International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

SALES PREDICTION ANALYSIS

Naveen Kumar R[1], Jegan J[2], Yogesh V[3], Kavitha S[4]

[1-3]Student, Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India
[4]Assistant Professor II Department of Information Technology, Kumaraguru College of Technology, Coimbatore, India

---

Abstract - *The business organizations depends on knowledge base and sales trends prediction. For enhancing accuracy and efficiency, Data mining techniques proved effective in extraction of hidden knowledge. Here we are going to briefly analyse the concept of sales data and sales forecast. In this research we have done a detailed study and analysis of models to improve the future sales*

***Key Words*: Accuracy, Algorithms, Boosting techniques, Machine learning, Regression, Prediction Model, Random Forest.**

1. INTRODUCTION

There is the need for the industry to estimate the behaviours of the customer about the acceptance of the product in the market. This helps the industries to launch the new products that can be great success and help the companies to keep them away from loss making investments. Many methods are generally used to forecast the future trends in the market. Each of these approaches has certain set of requirements that needs to be addressed for the successful application of the approach. In general, the forecasting is relying on the previous data available with the institutes related to their fields.

Similarly, when the new product is introduced in the market there is no data about the product is going to perform in the market. As in the initial phase of development of the product as there is very less data about the product the model developed for forecasting might not work. Therefore, in order to get the data about feedback of the customers is taken in the form of the survey. These surveys help in moving the correct direction product generation. Hence, the forecasting is the needed to upgrade the products and even to tune the production of the products in the industry.

2. Literature Survey

PAPER-1:

The analysis of comprehensible predictive models to improve future sales are carried out in this research. In this paper, the concept of sales data and sales forecast were briefly analysed. The maximum accuracy is shown by Gradient Boosting algorithm so it is considered as the best fit model.

PAPER-2:

Machine learning and deep learning are the methodology used in this paper for analyzing the data and fetch useful information from the patterns and trends. Neural network is the best algorithm.

PAPER-3:

The main aim of this paper is to determine a dimension for the prediction of the Big Mart Company. It is carried out by using the sales data of previous years. Random forest, Xgboost, K-N and linear regression were used in this paper and among them Random Forest proves best.

PAPER-4:

Here the data such as sales, weather, stock value are considered. Gradient boost, Random and Extremely Randomised Trees are used over here. Among them Random Forest is considered to be effective after considering several parameters.

PAPER-5:

For a company like Big Mart, prediction acts as a major part in the sales of the products. Over here, the algorithm Xgboost is used and provided them very good result and are effective.

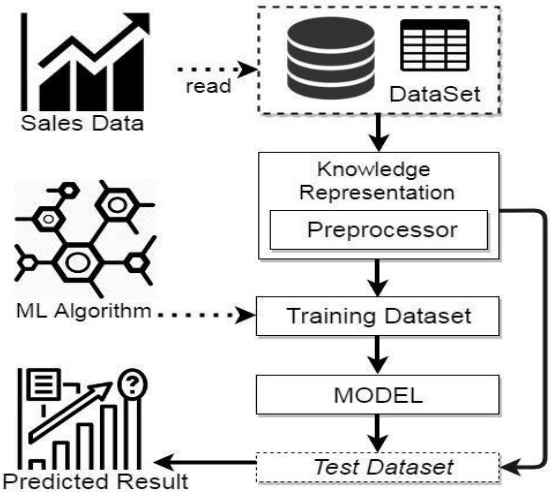
PAPER-6:

The major focus in this paper is to validate the average spending of the customer and carry out

the prediction on how the customer will spend in the future. The algorithm used over here are regression and Xgboost and Xgboost is the best fit for this scenario.

3. Proposed Approach:

The proposed work provides an opportunity for accuracy of findings by implementing data exploration and analysis algorithms. The algorithms ExtraTreeRegressor , Gradient Boosting, Random Forest , XgBoost and linear Regression are chosen as it provide better results.



RANDOM FOREST ALGORITHM

Random forest comprises of many decision trees. It is a supervised algorithm. This algorithm is used for regression to predict the future demand.

The Random forest algorithm, first processes the training set to develop a random forest model, and then produce K decision trees to set up a "forest" 101. Additionally, the algorithm categorizes the decision trees and generates a vote. The decision tree with maximum number of votes is the prediction.

Random forest algorithm is better than other algorithms in the following ways:

1. Random forest works great with multiple dimensional data.
2. Random forest can deal with the outliers.
3. This algorithm has a high leaming rate as compared to others.

XGBOOST ALGORITHM:

XGBoost is also referred as Extreme Gradient Boosting. It is widely used to obtain efficient model with high efficiency and accuracy. It provides a formula to avoid unnecessary errors of other decision tree models for optimization. This determines the value for maintaining the performance of the score and features.

LINEAR REGRESSION

Linear Regression is one of the common ML and data analysis technique. This algorithm is helpful for forecasting. The Linear regression technique is the type of regression, which combines the set of independent or dependent variable. It assigns a factor for each independent variable called coefficients represented by β.

GRADIENT BOOSTING

Gradient Boost is the one of the major boosting algorithm. Boosting is a ensemble technique in which the successive predictors learn from the mistakes of the previous or predecessor predictors. It is the method of improving the weak learners and create a combined prediction model. In this algorithm, decision trees are mainly used as base learners and trains the model in sequential manner. Week learners are recognized by the gradients in the loss function which is the training loss used to change the target variable for successive tree.

EXTRA TREE REGRESSOR

The Extra Tree Regressor creates a huge amount of decision tree from trained dataset. The average of the prediction is calculated of those decision trees on the basis of regression and maximum of the classification cases

4. DataSet

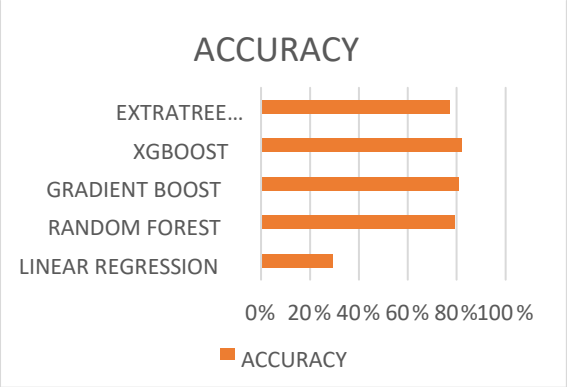
For the prediction of sales in the store we collected sales record for years. The dataset contains 12 variables which includes User ID, Gender, City Category, Product ID, Product Category2, ProductCategory3, Purchase amount etc.

5. Results

The performance of the machine learning algorithms are mainly focusing on accuracy. Companies uses the machine learning models with high accuracy for the practical business decisions. The other metrics used for these regression problems are RMSE where lesser the value is better.

Performance evaluation for different Machine Learning Techniques.

Algorithms	Accuracy	RMSE
Linear Regression	29%	4693
Random forest	79%	3051
XGBoost	82%	5023
Gradient Boost	81%	3004
ExtraTree Regression	77%	3137



Accuracy: In regression models, accuracy can be calculated by r2_score, which indicates the good fit of a set of predictions to the actual or real values.

Root Mean Squared Error (RMSE): It is one of the important metric for regression problems. It is defined by the square root of the average squared distance between the actual and the predicted score.

Based on the performance, we have concluded that the XGBoost showing 82% accuracy, RMSE value 5023 and therefore considered as the best fit comparing to other algorithms. This comparative evaluation will help the organizations to choose the better and efficient machine-learning model.

6. Conclusion

The performance is mainly focused on Accuracy. Machine learning techniques provides the effective mechanism in prediction and data mining as it overcome the problem with traditional techniques. These techniques enhances the data optimization along with improving the efficiency with better results and greater predictability. In our future work, we will use the other feature selection techniques and advanced ML algorithms to enhance the efficiency of the model with improved optimization.

REFERENCES

[1] Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan & Susan Treesa (2018) . Intelligent Sales Prediction Using Machine Learning Techniques.

[2] Xiangsheng Xie & Gang Hu (2008). Forecasting the Retail Sales of China's Catering Industry.

[3] Avinash kumar, Neha Gopal & Jatin Rajput(2020). An Intelligent Model For Predicting the Sales of a Product.

[4] Purvika Bajaj, Renesa Ray, Shivani Shedge & Shravani Vidhate(2020). SALES PREDICTION USING MACHINE LEARNING ALGORITHMS.

[5] Ching-Seh (Mike) Wu. Pratik Patil & Saravana Gunaseelan(2018). Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data.

[6] Nikhil Sunil Elias, 2Seema Singh(2019).FORECASTING of WALMART SALES using MACHINE LEARNING ALGORITHMS.

[7] Yuta Kaneko & Katsutoshi Yada(2016). A Deep Learning Approach for the Prediction of Retail Store Sales.

[8] Gopal Behera & Neeta Nain (2019). Sales Prediction for Big Mart.