

Introducing K-Means Algorithm to Predict and Detect Heart Attack Disease in Machine Learning

Zeinab Gazala Rafee¹, Gowramma G S²,

¹Dept. Computer Science & Engineering, Don Bosco Institute of Technology, VTU, Bangalore, India

²Dept. Computer Science & Engineering, Don Bosco Institute of Technology, VTU, Bangalore, India

Available online at: www.ijcseonline.org

Accepted: 18/Jun/2018, Published: 30/Jun/2018

Abstract— In this research paper we are going to show how we can predict heart attack diseases and condition by applying patient data to developed software. Early techniques have not been so much efficient in finding it even medical professors are not so much efficient enough in predicting the heart disease. We have plan to solve related this concept by this paper, here some of pre-defined heart related are stored in databases, according those databases value our algorithms (K-means) detect patient condition related heart issues, including Heart Attack, Stroke, Congestive Heart Failure, Angina and Cardiovascular Disease this is very help tool which is used by all humans in any diagnostics or hospitals, here we have plan to upgrade emergency doctor contact sharing technique, if any patient want immediate response from doctors they can take some immediate suggestion form doctors by data sharing technique. Finally, we have concluded to introduce some challenging issues in the design of efficient auditing protocols for prediction heart condition.

Keywords— Heart Attack, Predefined Parameters, Machine Learning, Patient Records, Doctors Records, Heart Disease, Prediction Model.

I. INTRODUCTION

Heart disease has created a lot of serious concern among researches; one of the major challenges in heart disease is correct detection and finding presence of it inside a human. Early techniques have not been so much efficient in finding it even medical professor are not so much efficient enough in predicating the heart disease. There are various medical instruments available in the market for predicting heart disease there are two major problems in them, the first one is that they are very much expensive and second one is that they are not efficiently able to calculate the chance of heart disease in human. According to latest survey the medical professional able to correctly predict only 67% of heart disease so there is a vast scope of research in area of predicating heart disease in human. With advancement in computer science has brought vast opportunities in different areas, medical science is one of the fields where the instrument of computer science can be used.

Application areas of computer science vary from metrology to ocean engineering. Medical science also use some of the major available tools in computer science; in last decade artificial intelligence has gained its moment because of advancement in computation power. Machine Learning is one such tool which is widely utilized in different domains because it doesn't require different algorithm for different dataset. Reprogrammable capacities of machine learning

bring a lot of strength and opens new doors of opportunities for area like medical science. In medical science heart disease is one of the major challenges; because a lot of parameters and technicality is involve for accurately predicating this disease. Machine learning could be a better choice for achieving high accuracy for predicating not only heart disease but also another diseases because this vary tool utilizes feature vector and its various data types under various condition for predicating the heart disease, algorithms receives some heart related parameters shows provides opportunities to minimize the error in predication of heart disease. All these techniques are using old patient record for getting predication about new patient. This predication system for heart disease helps doctors to predict heart disease in the early stage of disease resulting in saving millions of life.

This survey paper is dedicated for wide scope survey in the field of machine learning technique in heart disease. Later in implementation part will discuss about various machine learning algorithm for heart disease and their relative comparison on the various parameter. It also shows future prospectus of machine learning algorithm in heart disease. This paper also does a deep analysis on utilization of deep learning in field of predicting heart disease.

Here we are going to apply K-Means clustering algorithm, it is an unsupervised learning algorithm that, as the name hints,

finds a fixed number (k) of clusters in a set of data. A cluster is a group of data points that are grouped together due to similarities in their features. Clusters of data according to a heart related parameters. Some main parameter we are observed

- Cholesterol
- HDL Cholesterol (High Density Lipoprotein)
- LDL Cholesterol (Low Density Lipoprotein)
- Triglycerides
- Lipid Profile
- CK - Creatine Kinase
- CK - MB
- Myoglobin
- Brain Natriuretic Peptide (BNP)
- Troponin
- Homocysteine
- Cardiac Risk Assessment
- High-sensitivity C-reactive Protein
- Cardiac Biomarkers

Those parameter are input for our artificial intelligent concept, we look later out comes and working of our experimental concept.

II. RELATED WORK

Author analyzes various data mining tools and techniques in health care domain that can be employed in prediction of heart disease system and their efficient diagnosis. Methods/Statistical Analysis: A heart disease prediction model, which implements data mining technique, can help the medical practitioners in detecting the heart disease status based on the patient's clinical data. Data mining classification techniques for good decision making in the field of health care addressed are namely some parameters, protocol, prediction algorithm etc. Applying algorithms helps to make decisions quicker and more precise. Findings: Data mining is a powerful new technology for the extraction of hidden predictive and actionable information from large databases that can be used to gain deep and novel insights. Using advanced data mining techniques to excavate valuable information has been considered as an activist approach to improve the quality and accuracy of healthcare service while lowering the healthcare cost and diagnosis time. Using this technique presence of heart disease can be predicted accurately [1].

Consumers or non-medical professionals are progressively going online to seek health information. Despite the increasing number of health information search online, acquiring the correct and relevant information based on consumer's understanding remains a problem. The information retrieved from the Internet may not fit consumer's understanding because the consumer's familiarity with health topic varies. To improve the accuracy of health information search results, this paper investigates the impact of consumer's familiarity on the search behavior using language models approach. A user experiment was conducted with 60 participants searching on the topic tasks of dengue fever, diabetes mellitus, and gastro esophageal reflux disease. The participants also rated their familiarity with health task topics on the scale of 1 (not familiar at all) to 4 (familiar). This rating categorized the participants into four familiarity groups (F1, F2, F3, and F4). The data analysis involved the transcription of search task data into the sequence of search activities to identify unique search activity patterns between familiarity groups. The survey showed that the familiarity with health topics affected health information search behavior [2].

III. METHODOLOGY

This application can be divided into different methods each having unique functionality.

- Heart attack Data Importing
- Scheduling
- Processing
- Prediction

Heart attack Data Importing

In importing module the n number of heart related parameters are collected datasets are nothing but collection of set of information which are passed to the validating zone where each datasets are validated in the form of single and multiple jobs and are stored in the file systems, the datasets which are taken here are the unformatted data. The datasets are imported in the form of data, the parameter are stored in the database.

Scheduling

Our k-means scheduling module where the scheduler allocates the different resources to multiple tasks and that task should be completed within the scheduled time. Before that the system must be connected to the internet so that the links should be validated, the scheduling discipline may be first come first serve(FCFS), Earliest deadline first, Round Robin(RR), shortest time first etc.

Processing

After the heart related parameters scheduler schedules the jobs for different processor with their resources the next step is processing module where processing can be find the result of heart depending upon the user input. The results are stored in the databases for each job one separate database is allotted, here there are n number of jobs are taken so n number of databases are selected.

Prediction

In Prediction module the data stored in the databases which are reduction data are taken and passed to the allocation level where the reduction data are mapped in somewhere and it passed to the indexed level so that the indexed is created for each word to reduce the database capacity.

IV. K-MEANS ALGORITHM

The K-Means algorithm consists of the following steps:

Step (1): The algorithm reads the database in memory. The database contains several heart attributes instances.

Step (2): The algorithm initialize K empty heart prediction input data clusters. Each cluster has a prototype, which is a randomly generated instance according to updated parameters.

Step (3): Each instance in the database is assigned to the cluster having the closest prototype.

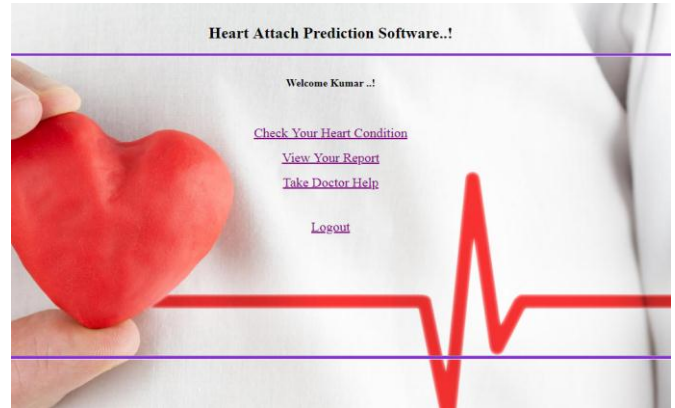
Step (4): Then, the prototype of each cluster is recomputed as the average of all the instances in that cluster.

Step (5): Then, Step3 and Step 4 are repeated several times, until the clusters become stable/predict heart related result (five different stages of result).

V. RESULTS AND DISCUSSION

Here we explain the results of heart failure prediction software project which is running in the Netbeans IDE tool using the java, java swing, AWT languages. In completion of web based project it takes four modules which are explained above here only the results of those modules are explained.

Here mainly we show some of main web pages related heart information and patient result information. Below images shows three filed are check your heart condition, view patient report and doctor help form for guideline after register and login by patient.



Next below form shows patient parameter requesting form after registered by patient. Here mainly 14 fields are used check your heart condition. This form is main part of this software those parameter used.. Cholesterol, HDL Cholesterol (High Density Lipoprotein). LDL Cholesterol (Low Density Lipoprotein), Triglycerides, Lipid Profile, CK - Creatine Kinase, CK – MB, Myoglobin, Brain Natriuretic Peptide (BNP), Troponin, Homocysteine, Cardiac Risk Assessment, High-sensitivity C-reactive Protein, Cardiac Biomarkers.

This image shows patient parameter processed form after registered by patient heart details. Here form display both two fields those are basic information and parameter fields including result. Here the software applied k-means algorithm to predict condition, this algorithm requires parameters are Cholesterol, HDL Cholesterol (High Density Lipoprotein). LDL Cholesterol (Low Density Lipoprotein), Triglycerides, Lipid Profile, CK - Creatine Kinase, CK – MB, Myoglobin, Brain Natriuretic Peptide (BNP), Troponin, Homocysteine, Cardiac Risk Assessment, High-sensitivity C-reactive Protein, Cardiac Biomarkers. This form also supports to report generation to patient mail id and doctor help forms.

Heart Attack Prediction Software..!

Welcome Zainab ..!

General Information		Personality Information	
Phone Number:	987463215	Age:	20
City:	Bangalore	Gender:	male
State:	Karnataka	Weight:	50
Country:	India	Height:	170CM
Heart Related Details			
1. Cholesterol:	5 mg/dL	2. HDL Cholesterol:	6 mg/dL
3. LDL Cholesterol:	3 mg/dL	4. Triglycerides:	2 mg/dL
5. Lipid Profile:	4 mg	6. CK - Creatine Kinase:	1 CK
7. CK - MB:	4 IU/L	8. Myoglobin:	55 nanogram/mL
9. Brain Natriuretic Peptide:	66 pentagram/mL	10. Troponin:	77 ng/dL
11. Homocysteine:	7 micromoles/Liter(mmol/L)	12. Cardiac Risk Assessment:	4 Range
13. High-sensitivity C-reactive Protein:	12 mg/L	14. Cardiac Biomarkers:	2 PO

Your heart related Possibility result Status : Heart Attack

[Generate Report to Mail](#)
 [Take Doctor Help](#)
 [Back](#)
 [Logout](#)

All patient registration, patient basic information, patient login information, doctor information, mail report information, heart parameter information and result stored in databases, k-means algorithm is very useful to find medical related problems mainly in heart attack future detection problems, here we take some reference paper mentioned in reference related this issue, 90% research and development work done by our own concept with reference by k-means algorithm.

VI. CONCLUSION

Early techniques have not been so much efficient in finding it even medical professor are not so much efficient enough in predicating the heart disease. There are various medical instruments available in the market for predicting heart disease there are two major problems in them, the first one is that they are very much expensive and second one is that they are not efficiently able to calculate the chance of heart disease in human. According to latest survey the medical professional are able to correctly predict only 67% of heart disease so there is a vast scope of research in area of predicating heart disease in human. In this concept we conclude about heart related problems we can predict by artificial intelligence applied machine learning concept adopted through k-means algorithm to generate almost 90% accuracy result. We have to show how we can predict heart attack deceases and condition by patient data. Some of pre defined heart related are stored in databases, according to those databases value our algorithms detect patient condition related heart issues, this is very help tool which is used by all humans, here we have plan to upgrade emergency doctor contact sharing technique, if any patient want immediate response from doctors they can take some immediate suggestion form doctors by data sharing technique. Finally, we have concluded to introduce some challenging issues in the design of efficient auditing protocols for prediction heart condition.

REFERENCES

- [1] Mai Shouman, Tim Turner, Rob Stocker, "Using data mining techniques in heart disease diagnosis and treatment", JapanEgypt Conference on Electronics, Communications and Computers 978-1-4673-0483-2 c_2012 IEEE.
- [2] N. Aaditya Sunder, P. PushpaLatha, "Performance analysis of classification data mining techniques over heart disease database" International Journal Of Engineering Science and Advance Technology"-vol-2 issue-3,470-478,May-June 2012
- [3] V. Kirubha and S. M. Priya, "Survey on Data Mining Algorithms in Disease Prediction" vol. 38, no. 3, pp. 124-128, 2016.
- [4] Y. H. Tam, H. S. Hassanein, S. G. Akl, and R. Prediction of Heart Problems. In Proc. of LCN, 2006.
- [5] Y. D. Lin and Y. C. Hsu. Multi-hop cellular: "A new method for prediction of heart disease. "In Proc. of INFOCOM, 2000.
- [6] P. T. Oliver, Dousse, and M. Hasler. "Prediction of Heart Disease using Machine Learning Algorithms. In" Proc. Of hdpt, 2002.
- [7] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm," Procedia Technology., vol. 10, pp. 85-94, 2013..
- [8] SellappanPalaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques," ©2008 IEEE
- [9] ShantakumarB.Patil, Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research" ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656 ©EuroJournals Publishing, Inc. 2009.
- [10] D. Park and M. Scott Corson. "A highly adaptive distributed routing algorithm for health care networks". In Proc. of INFOCOM, 1997
- [11] R. S. Chang, W. Y. Chen, and Y. F. Wen. "Hybrid wireless network and health care protocols". IEEE Transaction on Vehicular Technology, 2003.

Authors Profile

Mrs. Gowramma G.S Completed Bachelor of Engineering in department of Electronics and Communication & Engineering from Bapuji Institute of Technology, India in a year 1997, and completed Master of Technology in department of Computer Science & Engineering in year 2007. She is currently pursuing Ph.D. in VTU, Belgaum in "Data Mining & Artificial Intelligence" and currently working as Associate Professor in Don Bosco Institute of Technology, Department of Computer Science & Engineering, VTU, India since 2007. Her main research work focuses on Data Mining and Artificial Intelligence. She has more than 15 years of teaching experience, 5 years experience as a Software Engineering and 4 years of Research Experience. She has published 5 paper in international reputed journals.



Ms. Zeinab Gazala Rafee Completed Bachelor of Engineering in department of Computer Science & Engineering from HKBK College of Engineering, India. in a year 2014, and pursuing her Master of Technology in department of Computer Science & Engineering in Don Bosco Institute of Technology, VTU University, India in from 2016. She is interested to research in Data Mining, Big Data and Artificial Intelligence concepts and focuses same in Technology development field. She has published survey paper in international digital library.

