# Machine Learning Report

Machine learning techniques are used in this heart disease prediction assignment to forecast the likelihood of heart disease. After loading the dataset, it is preprocessed by using the Interquartile Range (IQR) method to remove outliers and fill in missing values with the mean. Using {StandardScaler}, continuous features including age, serum cholesterol, resting blood pressure, and maximum heart rate achieved are scaled, while one-hot encoding is used to modify categorical features. To ensure an objective assessment of the model, the dataset is then divided into training and testing sets.   
  
The Random Forest Classifier, Support Vector Machine (SVM) with a linear kernel, and Logistic Regression are the three machine learning models that are trained and assessed. The test accuracy for the Random Forest Classifier was 85.25%, the SVM was 83.61%, and the best accuracy was obtained by the logistic regression model.  
  
Using the pickle library, the Random Forest model is saved for deployment even though it isn't the best performance.   
To offer an interactive interface for heart disease prediction, a Streamlit application has been developed. A number of features, including age, sex, type of chest pain, and other medical indicators, are provided by users. The code interprets the input, applies the same scaler from the training phase to it, and then uses the Random Forest model that was saved to create a forecast. The model is now accessible and easy to utilize for real-world applications thanks to this deployment, which permits real-time predictions