**Machine Learning Classification**

**Problem Statement**

To predict the cardiovascular disease from the given dataset

**Recognition Based On The Problem Statement**

* For the provided Data set, the Machine Learning Classification can be used to predict the cardiovascular disease
* Dataset contains 1000 rows and 14 columns

**Dataset Pre-processed**

* No null values exist in the dataset.
* The dataset is free of outliers.

**Feature Engineering**

**Correlation method**

Out of the thirteen independent columns, only two exhibit a correlation with the dependent column of greater than 0.6

**Recursive Feature elimination method**

The highest accuracy value is 0.97 for k= 6 its obtained in both logistic method and support vector

**Kbest method**

The highest accuracy value is 0.97 for k=5 its obtained in random forest method.so k=5 in random forest is selected as best method

**Selected independent columns**

'gender', 'chestpain', 'restingBP', 'restingrelectro', 'oldpeak'

**Machine Learning Models**

**Random Forest classification**

A screenshot of a computer

Description automatically generated