BLACK FRIDAY SALES PREDICTION USING MACHINE LEARNING

Mr. R Siva [1], K. Ramya Sri [2], M. Megha Sri [3], M. Asha Sowjanya [4], K. Bhanu Sri [5]

[1] Assistant Professor, Department of CSE, Ramachandra College of Engineering, Eluru, A.P
[2,3,4,5] UG Student, Department of CSE, Ramachandra College of Engineering, Eluru, A.P

ABSTRACT

On Black Friday, every retail store is crowded. Customers rush in to purchase the majority of the items because they are heavily discounted. They offer discounts and deals on different highly branded products including electronics, kitchen appliances, Decor. In some cases, Customers may find it challenging to buy all of the products even with a decent arrangement. In retail industry, the number of sales play an important part that decide the loss a profit for the company. To overcome this problem we are creating a prediction model which predicts the customer choice of products and also identifies how much discount to be assigned to particular product. The main goal is to generate profit for retailers. Here, we are using some machine learning algorithms and Mean Squared Error (MSE) value, which is used as a performance evaluation measure in order to find error value.

1. INTRODUCTION

Black Friday" is the name given to the shopping day after thanksgiving. This day was actually considered as "Black Friday" on the grounds that the number of customers made auto collisions and some of the time even violence. Police begat the saying to depict the disorder encompassing the congestion of pedestrian and auto traffic in downtown shopping regions. In retail industry, the number of sales play an important part that decide the loss a profit for the company. Predicting the sales accurately gives the efficient industry management.

Black Friday is like a carnival sale in the USA. In this day huge sale occurs in a very less price for the products which are much demanded. To incur the sales, a prediction model is made to hover on the type of product which is sold in maximum numbers. A customer's behavior is to be analyzed in order to predict the amount of purchase to be done by him/her on a particular day. In this paper, we will predict the sales of a company on "Black Friday". The Friday that follows Thanksgiving Day, which is observed In the United States on the fourth Thursday of November, is referred to informally as "Black Friday." Although the phrase "Black Friday" did not become commonly used until more recent decades, the day after Thanksgiving has been observed as the start of the Christmas shopping season in the United States since 1952.
On Black Friday, many retailers offer heavily advertised bargains and open very early, like at midnight, or might even start their sales at some point on Thanksgiving .The main challenge for a retail store or e Commerce firm is to pick product prices so that they get the most profit at the end of the sales.

Regression analysis is a form of predictive modelling technique and it investigates the relationship between a target and predictor. This technique is used for time series finding, predicting, and modelling the causal effect relationship between the variables.

A Random Forest is an outfit method that can perform both the regression and classification tasks by using the multiple decision trees and Bootstrap Aggregation technique, generally known as bagging.

Decision Tree is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

2. LITERATURE SURVEY

Ample research is carried out on the analysis and prediction of sales using various techniques. There are many methods proposed to do so by various researchers. In this section, we will summarize a few of the machine learning approaches.

C. M. Wu et al. [1] have proposed a prediction model to analyze the customer's past spending and predict the future spending of the customer. The dataset referred is Black Friday Sales Dataset from analyticsvidhya. They have machine learning models such as Linear Regression, MLK classifier, Deep learning model using Keras, Decision Tree, and Decision Tree with bagging, and XGBoost. The performance evaluation measure Root Mean Squared Error (RMSE) is used to evaluate the models used. Simple problems like regression can be solved by the use of simple models like linear regression instead of complex neural network models.

Odegua, Rising [2] have proposed a sales forecasting model. The machine learning models used for implementation are K-Nearest Neighbor, Random Forest, and Gradient Boosting. The dataset used for the experimentation is provided by Data Science Nigeria, as a part of competitions based on Machine Learning. The performance evaluation measures used are Mean Absolute Error (MAE). Random Forest outperformed the other algorithms with a MAE rate of 0.409178.

Singh, K et al [3] have analyzed and visually represented the sales data provided in the complex dataset from which we ample clarity about how it works, which helps the investors and owners of an organization to analyze and visualize the sales data, which will outcome in the form of a proper decision and generate revenue. The data visualization is based on different parameters and dimensions. The result of which will enable the end-user to make better decisions, ability to predict future sales, increase the production dependencies on the demand, and also regional sales can be calculated.

S. Yadav et al [4] have analyzed and compared the performance of K-Fold cross-validation and hold-out validation method. The result of the experimentations where k-fold cross-validation gives more accurate results. The accuracy results of K - Fold cross-validation were around 0.1 - 3% more accurate as compared to hold-out validation for the same set of algorithms.

Purvika Bajaj et al. [5] have performed sales prediction based on a dataset collected from a grocery store. The algorithms used for experimentations are Linear Regression, K-Nearest Neighbors algorithm, XGBoost, and Random Forest. The result precision is based on Root Mean

Squared Error (RMSE), Variance Score, Training, and Testing Accuracies. The Random Forest algorithm outperforms the other three algorithms with an accuracy of 93.53%.

Ramasubbareddy S. et al. [6] have applied machine learning algorithms to predict sales. The dataset for the experimentation purpose is taken from Kaggle, named as Black Friday Sales Dataset. The algorithms used for the implementation of the system are linear regression, Ridge Regression, XGBoost, Decision Tree, Random Forest, and Rule-Based Decision Tree. Root Mean Squared Error is used as the performance evaluation measure. As per RMSE lower the RMSE value better the prediction. As a result, based on the RMSE rate Rule-Based DT outperforms other machine learning techniques with a RMSE rate of 2291.

Aaditi Narkhede et al.[7] has applied machine learning algorithm in tracking sales at places like shopping center big mart to anticipate the demand of customers and handle the management of inventory accordingly the methods presented here are an effective method for data shaping and decisionmaking. New ways that can better identify consumer needs and calculate marketing plans which will improve sales.

M.Sahaya Vennila et al. [8] have analyzed, preprocessed, and applied machine learning techniques to predict sales. The dataset used for the analysis and experimentation purpose is Black Friday Sales Dataset from Kaggle. The dataset is preprocessed. K - Fold method is used for the purpose of splitting the dataset into training and testing datasets. The prediction model is implemented using Linear Regression Decision Tree, Random Forest, Gradient Boost, and XGBoost. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are used as the accuracy evaluation measures. As a result of experimentation, the Random Forest performed significantly with an accuracy of 77%, with an RMSE value of 2730 and MAE value.

3. EXISTING SYSTEM

- ☐ This existing paper has the problem of internal bias and manual-driven process during the sales forecasting. The algorithm used in this is linear regression.
- ☐ The Black Friday sales are pervaded, all the retailers are packed, & many items are priced for sale at a significantly reduced price.
- ☐ The sales are good that it is hard for clients to obtain the items even with a decent arrangement.

The shop proprietors are more troubled with handling a constrained staff and focusing on prospective customers. While some approaches have been used, they have not been well received.

4. PROPOSED SYSTEM

- ☐ Machine Learning (ML) can be used for the various tasks. So,we are using some of the ML algorithm for the prediction of the amount that a customer is likely to spend on next "Black Friday" sale.
- ☐ It has been performed that the exploratory data analysis is used to find interesting trends from the dataset.
- ☐ In this we are going to predict the product that the customer is more likely to purchase, according to the customer's gender, age and occupation.

☐ The main goal is to generate profit for retailers. so, we are primarily identifying the products which are highly purchased by customers.

ADVANTAGES:

❖ Experiments states that our method can produce more accurate prediction when compared to the techniques like decision trees, ridge regression etc.

❖ A comparison of various methods are summarized. Also, we have concluded that our model with lowest RMSE perform better than exiting models.

The main advantage of this project is that we can use these models to check other sales prediction situation in the future.
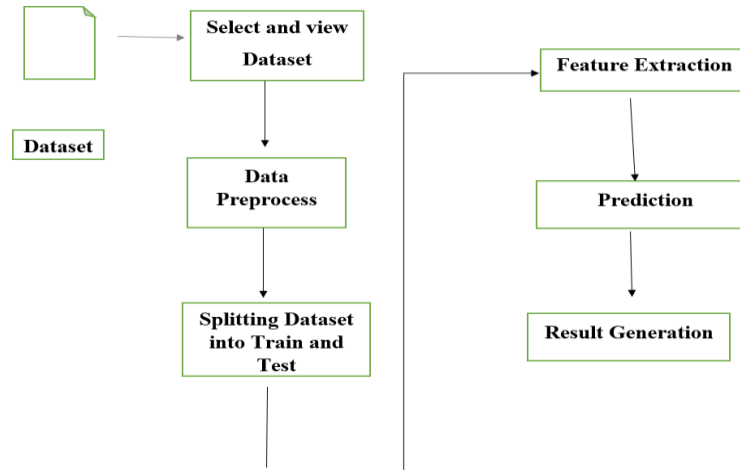
Fig 1 System Architecture

RANDOM FOREST

Random Forest is a classifier that contains a number of decision trees on various subsetsof the given dataset and takes the average to improve the predictive accuracy of that dataset. Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

In our project for random forest algorithm we used 21 estimators that is nothing but a decision trees to generate a model.

LINEAR REGRESSION:

Linear Regression is the supervised Machine Learning model in which the model finds the best fit linear line between the independent and dependent variable.
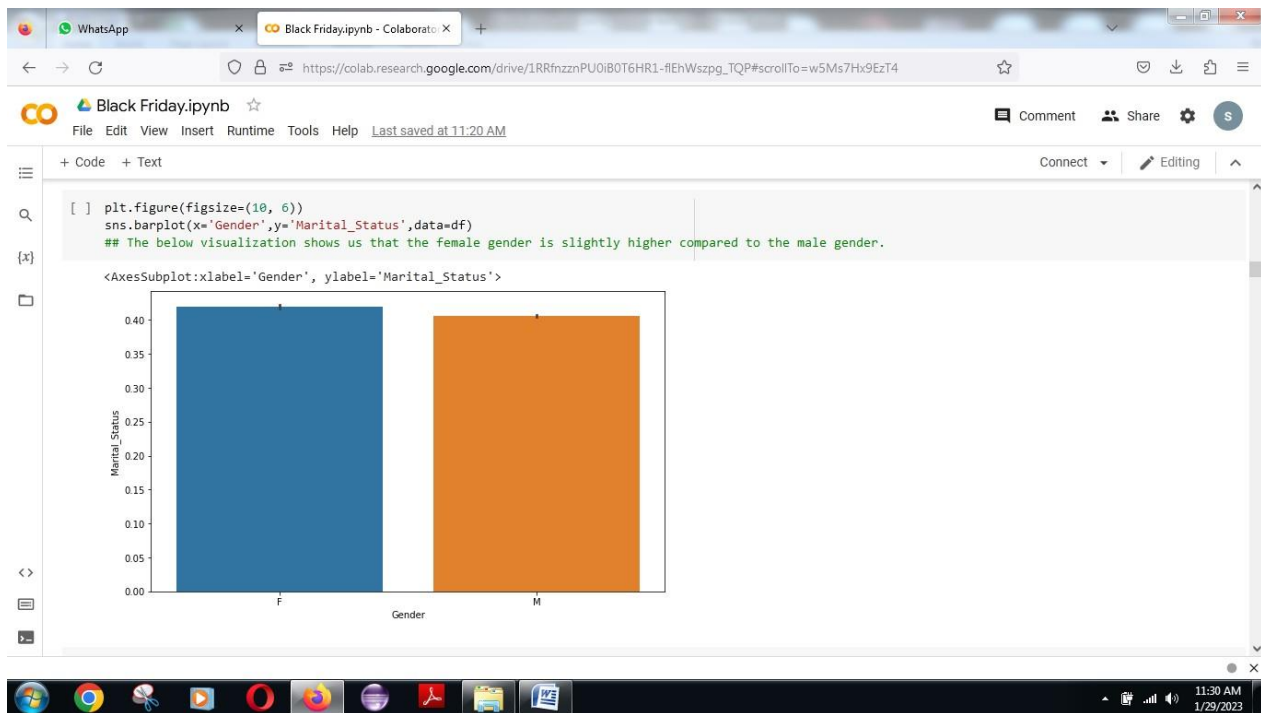
Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable

5. RESULTS

WWW.IJESAT.COM

International Journal of Engineering Science and Advanced Technology (IJESAT)

	Open Access Research Article
IJESAT (Enriching the Research)	Volume: 23 Issue: 07
	July, 2023
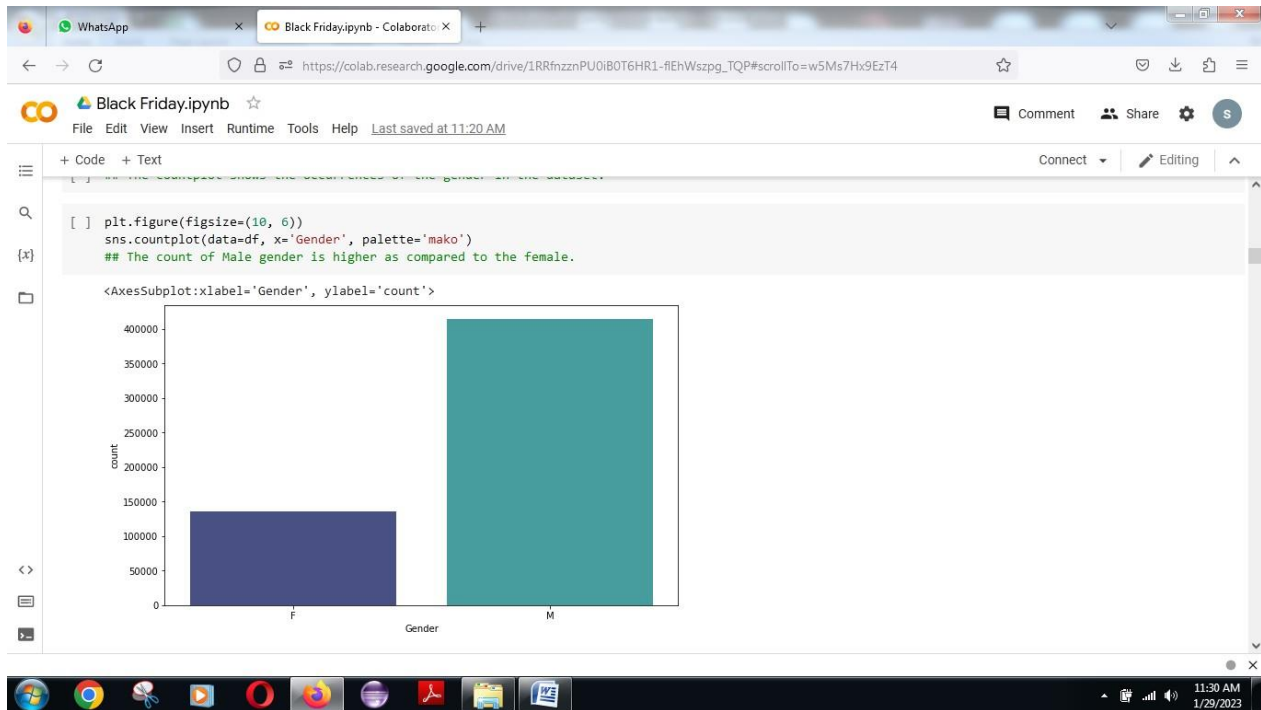
Screenshot 1 Barplot between gender and count

Screenshot 2 Barplot between gender and marital_status
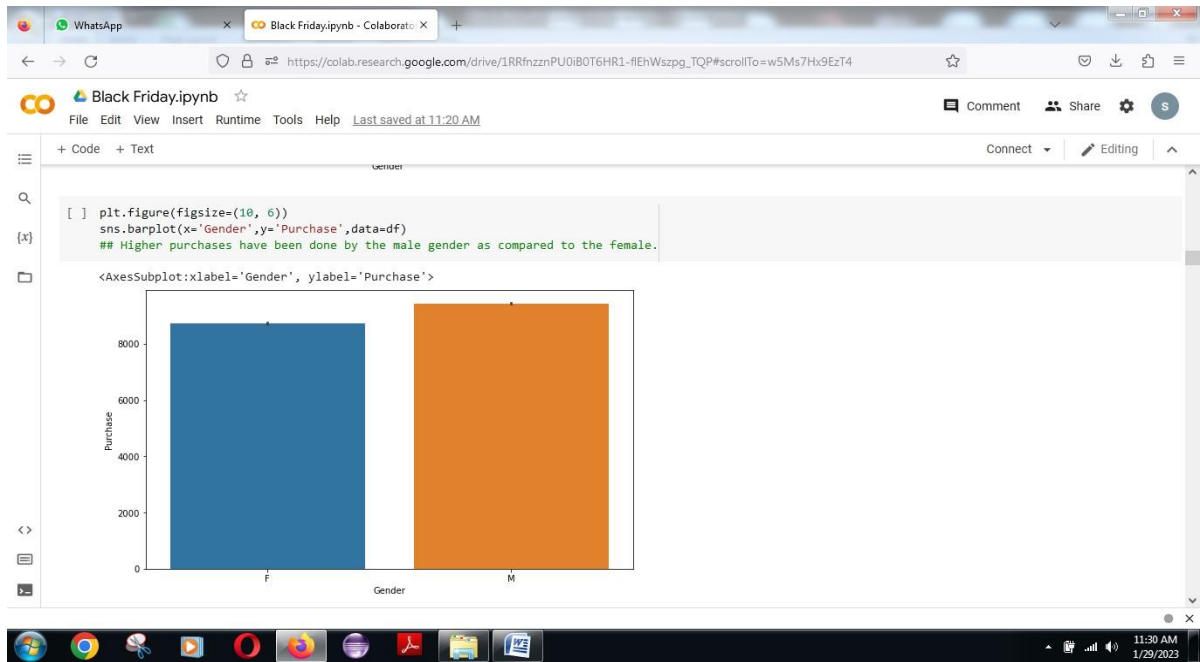
Screenshot 3 Barplot between gender and Purchase
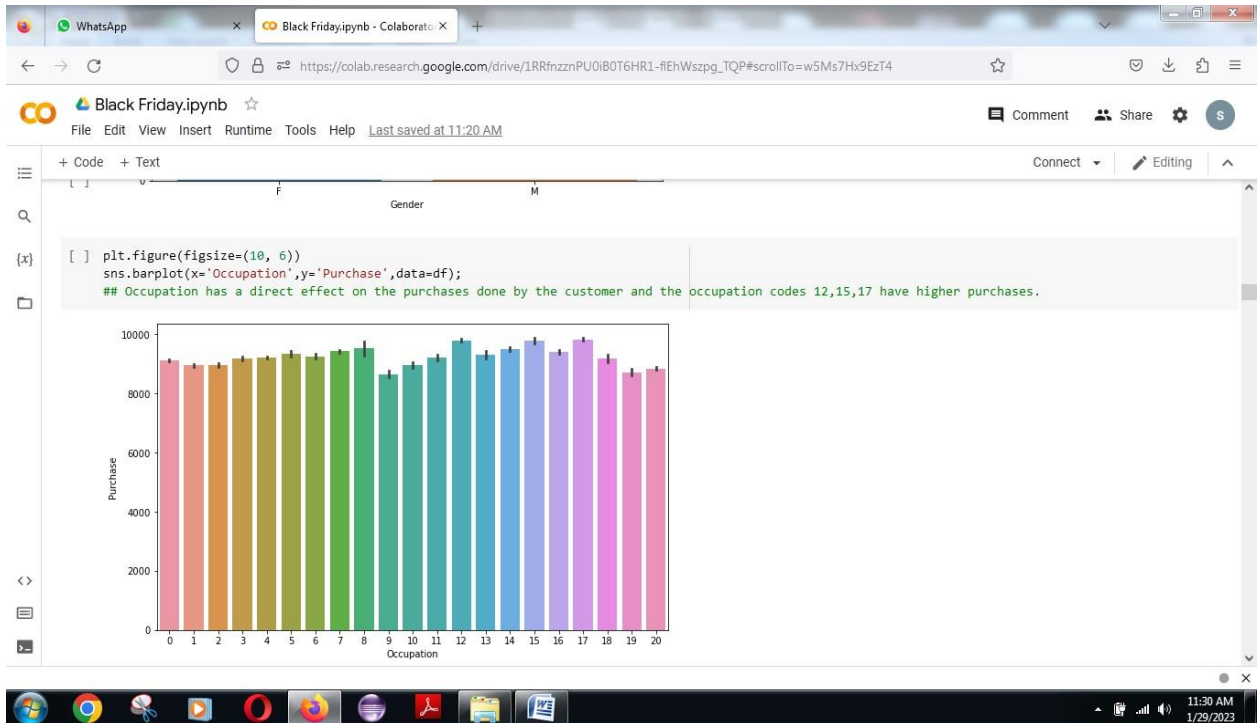
Screenshot 4 Barplot between occupation and purchase
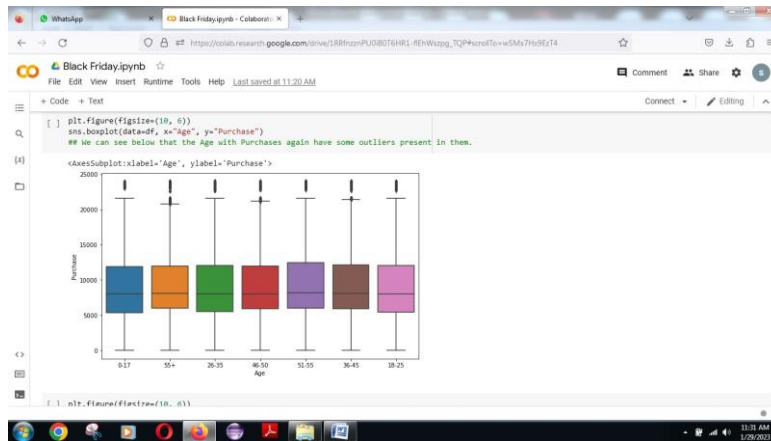
141
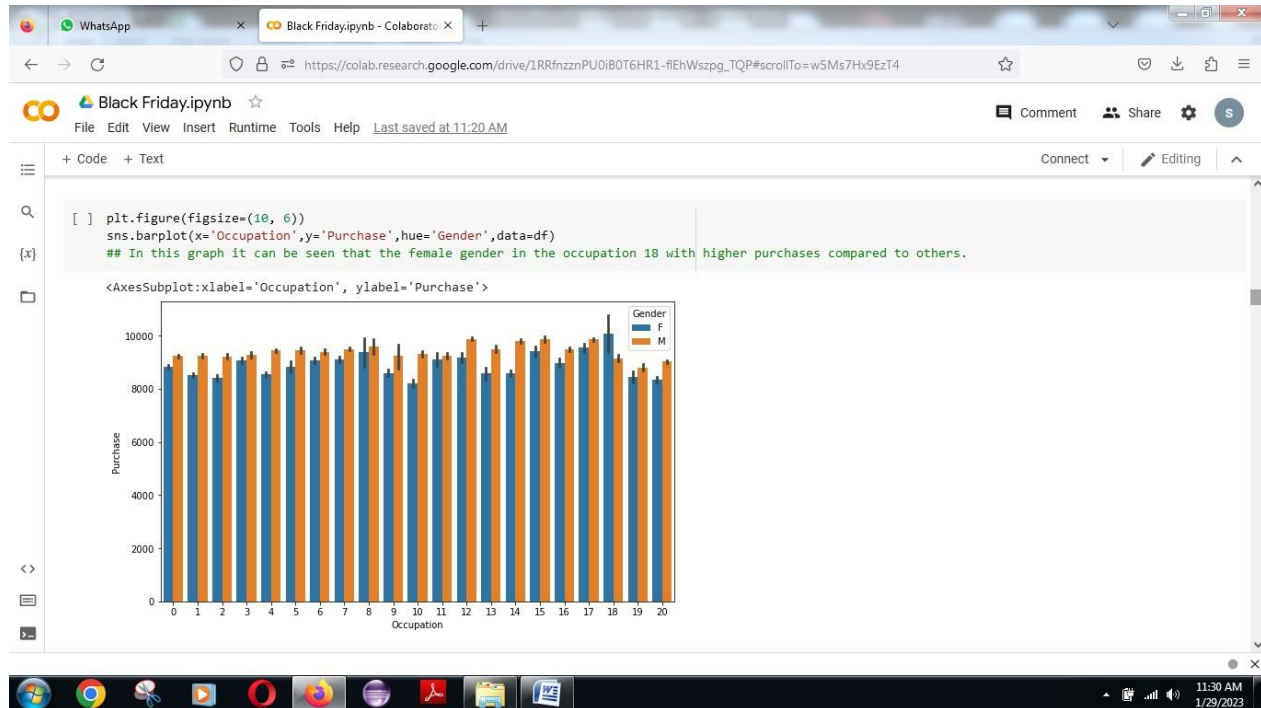
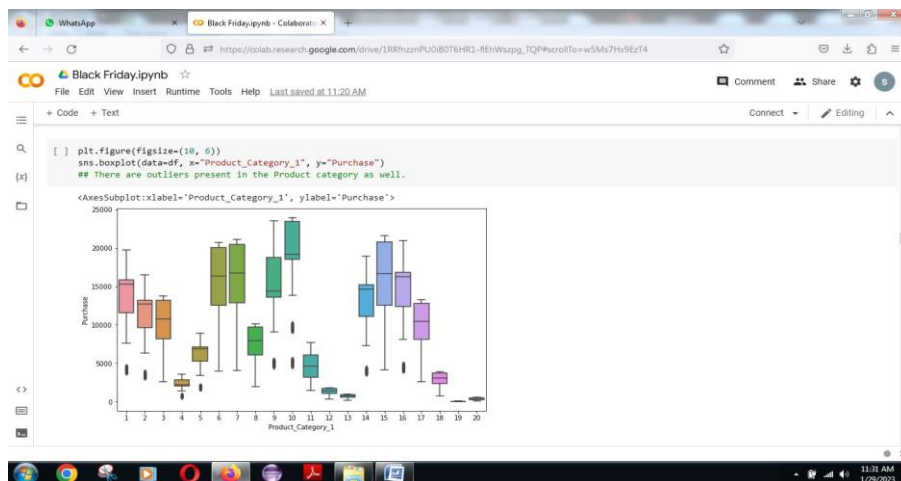**Screenshot 5 Barplot between occupation, purchase and gender**
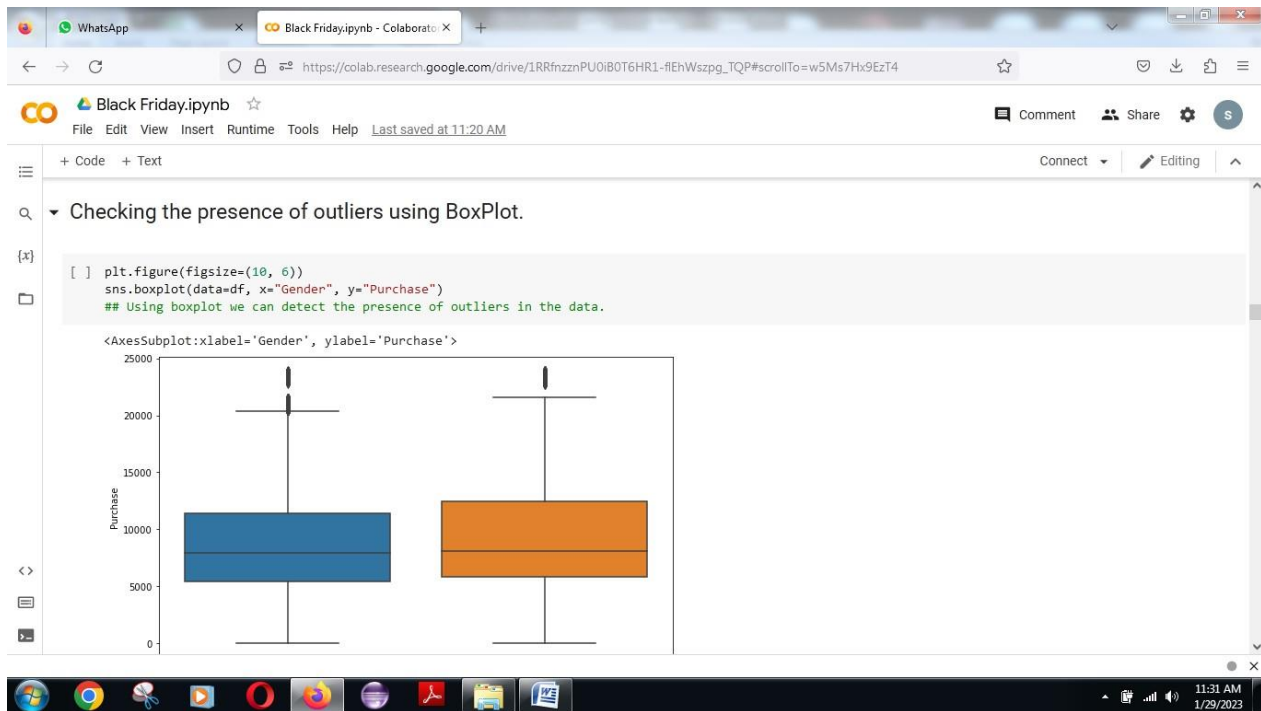
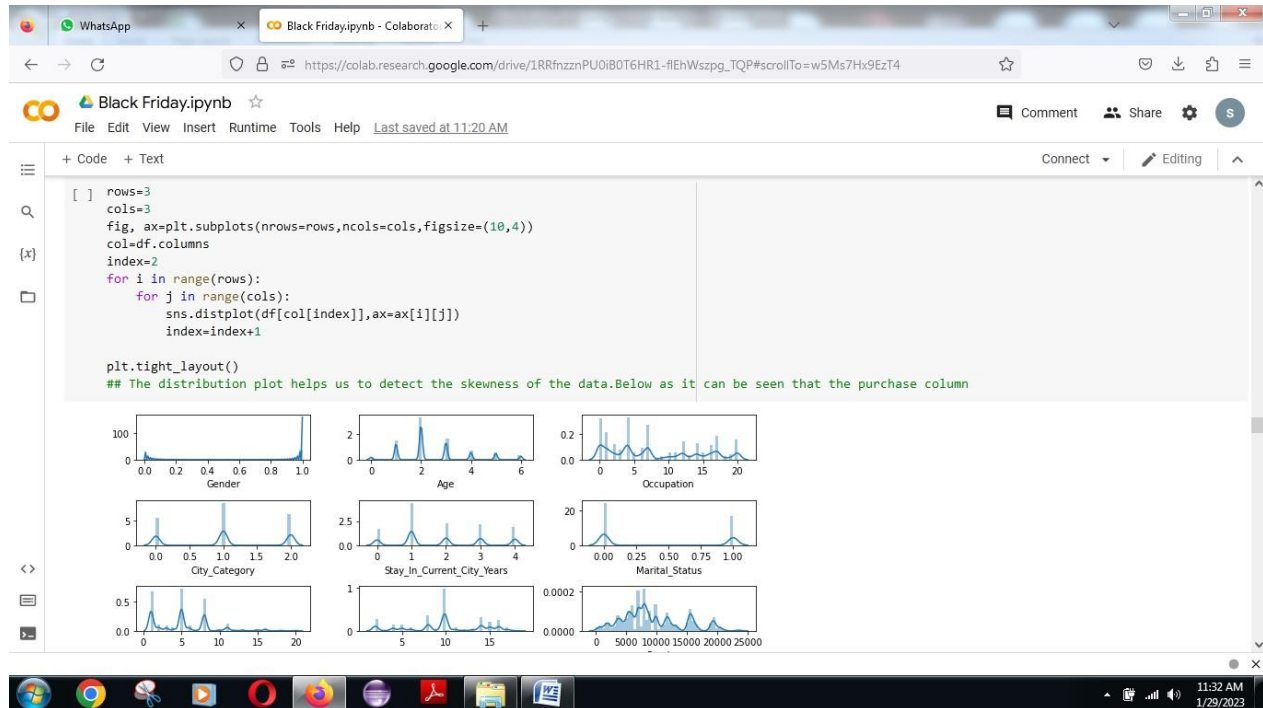**Screenshot 6 Boxplot between gender and purchase**
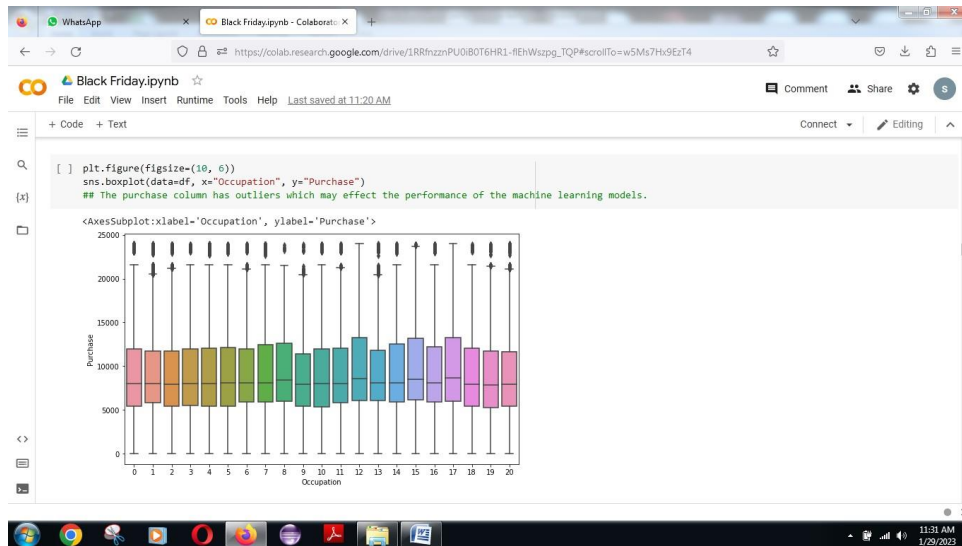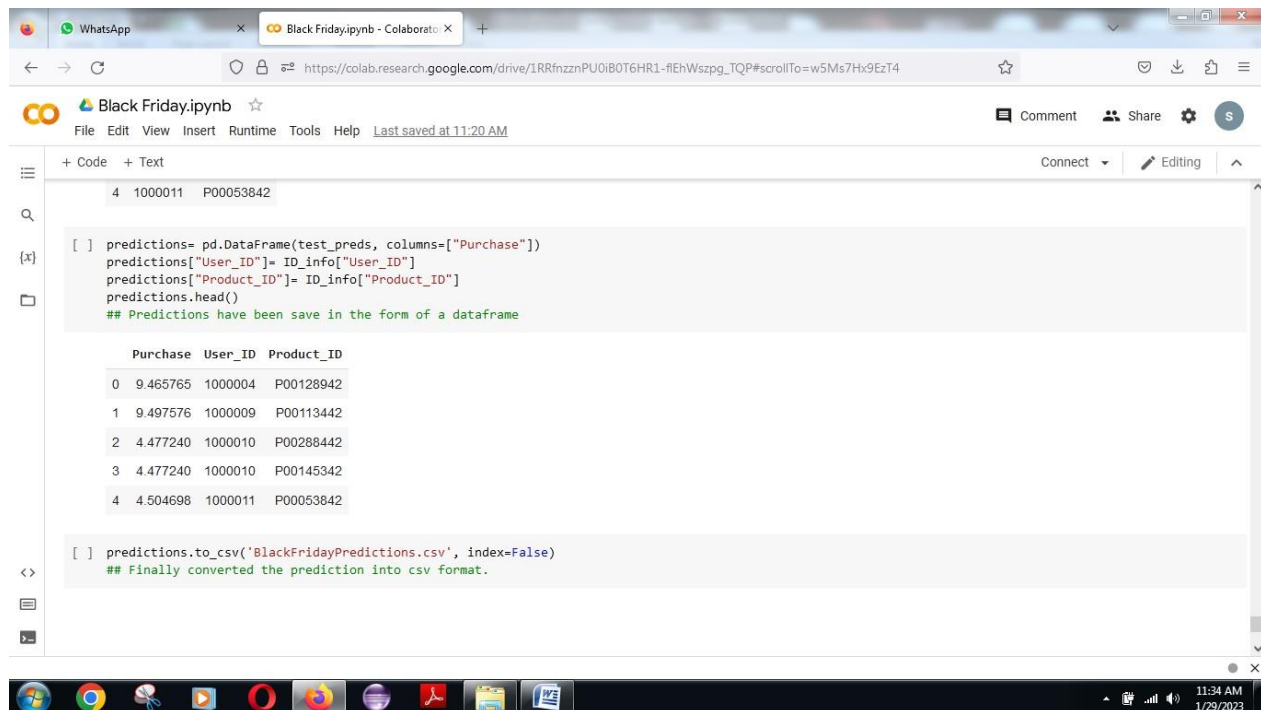
Screenshot 7 Boxplot between purchase and occupation

Screenshot 8 Boxplot between purchase and age

Screenshot 9 Boxplot between purchase and product_category_1

143

Screenshot 10 Distribution plot for all attributes

Screenshot 11 Predicting products by using Random Forest Regression

144

6. CONCLUSION

Machine Learning (ML) can be used for the various tasks. This research work presents the use of ML algorithm for the prediction of the amount that a customer is likely to spend on next "Black Friday" sale. It has been performed that the exploratory data analysis is used to find interesting trends from the dataset. Thus the dataset is used for the experimentation, Black Friday Sales Dataset from Kaggle. The models used are Linear Regression, Lasso Regression, Ridge Regression, Decision Tree Regressor, and Random Forest Regressor. The evaluation measure used is Mean Squared Error (MSE). Based on Table II Random Forest Regressor is best suitable for the prediction of sales based on a given dataset. Thus the proposed model will predict the customer purchase on Black Friday and give the retailer insight into customer choice of products. This will result in a discount based on customer-centric choices thus increasing the profit to the retailer as well as the customers. As future research, we can perform hyperparameter tuning and apply different machine learning algorithms.

REFERENCES

1. S. Beheshti-Kashi, H.R. Karimi, K.D. Thoben, M. Lutjen 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.
2. 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.
3. Goutam. Majumder, "ANALYSIS AND PREDICTION OF CONSUMER BEHAVIOUR ON BLACK FRIDAY SALES", Journal of the Gujarat Research Society, vol. 21.10, pp. 235-242, 2019.
4. Challagulla, B Venkata Udaya et al., "Empirical assessment of machine learning based software defect prediction techniques", International Journal on Artificial Intelligence Tools, vol. 17.02, pp. 389-400, 2008.
5. C.W. Chu and G.P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting", International Journal of production economics, vol. 86, no. 3, pp. 217-231, 2003.
6. S. Makridakis, S.C. Wheelwright and R.J. Hyndman, Forecasting methods and applications, John wiley & sons, 2008.
7. Alvaro Correia, Robert Peharz and Cassio P. de Campos, "Joints in Random Forests", Advances in Neural Information Processing Systems, vol. 33, 2020.
8. Kvalheim, Olav Martin et al., "Determination of optimum number of components in partial least squares regression from distributions of the root-mean-squared error obtained by Monte Carlo resampling", Journal of Chemometrics, vol. 32.4, pp. e2993, 2018.
9. Robert P. Sheridan et al., "Extreme gradient boosting as a method for quantitative structure– activity relationships", Journal of chemical information and modeling, vol. 56.12, pp. 2353- 2360, 2016
10. Kee Yuan Ngiam and Wei Khor, "Big data and machine learning algorithms for health-care delivery", The Lancet Oncology, vol. 20.5, pp. e262-e273, 2019.
11. P.M. Domingos, A few useful things to know about machine learning. Communacm, vol. 55, no. 10, pp. 78-87, 2012.
12. P. Langley and H.A. Simon, "Applications of machine learning and ruleinduction", Communications of the ACM, vol. 38, no. 11, pp. 54-64, 1995.