BigMart Sales Prediction Using Machine Learning Techniques

Aaditi Narkhede Mitali Awari Suvarna Gawali Prof.AmrapalMhaisgawali
*(**Department of Computer Science and Technology, Usha Mittal Institute of Technology,***
***SNDT Women's University, Santacruz (W), Mumbai Maharashtra 400049,India**).*
(aaditirajeshn@gmail.com,awarimitali98@gmail.com,suvarna3396@gmail.com,amrapali10@gmail.com)

---------------------------------------***********************-------------------------------

Abstract:

The sales forecast is based on BigMart sales for various outlets to adjust the business model to expected outcomes.The resulting data can then be used to prediction potential sales volumes for retailers such as BigMart through various machine learning methods. The estimate of the system proposed shouldtake account of price tag, outlet and outlet location. A number of networks use the various machine- learning algorithms, such as linear regression and decision tree algorithms, and an XGBoost regressor, which offers an efficient prevision of BigMart sales based on gradient. At last, hyperparameter tuning is used to help you to choose relevant hyperparameters that make the algorithmshine and produce the highest accuracy.

Index Terms**—**Machine Learning Algorithms, Prediction, Reliability,Sales forecasting ,Prediction model, Regression.

---------------------------------------***********************-------------------------------

I. INTRODUCTION

Every item is tracked for its shopping centers and BigMarts in order to anticipate a future demand of the customer and also improve the management of its inventory. BigMart is an immense network of shops virtually all over the world. Trends in BigMart are very relevant and data scientists evaluate those trends per product and store in order to create potential centres.Using the machine to forecast the transactions of BigMart 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 moment 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 BigMarts 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 an significant part in a retail center. A stronger prediction is always helpful in developingand enhancing corporate market strategies, which also help to increase awareness of the market.

II. LITERATURE SURVEY

Sales forecasts provide insight into how a firm should manage its workforce, cash flow, and the means. This is an important precondition for the planning and decision-making of enterprises. It allows businesses to formulate their business plans effectively[1]. Learning algorithms used in classification and model categories such as linear Regression, Ridge Regression, Random Forest, Decision Tree, XGBoost these algorithms are suitable for sales forecast. The technique of regression is used to forecast, model the time series, and find the relationship of cause-effect between variables. A linear regression model assumes that inputs X1, ..., XP is linear with the regression function E(Y). Because the continuous variables are not normally distributed, the regression model is constructed with transformed variables. Plotting the residuals against the variables makes it clear. From the model description, only the variables Item MRP, Outlet Identifier, Outlet Establishment Year, Outlet Size, Outlet Location Type, and Outlet Type are relevant at a significance level of 5 percent[6]. Complex models like neural networks are overkill for simple problems like regression. And simpler models alongwith proper data cleaning perform well for the regression[2].Linear regression is a very famous method for prediction and analysis but one drawback is it gives less accuracy[5].Using the Random Forest, prediction of the sales is made easier and care is taken in fixing the optimum number of trees[6]. Random Forest is a tree-based algorithm wherein a certain number of decision trees are combined to make a powerful prediction model. It was found that the general linear model using the principal component analysis and the random forest techniques produce better results which are been decided by the RMSE values[6]. The Decision Tree technique comes under the paradigm of artificial intelligence that creates a tree with the most significant function and subsequent nodes in the root node in a tree with features of lesser ranking[2]. Internally, the XGBoost model implements the stepwise, ridge the regression that dynamically selects the features, and excludes the features multicollinearity. This implementation yielded the best data set outcomes[2].

III. 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 lowfat 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.

IV. DATA PREPROCESSING

The dataset used is BigMart 2013 sales result and there are total 12 attributes. Item Outlet Sales is the target variable and the other remaining attributes are independent variable.The pre-processing of data is a method for preparing and adapting raw data to a model of learning. This is the first and significant step to construct a machine learning model. Real-world data generally contain noise, missing values and may not be used in an unusable format especially for machine learning models.

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 pre-processed, 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 products for customers.

V. FEATURE ENGINEERING

Feature Engineering is a method to exploit domain data understanding to construct functions that work with machine learning algorithms. When feature engineering is done correctly,the predictive capability of machine learning algorithms is enhanced by building raw data features that help facilitate the machine learning process. Feature engineering also includes the correction of inappropriate values. In the device dataset, the visibility of the item had a minimum value of 0 which is not acceptable, because the item should be accessible to all. And so it was replaced by the mean of the column. As Outlet Years , a new column is created so we must consider how long the store runs instead of the year it was formed. Item Type is another column in the dataset that has 16 categories and is combined under the Food , Drink and Non-Consumable category. Column Item fat content had various representations, which were divided into low fat and regular categories. Outliers present in Item Outlet Sales are often excluded for better performance.

VI. EVALUATION METRICS

Evaluation of the model is the vital part of creating an efficient machine learning model. Therefore it is important to create a model and get suggestions from it in terms of metrics. It will take and continue until we achieve good accuracy according to the value obtained from metric improvements. Evaluation metrics describe one model's results [3]. The ability to distinguish between model outcomes is an important feature of the evaluation metrics. Here, we used Root Mean Squared Error(RMSE)metric for evaluation process. RMSE is given by following formula-

RMSE=qPN

i=1(Predictedi□Actuali)2N.

Where, N is the Complete Number of Observations. RMSE is the most commonly used evaluation method for regression problems. The power of 'square root' causes this metric to display significant variation in percentages. The 'squared' aspect of this metric tends to deliver more stable outcomes that avoids the cancelation of positive or negative error values.

VII. 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[3]. Testing data is given as input after Model Building a target variable to predict. ***The models are build using:***

_ Linear Regression
_ Ridge Regression
_ Decision Tree
_ RandomForest
_ XGBoost

For all models based on the above algorithms, 20 fold cross validation is used. Essentially cross validation provides an indication of how well a model is generalizing to the unseen results.Description of different algorithms used as follows:-

A. 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)[9]. Some requirements for a successful linear regression model must be fulfilled by data. Some of those is

the lack of multicollinearity, i.e. the independent variables should correlate with each other.
RMSE: 1127 is accomplished by this algorithm.

B. Ridge Regression

Ridge Regression is a method used where multicollinearity (independent variables are highly correlated) affects outcomes. While the least square estimates (OLS) are objective in multicollinearity, their variances are broad and deviate from the true value. By applying a degree of bias to regression calculations, ridge regression eliminates standard errors[2]. The Linear Regression Loss function is increased in Ridge Regression so as not only to minimize the number of square residuals but also to penalize the estimates of the parameters.
This algorithm is achieves ***RMSE***:1129.

C. Decision Tree

Decision tree is a classifier referred to as a tuple recursive in instant-space. It is a powerful way of multi-variable analysis and is a powerful technique for data mining. Applications can be used in various fields, and this approach represents the variables involved in achieving a given purpose and the motives for achieving the target and the methods of execution. Let the objective be denoted as (O) and (Ci) the means of action to be followed and let (M ij) the means of action corresponding to those means, which can be indicated by qi, (i=P1 ... Pn), which corresponds to the relationship.[1] n i=1 qi = 1; cuqi > 0 With this algorithm, RMSE:1058 is achieved.

D. RandomForest

RandomForest 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. With this algorithm RMSE:1069 is reached.

E. XGBoost

XGBoost stands for eXtreme Gradient Boosting. The implementationof the algorithm was engineered for the efficiency of computing time and memory resources [9]. Boosting is a sequential process based on the principle of the ensemble.This ncorporates a collection of low learners and improves the accuracy of predictions.Model values are weighted at any
moment t, based on the effects of the preceding instant t-1. The correctly calculated results are given a lower weight, and the wrong ones are weighted higher. With this algorithm, the XGBoost model implements the stepwise, ridge regression internally, which automatically chooses the features and removes
the multi-colinearity.RMSE:1052 is achieved with this algorithm.

VIII. HYPERPARAMETER TUNING

Hyperparameter tuning selects an optimal range of hyperparameters for algorithm learning. A hyperparameter for this is a parameter the value of which is set before learning starts. Hyperparameters are not model parameters, and can not be directly derived from results. By planning,System parameters shall be equipped when using gradient descent minimize the function to loss. Whilst the model parameters specify how input data can be translated to the desired output, the hyperparameters explain how the model is actually being structured. The best way to think of hyperparameters is like an algorithm 's settings which can be modified to maximize performance. Models can have multiple hyperparameters and can be treated as a test problem in order to find the right combination of parameters. ***While there are now many hyperparameter optimization / tuning algorithms, simple strategies:*** 1.Grid Search, and 2.Random search. However, computational methods for both grid search and random search tuning take a very long time, from an hour to a day. Because of its quickest calculation, thus, the Bayesian Optimization approach is used for hyperparameter tuning.

Bayesian Optimization:
Bayesian methods, in contrast to random or matrix search, maintain track of previous test outcomes that they use to construct a probabilistic model mapping of hyperparameters to the likelihood of an objective function score:P***(score jhyperparameters):***
The simple theory is to spend a little more time choosing the next hyperparameter and allow fewer calls to the objective function. The goal of Bayesian reasoning is to become ”less accurate” by constantly updating the surrogate probability model after-objective function evaluation with more data than these methods do. Bayesian model-based approaches can find better hyperparameters in less time, since they purpose for determining the right range of hyperparameters based on previous experiments.

IX. RESULTS

To forecast BigMart’s revenue, simple to advanced machine learning algorithms have been implemented, such as Linear Regression, Ridge Regression, Decision Tree, Random Forest, XGBoost. It has been observed that increased efficiency is observed with XGBoost algorithms with lower RMSE rating. As a result , additional Hyperparameter Tuning was conducted on XGBoost with Bayesian Optimization technique due to its quick and fairly simple computation, which culminated in the acquisition of the lowest RMSE value and making the model better matched to the underlying results. The submission file detailing Item Outlet Sales for Item based on the Model is resulted.

Fig. 1. RMSE Table

ALGORITHM	RMSE
XGBoost	1052
Decision Tree	1058
Random Forest	1069
Linear Regression	1127
Ridge Regression	1129

Fig. 2. Hyper-parameter tuning

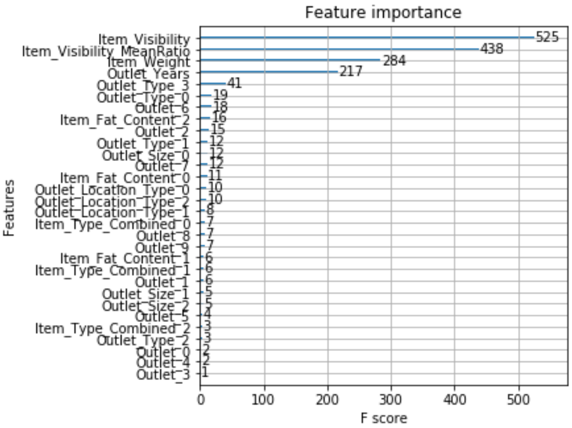
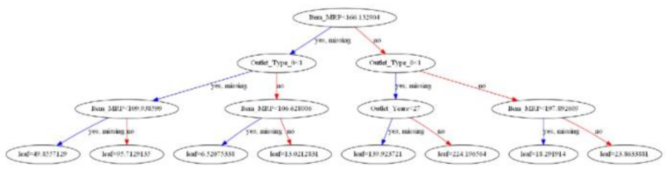


Fig. 3. XGBoost



X. CONCLUSION

Experts also shown that a smart sales forecasting program is required to manage vast volumes of data for business organizations. Business assessments are based on the speed and precision of the methods used to analyze the results. 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 BigMart will prepare its marketing campaigns.

REFERENCES

[1] S. Cheriyan, S. Ibrahim, S. Mohanan and S. Treesa, quot;Intelligent Sales Prediction Using Machine Learning Techniques,quot; 2018 International Conference on Computing, Electronics amp; Communications Engineering (iCCECE), Southend, United Kingdom, 2018, pp. 53-58.

[2] C. M. Wu, P. Patil and S. Gunaseelan, quot;Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data,quot; 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2018, pp. 16-20.

[3] A. Krishna, A. V, A. Aich and C. Hegde, quot;Salesforecasting of Retail Stores using Machine Learning Techniques,quot; 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 160-166.

[4] G. Nunnari and V. Nunnari, quot;Forecasting Monthly Sales Retail Time Series: A Case Study,quot; 2017 IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, 2017, pp. 1-6.

[5] Kadam, H., Shevade, R., Ketkar, P. and Rajguru, S. (2018). A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression. International Journal of Engineering Development and Research, 6(4), pp. 1-2.

[6] T. Alexander and D. Christopher, quot;An Ensemble Based Predictive Modeling in Forecasting Sales of Big Martquot;, International Journal of Scientific Research, vol. 5, no. 5, pp. 1-4, 2016. [Accessed 10 October 2019].

[7] G. Behera and N. Nain, quot;A Comparative Study of Big Mart Sales Predictionquot;, pp. 1-13, 2019. [Accessed 10 October 2019].

[8] S. Beheshti-Kashi, H. Karimi, K. Thoben and M. L¨utjen, quot;A survey on retail sales forecasting and prediction in fashion marketsquot;, Systems Science amp; Control Engineering, vol. 3, no. 1, pp. 154-161, 2014. Available: 10.1080/21642583.2014.999389 [Accessed 27 January 2020].

[9] A. 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].

[10] B. Pavlyshenko, quot;Machine-Learning Models for Sales Time Series Forecastingquot;, Data, vol. 4, no. 1, p. 15, 2019. Available: 10.3390/data4010015 [Accessed 27 January 2020].

[11] T. T. Joy, S. Rana, S. Gupta and S. Venkatesh, quot;Hyperparameter tuning for big data using Bayesian optimisation,quot; 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, 2016, pp. 2574-2579.

[12] M. Wistuba, N. Schilling and L. Schmidt-Thieme, quot;Hyperparameter Optimization Machines,quot; 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 41-50.

[13] M. Wistuba, N. Schilling and L. Schmidt-Thieme, quot;Learning hyperparameter optimization initializations, quot; 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Paris, 2015, pp. 1-10.

[14] K. Punam, R. Pamula and P. K. Jain, quot;A Two-Level Statistical Model for Big Mart Sales Prediction,quot; 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp. 617-620.

[15] S. Yadav and S. Shukla, quot;Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification,quot; 2016 IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, 2016, pp. 78-83.

[16] S. V. Patel and V. N. Jokhakar, quot;A random forest

based machine learning approach for mild steel defect diagnosis,quot; 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016, pp. 1-8.
[17] V. Shrivastava and P. Arya, quot;A study of various clustering algorithms on retail sales dataquot;, International Journal of Computing, Communications and Networking, vol. 1, no. 2, pp. 1-7, 2012. [Accessed 13 October 2019].