IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Intelligent Heart Attack Prediction System Using Big Data

Prof. U.H. Wanaskar[1], Prajakta Ghadge[1], Vrushali Girmev[1], Prajakta Deshmukh[1], Kajal Kokane[1]

Department of Computer Engineering, Savitribai Phule Pune University[1]

Abstract: Main objective is to develop a prototype Intelligent Heart Attack Prediction System with help of Big data and data mining modeling techniques, this system can extract and discover hidden knowledge (relationships and pattern) associated with heart disease from historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make efficient clinical decisions which traditional decision support systems can't do. The healthcare industry collects large amount of big data which are not mined. Using advanced data mining techniques remedies can be provided. Medical diagnosis is regarded as an important but complicated task that needs to be executed accurately and efficiently. By providing effective treatments, it also helps to reduce treatment costs. The automation of this system would be extremely helpful. Therefore, a heart attack prediction system would probably be more beneficial for medical diagnosis system

Keywords: Big data, Data mining, Hadoop, Healthcare, Knowledge-discovery, Risk prediction.

I. INTRODUCTION

The heart is most vital part of human body. Life is dependent on efficient working of heart. If operation of heart are not working properly, it will affect the other parts of human body such as kidney, brain etc. Now days a major cause of uncertain death in world is due to heart attack. Earlier only old aged and middle aged people were prone to having heart attack, because of today's unhealthy lifestyle it is affecting younger aged people also. Prediction of heart attack is a complicated task for medical practitioners. So heart attack prediction system would prove to be beneficial by using data mining techniques.

Data mining has already established as a novel field for exploring hidden patterns in the vast datasets. Data Mining is a non-trivial extraction of previously unknown, implicit and potential useful information about data. In short, it is a process to analyze the data from different views and gathering the knowledge from it.

The discovered knowledge can be used for different applications for example in healthcare industry (Heart attack prediction). So machine learning, big data and data mining techniques are used to develop software which will assist doctors and other people in making decision of heart attack in early stages.

II. APPROACH

Hadoop:

Hadoop handles large amount of structured as well as unstructured data more efficiently compared to the traditional enterprise data warehouse. As Hadoop is open source and can run on commodity hardware, the initial cost savings are dramatic and further continue to expand as the organizational data. In addition to it, Hadoop has a robust Apache community behind it which continues to contribute in its advancement.

Hadoop is open-source software which is developed and maintained by a network of developers from around the world. The Hadoop framework breaks big data into small parts that are stored on clusters of commodity hardware. Hadoop concurrently processes vast amount of data using multiple low cost computers for quick results.

Hbase:

HBase is a column based database management system that runs above HDFS. It suites well for sparse data sets, that are common in various big data use cases. HBase applications use java programming language much like a typical MapReduce application.
An HBase system consists a of sets of tables. Every table contains columns and rows, like a traditional database. An HBase column represents an attribute of an object. HBase allows several attributes to be grouped together into column families, such that the elements of a particular column family are stored together. In HBase you have to predefine the table schema and specify the column families. It is very flexible as new columns can be added to families at any time, making the schema flexible and hence able to adapt to the changing application requirements.

System architecture is given below:

- **Dataset:** System has two datasets. One is the original big data set and the other is used as updated data set.

- **HDFS:** HDFS is a Java oriented file system which provides reliable and scalable data storage.

Name node: Name Node is the core part of an HDFS. It stores the directory tree of all files in file system. It also tracks where across the cluster the file data is kept. It does not store the data of these files itself.

DOI 10.17148/IJARCCE.2016.55178

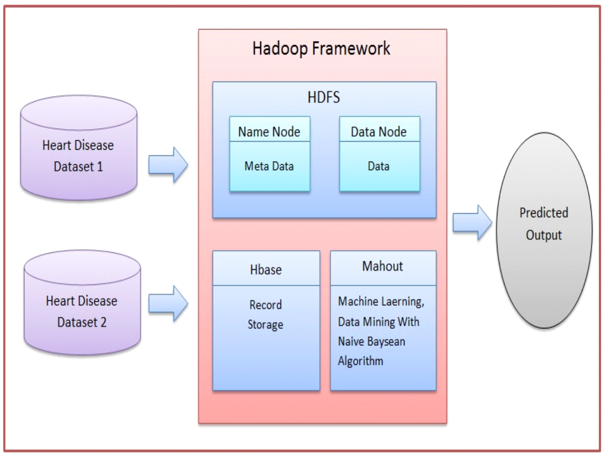


Fig. Architecture

Data node: A Data Node stores data in the Hadoop File System. A functional filesystem consists of one or more Data Node, with replicated data across them.

- **Data mining:** It is the analysis step of "Knowledge Discovery in Databases" process, or KDD. Data mining is an interdisciplinary subfield of computer science which is computational process of discovering patterns in large data sets i.e. big data.It includes the methods at intersection of machine learning, artificial intelligence, statistics and database systems. The primary objective of data mining process is to retrive information from a data set and transform it into an understandable format for future use.

- **Naive Bayes:** This is a type of classifier which is based on Bayes theorem. Such classifier algorithm uses conditional independence, this means it assumes that an attribute value on a given class is independent of the values of other attributes.

The Bayes theorem is described below:
Let X={x1, x2, ...,xn} be a set of n attributes. In Bayesian, X is considered as evidence and H be some hypothesis means, the data of X belongs to specific class C.
We have to determine P (H|X), the probability that the hypothesis H holds given evidence i.e. data sample X.
According to Bayes theorem the P (H|X) is expressed as P (H|X) = P (X| H) P (H) / P (X)

III. DATASET

The data was collected from the four following locations:

1.V.A. Medical Center, Long Beach, CA (long-beach-va.data)
2. Cleveland Clinic Foundation (cleveland.data)
3. University Hospital, Zurich, Switzerland (switzerland.data)
4.Hungarian Institute of Cardiology, Budapest (hungarian.data)

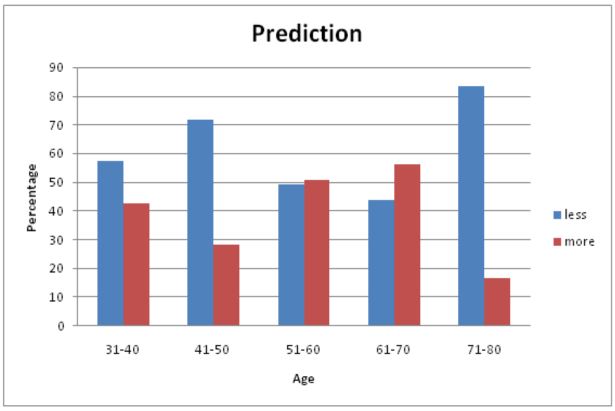
Databases have 76 raw attributes, only 14 of them are actually used. All attributes are numeric valued.

Attribute Information is as follows:
1) #3 : age
2) #4 : sex
3) #9 : cp
4) #10 : trestbps
5) #12 : chol
6) #16 : fbs
7) #19 : restecg
8) #32 : thalach
9) #38 : exang
10) #40 : oldpeak
11) #41 : slope
12) #44 : ca
13) #51 : thal
14) #58 : num i.e.the predicted attribute

IV. EXPERIMENTAL SETUP

Master node	
CPU	Intel Xeon E7420 2.13 GHz
RAM	16 GB
Hard disk	1T = 320GB * 4
Slave node 1	
CPU	Intel Xeon E7420 2.13 GHz
RAM	8 GB
Hard disk	500 GB
Slave node 2	
CPU	Intel Xeon E7420 2.13 GHz
RAM	8 GB
Hard disk	500

V. RESULT



Above graph examines the result of this project. This graph tells us about probability of having heart attack with age variation. The probability is shown in less or more percentage.

VI. CONCLUSION

In this work, we study the big data solution for predicting chances of having heart attack. Proposed solution gives big data infrastructure for both information extraction and predictive modelling. We study the effectiveness of our proposed solution with a set of experiment, considering

DOI 10.17148/IJARCCE.2016.55178

quality and scalability. As ongoing work, we aim at giving big data infrastructure for our designed risk calculation tool, for designing more sophisticated predictive modelling and feature extraction techniques, and extending our proposed solutions to predict other clinical risks.

REFERENCES

[1] Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction by Jyoti Soni, Ujma Ansari, Dipesh Sharma International Journal of Computer Applications

[2] Big Data Solutions for Predicting Risk-of-Readmission foR Congestive Heart Failure Patients by Kiyana Zolfaghar, Naren Meadem, Ankur Teredesai, Senjuti Basu Roy, Si-Chi Chin Institute of Technology, CWDS, UW Tacoma. IEEE 2013.

[3] Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques Chaitrali S. Dangare Sulabha S. Apte, PhD.

[4] Prediction of Heart Disease using Classification Algorithms by Hlaudi Daniel Masethe, Mosima Anna Masethe, USA

[5] Dr.Priti Chandrab, Dr.B.L Deekshatuluc, Research Scholar,JNTU Hyderabad,A.P INDIA

[6] Heart Disease Prediction System using Associative Classification and Genetic Algorithm by M.Akhil jabbar.

[7] Heart Disease Prediction System using Naïve Bayes and Jelinek-mercer smoothing by Ms.Rupali R.Patil Asst. Professor, Jawaharlal Nehru College of Engineering.

DOI 10.17148/IJARCCE.2016.55178