

A Machine Learning Approach to Heart Attack Prediction

Rohini Gawale

PG Student, Department of Computer Engineering Matoshri College of Engineering

Prof. R. M. Gawande

Assistant Professor, Department of Computer Engineering Matoshri College of Engineering

Abstract

In today's modern world cardiovascular disease is the most lethal one. This disease assaults an individual right away that may make surprising ramifications for the human life. So diagnosing patients accurately on time is the most testing task for the medicinal crew. The coronary illness treatment is very high and not reasonable by the vast majority of the patients especially in India. The examination extension is to build up an early forecast treatment utilizing information mining advances. Nowadays every hospital keeps the periodical medical reports of cardiovascular patients through a few clinic management gadget to manage their health-care. The data mining techniques namely decision tree and random forest are used to analyze heart attack dataset where classification of more common symptoms related to heart attack is done using C4.5 decision tree algorithm, alongside, random forest is applied to boost the certainty of the classification result of heart attack prediction. A decision tree is used for function selection system and SVM classifier for class. In this system various data mining technologies are applied to make a proactive approach against failures in early predictions diagnosis of the disease. Classification accuracy of SVM algorithm was better than DT algorithm. C4.5 generates a decision tree where each node splits the classes based on the gain of information. The overall accuracy of the SVM using four kernel types was above 73% and the overall accuracy of the DT method was 69%. We proposed an automated system for medical diagnosis that would enhance medical care and reduce cost. Our intent is to provide a ubiquitous service that is both feasible, sustainable and which also make people to assess their risk for heart attack at that point of time or later.

Index Terms—*Storke prediction, Random forest algorithm, KNN, ANN, C4.5 algorithm*

I. INTRODUCTION

Data mining has just settled as a novel field for investigating concealed examples in the colossal datasets. Medicinal science is another handle where enormous measure of information is created utilizing diverse clinical reports and other patient indications. Information mining can likewise be utilized vigorously for a similar reason in therapeutic datasets too. These investigated shrouded designs in therapeutic datasets can be utilized for clinical finding. In any case, therapeutic datasets are generally scattered, heterogeneous, and tremendous in nature. These datasets should be sorted out and coordinated with the clinic the executives frameworks. Cardiovascular infections are one of the most elevated flying ailments of the advanced world. As per world well being association about in excess of 12 million passings happens around the world, consistently because of heart issues. It is additionally one of the deadly sicknesses in India which causes greatest losses. The analysis of this sickness is complicated procedure. It ought to be analyzed precisely and accurately. Because of confinement of the capability of the restorative specialists and their inaccessibility at specific spots put their patients at high hazard. Regularly, it is analyzed utilizing instinct of the restorative pro. It would be profoundly profitable if the strategies will be incorporated with the restorative data framework

II. RELEVANT QUESTION ANSWERING

- 1) Aditi Gavhane, Isha Pandya "Prediction of Heart Disease Using Machine Learning" IEEE Conference Record 2018 IEEE The Heart Disease Prediction System using Machine learning algorithm, viz. MLP provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence author use Multi Layered Perceptron (MLP) in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users.

- 2) S. Kiruthika Devi, S. Krishnapriya and Dristipona Kalita “Prediction of Heart Disease using Data Mining Techniques” Indian Journal of Science and Technology, Vol9(39), October 2016
The accuracy of the algorithms used in each technique can be enhanced by hybridizing or combining algorithms to a single algorithm which may not be accurate for weakly classified sets of data, and is expected to make quicker and more precise decisions.
- 3) M. Lavanya, Mrs. P. M. Gomathi, “Prediction of Heart Disease using Classification Algorithms” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 7, July 2016 Medical related information are gigantic in nature and it tends to be gotten from various origination which are not so much pertinent in include. The exploration embraced an encounter on utilization of shifts information mining calculation to anticipate the coronary episodes and to analyze the based strategy for forecast. The predictive accuracy determined by REPTREE, naive bayes, neural networks, CART. The overall objective is to study the various data mining techniques available to predict the heart.
- 4) Hlaudi Daniel Masethe, Mosima Anna Masethe, “Pre- diction of Heart Disease using Classification Algo- rithms” WCECS 2014, 22-24 October, 2014, San Fran- cisco, USA
The research undertook an experiment on application of various data mining algorithms to predict the heart attacks and to compare the best method of prediction. The research results do not presents a dramatic differ- ence in the prediction when using different classification algorithms in data mining. The experiment can serve as an important tool for physicians to predict risky cases in the practice and advise accordingly. The model from the classification will be able to answer more complex queries in the prediction of heart attack diseases. The predictive accuracy determined by J48, REPTREE and SIMPLE CART algorithms suggests that parameters used are reliable indicators to predict the presence of heart diseases.
- 5) Abhishek Ta, “Heart Disease Prediction System Using Data Mining Techniques”
Right now, point was to structure mama prescient model for coronary illness location utilizing information min- ing procedures from Transthoracic Echocardiography Report dataset that is fit for upgrading the dependability of coronary illness determination utilizing reverberation cardiography.
- 6) Sellappan Palaniappan, Rafiah Awang “Intelligent Heart Disease Prediction System Using Data Mining Tech- niques”
A model coronary illness forecast framework is created utilizing three information mining order demonstrating strategies. The framework separates concealed informa- tion from a verifiable coronary illness database. DMX inquiry language and capacities are utilized to assemble and get to the models. The models are prepared and approved against a test dataset. Lift Chart and Clas- sification Matrix strategies are utilized to assess the viability of the models. Every one of the three models can separate examples in light of the anticipated state. The best model to foresee patients with coronary illness gives off an impression of being Naive Bayes followed by Neural Network and Decision Trees.
- 7) Aditya Methaila, Prince Kansal, “EARLY HEART DIS- EASE PREDICTION USING DATA MINING TECH- NIQUES”
Right now center is around utilizing various calculations and blends of a few objective properties for compelling cardiovascular failure expectation utilizing information mining. Choice Tree has outflanked with 99.62% ex- actness by utilizing 15 traits. Likewise the exactness of the Decision Tree and Bayesian Classification further improves in the wake of applying hereditary calculation to lessen the genuine information size to get the ideal subset of trait adequate for coronary illness expecta- tion. Affiliation grouping procedure apriori calculation, was alongside another calculation MAFIA was utilized. Straight Apriori-based calculations tally the entirety of the 2k subsets of every k-thing set they find, and in this way don't scale for long thing sets. They use "look a heads" to decrease the quantity of thing sets to be checked. MAFIA is an improvement when the thing sets in the database are long.

III. EXISTING SYSTEM

In Existing PHR system data is outsourced on data server over internet. For privacy and security of information, encryp- tion techniques and Cryptography is recycled to provide priv- ileges. For smoother access privileges, the traditional public key encryption (PKE) may costs excess key management over- head and scalability issue drawbacks associated with existing system, which include:

- In the existing gadget excessive upfront acquisition charges
- The ongoing protection expenses, and disruptions to work-flows that make a contribution to temporary losses in productivity.

IV. PROPOSED SYSTEM

In this paper, Data mining is the technique of finding previously unknown patterns and trends in databases and using that information to build predictive models. In this system using a android application user can uploads his/her health record in cloud storage. In cloud storage first user data is processed then training and testing data apply classification C4.5 algorithm to testing data and predict heart attack as shown in figure 1. The application of system studying and records mining techniques shows it provides better performance results. This app will expect the heart assault from the chest pain at an early stage and guide the person to take treatment early such as to get their ECG done as early as possible and get evaluated from a doctor to make diagnoses of heart attack.

V. SYSTEM ARCHITECTURE

In the Support Vector Machine algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

The basic idea behind any decision tree algorithm is as follows: Choose the best attribute(s) to split the remaining instances and make that attribute a decision node Repeat this process for recursively for each child Stop when: All the instances have the same target attribute value There are no more attributes There are no more instances

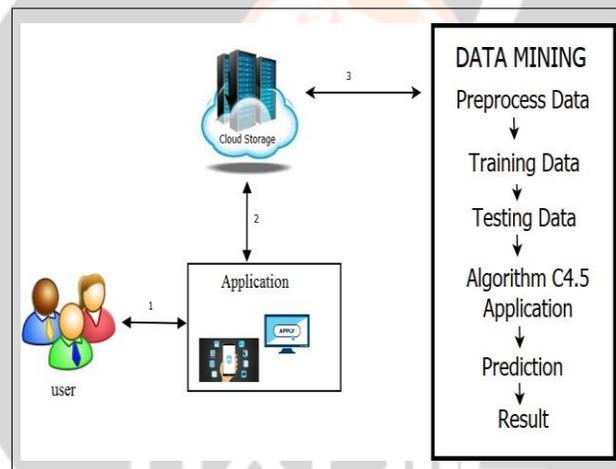


Fig. 1. System Architecture

VI. SYSTEM REQUIREMENTS

A. Software Requirement

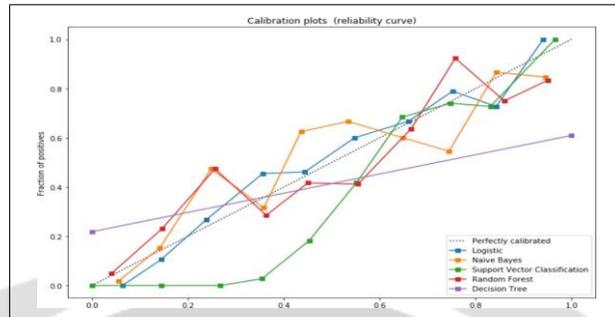
- 1) Operating System: Windows 7 or above
- 2) Programming Language: Python 3.7
- 3) IDE: Python IDLE

B. Hardware Requirement

- 1) Processor: Pentium Processor Core 2 Duo or Higher
- 2) Hard Disk: 250 GB (min)
- 3) RAM: 1GB or higher
- 4) Processor Speed: 3.2 GHz or faster processor

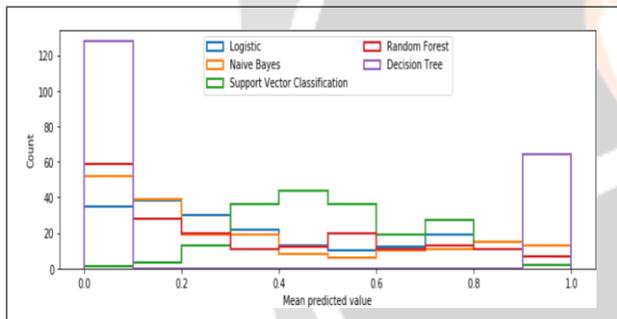
VII. EXPERIMENTAL ANALYSIS

Experimental Analysis show the comparison of various machine learning algorithms used for heart attack prediction. Our system which is developed in python has used decision tree classification algorithm C4.5. The graph shows decision tree analysis in purple line, this gives more accurate result than earlier existing results.



VIII. CONCLUSION

Hence we compare Decision Tree with SVM from the result it is found that SVM produce more accurate result in further system can be improved with some large dataset.



REFERENCES

- [1] Stefano Bromuri , Serban Puricel, Rene Schumann , Johannes rampf, Juan Ruiz and Michael Schumacher, -An expert Personal Health System to monitor patients affected by Gestational Diabetes Mellitus: A feasibility study, Journal of Ambient Intelligence and Smart Environ-ments 8 (2016) 219237G [1].
- [2] Gyorgy J. Simon, Member, IEEE, Pedro J. Caraballo, Terry M. Th- erneau, Steven S. Cha, M. Regina Castro and Peter W. Li , -Extending Association Rule Summarization Techniques to Assess Risk of Dia- betes Mellitus, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEER-ING [2].
- [3] Han Wu, Shengqi Yang , Zhangqin Huang, Jian He, Xiaoyi Wang, - Type 2 diabetes mellitus prediction model based on data mining, Informatics in Medicine Unlocked 10 (2018) 100107 [3].
- [4] Xue-Hui Meng a, Yi-Xiang Huang a, Dong-Ping Rao b, Qiu Zhang a, Qing Liu b, - Comparison of three data mining models for predicting diabetes or prediabetes by risk factors, Kaohsiung Journal of Medical Sciences (2013)29, 93e99 [4].
- [5] Kung-Jeng Wang, Angelia Melani Adrian a, Kun-Huang Chen a, Kung- Min Wang b, - An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus, Journal of Biomedical Informatics 54 (2015) 220229 [5].
- [6] Jinn-Yi Yeh, Tai-Hsi Wu b, Chuan-Wei Tsao, -Using data mining tech- niques to predict hospitalization of hemodialysis patients, Decision Sup- port Systems 50 (2011) 439448 [6].
- [7] Nongyao Nai-arun, Rungrutikarn Moungrmai , -Comparison of Classifiers for the Risk of Diabetes Prediction, 7th International Conference on Advances in Information Technology [7].

- [8] Asha Gowda Karegowda , M.A. Jayaram, A.S. Manjunath, - Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients, IJEAT Volume-1, Issue-3, February 2012 [8].
- [9] Darshan K R, Anandakumar K R, A Comprehensive Review on Us-age of Internet of Things (IoT) in Healthcare System 2015 IEEE ICERECS&T [9].
- [10] Chih-Hua Tai, Daw-Tung Lin, A Framework for Healthcare Every- where: BMI Prediction using Kinect and Data Mining Techniques on Mobiles, 2015 16th IEEE International Conference on Mobile Data Management[10].
- [11] Ritika Chadha , Shubhankar Mayank, Prediction of heart disease using data mining techniques, 2016 OF CSIT Springer [11].
- [12] Eleni I. Georga, Vasilios C. Protopappas, Stavroula G. Mougiakakou,, Short-term vs. Long-term Analysis of Diabetes Data: Application of Machine Learning and Data Mining Techniques, IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 2003 [12].

