

# MARKET RESEARCH: TAXONOMY AND PULLING OUT OF SALES PREDICTION USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

<sup>1</sup>G.Ramya, <sup>2</sup>K. Masoom Basha, <sup>3</sup>Dr.C.Srinivasa Kumar

<sup>1</sup>Assistant Professor, Department of CSE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana

<sup>2</sup>Assistant Professor, Department of CSE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana

<sup>3</sup>Professor & Dean, Department of CSE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Telangana

## ABSTRACT

Sales forecasting is the process of predicting future sales. It is the vital part of the financial planning of the business. Most of the companies heavily depend on the future prediction of the sales. Accurate sales forecasting empower the organizations to make informed business decisions and it will help to predict the short-term and long-term performances. A precise forecasting can avoid overestimating or underestimating of the future sales, which may leads to great loss to companies. The past and current sales statistics is used to estimate the future performance. But it is difficult to deal with accuracy of sales forecasting by traditional forecasting. For this purpose, various machine learning techniqueshave been discovered. In this work, we have taken Black Friday dataset and madea detailed analysis over the dataset. Here, we have implemented the different machine learning techniques with different metrics. By analysing the performance, we have trying to suggest the suitable predictive algorithm to our problem statement.

**Keywords:** Artificial Intelligence, Marketing Research, Sales Prediction Machine Learning Algorithms.

# **INTRODUCTION**

UGC CARE Group-1,



Sales play a key role in the business. At the company level, sales forecasting is the major part of the business plan and significant inputs for decision-making activities. It is essential for organizations to produce the required quantity at the specified time. For that, sales forecasting will gives the idea about how an organization should manage its budgeting, workforce and resources. This forecasting helps the business management to determine how much products should be manufacture, how much revenue can be expected and what could be the requirement of employees, investment and equipment. By analyzing the future trends and needs, Sales forecasting helps to improve the business growth. The traditional forecasting systems have some drawbacks related to accuracy of the forecasting and handling enormous amount of data. To overcome this problem, Machine-Learning (ML) techniques have been discovered. These techniques helps to analyses the bigdata and plays a important role in sales forecasting. Here we have used supervised machine learning techniques for the sales forecasting. Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. Sales forecasting is the process of estimating future sales. Accurate sales forecasts enable companies to make informed business decisions and predict short-term and long-term performance. Companies can base their forecasts on past sales data, industry- wide comparisons, and economic trends. Sales forecasts help sales teams achieve their goals by identifying early warning signals in their sales pipeline and course correct before it's too late. The goal is to improve the accuracy from the existing project. So that the sales and profit could be increased for the companies. Choosing an efficient algorithm from comparing different algorithms to improve the prediction further more.

### LITERATURE REVIEW

The detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting. The models implemented for prediction are Random Forest, Gradient Boosting and Extremely Randomized Trees (Extra Trees) Classifiers. The forecast of China's catering retail sales was studied in this paper. The seasonal impact was considered in the forecasting. The retail sales were predicted using the seasonal auto-regressive integrated moving



average (ARIMA) model. SVM method is obviously superior to the seasonal ARIMA method regardless of the long-term forecasting or the short- term forecasting. The approach shown in this paper is a systematic, accurate and precise model building to be used in computing and predicting current scenario and future projection of a product in market respectively. The aim of this paper is to propose a dimension for predicting the future sales of Big Mart Companies keeping in view the sales of previous years. A comprehensive study of sales prediction is done using Machine Learning models. Linear Regression, K-Neighbours Regressor, XGBoost, Regressor and Random Forest Regressor. This study focuses on the field of prediction models to develop an accurate and efficient algorithm to analyze the customer spending in the past and output the future spending of the customers with same features. The ability to predict data accurately is extremely valuable in a vast array of domains such as stocks, sales, weather or even sports. Presented here is the study and implementation of several ensemble classification algorithms employed on sales data, consisting of weekly retail sales numbers from different departments in Walmart retail outlets all over the United States of America. The models implemented for prediction are Random Forest, Gradient Boosting and Extremely Randomized Trees (Extra Trees) Classifiers. A retailer company wants a model that can predict accurate sales so that it can keep track of customers future demand and update in advance the sale inventory. In this work, we propose a technique to optimize the parameters and select the best tuning hyper parameters, further ensemble with Xgboost techniques for forecasting the future sales of a retailer company such as Big Mart and we found our model produces the better result.

### **METHODOLOGY**

The dataset has been collected from https://www.kaggle.com/ The training dataset contains 12 columns and 550069 rows. The Test dataset contains contains 12 columns and 233600. The dataset contains 12 variables which includes User ID, Gender, City Category, Product ID, Total count of years stayed in current city, Age, Occupation, Marital status, Product Category1, Product Category2, Product Category3 and Purchase amount This step is an important step in data mining process. Because it improves the quality of the experimental raw data.

i) Removal of Null values:

In this step, the null values in the fields Product Category2 and Product Category3 are filled with



the mean value of the feature.

ii) Converting Categorical values into numerical: Machine learning deal with numerical values easily because of the machine readable form. Therefore, the categorical values like Product ID, Gender, Age and City Category are converted to numerical values.

Step1: Based on its datatype, we have selected the categorical values.

Step2: By using python, we have converting the categorical values into numerical values.

iii) Separate the target variable: Here, we have to separate the target feature in which we are going to predict. In this case, purchase is the target variable.

Step1: The target lable purchase is assigned to the variable 'y'.

Step2: The preprocessed data except the target lable purchase is assigned to the variable 'X'.

iv) Standardize the features:

Here, we have to standardize the features because it arranges the data in a standard normal distribution. The standardization of the data is made only for training data most of the time because any kind of transformation of the features only be fitted on the training data.

Step1: Only trained data was taken.

Step2: By using the Standard Scaler API, we have standardize the features.

### **Linear Regression:**

Linear Regression is one of the common ML and data analysis technique. This algorithm is helpful for forecasting based on linear regression equation. The Linear regression technique is the type of regression, which combines the set of independent features(x) to predict the output value(y) or dependent variable. The linear equation assigns a factor to each independent variable called coefficients represented by  $\beta$ . XGBoost also known as Extreme Gradient Boosting has been used in order to get an efficient model with high computational speed and efficacy. The formula makes predictions using the ensemble method that models the anticipated errors of some decision trees to optimize last predictions. Production of this model also reports the value of each feature's effects in determining the last building performance score prediction. 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. Random forest is referred as a supervised machine learning ensemble method, which uses the multiple decision trees. It involves the technique called Bootstrap aggregation also known as bagging which aims to reduce the complexity of the models that overfit the training data . In this algorithm, rather than depending on individual decision tree it will combines the multiple decision trees to find the final outcome.

# **RESULTS AND DISCUSSION:**

The evaluation of the machine learning algorithms is an essential part of any prediction model building. For that, we should carefully choose the evaluation metrics . These metrics are used to measure or judge the quality of the model. 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 Based on the performance, we have concluded that the XGBoost and Gradient Boost algorithm 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.

ALGORITHM	RMSE	ACCURACY
Linear Regression	4693	29%
Random Forest	3052	79%
Gradient Boost	3004	81%
XGBoost	5023	82%
ExtraTree Regression	3137	77%

Figure1: Evaluating the Accuracy



Industrial Engineering Journal

ISSN: 0970-2555





In [74]:
barWidth = 0.25
fig = nlt.subnlots(figsize =(12. 8))

Figure2: Accuracy for different Machine Learning Techniques



Figure3: Accuracy Comparison for different Machine Learning Techniques



#### CONCLUSION

Sales forecasting is mainly required for the organizations for business decisions. Accurate forecasting will help the companies to enhance the market growth. Machine learning techniques provide the effective mechanism in prediction and data mining as it overcome the problem with traditional techniques. These techniques enhance the data optimization along with improving the efficiency with better results and greater predictability. After predicting the purchase amount, the companies can apply some marketing strategies for certain sections of customers so that the profit could be enhanced. In our future work, we will use the other feature selection techniques and advanced deep learning architecture algorithms to enhance the efficiency of the model with improved optimization. The goal is to improve the accuracy from the existing project. So that the sales and profit could be increased for the companies. Choosing an efficient algorithm from comparing different algorithms to improve the prediction further more.

#### REFERENCES

[1] Q. Huang and F. Zhou, "Research on retailer data clustering algorithm based on spark", AIP Conference Proceedings, vol. 1820, no. 1, pp. 080022, 2017, March..

[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 ModelFor 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, Seema Singh(2019).FORECASTING of WALMARTSALES 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.
- [9] Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan & Susan Treesa (2018) Intelligent Sales Prediction Using Machine Learning Techniques.