

Machine Learning for Black Friday Sales Prediction Framework

¹Prof. Usha Nandwani, ²Mr. Aniket S. Joshi, ³Mr. Hemant N. Gode, ⁴Mr. Prafulla G. Patil

¹Asst.Prof, ²³⁴UG Student, ¹²³⁴Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharashtra, India.

¹ushanandwani@gmail.com, ²aniketjoshi1040@gmail.com, ³hemant.gode777@gmail.com, ⁴prafullap4597@gmail.com

Abstract - understanding the purchasing patterns of various consumers (Dependent variable) In relation to various product utilize their demographic data (IS characteristics the majority of which our self explanatory. This data set is made up of redundant, unstructured, and all values. The retail industry's domain most frequently uses machine learning. As it will assist which their inventory management, Financial planning, promotion and marketing, this approach helps to produce predictor that has specific commercial value to store owners. Processing modelling, training, testing and assessing all steps in the development of a model. As a result framework will be created to automate some of these procedure, which will lessen there complexity. For the collection of Black Friday tire sales, the algorithm this chose was Random Forest Regressor, with a least RMSE(root mean square error) value of 2829 and an average score of 83.6%. [1]

1. INTRODUCTION

The shopping day after Thanksgiving is referred to as "Black Friday". Because to the high volume of Shoppers who cause vehicular collision and occasionally even violence, this day come to be known as "Black Friday". Police created idiom to describe the chaos caused by the heavy traffic in the downtown shopping areas, both by foot and by car. The volume of sales in the retail sector is crucial in determining the company's loss of profit. Accurate sales forecasting results in effective industry management. Using a refreshable Braille screen and black conversion of the picture to text and ASCII file. Also, a method for using optical character recognition to evaluate segmented characters from movies was suggested (OCR) software. Another investigation of image resizing and conversion was carried out. In the USA, Friday is comparable to a fun fair sale. The things that are in high popularity nowadays are sold in large quantities at extremely low prices. A production model is created to Focus on the kind of product that sells the most in order to increase sales. It is necessary to examine a customer's behavior in order to forecast how many perches they will make on a specific day. In this essay, this'll forecast a company's business on "Black Friday". This must thoroughly arrange the information and investigate the connections among various variables in order to anticipate the profits for different products based on their independent variable. [1]

2. AIMS AND OBJECTIVE

a) Aim

"Black Friday" is the shopping day in which comes right following Thanksgiving. Black Friday there in the USA is analogous to a fair sale. There are currently huge deals taking

place at absurdly low costs for the things that are in hot demand. To boost sales, a prediction models is developed and concentrates on the kind of product that sells best. To forecast the sales of different product depending on their independent variable, this must properly organize the data and study the correlations between specific parts.

b) Objective

The focus of this project is on two goals. Here are some of them :

1. Examining all consumer Data to determine the relationship between independent variables and the target variable doubt this must thoroughly arrange the data and examine the link between various variable in order to anticipate the sale of various products based on their independent variables. In order for a model to do computations and make precise sales predictions.
2. Estimating sales via testing and training the number of sales is in the retail sector.

3. LITERATURE SURVEY

Paper 1: A survey on retail sales forecasting and prediction in fashion markets

AUTHORS : Samanaha Beheshti-Kashi

The work of sales forecasting is crucial in consumer oriented industries like Fashion and electronics, forecasting is difficult due to erratic customer demand, short product life-cycle, and lack of historical sales data. Modern techniques for sales data Studies are given in the research report, with a particular emphasis on fashion and new product forecasting. Also, this study examining several methods for predicting the value of user generated material and search terms. [6]

Paper 2: Shopping with violence : Black Friday sales in the British context

AUTHORS : Oliver Smith, Thomas Raymen.

This article present some preliminary Analysis of the motivations and meanings attached to the behavior of those this begin to understand as “Extreme shoppers” through observation and qualitative interviews. It seeks to understand these behaviors in the context of the social harms associated with consumer culture.[5]

Paper 3: Analysis of consumer data on Black Friday sales using Apriori Algorithm

AUTHORS : Menuka Maharjan

Businesses may negative the layers of data that at first glance appear to be unconnected by identifying and following patterns in the data. Online merchants may easily identify the factors influencing the adoption of online shopping using the study and develop efficient marketing plans. With the ad of project with the aid of project 1 ML Application for the Black Friday sales prediction framework. This Study develops a path for studying consumers online purchasing behavior. Convenience, simplicity of use, and perceived rewards are

the main variable that influence consumers online purchasing decision. While deciding to do your shopping online, security is another important factor. In order to help merchants create effective marketing strategies for sale in their items online and further aid in the growth of the nation, this study will help further examine consumer online buying behavior two hours online shopping.[4]

4. EXISTING SYSTEM

The shopping day immediately after Thanksgiving is referred to as “Black Friday”. Because to the large number of people who cause vehicular crashes and occasionally even violence on this day, it was dubbed “Black Friday”. Police created the idiom describe the Chaos caused by the heavy traffic in the downtown retail areas, both by food and by car. This Looked at some of the things that were already there.

Then, comprehend the connection between the predictors goal variable as well as the connections between the predictors themselves taught a data set. Outlier is a component that stands out from the rest of the data in a glaring way. By utilizing the boxplot approach, this classify the outliers into the current class interval values.

5. COMPARATIVE STUDY

Table 5 : of comparative study

Sr. No.	Paper Name	Author	Tach.	Publications	Purpose
1	A survey on retail sales forecasting and prediction in fashion market	Samaneh Beheshti-Kashi	ML	Researchgate	State of the art method in the sales Forecasting and research With a focus on fashion A new product forecasting.
2	Shopping with violence :Black Friday sales in the British context	Oliver Smith And Thomas Raymen	ML	Researchgate	To understand as ‘extreme shoppers’ And seek To understand this behavior Against The context of the social harms associated with consumers culture
3	Analysis of consumer data on Black Friday sales using Apriori Algorithm	Menuka Maharjan	ML	Semanticssholer	Ability to recognize and track patterns in the data helps businesses shift through the layers of seemingly until later date of a meaningful relationships

6.PROBLEM STATEMENT

Given a person’s past purchasing behavior, You must make a prediction about how much they will pay for a product in the future in order to win the competition. For instance, if a person bought a health supplement the month before, he is most likely a health-conscious individual and is more likely to buy it again the following month. A firm needs this information since it enable them to understand consumer preferences and customize offers for customers when comparing various items.

7. PROPOSE SYSTEM

This will forecast a company’s business on Black Friday in this essay. This must examine the relationship between various factors and carefully organize the data in order to predict the sales of various items based on their independent variables. Such that a model can compute and reliably anticipate sales. By utilizing many decision trees and

bootstrap aggregation methodology, sometimes referred to as bagging, a random forest is an effective method that can carry out both the regression and classification tasks. The main idea behind this is to mix different decision trees in order to determine the final outputs rather than relying just on one decision tree.

ADVANTAGES OF PROPOSED SYSTEM :

Comparative values demonstrate that the suggest technique outperform other current methods in term of accuracy.

Random forest algorithm has a greatest accuracy in predicting Black Friday sales coming in at about 81 %.

Comparing the proper strategy to other recommended ways reveals how effective an appropriate it is. Algorithm: Sales Prediction, Data analysis, Random Forest Regression, Testing, and Training.

8. ALGORITHM

```

STEP 1 - START
STEP 2 - LOAD THE DATASET
    path = settings.MEDIA_ROOT +
    "/" + "train.csv"
    df = pd.read_csv(path)
STEP 3 - FEATURE ENGINEERING
    df.drop(['User_ID', 'Product_ID',
    'Gender', 'City_Category',
    Marital_Status',
    'Product_Category_3'],axis=1,
    inplace=True)
STEP 4 - SPLITTING DATA
    x=df.drop('Purchase',axis=1)
    y = df['Purchase']
STEP 5 - TRAINING THE MODEL
    x_train, x_test, y_train, y_test =
    train_test_split(x, y,
    train_size=0.70,
    random_state=0)
STEP 6 - PREDICTION TEST DATA SET
    from xgboost import
    xgbregressor
    model=
    xgbregressor(learning_rate=0.
    1, max_depth=8,
    min_child_weight=56,
    verbosity=0, random_state=42)
STEP 7 - EVALUTING TH MODEL
    rmse=np.sqrt(mean_squared
    error(y_test,y_pred))
    print("RMSE:%.2f" % rmse)
STEP 8 - END

```

9.MATHEMATICAL MODEL

The mathematical model or a random Forest regression can be expressed as follows:

Even a training data sent $X = \{x_1, x_2, \dots, x_n\}$ with corresponding output where is $Y = \{y_1, y_2, \dots, y_n\}$, where each x_i is a vector of input features and each y_i is a scaler output value:

- For each decision tree t in the random forest
- a. Create a bootstrap Data set X_t buy randomly sampling the training data set with replacement.
- b. Select a random subset of the input feature F_t call is it of decision tree.
- c. Ranger decision tree t on the data set X_t using the features of shit F_t .

For each input vector X , predict the output value Y buy averaging the output value predicted by each tree in the Random forest:

$$y = (1/m) * \sum_{t=1}^m y_t(x)$$

Where m is the number of decision tree in the random forest and $y_t(x)$ is the output value predicted by decision tree t for input vector x .

The Root Mean Square Error (RMSE) is a commonly use evaluation metric for regression models, including those use in Black Friday sales prediction frameworks. The mathematical model for RMSE can be expressed as follow: Given set of predicted values $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$. And corresponding true values $y = \{y_1, y_2, \dots, y_n\}$, where n is the number of predictions:

Calculate the difference between each predicted value and it's corresponding true value:

$$e = \{e_1, e_2, \dots, e_n\}, \text{ where } e_i = \hat{y}_i - y_i.$$

Square each difference:

$$e^2 = \{e_1^2, e_2^2, \dots, e_n^2\}, \text{ where } e_i^2 = (y_i - \hat{y}_i)^2.$$

Calculate the mean of the square differences:

$$MSE = (1/n) * \sum_{i=1}^n e_i^2.$$

Take the square root of the mean square error to obtain the RMSE:

$$RMSE = \sqrt{MSE}$$

Intuitively, the RMSE measures the average magnitude of the errors made by the prediction models, in the same units as the predicted values. A lower RMSE indicate that the model is better at predicting the true values.

In the context of black Friday sales prediction frameworks using machine learning, the RMSE can be used to evaluate the accuracy of the production model Anu validation data set or during training. The model can be trained and optimized minimise the RMSE, may involve tuning hyperparameters, feature engineering, are selecting different algorithms.

10. SYSTEM ARCHITECTURE

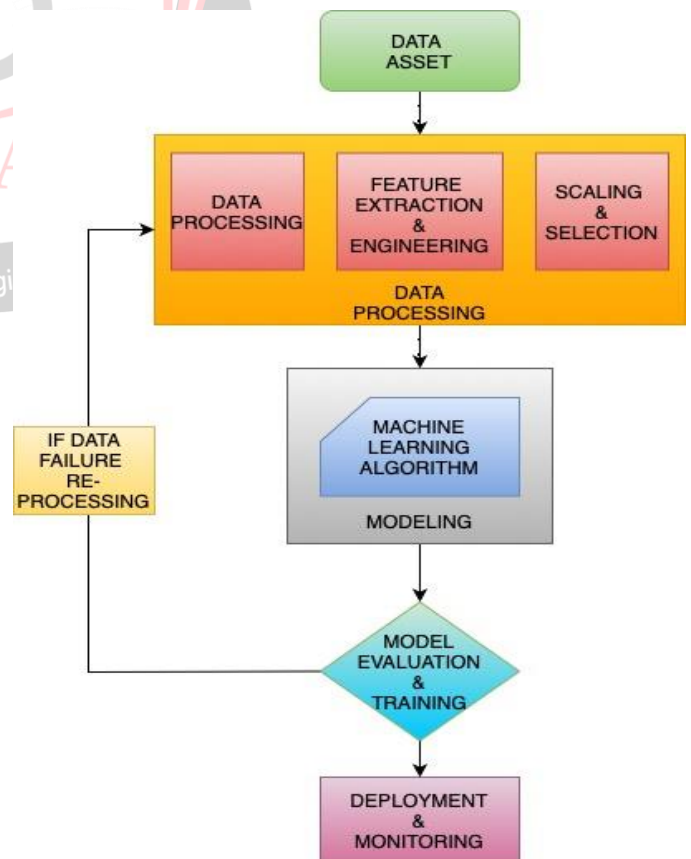


Fig.10 : System Architecture

Description

Step 1 – First registration of user and activate it , Then collect the data from user.

Step 2 – Data processing on user data asset

Step 3 – After data processing system model use algorithm.

Step 4 – Then modelling model evaluation and training on data asset.

Step 5 – If data note recognize properly it back to jump on step 2 .

Step 6 – Then system use deploy result and show monitoring

11. ADVANTAGES

1) in the US, Black Friday purchase still generate an outstanding \$6 billion in income.

2) brick and mortar retailers must work out how to increase sales on the most paper have concluded that our model with lowest XGBoost perform better than existing model.

12. DESIGN DETAILS

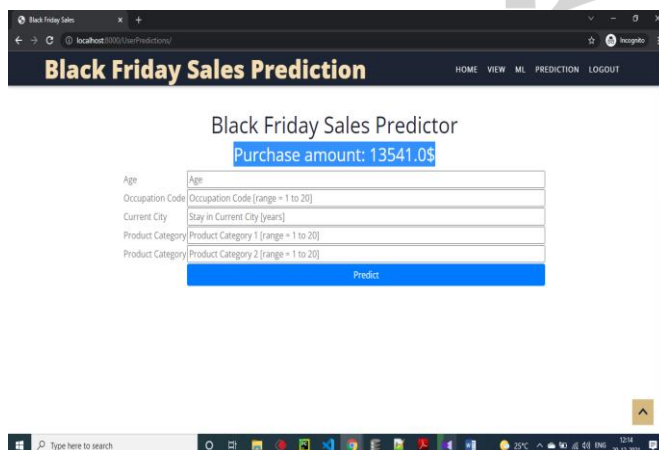


Fig 12 : Result of Design and Details

13. CONCLUSION

Thus we have tried to implement the paper “Black Friday sales prediction using machine learning”, [1] HV Ramchandra; G Balaraju; A Rajshekar; Harish Patil. “Machine learning application for black Friday sales prediction framework” 2021 International conference on engineering smart computing and informatics. IEEE 2021 and the conclusion is as follow:

This paper conclude the exploratory data analysis which is used to find interesting trains from the data set. This research work suggests that when the user tries to predict the product that the customer is more likely to purchase, according to customer’s gender, age and occupation etc., experiments States that our method can produce more accurate prediction when compare to the technique like decision trees, ridge

regression etc. Comparison of various methods are summarized. Also, according to paper have concluded that our model with lowest XGBoost perform better than existing model.

REFERENCE

[1] HV Ramachandra; G Balaraju; A Rajshekar, Harish Patil "Machine Learning Application for Black Friday Sales Prediction Framework" 2021 International Conference on Emerging Smart Computing and Informatics.

[https://ieeexplore.ieee.org/document/9396994/authors#authors]

[2]Mr.Vishal R. Shinde “Record linkage and data prediction” in IJREAM, ISSN: 2454-9150 Volume04 Special Issue-iCreate-2019, APR 2019. (indexed in scope database.Link:https://scopedatabase.com/documents/00000217/00000-95511.pdf)

[3]Alvaro Correia, Robert Peharz and Cassio P. de Campos, "Joints in Random Forests", Advances in Neural Information Processing Systems, vol. 33, 2020.

[4]Maharjan, M. (2019). Analysis of Consumer Data on Black Friday Sales Using Apriori Algorithm. SCITECH Nepal, 14(1), 17–21.

[5] Oliver Smith and Thomas Raymen, "Shopping with violence: Black Friday sales in the British context", Journal of Consumer Culture, vol. 17.3, pp. 677-694, 2017

[6] S.Beheshti – Kashi, H.R. Karimi , K.D. Thoben, M. Lutien 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