**Machine Learning Project Report**

**Name: Muhammad Bilal**

**ID: F2018266166**

**Section: V1**

My Project is E-Commerce Shipping Data. It has 11,000 instances. Basically, It will do Product Shipment Delivered on time or not? To Meet E-Commerce Customer Demand.

1. Purpose:

An international e-commerce company based wants to discover key insights from their customer database. They want to use some of the most advanced machine learning techniques to study their customers. The company sells electronic products. It is a Classification Problem

1. Content:

It has 12 Attributes:

1) ID: ID Number of Customers.

2) Warehouse block: The Company have big Warehouse which is divided in to block such as A,B,C,D,E.

3) Mode of shipment: The Company Ships the products in multiple way such as Ship, Flight and Road.

4) Customer care calls: The number of calls made from enquiry for enquiry of the shipment.

5) Customer rating: The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best).

6) Cost of the product: Cost of the Product in US Dollars.

7) Prior purchases: The Number of Prior Purchase.

8) Product importance: The company has categorized the product in the various parameter such as low, medium, high.

9) Gender: Male and Female.

10) Discount offered: Discount offered on that specific product.

11) Weight in gms: It is the weight in grams.

12) Reached on time: It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time.

Reached on time is a class variable.

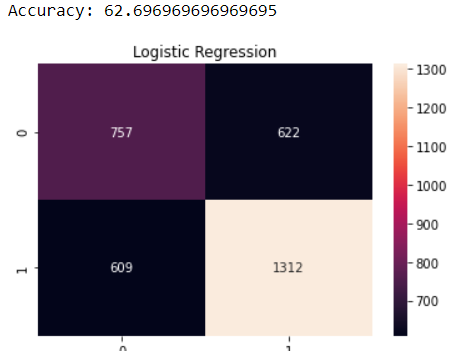
**Preprocessing:**

Here I’m converting string data into the numeric form by using One Hot Encodering. Attributes are **Product\_importance, Mode\_of\_Shipment, Warehouse\_block**

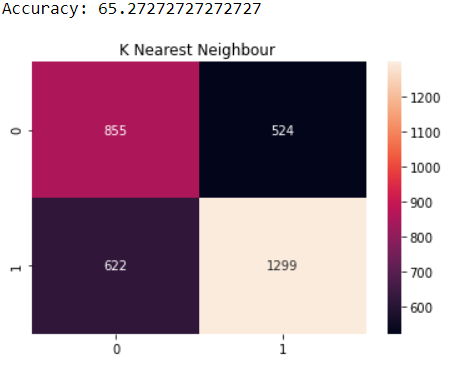
Secondly, I’m Scaling down the data by using standard scaler Afterward I’m implementing the models.

**Implementing Models and their Results:**

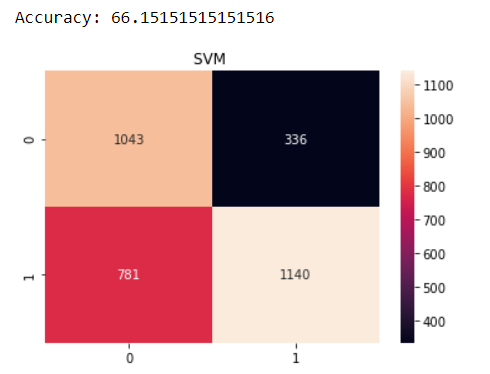
1. **Logistic Regression:**



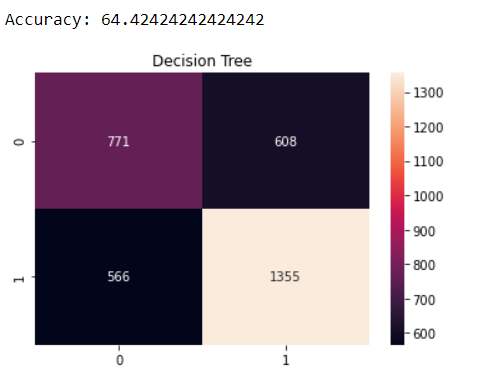
1. **KNN:**



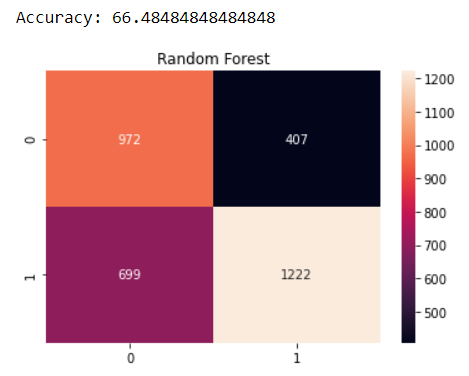
1. **SVM:**



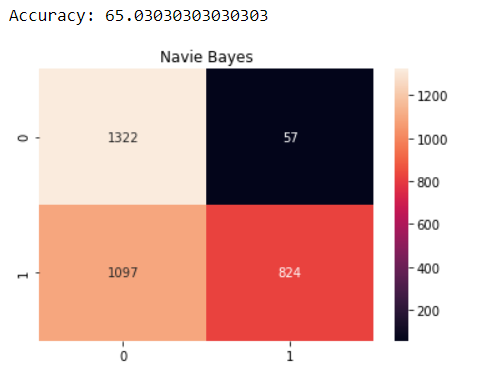
1. **Decision Tree:**



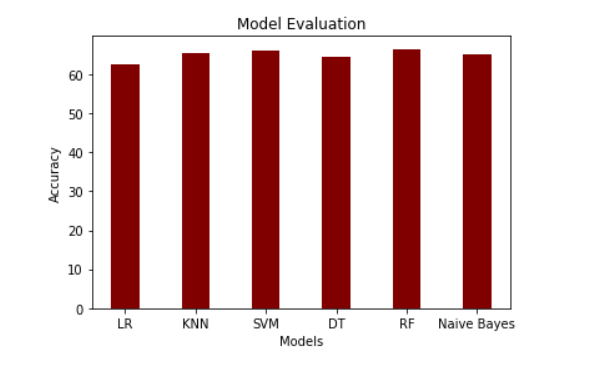
1. **Random Forest:**



1. **Naïve Bayes:**



**Result:**



**Comparative Analysis:**

We see that SVM and Random Forest have same and high score out of all models. so in this case we choose that model

which is simple and use low computational cost. so in our case we choose SVM model for the use because SVM is quiet simply

and it have low computational cost than Random forest.

why we didn't choose Random Forest?

because Random forest create 10 trees then we have same score with Random forest. The cost of create 10 trees is high than SVM

so SVM is Best in our case.