International Journal of Latest Trends in Engineering and Technology
Vol.(8)Issue(3), pp.045-050
DOI: http://dx.doi.org/10.21172/1.83.006
e-ISSN:2278-621X

PREDICTION OF HEART DISEASE AND STRATEGIC DECISION MAKING FOR PHI OF MEDICAL DATASET

Geetha Guttikonda[1], Sneha Cherukuri[2], Chandra Naga sravanthi[3], Mohammad Irfanullah[4] and Monica Korlapati[5]

Abstract- Individuals take standard medical examinations for the most part not for finding virus but rather to have genuine feelings of serenity with respect to their wellbeing status. Along these lines, it is imperative to give them a general criticism as for all the wellbeing markers that have been positioned against the entire populace. Here, we propose a framework for prediction of heart disease for a taken dataset. Especially, the highest health risk is revealed in the cases of people who are having heart disease and not predicting it before hand. The huge amounts of data generated for prediction of heart disease are too complex and voluminous to be processed and analyzed by traditional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. By using data mining techniques it takes less time for the prediction of the disease with more accuracy.

Keywords – Data mining, prediction, heart disease

I. INTRODUCTION

HEART attack diseases remain the main cause of death worldwide. In order to prevent these attacks they must be detected at earlier stages. Medical practitioners generate data but effective predictions are not done on it. For this purpose, the research converts the unused data into a dataset for modeling using different data mining techniques. People die due to various symptoms which were not taken into consideration. So the medical practitioners must predict heart disease before they occur in their patients. The features that increase the possibility of heart attacks are smoking, lack of physical exercises, high blood pressure etc[1].

Heart attack mainly occurs when there is irregularity in the flow of blood and heart muscle is injured because of inadequate oxygen supply. World Health Organization in the year 2008 reported that 30% of total global deaths are due to Cardio Vascular Disease (CVD). And by 2030, almost 25 million people will die from CVDs, mainly from heart disease and stroke. CVD is expected to be the leading cause of deaths in developing countries due to various changes in lifestyle, work culture and food habits. Therefore, more careful and efficient methods of cardiac diseases and periodic examination are of high importance [3].

From the Analysis of Heart Attack prediction system, a major challenge confronting healthcare organizations is arrangement of value administrations at moderate expenses. There are number of factors which increases risk of Heart disease [5].

I. Family history of heart disease
II. Smoking
III. Cholesterol
IV. High blood pressure
V. Obesity
VI. Lack of physical exercise
VII. Poor diet
VIII. Physical inactivity

[1] *Department of Information Technology V R Siddhartha Engineering College, AP, India*
[2] *Department of Information Technology V R Siddhartha Engineering College, AP, India*
[3] *Department of Information Technology V R Siddhartha Engineering College, AP, India*
[4] *Department of Information Technology V R Siddhartha Engineering College, AP, India*
[5] *Department of Information Technology V R Siddhartha Engineering College, AP, India*

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. The current research intends to predict the probability of getting heart disease from given patient data set [1].
Data mining is the solution to this serious problem. Data mining is an essential step in the process of knowledge discovery in databases. Thus data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.
Here we discuss about the classification of data mining technique to predict diagnosis of heart ailments efficiently, with reduced number of attributes that contribute more towards the cardiac disease and also Naïve Bayes, a data mining modeling technique which uses the concepts of conditional probability and a historical data set [2].
Data Mining involves some of the steps from raw data collection to some form of new knowledge. The iterative procedure comprises of Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation and Knowledge Representation. Medical history data comprise of a number of tests essentials to diagnose a particular disease [3].

II. LITERATURE SURVEY

In this research paper Data mining algorithms such as J48, Naive Bayes, REPTREE, CART, and Bayes Net are applied for predicting heart attacks. The coronary illness records to be the main reason for death around the world. It is troublesome for therapeutic specialists to anticipate the heart assault as it is a mind boggling errand that requires understanding and learning. The well being division today contains concealed data that can be critical in deciding. Information mining calculations, for example, J48, Naive Bayes, REPTREE, CART, and Bayes Net are connected in this investigate for anticipating heart assaults. The exploration result appears expectation exactness of 99%. Information mining empowers the wellbeing segment to foresee designs in the dataset [1].
This paper mainly focuses on two algorithms naive bayes, genetic algorithms which predict the risk level of heart disease. With the immensely developing populace, the specialists and specialists accessible are not in extent with the populace. Likewise, side effects of heart illness may not be noteworthy and in this manner, may regularly be disregarded. So they proposed an Intelligent Heart Disease Decision Emotionally supportive network to help the specialists connect with those individuals who are denied of these restorative administrations. When all is said in done, it can fill in as a preparation device to prepare attendants and therapeutic understudies to analyze patients having danger of coronary illness [2].

This paper describes a preprocessing technique and analyzes the accuracy for prediction after preprocessing the noisy data. It is also observed that the precision has been expanded to 91% after preprocessing. Swarm Intelligence systems hybrided with Rough Set Algorithm are to be taken as future work for correct lessening of applicable elements for forecast [3].
The author has focused on how these days individuals take a shot at PCs for quite a long time what's more, hours they do not have room, schedule-wise to deal with themselves. Because of wild calendars and utilization of garbage nourishment it influences the soundness of individuals and chiefly heart. So they are actualizing a coronary illness expectation framework utilizing information mining strategy Naive Bayes and k-Means bunching calculation. It is the mix of both the calculations. This paper gives a review for the same. It helps in foreseeing the coronary illness utilizing different characteristics and it predicts the yield as in the expectation frame. For gathering of different characteristics it utilizes k-Means calculation and for anticipating it utilizes Naive bayes calculation [4].
This paper aims at implementing the information digging procedure hereditary calculation for coronary illness forecast. The perceptions uncover that hereditary calculation with 14 traits give great outcome. Conclusion from this execution demonstrates that hereditary calculation has precision of 73.46% and takes elapsed time and energy 1285.99joules separately. Information mining instrument utilized for the usage is matlab2013 [5].

III. RELATED WORK

There were two existing systems we have studied for this paper. First approach is combining Naïve Bayes and Genetic algorithm to improve the classification accuracy of heart disease dataset. They utilized hereditary results as a decency measure to trim excess and unwanted characteristics and to rank the characteristics which help more towards order. Slightest positioned characteristics are uprooted and order calculation is based around the assessed characteristics. That classifier is prepared to order coronary illness information set as either sound or debilitated. Their existing system consists of 2 sections:
a. To Start with managing the attributes using genetic search.
b. Next it deals with building classifier and measuring accuracy of the genetic algorithm [9].
Second one is a pragmatic approach in which there is a preparatory stage, where the informational collection is preprocessed by utilizing numeric to nominal and replaces missing esteem strategies. In the wake of cleaning, the informational index is prepared for exactness. The next stage is the extraction of excess component for prediction. This is to be affected by utilizing the Swarm Intelligence Techniques hybrid with Rough set calculation. The informational collection is approved to secure forecast [3].

The following is the proposed work we have done:
- Firstly the patient's data collected is preprocessed.

- Next the naïve bayes classification is done on it and the data is classified, but the result obtained is not so clear.
- Later EM clustering is done on the data where it is divided into two clusters based on the ranges that we have set to it.
- After the clustering process we have generated a decision tree.
- Finally we have applied genetic algorithm on the data where the grouping is done according to age and fitness function is calculated and finally the prediction of heart disease is done.

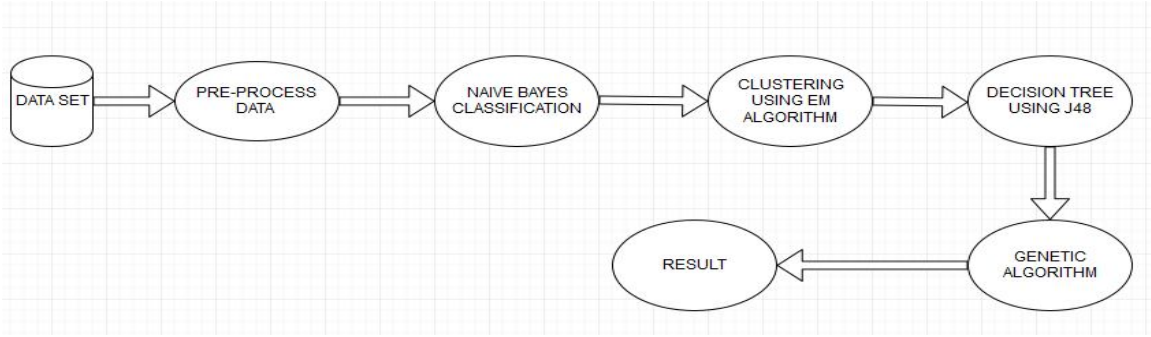


Figure 1:Processing steps of Proposed framework

Naïve Bayes Classifier:
This paper utilizes Naïve Bayes in order to build a heart disease prediction framework which is used to analyze the risk factor. Here it uses historical heart disease dataset in order to predict the risk level. It is a simple probabilistic classifier based on applying Bayes theorem with strong independence assumptions. It is not a single algorithm but family of algorithms based on a common principle that all the naive bayes classifier assumes that the value of particular feature is independent of the value of other feature. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. Here it uses the method of maximum likelihood [6].
Using Bayes,

In other words,
EM algorithm:
Starting from some initial guess, each iteration consists of
an E step (Expectation step)
an M step (Maximization step)
The key idea in EM is to obtain a set of working values for the missing data by substituting an expectation for each missing value. Compute probability densities for each value given a cluster and compute new belonging probabilities on the basis of these probability densities. Calculate the value of log likelihood based on the probability densities. Create the clusters by assigning each value to a cluster such that the belonging probabilities are maximized. EM is frequently used for data clustering in machine learning and computer vision.[7]

Decision Tree:
A Decision tree is a decision support tool which uses a tree like graph or model of decisions and their possible outcomes including their resource cost, utility etc. These decision tress are mostly used in decision analysis in order to reach a goal. The J48 algorithm forms a tree using the divide and conquers technique. Decision trees are most influential methodologies in learning disclosure and data mining. These are powerful instruments in numerous zones.

Genetic Algorithm:
A Genetic algorithm is a search heuristic which mimics the process of natural evolution. Genetic algorithms belong to the larger class of evolutionary algorithms which generates solutions to various optimization problems using inheritance, selection, cross over and mutation this is a adaptive search algorithm based on natural selection. This algorithm requires genetic representation of problem domain and fitness function. Fitness function is problem dependent. Initially many solutions are generated to form an initial population. During each successive generation, a proportion of the existing population is selected to breed a new generation. Individual solutions are selected through a fitness-based process, where best solutions are selected. The next step is to generate a second generation population of solutions from those selected through genetic operators: crossover (also called recombination) and mutation.
1. Make an arbitrary starting populace.
2. Current population can be used to find the fitness value and create new populace.
3. Selection of members, called parents, based on their fitness.
4. Some of the individuals in the current populace that have lower fitness are chosen as elite.
These are passed to the next populace.
5. Produces children from the parents and the operation is known as crossover. [8]

RESULTS

Figure 2: Input Dataset

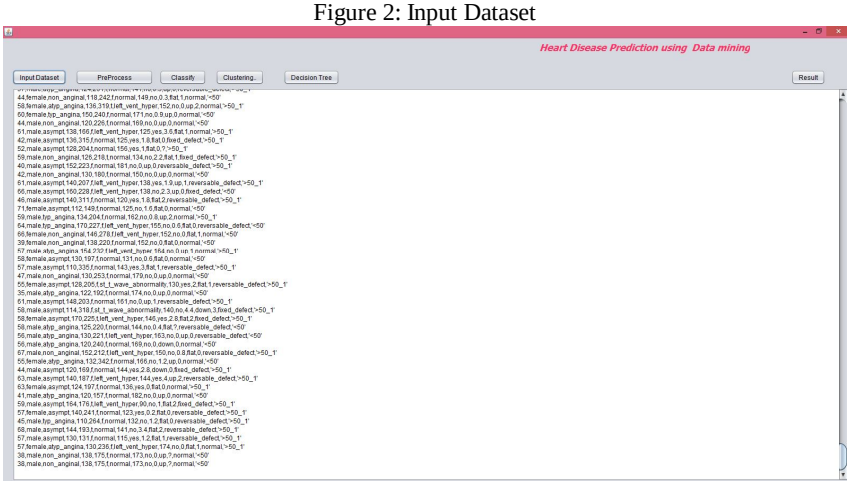


Figure 3: Preprocessing results

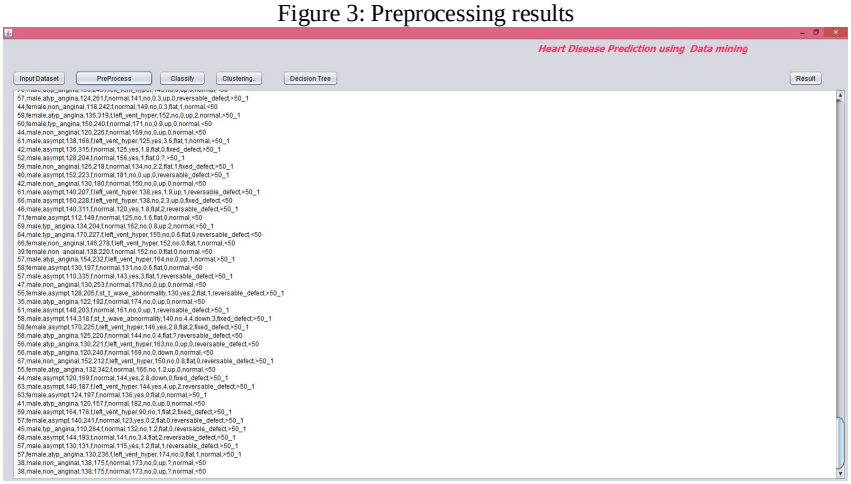


Figure 4: Classification results: Naïve Bayes Algorithm

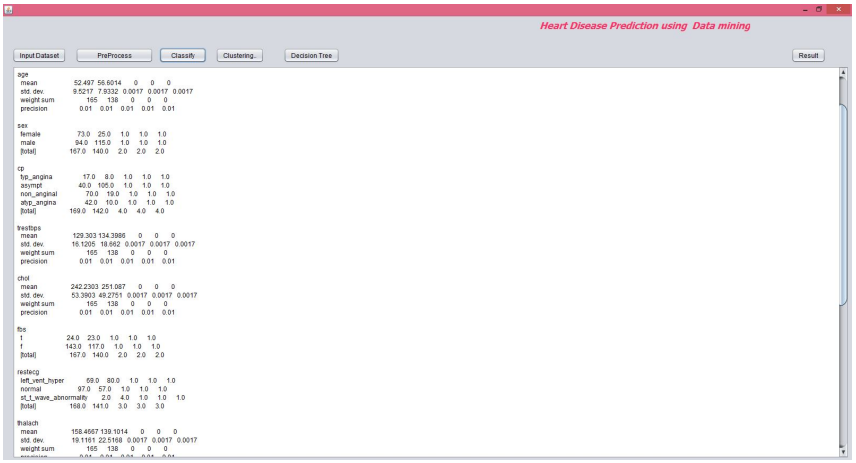


Figure 5: Clustering Results: EM clustering

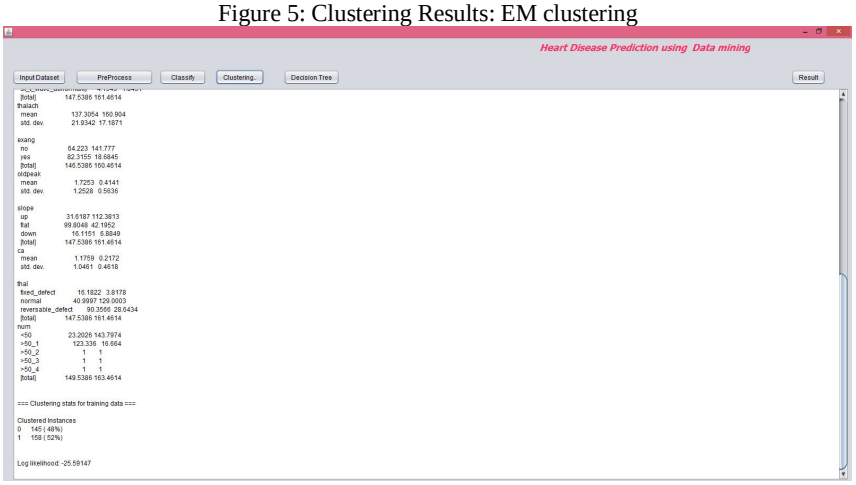


Figure 6: Decision Tree result:

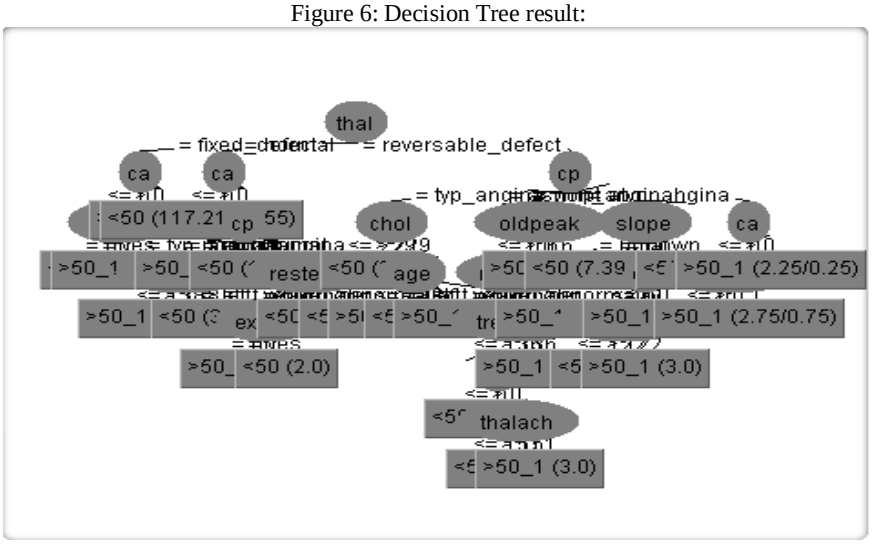
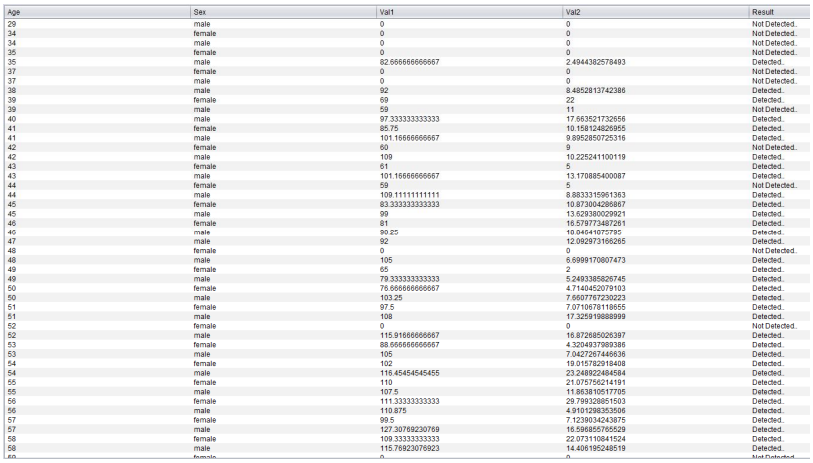


Figure 7: Prediction Results: Genetic Algorithm



Age	Sex	Val1	Val2	Result
29	male	0	0	Not Detected..
34	female	0	0	Not Detected..
34	male	0	0	Not Detected..
35	female	0	0	Not Detected..
35	male	82.666666666667	2.4944382578493	Detected..
37	female	0	0	Not Detected..
37	male	0	0	Not Detected..
38	male	92	8.4852813742386	Detected..
39	female	69	22	Detected..
39	male	59	11	Not Detected..
40	male	97.333333333333	17.663521732656	Detected..
41	female	85.75	10.158124826955	Detected..
41	male	101.16666666667	9.8952850725316	Detected..
42	female	60	9	Not Detected..
42	male	109	10.225241100119	Detected..
43	female	61	5	Detected..
43	male	101.16666666667	13.170885400087	Detected..
44	female	59	5	Not Detected..
44	male	109.11111111111	8.883331561363	Detected..
45	female	83.333333333333	10.873004286867	Detected..
45	male	99	13.629380029921	Detected..
46	female	81	16.579773487261	Detected..
46	male	99.25	10.64541975795	Detected..
47	male	92	12.092973166265	Detected..
48	female	0	0	Not Detected..
48	male	105	6.6999170807473	Detected..
49	female	65	2	Detected..
49	male	79.333333333333	5.2493385826745	Detected..
50	female	76.666666666667	4.7140452079103	Detected..
50	male	103.25	7.6607767230223	Detected..
51	female	97.5	7.0710678118655	Detected..
51	male	108	17.325919888999	Detected..
52	female	0	0	Not Detected..
52	male	115.91666666667	16.872685026397	Detected..
53	female	88.666666666667	4.3249937999386	Detected..
53	male	105	7.0427267446636	Detected..
54	female	102	19.015782918408	Detected..
54	male	116.45454545455	23.249922484584	Detected..
55	female	110	21.075756214191	Detected..
55	male	107.5	11.863810517705	Detected..
56	female	111.33333333333	29.799328851503	Detected..
56	male	110.875	4.9101298353506	Detected..
57	female	99.5	7.1239034243875	Detected..
57	male	127.30769230769	16.596855765529	Detected..
58	female	109.33333333333	22.073110841524	Detected..
58	male	115.76923076923	14.406195248519	Detected..
59	female	0	0	Not Detected..

REFERENCES

[1] Masethe, Hlaudi Daniel, and Mosima Anna Masethe. "Prediction of heart disease using classification algorithms." *Proceedings of the world congress on engineering and computer science*. Vol. 2. 2014.

[2] Shruti Ratnakar, K. Rajeswari, Rose Jacob "Prediction of heart disease using genetic algorithm for selection of optimal reduced set of attributes" *International Journal of Advanced Computational Engineering and Networking* ISSN (p): 2320-2106, Volume-1, Issue-2, April-2013

[3] Dr. Durairaj.M, Sivagowry.S "A Pragmatic Approach of Preprocessing the Data Set for Heart Disease Prediction" *International Journal of Innovative Research in Computer and Communication Engineering* Vol. 2, Issue 11, November 2014

[4] Shinde, Rucha, et al. "An Intelligent Heart Disease Prediction System Using K-Means Clustering and Naïve Bayes Algorithm." *IJCSIT) International Journal of Computer Science and Information Technologies* 6.1 (2015): 637-639.

[5] Beant Kaur , Dr. Williamjeet Singh "Analysis of heart attack prediction system using genetic algorithm" International Journal of Advanced Technology in Engineering and Science Vol.No.,3, Issue No.1, August 2015

[6] http://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes-classifier.pdf

[7] https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm.

[8] https://in.mathworks.com/help/gads/how-the-genetic-algorithm-works.html