Exploratory Analysis of Kidney Disease Data Set—A Comparative Study

Aniket Muley and Sagar Joshi

Abstract In this study, an applied and comparative study approach is proposed with a decision tree and random forest method. Here, secondary data associated with kidney-related disease was employed. Study explores that random forest is efficient as compared with decision trees. It is recommended that, in the future, random forest methods will be more fruitful in the classification related to the kidney patients.

Keyword Data mining · Chronic kidney disease · Decision tree · Random forest

1 Introduction

Nowadays, kidney disease patients are more in India. The functioning of it and causes associated with it are found in large scale in the nearby area. The main interest is to identify the major parameters that can cause chronic pains to patients. Therefore, developing a model to identify the disease can help the people to take precautions by identifying symptoms and overcome from the chronic disease earlier stage.

Various researcher works on the problems with different perspective. Some of the reviews are: Ahmed et al. [1] discovered that chronic kidney disease (CKD) is generally found in South Asian and in black skin people as compared to general population. They observed that it is due to diabetic people in south Asia and having maximum risk. Apart from that, other issues, viz. blood pressure, heart problems, family suffering from the same disease, it is more frequently observed during the age group of 60 and above. Data mining can be defined as a process of digging out until that time unidentified compelling and actionable information from big data and then using the information so derived to make vital tasks in industry and tactic judgment [2].

A. Muley (✉)
School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded, M.S. 431606, India

S. Joshi
Nutan Maharashtra Institute of Engineering and Technology, Talegaon Dabhade, M.S. 410507b, India

J. Nayak et al. (eds.), *Computational Intelligence in Data Mining*, Smart Innovation, Systems and Technologies 281, https://doi.org/10.1007/978-981-16-9447-9_10