

Hospital Selection Support System from Heterogeneous Data Source

Madhumathi B¹, Madhumitha V², Mahin Faliha S³, Maithili Parkavi⁴

^{1, 2, 3, 4}Students, RMK College of Engineering and Technology

Under the guidance of **G Indra**

Associate Professor, RMK College of Engineering and Technology

Abstract: *Medical institutions keep accumulating medical data, which is highly complex. Government agencies have been working hard to utilize such complex and diverse types of medical data to diagnose patient's diseases correctly and offer them the right treatment. Medical data comes from different sources, and most of it is unstructured. Big data computing is a new trend for future computing with a large-scale data set and can be divided into two paradigms: Batch-oriented computing and Real-time oriented computing (or stream computing). Batch computing is in general efficient in processing high volume data. The data are collected, stored, and processed in batches to produce the results. Big data analytics is often complex process of examining large and varied data sets to uncover information including hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions. Apache Hadoop is an example of batch-oriented computing. In this paper we are introducing a system which analyses the huge amount of clinical data using Hadoop and Hive to find the number of hospitals in an area, the best hospital in a particular area, the best hospital for a particular disease/problem etc. The analysis is done based on the users i.e. the patient's point of view. The resultant data is presented to the users through a web application which helps the users in deciding a hospital based on their disease and geographical location and fixing appointment with the particular hospital/doctor through the web application online.*

Keywords: Big data analytics, Hadoop, Hive, Web application

1. Introduction

Patients choose their doctors/hospitals based on three factors, a) other patient's Feedback b) Reputation and brand c) Location and accessibility. People who are new to an area may have heard a little about hospitals in that area. To aid them in choosing an appropriate hospital for them, we have introduced a system which makes use of the analytic techniques of big data to find the best hospitals that showed maximum success rate of curing a disease/problem and also present at their location. Also In today's world if someone wants to book a Doctor's Appointment, we need to call clinic or personally go to that place and book the appointment. This consumes precious time of the patient. Also, most of the time people get confused on which doctor to consult. The objective of this project is to build a system that will ease the process of choosing and booking appointment of the doctor. The establishment and improvement of doctor-patient interaction system is a very important requirement, especially now when the communication technology is developing rapidly. The advantages of the web can be made full use of to make up the time and distance gap between doctors and patients and to provide fast and adequate medical service. The main applications of this project can be in Towns, Village, and Urban area. It works as medical commerce project where doctors and hospitals can earn through this portal. It is where patients have some internet knowledge; they can get good treatment from skilled doctors. This portal may also be used in cities where maximum people are migrants. They can get good treatment from doctors even if they are new to a area. They can get health and fitness information. Hospitals, this portal is beneficial for hospitals where doctors can treat patients in different areas. This paper

contains Literature Survey in section III, study about bigdata in health care in section IV, The proposed system, and Tools and Techniques used in section V and VI followed by conclusion and Reference.

2. Literature Survey

- 1) This paper[1] summarizes the existing growth of Big Data Analytics in medical institution. It also examines some of the emerging role of Predictive Data Analytics (PDA), a few uses of Big Data Analytics in the medical field, the proposed generic architecture, in addition to some security solutions.
- 2) In this paper Reddy, Suresh et al., [2] had given an outline of loading and recovery procedures, Big Data techniques used in medical clouds, importance and need of Big Data Analytics in medical field and its merits, viewpoints in promising domain of predictive analytics, difficulties faced and the cure methods in medical domain. They carried out the trials on clinical data by Open source web interface with Hortonworks Data Platform. In these trials, they examined the hospital's over-all information such as common obstacles, sicknesses and clinical knowledge. The authors tried to research on sicknesses, which affects the patients and the type of hospital where the patients should join. They have also investigated on the types of difficulties encountered by the hospital(s).
- 3) Authors Van, Luis et al., [3] have proposed a novel Architecture which advises in using the Hadoop, Apache Storm, Kafka and NoSQL Cassandra. The groupings of the higher throughput publish-subscribe messaging, dispersed concurrent computing and information storing system will successfully evaluate the huge amount of

medical data arriving at a higher speed. The authors also provided the key technologies for the Stream Computing. Apache Kafka is a publish subscribe messaging scheme that is intended to be reckless, accessible and tough. It is designed for its accessibility, dependable, extensible, irreplaceable, well-organized and easy to arrange with and preserve. It is currently with twitter and in some real-time applications. "Not Only SQL" abbreviated as NoSQL denoting a varied and gradually acquainted cluster of the non-relational information administration schemes where the databases were not built mainly in the table format and usually does not use SQL for data handling. This will be convenient while working with an enormous amount of data where the nature of the data does not require relational model.

- 4) The establishment and improvement of doctor-patient interaction system is a very important requirement, especially now when the mobile communication technology is developing rapidly. The objective of this project is to build a system that will ease the process of booking appointment of the doctor. The patient will book the appointment through his/her mobile phone. The doctor will come to know the number of patients he has to attend whole day.
- 5) The author proposes a Medical system as a web portal and as an android application installing .apk file. The system is beneficial for both doctor and a patient and doctors. Admin perform their business and get 10% of doctors' fees or amount. Medical stores also get online business by providing home delivery of medicines to the patient. Thus paper work and lengthy process can be avoided. doctor and a patient and doctors. Admin perform

3. Big data in healthcare

Big Data refers to enormous amounts of data that cannot be processed by traditional techniques. The processing of Big Data begins with the raw data that is not grouped and is most often difficult to store in the memory of a single computer. Big data can be used to analyse the perceptions that can lead to better decision and strategic business moves. Big Data has been characterized by its three primary properties Volume, Velocity and Variety. Another important property in Big Data includes Veracity. Big Data analysis can be used for actual decision making in healthcare domain by altering the existing machine learning algorithms. Big data can be examined with the software tools which are usually used as a part of predictive analytics in medicine, data discovery, text mining and statistical analysis. Business Intelligence software and data visualisation tool can be a part of analysis process. In current years, the introduction of data analytics to large volumes of healthcare data collected on daily basis have unlocked abundant new chances and challenges in the field of medical informatics. New acceptance of Electronic Health Records (EHR) unlocks extra opportunities for data analytics, as we are able to contact structured and unstructured data which is analytically gathered for each event in the healthcare system. Medical information as well as the experimental judgement plays a major role. This medical information can be examined with the Big Data Analytics to envisage patient's illnesses and to advise the

appropriate desired medications. When associating big data analysis of other business areas with the healthcare, the health region is still in its initial phases due to abundant reasons.

4. Proposed System

The data was collected from websites like Kaggle. A total of 50,000 datasets were available. The data was analysed using Hadoop and Hive and the data was partitioned based on geographical location and disease. The hospitals which showed maximum number of cures for a particular disease was considered to be the best Hospital for that disease. The data sets which were resulted from this analysis was then exported to MySQL and used as input for the portal. Here in this process first admin has to collect all the hospital details and doctor details. After collecting, admin has to upload all the details in database. Now user has to register, create an account and they can search for which particular disease which hospital is best to get the treatment. For a particular city if they search admin will provide that details then user can get the result. After getting the result they have to send the request to that hospital. If they accept that request, the treatment will be started.

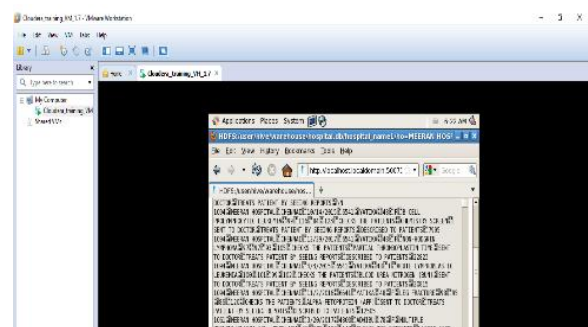
Assumptions of the project:

The goal of this project and application is to allow

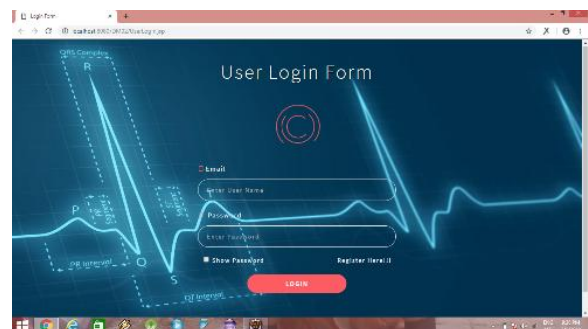
- Admin to add details about the hospital and its cure rate.
- Patients to search for their hospitals and request for an appointment.
- Doctors to accept or reject the appointment and notify patients about that.

Functional description of the System:

The system is divided into four parts



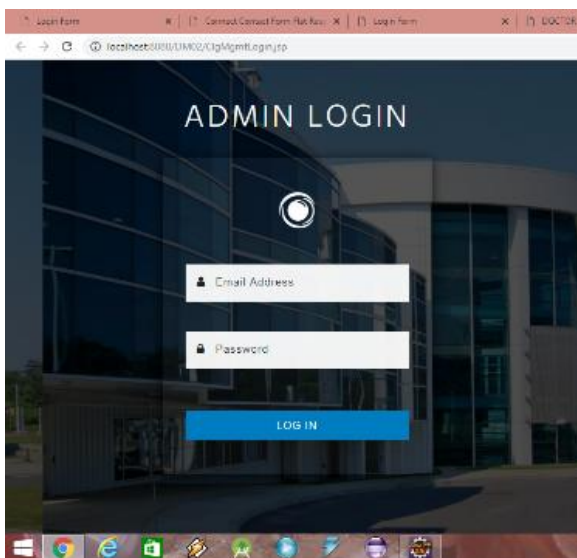
Bigdata analytics



Naïve user



Doctor / Hospital



System administrator

5. Tools and Technology

1) Hadoop Distributed File System(HDFS)

HDFS is a Hadoop based cluster for storage of huge data by dividing small parts and store them in distributed nodes. MapReduce The distribution of tasks onto file system can be done with MapReduce. Whenever information is gathering from the data store, it uses Map and Reduce techniques.

2) Hive

Hive is a query language it runs on Hadoop architecture. It is similar to SQL; the statements made in Hive are same as SQL statements.

3) HBase

Traditional databases are row-oriented database management systems but HBase is a column-oriented. It works on top of HDFS and it is not like SQL approach. It works on non-SQL based approach.

4) Sqoop

Sqoop supports incremental loads of a single table or a free form SQL query as well as saved jobs which can be a run multiple times to import updates made to a database since the last import. Imports can also be used to populate tables in Hive or HBase. Exports can be used to put data from Hadoop into a relational database. Sqoop got the name from "SQL-to-Hadoop."

5) MySQL:

MySQL is an open source relational database management system(RDBMS).MySQL is free and open-source software under the terms of the GNU General Public License.

6) VM Workstation

VMware workstation is a hosted hypervisor that runs on x64 versions of Windows and Linux operating systems(an x86 version of earlier releases was available);it enables users to setup virtual machines(VMs) on a single physical machine, and use them simultaneously along with the actual machine. Each virtual machine can execute its own operating system, including versions of Microsoft Windows, Linux, VMware Workstation Pro can save the state of a virtual machine (a "snapshot") at any instant. These snapshots can later be restored, effectively returning the virtual machine to the saved state, as it was and free from any post-snapshot damage to the VM.

7) Eclipse IDE

Eclipse is an integrated development environment (IDE) used in computer programming, and is the most widely used Java IDE. It contains a base workspace and an extensible plug-in system for customizing the environment. Eclipse is written mostly in Java and its primary use is for developing Java applications.

8) Apache Tomcat

Apache Tomcat, often referred to as Tomcat Server, is an open-source Java Servlet Container developed by the Apache Software Foundation (ASF). Tomcat implements several Java EE specifications including Java Servlet, Java Server Pages (JSP), Java EL, and WebSocket, and provides a "pure Java" HTTP web server environment in which Java code can run.

6. Conclusion

With the promotion of big data analysis in every industry, the medical industry is also widely using big data analysis as a tool, medical industry accumulated a large amount of data because of the wide ranging, It has an important role in creating new business value for the medical industry and enhancing healthcare industry by performing the analysis of big data. This paper focuses on the analysis of big data applications in the medical industry, and discussed the potential of its commercial value for the healthcare industry. Also this paper proposes a system which helps people in selecting and fixing appointment with a best and appropriate hospital for them.

References

- [1] Dhivyalakshmi .S, Umamakeswari. A "The role of Big data analytics in hospital management system."
- [2] A. R. Reddy and P. S. Kumar, "Predictive big data analytics in healthcare," Proc. - 2016 2nd Int. Conf. Compute. Intel. Common. Technol. CICT, pp. 623–626, 2016.
- [3] V. D. Ta, C. M. Liu, and G. W. Navinder, "Big data stream computing in healthcare real-time analytics," Proc. IEEE Int. Conf. Cloud Compute. BigData Anal. ICCCBDA, pp. 37–42, 2016.

- [4] Aishwarya Nandagawali, Komal Sute Prof. Dinesh Gawande."Online Medical System" India International Research Journal of Engineering and Technology (IRJET) Mar -2017.
- [5] Prof. S. B. Choudhari, Chaitanya Kusurkar, Rucha Sonje, Parag Mahajan, Joanna Vaz "Android Application for Doctor's Appointment."International Journal of Innovative Research in Computer and Communication Engineering
- [6] D. Chrimes, B. Moa, H. Zamani, and M.-H. Kuo, "Interactive Healthcare Big Data Analytics Platform under Simulated Performance, "IEEE 14th Intl Conf Dependable, Auton. Secur. Comput. 14th Intl Conf Pervasive Intell. Comput. 2nd Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congr., 2016.
- [7] S. M. Krishnan, "Application of analytics to bigdata in healthcare," Proc. - 32nd South. Biomed. Eng. Conf. SBEC, 2016.
- [8] D. Ramesh, P. Suraj, and L. Saini, "Big Data Analytics in Healthcare: A Survey Approach, " 2016.
- [9] Bernard Marr, how Big data is changing healthcare.
- [10] D.Peter Augustine, Leveraging Big Data Analytics and Hadoopin Developing Indian's Healthcare Services, International Journal of Computer Applications.
- [11] M. N. Vora, "Hadoop- HBase for large-scale data, " Proc. International ConferenceOn Computer Science Network Technology ICCSNT, vol. I, Dec. 2011.