**MACHINE LEARNING ALGORITHMS PROJECT**

**COURSE: PROGRAMMING FOR ARTIFICIAL INTELLIGENCE**

**MEMBERS:**

**SP23-BAI-039**

**SP23-BAI-011**

**SP23-BAI-005**

**SUBMITTED TO: SIR AKSAM IFTIKHAR**

This report covers the implementation and results of three different algorithms based on their performance. The algorithms used in this project are:

* Random Forest Classifier
* Support Vector Machine (SVM)
* Logistic Regression

The goal of this project is to evaluate the status of people’s cardiovascular disease as present:1 or absent:0.

**The key features of this project are:**

Demographics: Age (in years) and gender (1: Female, 2: Male).

Anthropometric Data: Height (cm), weight (kg), and BMI (calculated).

Blood Pressure: Systolic (ap\_hi) and diastolic (ap\_lo).

Health Indicators: Cholesterol, glucose, smoking, alcohol use, physical activity.

Target Variable: Cardiovascular disease (1: Present, 0: Absent).

**Dataset Overview:**

Size: Thousands of individual records.

Type: Mixed numeric and categorical data.

**Data Preprocessing:**

The data preprocessing involves data Scaling used in SVM and Logistic Regression as the random forest does not basically require scaled data.

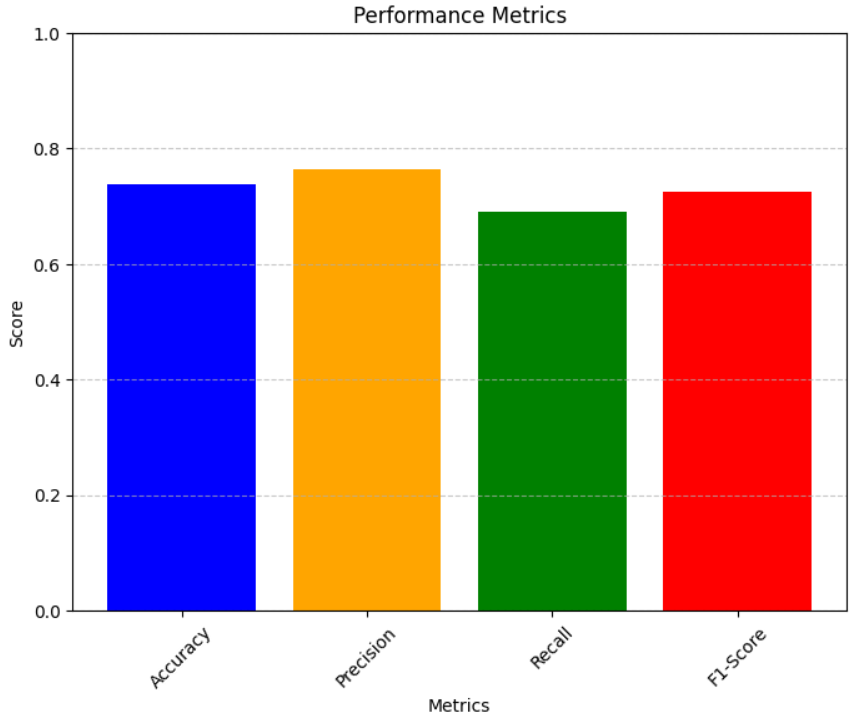
There was no missing data in the dataset.

Also, there were no features to be extracted. All the features are to be used for evaluation.

**Random Forest Algorithms:**

A Random Forest Classifier is a popular machine learning algorithm that belongs to the family of ensemble methods. It is used for classification tasks and is based on constructing a collection ("forest") of decision trees during training time. The algorithm combines the predictions of multiple decision trees to produce a more robust and accurate prediction.

The bar chart for different evaluation metrics is:



**LOGISTIC REGRESSION:**

Logistic Regression is a statistical method used in machine learning for binary classification problems, where the goal is to predict the probability of one of two possible outcomes. Despite its name, it is a classification algorithm, not a regression algorithm.

The bar chart for different evaluation metrics is:

A bar chart with different colored squares

Description automatically generated

**SVM:**

A Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for classification and regression tasks. Its primary goal is to find the optimal hyperplane that separates data points into different classes with the largest margin.

The bar chart for the evaluation metrices is:

A chart with different colors

Description automatically generated