

BLACK FRIDAY SALES PREDICTION

1.INTRODUCTION:

1.1Introduction:

Black Friday is a shopping day for a combination of reasons. As the first day after the last major holiday before Christmas, it marks the unofficial beginning of the Christmas shopping season. Additionally, many employers give their employees the day off as part of the Thanksgiving holiday weekend. Thanksgiving was the first holiday celebrated in America. It was first celebrated in the autumn of 1621 when the Wampanoag Indians and the pilgrims got together for a three-day feast and festival of fun. Today, families celebrate Thanksgiving by eating a big turkey dinner and enjoying pumpkin pie.

Black Friday has routinely been the busiest shopping day of the year in the United States since 2005 although news reports, which at that time were inaccurate have described it as the busiest shopping day of the year for a much longer period of time Similar stories resurface year upon year at this time portraying hysteria and shortage of stock, creating a state of positive feedback

1.2 Objectives of Research:

The primary objective of this project is to predict whether the customer the purchased the item that was in sale . It gives a brief details about each and every person who purchased the particular brand was belongs to which city There will a particular branded products that will reduce the prices up to 50% and that particular product will have more demand in that particular day. So this project gives the brief details about each and every person who purchased in the Black Friday belongs to which city.

1.3 Problem Statement:

The problem statement of this project is if a person so called 'X' as purchased something on Black Friday Then to which city that particular person X belongs to.

2. REVIEW OF LITERATURE:

Since 2016, we've noticed a trend. Black Friday ads are being released earlier. Ads are showing up as early as October 1. And, while most major retailers still wait

until November, many of the big players (including Target, Macy's and Best Buy) crept up their ads last year by several days (from after Nov. 10 to second week of November).

This year, we predict the major players will start releasing their ads around November 1, with all of the major big-box retailers releasing by Nov. 8. Amazon, though, will continue to bide its time and wait until around Nov. 15. And eBay will continue its tradition of releasing its ad at the very last minute.

This year, we're hoping for discounts on the nearly-\$400 Google Max, which is known for its great sound quality (see the review). Meanwhile, Apple's \$350 Home Pod (new this year) will probably remain full price.

3. DATA COLLECTION:

Prediction models were generated by logistic regression (LR), Support vector machine (SVM), Random Forest Classifier and Multiple Linear regression . A comparison of these models was conducted to determine which method produced the best predictors. To assess the likelihood of prediction . so when compared the optimal solution with high accuracy is by using Random Forest Classifier . by using that we can optimally predict the accuracy that the items purchased by a person belongs to which city.

4. METHODOLOGY:

Brief Description of Algorithms Used

Support Vector Machine (SVM):

SVM is one of the standard set of supervised machine learning model employed in classification. Given a two-class training sample the aim of a support vector machine is to find the best highest-margin separating hyperplane between the two classes. For better generalization hyperplane should not lies closer to the data points belong to the other class. Hyperplane should be selected which is far from the data points from each category.

Decision Tree Classifier:

Decision Tree is a supervised machine learning algorithm used to solve classification problems. The main objective of using Decision Tree in this research work is the prediction of target class using decision rule taken from prior data. It uses nodes and internodes for the

prediction and classification. Root nodes classify the instances with different features. Root nodes can have two or more branches while the leaf nodes represent classification. In every stage, Decision tree chooses each node by evaluating the highest information gain among all the attributes.

Random Forest:

Random Forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because it's simplicity and the fact that it can be used for both classification and regression tasks. In this post, you are going to learn, how the random forest algorithm works and several other important things about it.

K-Nearest Neighbors:

K-Nearest Neighbors (KNN) is one of the simplest algorithms used in Machine learning for regression and classification problem. KNN algorithms use a data and classify new data points based on a similarity measures (e.g. distance function). Classification is done by a majority vote to its neighbors. The data is assigned to the class which has the most nearest neighbors. As you increase the number of nearest neighbors, the value of k, accuracy might increase.

Logistic Regression:

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1.

Accuracy Measures:

SVM, Decision Tree, Random Forest, KNN, Logistic Regression algorithms are used in this research work. Experiments are performed using internal cross-validation 10-folds. Accuracy, confusion matrix and ROC (Receiver Operating Curve) measures are used for the classification of this work.

From all the methods we used the best model that is **random forest** model, because the model gives the correct accuracy

4.1 Explanatory Data Analysis:

The dataset here is a sample of the transactions made in a retail store. Classification problem can be settled in this dataset since several variables are categorical, and some other approaches could be even "Predict the category of city". This dataset is also particularly convenient for clustering and maybe find different clusters of consumers within it. The dataset having the 12 attributes in that we take only 8 attributes.

- Gender - It is a Boolean value represent 0 as Female and 1 as Male.
- Age - It gives the age of a particular customer.
- Occupation - Id Occupation of each customer.
- Stay in years - How many years customer stay in that city.
- Marital Status - Gives the status of his marriage.
- Product Category - Giving the product category with numbers.
- Purchase - Purchase amount in dollars.
- City Category - Boolean values convey the city category as 1's and 0's

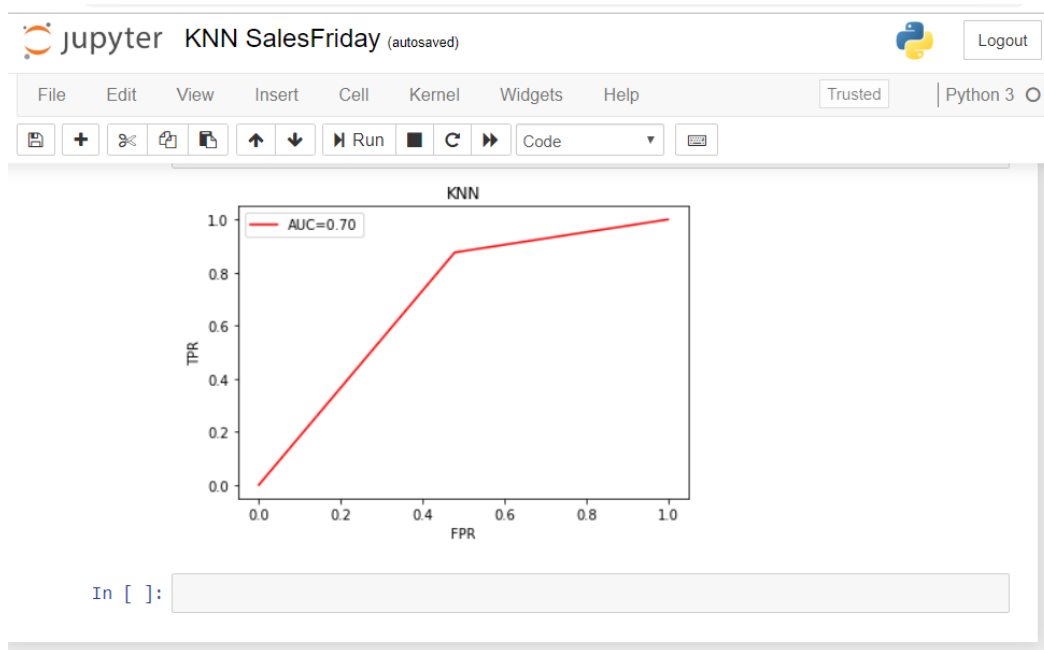
4.1.1 Figures and Tables:

Dataset for Black Friday:

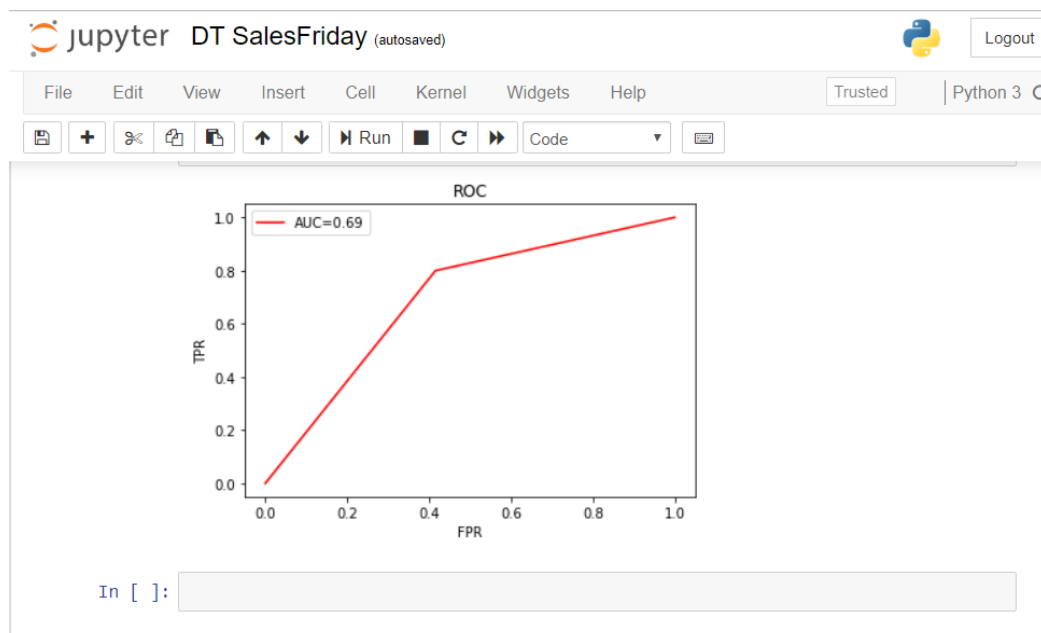
	Gender	Age	Occupation	Stay in Years	Marital Status	Prod_category	Purchase	Category
0	F	10	10	2	0	3	8370	1
1	F	10	10	2	0	1	15200	1
2	F	10	10	2	0	12	1422	1
3	F	10	10	2	0	12	1057	1
4	M	60	16	5	0	8	7969	0
5	M	30	15	3	0	1	15227	1
6	M	48	7	2	1	1	19215	1
7	M	48	7	2	1	1	15854	1
8	M	48	7	2	1	1	15686	1
9	M	30	20	1	1	8	7871	1
10	M	30	20	1	1	5	5254	1
11	M	30	20	1	1	8	3957	1
12	M	30	20	1	1	8	6073	1
13	M	30	20	1	1	1	15665	1
14	F	52	0	1	0	5	5270	1

ROC Curves:

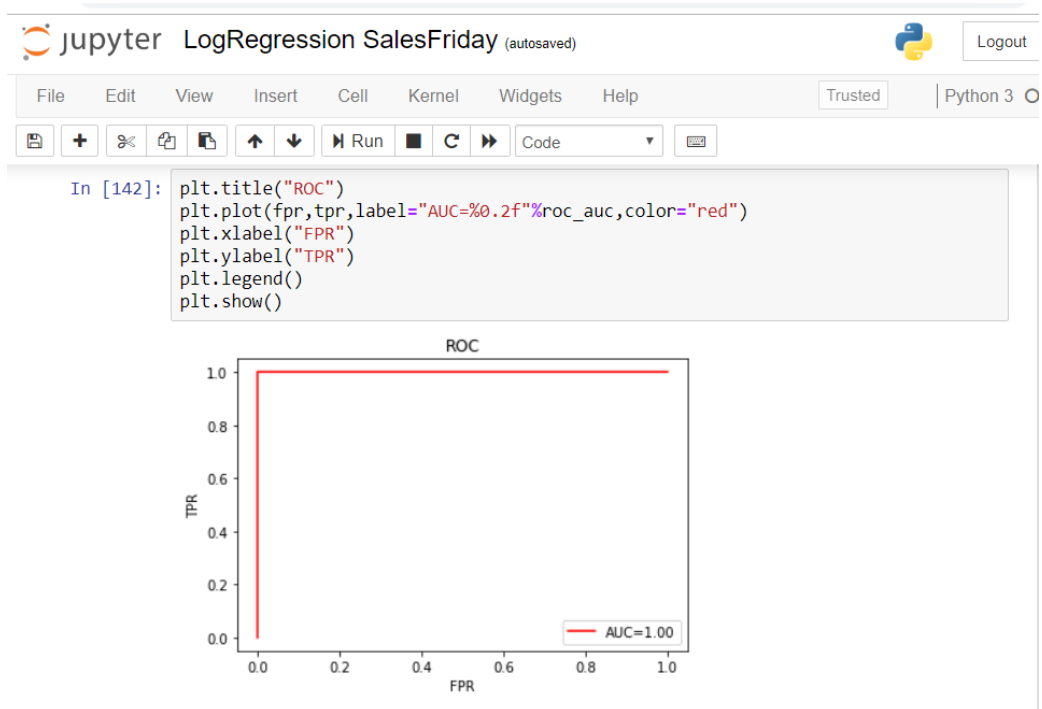
KNN:



Decision Tree:



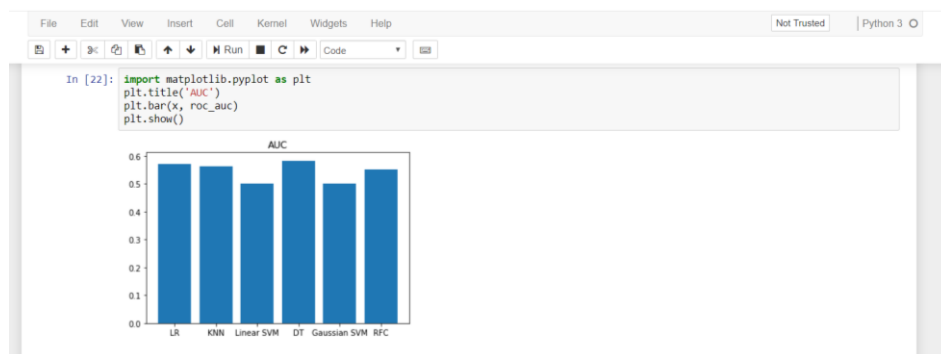
LogisticRegression:



Random Forest:

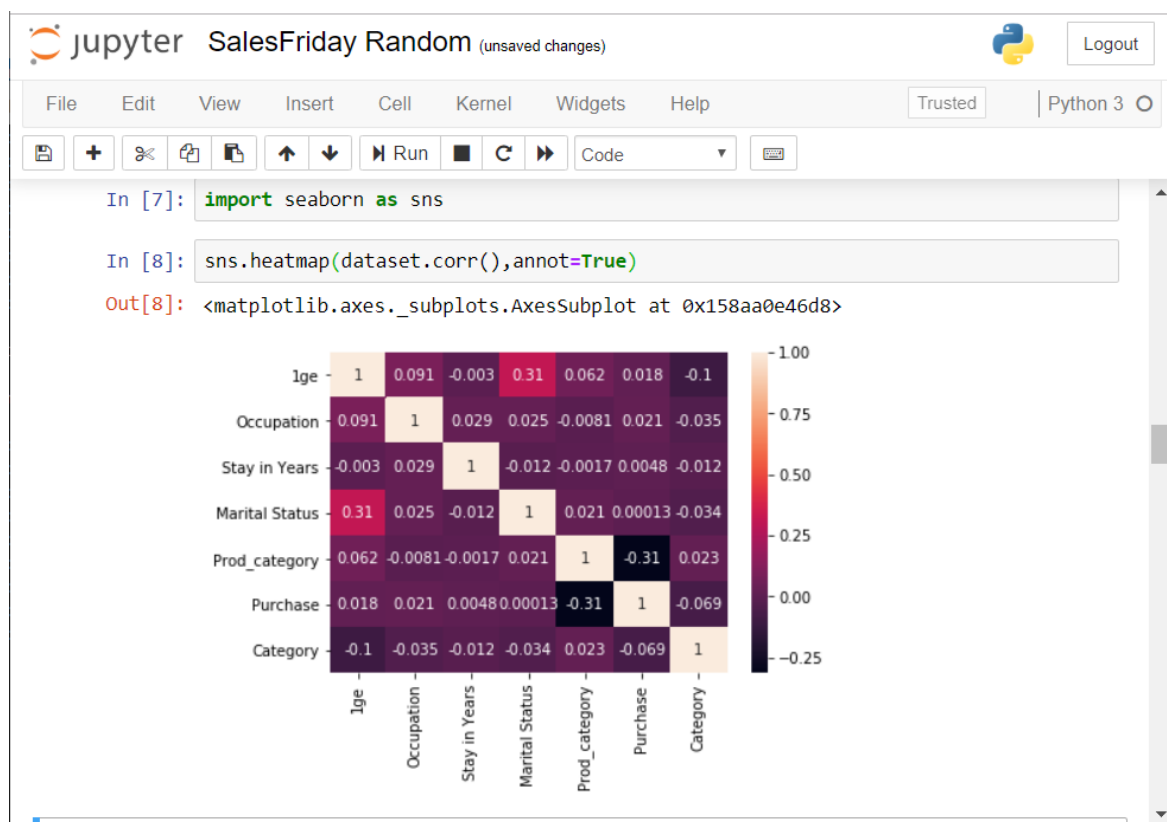


Bar Graphs:



4.2 Statistical Techniques & Data Visualization:

By importing matplotlib.pyplot library we have drawn graphs to demonstrate the AUC-ROC curves and by using bar graphs we have visualized the percentage levels of different techniques. And we have used the co-relation function to demonstrate the impact of every factor on each other.



4.3 Data Modelling using Supervised ML Techniques:

In general we have two types of learning algorithms, supervised and unsupervised learning algorithms. and in detail it consists of different techniques like,

Support Vector Machines

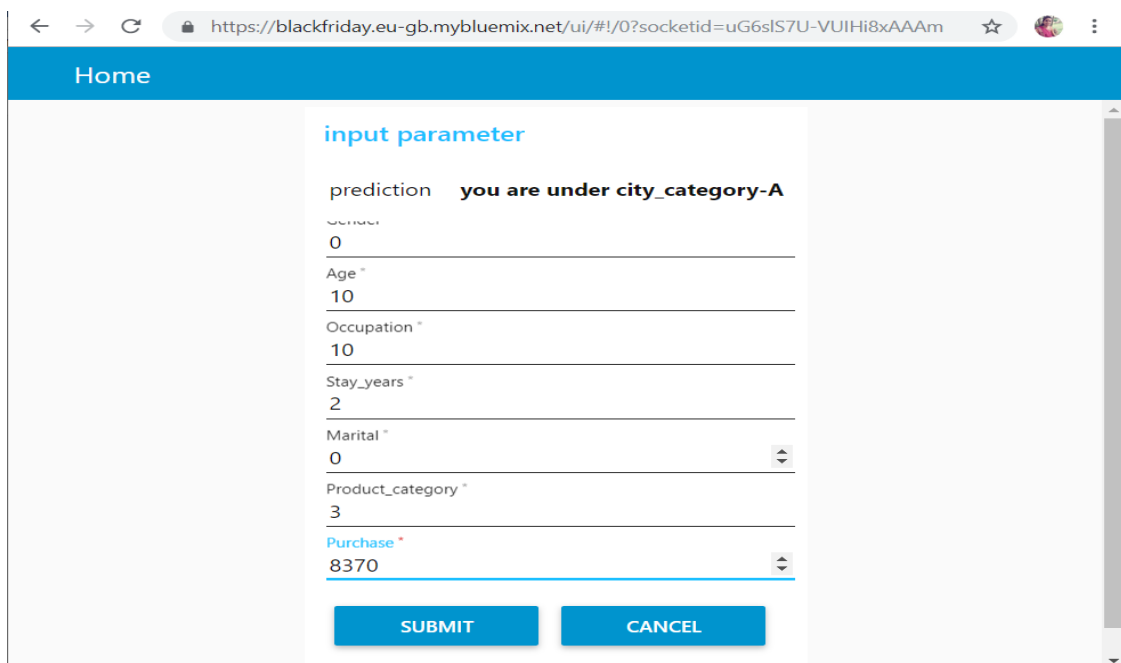
logistic regression

decision trees

k-nearest neighbour algorithm

since our model comes under supervised learning algorithm we applied every technique of the algorithm and based on the accuracy values we obtained we chosen random forest technique. and our model has dependent variable which is dichotomous(binary) means the output can be either a person is having diabetes or not having diabetes.

5. FINDINGS AND SUGGESTIONS:



The screenshot shows a web browser window with the URL <https://blackfriday.eu-gb.mybluemix.net/ui/#!/0?socketid=uG6slS7U-VUIHi8xAAAm>. The page has a blue header with the text "Home". Below the header, there is a section titled "input parameter" in blue. This section contains several input fields with labels and values: "prediction" with the value "you are under city_category-A", "Gender" with the value "0", "Age" with the value "10", "Occupation" with the value "10", "Stay_years" with the value "2", "Marital" with the value "0", "Product_category" with the value "3", and "Purchase" with the value "8370". At the bottom of the input section, there are two blue buttons: "SUBMIT" and "CANCEL".

Home

input parameter

prediction **you are under city_category-B**

City
1

Age *
22

Occupation *
3

Stay_years *
4

Marital *
1

Product_category *
1

Purchase *
19499

SUBMIT CANCEL

6. CONCLUSION:

Black Friday has integrated itself into our family culture, work culture, and consumer culture. Global capitalism is deeply rooted in consumerism and it is very prevalent in our society whether we know it or not. Some families, such as my own, accept Black Friday as a part of Thanksgiving traditions: looking through ads, talking about shopping at dinner, shopping with family members, etc. While other families try to adapt so Black Friday does not interfere with family plans. American family dynamics are vastly different from household to household. The importance of the consumer holiday known as Black Friday is highly dependent on the family and the people in it. But one cannot deny that over the past few decades, Black Friday has “creeped” its way into Thanksgiving gradually. In this Machine Learning Model we predict the city category in the black Friday sales.

