

Chapter 14

Risk Analysis of Diabetic Patient Using Map–Reduce and Machine Learning Algorithm

Nagaraj V. Dharwadkar

 <https://orcid.org/0000-0003-3017-0011>

Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Sakhrale, India

Shivananda R. Poojara

Department of Computer Science and Engineering, University of Tartu, Estonia

Anil K. Kannur

Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Sakhrale, India

ABSTRACT

Diabetes is one of the four non-communicable diseases causing maximum deaths all over the world. The numbers of diabetes patients are increasing day by day. Machine learning techniques can help in early diagnosis of diabetes to overcome the influence of it. In this chapter, the authors proposed the system that imputes missing values present in diabetes dataset and parallel process diabetes data for the pattern discovery using Hadoop-MapReduce-based C4.5 machine learning algorithm. The system uses these patterns to classify the patient into diabetes and non-diabetes class and to predict risk levels associated with the patient. The two datasets, namely Pima Indian Diabetes Dataset (PIDD) and Local Diabetes Dataset (LDD), are used for the experimentation. The experimental results show that C4.5 classifier gives accuracy of 73.91% and 79.33% when applied on (PIDD) (LDD) respectively. The proposed system will provide an effective solution for early diagnosis of diabetes patients and their associated risk level so that the patients can take precaution and treatment at early stages of the disease.

DOI: 10.4018/978-1-7998-3053-5.ch014

INTRODUCTION

In last one decade, the health care systems in all over the world have grown and in the same phase, the large volume of data has proportionally increased in the form of patients' records. These patients' records include the x-rays, images, clinical information reports, diagnosis, prescriptions, etc. Nowadays the health care systems are generating and maintaining the data in digital form. This data has particular type such as structured, unstructured, semi-structured or combination of these types. The data generated at health care systems will be helpful not only for the treatment of the patients but also to track what possible may happen to patients' health in future. Health care systems are concerned with all the diseases and with possible diagnosis and treatment of the patient. The Health care system faces many challenging health hazards in day-today, and one of such challenging health hazards is diabetes. Diabetes is one among the four major non-transmissible diseases that is under attention of healthcare sector. Diabetes is caused when a human body does not generate enough insulin. Diabetes has become the most dangerous health issue in this very world. The numbers of diabetes patients are increasing day by day since past few years. On the basis of WHO (World Health Organization) worldwide survey report on diabetes says that, in the year 1980, 108 million people were Diabetic and in 2014, the count of people increased up 422 million suffering from diabetes. From 1980 to 2014, the average percentage in increase of diabetic patients randomly changed and as on today more than 600 million suffering from diabetes. There are three types of diabetes that can be characterized into: type-1, type-2 and gestational diabetes. The type-1 diabetes is an insulin-dependent diabetes that occurs due to the inability of patient's pancreas to generate insulin. In this case, the patient needs to inject insulin externally for the survival. The type-2 diabetes is non-insulin dependent diabetes, caused due to unsuccessful utilization of insulin produced by human body cells. The gestational diabetes is third type which is occurs during pregnancy, this is a temporary condition from which pregnant woman may suffer. This condition can lead patient towards the development of type-2 diabetes. It may cause complications in pregnancy (World Health Organization, 2016). Diabetes of each category can make complications in the human body. It leaves long term bad impact on human body. Due to diabetes patient may suffer from different diseases such as heart stroke, attack, eye blindness, leg amputations, kidney failure etc. For pregnant women, diabetes may lead to the possibility of fatal death or other complications in pregnancy. As diabetes has long-term complications associated with it, the patient may suffer from economic loss through regular medical checkups cost, prescription cost (World Health Organization, 2016). Hence it is necessary to control this rising serious disease. There is no permanent and complete solution for diabetes as on date. Hence to make an easy living with diabetes better solution is an early diagnosis of diabetes or diseases that can be caused due to diabetes. To reduce the influence of diabetes contribution of everyone is required. The government, health care industries, technology specialist, the scientists must provide solutions to overcome this issue (World Health Organization, 2016). With the help of technologies such as machine learning, data mining it is possible to analyze diabetes data and accordingly provide early diagnosis, precautions and pre-treatment to the patient (Eswari, T, 2015). The digital data generated in health care systems will be helpful in analyzing the risks in patients. Not only the health care systems researching on the diabetics but also the engineering and biotechnology disciplines are working in the same sector of diabetes to provide the solutions in preventing the risks in patients. One of the engineering disciplines is computer science, there are many techniques available for analyzing the digital data of patients and predict what may possibly happen. The problem statement for the work is to analyze the risk of the diabetic patient, and we propose the methodology for the same using Map-reduce and decision tree-based algorithms. In

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/risk-analysis-of-diabetic-patient-using-map-reduce-and-machine-learning-algorithm/264314

Related Content

A High-Level Interactive Query Language for Big Data Analytics Based on a Functional Model

Symphorien Monsia and Sami Faiz (2020). *International Journal of Data Analytics* (pp. 22-37).

www.irma-international.org/article/a-high-level-interactive-query-language-for-big-data-analytics-based-on-a-functional-model/244167

A Predictive Analytics Framework for Blood Donor Classification

Kavita Pabreja and Akanksha Bhasin (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-14).

www.irma-international.org/article/a-predictive-analytics-framework-for-blood-donor-classification/277644

Linked Open Statistical Metadata

Franck Cotton and Daniel Gillman (2017). *Data Visualization and Statistical Literacy for Open and Big Data* (pp. 297-320).

www.irma-international.org/chapter/linked-open-statistical-metadata/179971

Predictive Modeling of Surgical Site Infections Using Sparse Laboratory Data

Prabhu RV Shankar, Anupama Kesari, Priya Shalini, N. Kamalashree, Charan Bharadwaj, Nitika Raj, Sowrabha Srinivas, Manu Shivakumar, Anand Raj Ullend and Nagabhushana N. Tagadur (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 13-26).

www.irma-international.org/article/predictive-modeling-of-surgical-site-infections-using-sparse-laboratory-data/209738

The Delphi Method Application in the Analysis of Postgraduate Students' Attitudes on the Environmental Performance Index

Natasa Petrovic, Jasna Petkovic, Dragana Makajic-Nikolic, Maja Levi Jaksic and Marko Cirovic (2017). *Emerging Trends in the Development and Application of Composite Indicators* (pp. 248-276).

www.irma-international.org/chapter/the-delphi-method-application-in-the-analysis-of-postgraduate-students-attitudes-on-the-environmental-performance-index/165655