Big Mart Sales Prediction using Machine Learning

Prof. Chandani Lachake[1], Manmohan Kumar[2], Nilesh Kumawat[3], Angatjot Singh[4], Rutik Nanaware[5]
Professor, Department of Computer Science Engineering[1]
B.E Students, Department of Computer Science Engineering[2,3,4,5]
SKN Sinhgad Institute of Technology & Science, Lonavala, Maharashtra, India

Abstract: *Nowadays shopping malls and Big Marts keep the track of their sales data for each and every individual item for predicting future demand of the customer and updating 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 stored in 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 paper, we propose a predictive model using XG boost Regressor technique for predicting the sales of a company like Big Mart and found that the model produces better performance as compared to existing models. A retail company wants a model that can predict accurate sales so that it can keep track of customers' future demand and update them in advance of the sale inventory. In this work, we propose a Grid Search Optimization (GSO) technique to optimize the parameters and select the best tuning hyper parameters, the further ensemble with Xgboost techniques for forecasting the future sales of a retail company such as Big Mart and we found our model produces the better result.*

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

I. INTRODUCTION

Big Mart is a big supermarket chain, with stores all around the country and its current board set out a challenge to all Data Scientists out there to help them create a model that can predict the sales, per product, for each store to give accurate results. Big Mart has collected sales data from the year 2013, for 1559 products across 10 stores in different cities.

With this information, the corporation hopes we can identify the products and stores which play a key role in their sales and use that information to take the correct measures to ensure the success of their business. 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 behaviors, 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.

Everyday competitiveness between various shopping centers as and as huge marts is becoming higher intense, violent just because of the quick development of global malls also online shopping. Each market seeks to offer personalized and limited-time deals to attract many clients relying on a period of time, so that each item's volume of sales may be estimated for the organization's stock control, transportation and logistical services.

II. RELATED WORK

(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. (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' behaviour. (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.

III. LITERATURE SURVEY

A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression (2018) Kadam, H., Shevade, R., Ketkar, P. and Rajguru. A Forecast for Big Mart Sales Based on Random Forests and Multiple Linear Regression used Random Forest and Linear Regression for prediction analysis which gives less accuracy. To overcome this, we can use XG boost Algorithm which will give more accuracy and will be more efficient.

Forecasting methods and applications (2008) Makridakis, S., Wheelwrigh. S. C., Hyndman. R.J. Forecasting methods and applications contain a Lack of Data and short life cycles. So some of the datalike historical data, and consumer-oriented markets face uncertain demands and can be predicted for accurate results.

Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data (2018) C. M. Wu, P. Patil, and S. Gunaseelan. Comparison of Different Machine Learning Algorithms for Multiple Regression on Black Friday Sales Data Used Neural Network for comparison of different algorithms. To overcome this Complex model like neural networks is used for comparison between different algorithms which is not efficient so we can use simpler algorithms for prediction. Prediction of retail sales of footwear using feed-forward and recurrent Neural Networks (2018)

by Das, P., Chaudhury. Prediction of retail sales of footwear using feed-forward and recurrent neural networks used neural networks for prediction of sales. Using a neural network for predicting weekly retail sales, is not efficient, So XG boost can work efficiently.

IV. PROPOSED SYSTEM

4.1 Data Processing and Methodology

   a. Data Collection: We have collected the data securely in accordance with an agreed methodology. The procedure for the collected data may differ from client to client and is dependent on the type, quantity, availability, and need of data.

   b. Data Cleaning and Preprocessing: The collected data is passed through a 'cleaning' process, so as to make sure that the data is segregated properly and identified gaps in the data are filled with the appropriate information, making data compatible and also fixing errors in storage systems which can cause data redundancy.

   c. Data Modeling: This is primarily a process in which the given dataset and the objects in it are analyzed to get a clear view of the requirements that may help us support our business model. Based on the analysis of patterns present in the data, models are then created on the established flow of the project. This flow offers better assistance in the utilization of the previously agreed upon the semi-formal model that showcases the features of the project. It also provides guidance to follow the relation between the data objects and other objects.

d. Data Prediction: Machine Learning prediction models are trained in this process and then later on evaluated using the data. This will then be applied to the preprocessed dataset. Some of the Models to be used for the prediction are:

e. Linear Regression

f. Random Forest

g. Decision tree

h. XG Boost Regressor

i. Data Visualization: Data Analyzed is then further picturized for customers and the admin to reach out conclusions and take effective decisions.
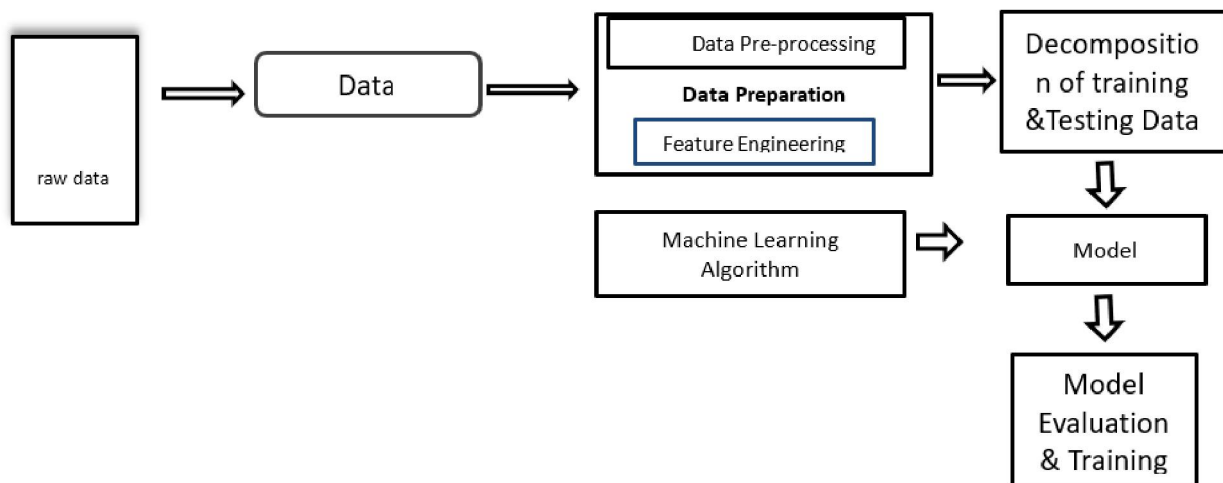

Fig:-Working procedure of proposed model

The figure represents the use case model of our system "Big Mart Sales Prediction Using Machine Learning"to find out the sales of each product at a particular store. Using this model, Big Mart will try to understand the properties of products and stores which play a key role in increasing sales.Huge shopping centres such as big malls and marts are recording data related to sales of items or products with their various dependent or independent factors as an important step to be helpful in prediction of future demands and inventory management

V. CONCLUSION

In this work, the effectiveness of various algorithms on the data on revenue and review of, best performance-algorithm, here propose a software to using regression approach for predicting the sales centred on sales data from the past the accuracy of linear regression prediction can be enhanced with this method, polynomial regression, Ridge regression, and Xgboost regression can be determined. We have designed a predictive model using ensemble techniques with this algorithm in the Big Mart dataset for forecasting future sales of a particular store or outlet of Big Mart. 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.

VI. ACKNOWLEDGMENT

REFERENCES

[1]. Rohit Sav, Pratiksha Shinde, Saurabh Gaikwad (2021, June). Big Mart Sales Prediction using Machine Learning.2021 International Journal of Research Thoughts (IJCRT).

[2]. Inedi. Theresa, Dr. Venkata Reddy Medikonda,K.V.Narasimha Reddy. (2020, March). Prediction of Big Mart Sales using Exploratory Machine Learning Techniques 020 International Journal of Advanced Science and Technology (IJAST).

[3]. https://www.javatpoint.com/machine-learning-random-forest-algorithm

[4]. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[5]. Bohdan M. Pavlyshenko (2018, August 25). Rainfall Predictive Approach for La Trinidad, Benguet using Machine Learning Classification. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP).

[6]. Pavan Chatradi, Meghana, Avinash Chakravarthy V, Sai Mythri Kalavala , Mrs.Neetha KS (2020), Improvizing Big Market Sales Prediction, Volume 12 Issue 4 (https://www.xajzkjdx.cn/gallery/423-april2020.pdf)

[7]. K. Punam, R. Pamula, and P. K. Jain, "A two-level statistical modelfor big mart sales prediction," in 2018 International Conference on Computing, Power and Communication Technologies (GUCON).

[8]. IEEE, 2018, pp. 617–620.

[9]. P. Das and S. Chaudhury, "Prediction of retail sales of footwear usingfeedforward and recurrent neural networks," Neural Computing and Applications, vol. 16, no. 4-5, pp. 491–502, 2007.

[10]. C.-W. Chu and G. P. Zhang, "A comparative study of linear andnonlinear models for aggregate retail sales forecasting," International Journal of production economics, vol. 86, no. 3, pp. 217–231, 2003.

[11]. S. Beheshti-Kashi, H. R. Karimi, K.-D. Thoben, M. Lutjen, and M. Teucke, "A survey on retail sales forecasting and prediction in fashion markets," Systems Science & Control Engineering, vol. 3,no. 1, pp. 154–161, 2015.

[12]. S. Asur and B. A. Huberman, "Predicting the future with socialmedia," in Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent TechnologyVolume 01. IEEE Computer Society, 2010, pp. 492–499.

[13]. V. Dhar and E. A. Chang, "Does chatter matter? the impact of usergenerated content on music sales," Journal of Interactive Marketing, vol. 23, no. 4, pp. 300–307, 2009.

[14]. J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stockmarket," Journal of computational science, vol. 2, no. 1, pp. 1–8,