



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT

2018 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 11th Febraury 2018. Link :

<http://www.ijiemr.org/downloads.php?vol=Volume-7&issue=ISSUE-02>

Title: Analysis of Heart Diseases Using Data Mining Classification Technique.

Volume 07, Issue 02, Page No: 190 – 196.

Paper Authors

***SARANGAM KODATI, DR. R VIVEKANANDAM, B. MADHUSUDHAN REDDY.**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

ANALYSIS OF HEART DISEASES USING DATA MINING CLASSIFICATION TECHNIQUE

¹SARANGAM KODATI, ²DR. R VIVEKANANDAM, ³B. MADHUSUDHAN REDDY

¹Research Scholar, Dept of CSE, Sri Satya Sai University of Technology and Medical Science, Sehore, Bhopal, Madhya Pradesh, India

²Professor, Director in Muthayammal Engineering College, Namakkal, India

³Assistant Professor, Dept of ECE, Vignan Institute of Technology and Science, Telangana, India

ABSTRACT

Heart disease is the leading cause of death in the world over the past 10 years. Researchers have been using numerous data mining techniques according to help health care specialists into the diagnosis on heart disease. The important objective of this research paper is to discuss about the classification algorithms applied regarding different types on real time medical data sets and Analysis its performance. Classification algorithms providing maximum precision on various kinds on medical data sets are performed namely optimum result for Performance evaluation report. Performance evaluation file comprises about the most frequently used algorithms regarding respective medical dataset and efficient classification algorithm according to analyze the particular disease. The Analysis on study Naïve Bayes Method.

Keywords: Data Mining, Heart Disease, Naïve Bayes Method, WEKA .

I.INTRODUCTION

Data mining is a large area up to expectation integrates techniques from numerous fields consisting of machine learning, statistics, pattern recognition, artificial intelligence, and database systems, for the analysis of large volumes concerning data. There have been a huge range concerning data mining algorithms rooted of these fields to perform different data analysis tasks. Data Mining is the process over extracting hidden knowledge from large volumes concerning raw data. The knowledge must keep new, not obvious, and one must be able in imitation of uses it. Data mining has been defined as “the nontrivial extraction of earlier unknown, implicit and potentially useful information from data. It is “the science of extracting useful data from large databases”. It is some about the tasks into the procedure on

knowledge discovery from the database. Data Mining is used to discover knowledge out about data and presenting it into a form as is easily understood by humans. It is a method according to examine huge quantities about data routinely collected. Data mining is almost useful of an exploratory analysis because regarding nontrivial information of huge volumes regarding data. It is a cooperative effort of humans and computers. Best results are achieved through balancing the knowledge on human specialists of describing problems and goals with the search capabilities concerning computers. There are two primary goals concerning data mining tend to keep prediction and description. Prediction involves some variables or fields among the data set to predict unknown or future values regarding

other variables about interest. On the other hand, Description focuses on finding patterns describing the data that can be interpreted via

humans. The Disease Prediction plays an important role in data mining.

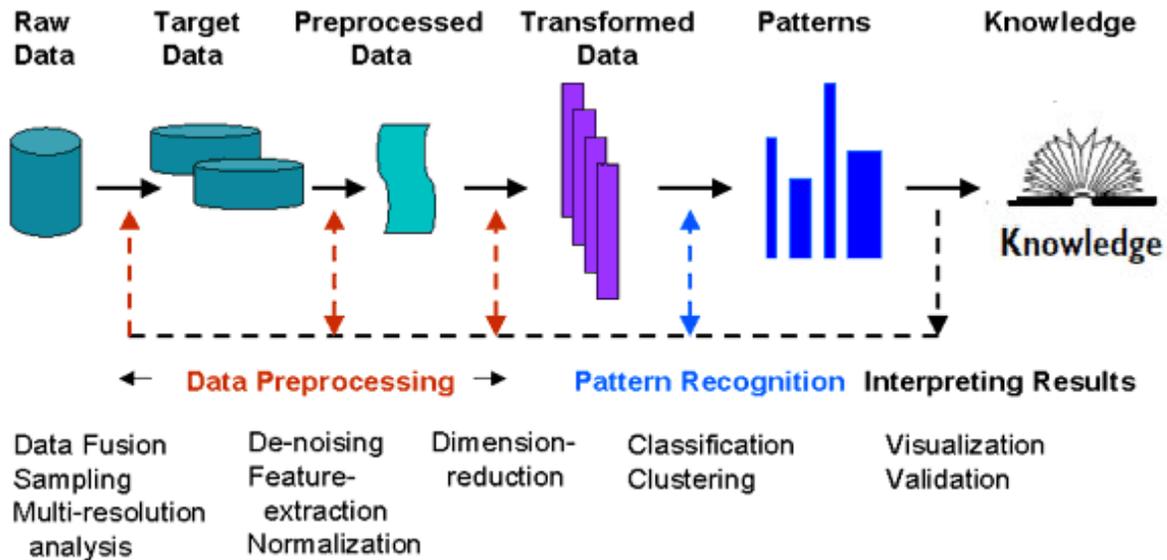


Fig1: DATA MINING KDD PROCESS

Heart disease is the leading cause of death in the world over the past 10 years. The World Health Organization referred to that heart disease is the first lead purpose on death within high and low income countries [1]. The European Public Health Alliance reported that heart attacks and other circulatory diseases account because of 41% of all deaths [2]. The Economic and Social Commission over Asia or the Pacific reported so much in one fifth regarding Asian countries, most lives are lost according to non-communicable diseases such as cardiovascular, cancers, and diabetes diseases [3]. The Australian Bureau of Statistics reported to that amount heart and circulatory system diseases are the first leading cause concerning death in Australia, causing 33.7% all deaths [4].

Motivated by the world-wide increasing mortality of heart disease patients each year and the presence of large amount of patients' data that could keep used after extract useful knowledge, researchers have been the use of data mining methods to help healthcare specialists in the analysis regarding heart disease [5],[6]. Data mining is an essential step of knowledge discovery. It is the exploration of huge datasets to extract hidden and before unknown patterns, relationships and knowledge to that amount are difficult according to keep detected with traditional statistical methods [7],[11]. The application about data mining is rapidly spreading into a wide range of sectors such as much analysis concerning organic compounds, financial forecasting, healthcare and weather forecasting. Data mining into

healthcare is an emerging field on high importance for providing prognosis and a deeper understanding on medical data. Healthcare data mining attempts to resolve real world health problems among analysis or treatment of diseases [13]. Researchers are the usage of data mining methods between the medical diagnosis on a number of diseases such as diabetes [14], stroke [15], cancer [16], and heart disease [17]. Several data mining strategies are used in the diagnosis about heart disease showing different levels on accuracy.

II. Naïve Bayes:

This algorithm is derived from the Bayesian theory that is based totally on the probability of occurrence or non-occurrence about a phenomenon for classification. Based on the

III. WEKA

The WEKA ("Waikato Environment for Knowledge Analysis") tool is used for Data mining. Data mining finds valuable information hidden in huge volumes of data. Weka is a collection of machine learning algorithms because of data mining tasks, written in Java and such contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. [18] The key features regarding Weka are it is open source and platform independent. It provides many different algorithms for data mining and machine learning [12]. We have used Naïve Bayes method to perform the mining and classification process. We have used 10 folds cross validation to minimize any bias in the process and improve the affectivity of the process.

intrinsic properties concerning the possibility (especially Share probability), Naive Bayes will provide excellent results including received initial training. In Naive Bayes, the approach concerning study is a kind of learning by using observer and controller. In fact, this method can keep formed models because of classification and prediction on several purposes. Also Naive Bayes is used according to solving problems, consciousness and classification of data among the different categories of data, and as soon as possible this method can form the model so lead to issues such as like classification together with 2 or more classes. However, with design issues and assumptions about the Naive Bayes, this technique is more suitable for classification about more issues in the actual world [8, 9, 10].

IV. HEART DISEASE DATASETS

The data used in this study is the benchmark Cleveland Clinic Foundation Heart disease data set available at <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>. The data set has 76 raw attributes. However, all of the published experiments only refer to 13 of them. The data set contains 303 rows of which 297 are complete. Six rows contain missing values and they are removed from the experiment.

Heart Disease Attribute Information: Only 14 Used

| S.No | Attribute Information |
|------|-----------------------|
| 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) (the predicted attribute) |

Heart disease Class Distribution

| Database | 0 | 1 | 2 | 3 | 4 | Total |
|---------------|-----|----|----|----|----|-------|
| Cleveland | 164 | 55 | 36 | 35 | 13 | 303 |
| Hungarian | 188 | 37 | 26 | 28 | 15 | 294 |
| Switzerland | 8 | 48 | 32 | 30 | 5 | 123 |
| Long Beach VA | 51 | 56 | 41 | 42 | 10 | 200 |

Heart Disease Database Of Instance

| Database | #of instances |
|---------------|---------------|
| Cleveland | 303 |
| Hungarian | 294 |
| Switzerland | 123 |
| Long Beach VA | 200 |

V.HEART DISEASE DATASET ANALYSIS

The distribution of the attribute values with respect to the class attribute num

The blue colored regions in the graphs in Figure 2 denote high cholesterol values. From the graphs we can see that, most of the diabetic patients with high cholesterol values are in the age group of 29 – 77, have a trestbps in the range of 94 – 200, have Chol value of 126 to 564, have a value in the range of 102 – 135, have a PP value in the range of 88 – 107

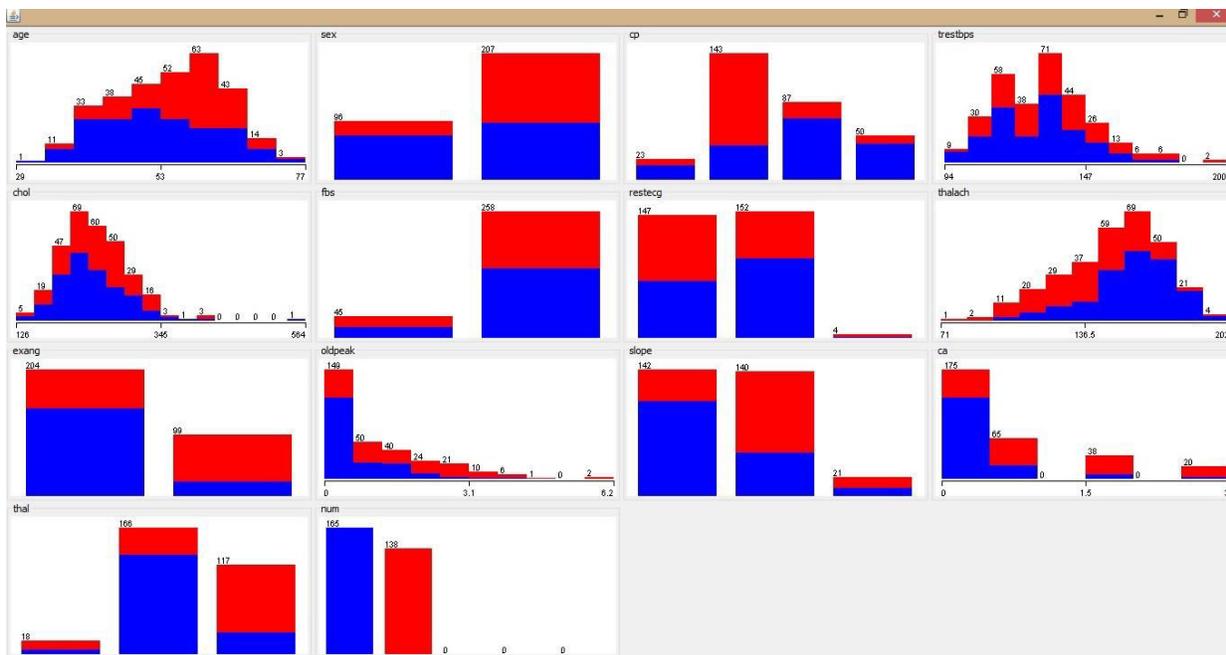


Fig2: Heart Disease Attribute Value Distributions With Respect Class Attribute

VI.RESULTS

The results of our experimentation are shown in Figure 3. The proposed naïve bayes model was able to classify 83.7% of the input instances correctly. It exhibited a precision of 83.7% in average, recall of 83.7% in average, and F-

measure of 83.7% in average. The results show clearly that the proposed method performs well compared to other similar methods in the literature, taking into the fact that the attributes taken for analysis are not direct indicators of heart disease.

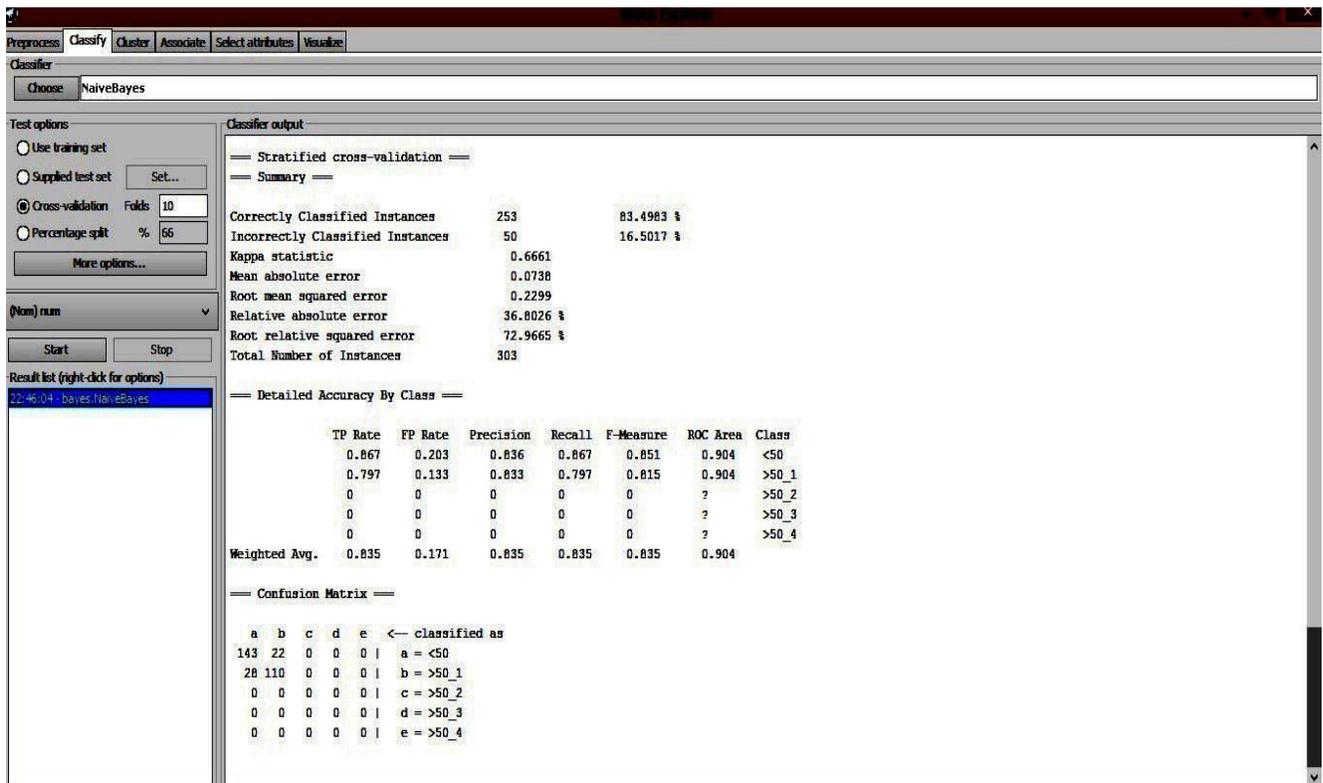


Fig 3: WEKA Naïve Bayes Num Result Window

VII.CONCLUSIONS AND FUTURE ENHANCEMENTS

Heart disease is the leading cause of death all over the world in the past ten years. Motivated by using the world-wide growing mortality concerning heart disease patients each year and the availability of large amount about patients' data up to expectation could be used to extract useful knowledge, researchers have been the use of data mining methods to help healthcare

specialists in the diagnosis over heart disease. Application about Data mining within analyzing the medical data is a good approach for considering the existing relationships within variables. From our proposed method we have shown that mining helps in conformity with retrieve useful correlation also from attributes which are not direct indicators about the category we are trying according to predict. In our work, we have tried in imitation of predict the chances concerning getting a heart disease

using attributes from diabetics diagnosis. This can remain extended according to predict other types over ailments as arise from diabetes, such namely visual impairment in future. Further,

VIII. REFERENCES

- [1] World Health Organization. (July 2007-February 2011). [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs310.pdf>.
- [2] European Public Health Alliance. (July 2010-February 2011). [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs310.pdf>.
- [3] ESCAP. (July 2017-February 2011). [Online]. Available: <http://www.unescap.org/stat/data/syb2009/9.Health-risks-causes-of-death.asp>.
- [4] Australian Bureau of Statistics. (July 2017-February 2011). [Online]. Available: [http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/E8510D1C8DC1AE1CCA2576F600139288/\\$File/33030_2008.pdf](http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/E8510D1C8DC1AE1CCA2576F600139288/$File/33030_2008.pdf)
- [5] C. Helma, E. Gottmann, and S. Kramer, "Knowledge discovery and data mining in toxicology," *Statistical Methods in Medical Research*, 2000.
- [6] V. Podgorelec *et al.*, "Decision Trees: An Overview and Their Use in Medicine," *Journal of Medical Systems*, vol. 26, 2002.
- [7] J. Han and M. Kamber, "Data Mining Concepts and Techniques,"
- the data analysis results can be used for further research in enhancing the accuracy of the prediction system in future.
- Morgan Kaufmann Publishers*, 2006.
- [8] R. Alizadehsani, J. Habibi, M.J. Hosseini, R. Boghrati, A. Ghandeharioun, B. Bahadorian, Z. AlizadehSani, "Diagnosis of Coronary Artery Disease Using Data Mining Techniques Based on Symptoms and ECG Features", *European Journal of Scientific Research*, Vol.82, No.4, pp.542-553, 2012.
- [9] G. Parthiban, A. Rajesh, S.K.Srivatsa, "Diagnosis of Heart Disease for Diabetic Patients using Naive Bayes Method", *International Journal of Computer Applications*, Vol. 24, No.3, pp. 7–11, June 2011.
- [10] R. Alizadehsani, J. Habibi, Z. Alizadeh Sani, H. Mashayekhi, R. Boghrati, A. Ghandeharioun, B. Bahadorian, "Diagnosis of Coronary Artery Disease Using Data mining based on Lab Data and Echo Features", *Journal of Medical and Bioengineering*, 2012.
- [11] B. Thuraisingham, "A Primer for Unders
- [12] D. Ashby and A. Smith, *The Best Medicine? Plus Magazine - Living Mathematics*, 2005.
- [12] An Introduction to the WEKA Data mining system
<http://www.cs.ccsu.edu/~markov/weka-tutorial.pdf>



[13] S. C. Liao and I. N. Lee, "Appropriate medical data categorization for data mining classification techniques," *MED. INFORM*, vol. 27, no. 1, pp. 59–67, 2002.

[14] T. Porter and B. Green, "Identifying Diabetic Patients: A Data Mining Approach," *Americas Conference on Information Systems*, 2009.

[15] S. Panzarasa et al, "Data mining techniques for analyzing stroke care processes," in *Proc. of the 13th World Congress on Medical Informatics*, 2010.

[16] L. Li, T. H., Z. Wu, J. Gong , M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," *Artificial Intelligence in Medicine*, Elsevier, 2004.

[17] R. Das, I. Turkoglu, and A. Sengur, *Effective diagnosis of heart disease through neural networks ensembles*, Expert Systems with Applications, Elsevier, pp. 7675–7680, 2009.

[18] Weka Data mining software
<http://www.cs.waikato.ac.nz/ml/weka>