iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50059

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Forecasting Future Sales of Bigmarts

P Srinivasa Rao[1], N Sanjay[2], K Anusha Reddy[3]
[1]*Associate Professor, Dept of IT, Malla Reddy College of Engineering & Technology, Hyderabad, Telangana, India*
[2, 3]*UG Student, Dept of IT, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India*

Abstract: ***In this contemporary era, supermarket and general stores have been scrutinizing the sales record for knowing the demands of customers and to find the straggles in general trend. So, as the data is available a predictive model has been built using various algorithms such as Polynomial regression, XGBoost, Linear regression, and Ridge regression techniques for forecasting the sales of a business in advance. The prediction is based on sales of supermarket for various outlets to calibrate the business model to expected outcomes. With the results and analysis provided by the model retailers can know the sales volume in advance.***

Keywords: Machine Learning Algorithms, Prediction, Hyperparameter Tuning, Sales Forecasting.

I. INTRODUCTION

In this world of entrepreneur, a new rise of war between leading firms has taken of in order to have a monopoly in customer acquisition. Previously big firms and producers use to follow various techniques like marketing, discounting which lead to extensive co-set cutting or compromise of product quality this would neither benefit customers nor retailers. In this new era has the market size increased, world wide access of goods due to globalization and accessibility of internet has given rise to a new gateway where new techniques, technologies, software's, algorithms, third party services have made these work simple but effective. So as a part of this new revolution we propose new method for forecasting future sales based on the data available at outlets which is customer and product specific. This helps in better management of inventory(stocks).

This helps firms to act or produce or manage according to the prediction. This results in more profits, less wastage (avoiding expiry), better cost cutting, helps in research and development of product based on customers behavior. So basically, we have implements some of the new ML algorithms in this paper to avoid the problems in the existing system. We have removed the null values from the data sets and data duplication has also been resolved. By resolving these problems, the results achieved are very precise with negligible error. In this project study we have used some of the machine learning algorithms like Linear regression, Random Forest and XGBoost techniques and the accuracy are compared individually. After observing the RMSE values of all the algorithms, the accuracy of the Random Forest is comparatively high, so we have used the Random Forest algorithm using pickle. dump method from machine learning package to forecast the future sales.

II. METHODOLOGY

To train the model and to deploy the trained model, there are basically six Modules present:

A. Data Gathering

The data gathering is generally done manually by collecting entire data. But as entire data is available online in data resources on the web, we can directly get the data by downloading and use it for training the model. Considering the problem, the data required for sales forecasting includes item identifier, visibility, Outlet Identifier, Item outlet sales, etc. Such required data should be stored or collected for analysis.

B. Data Preparation

Generally, the data collected will never be in correct format we will encounter a lot of inconsistencies like duplicate values, missing values, redundant variables, etc. Overcoming such inconsistencies is crucial because they might lead to wrong predictions. With this stage, we scan the data set for any inconsistencies and fix them.

C. Exploratory Data Analysis

Knowing the hidden patterns is one of the crucial steps which is done at this stage. Data Exploration involves understanding the hidden patterns and trends in the data. This stage helps in getting all the serviceable insights are extracted and correlations among the variables are understood.

©IJRASET: All Rights are Reserved

D. Model Building

After Data Exploration all the patterns and insights are derived which are helpful to build the Machine Learning Model. Model building begins by categorizing the data set into testing data, and training data. For building and analyzing the training data is used. Considering the suitable algorithm is dependent on the data set, the level of complexity and the type of problem to be solved.

E. Model Evaluation And Optimization

After using the training data set for model building, testing should be done. This stage involves in testing the model with the help of testing data which is categorized earlier. Testing data is helpful in checking the efficiency of the model. Accuracy is calculated, and if any further improvements are required, they can be implemented in this stage.

F. Predictions

After model evaluation and optimization, model is used to make the predictions. In this stage the predictions are made by programming hence this step is also considered as the programming phase.

G. Architecture

The following technical architecture represents the entire top-down work flow of model building including selection.

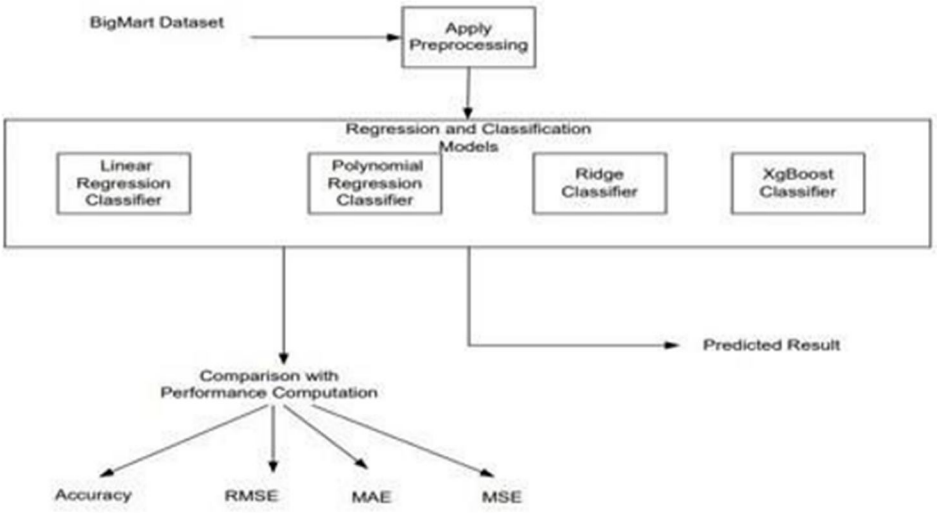


Figure 1: Representing Technical Architecture and Execution Flow

III. MODELING AND ANALYSIS

Entire data is categorized into training and testing data. We have trained our machine learning model with the available training data. Whenever new input data is given to the ML algorithm, it makes a prediction based on the model that have been trained. The predictive accuracy of machine learning algorithms is enhanced by building given data features that help facilitate the machine learning process, when feature engineering is done correctly. Optimal range of hyperparameters for algorithm learning is selected by hyperparameter tuning.

There can be multiple hyperparameters for the model to get the correct combination of parameters in the training data. Following are the strategies that are used for hyperparameter optimization.

1) Grid Search

2) Random search

Despite the advantages, for grid search and random search hyperparameter tuning take more time. Hence the Bayesian Optimization approach is opted for hyperparameter tuning as the calculations are done quickly calculation. Multiple algorithms are used and accuracy is considered as the key criteria for prediction.

The location with large area sales is not the highest. OUT027 location produced the highest sales , which was in turn a Supermarket Type3, having its size recorded as medium in our dataset. It can be said that this outlet's performance was much better than any other outlet location with any size provided in the considered dataset.
The median of the target variable Item Outlet Sales was calculated to be 3364.95 for OUT027 location. The location with second highest median.

ALGORITHM	RMSE
Linear Regression	1085.7
Random Forest	1029.2
XGBoost	1191.8

Figure 2: Final RMSE Values For Implemented ML Algorithms

IV. RESULTS AND DISCUSSION

When compared with the other algorithms Random Forest algorithm got the lowest Root Mean Square Error. Hence, additional Hyperparameter Tuning was conducted on Random Forest with Bayesian Optimization technique due to efficient and fast computation, and lowest RMSE value and making the model best with accurate results. Basing on the data it can be concluded that to increase the sales as many possible locations should be shifted to Supermarket Type3.

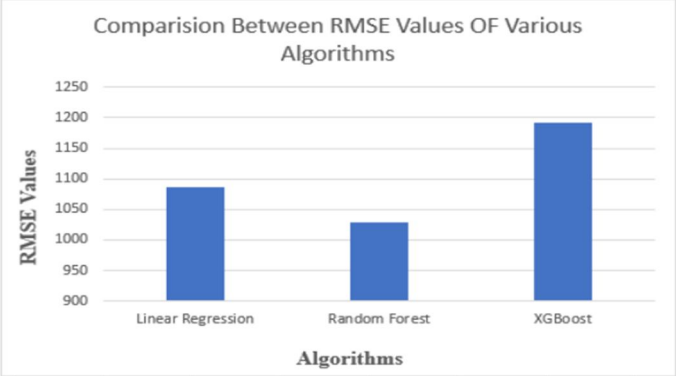


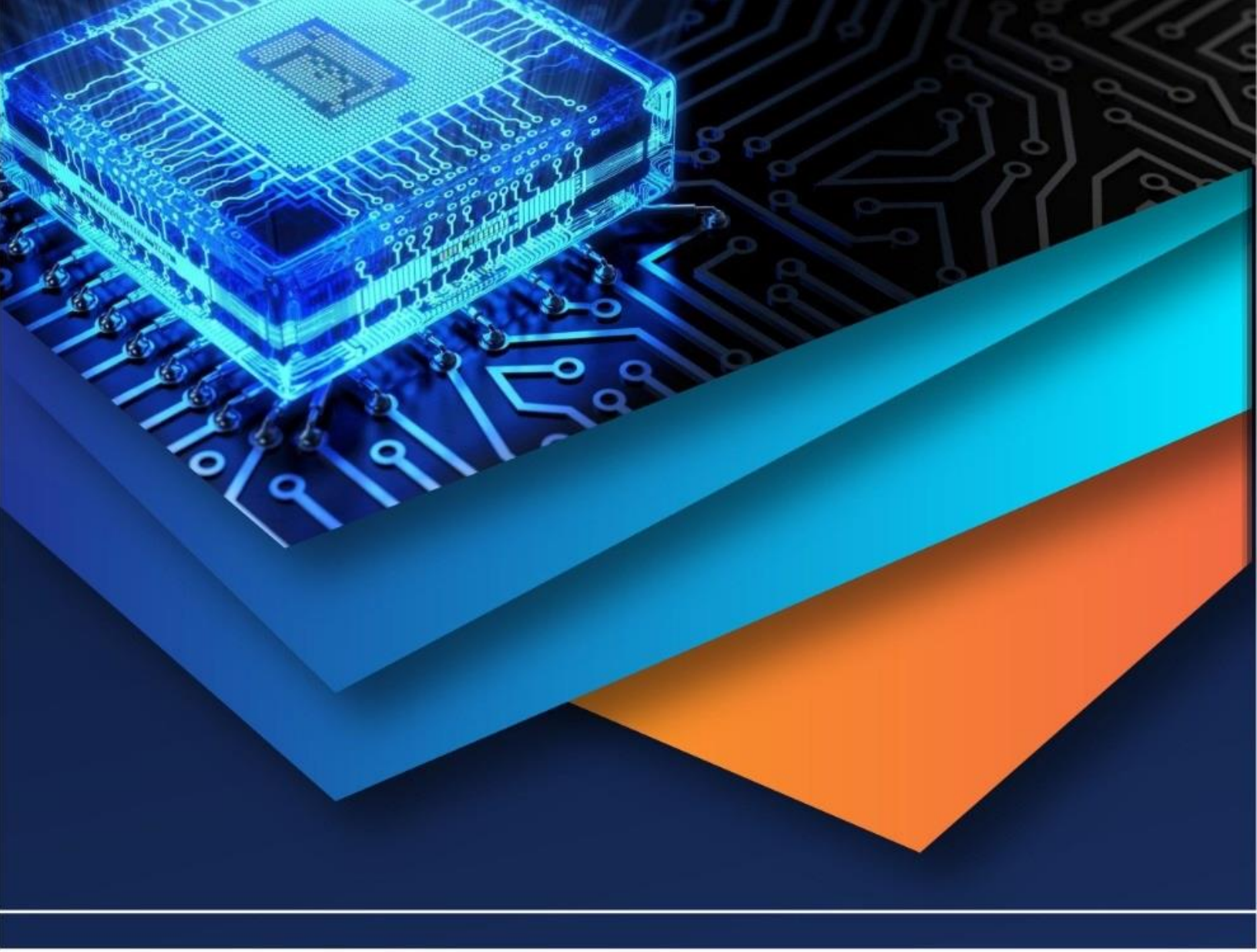
Figure 3: Comparison Graph

V. CONCLUSION

Now a days because of the demand and competitiveness among various super markets the store owners are not maintaining the stock correctly, Hence basing on the accuracies of mentioned algorithms on the data that is maintained by stores, we propose software using the regression approach for predicting the sales. And the sales data is tested on multiple algorithms and enhanced. The methodologies of Machine Learning represented in this research paper will provide an efficient approach to acquire proper insights from the data and deciding what actions to be performed to overcome the difficulties. RMSE and other error calculations results the best machine learning algorithms which help to consider the most suitable sales prediction algorithm.

REFERENCES

[1] Chu, C. W., Zhang, G. P. (2013). A comparative study of linear and nonlinear models for aggregate retail sales forecasting. International Journal of production economics, 86(3), 217-231.
[2] Hadavandi, E., Shavandi, H., Ghanbari, A. (2011). An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: Case study of printed circuit board. Expert Systems with Applications, 38(8), 9392-9399. Suma, V., Hills, S. M. (2020). Data mining based prediction of demand in Indian market for refurbished electronics. Journal of Soft Computing Paradigm (JSCP), 2(02), 101-110.
[3] Anggraeni, W., Vinarti, R. A., Kurniawati, Y. D. (2015). Performance comparisons between arima and arimax method in moslem kids clothes demand forecasting: Case study. Procedia Computer Science, 72, 630-637.
[4] Ranjitha, P., Spandana, M. (2021, May). Predictive Analysis for Big Mart Sales Using Machine Learning Algorithms. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1416-1421). IEEE



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)