e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

PREDICTION OF BIG MART SALES USING MACHINE LEARNING

Naveenraj R*1, Vinayaga Sundharam R*2

*1Student, Department Of Management Studies, Kumaraguru College Of Technology, Coimbatore, Tamil Nadu, India.

*2Associate Professor, Department Of Management Studies, Kumaraguru College Of Technology, Coimbatore, Tamil Nadu, India.

ABSTRACT

In this study, exploratory machine learning approaches are used to forecast big-box store sales. In general, sales forecasting is crucial for advertising, merchandising, warehousing, and production, and it is done in a variety of organizations. To modify the business strategy to predicted results, the sales estimate is based on Big Mart sales from different stores. Different machine learning approaches may then be applied to forecast possible sales volumes for stores like Big Mart. Machine Learning models such as Linear, Ridge and lasso regression model, Random Forest, Gradient Boosted Decision Tree, AdaBoost regressor, Xgboost, Light Gradient Boosting Machine are used in detailed research of sales prediction. In order to anticipate correct outcomes, data exploration, data transformation, and feature engineering are essential. The method is tested using Big Mart Sales data from the year 2013.

Keywords: Machine Learning, Data Exploration, Sales Forecast, Random Forest, Linear Regression.

I. INTRODUCTION

Forecasting sales always has been a critical area to focus on. In order to maintain the efficacy of marketing organizations, all suppliers must use an efficient and optimum forecasting method. Manual material handling of this work may result in significant mistakes, leading to poor organization management, and, more significantly, that would be time consuming, which is not desired in today's fast-paced environment. The main aim of business sectors is to attract the target audience. As a result, it's critical that the firm has already been capable of reaching this goal through the use of a prediction model.

Big Mart is a massive network of stores that spans the globe. Big Mart's trends are extremely important, as data scientists analyze them by product and location to identify potential centers. Using a computer to predict Big Mart sales allows data scientists to explore different patterns by shop and product to get the best results. Many businesses rely largely on their information base and require market forecasting. Forecasting involves evaluating data from a wide variety of sources, including consumer trends, buying behavior, and other considerations. This research would also assist businesses in properly managing their financial means.

And that is where machine learning can really be put to good use. In this paper, we employ data mining approaches including discovery, data transformation, feature development, model construction, and testing to forecast sales using various machine learning algorithms. This approach involves pre-processing raw data acquired by a large mart for missing data, abnormalities, and outliers. After that, an algorithm will be trained to create a model depending on the data.

II. RELATED WORK

(Fawcett, Tom and Foster J. Provost) The method of identifying suspicious behavior using an automated prototype is described in this study. For the purpose of completing this acceptable prototype, many machine learning methods were used. Here, data mining and constructive induction approaches are used to uncover the disparity in cell phone owners' behavior.

(Demchenko et al.) To forecast sales, a generic linear method, a decision tree approach, and a decent gradient approach were employed. The original data set evaluated included a large number of entries, but the final data set utilized for analysis was significantly less than the original since it included non-usable data, duplicate entries, and unimportant sales data.

(Ragg et al.) Many vendors would profit from the forecast of a single transaction rate, as shown in this study, which implies the knowledge collected may be useful for the design of a set-up that would predict a large

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

number of results. The neural network technique is used to make the prediction. They used Bayesian learning to acquire insights in this situation.

(Augusto Ribeiro et al.) A pharmaceutical distribution company's sale forecast is described in this study. The article tackles two issues: one, it conducts stock proration to avoid going out of stock, and two, it focuses on sales forecast to control the quantity of medication stock that the firm must retain in order to minimize customer discontent.

(Cheriyan et al.) This study looks into the judgments that should be made experimental results and the insights gained via data visualization. It made use of data mining methods. The Gradient Boost method has been found to be the most accurate in predicting future transactions.

(Armstrong J) Three modules, hive, R programming, and tableau, were used to forecast sales. By looking at the store's past, you may have a better knowledge of the income and make changes to the objective to make it more successful. To achieve the findings, key values are retrieved inside the diagram to decrease all intermediate values by lowering the intermediate key feature.

(Panjwani et al.) The aim of the study is to provide appropriate findings for predicting a firm's future sales or needs using approaches such as Clustering Models and metrics for sales forecasts. The algorithmic approaches' potential is assessed and employed in further study as a result.

(Manpreet Singh et al.) Inspection of data obtained from a retail store and projection of future store management techniques are carried out in this study. The impacts of numerous sequences of events, such as meteorological conditions, vacations, and so on, may genuinely change the status of various departments, therefore it also analyses and evaluates these effects and their impact on sales.

III. METHODOLOGY

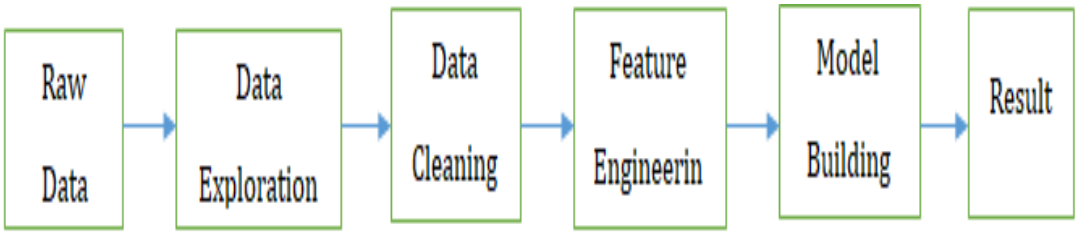

Fig 3.1: Steps followed to obtain the results

IV. DATA VISUALIZATION

Visualizing the data gives a better idea of what that means by placing data in a graphical context such as graphs. This allows us to understand the data more naturally, making it simpler to identify trends, trends, and anomalies in big datasets.

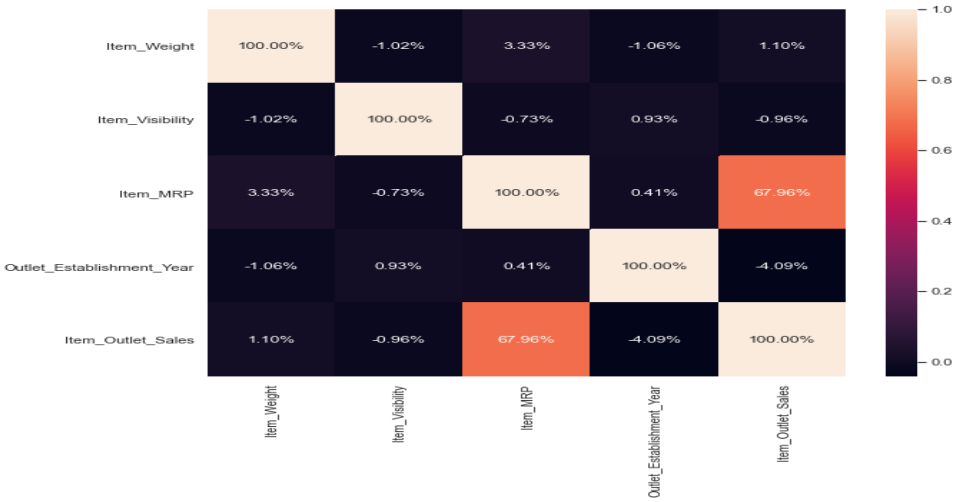

Fig 4.1: Heatmap showing the correlation between attributes

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

The above heatmap shows that lower the dependability of the target variable on the corresponding attribute, higher the intensity of the colour of the attribute in relation to the target variable. Therefore, the Item_outlet_sales relies less on Item_visibility and more on Item_MRP.

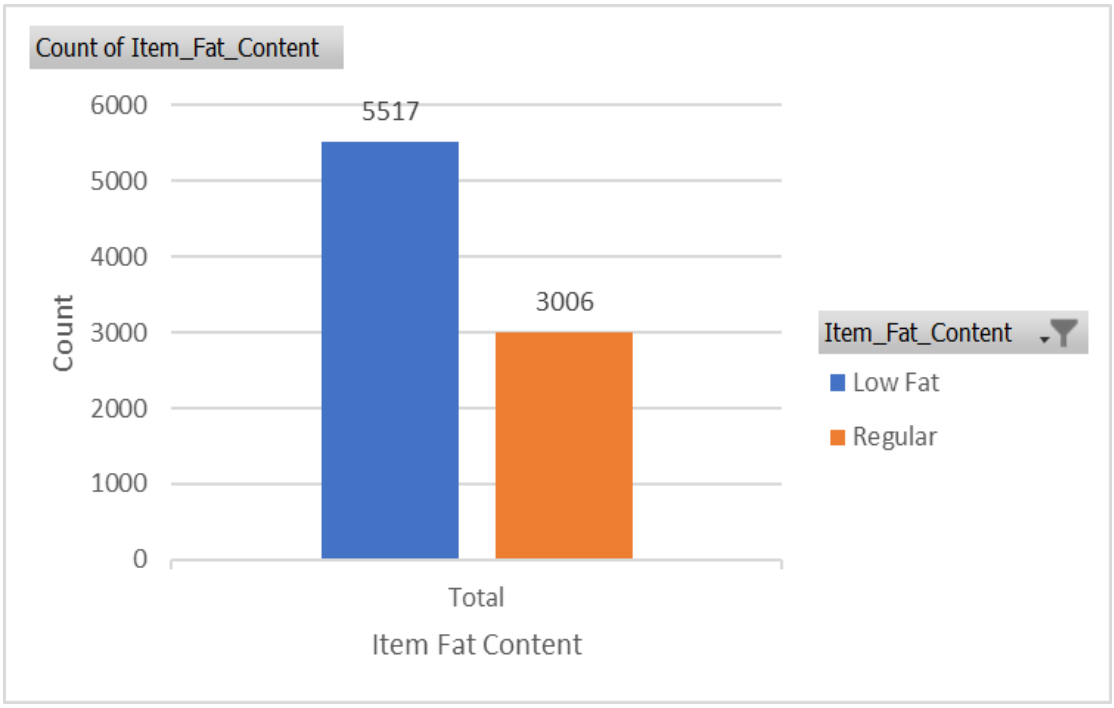

Fig 4.2: Number of items having different fat content

The above graph is plotted to indicate the fat content among different items and it indicates that majority of the items have lower fat content.

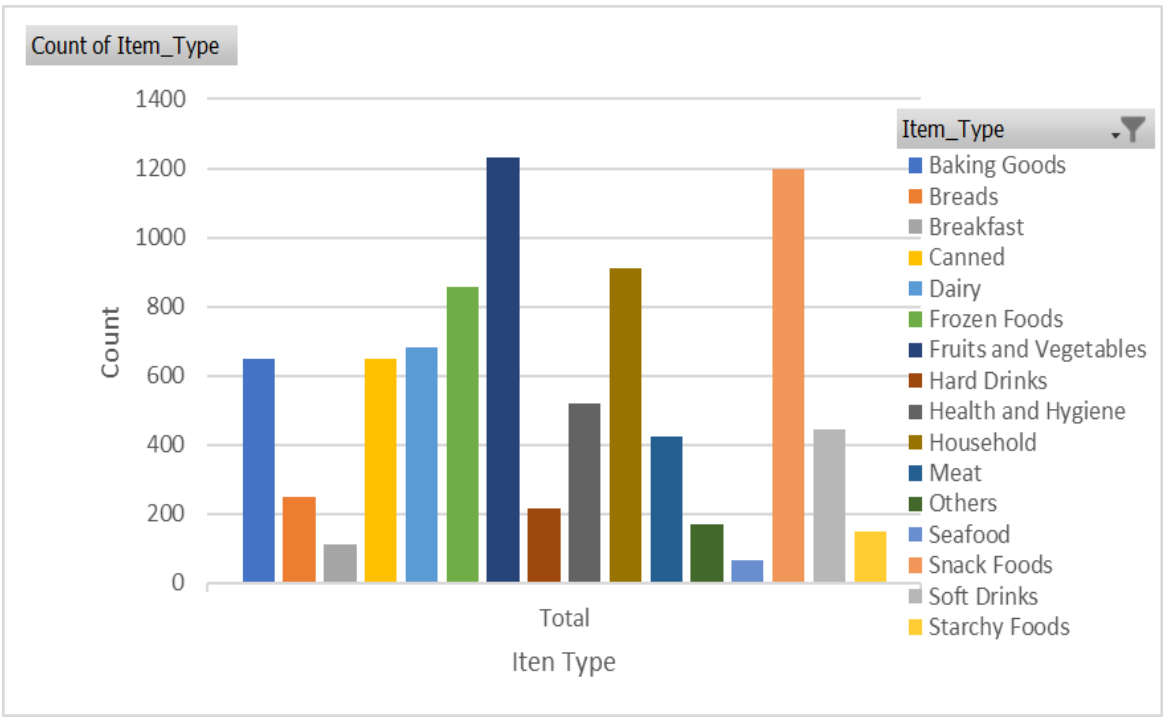

Fig 4.3: Chart showing number of items in each type

The above graph indicates the different types of items that are in the outlet. This shows the wide range of products that are available in the outlet. Fruits and snacks are the most preferred items from the outlet.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

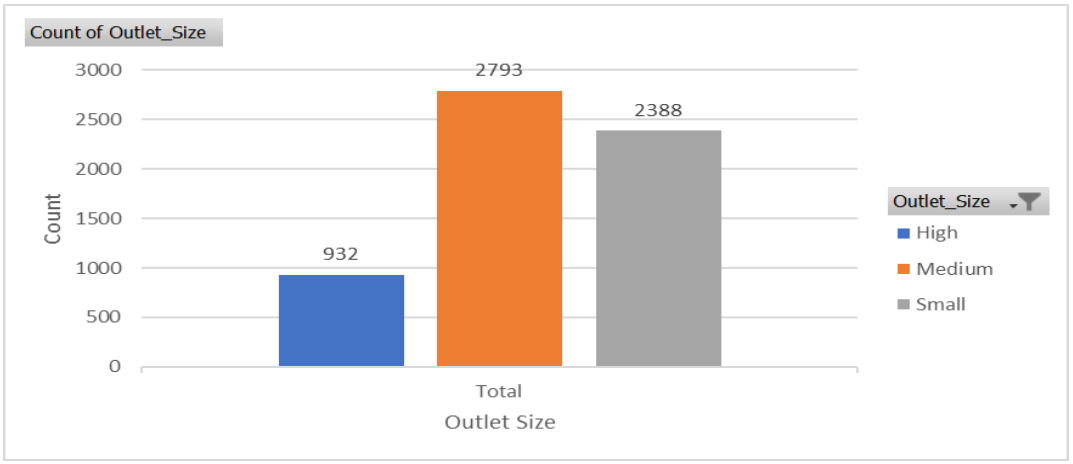

Fig 4.4: Chart showing different outlet sizes

The above graph indicates the size of the outlet. This clearly indicates that most of the outlet size are either small or medium. Only a few of the outlets are large in size.

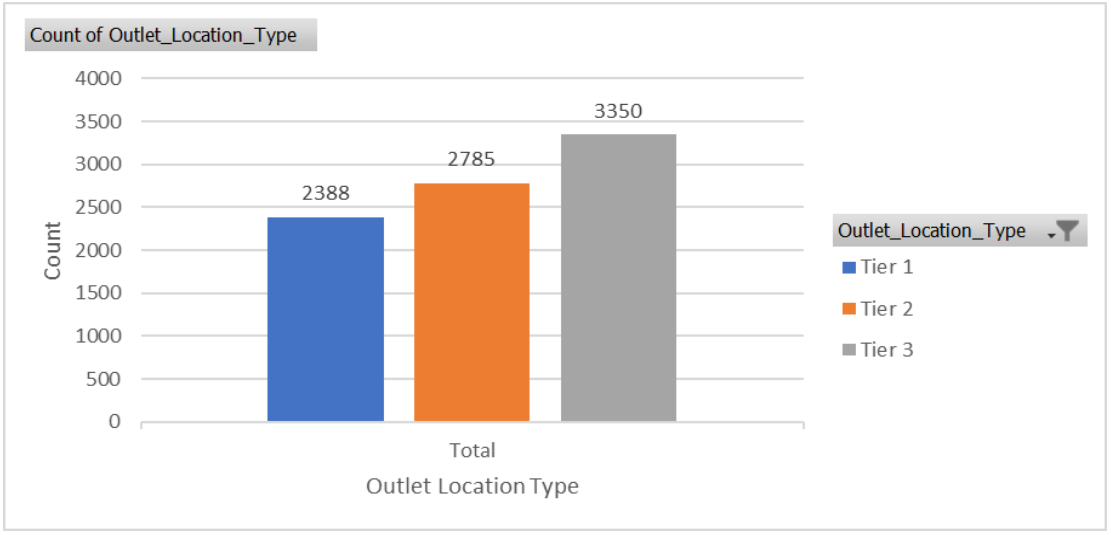

Fig 4.5: Chart showing number of outlets in different locations

The above graph represents the location type the outlet is present. It is classified as Tier1, Tier 2 and Tier 3. Most of the outlets are present in Tier 3 cities.

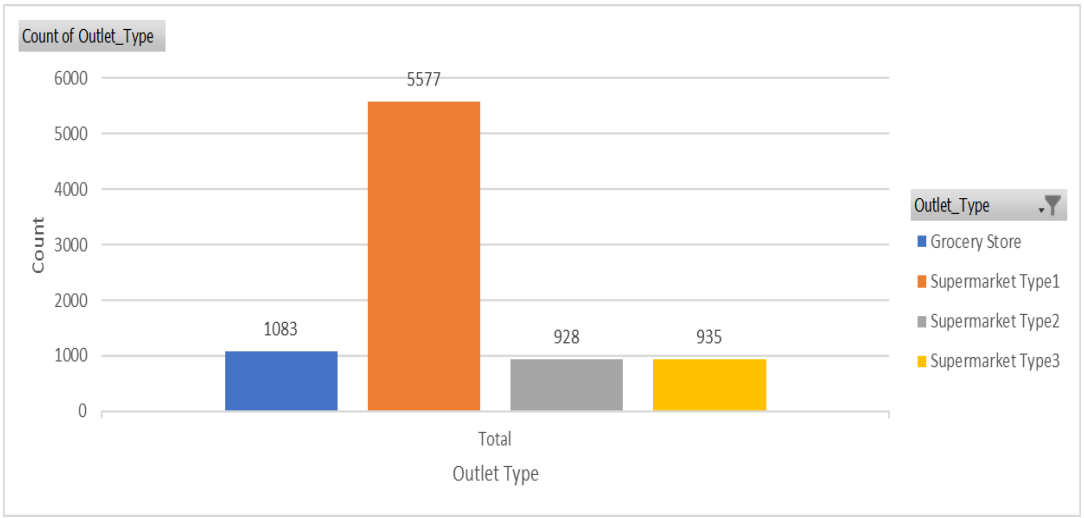

Fig 4.6: Chart showing different outlet types

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

The above graph represents the different type of outlets that are present. The graph shoes that among the different types like Supermarket Type 1, Supermarket Type 2, Supermarket Type 3, Grocery Store that Supermarket Type 1 is the most common type of outlet.

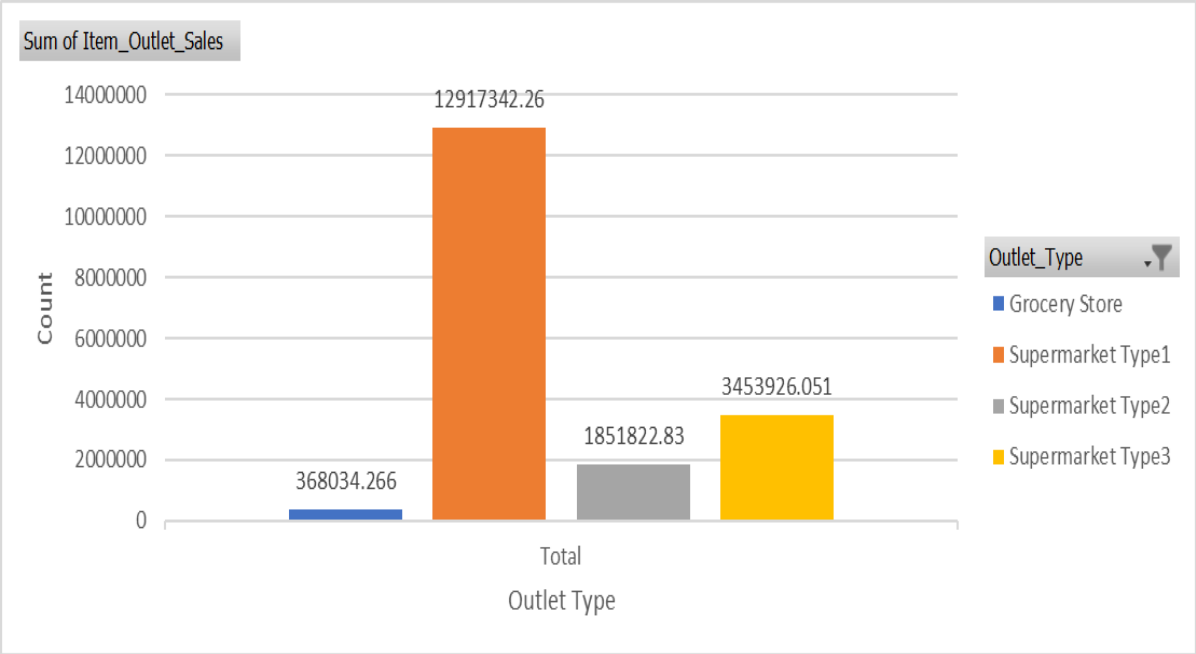

Fig 4.7: Effect of Outlet type on outlet sales

The above graph represents the sales revenue of the different type of outlets. The super market type 1 has the best sales revenue among the other outlet types and it shows that the type of the outlet has significant effect on the sales revenue.

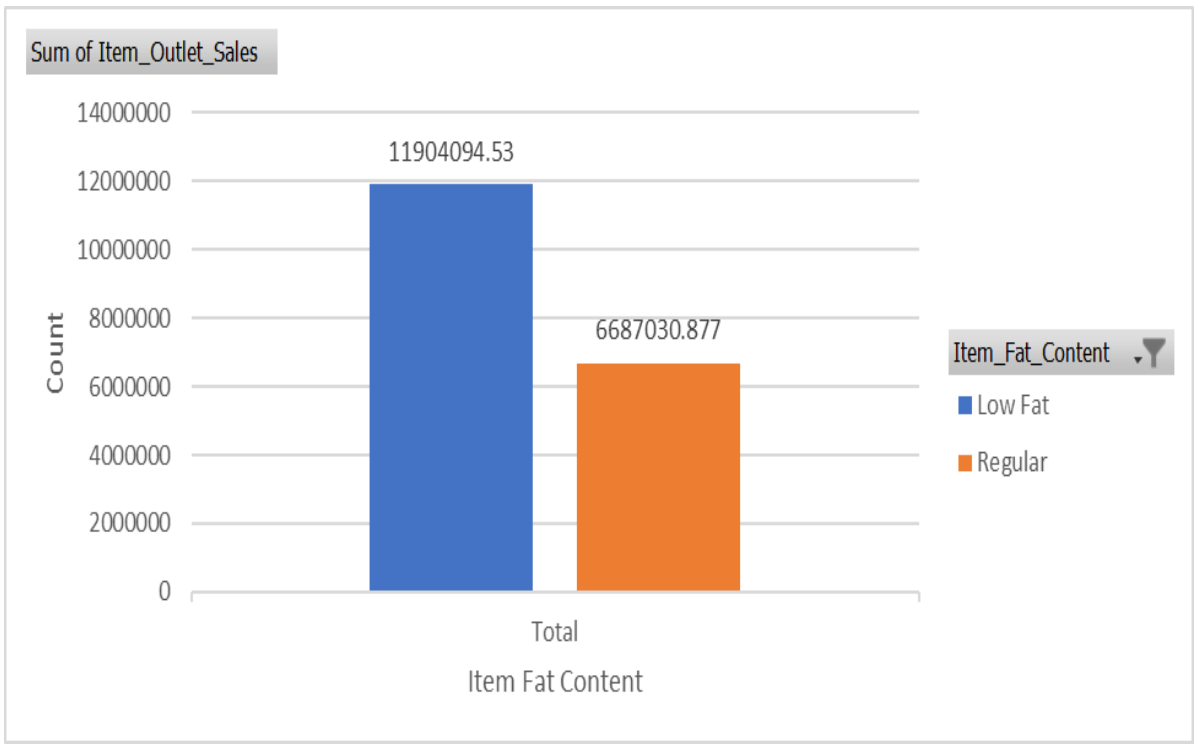

Fig 4.8: Effect of item fat content on outlet sales

The above graph represents the sales revenue of items having different fat content. From the graph we can conclude that low fat content results in high Sales revenue.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

V. DATA PREPROCESSING

During visualizing the data in the data visualization phase, it is found that the attributes Item Weight and Outlet Size have missing values. Pre-processing of data is required in order to fill the data with missing values so that it can adopt to a machine learning model which helps in increasing the efficiency of the model. The values that are missing which corresponds to the Item Weight was filled by averaging the weight of the particular item on the other hand the missing values that corresponds to the outset size was filled by using the mode of the outlet size of a specific type of outlet. Big Mart 2013 sales results were utilized as the dataset, and there are a total of 12 attributes.

VI. FEATURE ENGINEERING

Feature scaling is a technique for converting data into a precise and adaptable size in order to improve the accuracy and reduce error. It essentially stops the algorithm from using a wide variation of data points, allowing us to get better outcomes. In the dataset, the Item Visibility has a least value of Zero, which is desirable because everyone should able to access all the items so the missing values are replaced by taking the mean value of the column. In the Item Outlet Sales, the outliers are often removed or excluded for better performance.

VII. SPLITTING THE DATASET

The dataset was about to split into training and testing set. To avoid the process of overfitting, for train and test, two distinct datasets are not loaded. Therefore, the single dataset is split into train and test sets. The training dataset is the one that we are going to train our model on and the testing dataset is the one which is used to predict the outcome of the test.

VIII. EVALUATION METRICS

The validation of the model is a critical component in developing a successful machine learning model. As a result, it's critical to build a model and receive metrics recommendations from it. It will take time and effort until we reach high precision based on the results of metric improvements. The results of one model are described using evaluation metrics. A key characteristic of the evaluation metrics is the capacity to differentiate between model results. For this study, we employed the Root Mean Squared Error (RMSE) metric. For regression issues, the RMSE is the most widely used evaluating approach. Because of the square root's power, this measure has a lot of variances in percentages. The squared feature of this measure tends to produce more consistent results by preventing the cancellation of positive and negative error values.

IX. MODEL BUILDING

After completing Data Preprocessing and Feature Transformation, the dataset is now ready to build a predictive model. The algorithm is fed into the training set in order to learn how to forecast values. After Model Building a target variable to forecast, testing data is supplied as input. The predictive models are built using

- Linear Regression
- Ridge Regression
- Lasso Regression
- Random Forest
- Decision Tree
- AdaBoost
- XGBoost

9.1 Linear Regression:

One of the most essential and commonly used regression techniques is linear regression. It's one of the most basic regression techniques. The simplicity with which the results may be interpreted is one of its primary merits.

$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon.$

Where Y - Variable to be Predicted

X – Variables used for making a prediction

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:03/Issue:09/September-2021 Impact Factor- 6.752 www.irjmets.com

$\beta_0, \beta_1 ... \beta_r$ - Regression Coeffecients

ε - Random Error

Regardless of how well the model is trained, tested, and validated, there will always be a variation between observed and predicted, which is irreducible error, so we cannot rely entirely on the learning algorithm's predicted results. Data must meet several conditions for a successful linear regression model. One of them is the lack of multiple linear regression, which means that the independent variables should be correlated.

The RMSE value obtained from this algorithm is 1200.37

9.2 Ridge Regression:

When multiple regression impacts results, Ridge Regression is employed. In multiple regression, the least square estimates are objective, but their variances are large and vary from the real value. Ridge regression eliminates standard errors by introducing a degree of bias to regression computations. In Ridge Regression, the Linear Regression Loss function is extended to punish not just the number of square residuals but also the parameter estimations.

The RMSE value obtained from this algorithm is 1200.37

9.3 Lasso Regression:

The Least Absolute Shrinkage Selector Operator regression makes some Coefficients to be zero, given a chance and improves the model. Thus, Lasso regression enables feature selection. Even at small alpha's, our coefficients are reducing to absolute zero. Therefore, Lasso selects only some features while reduces the coefficients of others to zero. This property of Lasso regression is called feature selection.

The RMSE value obtained from this algorithm is 1200.55

9.4 Random Forest:

The random forest algorithm is a highly accurate sales prediction method. It's simple to use and comprehend for forecasting the outcomes of machine learning projects. Random forest classifiers are employed in sales prediction because they have decision tree-like hyperparameters. The tree model is similar to a decision-making tool. A random forest model is created for each individual learner using a random set of rows and a few randomly selected factors. The final forecast may be based on all of the individual learners' guesses. In the case of a regression problem, the final forecast may be the average of all previous predictions.

The RMSE value obtained from this algorithm is 1093.94

9.5 Decision Tree:

It's a simple model with little bias that may be used to create a classifier model, with the root node being the first to be considered in a top-down approach. It is a well-known machine learning model. A decision tree is referred to as a tuple recursive classifier. It is a potent approach for data mining and a powerful method of multi-variable analysis. This approach depicts the variables involved in accomplishing a particular goal, as well as the motivations for obtaining the goal and the means of execution, in a variety of areas.

The RMSE value obtained from this algorithm is 1092.60

9.6 AdaBoost:

AdaBoost is a boosting ensemble approach that works particularly well with decision trees, such as misclassified data points. Learning from prior mistakes is the key to boosting models. By raising the weight of misclassified data points, AdaBoost learns from its mistakes. The steps involved in AdaBoost are 1. Initialize the weights and train a decision tree. 2. Calculate the weighted error rate of the decision tree. 3. Calculate the decision tree weight in the ensemble. 4. Update the weights of the wrongly classified data points. 5. Repeat step 1 and make final prediction.

The RMSE value obtained from this algorithm is 1292.75

9.7 XGBoost:

Decision trees and gradient boosting are used to create the XG Boost method. The algorithm's construction was designed to maximize the efficiency of computation time and memory resources. Boosting is a sequential procedure based on the ensemble concept. This involves a group of low learners and increases accuracy rate. At

every time t, model variables are weighted depending on the impacts of the previous instant. Correctly computed findings are given a lesser weight, whereas incorrectly calculated results are given a greater weight. The XGBoost model uses this method to internally perform stepwise ridge regression, which automatically selects features and eliminates multiple regression.

The RMSE value obtained from this algorithm is 1161.00

X. RESULT

Various machine learning algorithms like Linear Regression, Ridge Regression, Lasso Regression, Random Forest, Decision Tree, AdaBoost, XGBoost have been built to predict the sales revenue of Big Mart. It's been found that the most efficient algorithm to predict the sales revenue of Big mart is observed with Gradient Boosted Decision Tree and Random Forest algorithms having the least RMSE value among other algorithms.

XI. CONCLUSION

The goal of this study is to use machine learning techniques to forecast future sales of Big Mart based on previous year's data. With conventional methods failing to assist businesses in increasing revenue, the application of Machine Learning methodologies proves to be a significant factor in creating company plans, that take consumer purchasing habits into account. Predicting sales based on a variety of criteria, including past year's sales, allows firms to develop effective sales plans and enter the competitive market unafraid.

XII. REFERENCE

[1] Fawcett, Tom, Foster J. Provost. "Combining Data Mining and Machine Learning for Effective User Profiling" In KDD, pp. 8-13.1996.

[2] "Applied Linear Statistical Model", Fifth Edition by Ragg, Thomas, Wolfram Menzel, Walter Baum and Michael Wigbers. "Bayesian learning for sales rate prediction for thousands of retailers." Neurocomputing 43, no. 1-4 (2002):127-144.

[3] Augusto Ribeiro, Isabel Seruca and Natrcia Duro. "Improving Organizational Decision Support: Detection of Outliers and Sales Prediction for a Pharmaceutical Company." Procedia Computer Science, Vol. 121, pp. 282-290, December 2017.

[4] Cheriyan, Sunitha, Shaniba Ibrahim, Sanju Mohanan and Susan Treesa. " Intelligent Sales Prediction Using Machine Learning Techniques." In 2018 International Conference on Computing, Electronics and Communication Engineering (iCCECE), pp. 53 – 58, IEEE, 2018.

[5] Armstrong J. "Sales Forecasting", SSRN Electronic Journal 2008.

[6] Panjwani Mansi, Rahul Ramrakhiani, Hitesh Jumnani, Krishna Zanwar and Rupali Hande. "Sales Prediction System Using Machine Learning." No. 3243. EasyChair, 2020.

[7] Singh Manpreet, Bhawick Ghutla, Reuben Lilo Jnr, Aesaan FS Mohammed and Mohmood A. Rashid "Walmart Sales Data Analysis – A Big Data Analytics Perspective." In 2017 4th Asia-Pacific World Conference on Computer Science and Engineering (APWC on CSE), pp. 114 – 119, IEEE, 2017.