

List of Figures

| Contents | Pageno: |
|--|----------------|
| Fig 1: Methods for handling the missing data | 12 |
| Fig 2: Working principle of k-NN | 17 |
| Fig 3: Cluster when k=3 | 17 |
| Fig 4: Random forest with two trees | 18 |

ABSTRACT

Heart disease has been identified as the leading cause of death across the world. Not all people with coronary artery disease have chest pain as a symptom. There are other factors such as R-Blood Pressure, S-Cholesterol, F-Blood Sugar and R ECG. So for the proper treatment of the disease we need a reliable, accurate and feasible system for the prediction of the heart disease. The data is to be pre-processed to convert the raw data into a clean dataset, some of the techniques like removal of noisy data, replacing the missing data, filling default values and classification of attributes for prediction and decision making at various levels and then the dataset is given as the input to the algorithm. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. k-Nearest Neighbour (k-NN) and Random forest are used in the prediction of heart disease based on parameters effectively.

Key words: Heart diseases, Machine learning, Classification, Prediction, Medical Diagnosis.