Diabetes Mellitus Prediction in Big Data- Using Hadoop/Map Reduce Frame work

K S Praveenkumar[1], Joyal Paul[2], Dr. K J James[3],

[1]PhD Research Scholar, KAHE, Coimbatore (Asst. Professor, Dept. of CS, SNGIST, Kochi),
[2]Assistant Professor, Dept. of ME, SNGIST
[3]Professor, SNGIST Group of Institutions
praveen7387@gmail.com[1], joyalpaul@gmail.com[2], kjjames111@gmail.com[3]

Presenter: K S Praveenkumar, Contact Number: XXXXXXXXXX
Subject Area: Computer Science

Abstract

Diabetes Mellitus disease prediction is a developing field of research in the healthcare sector. Even though many data mining approaches have been implemented to assess the important causes of diabetes, but only limited sets of clinical risk parameters are considered. Diabetes mellitus (DM), commonly referred to as diabetes, is a group of metabolic disorders in which there are high blood sugar levels over a prolonged period. Diabetes Mellitus is a condition wherein a person is either incapable of producing insulin or the body is not able to utilize the insulin present in the body. While many consider diabetes as an ailment with genetic predisposition, today it has become one of the leading lifestyle ailments. There are three main types of diabetes mellitus: Type 1 DM results from the pancreas's failure to produce enough insulin. Type 2 DM begins with insulin resistance, a condition in which cells fail to respond to insulin properly. As the disease progresses a lack of insulin may also develop. The most common cause is excessive body weight and insufficient exercise. Gestational diabetes is the third main form, and occurs when pregnant women without a previous history of diabetes develop high blood sugar levels. Prevention and treatment involve maintaining a healthy diet, regular physical exercise, a normal body weight, and avoiding use of tobacco. Control of blood pressure and maintaining proper foot care are important for people with the disease.

So the results produced by such techniques may not represent correct diabetes. In this study, the proposed system can efficiently find out the rules to predict the risk level of a patient based on the given parameter about their health. Here we evaluate many factors Hereditary and genetics factors, Stress, Body Mass Index, Increased cholesterol level, High carbohydrate diet, Nutritional deficiency, Nature of Exercises, Tension and worries, High blood pressure, Insulin deficiency, Insulin resistance. Then we evaluate and compare this system using suitable rules and Map Reduce algorithm. The performance of the system is evaluated in terms of different parameter like rules used, classification accuracy, and classification error. By considering all these parameters, the system can predict diabetics in a great accuracy. Also this paper surveys about various techniques and tools available in Big Data to predict Diabetes mellitus. Big Data can significantly diabetes research and ultimately improves the quality of health care for diabetics patients.

*Keywords: **Diabetes Mellitus ; Big Data, Hadoop/Map Reduce; C4.5 algorithm***;