WEB APPLICATION FOR BIG MART SALES PEDICTION USING MACHINE LEARNING

[1]Jangali Induja, [2]Elpula Pranitha, [3]Ettireddy Harish Reddy, [4]Mrs.Adluri Vijaya Lakshmi

Computer Science and Engineering,
B V Raju Institute of Technology, Medak, India

***Abstract*: These days shopping malls and Big Marts keep track of their sales data of each and every individual item for predicting future demand of the customer and update the inventory management as well. These data stores basically contain a large number of customer data and individual item attributes in a data warehouse. Further, anomalies and frequent patterns are detected by mining the data store from the data warehouse. The resultant data can be used for predicting future sales volume with the help of different machine learning techniques for the retailers like Big Mart. In this project, we propose a predictive model using Linear Regression technique for predicting the sales of a company like Big Mart and found that the model produces better performance as compared to existing models. We are developing this project using Machine Learning and with Python programming language. We are developing the web application for the prediction using flask (python)**

***Keywords*: Python, Machine Learning, Flask, Linear Regression, Random Forest.**

I. INTRODUCTION

Every item is tracked for its shopping centers and Big Mart in order to anticipate a future demand of the customer and also improve the management of its inventory. Big Mart is an immense network of shops virtually all over the world. Trends in Big Mart are very relevant and data scientists evaluate those trends per product and store in order to create potential centers. Using the machine to forecast the transactions of Big Mart helps data scientists to test the various patterns by store and product to achieve the correct results. Many companies rely heavily on the knowledge base and need market patterns to be forecasted. Each shopping center or store endeavors to give the individual and present proprietor to draw in more clients relying upon the day, with the goal that the business volume for everything can be evaluated for organization stock administration, logistics and transportation administration, and so forth. To address the issue of deals expectation of things dependent on client's future requests in various Big Mart across different areas diverse Machine Learning algorithms like Linear Regression, Random Forest, Decision Tree, Ridge Regression, XGBoost are utilized for gauging of deals volume. Deals foresee the outcome as deals rely upon the sort of store, populace around the store, a city wherein the store is located i.e., it is possible that it is in an urban zone or country. Population statistics around the store also affect sales, and the capacity of the store and many more things should be considered. Because every business has strong demand, sales forecasts play a significant part in a retail center. A stronger prediction is always helpful in developing and enhancing corporate market strategies, which also help to increase awareness of the market.

II. METHODOLOGY

The steps to be followed in this work, right from the dataset preparation to obtaining results are represented in fig.2.1.

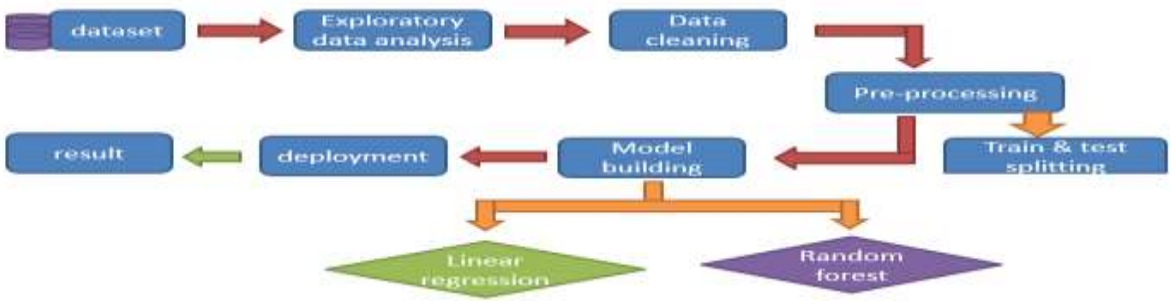


Fig 2.1: steps to be followed to obtain result

Implementation Process:

1. System:

1.1 Store Dataset:

The System stores the dataset given by the user.

1.2 Pre-processing

To treat null values, outliers and Label Encoding.

1.3 Model Training:

The system takes the data from the user and fed that data to the selected model.

1.4 Graphs Generation:

The system takes the dataset given by the user, selects the model and generates the accuracy corresponding to the selected model

2. User:

2.1 Upload Dataset:

The user can load the dataset he/she want to work on.

2.2 View Dataset:

The User can view the dataset.

2.3 Select model:

User can apply the model to the dataset for accuracy.

2.4 Graphs:

User can evaluate the model performance using the graphs.

2.5 Prediction:

Passing parameters to predict the output

III. MODELING AND ANALYSIS

EXPLORATORY DATA ANALYSIS:

It is beneficial to add test data to train data to explore data in every dataset and thus to merge train and test data with a view to data visualization, feature engineering. For the exploratory method, univariate analysis and bivariate analysis are to be conducted to obtain data information. Few observations have been made during the Univariate Analysis and are as follows: The categories 'LF', 'low fat', and 'Low Fat' are the same and 'reg' and 'Regular' are the same category. As a result, they can merge into one, and Low fats are almost twice that of regular items. The main sales in the Item Type column are Fruit and Snack. The variable goal is skewed to the right. These items are not consumable, but all items are labelled either as low fat or regular items. Through the study of Bivariate, a clear relationship between product weight and sales and between item fat content and sales has been found. A significant amount of sales is obtained from products with visibility below 0.2. Individuals have selected a low fat category over other groups. In the relationship between the item identifiers and the outlet size, the items are purchased more frequently as the outlet size increases. The exposure of the item means that more visible items have less sales.

DATA CLEANING:

The attributes Outlet Size and Item weight has missing values. In our work in case of Outlet Size missing value we replace it by the mode of that attribute and for the Item Weight missing values we replace by mean of that particular attribute. The missing attributes are numerical where the replacement by mean and mode diminishes the correlation among imputed attributes. For our model we are assuming that there is no relationship between the measured attribute and imputed attribute. The data cleaning can be implemented with the Klib library.

DATA PREPROCESSING:

Data pre-processing needs to be performed in order to purify data and adapt it to the machine learning model of a system which also makes a machine learning model more accurate and efficient. The first thing for data preprocessing is to collect the required dataset, and then check the missing values once the dataset is imported. Correcting missed values is necessary, or else the data would be difficult to access and maintain. Then calculate the mean of the column containing missing values to rectify the missed values, and substitute it with the measured mean. When the dataset is preprocessed, the dataset is separated into the dataset of train and test. Now, this dataset can be used to train a machine learning algorithm to predict Item Outlet Sales against a variety of items that will help retailers create personalized offers against specific.

MODEL BUILDING:

The dataset is now ready to fit a model after performing Data Preprocessing and Feature Transformation. The training set is fed into the algorithm in order to learn how to predict values. Testing data is given as input after Model Building a target variable to predict. The models are build using:

- Linear Regression
- Random Forest

Linear Regression

The most common and simplest statistical approach for predictive modeling is linear regression. Below is the linear regression equation: Y = _1X1 + _2X2 + ::::_nXn Where X1, X2,..., Xn are the independent variables, Y is the target variable and all the coefficients are the thetas. The magnitude of a coefficient as compared to the other variables determines the importance of the corresponding independent variable. This algorithm's basic principle is to match a straight line between the chosen training dataset features and a constant target variable, i.e. sales. The algorithm chooses a line which fits better with the data. Linear regression performs the task of predicting a dependent variable value (y) based on a given independent variable (x). This regression technique considers a linear relationship between x (input) and y(output).

Random Forest

Random Forest is a tree-based bootstrapping algorithm that combines a certain number of weak learners (decision trees) to construct a powerful model of prediction. For each person learner, a random set of rows and a few randomly selected variables are used to create a decision tree model. Final prediction may be a function of all the predictions made by the individual learners. In the event of a regression problem, the final prediction may be the mean for all predictions.

IV. RESULTS AND DISCUSSION

Predictive analytics is the use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.

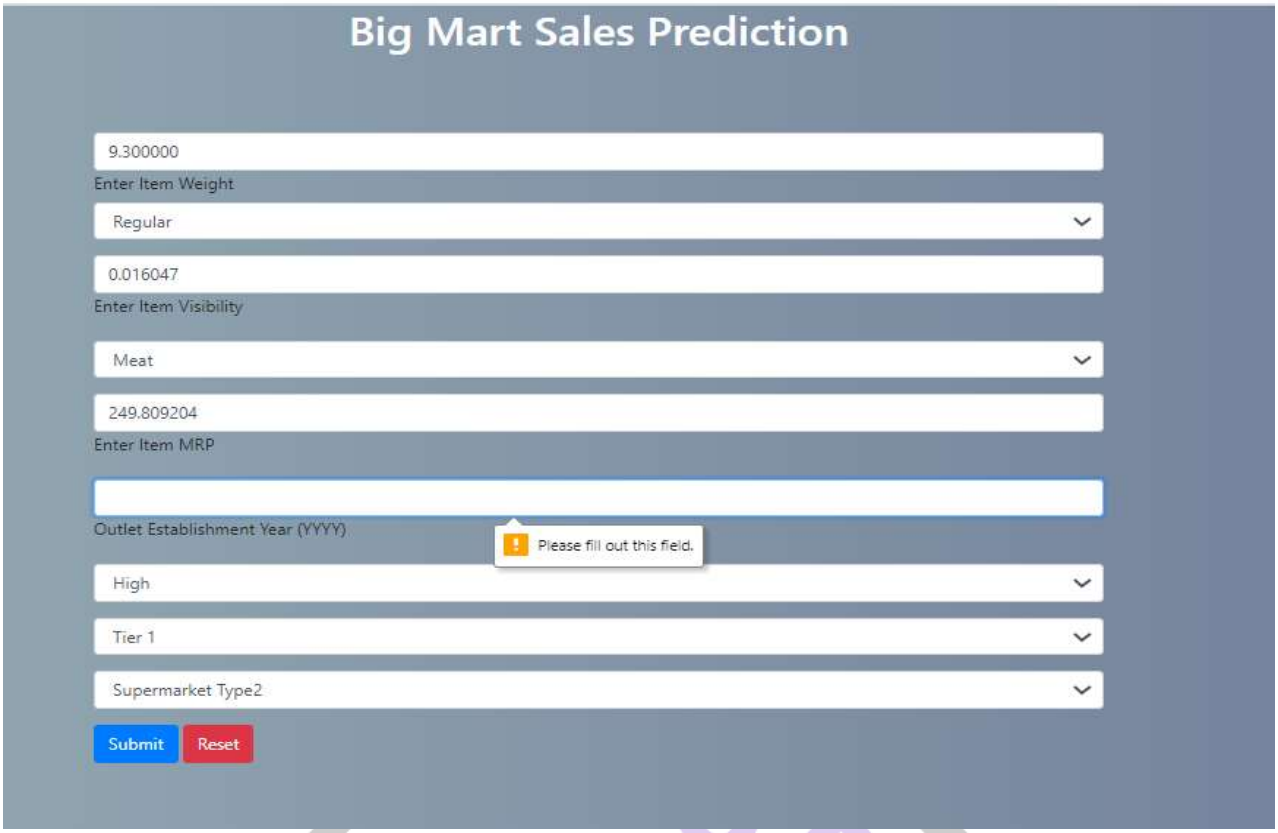
If classification is about separating data into classes, prediction is about fitting a shape that gets as close to the data as possible. In this we are predicting the value using the linear regression algorithm.

When we run the code we get the application page and we need enter the data from the dataset and submit then we get the prediction value. In another case when any field is missing then shows the error like fill out this.

Test Case: 1

{"Prediction":5640.871342769362}

Test Case: 2



V. CONCLUSION

Our predictions help big marts to refine their methodologies and strategies which in turn helps them to increase their profit. The results predicted will be very useful for the executives of the company to know about their sales and profits. This will also give them the idea for their new locations or Centre's of Big-mart. The Machine Learning Methods presented in this research paper should provide an effective method for data shaping and decision-making. New approaches that can better identify consumer needs and formulate marketing plans will be implemented.
The outcome of machine learning algorithms will help to select the most suitable demand prediction algorithm and with the aid of which Big Mart will prepare its marketing campaigns.

REFERENCES

[1] Chandel, A. Dubey, S. Dhawale and M. Ghuge, quot;Sales Prediction System using Machine Learningquot;, International Journal of Scientific Research and Engineering Development, vol. 2, no. 2, pp. 1-4, 2019.[Accessed 27 January 2020].

[2] Makridakis, S., Wheelwright, S.C., Hyndman, R.J.: Forecastingmethods and applications. John wiley & sons.

[3] C. M. Wu, P. Patil and S. Gunaseelan: Comparison of DifferentMachine Learning Algorithms for Multiple Regression on Black Friday Sales Data.

[4] S.Cheriyan, S.Ibrahim, S.Mohanan and S.Treesa, quot;Intelligent Sales Prediction Using Machine Learning Techniques,quot; International Conference on Computing.

[5] Kadam, H., Shevade, R., Ketkar, P. and Rajguru.: "A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression".

[6] C. M. Wu, P. Patil and S. Gunaseelan: Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data.

[7] Das, P., Chaudhury: Prediction of retail sales of footwear using feed forward and recurrent neural networks.

[8] Das, P., Chaudhury, S.: Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data.