**Assignment: 1**

**Supervised learning**

The dataset was uploaded in the Jupiter notebook and was examined for null values first. After using isnull().sum() we fount that there are no null values in the dataset. There are 569 rows and 32 columns (features) in the dataset.

There are three algorithms applied in the dataset to understand the best model for fit.

**KNN**

KNN algorithm is used to predict the correct values out of all the instances that are available in the data, for instance. The KNN calculation expect that comparable things exist in nearness. As it were, comparable things are close to one another.

**Classification in KNN:**

precision recall f1-score support

B 0.95 0.99 0.97 105

M 0.98 0.91 0.94 66

accuracy 0.96 171

macro avg 0.96 0.95 0.96 171

weighted avg 0.96 0.96 0.96 171

**Decision tree:**

As it is a predictive model, Decision Tree Analysis is done via algorithmic approach where a data set is split into subsets as per conditions. It is mainly used to forecast the target variable.

**Classification in Decision tree:**

precision recall f1-score support

B 0.92 0.93 0.92 105

M 0.89 0.86 0.88 66

accuracy 0.91 171

macro avg 0.90 0.90 0.90 171

weighted avg 0.91 0.91 0.91 171

**Naïve Bayes**

Naive Bayes is a probabilistic machine learning algorithm based on the Bayes Theorem, used in a wide variety of classification tasks.

This is mainly used for Natural language processing algorithms. This theorem generally comes handy in case of conditional probabilities

**Classification in Naïve:**

precision recall f1-score support

B 0.93 0.94 0.93 105

M 0.91 0.88 0.89 66

accuracy 0.92 171

macro avg 0.92 0.91 0.91 171

weighted avg 0.92 0.92 0.92 171

**Interpretations:**

Depending on the f1 scores it can be observed that weighted average of KNN is better than decision tree and naïve algorithm.

The k value for the lowest error rate is 5 and a comparison showed the difference between the accuracy by one percent

In terms of accuracy also KNN shows a better result of 96% than the other two models.

It can also be observed that a slight change in the data can raises an extensive large changes in the conclusive outcome of the model, so utilization of strategies like sacking irregular timberlands and boosting in decision tree will assist with improving the exhibition.

Utilizing log probabilities and fragmenting information encourages in Naïve Bayes to improve the exhibition. Cross approval is the most usually utilized method to test the viable.