Grenze International Journal of Engineering and Technology, June Issue

Big Mart Sales Prediction using Machine Learning

Saritha Shetty[1] and Savitha Shetty[2]
[1]Nitte deemed to be University/Department of MCA, Karkala, India
Email: shettysaritha1@gmail.com
[2] Nitte deemed to be University/Department of CSE, Karkala, India
Email: shettysavi1@gmail.com

***Abstract*— The traditional method of achieving sales and marketing objectives is outdated and no longer effective at helping businesses keep up with the fast-moving, competitive industry. With the advancement in machine learning, we looked into a number of machine learning techniques in our study that help the sales team design plans for expanding their business. The generated information can then be utilized to forecast possible sales volumes for merchants like Big Mart using a wide range of machine learning techniques. To estimate the output sales, machine learning models are fed this dataset of 8523 tuples and 12 different attributes that was obtained for big mart from the data source named Kaggle. We compared Linear Regression, AdaBoost Regressor and Random Forest regressor algorithms in this paper. Random Forest regressor performed best with the least Mean Squared Error (MSE) value of 1107.9451.**

***Index Terms*— sales prediction, adaboost regressor, linear regression, random forest regressor and performance metrics.**

I. INTRODUCTION

In the modern world, large malls and markets gather information about the sales of goods or items with a variety of dependent or independent components as an essential step to help predict future demand and manage inventories. The dataset was created by combining information about the attributes of the items with information about inventory management data and customer records in a data warehouse. The data is then altered and further examined in order to produce forecasts with greater accuracy and find new discoveries. Thereafter, future sales might be predicted using random forests, AdaBoost Regressor, and simple or complicated linear regression models. Big Mart patterns are carefully examined by data experts in order to develop new hubs. Data scientists test various patterns per shop and product in order to achieve the best results. They first use a machine to predict the transactions of Big Mart. Many businesses that heavily rely on their information base need to be able to predict market trends. Shopping malls and retail establishments are constantly seeking for ways to help store owners draw in more customers throughout the day so that the total volume of sales can be determined for use in inventory management, logistics, and transportation management.

In order to detect characters from the ancient Japanese text Kuzushiji, various machine learning strategies are compared [1]. Visuals of molecules can be segmented using deep learning techniques[2].

Today's businesses find it harder to comprehend and anticipate commercial developments [3]. The number of trades in various Big Mart sectors is estimated using a variety of methods of machine learning, like linear regression, random forest, and AdaBoost, depending on expected client demand. The completion of the transaction will likely rely on the type of business, the population in the neighborhood surrounding the store, and

Grenze ID: 01.GIJET.9.2.426

the location's city because it's entirely feasible for the site of the shop to be in either an urban or a rural area. When examining sales, it is important to consider the store's capacity as well as the demographics and other aspects of the surrounding area. An important factor in a shopping mall's sales forecast is the high demand that each business must meet. It's crucial to take into account the store's capacity as well as the neighborhood's demographics and other factors while examining sales. The enormous demand that each store must meet is a significant aspect in a shopping mall's sales prediction. Employing the mean square root error, the recommended Big Mart methods' performance has been assessed [4].

To better gauge future client demand and plan for it, a store would benefit from having an approach that can reliably forecast sales. In this research, we propose three machine learning-based models for estimating Big Mart's future outlet sales. We discovered that our model yields the best results. The optimal algorithm for predicting sales with the best r2 score, the least MAE, and the least MSE must be found by comparing it to the various other models.

II. Literature Survey

The sales data for each distinct item is presently tracked by superstore run-centers, Big Marts, in order to forecast possible buyer demand and revise inventory control. Due to the fact that sales are the foundation of any retailing firm, accurate sales forecasting and analysis have been carried out by many researchers, which are summarized here. Several applications employ artificial intelligence algorithms.

Akshay Krishna and co-authors have presented a Big Mart sales prediction system that considers a range of forecasting approaches, including Ridge regression, Lasso regression, multiple regression, polynomial regression, etc [5].

Regression strategies to handle tough problems were proposed by Behera and co-authors. In order to forecast a store's weekly sales, they further research, assess, and explore the benefits and drawbacks of several regression-predicated forecasting methodologies [6].

Big Marts monitor specific item sales in order to forecast possible buyer demand and revise control over inventory. Regression, neural networks, decision trees, random forests are just some of the data-driven methods that have been studied extensively for their potential to accurately predict big-box retailer sales. These methods have been shown to be effective at capturing the interplay between multiple variables and making accurate forecasts [7].

Sales forecasting is made possible by the application of statistical methods and automated learning [8].

Gradient Boosted Tree is employed for predicting purchases, or which products will be marketed subsequently and in what quantities [9].

Author's suggested approach outlines an effective technique for converting Braille document to Kannada and vice versa [10]. In the limited resource area of literature, study could be expanded [11].

III. Data Pre - Processing

Data must be created before being utilized in machine learning models since it cannot be used in machine learning algorithms in its natural state due to the way it is currently gathered. This method is employed to resolve issues that the knowledge extractor is still learning about. Preprocessing requires properly formatted and cleaned data.

A. Data Collection

The dataset is crucial for a model to effectively forecast future sales demand in a retail environment. The dataset obtained is Big Mart 2013 sales result which has 12 features with an average of 8523 tuples per feature, giving it a rough size of 102276 instances.

We gathered the Big Mart sales dataset for our research, which is displayed in Table I.

B. Data Exploration

During data exploration, it is necessary to examine the raw dataset in greater depth in order to learn important information. At the step of data processing where minimum values are determined, we learn that missing value characteristics have Nan or zero as the minimal value, which affects the reliability of our forecasts. Because these fields need to be fixed before being supplied to the model, a data cleaning technique is needed to manage them.

C. Data Cleaning

Data examination revealed that Outlet Size, Item Weight, and Item_ Outlet_ Sales were the properties lacking

TABLE I. FEATURES OF DATASET

3	Attribute names	Count
object	Item_Identifier	8523
float64	Item_weight	7060
object	Item_Fat_content	8523
float64	Item_visibility	8523
object	Item_Type	8523
float64	Item_MRP	8523
object	Outlet_Identifier	8523
int64	Outlet_Establishment_Year	8523
object	Outlet_Size	6113
object	Outlet_Location_Type	8523
object	Outlet_Type	8523
float64	Item_Outlet_Sales	8523

values. By replacing missing values with zeros, the mode, and the mean of the corresponding attribute, we can lower the correlation between the input attributes.

D. Encoding categorical data

Since mathematical concepts rely heavily on numerical data for issue solving, we must ensure If there is enough evidence in the form of numbers to back up our claim. in order to make the most of machine learning techniques. The sklearn library's LabelEncoder class was brought in so that we could finish this project. Store information such as "Outlet_Size," "Outlet_Type," and "Outlet_Establishment_Year" are all stored in an encoded format.

E. Feature Engineering

The next step is to select relevant features from the data. This may involve transforming the data, aggregating the data, or creating new features based on the existing data. The goal is to select the most relevant features that are likely to have the greatest impact on the sales predictions.
It was discovered to be challenging to give significance to the Item Identifier property because the unique ID begins with either DR, FD, or NC. To get around this, we've started writing "Drinks," "Food," and "Non-Consumables" instead of "DR," "FD," and "NC." In order to reduce clutter, the attribute Item_Fat_Content was simplified by replacing previously used values with just two new ones: "Low Fat" and "Regular." Finally, we add an additional attribute called Outlet_Years to the dataset in order to calculate a particular outlet's age.

F. Identifying Variables of Interest and Their Dependences

Dependent variables are the goals or output variables that must be evaluated and compared at the end of the day. Independent variables are those qualities or input components that cannot be changed, allowing for the forecasting of outcomes. Independent variables include "Item_fat_content", "Item_weight", "Item_type", "Item_visibility","Item_mrp","Outlet_establishment_year","Outlet_size","Outlet_location_type,"and "Outlet_type." The value of 'Item_Outlet_Sales' is the criterion used to determine statistical significance.

G. Datasets are separated into test and training sets.

To avoid overfitting, we only import a single dataset for both training and evaluation. Therefore, partitioning is performed internally to the same dataset. The model must be trained using the dataset designed for that purpose.The results of a test can be predicted using test datasets.

IV. METHODOLOGY

The dataset is ready to be used to create the predictive model to predict sales for Big Mart after the earlier phases have been completed. In our study we have proposed three machine learning models such as linear regression, AdaBoost Regressor and Random Forest regressor.

A. Linear Regression

The linear regression algorithm is by far the most well-known and commonly utilized one for machine learning. This method is carried out so that a linear relationship may be constructed between the response, also known as

the independent variables, and the target, also known as the dependent variable. The linear regression model is built on the following equation, which serves as the basis for the model:

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \ldots\ldots + \theta_n x_n \quad (1)$$

where y is the variable of interest, θ represents the intercept, x_1, x_2, x_3,...., x_n are independent variables, and the coefficients 1, 2, 3,....., n are associated with each variable individually.
Finding the optimal line of greatest fit between the dependant and dependent variables is the primary goal of this approach. Finding the best possible values for each leads to success. By "best fit," we mean a model in which the projected value is as near to the true value as possible with a small margin of error.
The following equation can be used as a general rule to determine the amount of error present when comparing a set of data points to a fitted regression line.

$$\text{Error} = y - \hat{y}, \quad (2)$$

where y is the observed value.

B. AdaBoost Regressor

Yoav Freund and Robert Schapire developed the Adaptive Boosting algorithm. Specifically used for binary classification, it is one of the earliest and most effective algorithms in the Boosting subfield of machine learning. If the original data is used to appropriately train a regressor, then more copies of the regressor are fitted to the same data using the AdaBoost regression technique. However, the mistake of this forecast is then used to enhance the weight of each instance. As a result, advanced regression models pay closer attention to the challenging cases.
Mathematical summary of the AdaBoost Regressor:
1. To begin, we give each data point a weight of

$$w_i = 1/N \quad (3)$$

In (3) N represents the overall sample count in the dataset.
2. For m=1 to M :
(a) In order to obtain the training samples x_i, take a sample from the dataset using the weights $w_i^{(m)}$.
(b) Create a model for the data using a classifier K_m and all of the training samples x_i.
(c) Compute

$$\epsilon = \frac{\sum y_i \neq K_m(x_i) w_i^{(m)}}{\sum y_i w_i^{(m)}} \quad (4)$$

where y_i represents the value that should be assigned to the target variable based on the ground truth, and $w_i^{(m)}$ represents the weight of the sample i at the m^{th} iteration of the process.
(d) Compute

$$\alpha_m = \frac{1}{2} \ln \frac{1-\epsilon}{\epsilon} \quad (5)$$

(e) Update all weights

$$w_i^{(m+1)} = w_i^{(m)} e^{-\alpha_m y K_m(x)} \quad (6)$$

3. New predictions computed by

$$K(x) = \text{sign}\left[\sum_{m=1}^{M} \alpha_m K_m(x)\right] \quad (7)$$

When used for classification, AdaBoost falls under the realm of supervised machine learning. It involves training several stumps iteratively with feature data(x) and labeled target data(y). After being trained, the model's final predictions are derived from a weighted sum of the predictions made by the stump.

C. Random Forest Regressor

Decision trees are what make up a Random Forest that employs a bagging strategy to improve accuracy. To deal with increased variability and bias, boosting and bagging are two of the most widely used ensemble methods. During bagging, many separate base learners or base models randomly select records from the training dataset. Random Forest's beginner students Data collected by a regressor is used to train a decision tree. Given the significant likelihood of over fitting when training accuracy is high but test accuracy is low when using decision

trees in their entirety, the trees themselves cannot be considered accurate learners. Therefore, we apply a method called bootstrapping, which involves randomly assigning parts of the original data file to each of the decision trees via a combination of row sampling and feature sampling with replacement. For this reason, all of the models have been trained using all of the data files, and when we feed test data into one of the trained models, we average the results from all of the models. Aggregation describes the process by which the separate outcomes are combined. The number of decision trees used in the random forest is the hyper parameter that must be adjusted in this technique.
Let's determine the Gini value of a single branch in a decision tree:

$$Mi_j = w_j C_j - w_{l(j)} C_{l(j)} - w_{r(j)} C_{r(j)} \quad (8)$$

In (8) Mi_j is the weighted number of samples that made it to node j and C_j denotes the entropy of node j Node j's left-split offspring is denoted by l(j), and its right-split offspring by r(j).
The relative weight of each characteristic on a generic learner is then calculated by using the formula:

$$Ni_j = \sum_{j:\text{node j splits on feature i}} Mi_j / \sum_{k \in \text{all nodes}} Mi_k \quad (9)$$

In (9) Ni_j denotes the significance of the characteristic I

The adjusted mean is:

$$\text{norm}Ni_i = Ni_i / \sum_{j \in all\ features} Ni_j \quad (10)$$

At the level of the Random Forest, the relevance of the characteristic is determined by taking the mean across all of the trees.

$$\text{RFO}Ni_i = \sum_{j \in all\ trees} normNi_{ij} / T \quad (11)$$

In (11) $RFONi_i$-feature importance I used the random forest model's entire tree set to derive $normNi_{ij}$, the normalized feature significance for i in tree j, and T, the total number of trees.

V. Results And Discussions

Linear Regression, AdaBoost Regressor and the Random Forest algorithm were only some of the machine learning methods used to forecast sales at various big mart locations. The accuracy of the suggested models is measured using a number of different parameters, consisting of the MSE, MAE, and R2_score (Coefficient of Determination). In equations (12), (13), and (14), the MAE, MSE, and R2_score are each defined.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{n} |y_i - \hat{y_i}| \quad (12)$$

$$\text{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (y_i - \hat{y_i})^2} \quad (13)$$

where y_i represents an observed quantity and y_1 is forecasted.

$$R^2 = 1 - SS_{res} / SS_{tot} \quad (14)$$

where, SS_{res} is the sum of the residual errors and SS_{tot} is the total sum of the errors
Table II shows how the models compare with respect to MAE, MSE, and R2_score. When compared to linear regression and AdaBoost Regressor, Random Forest regressor provides more accurate predictions in terms of MAE, MSE, and R2_score.

TABLE II. COMPARISON OF MAE, MSE, R2_SCORE WITH THE MODEL

Model	MAE	MSE	R2_score
Linear regression	880.999904	1162.4412	0.5041875
AdaBoost regressor	928.108091	1194.4458	0.4765100
Random forest regressor	781.538485	1107.9451	0.5495860

VI. Conclusions

The Big Mart problem of sales forecasting has been tackled after the study's discussion of data processing and modelling methodologies. The major markets desire more accurate forecasting techniques that employ regression-based strategies to safeguard their investment from losses. A company's earnings is directly correlated with how well it predicts its sales. It is crucial to select a model that balances precision and computational effectiveness, as well as to continuously assess and improve the model as new data becomes available. Therefore, we may draw the conclusion from our study that Random Forest regressor offers a better prediction of Big Mart outlet sales than Linear and AdaBoost regression methodologies based on MAE, MSE, and R2_score.

References

[1] CV, D. A., Reddy, K. H., Shetty, J., & Shetty, S. Kuzushiji Character Recognition Using Deep Learning: A Local Binary and Phase Quantization Patterns. *Available at SSRN 4351166*. CV, D. A., Reddy, K. H., Shetty, J., & Shetty, S. Kuzushiji Character Recognition Using Deep Learning: A Local Binary and Phase Quantization Patterns. *Available at SSRN 4351166*.

[2] D. A. CV, K.H. Reddy, J. Shetty & S. Shetty, Kuzushiji Character Recognition Using Deep Learning: A Local Binary and Phase Quantization Patterns. *Available at SSRN 4351166*. CV, D. A., Reddy, K. H., Shetty, J., & Shetty, S. Kuzushiji Character Recognition Using Deep Learning: A Local Binary and Phase Quantization Patterns. *Available at SSRN 4351166*.

[3] A.H. Alharbi, C.V. Aravinda, M. Lin, P.S. Venugopala, P. Reddicherla & M.A. Shah, (2022). Segmentation and classification of white blood cells using the unet. *Contrast Media & Molecular Imaging*, *2022*.

[4] T. K Thivakaran. & M. Ramesh, (2022, September). Sales Data Analysis and Prediction System for Big Mart using Deep Recurrent Reinforcement Principles. In *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 1715-1721). IEEE.

[5] R.F. Ali, A. Muneer, A. Almaghthawi, S.M. Fati, & E.A.A. Ghaleb, (2023). BMSP-ML: big mart sales prediction using different machine learning techniques. *Int J Artif Intell*, *12*(2), 1-1x.

[6] A. Krishna, V. Akhilesh, A. Aich, & C. Hegde, (2018, December). Sales-forecasting of retail stores using machine learning techniques. In *2018 3rd international conference on computational systems and information technology for sustainable solutions (CSITSS)* (pp. 160-166). IEEE.

[7] G.Behera, A. Bhoi, & A.K. Bhoi, (2022). A Comparative Analysis of Weekly Sales Forecasting Using Regression Techniques. In *Intelligent Systems: Proceedings of ICMIB 2021* (pp. 31-43). Singapore: Springer Nature Singapore.

[8] P. Ranjitha, & M. Spandana, (2021, May). Predictive analysis for big mart sales using machine learning algorithms. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1416-1421). IEEE.

[9] T. Deng, Y. Zhao, S. Wang, & H. Yu, (2021, January). Sales Forecasting Based on LightGBM. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 383-386). IEEE.

[10] S. N. Gunjal, D.B. Kshirsagar, B.J. Dange, & H.E. Khodke, (2022, September). Fusing Clustering and Machine Learning Techniques for Big-Mart Sales Predication. In *2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)* (pp. 1-6). IEEE.

[11] S. Shetty, S. Shetty, S. Hegde & K. Pandit, (2019, August). Transliteration of text input from Kannada to Braille and vice versa. In *2019 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (pp. 1-4). IEEE.

[12] S. Shetty, S. Shetty, S. Hegde & K. Pandit (2021). English transliteration of Kannada words with Anusvara and Visarga. In *Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2019* (pp. 349-361). Springer Singapore.