

# Black Friday Sales Detection Using Machine Learning

Ms. TAHERA ABID, MIR MOHAMMED  
RAHMATH ULLAH<sup>2</sup>, SAYED AHMED ALI<sup>3</sup>,  
MOHAMMED FARHAN KHAN<sup>4</sup>

Assistant Professor<sup>1</sup> UG Scholar<sup>2,3,4</sup>  
Department of Information Technology

NAWAB SHAH ALAM KHAN COLLEGE OF ENGINEERING & TECHNOLOGY  
New Malakpet, Hyderabad-500024.

**Abstract:** -We may be able to ascertain, with the help of their demographic data, the manner in which various customers—the dependent variable—purchase various products (IS characteristics, the majority of which are self-explanatory). There are unstructured, duplicate, and null values in this dataset. In the retail sector, machine learning is the most frequently used application. Because it will assist with inventory management, financial planning, advertising, and marketing, this method contributes to the production of a predictor that will be of great economic benefit to store owners.

Preprocessing, modeling, training, testing, and evaluation are the phases of building a model. Systems will be made to smooth out and to some extent mechanize this cycle. We suggested the Random Forest regressor with an average accuracy of 83.6 percent and a minimum RMSE (Root Mean Squared Error) value of 2829 using the Black Friday sales dataset.

**Keywords:** - Frameworks, minimum RMSE, self-explanatory data, and repetitive demographic information.

## INTRODUCTION

This occurred when gold market monopolists Jay Gould and James Fisk used their connections to the Award Organization as a means of financial gain.

President Grant after being made aware of the deception, ordered the Treasury to release a sizeable amount of gold. The run ended, and gold prices fell by as a result. 18%. Abel Corbin, the president's own brother by marriage, was destroyed, and fortunes were earned and lost in a single day. The phrase "The greatest shopping day of the year" was first used in the publication Assembling Plant The Leaders and Backing in November 1951 and again in to describe the day following Thanksgiving in 1952. In this instance, the technique of having employees call in sick the day following Thanksgiving to receive a four-day holiday was being discussed. No indication is given that this use has proliferated in any way. Around the same time, the police in Philadelphia and Rochester started calling the start of the holiday shopping season "The shopping extravaganza following Thanksgiving" and "Dark Saturday" to describe the crowds and traffic jams. In 1961, an advertising expert suggested

renaming the days of the week as "Large Friday" and "Enormous Saturday" as Philadelphia's civic and business community worked to improve circumstances. These words, however, were swiftly forgotten. The forecast model developed will provide an expectation based on the client's age, city class, occupation, and other factors. The prediction model is implemented using models like linear regression, ridge regression, lasso regression, decision tree regression, and random forest regression. The word "dark" has been used to describe disaster-filled days for a very long time. Many events have been described to as "Black Friday," which occurred when lenders Jay Gould and James Fisk took advantage of their connections with the Award Organisation in an effort to control the gold market, although the Panic of 1869 was the most significant such event in American history. President Grant instructed the Treasury to release a significant amount of gold after learning about the fraud. This stopped the run and made gold prices go down by 18%. Fortunes were made and lost in a solitary day, and the president's own brother by marriage, Abel Corbin, was demolished.

### LITERATURE SURVEY

**Title:** "Black Friday Sales - An effective analytical study to understand its impact"

**Year:** 2021

**Authors:** Ninad Loya, Shreya Nipane, Tania Dakua, Vaibhav Sharma

**Abstract:**

The day after Thanksgiving, which is observed in the Assembled State, is the busiest shopping day of the year. The holiday shopping season is the busiest, which aids retailers in increasing sales. Despite being widely disseminated over an extended period of time, the buyers didn't pay much attention to this instruction. In order to determine how well-liked this deal is in India, an online survey was carried out using a form. 280 users contributed information. The results of this study suggest that, for a variety of reasons, Indians are less aware of the Black Friday bargains.

**Key words:** - Black Friday, marketing, social media.

**Title:** "Black Friday A study of consumer behavior and sales predictions"

**Year:** 2021

**Authors:** Keerthan, Potturi

**Abstract:**

This study was conducted with the intention of observing and analysing consumer purchasing and behaviour patterns. The US's busiest shopping day for Christmas is the day right after Thanksgiving because it gives companies a wonderful opportunity to extend their sales. This method has become popular in various countries over time.

Despite being a worldwide phenomenon, little study has been done on customers' attitudes and behaviour on Black Friday. The goal of the study was to learn more about Greek customers' Black Friday-related actions, attitudes, and intentions. Also included in this study's objectives were the identification of marketing variations between Black Friday shoppers and non-shoppers as well as ways for merchants to take advantage of these variations. The information obtained from clients and the strategy used in the online survey research The study's findings show that Black Friday buying habits have significantly changed.

**Key words:** Black Friday, retail, consumer behavior, Sales patterns, Purchasing, Patterns.

**Title:** "Effect of Online Marketing on Black Friday Sales in India"

**Year:** 2019

**Authors:** Shubham Bantia<sup>1</sup>, Swapnil Joshi<sup>2</sup>, Pooja Darda<sup>3</sup>.

**Abstract:**

The modern buying experience now includes an online component. Businesses using online purchasing platforms have multiplied in the twenty-first century. The Friday following Thanksgiving, known as the big shopping day after Thanksgiving, and the Monday following Thanksgiving, known as the online Christmas sales extravaganza, are two key seasonal

shopping days that many U.S. retailers view as usually positive. The data used in this study came from a poll of 100 or so Indian consumers who were familiar with the idea of Black Friday and had probably made an online purchase during the sale. This article also touches on the results of numerous companies' online marketing initiatives to increase Black Friday sales. discounts, sales, social media marketing, online marketing, Black Friday, etc.

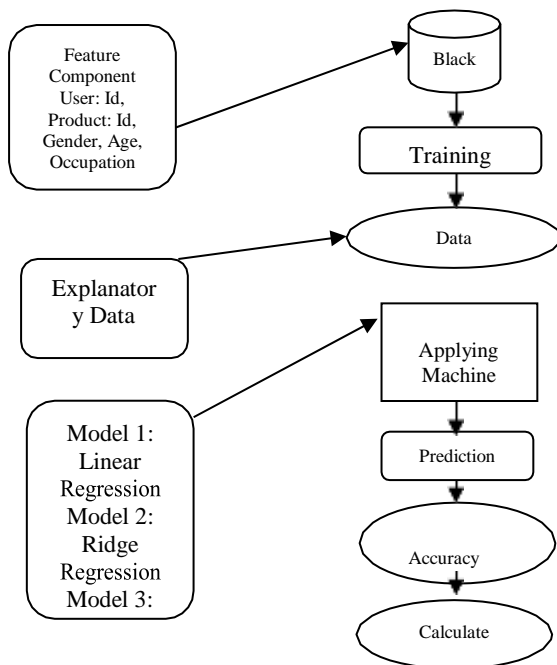
## **SYSTEM ANALYSIS**

### **Objectives:**

To understand the inventory management system by online retailers

### **Proposed System:**

The input data set is the Black Friday sales data set. Information Perception Of The Different Ascribes Of This Dataset Is Performed. Data preprocessing primarily entails filling in missing values. The categorical values are encoded with a label into a number. The Classifications, For example, Orientation Where F Addresses Female And M Addresses Male Is Switched Over completely to Mathematical Structure As 0 And 1 Additionally Other Unmitigated Qualities, For example, City Class, Stay\_In-CurrentCity, Age Are Changed Over completely to



**Fig: proposed system**

Numerical Form By Applying Label Encoding.

The Qualities Like Client: ID and the Good; To train the model without bias based on user: Id are removed. Id or Good: id, as well as to improve performance. Linear regression, ridge regression, lasso regression, decision tree regression, and random forest regressor are the algorithms utilized in the system's implementation. Fivefold cross-validations are used to train the models. Mean Squared Error (MSE) is the performance evaluation metric that is utilized. With a MSE Score of 3062.719, Random Forest Regress or performs better than the other algorithms.

**Advantages:**

- Increasing the accuracy score.
- For the live example android app in our project.
- Large amount of feature we are taking for the training and testing.

**Existing System:**

Machine learning models like Linear Regression, MLK classifier, Deep learning model employing Keas, Decision Tree, Decision Tree with bagging, and XGBoost are available.

**Drawbacks:**

- Small level dataset.
- Less amount of accuracy score.
- Applicable on small level prediction work.

**Modules And Their Functionalities:**

1. Data Ingestion.
2. Data Preprocessing.
3. Exploratory data analysis.
4. Correlations.
5. Evaluation Metric.

**1.Data Ingestion:**

Information ingestion is the process of moving data from a number of sources to a storage medium, where it is often accessed, utilised, and assessed by a company. Usually, the destination is a database, data warehouse, data mart, or document storage facility. A source could be almost anything, including SaaS data, internal applications, data sets, accounting sheets, or even data that has been scraped from the internet. The foundation of any analytics architecture is the data ingestion layer. The reporting and

analytics procedures that follow require readily available and trustworthy data. There are numerous ways to ingest information, and there are numerous models or structures that can be used to construct a specific information ingestion layer.

## **2 Data Preprocessing:**

Data pre-processing is a data mining technique used to convert raw data into a format that is both practical and effective. Here, the data travels through two steps.

**Data Cleaning:** Data accuracy and the absence of unwanted information are essential. As a result, the data is cleaned up prior to moving on to the next stage. Information must be cleaned up by looking for missing components, copying records, organising invalid data, and getting rid of it. Data Pre-processing is a data mining technique used to convert the raw data into a format that is both practical and effective. Here, the data travels through two steps.

### **Data Transformation:**

Data transformation is the mathematical transformation of the datasets. The data is altered into the proper structures required for the information mining process. We are able to understand the data even better by meticulously organising the hundreds of documents. Change can be broken down

into standardisation, normalisation, and quality selection.

### **Exploratory Data Analysis:**

Exploratory data analysis (EDA) is a strategy for better understanding datasets through the use of visual components including scatter plots, bar graphs, and other diagrams. As a result, we are better equipped to do the proper analysis and identify the data trends with greater accuracy. The graphs of yearly patterns demonstrate how the annual planting and harvesting areas have an impact on US exports. In spite of the worldwide yield, China's imports have increased over the past 12 years despite a steep decline in exports in 2009. This shows that China has a sizable and ongoing need for soybean crops, but that it currently depends on the supply from other countries to meet its needs.

### **Software and Hardware Requirements:**

The software and hardware requirements are listed below.

#### **Software Requirements:**

- Operating System :Windows 7, 8 and 10  
(32 and 64 bit)
- Front End :Python
- Packages : Numpy, Pandas, Matplotlib, Sklearn.

#### **Hardware Requirements:**

- Processor - Dual Core

- RAM - 2 GB or above
- Hard Disk - 100 GB or above.

### System Architecture:

### Data Ingestion:

Data ingestion is the process of transferring data from numerous sources to a storage medium so that an organisation may access, utilise, and analyse it. The usual places to go are a database, a data warehouse, a data mart, or a document store. Nearly anything can be used as a source, including Sea's data, internal software, databases, spreadsheets, and even content that was plucked from the internet.

### Data Preprocessing:

1. Data pre-processing is a data mining technique used to convert raw data into a format that is both practical and effective. This information travels through two stages.

#### 2. Data Cleaning:

The absence of errors and undesired data in data is crucial. So, before moving on to the next phases, the data is cleaned. Cleaning up data involves looking for and eliminating any duplicate records, missing information, and improper formatting..

#### 3. Data Transformation:

Data transformation is the mathematical transformation of datasets; data is changed into usable formats for the data mining

process. By organising the hundreds of records in an organised manner, this enables us to comprehend the material more thoroughly. Normalisation, standardisation, and attribute selection are all parts of transformation.

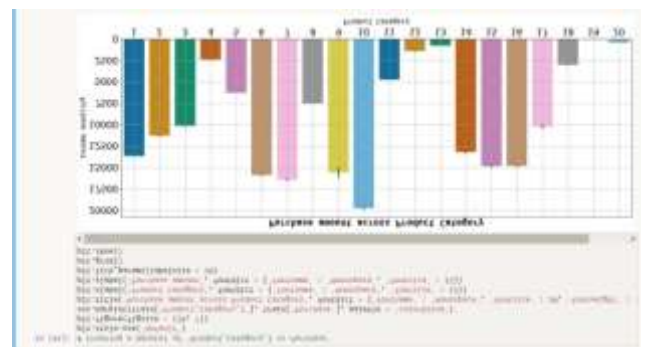
### Results Screenshots:



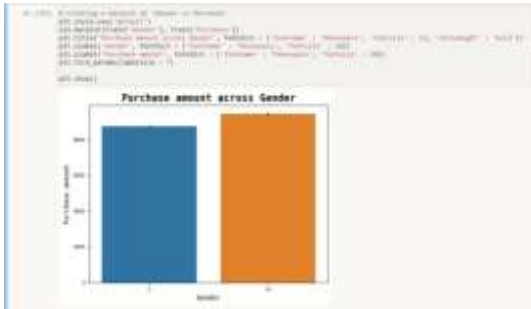
Fig : DistPlot



Fig : Bar Plot for Gender



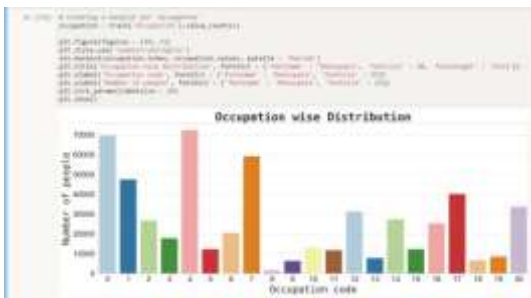
**Fig : Bar Plot of Product Category Vs Purchase**



**Fig : Bar Plot of Gender Vs Purchase**



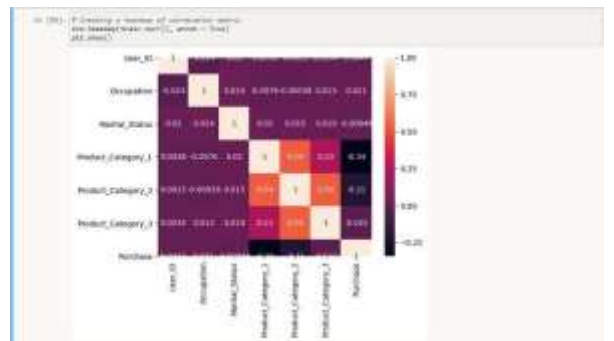
**Fig: Tree Map for Stay In Current City Years**



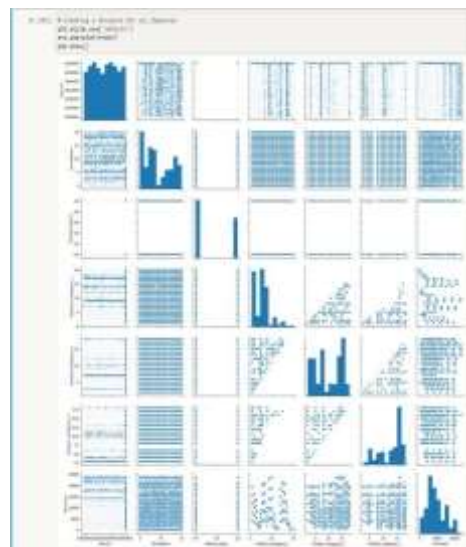
**Fig: Bar Plot For Occupation**



**Fig : Pie Chart For Cit Category**



**Fig : Heat Map of Correlation Matrix**



**Fig : Pair plot For All Features**

**CONCLUSION AND FUTURE ENHANCEMENT**

**Conclusion:**

Utilizing a machine learning algorithm to predict a customer's "Black Friay" spending

is demonstrated in this study. To find interesting patterns in the dataset, it has been decided to use exploratory data analysis. According to this study, the user's gender, age, and occupation all play a role when they try to predict which product a client is more likely to buy based on their gender, age, and occupation. Tests show that our system, when contrasted with methods like choice trees and edge relapse, can yield more precise expectations.

In this project, we attempted to either construct a model employing various algorithms like linear regression, KNN regression, decision tree regression, random forest regression, and XGB regress, or to make the most accurate prediction.

The hyper boundary tuned XGB relapse or gives us the best rimes esteem and r2 score for this issue. Machine learning, or ML, can be used for a wide range of tasks.

#### **Future Enhancements:**

We have an enormous enough dataset, so we can utilize brain organizations, for example, a fake brain organization to fabricate a model which can bring about better execution.

If retailers want Black Friday to continue to be a profitable event, they must make effective use of this strong online trend.

The huge shopping day after Thanksgiving is never going to lose its name and popularity in the shopping aspect. Shopping will see numerous additional advancements. Robotic arms, self-driving cars, and drones that drop off packages at our homes will soon be commonplace.

We don't have to go to the store to buy the products, but there will be people who prefer to shop in person because it feels more natural. This will help boost in-store sales.

#### **REFERENCES**

1. Beheshti-Kashi, S., Karimi, H.R., Thoben, K.D., Lutjen, M., Teucke, M.: "A survey on retail sales forecasting and prediction in fashion markets," *Systems Science & Control Engineering* 3(1), 154, 161(2015).
2. Smith, Oliver, and Thomas Raymen. "Shopping with violence: Black Friday sales in the British context." *Journal of Consumer Culture* 17.3 (2017): 677-694.
3. Majumder, Goutam. "ANALYSIS AND PREDICTION OF CONSUMER BEHAVIOUR ON BLACK FRIDAY SALES." *Journal of the Gujarat Research Society* 21.10s (2019): 235-242.
4. Challagulla, Venkata Udaya B., et al. "Empirical assessment of machine

- learning based software defect prediction techniques.” *International Journal on Artificial Intelligence Tools* 17.02 (2008): 389-400.
5. Chu, C.W., Zhang, G.P.: “A comparative study of linear and nonlinear models for aggregate retail sales forecasting,” *International Journal of production economics* 86(3), 217{231(2003)}.
  6. John Makridakis, S., Wheelwright, S.C., Hyndman, R.J.: “Forecasting methods and applications,” wiley& sons (2008).
  7. Correia, Alvaro, Robert Peharz, and Cassio P. de Campos. “Joints in Random Forests.” *Advances in Neural Information Processing Systems* 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* 32.4 (2018)...
  9. Sheridan, Robert P., et al. “Extreme gradient boosting as a method for quantitative structure–activity relationships.” *Journal of chemical information and modeling* 56.12 (2016).
  10. Ngiam, Kee Yuan, and Wei Khor. “Big data and machine learning algorithms for health-care delivery.” *The Lancet Oncology* 20.5 (2019): e262-e273.