MEDOOP: Medical Health Information System Based on Hadoop

Komal Mandal, Pooja Maller, Darshana Mohite, Sameera Shaikh
Under Graduate Students, Siddhant College Of Engineering, Pune-412109.
komalmandal90@gmail.com, poojamaller25@gmail.com,
darshana.mohite93@gmail.com, sameerasweetu98@gmail.com.

***Abstract*—** Medical information exchange is an imminent stage of medical informatics development. Integrating the administration feature, the structure of HIE platform accepts a centralized architecture which has to serve a big volume of medical information and abut multiple types of query and information extraction for centralized design. Cloud computing is a technological requirement. In this paper, we put forward a medical information platform based on Hadoop, which is named after Medoop. The current Medoop make effective use of HDFS to store the integrated CDA records more effectively, arrange the content data in CDA record as per the continuous business queries and calculate the statistic distributedly in Map Reduce model. The aim of Medoop is to implement a complete platform for medical data to be stored, exchanging and using the element in Hadoop environment. The project presents Medical Hadoop which consists of cloud server and predicts wether the patient is affected by diabetes or not. For predicting this we use Naïve Bayes algorithm. Naïve Bayes is provided with static dataset in its training phase and trained accordingly for detection. Naïve Bayes is provided 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).*

*****

I. INTRODUCTION

The project presents medical Hadoop which consists of cloud server and predicts whether the patient is affected by diabetes or not. All the data will be stored in the cloud server. The database will be stored in HDFS format.

The various operations like add/manage, searching, applying analysis on dataset and mining will be performed. In add/ manage operations patients data is accepted as an input. And in searching operations key pair values are used to find a particular patients record and new/current records. This system will predict if patient is affected by the diabetes or not. This system will detect and suggest the available doctors in the vicinity of the patients with the help of Global Positioning System. And this system will also list the medicines according to the detected diabetes result. Only the authorized client is allowed to login to the cloud server. The authorized client can be a 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 evolution of Hadoop gives us a new storage analyzing and processing result for the health big data in centralized HIE. HDFS abut large volume storage requirement and Map Reduce uses parallel computing in cluster to produce the medical big data.

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

The above architecture consists of a cloud server and clients. All the databases will be stored in a HDFS format in Hadoop file system near the cloud server.

There are various operations in architecture like ADD/Manage patient dataset on cloud, searching operation which is applied using key/pair value, applying analysis on dataset, applying mining (Diabetic). The system architecture also suggest related treatment and medicine for diabetes. The authenticate client has only the right to login, perform the operations like searching a patient and viewing the prediction result. The prediction results will be given by the Naïve Bayes according to the provided dataset of a particular patient who is to be diagnosed.

 ISSN: 2321-8169
6419 - 6421

V. FLOWCHART

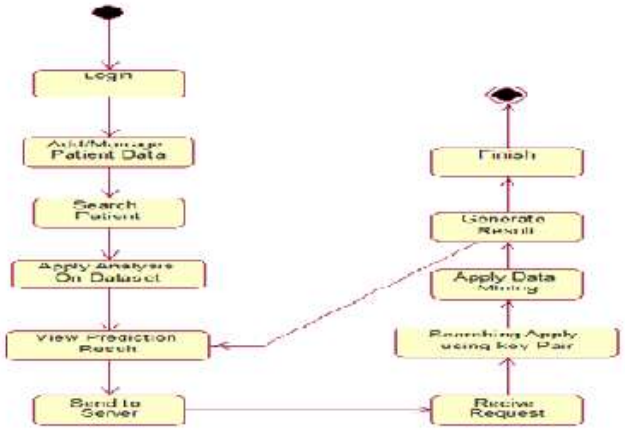

Figure 2: Flowchart

A. *Steps of Flowchart*

- Firstly the authenticate client logins to the cloud server and performs various operations.
- The first task he performs is Add/Manage in which we can add the new patient or we can manage the patients records from past history.
- The second task he performs is search a particular patients history dataset whether he is affected with some other disease or allergy.
- The third task he performs is apply analysis on dataset by using k-means and naïve Bayes algorithm.
- The server receives the request and performs the key pair operation on it.
- Then he applies the data mining operations and generate the result
- After all the operations are performed he finishes the task.

VI. 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*
   Low<5
   Medium<8
   High>8

2) *Plasma Glucose*
   Low<94
   Medium<148
   High<148

3) *Diastolic Blood Pressure*
   Low<60
   Medium<85
   High>85

4) *Triceps skin fold thickness.*
   (k-means)

5) *Insulin*
   Low<89
   Medium<194
   High>194

6) *Body Mass Index*
   Body mass index(kg/$m^{2)}$
   Underweight=<18.5
   Normal weight=<18.5-24.9
   Overweight=25-29.9
   Obesity=BMI of 30 or greater

7) *Diabetic Pedigree Function*
   Low<0.5
   Medium<1.2
   High>1.2

8) *Age*
   (k-means)

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

VII. 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);

VIII. 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.

IX. APPLICATIONS

1) Detection of diabetes in patients.

2) Fast retieval of data.

3) Suggesting available doctors.

4) Suggesting relevant drugs and therapies

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