JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DIWALI SALES PREDICTION USING MACHINE LEARNING

[1]Mrs. Swapna G, [2]Adarsh K, [3]Aniketh H, [4]Latha V, [5]M Sreelakshmi

[1]Assistant Professor, [2,3,4,5] Undergraduates
[1]Department of Computer Science and Engineering
[1]K S Institute of Technology, Bangalore, India

Abstract: Understanding the purchase behavior of various customers (dependent variable) against different products using their demographic information (IS features where most of the features are self – explanatory). This dataset consists of null values, redundant and unstructured data. Machine learning is the most common application in the domain retail industry. This concept helps to develop a predictor that has a distinct commercial value to the shop owners as it will help with their inventory management, financial planning, advertising and marketing. This entire process of developing a model includes preprocessing, modeling, training testing and evaluating. Hence, frameworks will be developed to automate few of this process and its complexity will be reduced. The algorithm that we will be using is Random forest Regression approach which is the one of the most important Supervised Machine Learning algorithm.

Index Terms **- Sales Prediction, Data Analysis, Random Forest Regressor, Testing and Training**

I. INTRODUCTION

"Diwali sales" is the biggest sales which usually occurs during Diwali festival which falls in the month of November. In the retailing sector, the amount of sales determines whether the company turns a profit or loses money. Accurate sales forecasting allows for impactful industry management. To increase sales, an estimation model is created that glides on the sort of product that sells the most. The behavior of a purchaser should be studied in order to forecast how much money he or she will spend on a given day. We will anticipate the purchases of a "retail shop" on "Diwali sales" in this paper. In order to predict the sales of various products based on their predictor factors, we must first analyze and organize the correlation between different variables. So that a prototype can perform classification and reliably estimate sales.

Two goals are addressed in this paper.

1.Evaluating all consumer data and determining correlations between
Independent factors and the target output.

2. Evaluating and mentoring to predict expected revenue

II. LITERATURE SURVEY

[1] Sumit Kalra, Boominathan, Perumal, Samar Yadav and Swathi Jamjala Narayanan proposed a Workflow on Analyzing and Predicting the purchases done on the day of Black Friday.The primary objective is to analyze the data and estimate what consumers will buy based on different productids.The Xgboost, Tfid transformer, Model1+ Model 2, and Extra tree regressor are the four algorithms used. The four algorithms mentioned above significantly outperforms in terms of precision and RMSE accuracy.

[2] Amruta Aher, Dr. K. Rajeswari and Prof. Sushma Vispute proposed a Workflow on Data Analysis and Price Prediction of Black Friday Sales using Machine Learning Techniques. The major goal wasto predict the salesthat occurred on Black Friday. Ridge regression, Lasso regression, Decision Tree Regressor, and Random Forest Regressor are the algorithms used for prognostication. The 5-fold cross-validation is utilized to validate the algorithm. The Mean Squared Error (MSE) is the performance review metric used. With an MSE score of 3062.719, the Random Forest Regressor surpasses the other algorithms.

[3] Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan and Susan proposed a Trees a Intelligent Sales Prediction Using Machine Learning Techniques.The major goal was to examine and predict decisions. Generalized Linear Model,Decision Tree, and Gradient Boost Tree are the algorithms used for estimation.The results indicate that the Gradient Boost algorithm has the highest reliability in forecasting and future revenues predictions, with an accuracy rate of around 98 percent.

[4] Ramachandra H V, Balaraju G, Rajashekar A and Harish Patil proposed Workflow on Machine Learning Application for Black Friday Sales Prediction Framework. The main goal is to analyze all purchaser data and find relationships between independent factors and the target value in ability to forecast expected revenue. Linear regression, Random Forest, Decision Tree, and Xgboost regression, as well as Random Forest Regression, are the algorithms used to make predictions .Root Mean Squared Error was predicted using five algorithms (RMSE). With an average accuracy of 83.6 percent and a minimum RMSE, the Random Forest regressor performed well.

[5] B.Sri Sai Ramya and K. Vedavathi proposed An Advanced Sales Forecasting Using Machine Learning Algorithm. The main goal is to forecast and analyze sales, as well as to predict potential sales. The prediction was made using the Extreme Gradient Boosting algorithm. The Extreme Gradient Boosting algorithm which gave better efficiency to manipulate the sales analysis. It predicts the sales occurred in big mart using machine learning concepts. The Extreme Gradient Boosting algorithm was used for prediction. The Extreme Boosting algorithm gave good accuracy. Linear regression, Support Vector regression, Random Forest, and Decision Tree are the estimation algorithms used. Random Forest achieved a positive accuracy score with a low error rate.

[6] Karandeep Singh, Booma P M and Umapathy Eaganathan proposed on E-Commerce System for Sales Prediction Using Machine Learning. The main aim to predict and analysis the sales occurred in ecommerce. Random Forest and Gradient Boostingare the algorithms used for prediction, and it was discovered that Gradient Boosting has higher train and test accuracy than Random Forest.

[7] Purvika Bajaj, Renesa Ray, Shivani Shedge, Shravani Vidhate and Prof. Dr. Nikhilkumar Shardoor proposed on Sales Prediction using Machine Learning Algorithms. The main aim is to predict the products sold. Linear regression, K-Neighbours regression, Xgboost regression, and Random Forest regression are the algorithms used for estimation. The algorithms used to forecast the Variance Score and Root Mean Squared Error (RMSE).

[8] Rohit Sav, Pratiksha Shinde and Saurabh Gaikwad proposed a Big Mart sales prediction using Machine Learning The main aim to predict the sales occurred in big mart using machine learning concepts. The Extreme Gradient Boosting algorithm was used for prediction. The Extreme Boosting algorithm gave good accuracy.

[9] Meghana N, Pavan Chatradi, Avinash Chakravarthy V, Sai Mythri Kalavala and Mrs.Neetha K S proposed on Improvising Big Mart Sales Prediction. The main aim is to predict and analysis of the sales.The algorithms used for prediction are Linear regression, Random Forest and Xg Boost. The Random Forest and Xg Boost approach give good accuracy score compared to other models used.

[10] Bodduru Keerthana and Dr K. Venkata Rao Sales prediction on Video Games using Machine Learning. The main aim is to predict and analysis the sales occurred in video game sales. The algorithm used for prediction are Linear regression, Support Vector regression, Random Forest and Decision Tree. Random Forest gave the good accuracy score results with minimum error.

III. DATASET

The dataset of the Sales transactions from a retail store are included in this collection. There are 537577 rows (transactions) in this data collection, with 12 columns (features). User ID is the user's unique identifier. Product ID is the product's unique identifier. The gender of the person making the transaction is indicated by gender. The age of the person making the transaction is indicated by their age group. Occupation displays the user's job title. The user's living city category is city category. Stay In Current City Years is a variable that indicates how long the user has lived in the current city. Product 1, Product 2, and Product 3 are all already numbered. Purchase price of the purchase.

Sl. No	Attribute	Data Types
1	User_ID	integer
2	Product_ID	object
3	Gender	object
4	Age	object
5	Occupation	integer
6	City_Category	object
7	Stay_In_Current_City	object
8	Martial_Status	integer
9	Product_1	integer
10	Product_2	float
11	Product_3	float
12	Purchase	integer

Fig 1: Description of dataset

IV. METHODOLOGY

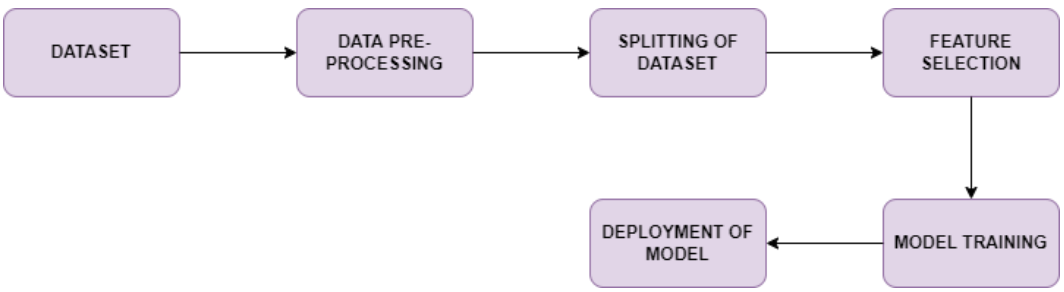

Fig 2: Methodology used for our proposed Diwali Sales Prediction Using Machine Learning

4.1 Dataset: Since a machine learning model is entirely based on data, the very first thing we need to generate one is a dataset. The dataset is a collection of data in a specific format for a perceived issue. The Diwali Sales dataset was gathered from a Bangalore-based corporate gift store.

4.2 Data Pre-Processing: Pre-processing data is the first and most significant step in building a machine learning model. Data cannot be used in its natural form in machine learning algorithms because of the way it was obtained; therefore, data must be developed before being used in machine learning models. And, before performing any data-related operation, it is necessary to clean the data and format it. As a result, we employ data pre-processing tasks. It entails dealing with inconsistent datasets and incomplete data.

- A few columns must be excluded as part of data cleaning because they do not add value to the algorithm's final outcome.
- The missing values must be manipulated so that the data to be fed into the model does not contain any discrepancies.
- By organizing the data in increasing order, we can find the adequate value by calculating the median value and substituting the incorrect entries. This demonstrates that we used data analytics, statistics, and probability.

4.3 Splitting of Dataset: Two distinct datasets are not imported for train and test to prevent overfitting. As a result, the partitioning is done within a single dataset. The data we'll use to train the model is known as the training dataset. Test datasets are those that can be used to predict a test's outcome.

4.4 Scaling of Feature: The scaling of features is a technique for converting data into a precise and customizable size in order to improve the precision and avoid errors. It essentially inhibits the algorithm from using a large variance of data points, allowing us to achieve better results. The class Standard Scaler was imported from the sklearn library.

4.5 Model Construction: The dataset is now capable of building the conceptual approach after completing the initial phases. The model is then used as a predictive model for Diwali Sales Prediction after it has been built.

```
#Reading modified data
train2 = pd.read_csv(“train_modified.csv”)
test2 = pd.read_csv("test_modified.csv”)
inp: - train2.head()
```

In this project, we presented a framework that employsthe Random forest regression algorithm.

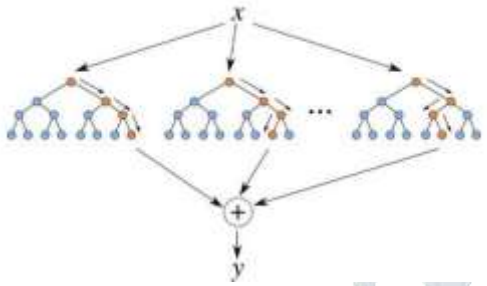

Fig 3: Random Forest Regression

During the last quarter of the year the sales will be high. We will use Random forest regression algorithm under supervised machine learning. There will always be some residual error and we have to find what that error is. We will know what the actual result is and what the predicted result will be. We bind those results to create a confusion matrix. To check the accuracy of our model we have to create the confusion matrix

V. RESULTS AND DISCUSSION

As conventional means failing to help companies thrive in terms of income, the use of machine learning techniques becomes a pertinent reference to deem when profoundly changing a marketing strategy that considers consumers purchasing patterns. Representation of sales based on a variety of aspects, throughout last year's sales, aids business owners in developing appropriate ways to boost demand-driven goods sales. According to this study, when a user tries to estimate which item a purchaser is more likely to buy based on gender identity, age, and profession all play a significant role. Other machine learning techniques can be implemented to the approach to enhance precision, and data preprocessing and interpretation can be improved. More data will be available as the dataset grows. Simulation that are exact for the purpose of bettering the data or obtaining a wider range of data. As an outcome, we'll have to update the dataset with the new information. It has a wide range of aspects. We must improve our performance in order to achieve a better result. Make use of a dataset that is perfectly balanced, with each field having its own value. It contains different values. It is possible to use the same algorithms in different ways. The dataset, however, is the only thing that needs to be changed. Datasets that are large are difficult to manage. For the best results, large datasets are advised

VI. ACKNOWLEDGEMENT

We would like to express our special thanks and gratitude to **Mrs. Swapana G** for her valuable suggestion and useful advice during the planning and development of this project. We would also like to thank all the professors, staff and management of KSIT for their continuous support and encouragement.

REFERENCES

[1] Sumit Kalra, Boominathan, Perumal, Samar Yadav and Swathi Jamjala Narayanan," Analysing and Predicting the purchases done on the day of Black Friday",IEEE (2020) International Conference on Emerging Trends in Information Technology and Engineering (icETITE).

[2] Amruta Aher, Dr. K. Rajeswari and Prof. Sushma Vispute," n Data Analysis and Price Prediction of Black Friday Sales using Machine Learning Techniques", International Journal of Engineering Research and Technology (IJERT) (2021).

[3] Ramachandra H V, Balaraju G, Rajashekar A and Harish Patil," Machine Learning Application for Black Friday Sales Prediction Framework", IEEe (2021) Pune, India

[4] Purvika Bajaj, Renesa Ray, Shivani Shedge, Shravani Vidhate and Prof. Dr. Nikhilkumar Shardoor," Sales Prediction using machine learning algorithms", International Research Journal of Engineering and Technology (IJRET) (2020).

[5] Sunitha Cheriyan, Shaniba Ibrahim, Saju Mohanan and Susan Treesa," Intelligent Sales Prediction Using Machine Learning Techniques", IEEE (2019).

[6] Karandeep Singh, Booma P M and Umapathy Eaganathan," E-Commerce System For Sales Prediction Using Machine Learning",ICCPET (2020).

[7] B.Sri Sai Ramya and K. Vedavathi," Intelligent Sales Prediction Using Machine Learning Techniques", International Conference on Computing, Electronics & Communications Engineering (iCCECE) (2018).

[8] Rohit Sav, Pratiksha Shinde and Saurabh Gaikwad," Big Mart sales prediction using machine learning",International Journal of Creative Research Thoughts(IJCT)(2021).

[9] Meghana N, Pavan Chatradi, Avinash Chakravarthy V, Sai Mythri Kalavala and Mrs.Neetha K S," n Improvising Big Mart Sales Prediction", Journal of Xi'an University of Architecture & Technology (2020).

[10] Bodduru Keerthana and Dr K. Venkata Rao," Sales prediction on Video Games using Machine Learning", Journal of Emerging Technologies and Innovative Research (JERTIR) (2019).