, Yuhao Yang (2024) achieved the highest R-Square of 62.7% using XGBoost, reinforcing its effectiveness for complex retail data. In a different approach, P. Guru (2024) employed KNN classification, achieving 84.7% accuracy, which demonstrates the versatility of machine learning in categorization tasks. These studies collectively demonstrate that while several algorithms show promise, ensemble methods like XGBoost and Random Forest consistently deliver superior accuracy, especially when paired with careful preprocessing and hyperparameter tuning.

Table 1.1 Related Work Results for Quantitative Features

Publications	Year	Method	Dataset	R-Square
Yuhao Yang [30]	2024	XGB Classifier Model	Supermarket dataset	62.67%
Koh Ya Wen1 [32]	2024	XGB Regressor	Big Mart 2013 dataset	61.81%
Kunal [33]	2024	XGB Regressor	Big Mart 2013 dataset	55.88%
Sugiyanti [12]	2024	XGB Regressor	Big Mart 2013 dataset	57.6%
Aakanksha Ramesh [31]	2023	Gradient Boosting Model	Big Mart 2013 dataset	70%
E. Durga Prasad [28]	2023	XGB Regressor	Big Mart 2013 dataset	58%
M.Saad Bin Ilyas [14]	2023	M5 Model Tree	Big Mart 2013 dataset	58.90%
Shantanu Choudhary [8]	2022	Random Forest Regressor	Big Mart 2013 dataset	61%
Vidya Chitre [1]	2022	XGB Regressor	Big Mart 2013 dataset	61.14%

In Table 1.1, models such as XGB Regressor, Gradient Boosting, Random Forest, and M5 Model Tree were applied to predict sales. Although these models handle large datasets well, their performance here was weak, as all produced low R-squared values. R-squared shows how well a model explains data variation; a low value means poor prediction accuracy. Past studies [32,33,12,31,28,14,8,1] also reported this issue and suggested improving models with better techniques or parameter tuning. Similarly, the XGB Classifier, used for categorization, also showed low accuracy. Future improvements are needed to enhance both regressors and classifiers for reliable predictions.

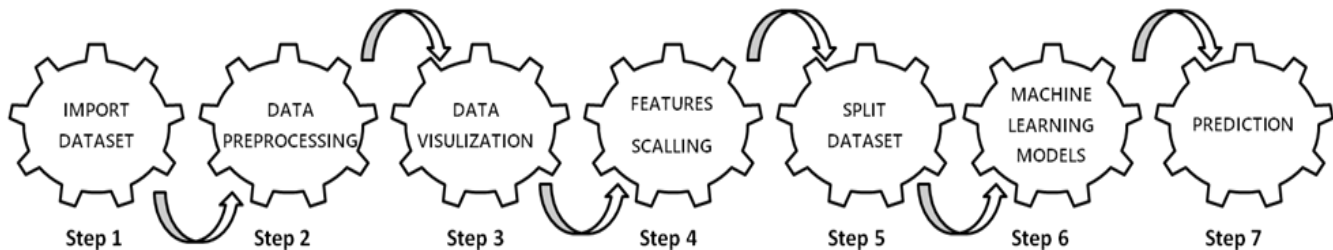
Table 1.2 Related Work Results for Qualitative Features

Publications	Year	Method	Dataset	Accuracy
P. Guru [13]	2024	KNN Classifier	Big Mart 2013 dataset	84.7%

In the Big Mart dataset, KNN achieved 84.7% accuracy, showing strong performance. However, the study ignored its limitations, such as dependence on distance metrics, parameter tuning, scaling, and sensitivity to noise. In large datasets, KNN becomes slow and computationally heavy, raising doubts about the reported accuracy and its generalization

**METHODOLOGY**

The following steps are included in the preparation or obtaining results, and creating models by a flow chart.



**Data Collection:**

Firstly, get the dataset from Kaggle. This dataset link is

<https://www.kaggle.com/datasets/brijbhushannanda1979/bigmart-sales-data>

This dataset has 8523 values and 12 features, of which 7 are qualitative features and 5 are quantitative features. The column names are table below.

Table 1.3 Dataset Features

Qualitative Features	Explanation
Item_Identifier	Product Identification
Item_Fat_Content	Product low fat, high fat or regular-fat
Item_Type	Product Category
Outlet_Identifier	Store Identification
Outlet_Size	Size of Store
Outlet_Location	The type of city in which the store is located
Outlet_Type	Grocery store or supermarket
Quantitative Features	Explanation
Item_Weight	Product Weight
Item_Visibility	The % of the total display area of all similar products in a store
Item_MRP	Product Price
Outlet_Establishment_Year	Store Established Year
Item_Outlet_Sales	Sales of the product in the particular store.

Handling Missing Values:

Table 1.4 Dataset Missing Values

Item_Identifier	0 %
Item_Weight	17.16 %
Item_Fat_Content	0 %
Item_Visibility	0 %
Item_Type	0 %
Item_MRP	0 %
Outlet_Identifier	0 %
Outlet_Establishment_Year	0 %
Outlet_Size	28.27%
Outlet_Location_Type	0 %
Outlet_Type	0 %
Item_Outlet_Sales	0 %

After the dataset was read, we checked the null or missing values in the dataset. By using python we checked the missing values in the dataset so showing the missing values in Table 1.4. This table shows us that Item\_Weight has 17.16% missing values, and it is a quantitative feature; Outlet\_Size has 28.27% missing values, and it is a qualitative feature. Many Data scientists said that if 30% missing values in an individual feature, then drop it if it is not our primary key because 30% above missing values affect the efficiency of the model and result, but in some cases, our feature's missing value is less than 30% so we fill

the missing values by the method of mean and mode. We used the mean of the Missing values in the Dataset. Weight is based on quantitative data, so we used to mean it to fill the missing values, and Outlet is qualitative data, so we used the mode to fill the missing values. After filling the missing values, the dataset fills 100% so now we checked outliers in the dataset by using the statistical method inter-quartile range, which is the difference between the lower quartile and upper quartile outlier of the quantity data with the log method. Drop the two features Item\_Identifier and Outlet\_Identifier.

$$\text{Mean Weight\_Item} = \frac{\text{Sum of all weight item}}{\text{Total item}}$$

Mode Outlet\_Size = " Most repeated value in the outlet\_Size"

**Data Cleaning**

To remove outliers from a dataset, particularly when working with the logarithmic transformation (e.g., np.log1p), you can use statistical methods like the **Interquartile Range (IQR)**. We use the method of log1p. The log1p identifies a small negative number of the floating-point negative zero as it

$$IQR = Q_3 - Q_1 \quad \text{Where } Q_1 = 1\left(\frac{n+1}{4}\right)^{\text{th}} \quad Q_3 = 3\left(\frac{n+1}{4}\right)^{\text{th}}$$

corresponds to the following. These steps increase our results.

The qualitative feature Item\_Visibility and quantitative features Item\_Outlet\_Sales remove outliers using the IQR statistical method.

**Data coding of qualitative features:**

Now we code the qualitative data by its category, and we have done their coding. Now all feature has quantitative values. We have 5 qualitative features; each feature has some category we code it by its category: Item fat content, Item type, Outlet size, Outlet location type, Outlet type. Item\_MRP

**Data Standard Scaling:**

we reshape the data by the method of standard scaling (-1, +1), which applies to quantity features like to apply on item weight, item visibility, item MRP, and item outlet sales. When we condign the data, then it is necessary to standardize scaling it required for balance the dataset. It is the last method in data cleaning. Standard scaling applies to 5 quantitative features: Item\_Weight, Item\_Visibility, Outlet\_Establishment\_Year, and Item\_Outlet\_Sales.

**Data visualization:**

After cleaning the data, we visualize the data both qualitatively and quantitatively, as shown below

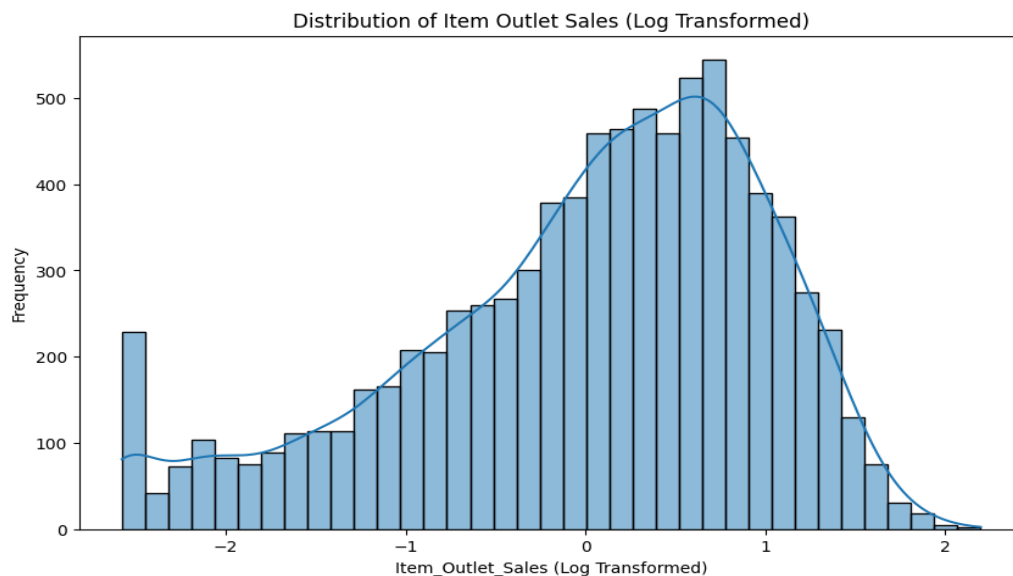
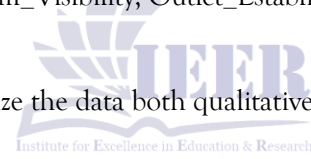


Figure 1.1 Histogram of Item\_Outlet\_Sales

The histogram uses log-transformed sales values on the X-axis and frequency on the Y-axis. Log transformation reduces skewness, making the data smoother and closer to a normal distribution.

This improves prediction accuracy, model strength, and interpretability. Most sales fall between -1 and 1, with a few outliers at extreme values

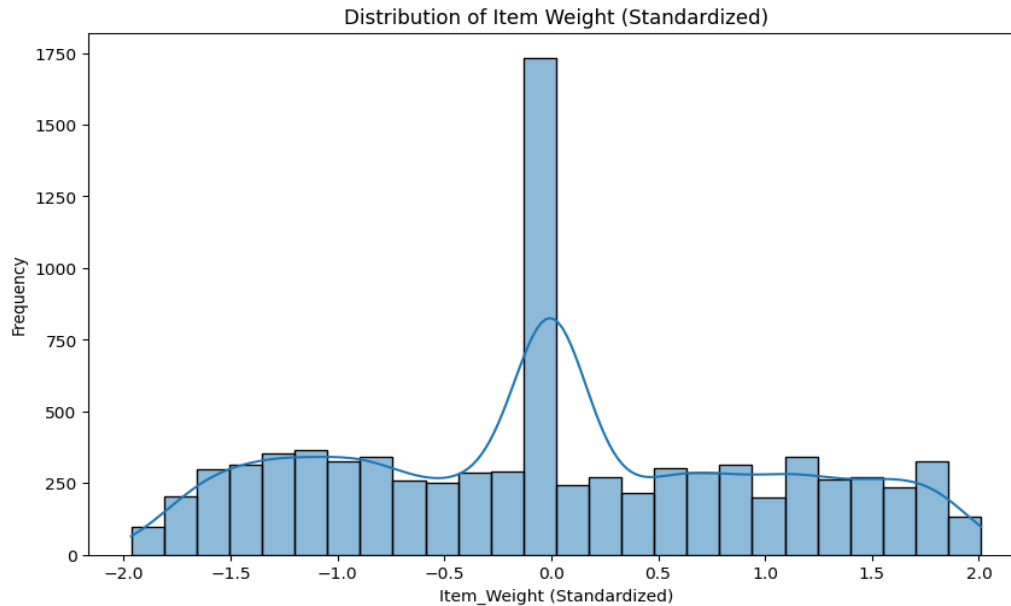


Figure 1.2 Histogram of Item\_Weight

This graph shows item weights after standardization (average = 0, spread = 1). Bars show how often each weight appears. A big spike at 0 may mean missing weights were replaced with averages.

**Splitting the dataset:**

After cleaning, coding, and scaling, we split the dataset into two parts; the first part is based on 70% and the second part is based on 30%. We train the data in 70% and test the data in 30%. We have total values 8523, and then split the data, then we get 5966 values for the Train dataset and 2557 values for the Test dataset.

**Model building:**

After we finish preparing and transforming the data, it's ready to be used for building a prediction model. We train the model using the training data so it can learn to make

predictions. Then, we use the model to predict the target variable using the test. The predictive models are built using the best technique, Lazy Predict, which shows us 39 models. These models have different accuracies, but the best accuracy of the model is the Gradient Boosting Regressor  $R^2 = 74\%$ . This regression model is predictive, and it converts the weak classification into a strong classification. Its speed and accuracy are very impressive, so it is called the Boost model. It boosts the accuracy of data, and easy to predict the dataset.

$$R^2 = 1 - \frac{\text{unexplained variation}}{\text{Total variation}}$$

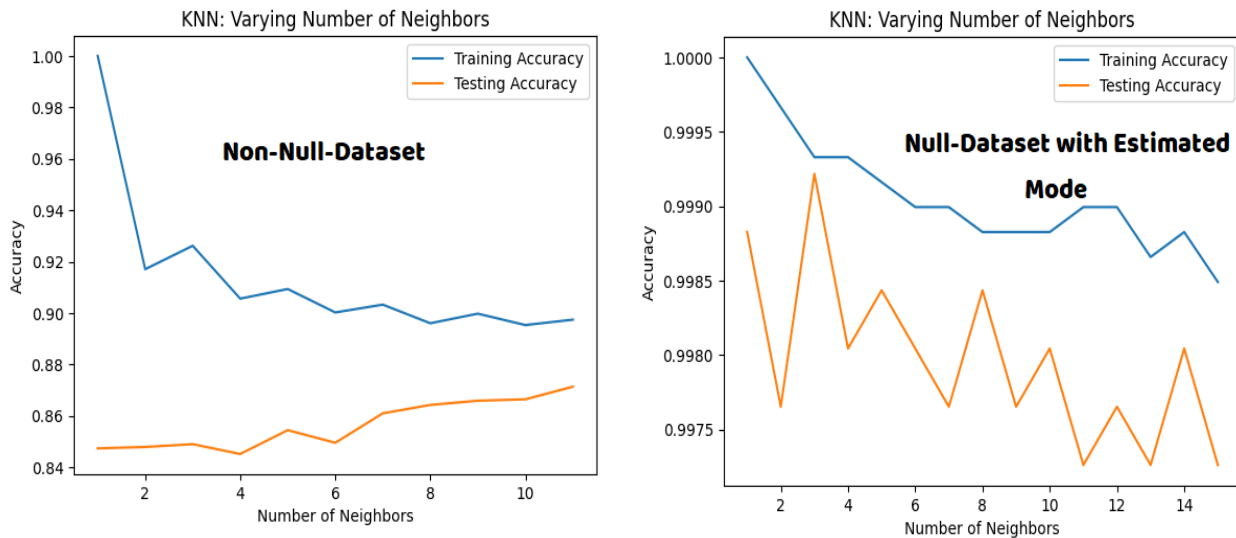
$$\text{Adj } R^2 = 1 - \frac{(1 - R^2)(n - 1)}{(n - k - 1)}$$

**KNN Classifier:**

We apply the KNN Classifier on the non-null dataset, in which 9 features and 6113 rows, our target is 'Outlet\_Size' because the KNN Classifier applies to categorical features, then our model accuracy is 85% with k=12, which is very

high and good after data pre-processing. But when we get a null dataset with 9 features and 2410 rows, we estimated null values with the mode. In this situation, we get higher accuracy, so we conclude that accuracy is higher when estimating missing values in a dataset with k=12, and the accuracy rate is 99% after data pre-processing.

Figure 1.3 Line Plot (Accuracy vs Number of Neighbors in KNN)



RESULTS

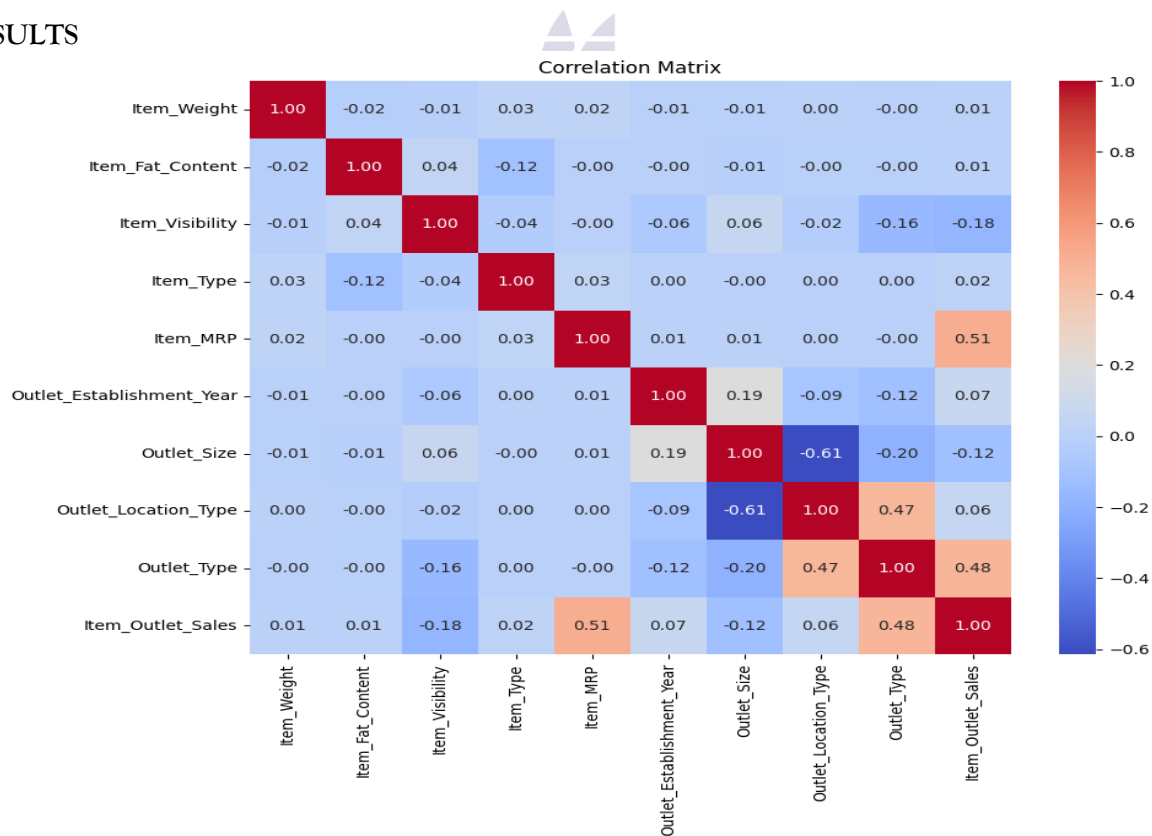


Figure 1.4 Correlation Matrix Between Different Features

This heatmap shows how factors relate to sales. Strongest links: Item MRP (0.51), Outlet Type (0.48), and Location (0.46). Negative links: Item Visibility (-0.18), Outlet Size (-0.12). Other features show little effect. Main drivers: price, outlet type, and location.

Actual vs. Predicted Item\_Outlet\_Sales with Regression Line (Gradient Boosting Regressor)

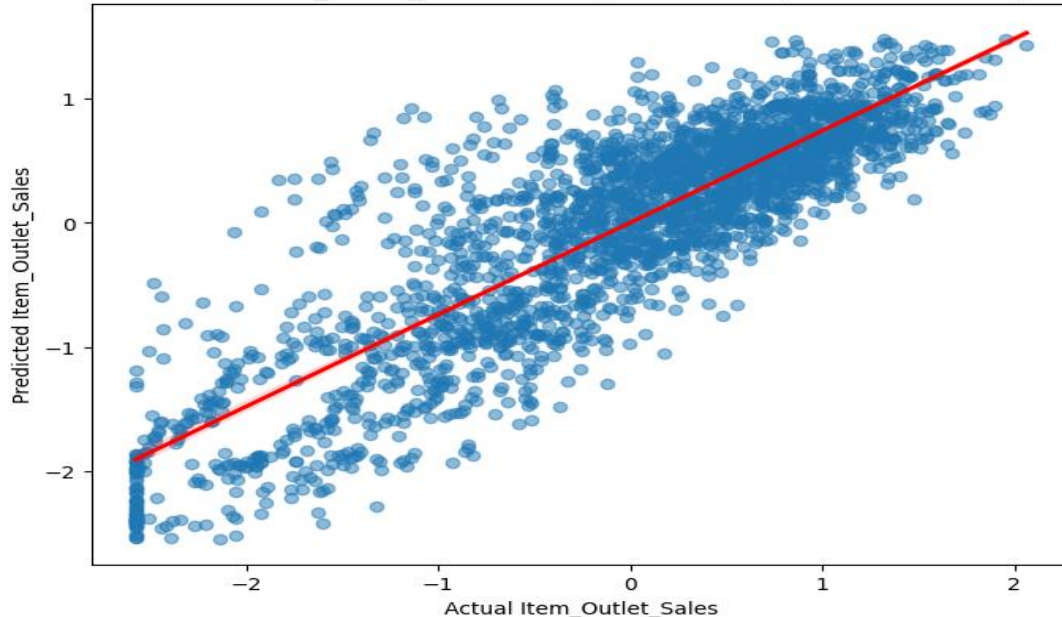
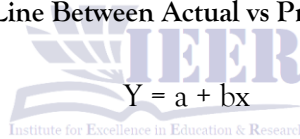


Figure 1.5 Regression Line Between Actual vs Predictive Item\_Outlet\_Sales



$$b = \frac{n \sum XY - (\sum X)(\sum Y)}{n \sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum Y - b \sum X}{n}$$

This scatter plot compares actual vs. predicted sales using Gradient Boosting. Blue dots near the red line show accurate predictions, proving strong model performance. Deviations occur at extreme sales, where errors are larger. Overall, Gradient Boosting captures sales patterns well, though further improvements with feature engineering or ensembles remain possible. For this data set, we will use Machine Learning Tools and techniques based on the attributes of the data set use 3 regression models for prediction. Colab Python Machine Learning models are used for training and Testing of the data for prediction and accuracy.

Table 1.4 Regression Results.

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
GradientBoostingRegressor	74%	74%	0.52	2.88
MLPRegressor	73%	73%	0.52	5.51
LGBMRegressor	73%	73%	0.52	0.17

Table 1.5 Classification Results.

Model	Dataset	Accuracy
KNN Classifier	Non-Null-Dataset	85%
KNN Classifier	Null-Dataset Estimated with mode value	99%

## CONCLUSION

The Gradient Boosting Regressor shows an accuracy of 74%, with an RMSE of 0.52. It is our best model. In many previous research papers on this dataset, the maximum reported accuracy was around 70%, but we improved it to 74%. In the future, a lot of work can be done in this field to further increase accuracy. Deep learning can also be applied to this dataset to achieve even higher accuracy, which will help in making better predictions. The KNN Classifier method also gave us high accuracy—85% with the non-null dataset and 99% accuracy with the null dataset (estimated using mode values). This is a significant improvement compared to previous research papers. Future work should focus on deep learning models, as they have the potential to further improve the accuracy of predicting Big Mart prices.

## REFERENCE

- 1.V. Chitre, "Big Mart Sales Analysis," Int. J. Innov. Technol. Explor. Eng., vol.11, no.5, pp. 8-11, 2022, doi: 10.35940/ijitee.c9833.0411522.
- 2.Kumar, Navneet, and Suraj Choudhari. "BigMart Sale Prediction using Machine Learning." International Journal of Innovative Science and Research Technology, Volume 6, Issue 9, (2021).
- 3.Behera, Gopal, and Neeta Nain. "Grid search optimization (GSO) based future sale prediction for Big Mart." 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2019.
- 4.Silva, Ana Lúcia, and Margarida GMS Cardoso. "Predicting supermarket sales: The use of regression trees." Journal of Targeting, Measurement, and Analysis for Marketing 13.3 (2005): 239-249.
- 5.A.Pant, S. Sharma, M. Bansal, and M. Narang, "Comparative Analysis of Supervised Machine Learning Techniques for AQI Prediction," 2022 Int. Conf. Adv. Comput Technol. Appl. ICACTA 2022, vol. 12, no. 11, pp. 102-110, 2022, doi:10.1109/ICACTA54488.2022.9753636.
- 6.P.Kumar, "BIGMART SALES PREDICTIVE MODEL ANALYSIS USING," vol. VIII, no. I, 2023.
- 7.Syed. Nagoorbee, B. Lakshmi Kanth. "Machine Learning Algorithms for Prediction of BigMart Sales." Journal of Engineering Sciences, Vol 13, Issue 12, 2022.
- 8.S. Choudhary, U. Singh, N. Saxena, and S. Jain, "Big Mart Sales Prediction and Analysis," Int. J. Innov. Res. Technol., vol. 8, no. 12, pp. 1003-1007, 2022, [Online].Available:https://www.ijirt.org/master/publishedpaper/IJIRT155006\_PAPER.pdf
- 9.A. Muneer and S. M. Fati, "A Comparative analysis of machine learning techniques for cyberbullying detection on Twitter," Future Internet, vol. 12, no. 11, p. 187, Oct. 2020, doi: 10.3390/fi12110187.
- 10.Gopalakrishnan, T., Ritesh Choudhary, and Sarada Prasad. "Prediction of Sales Value in Online Shopping Using Linear Regression." 2018 4th International Conference on Computing, Communication and Automation (ICCCA). IEEE, 2018
- 11.S.R Vinayaga and R. Naveenraj, "Prediction of Big Mart Sales Using Machine Learning," Int. Res. J. Mod. Eng. Technol. Sci. www.irjmets.com @International Res.J. Mod. Eng., no. 09, pp. 2582-5208, 2021.
- 12.M. Haris and J. Pusat, "Analysis of the application of hyperparameter tuning in machine learning to increase the accuracy of sales-level," vol. 7, no. 1, pp. 98-106, 2024.
- 13.P. Guru, J. Sathyapriya, K. V. R. Rajandran, J. Bhuvanewari, and C.Parimala, "Product Sales Forecasting and Prediction Using Machine Learning Algorithm," Int. J.Intell. Syst. Appl. Eng., vol. 12, no. 4s, pp. 355-366, 2024.

14. Muhammad saad bin ilyas, Atif Ikram, Muhammad Aadil Butt, and Iqra Tariq "Comparative Analysis of Regression Algorithms Used to Predict the Sales of Big Marts," *J. Innov. Comput. Emerg. Technol.*, vol. 3, no. 1, 2023, doi: 10.56536/jicet.v3i1.53.
15. P. S. Rao, "Forecasting Future Sales of Bigmarts," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 4, pp. 226-228, 2023, doi: 10.22214/ijraset.2023.50059.
16. S. Shetty and S. Shetty, "Big Mart Sales Prediction using Machine Learning," 14th Int. Conf. Adv. Comput. Control. Telecommun. Technol. ACT 2023, vol. 2023-June, no. 5, pp. 1556-1561, 2023, doi: 10.17148/ijarce.2023.124112.
17. B. Yao, "Walmart Sales Prediction Based on Decision Tree, Random Forest, and K Neighbors Regressor," *Highlights Business, Econ. Manag.*, vol. 5, pp. 330-335, 2023, doi: 10.54097/hbem.v5i.5100.
18. R. F. Ali, A. Muneer, A. Almaghthawi, A. Alghamdi, S. M. Fati, and E. A. A. Ghaleb, "BMSP-ML: big mart sales prediction using different machine learning techniques," *IAES Int. J. Artif. Intell.*, vol. 12, no. 2, pp. 874-883, 2023, doi:10.11591/ijai.v12.i2.pp874-883.
19. Bhavana T and Lakshmi K, "Machine Learning Algorithm for Predicting Big-Mart Sales," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, no. 6, pp. 3457-3461, 2022, [Online]. Available: [www.irjmets.com](http://www.irjmets.com)
20. J. Jyothi, U. Chandana Madhuri, B. Ome, S. Srija, J. Sravani, and P. Mounica, "Sales Prediction for Big Mart Outlets," no. 14, pp. 144-148, 2021.
21. A. Keyaben Patel, N. Kumar, and S. Choudhari, "BigMart Sale Prediction using Machine Learning," *Int. J. Innov. Sci. Res. Technol.*, vol. 6, no. 9, pp. 427-428, 2021, [Online]. Available: <https://www.xajzkjdx.cn/gallery/423-april2020.pdf>
22. A. Konda, R. Bandaru, M. Manchala, K. T. Naraharisetty, and A. S. Thankachan, "Predictive analysis for big mart sales using machine learning," *AIP Conf. Proc.*, vol. 2754, no. 1, pp. 1416-1421, 2023, doi: 10.1063/5.0161053.
23. X. Dairu and Z. Shillong, "Machine Learning Model for Sales Forecasting by Using XGBoost," 2021 IEEE Int. Conf. Consum. Electron. Comput. Eng. ICCECE2021, no. Iccece, pp. 480-483, 2021, doi:10.1109/ICCECE51280.2021.9342304
24. Elcio Tarallo, Getúlio K. Akabane, Camilo I. Shimabukuro, Jose Mello, Douglas Amancio, "Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research" <https://www.researchgate.net/publication/338172563> 3-2019.
25. Raizada, Stuti, and Jatinder Kumar R. Saini. "Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting." *International Journal of Advanced Computer Science and Applications* 12.11 (2021).
26. E. D. Wahyuni, A. A. Arifiyanti, and M. Kustyani, "Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining," *Pros. Nas. Rekayasa Teknol. Ind.. dan Inf. XIV Tahun 2019*, vol. 2019, no. November, pp. 263-269, 2019, [Online]. Available: <http://journal.itny.ac.id/index.php/ReTII>
27. E. Durga Prasad, U. Harsha Vardhan, M. Dhiraj Yadav, Dr. Krishna Samalla, Y. Srinivasulu, "Big Mart Sales Prediction using Machine Learning and Python," *International Journal of Creative Research Thoughts (IJCRT)*, Volume 11, Issue 5, May 2023, ISSN: 2320-2882, Pages 773-778.

- 28.Hrishabh Upadhyay, Sushant Shekhar, Anurag Vidyarthi, Rishi Prakash, Gower."Sales Prediction in the Retail Industry Using Machine Learning: A Case Study of BigMart." 2023 International Conference on Electrical, Electronics, Communication, and Computers (ELEXCOM)
- 29.Y. Yang, "Supermarket Sales Prediction Based on Xgboost Classifier Model," vol. 88, pp. 169-173, 2024.
- 30.A. Jadhav, "Machine Learning for Sales Prediction in Big Mart," SSRN Electron. J., pp. 2-6, 2023, doi: 10.2139/ssrn..4530624.
- 31.Wen, Koh Ya, Minnu Helen Joseph, and V. Sivakumar. "Big Mart Sales Prediction using Machine Learning." EAI Endorsed Transactions on the Internet of Things 10(2024).
- 32.Kunal, & Singh, Muskaan & Sharma, Harshit & Roy, Shweta & Agarwal, Kadambri. (2024). Big Mart Sales Analysis. 554-561.  
10.1109/CCICT62777.2024.00092.
- 33.S. V. Chandu, R. R. Anumula, P. Chandu, and U. S. Varri, "Evaluating the Effectiveness of Machine Learning Algorithms for Network Intrusion Detection," *Commun. Comput. Inf. Sci.*, vol. 2333 CCIS, no. 04, pp. 325-344, 2025, doi: 10.1007/978-3-031-83783-8\_19.
- 34.A. Konda, R. Bandaru, M. Manchala, K. T. Naraharisetty, and A. S. Thankachan, "Predictive analysis for big mart sales using machine learning," *AIP Conf. Proc.*, vol. 2754, no. 1, pp. 580-584, 2023, doi: 10.1063/5.0161053.
- 35.S. Agrawal, I. Agarwal, and K. Khandelwal, "Big Mart Sales Predictive Analysis using ML," vol. 6495, pp. 1-4, 2025, doi: 10.22161/ijaers.ncitsa.2025.27.

