



A Review on Prediction of Heart Attack Using Data Mining Algorithms

Prof. Thakur Ritesh¹, Bendre Gauri², Darekar Ashwini², Shinde Priyanka²

Asst. Professor, Department of Computer Engineering, Savitribai Phule Pune University, Institute of Knowledge
College of Engineering, Pimpale Jagtap, Pune, India¹

Student, Department of Computer Engineering, Savitribai Phule Pune University, Institute of Knowledge College of
Engineering, Pimpale Jagtap, Pune, India²

ABSTRACT : The Heart Attack Prediction System is an web-based application planned for the therapeutic field. The human services condition is observed to be rich in data, however poor in extricating learning from the data. Information mining instruments customarily were taken much tedious to determine. The enormous measures of information created for the forecast of coronary illness are excessively mind boggling and voluminous, making it impossible to be prepared and broke down by customary strategies. The proposed framework is planned to build up an Intelligent System utilizing information mining procedure, using Naive Bayes, C4.5 and k-means calculation and actualized as Java application. The test outcomes are utilized to figure to register forecast result.

KEYWORDS: data mining, classification, entropy, gain, health informatics, Naive Bayes, C4.5, K-means Clustering

I. INTRODUCTION

The coronary ailment is the best explanation behind death nowadays. Blood weight, cholesterol, beat rate are the critical clarification behind the coronary disease. Some non-modifiable segments are also there. For example, smoking, drinking similarly clarification behind coronary sickness. The heart is a working course of action of our human body. In case the limit of the heart is not done really suggests, it will impact another human body part in addition. Some threat factors of coronary disease are Family history, High circulatory strain, Cholesterol, Age, Poor eating standard, Smoking. Right when veins are overstretched, the peril level of the veins are extended. This prompts the blood weight. Heartbeat is commonly measured similar to systolic and diastolic. Systolic exhibits the weight in the passageways when the heart muscle contracts and diastolic demonstrates the weight in the hallways when the heart muscle is in resting state. The level of lipids or fats extended in the blood causes the coronary ailment. The lipids are in the veins consequently the passageways get the chance to be confined and circulatory system is furthermore ended up being moderate. Age is the non-modifiable threat segment which furthermore a clarification behind the coronary disease. Smoking is the reason behind 40% of the death of heart diseases. Since it compels the oxygen level in the blood by then it damages and settles the veins. Distinctive data mining techniques, for instance, Naive Bayes, KNN estimation, Choice tree, Neural Network are used to anticipate the risk of coronary ailment. The KNN count uses the K customer described quality to find the estimations of the components of coronary sickness. Decision tree figuring is used to give the gathered answer to the coronary sickness. The Naive Bayes technique is used to envision the coronary sickness through probability. The Neural Network gives the limited screw up of the desire of coronary sickness. In this already said techniques the patient records are requested and foreseen reliably. The patient activity is checked reliably if there is any movements happen, then the danger level of disease is taught to the patient and master. The authorities can anticipate heart diseases at an earlier stage in perspective of machine learning figurines and with the help of PC development.

Information mining is an interdisciplinary subfield of programming designing. It is the computational method of discovering cases in limitless data sets including strategies at the intersection purpose of fake awareness, machine learning, estimations, and database structures. The general goal of the data mining technique is to think information from a data set and change it into a sensible structure for further use. Other than the unrefined examination step, it

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

incorporates database and data organization edges, data preprocessing, model and induction considerations, interesting estimations, disease quality thoughts, post-planning of discovered structures, portrayal, and the web overhauling. Information mining is the examination wander off the "learning disclosure in databases" prepare.

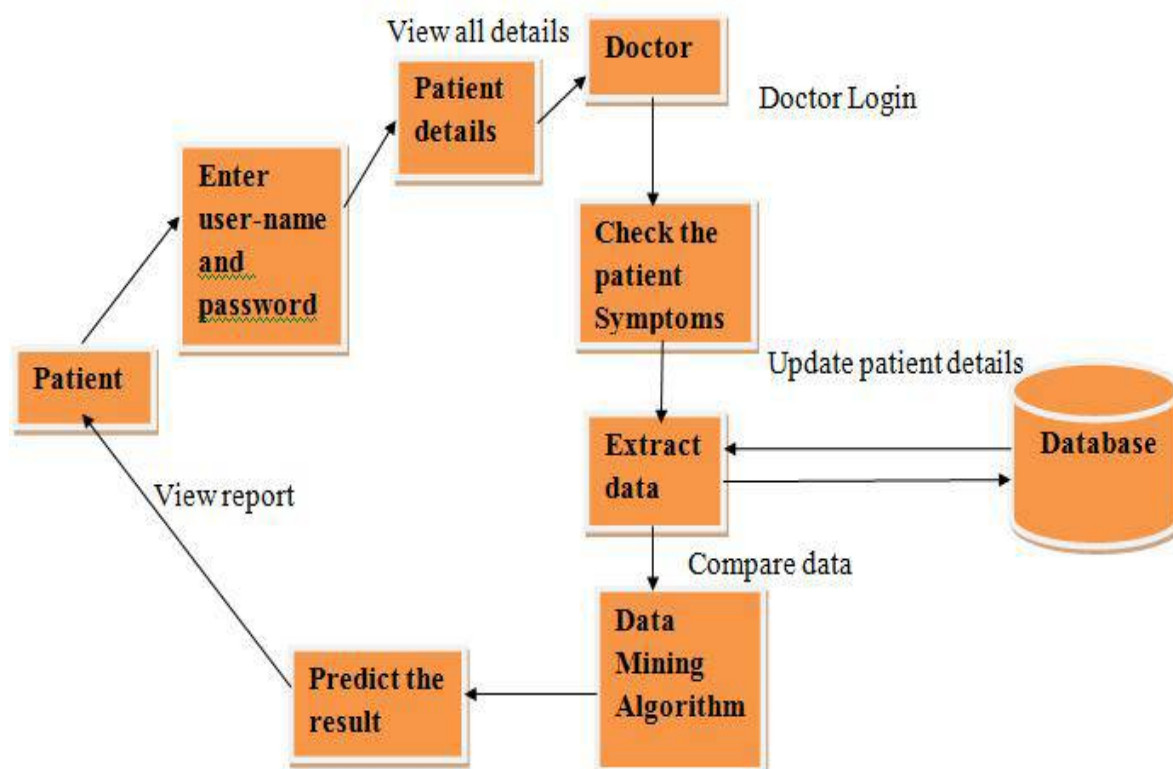
II. EXISTING SYSTEM

In existing framework single information mining strategy is utilized to analyze the coronary illness. There is no past research that distinguishes which information mining method can give more dependable exactness in recognize reasonable treatment for heart illnesses patients. Down to earth utilization of medicinal services database framework and learning disclosure is troublesome in heart maladies analysis.

III. PROPOSED SYSTEM

In Proposed System, proper pharmaceuticals for heart patients is gainful and needs support examination. To evaluate if applying information mining systems to heart assault identification can give as tried and true execution as refined in heart assault conclusion.

In this system we are using Naive Bayes Theorem, C4.5 Algorithm, K-means Algorithm.



Algorithms:

A. Naive Bayes:

Bayesian network (BN) are parametric classifiers and have problems with multi source data. Sungbo et al. have proposed a new classification models that exploit temporal relations among features which within and across data



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

stream. . The consideration of such new relations improve significantly the classification accuracy. It also improves the interpretability of the resulting models. Data streams are collected using diverse sensors. These multiple diverse sensors are used to monitor changes. Sensor readings are obviously dependent. Hence, changes detected in one sensor might affect readings in others. However, the interdependency among sensor readings and their temporal relations have not been treated impervious classifications work. To handle the problem of temporal relations, authors provide a monitoring scenario of mobile robots with many sensors. The robot is engaged in various tasks and sends to a central node.

B. C4.5:

The C4.5 algorithm generates a decision tree for the given data by recursively splitting that data. The decision tree grows using Depth-first strategy. The C4.5 algorithm considers all the possible tests that can split the data and selects a test that gives the best information gain (i.e. highest gain ratio). This test removes ID3's bias in favor of wide decision trees . For each discrete attribute, one test is used to produce many outcomes as the number of distinct values of the attribute. For each continuous attribute, the data is sorted, and the entropy gain is calculated based on binary cuts on each distinct value in one scan of the sorted data. This process is repeated for all continuous attributes. The C4.5 algorithm allows pruning of the resulting decision trees.

C. K-means:

Grouping a set of objects in such a way that objects in the same group is more similar to each other than to those in other groups. Clustering is an unsupervised learning. The algorithm clusters information's into k groups, where k is considered as an input parameter. Next it assigns each information's to clusters based upon the observation's proximity to the mean of the cluster. The cluster's mean is then more computed and the process will continue again. The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging and related fields. The steps involved in Kmeans algorithms are as follows: - Choose the number of clusters, k. - Randomly create k clusters and find the cluster midpoint. - Consign each point to the closest cluster midpoint - Recompute the new cluster midpoints. - Iterate the above two steps until various convergence condition is met.

Modules :

A. Patient Module

Patient have to register in system for his/her heart prediction. After registration login id and password will generate. Patient should login in this system for upload his/ her result. This result will send to doctor. Patient can take appointment from doctor which will accept by particular doctor only. Patient can also give feedback to particular doctor.

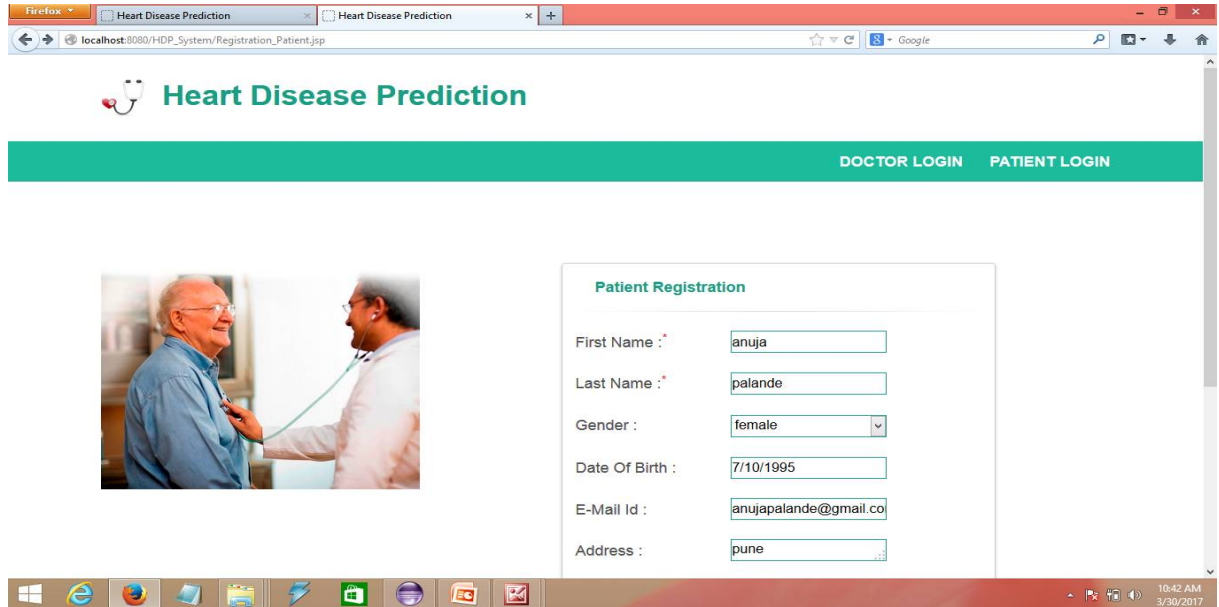


International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

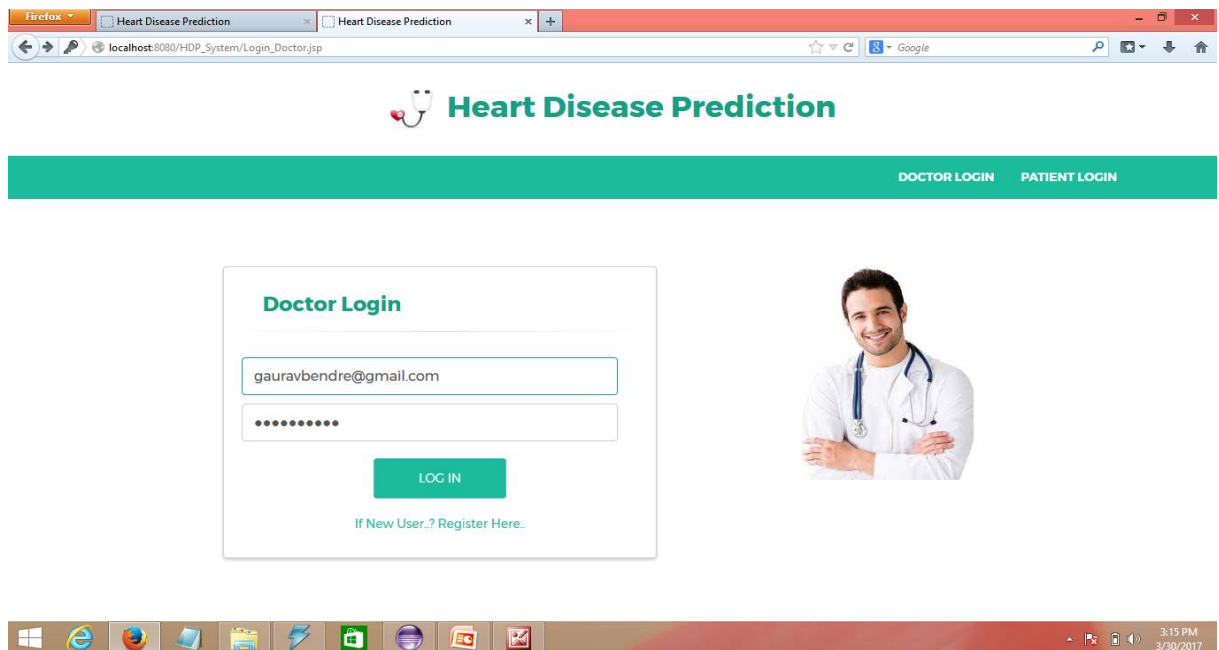
Website: www.ijircce.com

Vol. 5, Issue 3, March 2017



B. Doctor Module

Doctor registration is there to use this system. Doctor will get patient report after that doctor analyse report using Naive Bayes Algorithm, C4.5 Algorithm, K-means Algorithm. Doctor can accept appointment and can view feedback given by patient.



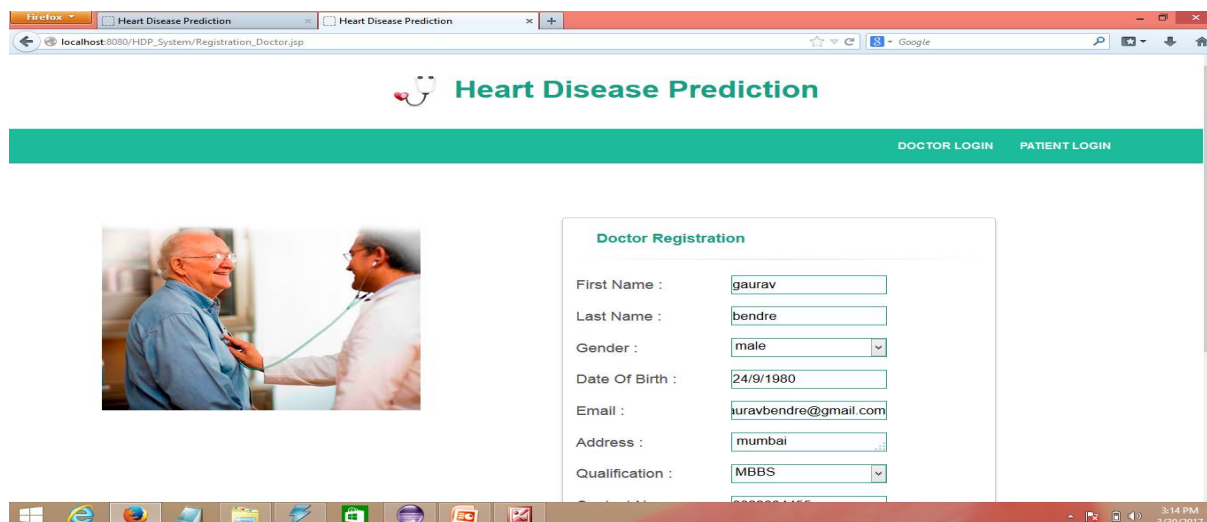


International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017



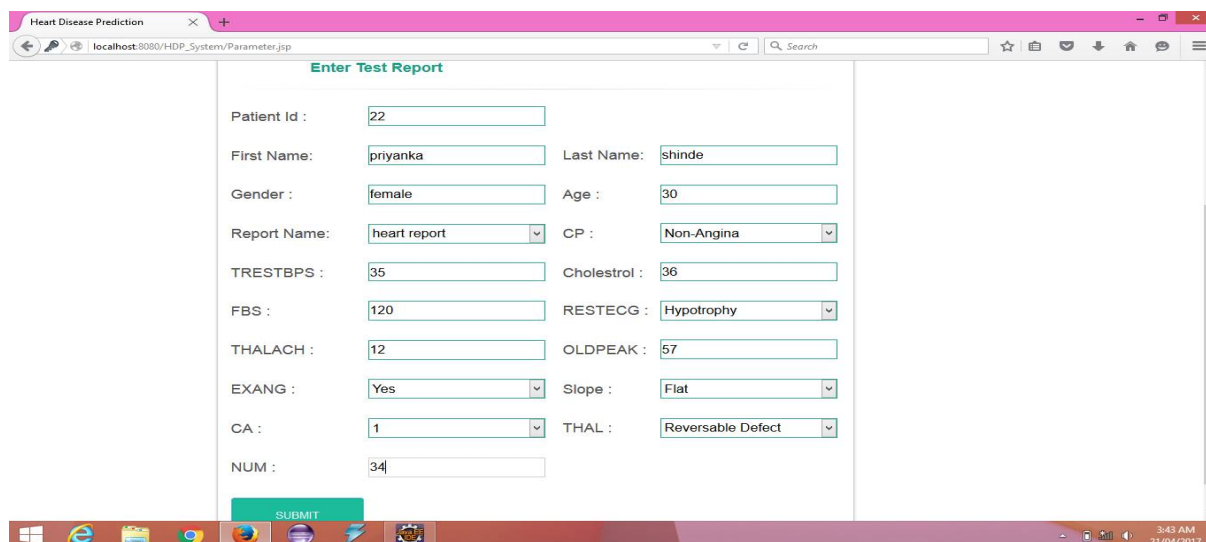
C. Database

MySQL is used as a backend database which store all the detailed information in table format. So it is user friendly and easy to access.

IV. APPLICATIONS

1. In medical and health care areas.
2. Disease diagnosis is one of the applications where data mining tools are proving successful results.
3. Researchers have long been concerned with applying statistical and data mining tools to improve data analysis on large data sets.
4. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care.

V. RESULT





International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 3, March 2017

VI. CONCLUSION

In this paper recently Heart Attack Prediction framework in light of Naive Bayes, C4.5, K-means calculation is presented. For this Data mining idea is utilized, so this paper gives a risk and fundamental quick of different desired models in information mining and finds a most unmistakable model for further work.

This framework is advantageous ,powerful and gives great forecast of sickness to tolerant. This work can be enhanced by extending the amount of attributes for the present game plan of our past work.



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirce.com

Vol. 5, Issue 3, March 2017

REFERENCES

- [1] Ranganatha S., Pooja Raj H.R., Anusha C., Vinay S.K., "MEDICAL DATA MINING AND ANALYSIS FOR HEART DISEASE DATASET USING CLASSIFICATION TECHNIQUES".
- [2] Shan Xu, Haoyue Shi, Xiaohui Duan, Tiangang Zhu, Peihua Wu, Dongyue Liu, "Cardiovascular Risk Prediction Method Based on Test Analysis and Data Mining Ensemble System".
- [3] Theresa Princy. R, J. Thomas, " Human Heart Disease Prediction System using DataMining Techniques", in International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016.
- [4] Sana Shaikh, Amit Sawant, Shreerang Paradkar, Kedar Patil, " Electronic Recording System-Heart Disease Prediction System", in International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04 – 06, 2015, Mumbai, India.
- [5] Mai Shouman, Tim Turner, Rob Stocker, "USING DATA MINING TECHNIQUES IN HEART DISEASE DIAGNOSIS AND TREATMENT", IEEE, 2012.
- [6] Devendra Ratnaparkhi, Tushar Mahajan, Vishal Jadhav, " Heart Disease Prediction System Using Data Mining Technique ", in International Research Journal of Engineering and Technology (IRJET) Volume: 02 Issue: 08 | Nov-2015, IRJET ISO 9001:2008 Certified Journal Page 1553.
- [7] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, " Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review ", in International Journal of Computer Applications (0975 – 8887) Volume 136 – No.2, February 2016.
- [8] Deepali Chandna, "Diagnosis of Heart Disease Using Data Mining Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1678-1680.