**PROBLEM SOLUTION FIT**

Heart disease is also referred as cardiovascular disease. Many research scholars propose various data mining tool to identify its cause. Initially 13 different parameters such as age factor, gender, blood sugar, heart rate achieved are needed. The outcome of various data mining tools yields a different level of accuracy with respect to the number of instances and parameters used.

Data mining finds its application in various domains such as health care, biomedical research, computer research, etc. To extract hidden information, data mining is very useful in interpreting the cause of heart disease. Complex data can also be processed with various classifiers such as neural networks, genetic algorithm, decision tree, rule induction, data visualization and nearest neighbour.

FEATURE DESCRIPTION

PARAMETERS

There are various parameters used in training and testing which includes age, gender, resting blood pressure, maximum heart rate achieved, resting electrocardiographic results, and exercise induced angina. Attributes are defined under 4 types, namely real, binary, ordered, and nominal. Gender comes under binary type. Either it has to be male or female. Resting electrocardiographic results comes under nominal (i.e.) value should lie between 0 and 2. Real values include both integer and decimal values. The list of features is listed in Table 1. Initially for classification, consider the 13 features listed. Later attributes can

be removed for better accuracy.

Features such as serum cholesterol, chest pain type, thalassemia, colored fluoroscopy, resting blood sugar plays vital role in heart disease detection. Serum cholesterol, soft and waxy substances found in blood cells. It indicates the total amount of cholesterol present in the body. The measure of cholesterol present in the blood has the impact to cause cardiovascular disease. The unit is measured in terms of mg/dl. Thalassaemia (Thal) is a blood disorder identified by abnormal formation of haemoglobin. People with thalassemia have more chances to have cardiovascular diseases such as heart failure and abnormal rhythms. Based on thal value the level of disorder in heart can be identified. The value 3, 6 and 7 indicates normal, fixed and reversible defects respectively. Chest pain type (CP) includes typical angina, atypical angina, non- angina pain and asymptomatic. Angina is blockage in the flow of blood and oxygen to the heart. Resting electrocardiographies results should be nominal, i.e., the values should lie between 0 and 2. Measuring the electrical activity in your

heart is termed as electrocardiograph.

DECISION TREE

The decision tree is a data mining tool which plays a vital role in knowledge discovery. It extracts meaningful hidden information. The large data set can also be processed to obtain new target patterns. Decision tree finds its application in various domains such as machine learning, information extraction, bio-medical applications and scientific research for classification. The description of J48 decision trees can be found indecision tree does not require more domain knowledge. It is easy to interpret and fast. Decision trees can handle different data types such as nominal, ordinal, binary, real values. The research paper is

based on J48 algorithm. New dataset records can be formed by applying a decision tree like J48. The research paper takes a data set that contains 240 instances which has training and testing data of 120 each. Initially, consider all 13 parameters like age factor, gender, chest pain type, resting electrocardiographic results which obtain 75.83% accuracy of correctly classified parameters.

Decision tree allows adding or removing attributes to yield better results and accuracy. Remove two attributes such as resting electrocardiography’s results and maximum heart rate, which reduces overall performance. It is seen that accuracy rate has improved to 76.67% similarly attributes can be removed in order to achieve better results. On repeated removal of attribute study achieves 85%. Consolidate the attributes based on their accuracy achieved. Three attributes thalassemia, chest pain type and number of major vessels colored by fluoroscopy contribute maximum accuracy. Once the feature has been selected, the next step is to find the minimum number of objects required to attain better results. On repeated trial and error method, it is noted that 4 minimum objects are required to perform heart disease diagnosis. The next step is to identify the confidence factor. In the research work it is fixed ad 0.25 based on various trial