Dr. Manju Arora, 2024, 12:3
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology

An Open Access Journal

Black Friday Sales Prediction Using Machine Learning: An Overview

Dr. Manju Arora, Shanoor
Department of Information Technology
JaganNath University, Bahadurgarh Haryana

Abstract- With a major impact on consumer behaviour and sales patterns, Black Friday has emerged as a key event in the retail calendar. Through an analysis of past sales data, economic variables, and consumer emotion, this study seeks to anticipate Black Friday sales. Large online retailers like Amazon, Flipkart, and others entice buyers with sales and discounts across a variety of product categories on Black Friday. This day starts early in the evening, sometimes even a few days in advance, and the stores provide heavily marketed and discounted bargains. This study aims to comprehend, using their demographic data, the purchasing patterns of a varied range of consumers (dependent variable) with respect to different products. Computers can make better decisions because machine learning places a strong emphasis on "learning." Machine learning models can make more accurate predictions and make better judgments based on past experiences. The philosophy covered in this work aids in the creation of a prediction model that will be very helpful to sales administration on Black Friday.

Keywords- Sales prediction, Regressor, Mean Squared Error, Random Forest, Machine Learning.

I. INTRODUCTION

The Internet revolution has brought about significant changes in the shopping sector. Retailers are fighting for market share as this phenomenon has gained global traction, with consumers expecting substantial reductions. Precise forecasts of Black Friday sales are crucial for firms looking to streamline their processes and profit from this crucial time. The main advantages of shopping online are its convenience, increased choice, cheaper costs, ease of price comparison, lack of crowds, etc. The pandemic has increased internet purchasing.

Businesses are increasingly using data-driven strategies to improve sales forecasts on Black Friday. Through meticulous organization and analysis of consumer data, their goal is to identify correlations between independent factors and the target variable—here, sales of different products—in this context. The developed prediction model will assist in examining the connections between different attributes, so optimizing their earnings around this crucial purchasing occasion. Finding significant correlations between many variables requires careful investigation and effective data organizing.

It makes it possible to estimate sales for different products with accuracy depending on their independent variables. Dataset is used for training and prediction. The developed prediction model will offer a forecast depending on the customer's age, city category, occupation, discount, and other factors.

Models such as random forest regression, ridge regression, lasso regression, and linear regression are the foundation for the implementation of the prediction model.

II. LITERATURE

Using a dataset of tire Black Friday sales, the Random Forest regressor produced an average accuracy of 83.6%. The research illustrates the value of the machine learning framework for shops by predicting Black Friday sales with an average accuracy of 83.6% and an average RMSE of 2829. It does this through using demographic data and customer purchasing time[1]. Machine learning models are more precise in making decisions and estimate events in the future. The research project covers the developing of a forecasting model with machine learning to help sales administration forecast Black Friday sales outcomes [2]. With a success rate of 99.21%, the prediction model utilizing multiple regression methods predicts customer requests and raises retail store revenues. In this research, we investigate ways retail stores can increase sales success, profits, and customer behaviour forecasting with high accuracy through using machine learning techniques [3]. The field of sales and marketing is being significantly influenced by advancements in machine learning. Using multiple approaches to machine learning and factors such as item and outlet details, the article explores the projected level of sales in supermarkets [4].

III. METHODOLOGY

The goal of this research endeavour is to provide a unified framework that combines age and gender estimation with cutting edge object detection. Our aim is to improve the precision and effectiveness of demographic analysis across a range of applications by merging these processes. Our method combines both conventional and deep learning techniques to estimate age and gender from identified objects using a forest regressor, among other sophisticated techniques.

1. Sales Data

Sales transactions from a retail store are included in the dataset, which presents a great chance to explore feature engineering and learn important lessons from a variety of shopping experiences. Attributes in the dataset include user_id, product_id, marital_status, city_category, occupation, and more.

Table 1 mentions the definition of the dataset.

Sr No	Variable	Definition
1	User_Id	Unique Id of Customer
2	Product_Id	Unique Product Id
3	Gender	Sex of Customer
4	AGE	Customer Age
5	Occupation	Occupation of Customer
6	City_Category	City Category of Customer
7	Stay_In_ Current City	Number of Years Customer Stays In City
8	Maritial_Status	tomer Marital Status
9	Product_Category_ 1	Product Category
10	Product_Category_ 2	Product Category
11	Product_Category_ 3	Product Category
12	Purchase	Amount of Customer Purchase
13	Discount	Amount of Discount Given to Customer

The Black Friday deals dataset is used to forecast the number of purchases made by customers on Black Friday deals and to train different machine learning models based on discount.

Sales Data Source

https://www.kaggle.com/code/midouazerty/black-friday-sales-prediction/input?select=train.csv

2. Data Preprocessing

In order to align with the "sales" dataset's structure, a new column called "DISCOUNT" is introduced to

the "test" dataset. This column contains NaN values for test data. The "DISCOUNT" column is a new addition. The distribution of categories in the dataset is used to impute random values to the missing data in the "Product_Category_2, Product_Category_3" column.

This entails actions such as dividing data into training and testing sets, encoding categorical variables, scaling numerical features, and handling missing values. These procedures guarantee that the dataset is prepared for additional modelling and analysis in the research.

The next step after data preparation is Exploratory Data Analysis (EDA).

I created a visual representation of the client base's distribution according to factors like gender, age, occupation, city category, length of stay, discount, and marital status. This makes it possible to learn more about the client base's gender distribution, age distribution, professional experiences, place, residence stability, and marital statistics.

3. Sales Estimate

Analyzing a variety of variables, including past sales data, economic indicators, and consumer behaviour patterns, is necessary for sales estimation. Use Random Forest Regressor to estimate the feature on the object. This calls for the extraction and assessment of high accuracy.

4. Assessment and Evalution

To guarantee the precision and dependability of predictive models, historical data must be compared for validation. Metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used in this procedure to measure how much the actual and forecasted sales statistics differ from one another, to understand your model performance.

5. Comparison with the Foundational Model

This baseline usually denotes a rudimentary methodology, frequently employing elementary statistical metrics such as historical means or basic linear regression. Through comparing the predicted accuracy of sophisticated models with this reference point, we can understand the small improvement that has been made. By comparing them, this comparison makes sure that the predictive models used in the retail industry are not only highly developed but also much better than less complex alternatives, maximizing their usefulness in guiding strategic decisions.

6. Implementation of User Interface

To guarantee functionality and user-friendliness, a sequential process is involved in using Tkinter for a Black Friday sales forecast program. After that, the UI layout is created using Tkinter widgets, which include buttons, entry fields, and labels, arranged to encourage user interaction and input.

Implementation of Workflow Diagram and Gantt Chart

Project Initial Planning (12-Mar)

- Establish the project's objectives, deliverables, scope, and schedule.
- Assign team members jobs and duties.
- List the resources, tools, and datasets that are necessary.

Data Preprocessing (22-Mar)

- Collect and clean the dataset.
- Handle missing values, outliers, and inconsistencies.
- Normalize or standardize data as needed.
- Convert categorical variables to numerical formats if necessary.

Exploratory Data Analysis (EDA) (01-Apr)

- To comprehend data distribution and patterns, conduct preliminary data investigation.
- Use graphs and charts to visually represent data in order to spot patterns and correlations.
- Condense important findings to guide the creation of the model.

Model Development and Training (11-Apr)

- For prediction, choose the proper machine learning algorithms.
- Divide the data into sets for testing and training.

- Train several models with the use of training data.
- For the best possible model performance, adjust the hyperparameters.

Model Evaluation and Selection (21-Apr)
- Evaluate models using appropriate metrics (e.g., MAE, RMSE).
- Compare model performance against the baseline.
- Select the best-performing model based on evaluation results.
- Perform cross-validation to ensure model robustness.

Prediction Analysis and Scenario Testing (01-May)
- Use the selected model to make predictions on new data.
- Analyze prediction results and interpret findings.

Deployment Preparation (06-May)
- Develop a user interface (UI) using Tkinter for the prediction application.
- Integrate the trained model into the UI.
- Test the application for usability and functionality.
- Prepare deployment documentation and user guides.

Project Wrap-Up and Presentation (11-May)
- Prepare a comprehensive project report summarizing the entire process.
- Create a presentation highlighting key findings, methodologies, and results.

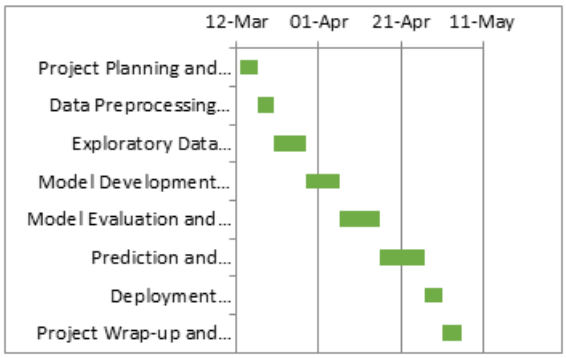

Fig 1: Gantt Chart

Algorithm

Gathering and Preparing Data
- Gather previous sales information and any pertinent data.
- Handle missing numbers, outliers, and inconsistent data to make the data cleaner.
- To guarantee consistency, normalize or standardize numerical properties.
- Use methods like one-hot encoding to convert categorical variables (such gender, city, and current city) to numerical representations.

Analyzing exploratory data (EDA)
- With the use of statistical summaries and visualizations, examine the distribution and trends of important properties.
- To help with feature selection, find relationships between features and sales.
- Obtain knowledge from data visualizations to direct the creation of models.

Engineering Features
- Provide additional features (such as age groups and the discount percentage) that could improve the model's ability to forecast outcomes.

Model Choice
- Select suitable machine learning techniques (e.g., random forests, linear regression, etc.) for sales prediction.
- To assess the performance of the model, divide the dataset into training and testing sets.

Training Models
- Utilizing specific algorithms, train several models on the training dataset in order to maximize performance.

Model Assessment
- Utilize measures such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to assess each model's performance on the testing dataset. additionally Model performance is compared to a baseline model.

Model Choice

- Based on evaluation, choose the model that performs the best in order to increase accuracy.

The Planning of Deployment

- Using Tkinter, create an intuitive user interface (UI) that lets users enter data and get sales projections in real time by integrating the trained model into the UI.

Final Presentation and Deployment

- Release the program to users, and write a report on the project.

IV. DATA AND RESULT

1. Dataset

We used a dataset that records many facets of consumer behaviour and transaction details during Black Friday events for our study on Black Friday sales prediction. User and product IDs, customer demographics (including gender, age, marital status, and occupation), city and product categories, discount and the amount spent by customers on purchases are some of the dataset's key elements.

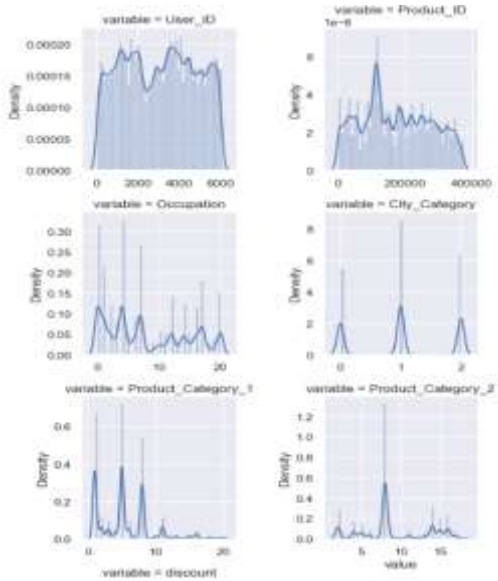

Fig 2: Sales Prediction

A study of our prediction models' output and the learnings from their evaluation. Key results like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared showed the models' predictive power, as did their increased accuracy in Black Friday sales forecasts.

We evaluated the significance of different factors in forecasting Black Friday sales and compared their performance.

2. Result

The evaluation of the integration process was successful in terms of sales prediction based on discount.

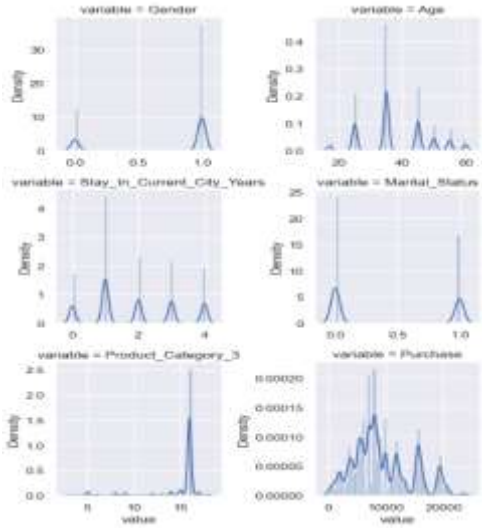

Fig 3: Sales Prediction_Using variables

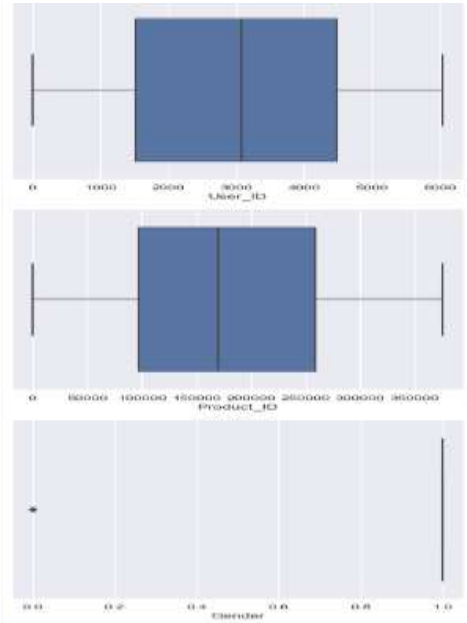

Fig 4: Box Plot(USER_ID, PRODUCT_ID ,gender)

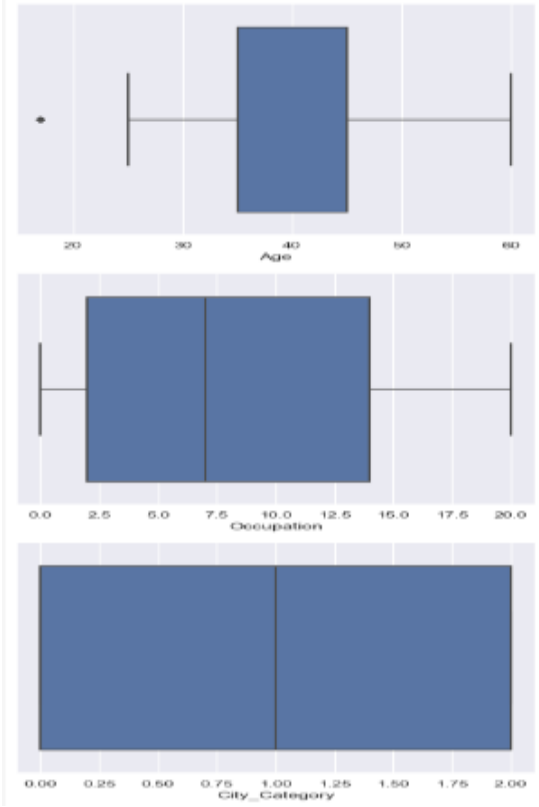

Fig 5: Box Plot (Age, Occupation, City_Category)

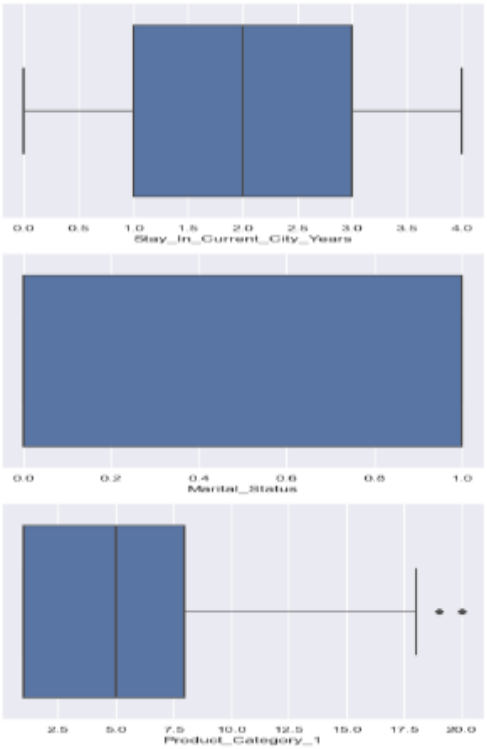

Fig 6: Box Plot (Stay_in_Current_City_Year, Maritial_Status, Product_Category_1)

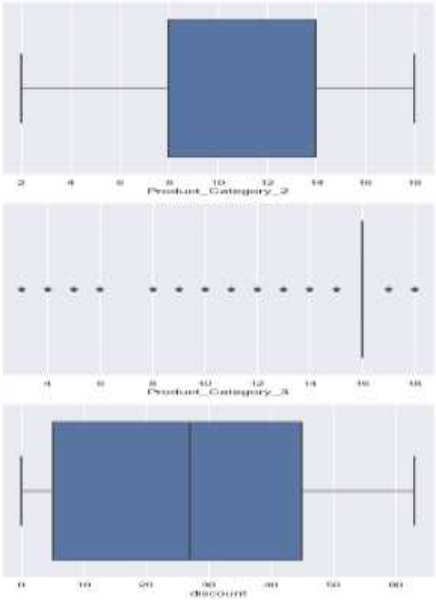

Fig 7: Box Plot (Product_Category_2, Discount)

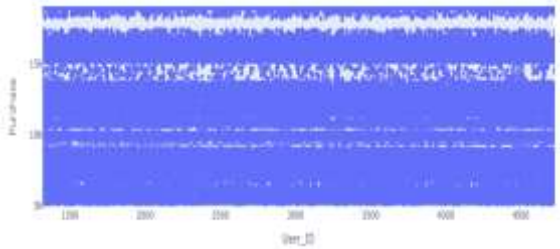

Fig 8: Scatter Plot between Product_ID and Purchase

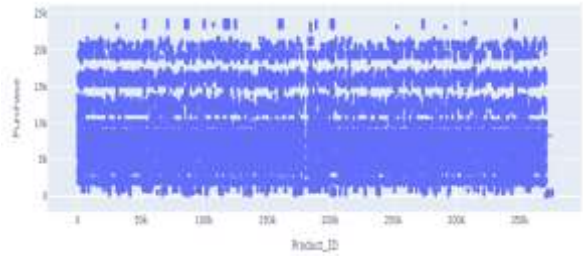

Fig 9: Scatter Plot between Product_ID and Purchase

Fig 9: Scatter Plot between Gender and Purchase

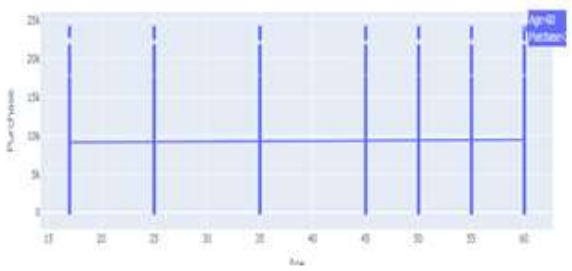

Fig 10: Scatter Plot between Age and Purchase

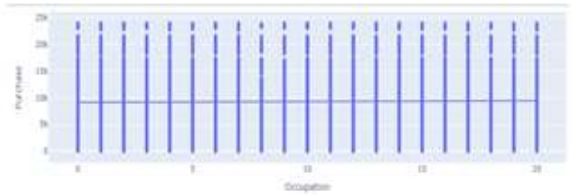

Fig 11: Scatter Plot between Occupation and Purchase

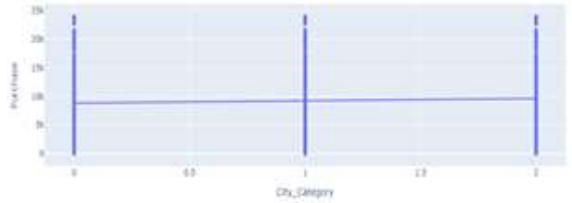

Fig 12: Scatter Plot between City_Category and Purchase

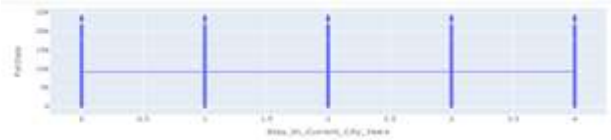

Fig 13: Scatter Plot between Stay_In_Current_City_Years and Purchase

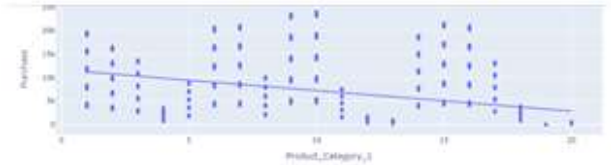

Fig 14: Scatter Plot between Product_Category_1 and Purchase

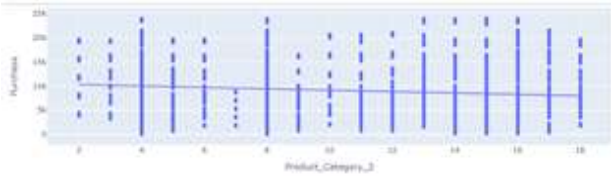

Fig 15: Scatter Plot between Product_Category_2 and Purchase

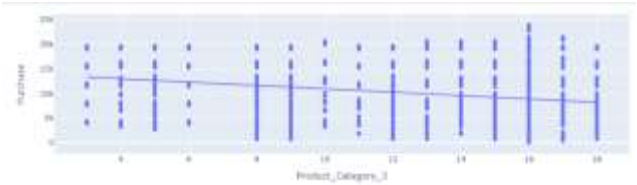

Fig 16: Scatter Plot between Product_Category_3 and Purchase

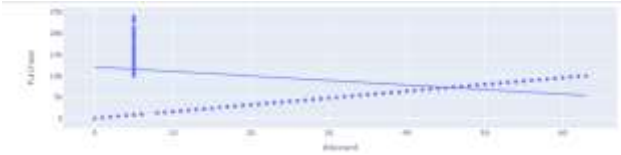

Fig 17: Scatter Plot between discount and Purchase

	Score in percentage
Linear Regression	26.562919
Lasso Regression	26.562929
Ridge Regression	26.562919
Random Forest Regressor	89.711642

Fig 18- Output 1

Fig 19- Output 2

V. CONCLUSION

In conclusion, our study on Black Friday sales prediction has provided insightful information on the workings of the retail industry and customer behavior during this important shopping event. We have shown how successful machine learning algorithms are in accurately forecasting sales through the creation and assessment of predictive models. Retailers may respond to shifting market conditions and provide customers with a more customized purchasing experience by incorporating

predictive models into their decision-making processes.Our research concludes by highlighting the importance of data-driven strategies in retail analytics and the potential of predictive modeling to propel business success in the hectic environment of Black Friday sales.

Limitations

Although our research on Black Friday sales prediction provided valuable data, it is crucial to acknowledge some inherent drawbacks of the study.

- The extent of analysis and predictive model accuracy may have been limited by the dataset's quality and accessibility.
- Predictive models may find it difficult to reflect the complex relationship between personal preferences, social expectations, and outside effects on purchase decisions due to the complexity of customer behavior.
- Another drawback is temporal behavior that is Black Friday sales patterns are subject to change over time as a result of consumer choices, market trends, and economic situations. These temporal modifications may not be fully captured by predictive models based on historical data, which could restrict their ability to accurately forecast future sales.

REFERENCES

1. Ramachandra, H. V., Balaraju, G., Rajashekar, A., & Patil, H. (2021, March). Machine learning application for black friday sales prediction framework. In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 57-61). IEEE.
2. Patil, S., Nankar, O., Agrawal, R., Sharma, K., Awasthi, S., & Jha, N. (2023, January). Black Friday Sales Prediction using Supervised Machine Learning. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 1006-1012).
3. Alagarsamy, S., Varma, K. G., Harshitha, K., Hareesh, K., & Varshini, K. (2023, January). Predictive analytics for black friday sales using machine learning technique. In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 389-393).
4. Kiran, J. S., Rao, P. S., Rao, P. P., Babu, B. S., & Divya, N. (2022, January). Analysis on the prediction of sales using various machine learning testing algorithms. In 2022 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.