

# D MART SALES PREDICTION

Sangeetha G<sup>1</sup>, Dr. Shilpa Abhang<sup>2</sup>

(18MCA16), Student, Dept. of MCA, Jyoti Nivas College Autonomous, Hosur Road, Bengaluru – 560095<sup>1</sup>

Asst Prof, Dept. of MCA, Jyoti Nivas College Autonomous, Hosur Road, Bengaluru – 560095<sup>2</sup>

**Abstract:** Nowadays shopping malls and D Marts keep the 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 D Mart. In this paper, we propose a predictive model using Xgboost technique for predicting the sales of a company like D Mart and found that the model produces better performance as compared to existing models. A comparative analysis of the model with others in terms performance metrics is also explained in details. The aim is to build a predictive model and find out the sales of each product at a particular store. Using this model, D Mart will try to understand the properties of products and stores which play a key role in increasing sales.

**Keywords:** D Mart Sales Prediction.

## I INTRODUCTION

Day by day competition among different shopping malls as well as D marts is getting more serious and aggressive only due to the rapid growth of the global malls and on-line shopping. Every mall or mart is trying to provide personalized and short-time offers for attracting more customers depending upon the day, such that the volume of sales for each item can be predicted for inventory management of the organization, logistics and transport service, etc. Present machine learning algorithm are very sophisticated and provide techniques to predict or forecast the future demand of sales for an organization, which also helps in overcoming the cheap availability of computing and storage systems. In this paper, we are addressing the problem of D mart sales prediction or forecasting of an item on customer's future demand in different D mart stores across various locations and products based on the previous record. Different machine learning algorithms like linear regression analysis, random forest, are used for prediction or forecasting of sales volume. As good sales are the life of every organization so the forecasting of sales plays an important role in any shopping complex. Always a better prediction is helpful, to develop as well as to enhance the strategies of business about the marketplace which is also helpful.

## II RELATED WORK

In paper [1], This paper took through the entire journey of solving a data science problem. started with making some hypothesis about the data without looking at it. Then moved on to data exploration where found out some nuances in the data which required remediation. Next, performed data cleaning and feature engineering, where imputed missing values and solved other irregularities, made new features and also made the data model-friendly by one-hot-coding. Finally made regression, decision tree and random forest model and got a glimpse of how to tune them for better results.

In paper [2], They have designed a predictive model by modifying Gradient boosting machines as Xgboost technique and ex-perimeter it on the 2013 Big Mart data set for predicting sales of the product from a particular outlet. Experiments support that our technique produce more accurate prediction compared to than other available techniques like decision trees, ridge regression etc. Finally a comparison of different models is summarized in Table 2. From Table 2 it is also concluded that our model with lowest MAE and RMSE performs better compared to existing.

In paper [3], They use random forest regression and XG-booster approach to predict sales where data mining techniques such as discovery, data transformation, feature development, model creation and testing are used. In this technique raw data collected by a big mart will be prepossessed for missing data, anomalies and outlier. An algorithm will then be trained to construct a model on that data. they use this model to forecast the end results. It is a system in which three functions are combined. It is used to extract and transform the data from one database into an appropriate format.



Generally, in preprocessing of data raw data is converted into useful form of data. Data preprocessing involves the following steps

- Data-cleaning.
- Data-transformation.
- Data-reduction

In this proposed system they have used Random Forest Algorithm to incorporate predictions from multiple decision trees into a single model.

In paper [4], Sales forecasting is very crucial for every company, especially big ones and this process is very complex because there are lots of factors that should be taken into consideration. In order to implement achievable goals and successfully implement them, supermarkets chains always want to forecast sales. In this study, they used three machine learning algorithms (K-Nearest Neighbor, Random Forest and Gradient Boosting) for sales forecasting, RF performed better, as it had a lower mean absolute error than the other two models. They also observed that getting more data would generally increase the predictive power of our models.

In paper [5], There are two noteworthy types of forecasting. Macro Forecasting and Micro Forecasting. Macro Forecasting is unstable with foreseeing commercial centers in entirety. This will be identified with close the present level of Market Plea and assessing the results of market request in the up and coming ages. Micro Forecasting is concerned with entire segment deals figures. This will be tied in with molding a thing's business sector parcel specifically industry and thinking about the outcomes of piece of the overall industry later on period. A gauge is an evaluation of an episode which will occur in future. The forecast esteem isn't a deterministic measure. Since it is just a gauge in light of the past information identified with a specific occasion, legitimate care must be given in evaluating it.

### III METHODOLOGY

The aim is to build a predictive model and find out the sales of each product at a particular store. Using this model, D Mart will try to understand the properties of products and stores which play a key role in increasing sales.

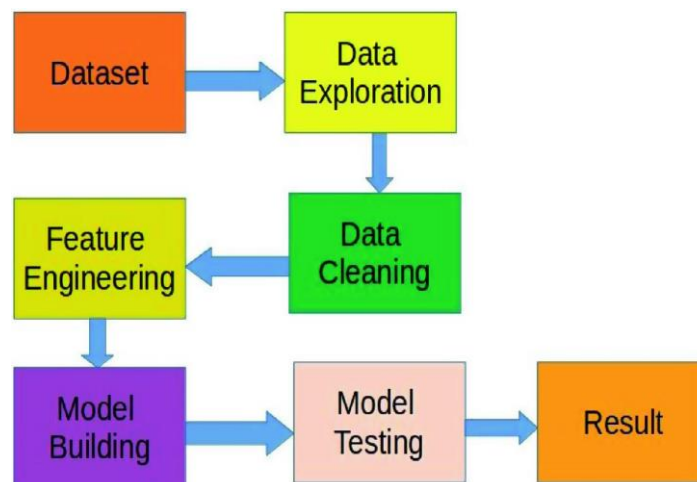


Figure 1

### IV IMPLEMENTATION

#### Experimental setup

Python SciPy environment installed ideally with Python 3.6.3, and other libraries like Scikit-learn, Pandas, NumPy, and Matplotlib is installed. Minimum of 4GB RAM is required.

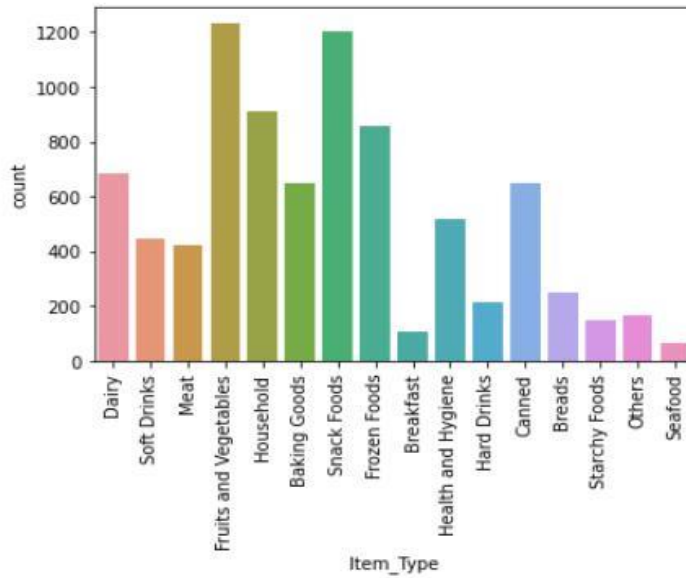
#### Method to Run this project

1. Install Python 3.6.3
2. First Train the model.



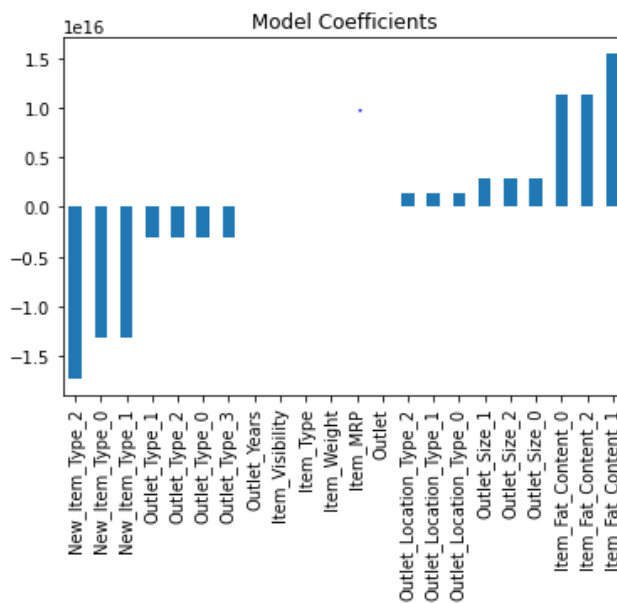
3. Second give the Data set.
4. imputing missing values in the data and checking for outliers
5. modifying existing variables and creating new ones for analysis
6. Do predictive models on the data.

**Experimental Results**



Model Report  
 MSE: 1270969.2690816731  
 CV Score: 1278631.429283423

Out[148]: <AxesSubplot:title={'center': 'Model Coefficients'}>



**V CONCLUSION**

In this paper, basics of machine learning and the associated data processing and modeling algorithms have been described, followed by their application for the task of sales prediction in D Mart shopping centers at different locations. On implementation, the prediction results show the correlation among different attributes considered and how a particular location of medium size recorded the highest sales, suggesting that other shopping locations should follow similar patterns for improved sales. Multiple instances parameters and various factors can be used to make this sales prediction more innovative and successful. Accuracy, which plays a key role in prediction-based systems, can be significantly increased as the number of parameters used are increased. Also, a look into how the sub-models work can lead to increase in productivity of system. The project can be further collaborated in a web-based application or in any device supported with an in-built intelligence by virtue of Internet of Things (IoT), to be more feasible for use. Various stakeholders concerned with sales information can also provide more inputs to help in hypothesis generation and more instances can be taken into consideration such that more precise results that are closer to real world situations are generated. When combined with effective data mining methods and properties, the traditional means could be seen to make a higher and positive effect on the overall development of corporation's tasks on the whole. One of the main highlights is more expressive regression outputs, which are more understandable bounded with some of accuracy. Moreover, the flexibility of the proposed approach can be increased with variants at a very appropriate stage of regression.

**REFERENCES**

1. Nikita Malik and Karan Singh, New Delhi, Sales Prediction Model for Big Mart, Issue January-June 2020.
2. Magana N, GITAM School of Technology, Bengaluru campus, Karnataka, India, Improvizing Big Market Sales Prediction, Issue 2020
3. Gopal Behera and Neeta Nain, MNIT Jaipur, A Comparative Study of Big Mart Sales Prediction, Issue sep-2019.
4. Charles Westby, Big Mart Sales Predictions-Issue-2017
5. Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge, Sales Prediction System using Machine Learning Mumba, issue-mar-2019