ORIGINAL RESEARCH PAPE

AN ANALYTICAL SURVEY OF USAG DATA AND HADOOP FOR PREDICT DISEASES

| Ms Priti D. Sadaria | Department of Computer Science |
| Dr. Achyut C Patel* | Smt.M.T.Dhamsania Commerce (|

ABSTRACT

Now a day no field remain untouched with Information Techr Technology for different purpose. Health data growing quickly k Extraction of useful information by analyzing this rapidly grov applicable in real life is really a challenging task. Knowledge disc is a new trend that is Big Data Computing. Machine learning techr computing provides computing services over the internet whic analytics for big data processing. Now a day, analysis of diab unpredictable growth of data which leads to a big challenge in pro

Hypothesis: To propose an algorithm for prediction of diseases like Nephropathy, Retinopathy and Cardio Vascular Diseases and as a result treatment can be given at proper time.

Methodology: This research focuses on the significance of Big Data tools and machine learning algorithms for clustering and classification to develop a predictive algorithm for predicting diabetic related diseases by using Hadoop platform.

INTRODUCTION:

The use of public healthcare data and analysis plays an important role for the healthcare planners. Since last few years due to rapid adoption of Information Technology in healthcare systems, the health data increase exponentially and data is available in different forms. Healthcare providers can use healthcare data and analytics to learn more about patients and can enhance preventive care by utilizing key data. Healthcare data and information are major helpful sources for taking effective decisions. From such rapidly growing huge data, Knowledge discovery and decision-making is a challenge concerning both data organization and timely processing. Big Data computing combines large-scale computing with machine learning techniques is used to build predictive analytics for prediction of some chronic diseases.

To process large amount of Big Data, by human inspection is impossible and because of this reason, it is required to develop a high-speed system to manage such Big Data. A variety of Big Data tools are presently being used to scrutinize data at a faster rate and present users with essential knowledge for decision-making and prediction.

Storing and processing quickly growing data within the specific time using conventional tools is a difficult task. Capable tools are required for data management and analysis for the data which is collected from a lot of applications in scientific and business field [1]. Big Data provides tools to create, manipulate and manage large datasets.

In recent era a service-oriented computing model – Cloud computing is mostly used for processing large volumes of rapidly increasing data at a faster scale and it is actually a demand for Big Data computing. So to fulfill this requirement, Big Data framework Hadoop and Spark are used to complete Big Data tasks with machine learning techniques. This proposal focuses on predictive analytics by using different machine learning strategies to analyze Big Data and as a result of this one can make decisions about future

compl resear with H

The o ineffe utilizir and ca based semi s health

Deper be col a new diseas amour parall« enviro

Now a It is or The p diabe Diseas new a new a before

Objec
• To usi
• To
• Th to K-i
• To

Resea
The li standa Cloud algori archite and t structi

Revie
Wei Fa