International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 5 416 - 418

Medical Affiliated Hadoop Application (MAHA)

Prof. Sunil Yadav, Pooja Maller, Darshana Mohite, Sameera Shaikh
Under Graduate Students, Siddhant College Of Engineering, Pune-4044.
sunilyadav14@gmail.com, poojamaller25@gmail.com, darshana.mohite93@gmail.com, sameerasweetu98@gmail.com.

***Abstract*—** Medical information exchange is an inevitable stage of medical informatics development. Administration features are integrated and the structure of HIE platform receives a centralized architecture which serves a enormous volume of medical information and terminates with multiple types of query and information extraction for centralized design. Cloud computing is a technological prerequisite. The purpose of this paper is that, we presents a medical information platform based on Hadoop, which is named after Medoop. The current Medoop project make essential use of HDFS to accumulate the integrated CDA records more effectively, the data content arranged in CDA record as per the cessation business queries and calculates the statistic data distributively in Map Reduce model. The scope of Medoop is to bring about a complete platform for medical data for accumulating, exchanging and using the element in Hadoop ecosystem. The project presents Medical Hadoop which consists of cloud server and estimates whether the patient is affected by diabetes or not. For prediction purpose we use Naïve Bayes algorithm. With the help of Naïve Bayes static data is provided in training phase and it trains accordingly for detection. Naïve Bayes is stipulated with normalized dataset using k-means(or any other normalization algorithm).

Keywords- *Hadoop, Health Information Exchange (HIE), Mapper, Reducer, cloud Computing, HDFS, Naïve Bayes, k-means, Clinical Document Architecture (CDA), Android.*

*****

I. INTRODUCTION

The project compasses of medical Hadoop which comprises of cloud server and the android client which estimates whether the patient is affected by diabetes or not. All the data will be under restricted access in the cloud server. The patient's information will be accumulates in HDFS format.

The different operations like add/manage, searching, applying analysis on dataset and mining will be performed. In add/manage operations patients data is approved as an input. Key pair values are used to look out for a particular patients record and new/current records in searching operation. This system will estimate if patient is affected by the diabetes or not. The available doctors in the vicinity of the patients area will be discovered and suggested by this system with the help of Global Positioning System. According to the detected diabetes result this system will list the medicines. Only the authorized client is allowed to access the cloud server. The authorized client can be an experienced doctor or any other medical expert. With the help of Global Positioning System with the android client this system will allow the user to locate themselves on a map and even find and navigate to the hospital nearby.

The processing and new storage analyzing results are given in extinction of Hadoop, the centralized HIE. To yield the medical big data parallel computing in cluster is used by map reduce.

Juvenile: it is called as onset insulin-dependent diabetes. It accounts for 5-10 out of 100 people who is affected by diabetes. Diabetes can be categorized in to two types 1. Juvenile and 2. Resistive. In type 1 diabetes the cells that release insulin is destroyed by the body's immune system. Resistive: it is called as adult onset or non insulin dependent diabetes. It develops at any stage of age. Apparently, at adulthood it becomes most common. Here the output is displayed as 0 and 1. If the patient is affected by diabetes then output is given as 1, along with these it displays the type of diabetes and the relevant treatment. If patient is not affected by diabetes then output is given as 0.

II. PROBLEM STATEMENT

To develop a prediction system for applying data mining techniques on Big dataset using Hadoop, by accepting various parameters of patient as input and predicting whether the person is affected by diabetes or not, and provides a information of available doctors in vicinity of the person.

III. ARCHITECTURE

Figure 1: System Architecture

IV. DESCRIPTION OF SYSTEM ARCHITECTURE

This architecture comprises of cloud server and clients. All the databases will be accumulated in Hadoop file system near the cloud server.

There are distinct operations in system like ADD and Manage dataset , searching operation which is applied using key/pair value, analysis on dataset, mining (Diabetic). The system also suggest relevant treatment and medicine for diabetic patients. Only the authorized client has the right to access, perform the operations like searching a patient and viewing the prediction result. The prediction results are given to the particular patient according to the provided data set given by the Naïve Bayes .

A. Diabetes and its Type are:

a. .Juvenile
b. Resistive

1. Juvenile –
   - It occurs when the immune system destructs cells in prancreas called bets cells.
   - They are responsible for insulin

2. Resistive-
   - Also known as insulin resistance.
   - Type 2 diabetes.
   - If a patient body does not use insulin properly then he is affected by type 2 diabetes..

V. OUTCOME

Specifically, we have developed a system in which the static dataset is taken as an input in the training set under the surveillance of a medical expert. In the detection phase the system should give predictions by referring the dataset of the patient who is to be diagnosed.

Parameters that should be given to the system-

1) *Number of times pregnant*
2) *Plasma Glucose*
3) *Diastolic Blood Pressure*
4) *Triceps skin fold thickness.*
5) *Insuline*
6) *Body Mass Index*
7) Diabetic Pedigree Function
8) *Age (k-means)*

OUTPUT → 0, 1
Where 0→ patient is not affected by diabetes
1→ patient is affected by diabetes.

VI. SYSTEM FUNCTION

Morphism-
Server(DB)←Registration (User, Detaials);
Yes/No←Login(UserId, Password);
Result<set>←Search for patient diseases(Patient details);
Result<levels>←Search for patient diseases(disease attributes);
Result<set>←Apply prediction(current parameters<set>);
Treatment and Medicine←Treatment Suggestion(disease level);
Result From Mapand Reduce←Search for Patients(details);

VII. ADVANTAGES AND DISADVANTAGES

A. *Advantages*

There are multiple advantages for Hadoop which are used in medical information platform-

1) High vigorous scalability without stop. In HDFS, metadata resides on a Namenode server and data blocks occurs throughout an area which resides on Datanode servers.

2) High reliability by automatic data detection and data replication. The data is duplicated to several copies and is distributed to various Datanode servers.

3) It provides us with cheap data storage and also with high performance distributed computing.

4) It provides fast processing
5) High usability due to Android client.
6) Fault tolerance due to multinode server structure. .

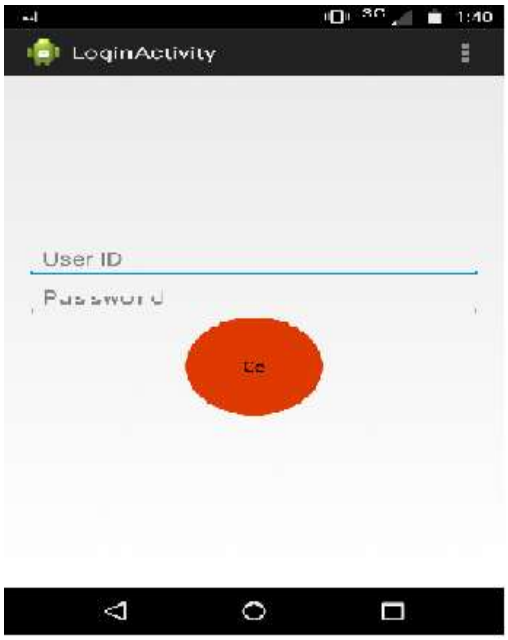
B. *Disadvantages*

1) Improper training dataset can cause wrong results
2) Server should be running.

VIII. APPLICATIONS

1) Detection of diabetes in patients.
2) Fast retieval of data.
3) Suggesting available doctors.
4) Suggesting relevant drugs and therapies

IX. RESULT



X. FUTURE SCOPE

In near future if any better algorithm than Naïve Bayes is available for the prediction purpose can be used to achieve high efficiency. This system can be implemented for detecting other diseases and providing patients with relevant therapies and drug information.

XI. CONCLUSION

In this paper we make a prediction system for diabetic patients, suggest available doctors, therapy, drugs. The issues related to the current systems are also discussed. We have review the application also respectively. We have also given the idea that how the system will function.

XII. REFERENCES

[1] Amazon, Amazon elastic compute cloud,http://aws.amazon.com/ec2[2]Microsoft, Windows Azure, http://www.windowsazure.com

[2] Dolin, R.H., et al., HL7 Clinical Document Architecture, Release 2.J Am Med Inform Assac, 2006, 13(1): p. 30-39.

[3] Ministry of Health, Regional Health Information Platform Guide(Trial), in Chinese, 2009.

[4] National Institute of Standards, The NIST Definition of cloud computing, 2011.

[5] Infoway, Emerging Technology Series: Cloud Computing in Health, 2012.

[6] Apache, Hadoop, http://hadoop.apache.org.

[7] Health Level 7.http://www.hL7.org.

[8] Rodriguez-Matrinez, M.,etal., MedBook: A Cloud-Based Healthcare Billing and Record Management System, in IEEE International Conference on Cloud Computing 2012.

[9] MEDOOP - Medical Information Platform Based On Hadoop.