Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques

Published by: Prof. Pooja Kamat Assistant Professor (IT),

Symbiosis Institute of Technology, Pune, India

pooja.kamat@sitpune.edu.in in 2020

I. INTRODUCTION

Data mining is the way of discovering meaningful patterns and knowledge from a vast amount of data in the database, data warehouses, web or data stored in different information repositories. It is essential in many fields of studies to discover hidden information from massive datasets that help stakeholders to understand and retrieve their data within a short period. Different data mining techniques are used to classify, predict and cluster data to make correct or accurate decision-making in many organizations. In medical centers (hospitals, or other clinical centers) using data mining techniques helps to treat patients to identify if the individual has diseases or not and for early automatic diagnosis of patients from their diseases within the result retrieve in short time.

Methodology:

The main objective of the proposed methodology is to predict the occurrence of heart disease for early automatic diagnosis of the disease within retrieve result in short time [23]. This play vital roles for healthcare experts to treat their patients based on accurate decision-making and give qualities of services to the people. The proposed methodology is also critical in healthcare Organization with experts that have no more knowledge and skill. One of the main limitation of existing methodology is the ability to give accurate result as needed. This system use data mining techniques and machine learning algorithms J48, Naïve Bayes and Support Vector Machine, with k-fold cross-validation to predict the occurrence of heart disease. It uses different medical attributes that are more relevant such age, sex, blood pressure, cholesterol, blood sugar and heart rate are some of the attributes are included to identify if the person has heart disease or not. Analyses of data set are computing (implementing) using WEKA software. WEKA is open source software which includes a collection Machine learning algorithms for the data mining tasks. WEKA implementation is based on Java code. WEKA contains tools for data preprocessing, regression, clustering, classification, association rule and visualization that are important in data mining tasks.

CONCLUSION:

Heart disease is the most common disease that leads to death in our world according to World Health Organization(WHO) reports, especially in under developing countries. All medical experts do not have equal knowledge and skill to make an accurate decision, in which some experts give a poor reasonable decision that leads people to dangerous situations. To overcome this problems prediction of the occurrence of diseases is necessary. One of the benefit of survey papers is to improve the existing methodology for better decision making by using different algorithms and feature selection methods. Proposed methodology uses J48, Naïve Bayes and Support Vector Machine algorithms for predicting the occurrence of heart disease for early automatic diagnosis and short time retrieve result that helps to give the quality