ISSN (e): 2250 – 3005 || Volume, 08 || Issue, 6|| Jun – 2018 ||
International Journal of Computational Engineering Research (IJCER)

Analysis of Classification Algorithms on Heart Diseases Data using Association Rule Mining

KolluruVenkata Nagendra, Dr. Maligela Ussenaiah

Research Scholar, Department of Computer Science Vikrama Simhapuri University. Nellore, Andhra Pradesh, India

Assistant Professor,Department of Computer Science,Vikrama Simhapuri University. Nellore, Andhra Pradesh, India

Corresponding Auther: Kolluruvenkata Nagendra

ABSTRACT:

Association Rule Mining (ARM) is a promising technique to give experiences to better administration of incessant illnesses. In any case, ARM tends to give a staggering number of principles, prompting the long-standing issue of distinguishing the 'fascinating' guidelines for information revelation. Classification Association Rules (CARs) demonstrative of the improvement of Cardio Vascular Diseases (CVD) are produced from preparing information and bunched in light of shared characteristic of cases fulfilling the govern predecessors. The execution effect of with all components is exhibited on different classification calculations, for example, Neural Network (NN), Support Vector Machines (SVM), eXtreme Gradient Boosting (XGBoost) and Random Forest (RF). We examined the computational time and measurable measurements with Accuracy and Recall. The affiliation rules are likewise discovered the better execution of a calculation. The trial comes about show that XGBoost calculation perform superior to anything the rest of the calculations in the medicinal informational collection.

KEYWORDS: Association rule mining; chronic disease management; Classification, Random Forest

---

Date of Submission: 15-06-2018 Date of acceptance: 30-06-2018

I. INTRODUCTION

Data Mining refers to using a variety of techniques to identify information or decision making knowledge from the data base and extracting the information in a way that they can be put to use in areas such as Decision Support System, Prediction, Forecasting and Estimation. The data mining system self-learn from the previous history. When valuable knowledge is discovered, it can be helpful to manage and make good decisions. It is also one of the steps in Knowledge Discovery in Databases (KDD) process which is concerned with the algorithm means by which patterns or structures are enumerated from the data under acceptable computational efficiency limitations. Data mining research involves two fundamental goals Prediction and Description. Prediction makes use of existing variables in the data base in order to predict future values. Description focuses on finding patterns describing the data and subsequent presentation for user interpretation. Class/Concept Description Association analysis, Classification, Clustering are the most important functionalities of Data Mining. The present work focuses on classification.

The major importance given to the Data Mining are Research and surveys, Information collection, Customer opinions, Data scanning, extraction of information, preprocessing of data, web data, Competitor analysis and online research, news and Updating data.

Data mining can be used for product research, surveys, market research, and analysis. Information can be gathered that is quite useful in driving new marketing campaigns and promotions. Through the web scraping process, it is possible to collect information regarding investors, investments, and funds by scraping through related websites and databases. Customer views and suggestions play an important role in the way a company operates. The information can be readily being found on forums, blogs and other resources where customers freely provide their views. Data collected and stored will not be important unless scanned. Scanning is important to identify patterns and similarities contained in the data. This is the processing of identifying the useful patterns in data that can be used in decision-making process. This is so because decision making must be based on sound information and facts.

Usually, the data collected is stored in the data warehouse. This data needs to be pre-processed.by pre-processing it means some data that may be deemed unimportant may therefore removed manually be data mining experts. Web data usually poses many challenges in mining. This is so because of its nature. For instance, web data can be deemed as dynamic meaning it keeps changing from time to time. Therefore it means the process of data mining should be repeated at regular intervals. There is a need to understand how the competitors are fairing on in the business market. The methods of marketing and distribution can be mined. The internet is highly regarded for its huge information. It is evident that it is the largest source of information. It is possible to gather a lot of information regarding different companies, customers, and business clients. It is possible to detect frauds through online means. Nowadays with almost all major newspapers and news sources posting their news online, it is possible to gather information regarding trends and other critical areas. In this way, it is possible to be in the better position of competing in the market. This is quite important. Data collected will be useless unless it is updated. This is to ensure that the information is relevant so as to make decisions from it.

Classification is a data mining technique that assigns categories to a collection of data in order to aid in more accurate predictions and analysis. Classification is one of the several methods intended to make analysis of very large data sets effectively. It is used to find out in which group each data instance is related within a given dataset. It is used for classifying data into different classes according to some constraints. Artificial Neural Network (ANN), Bayesian Networks (BN), Decision Tree (DT), Nearest Neighbor (NN), Support Vector Machine (SVM), Rough Sets, Fuzzy Logic, Genetic Algorithms are different classification techniques for discovering knowledge. The goal of classification is to accurately predict the target class for each case in the data. The major issue is preparing the data for Classification involves the Data cleaning, Relevance Analysis, Data Transformation and reduction, Normalization and Generalization activities.

Data cleaning involves removing the noise and treatment of missing values. The noise is removed by applying smoothing techniques and the problem of missing values is solved by replacing a missing value with most commonly occurring value for that attribute. Database may also have the irrelevant attributes. Correlation analysis is used to know whether any two given attributes are related. The data can be transformed by any of the following methods. The data is transformed using normalization. Normalization involves scaling all values for given attribute in order to make them fall within a small specified range. Normalization is used when in the learning step, the neural networks or the methods involving measurements are used. The data can also be transformed by generalizing it to the higher concept. For this purpose we can use the concept hierarchies.

Association Rule Mining is presented in below section. Section 3 described classification algorithm models. Performance Metric and data set description discussed in section 4 and 5 respectively. Section 6 describes the experimental results. Finally the last section concludes this work.

II. ASSOCIATION RULE MINING

Association rule mining finds interesting association or correlation relationship among a large set of data items. With huge volume of data collected and stored, many industries are becoming interested in mining association rules from their data bases. The discovery of interesting association relationship among huge amounts of business transaction records can help in many business decision making process, such as catalog design, cross marketing, and loss leader analysis. The best application of association rule is Market Basket Analysis. The different association rules mining algorithm are Apriori Algorithm (AA), Partition, Dynamic Hashing and Pruning (DHP), Dynamic Item set Counting (DIC), FP Growth (FPG) etc.

i).Classification Association Rule Mining (CARM): Classification Association Rule Mining (CARM) is a Classification Rule Mining approach by means of ARM. CARM mines a set of Classification Association Rules (CARs) from a classified transaction database, where each CAR describes an implicative (although not necessarily causative) relationship between a set of data attributes and a pre-defined class. The following are the different association rules on Description of Framingham CHD dataset

1. If age <= 48 AND prevalentHyp <= 0 AND cigsPerDay <= 9 AND gender <= 0 AND education <= 3 AND age <= 39 then : N
2. If age <= 48 AND prevalentHyp <= 0 AND currentSmoker <= 0 then : N
3. If glucose <= 121 AND sysBP <= 144 AND age <= 47 AND education > 2 AND currentSmoker > 0 AND cigsPerDay <= 35 AND prevalentHyp <= 0 AND gender <= 0 then : N
4. If glucose <= 121 AND age <= 55 AND diaBP <= 111 AND gender <= 0 AND education > 1 AND currentSmoker <= 0 AND BPMeds <= 0 AND education <= 3 AND education > 2 AND prevalentHyp <= 0 then : N
5. If glucose <= 121 AND age <= 46 AND currentSmoker <= 0 AND BPMeds <= 0 AND gender <= 0 then : N
6. If glucose <= 121 AND sysBP <= 144.5 AND BPMeds <= 0 AND age <= 47 AND currentSmoker <= 0 AD education > 2 then : N

7. If glucose <= 121 AND sysBP <= 144.5 AND BPMeds <= 0 AND age <= 50 AND glucose > 65 AND currentSmoker > 0 then: N
8. if glucose <= 121 AND sysBP <= 155 AND age <= 57 AND BPMeds <= 0 AND currentSmoker <= 0 AND gender <= 0 AND education > 1 AND age > 50 AND education <= 3 then: N
9. if glucose <= 122 AND prevalentHyp <= 0 AND gender <= 0 AND currentSmoker <= 0 AND heartRate > 67 AND glucose > 63 then: N
10. if glucose <= 122 AND sysBP <= 166 AND age <= 57 AND BPMeds <= 0 AND currentSmoker <= 0 AND sysBP > 112.5 AND prevalentHyp <= 0 AND gender > 0 then: N
11. if glucose <= 122 AND diaBP <= 111 AND age <= 55 AND gender <= 0 AND prevalentHyp > 0 AND sysBP > 139.5 AND BPMeds <= 0 AND glucose <= 86 AND currentSmoker <= 0 AND education <= 3 then: N
12. if glucose <= 202 AND diaBP <= 114 AND age <= 46 AND BPMeds <= 0 AND currentSmoker <= 0 AND education <= 2 AND education > 1 then: N
13. if glucose <= 202 AND diaBP <= 114 AND age <= 46 AND BPMeds > 0 then: N
14. if glucose <= 202 AND diaBP <= 111 AND sysBP <= 113 AND cigsPerDay <= 20 AND diaBP > 62.5 AND currentSmoker <= 0 AND gender > 0 then: N
15. if glucose <= 202 AND diaBP <= 111 AND age <= 46 AND currentSmoker <= 0 AND education > 2: then N

16. if glucose <= 202 AND diaBP <= 111 AND age <= 62 AND gender <= 0 AND BPMeds <= 0 AND prevalentHyp <= 0 AND currentSmoker > 0 AND age <= 58 AND education <= 2 then: N
17. if glucose <= 122 AND sysBP <= 166 AND gender <= 0 AND sysBP > 161 then: N
18. if glucose <= 122 AND sysBP <= 160 AND age <= 64 AND BPMeds <= 0 AND education <= 3 AND education > 1 AND prevalentHyp > 0 AND heartRate > 67 AND currentSmoker <= 0 AND diabetes <= 0 AND sysBP > 132.5 then: N
19. if sysBP <= 155 AND BPMeds <= 0 AND age <= 46 AND currentSmoker > 0 AND gender > 0 AND diabetes <= 0 AND prevalentHyp <= 0 then: N
20. if glucose <= 122 AND sysBP <= 160 AND BPMeds <= 0 AND currentSmoker > 0 AND gender <= 0 AND education <= 3 AND education <= 2 AND prevalentHyp <= 0 AND diaBP > 71.5 then: N
21. if glucose <= 122 AND sysBP <= 160 AND BPMeds <= 0 AND age <= 57 AND gender <= 0 AND education <= 2 AND BMI <= 36.29 AND glucose <= 72 then: N
22. if diabetes <= 0 AND sysBP <= 143.5 AND currentSmoker > 0 AND prevalentHyp <= 0 AND gender <= 0 AND education <= 3 AND education > 2 AND cigsPerDay > 12 then: N
23. if iabetes <= 0 AND sysBP <= 160 AND currentSmoker > 0 AND gender > 0 then: N
24. if cigsPerDay <= 20 AND diabetes <= 0 AND gender <= 0 AND currentSmoker > 0 AND education <= 3 AND age <= 51 AND BPMeds <= 0 AND sysBP > 143 then: N
25. if cigsPerDay <= 12 AND diabetes <= 0 AND currentSmoker > 0 AND education > 3 then: N
26. if currentSmoker <= 0 AND education <= 3 AND diabetes <= 0 AND gender > 0 AND glucose > 65 AND prevalentHyp > 0 AND BPMeds <= 0 then: N
27. if currentSmoker <= 0 AND glucose <= 69 AND gender > 0 then: N
28. if gender <= 0 AND prevalentHyp <= 0 AND diabetes <= 0 AND currentSmoker <= 0 AND age <= 65 AND education > 1 AND sysBP <= 123.5 then: N
29. if gender <= 0 AND education <= 2 AND prevalentHyp <= 0 AND diabetes <= 0 AND totChol > 195 then: N
30. if gender <= 0 AND prevalentHyp > 0 AND glucose <= 117 AND BPMeds > 0 AND prevalentStroke <= 0 AND cigsPerDay <= 15 AND heartRate <= 92 AND education > 1 then: N
31. if gender <= 0 AND education <= 1 AND currentSmoker <= 0 AND diabetes > 0 AND age > 61 then: N
32. if gender <= 0 AND prevalentHyp > 0 AND glucose > 123 ANDeducation <= 1 AND currentSmoker <= 0 AND totChol <= 279 then: Y
33. if gender <= 0 AND prevalentHyp > 0 AND education <= 1 AND currentSmoker <= 0 AND diaBP <= 76 AND heartRate <= 81 then: Y
34. if gender <= 0 AND education <= 1 AND currentSmoker <= 0 AND BPMeds <= 0 then: N
35. if gender <= 0 AND currentSmoker > 0 AND education > 3: N
36. if gender <= 0 AND currentSmoker > 0 AND BPMeds > 0 AND diabetes <= 0 AND education > 1 then : N
37. if prevalentHyp <= 0 AND age > 59 AND sysBP > 128.5 AND sysBP <= 140.5 then: N
38. if gender <= 0 AND prevalentHyp > 0 AND glucose > 117 then: Y

39. if gender <= 0 AND prevalentHyp > 0 AND currentSmoker > 0 AND cigsPerDay <= 12 AND BPMeds <= 0 AND diabetes <= 0 AND heartRate > 79 then: N

40. If gender <= 0 AND prevalentHyp > 0 AND heartRate <= 64 AND age > 55 then: N
41. If diabetes > 0 AND gender <= 0 then: N
42. If heartRate > 99 AND diaBP > 84.5 AND age <= 56 then: N
43. If currentSmoker > 0 AND age <= 43 AND gender > 0 then: N
44. If currentSmoker > 0 AND BMI <= 32.8 AND gender > 0 AND heartRate > 84 AND glucose > 71 then: Y
45. If currentSmoker > 0 AND BMI <= 32.8 AND education > 1 AND diabetes <= 0 AND prevalentHyp > 0 AND diaBP > 88 then: Y
46. If prevalentHyp <= 0 AND diabetes <= 0 AND gender <= 0 AND education > 2 AND education <= 3 AND currentSmoker > 0 AND glucose <= 70 then: N
47. if prevalentHyp <= 0 AND gender <= 0 AND education <= 3 AND age > 54 then: Y
48. if gender > 0 AND BPMeds <= 0 AND currentSmoker <= 0 AND diabetes <= 0 AND prevalentHyp > 0 AND glucose <= 89 AND diaBP > 92.5 then : Y
49. if diabetes > 0 AND prevalentHyp > 0 AND glucose > 84 AND totChol > 229 then: Y
50. if diabetes > 0 AND age <= 64 AND sysBP > 127.5 then : N
51. if glucose > 112 then: Y
52. if prevalentHyp > 0 AND BPMeds <= 0 AND BMI > 32.73 then: Y
53. if currentSmoker > 0 AND age <= 62 AND BMI > 23.43 AND BPMeds <= 0 AND diaBP <= 109 then: N
54. if currentSmoker > 0 AND age <= 61 AND BPMeds <= 0 AND cigsPerDay > 12 then: Y
55. if currentSmoker > 0 AND age > 61 then : N
56. if currentSmoker > 0 AND totChol <= 245 then: N
57. if BPMeds > 0 AND currentSmoker <= 0 AND education > 1then : Y
58. if prevalentHyp <= 0 AND totChol > 245 AND age <= 61 then : Y
59. if gender > 0 AND prevalentHyp > 0 AND totChol <= 306 then: N
60. if prevalentHyp <= 0 then: N
61. if BPMeds > 0 AND prevalentStroke <= 0 AND currentSmoker <= 0 AND totChol <= 281 then: Y
62. if BPMeds > 0 AND currentSmoker <= 0 AND age > 57 then: N
63. if gender <= 0 AND currentSmoker <= 0 then : N

III. CLASSIFICATION ALGORITHMS

The present work focuses on the following classification algorithms

- Artificial Neural Network
- Support Vector Machines
- EXtreme Gradient Boost
- Random Forest

a).Artificial Neural Network (ANN)

Unlike human brain the Artificial Neural Network (ANN) has heuristic knowledge. The main characteristic of such a computing system is the number of highly interconnected processing elements (neurons) working together to solve specific problems without being programmed with step-by-step instructions. Instead, ANN's are capable of learning on their own or by example through a learning process that involves adjustments to the connections that exist between the neurons. Artificial Neural Networks (ANNs) do not require restrictive assumptions and its parallel processing capability they work well on large size training samples. ANN has detected complex nonlinear relationships between dependent and independent variables and also traditional methods works on linear as well as non linear data. Due to these reasons many researchers often use ANN for Heart diseases prediction.

b).Support Vector Machines (SVM)

In recent years, Support Vector Machines (SVM) with linear or nonlinear kernels have become one of the most promising learning algorithms for classification as well as for regression which are two fundamental tasks in data mining via the use of kernel mapping, Variants of SVMs have successfully incorporated effective and flexible nonlinear models Kernel-based techniques (like support vector machines, kernel principal component analysis, Bayes point machines, and Gaussian processes) represent a major development in machine learning algorithms. SVM (support vector machines)is a group of supervised learning techniques or methods, which is used to do for classification or regression. SVM (support vector machines) represents an extension to nonlinear models of the generalized portrait algorithm. The advantages of SVM are Provides a solid description of the learned model, Can be used for forecasting and classification, Extremely precise and ability to model complex nonlinear decision boundaries.

c).Extreme Gradient Boost (XGBoost)

XGBoost (Extreme Gradient Boosting) is an advanced and more efficient implementation of Gradient Boosting Algorithm .Extreme Gradient Boosting (XGBoost) is a supervised classification algorithm and it is very popular in various data science applications. The term "gradient boosting" come from "greedy function approximation: A gradient boosting machine". It supports various objective functions, linear models, tree learning algorithms and ranking. The big prosperity and popularity of XGBoost is its scalability on a single machine by executing parallel computations which allow quicker model exploration. In this XGBoost linear model is used in our experiments. It is 10 times faster than the normal Gradient Boosting as it implements parallel processing. It is highly flexible as users can define custom optimization objectives and evaluation criteria, has an inbuilt mechanism to handle missing values. Unlike gradient boosting which stops splitting a node as soon as it encounters a negative loss, XG Boost splits up to the maximum depth specified and prunes the tree backward and removes splits beyond which there is an only negative loss?

d).**Random Forest (RF)**

In recent times, Random Forest has gained a lot of importance as more data science problems are in place. Random forest is similar to "decision tree" models but have multiple decision tree constructs in training set. The drawback of single decision tree is to over-fitting training examples due to highly irrelevant patterns and the tree grows very deep. It leads to low bias, but high variance. Random Forest (RF) or random decision forest is an ensemble method of classification and regression. It is a supervised learning algorithm. It constructs several decision trees on training examples and output the mean predictions of all class labels. It reduces the variance error. The RF splits the training set randomly with replacement and fit the trees by averaging multiple decision trees or majority vote. The forest converges when the limit of trees in the forest becomes large. By default, RF finds the importance of variables in both classification and regression problems.

IV. PERFORMANCE METRICS FOR MODEL

To evaluate the performance of a model, we use diff erent metrics are computed from confusion matrix. Where TP - True Positive, FP - False Positive, TN - True Negative and FN - False Negative.

$$\text{Accuracy} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

V. DESCRIPTION OF FRAMINGHAM CHD DATASET

Framingham CHD dataset comprises of diff erent patients qualities as exhibited in Table 1. The ascribes F1 to F15 are illustrative factors of multi decade perception of each patient and F16 utilized as class mark. The dataset is gathered from the Boston University from clinical trials. The dataset comprises of 4,240 patients. Amid the pre-process stage, we overlooked columns if any missing qualities from dataset. After pre-preparing stage there is 14% decrease and diminished dataset comprises of aggregate of 3,658 lines and each with an arrangement of 15 traits. The dataset contains non-CHD patients are 84.77% and 15.23% of CHD patients.

Table 1: CHD Dataset Description

Test Sample	Feature	Attribute Type
F1	Gender	Discrete
F 2	Age	Continuous
F 3	Education	Discrete
F4	CurrentSmoker	Discrete
F5	CigsPerDay	Continuous
F6	BPMeds	Continuous
F7	PrevelentStroke	Discrete
F8	PrevelentHyp	Discrete
F9	Diabetes	Discrete
F10	TotChol	Continuous
F11	Glucose	Continuous
F12	DiaBP	Continuous
F13	BMI	Continuous

F14	HeartRate	Continuous
F15	SysBP	Continuous
F16	TenYearCHD	Discrete

Discreate Paremeters: A discrete variable over a particular range of real values is one for which, for any value in the range that the variable is permitted to take on, there is a positive minimum distance to the nearest other permissible value. The number of permitted values is either finite or countably infinite.
Continious Parameters: A continuous variable is one which can take on infinitely many, uncountable values.
The CHD patients are extremely in little numbers than non-CHD. The dataset has 8 continuous and 7 discrete factors is appeared in Table1. Perception is a vital undertaking in information mining system, which gives us bits of knowledge of, how the factors are appropriated before adapting any classification calculation. For the dataset, we watched couple of discrete factors i.e. Yes/No sort questions are gathered from the patients.

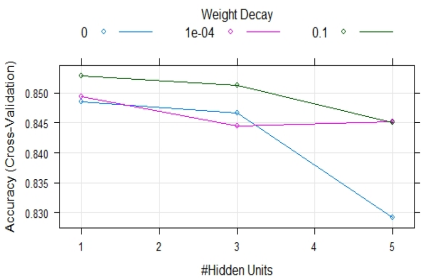
VI. EXPERIMENTAL RESULTS

In this work the positive class is no-CHD patients and negative class is CHD patients. Our principle goal of this work is to precisely recognize those patients are having CHD illness, i.e. need to diminish the false negatives, false negatives and need to manage class awkwardness issue. To begin with we talk about the outcomes With out component area. Every one of the investigations are actualized utilizing R dialect and executed on Intel i3360 4-center machine with 4GBRAM PC. To quantify the execution of classifiers straightforward we defined a moderate classification blunder rate or higher exactness i.e., the proportion of whole of genuine positive and genuine negative to add up to number of tests. Be that as it may, to quantify the strong metric for CHD forecasts, we may inspired by accuracy, i.e., the proportion of number of genuine positive to the whole of the genuine positives and false positives .
The CHD dataset is standardized and k-fold cross-validation is performed on the informational index, where k=10 in our work. We for the most part center around the assessment of the classifier with two measurements examined in Section3.1 displayed in Table2. In-terms of Accuracy, NN yielded the mean of 85.24% and running time is around 744.89 sec which is very high during the time spent for tuning. Essentially SVM and RF delivered 84.82% and 84.96% separately. XGBoost yielded 87.99% of mean Accuracy. The high accuracy is considered as the best metric to recognize the CHD patients.

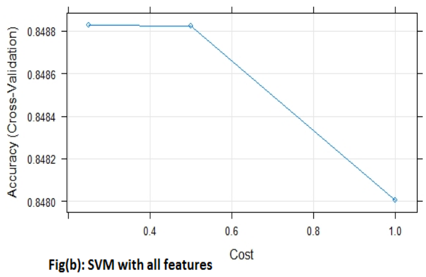
Table 2: All Features

Algorithm	Accuracy	Recall	Time (sec)
NN	85.46	99.52	744.89
SVM	84.82	100	2974.6
RF	84.96	99.65	88.03
XGBoost	**86.99**	96.19	155.9

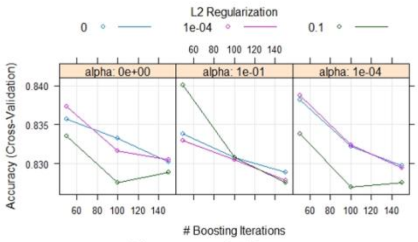
a).Resultant Graphs for Classification Algorithms: The below figures indicates the resultant Graphs of different classification algorithms. Fig(a): Neural Network algorithm with all features of CHD Dataset. Fig(b): SVM algorithm with all features of CHD Dataset, Fig(c). XGBoost algorithm with all features of CHD Dataset Fig(d): RF algorithm with all features of CHD Dataset.



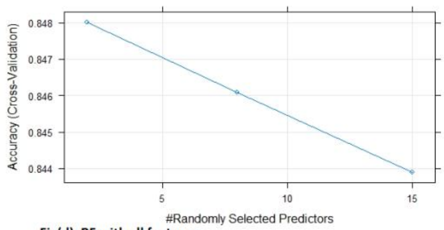
Fig(a):Neural Network with all features



Fig(b): SVM with all features



Fig(c): XGBoost with all features



Fig(d): RF with all features

VII. CONCLUSION

In this paper we display discovery of CHD utilizing machine learning forecast models. Our work was done on Framingham heart diseases dataset by employing diversified classifiers viz., NN, SVM, RF and XGboost Linear. We acquired normal exactnesses of these classifiers from 10-fold cross-validations on dataset. The exactness observed to be high in some classifiers. At last we infer that in the medicinal field and the measurements should be enhanced by additionally utilizing more patients' examples and furthermore fuse a few other hazard elements to precisely identify CHD utilizing different machine learning methods and XGBoost gives the exact outcomes contrasted with every single other technique.

REFERENCES

[1]. Shen Song, Jim Warren, Patricia Riddle, Profiling Cardiovascular Disease Event Risk through Clustering of Classification Association Rules, 2014 IEEE 27th International Symposium on Computer-Based Medical Systems

[2]. Anderson, K., Odell, P., Wilson, P. and Kannel, W. (1990): Cardiovascular disease risk profiles. American Heart Journal 121:293-8.

[3]. Collica, R.(2007): CRM Segmentation and Clustering Using SAS(R) Enterprise Miner(TM), Cary, NC: SAS Institute.

[4]. Ho, K., Pinsky, J., Kannel, W. and Levy, D. (1989): The Epidemiology of Heart Failure: The Framingham Study. Journal of the American College of Cardiology, 22(4), Suppl. 1, pp.A6–A13.

[5]. Hubert, H., Feinleib, M., McNamara, P. and Castelli, W. (1983): Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. Circulation.67(5):968-77.

[6]. Mabotuwana, T., Warren, J. and Kennelly, J. (2009): A computational framework to identify patients with poor adherence to blood pressure lowering medication. International Journal of Medical Informatics 78(11):74556.

[7]. Mannan, H., Stevenson, C. , Peeters, A. and McNeil, J., (2012): A new set of risk equations for predicting long term risk of all-cause mortality using cardiovascular risk factors. Preventive Medicine, 56(1), pp.41–45.

[8]. Mokdad, A., Ford, E., Bowman, B., Dietz, W., Vinicor, F., Bales, V. and Marks, J. (2003): Prevalence of obesity, diabetes, and obesity-related health factors, 2001. JAMA 289(1):76-9.

[9]. National Heart, Lung, and Blood Institute: Biologic Specimen and Data Repository Information Coordinating Center, https://biolincc.nhlbi.nih.gov/static/studies/teaching/fram doc.pdf, last accessed 26 Dec 2013.

[10]. New Zealand Guidelines Group (2012): New Zealand Primary Care Handbook 2012. 3rd ed. Wellington: New Zealand Guidelines Group.

[11]. Ochiai, A. (1957): Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions. Japonese Society of Scientific Fisheries, 22, pp.526–530.

[12]. Ordonez, C., Ezquerra, N. and Santana, C. (2006): Constraining and summarizing association rules in medical data. Knowledge and Information Systems, 9(3), pp.259–283.

[13]. [Perumal, L. Wells, S. and Ameratunga, S. et al., (2012): Markedly different clustering of CVD risk factors in New Zealand Indian and European people but similar risk scores (PREDICT-14). Australian and New Zealand Journal of Public Health, 36(2), pp.141–144.

[14]. Quinlan, J.R. (1993): C4.5: programs for machine learning, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

[15]. Riddle, P., Fresnedo, R. and Newman, D. (1995): Framework for a generic knowledge discovery toolkit. Preliminary papers of the 5th International Workshop on Artificial Intelligence and Statistics. Ft. Lauderdale, Florida, pp. 457–468.

[16].] Riddle, P., Segal, R. and Etzioni, O. (1994): Representation design and brute-force induction in a Boeing manufacturing domain. Applied Artificial Intelligence, 8, pp.125–147.

[17]. Riddell, T., Wells, S. and Jackson, R. et al., (2010): Performance of Framingham cardiovascular risk scores by ethnic groups in New Zealand: PREDICT CVD-10. Journal of the New Zealand Medical Association, 123(1309), pp.50–61.

[18]. Rodondi, Locatelli, N. and Aujesky, D. et al. (2012): Framingham risk score and alternatives for prediction of coronary heart disease in older adults. PLoS One, 7(3), p.e34287.

[19]. SAS Institute (2013), the data analysis for this paper was generated using SAS software, Version 9.3 of the SAS System for Windows7. Copyright 2013. SAS Institute, Inc. Cary, NC, USA.

[20]. Song, S., Warren, J. & Riddle, P.(2014): Developing High Risk Clusters for Chronic Disease Events with Classification Association Rule Mining. Proceedings, 7th Australasian Workshop on Health Informatics and Knowledge Management (HIKM), Auckland. (Conferences in Research and Practice in Information Technology (CRPIT), Vol. 153).

[21]. Tan, P.-N., Kumar, V. and Srivastava, J. (2002): Selecting the right interestingness measure for association patterns. Proc. the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. New York, USA. pp. 32-41, ACM Press.

[22]. Wang, Y.J., Xin, Q. and Coenen, F. (2007): A Novel rule ordering approach in classification association rule mining. Proc. 5th International Conference on Machine Learning and Data Mining in Pattern Recognition. Leipzig, Germany, pp. 339–348.

[23]. Webb, G. (2000): Efficient search for association rules. Proc. the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. New York, USA. pp. 99-107, ACM Press.

Biographies

K.Venkata Nagendra, working as Assistant Professor in the Department of Computer Science Engineering at Geethanjali Institute of Science and Technology, Nellore, Andhrapradesh, India. He has 7 years of experience in the field of teaching. He is a research scholar in Vikrama Simhapuri University. He did his M.Tech in ANU, Guntur. His areas of interests are Data warehousing and Data Mining and Cloud Computing.

Dr.Maligela Ussnaiah, working as Assistant Professor in the Department of Computer Science in Vikrama Simhapuri University, Nellore, Andhra Pradesh, India. He is having 9 years of teaching experience. He did his PhD in computer science from SriKrishna Devaraya Univeristy, Ananthapur, Andhra Pradesh. His areas of interests are Networks, Mobile Wireless Networks, Data warehousing and Data Mining and Image processing.

Kolluruvenkata Nagendra "Classification Algorithm Accuracy Analysis On Heart Diseases Data Using Association Rule Mining "International Journal of Computational Engineering Research (IJCER), vol. 08, no. 06, 2018, pp. 39-46.