



Heart Disease Prediction using Machine Learning Algorithms

T.Sowndhariyaa^{1*}, T.M.Nithya²

¹PG Scholar, Department of CSE, K.Ramakrishnan College of Engineering, Trichy

²Assistant Professor, Department of CSE, K.Ramakrishnan College of Engineering, Trichy

Email Id: sowndhariyaaathangarasun@gmail.com

To Cite this Article

T.Sowndhariyaa and T.M.Nithya². Heart Disease Prediction using Machine Learning Algorithms. International Journal for Modern Trends in Science and Technology 2022, 8(07), pp. 51-57. <https://doi.org/10.46501/IJMTST0807008>

Article Info

Received: 27 May 2022; Accepted: 25 June 2022; Published: 29 June 2022.

ABSTRACT

In the trade of care data wealthy square measure on the market with the structured and also the non structured knowledge. For this the information mining is required to spot the hidden pattern and for creating economical call. Within the health field the prediction of the center diseases victimization this knowledge can greatly facilitate the human to spot the diseases. To find the center diseases victimization the giving the input with the a lot of attribute are going to be analyzed during this paper. The system uses medical terms like sex, vital sign, sterol case history, Smoking , Poor diet , High vital sign , High blood sterol Physical inactivity , Obesity , Hyper tension etc like thirteen attributes to guess the chance of patient obtaining a heart condition. This analysis thesis side 2 a lot of attributes i.e. avoirdupois and smoking. The information mining classification techniques, particularly call Trees, Naive Bayes, and Support vector machine square measure analyzed on heart condition info. The show of those techniques is compared, supported accuracy. As per our results accuracy of Support Vector machine, call Trees, and Naive Bayes square measure ninety eight, 85.5%, and 65.74% severally. Our analysis shows that out of those 3 classification models support vector machine predicts heart condition with highest accuracy.

Keywords :Data Mining, Heart Diseases, Navie Bayes ,Svm, Decision Trees

1. INTRODUCTION

A method of extracting attention-grabbing data or patterns from massive databases is outlined as data processing. There square measure various techniques employed to find all reasonably data, largely they ensuing from machine learning and statistics. The superior division of those approaches focuses on the identification of correct data. Even supposing this data could also be ineffective if it didn't provide any establish any helpful data for the top user. The errands performed within the data processing square measure supported what reasonably data

somebody has to mine. Data mining data methoding technique square measure the results of a long process of study and merchandise development.

The center is critical organ of form half. Once the center stops beating, blood circulation won't sent the aerated blood to the brain and also the death can occur among couple of minutes. Life depends on healthy operating of the center. The term malady heart condition cardiopathy cardiovascular disease refers to vessel system & disease of heart among it.

A number of things are shown that will increase the chance of Heart disease: case history, Smoking, Poor

diet, High vital sign, High blood sterol, Obesity, Physical inactivity, Hyper tension, Factors like these square measure wont to examine the center unwellness. identification is predicated on patient's gift check results & doctor's data. Therefore the associatealysis is an tangled task that needs abundant understanding & high ability.

2. LITERATURE SURVEY

In [1] discuss the show analysis of classification data processing technique for cardiopathy prediction. The 3 algorithms utilized in this work area unit naïve Thomas Bayes, adult female and apriori. The performance valuation is predicated on classification matrix, it show the frequency of correct and incorrect prediction. The analyze model is read in raise charts, Bar charts and Pie charts. This struggle may be more increased and enlarged by victimisation different data processing techniques like statistic, clump and Association rules. rather than categorical information, the continual information may be used.

In [4] the fuzzy specialist system is planned for cardiopathy diagnosing with reduced variety of attribute. The author realize that however genetic rule and mathematical logic mix along for economical and price effectual diagnosing of cardiopathy. The GA and 2 models of fuzzy system Mamdani and Takagi-sugeno were wont to realize the price. The dataset were taken in Cleveland clinic institution dataset. The input field could be a set of all the bound kind and therefore the output of the system is to induce a worth '1' or '0' that indicates the presence or absence of unwellness. it's further improved by victimisation art classifiers like call tree, Naïve Thomas Bayes, Classification via clump and SVM classifier.

In [5] data processing classification technique particularly manslayer classifier, call tree, Artificial Neural Network and Support vector machine area unit used for predicting vas cardiopathy. The arrangement issue used for compare these technique area unit sensitivity, accuracy, specificity, error rate, true positive rate and false positive rate. to measure the unbiased estimate of prediction models the author used ten fold cross validation technique. This model was industrial by victimisation data processing classification tool rail version three.6. It contains fourteen attributes and 303 instances. Error rates for manslayer, Artificial Neural

Networks, Support vector machine and call tree area unit 02.756, 0.2248, 0.1588 and 0.2755 severally. The accuracy of manslayer, Artificial Neural Networks, Support Vector Machine and call tree area unit eighty one.10%, 80.07%, 84.13% and 79.15% severally. Four classification models, the Support Vector Machine has given least error rate and highest accuracy. the author complete that the Support Vector Machine is that the best technique for predicting the disorder. In future so as to boost the potency of the classification techniques by making metamodels.

In [6] heart failure symptom is foretold victimisation medical specialty data processing technique. The creator used information classification that is predicated on supervised machine learning algorithms. For information classification the Tanagra tool is employed. victimisation entropy primarily based cross validations and partitioned off techniques, the information is evaluated and therefore the results area unit compared. The algorithms utilized in these techniques area unit K-nearest neighbours, K-means and Mean clump rule (EMC) is that the extension of the K-mean rule for clump method that reduces the amount of iterations. During this paper result the author analyzed that the mean clump rule perform well once compare to different rule. To run the information the time taken is extremely quick and it offers the results of accuracy regarding eighty three.25%. Associate in Nursing increased by applying unsupervised machine learning rule.

In [8] association rule mining technique is employed for predicting heart failure. During this paper creator future a unique technique CBARBSN for association rule mining supported sequence numbers and clump the transactional info for predict heart failure. The 2 necessary step of this method initial the medical information is remodelled into binary and therefore the planned technique is applied to the binary transactional information. the information is collected from Cleveland info. The medical information contains fourteen attributes. From the results author complete that the planned rule performs higher than the present ARNBSN () rule. The performance of the algorithms is compared supported the execution time.

In [9] the arterial blood vessel unwellness was with efficiency diagnosed by rotation forest rule so as to support clinical decision-making method. It utilizes the

ANN (Artificial Neural Networks) with back propagation rule of rotation forest ensemble technique as base classifier. The rule is enforced in matlab. From Associate in Nursing experiment, the author diagnosed the unwellness by scrutiny the performance of base classifiers in terms of sensitivity, accuracy, United Self-Defense Group of Colombia and specificity on 2 things i) while not rotation forest classifier, the best performance of classifiers and ii) with rotation forest rule it truly improve the performance of classifier. Result it's experiential that Levenberg-Marquardt was the simplest classifier with or while not random forest. The accuracy is improved to ninety four.2% of original classification accuracy that is Associate in nursing improvement of 8 May 1945. In prospect the planned work could also be wont to develop economical knowledgeable systems for the diagnosing of cardiopathy.

3. METHODOLOGY

3.1Naïve Bayes

Naïve Thomas Bayes could be a supervised likelihood machine learning classifier technique that assumes terms occur severally. this will be accustomed in classifying matter documents in simplest technique. The Naïve Thomas Bayes model computes the posterior likelihood of a category, supported the allocation of words within the document this illustration works with the BOWs feature extraction that ignores true of the word within the document .Bayesian classification represent a supervised learning technique as well as applied mathematics technique for classification. it's straightforward probabilistic classifier supported theorem theorem with robust independence assumption. it's for the foremost half suited once the spatiality of input is high. they will predict the likelihood that a given tuple belongs to a specific category. The Thomas Bayes theorem.formula will be written as

$$P(\text{Hy} \mid \text{Ev}) = [P(\text{Ev} \mid \text{Hy}) * P(\text{Hy})] / P(\text{Ev})$$

The basic plan of Bayes's rule is that the end result of a hypothesis or a happening (Hy) will be foreseen supported some evidences (Ev) which will be discovered from the Bayes's rule.

3.2Support Vector Machine

1)SVM perform classification by finding the hyper plane that maximizes the margin between the 2

categories. The vectors (cases) that establish the hyper plane area unit the support vectors. In machine learning, SVMs, additionally (support vector networks) area unit supervised learning models with associated learning algorithms that analyze knowledge used for classification and multivariate analysis.

2) Classification SVM Type 1

For this kind of SVM, preparation involves the diminution of the error function:

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i$$

subject to the constraints:

$$y_i (\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, i = 1, \dots, N$$

Where C is that the capability constant, w is that the vector of coefficients, b could be a constant, and represents parameters for handling no divisible knowledge (inputs). The index i labels the N coaching cases. represents the category labels and xi represents the freelance variables. The kernel is employed to remodel knowledge from the input (independent) to the feature area. It ought to be noted that the larger the C, the additional the error is penalised. Thus, C ought to be chosen with care to avoid over fitting.

3.3 Decision Trees

The decision tree is further authoritative for categorization effort. the fundamental steps during this technique area unit to assembling a tree & applying the tree to the dataset. There area unit several fashionable call tree algorithms C4.5. From these C4.5 algorithmic rule is employed for this technique. C4.5 algorithmic rule uses pruning technique to create a tree. the tactic of reduces size of tree by removing over fitting knowledge, that ends up in poor accuracy in predication is termed as pruning. The C4.5 algorithmic rule recursively classifies knowledge till it's been classified as dead as potential. this system offers most accuracy on coaching knowledge. the same old construct is to create a tree that gives balance of flexibility & accuracy.

3.3 WEKA TOOL

Weka could be a acquainted sight system within the history of the machine learning analysis communities and data processing, as a result of it's the

sole toolkit that has gained such in depth acceptance and survived for an intensive amount of your time. It provides many numerous algorithms for data processing and machine learning. maori hen is open supply and freely on the market. it's additionally platform-independent.

Advantages

- As WEKA is absolutely enforced in java programming languages, it's platform freelance & moveable.
- It is freely on the market below wildebeest General Public License.
- Weka s/w contain terribly graphical computer programme, therefore the system is extremely straightforward to access.
- There is extremely giant assortment of various data processing algorithms.

Disadvantages

- Lack of potentialities to interface with alternative package
- Performance is commonly sacrificed in favour of movability, style transparency, etc.
- Memory limitation, as a result of the info has got to be loaded into main memory utterly.

4. EXPERIMENT AND RESULTS

In this study mainly classification algorithms such as Naive Bayes, Decision Trees and SVM algorithm is used for predicting the Heart Disease from the given data set instances and the proposed algorithms are applied on type Heart Disease dataset in the WEKA tool and the performance is measured. The heart is main organ of human body part. It is nothing more than a pump, which pump blood through the body. If circulation of blood in body is incompetent the organ like brain suffers and if heart stops working altogether, death occurs within minutes. The term Heart disease refers to blood vessel system & disease of heart within it. A number of factors have been shown that increases the risk of Heart disease: Family history, Smoking , Poor diet , High blood pressure , High blood cholesterol , Obesity , Physical inactivity , Hyper tension etc., Factors like these are used to analyze the Heart disease. In many cases, diagnosis is generally based on patient's current test results & doctor's experience. Thus the diagnosis is a complex task that requires much experience & high skill.

DATA SOURCE

The in public on the market heart condition information is employed. The Cleveland heart condition information consists of a thousand records&Statlog heart condition information consists of 970 records [12]. the info set contains three kinds of attributes: Key, Input & foreseeable attribute that area unit listed below.

Attribute Information:

Only fourteen attributes used:

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 foretold attribute)

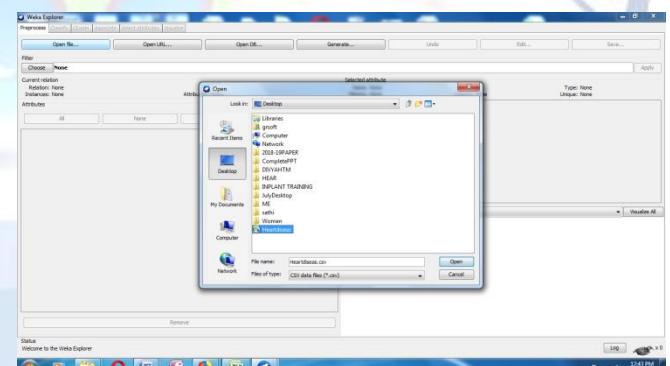


Figure 4.1.1 Upload Heart Disease Dataset

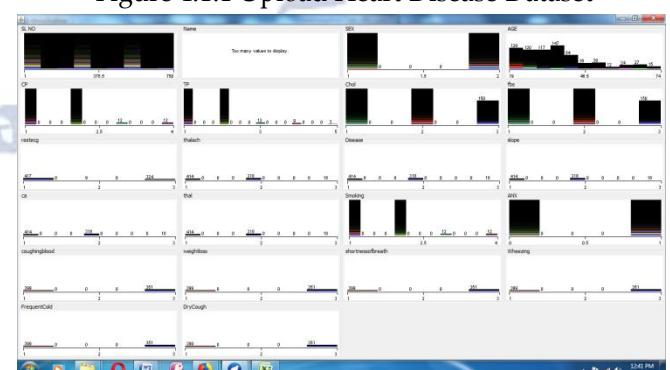


Figure 4.1.2 ViewHeart Disease Chart

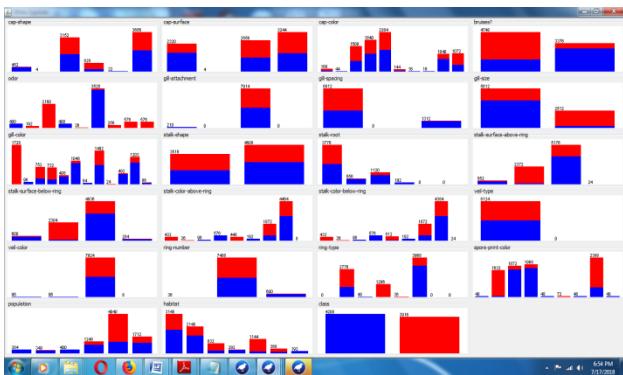
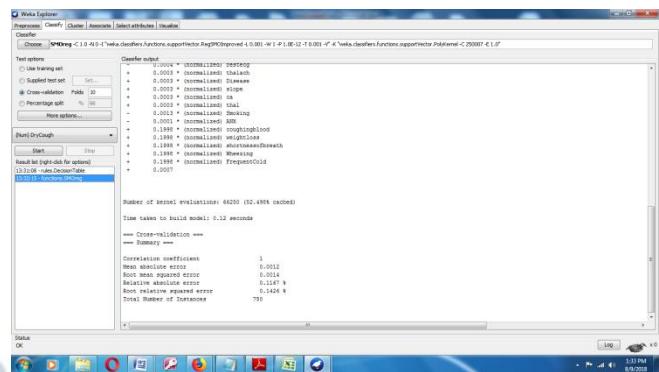


Figure 4.1.3 Over All Chart



4.1.7 SVM Heart Disease Dataset Classification techniques

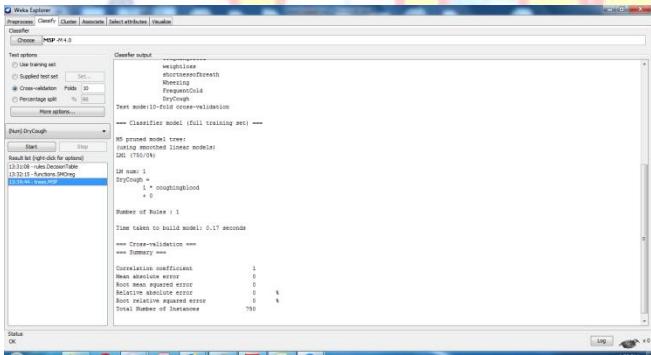
4.2 Performance Calculated Using Correlation Coefficient

Experiments are performed on the heart malady datasets. Machine with configuration of windows seven system and 2-GB of RAM is utilized. The results were compared to experiments with rail implementations of Navie scientist, decision Tree and Support Vector machine the techniques run to form certain that the results are comparable for statistic.

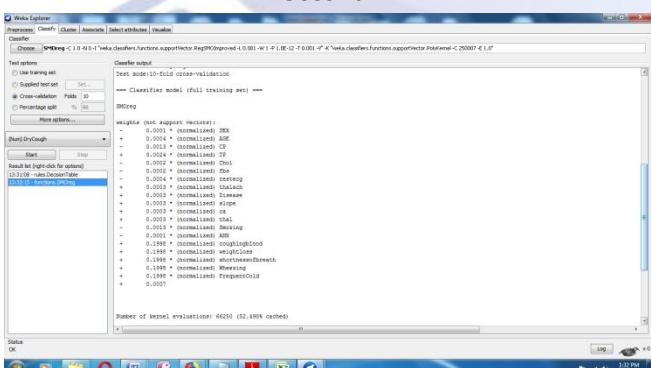
Table 4.2.1: The table shows Correlation coefficient

ALGORITHM	CORRELATION COEFFICIENT
Navie Bayes	0.9485
Decision Tree	0.9961
SVM	1

4.1.4 Decision Tree Heart Disease Dataset Execution Second



4.1.5 Navie Bayes Heart Disease Dataset Execution Second



4.1.6 SVM Heart Disease Dataset Execution Second

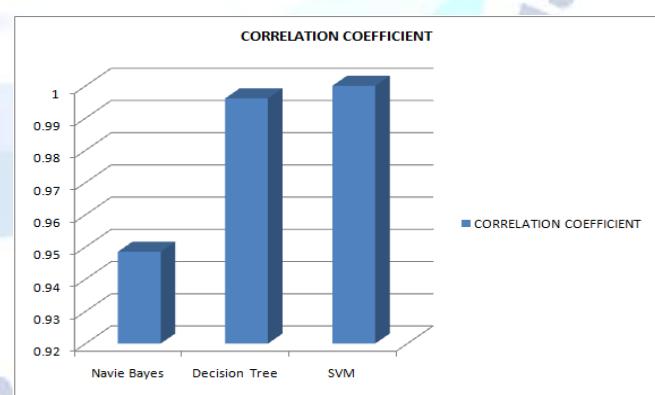


Fig 4.2.1 Evaluation measures of Correlation coefficient

4.2.2 Relative Absolute Error

Experiments square measure performed on the guts illness datasets. Machine with configuration of windows seven system and 2-GB of RAM is employed. The results were compared to experiments with wood hen

implementations of Navie mathematician, call Tree and Support Vector machine the techniques run to make sure that the results square measure comparable for Relative absolute error.

Table: 4.2.2 The table shows relative absolute error

ALGORITHM	RAE
Navie Bayes	0.9632
Decision Tree	0.9971
SVM	0.3707

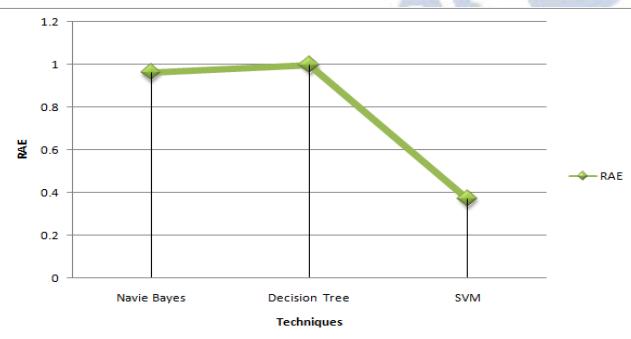


Fig: 4.2.2 Evaluation measures of Relative absolute error

4.2.3 EXECUTION TIME SECOND

Experiments square measure performed on the guts illness. Machine with configuration of windows seven system and 2-GB of RAM is employed. The results were compared to experiments with wood hen implementations of Navie mathematician, call Tree and Support Vector machine the techniques run to make sure that the results square measure comparable for Relative absolute error.

Table 4.2.3 Execution Time Second

ALGORITHMS	TIME TAKEN (IN MIL SECS)
Navie Bayes	0.03
Decision Tree	0.13
SVM	0.02

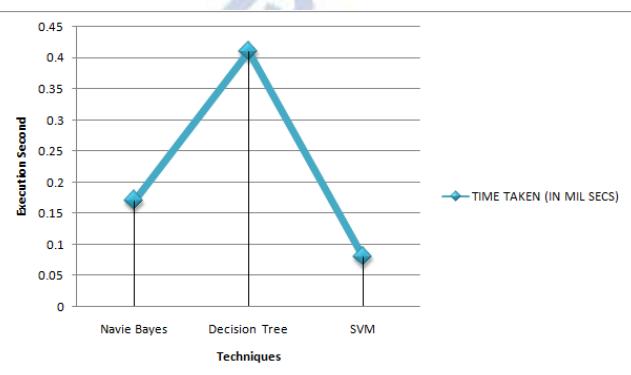


Figure: 4.2.3 Representing Dataset measured in Execution Second(ms)

5. CONCLUSION

Around eighteen million people 7-membered of the Indians ar laid low with upset. upset is typically affected sixty 5 above age group. This paper primarily focuses on three all completely different categories of heart illnessss specifically blood vessel illness, upset and heart disease. the aim of this work is to forecast exactly the presence of upset. throughout this thesis extra input attributes fat and smoking ar used to get extra correct results. three processing classification techniques were applied specifically decision trees, Naive mathematician & Support vector machine. From results it has been seen that SVM provides correct results as compare to decision trees & Naive mathematician.

Future Enhancement

As a future work the someone has planned to perform further experiments with extra dataset and algorithms to spice up the classification accuracy and to form a model which will predict specific upset varieties .In the planned formula, it's assumed that each operation of a special attributes of dataset is to be completed on a wood hen Tool. the long-standing time works may even be to look out the ways to unravel text, number,boolean setting and additionally the dynamic dataset. very high configuration machines ar used to acquire the GA at intervals the minimum amount. SVM based totally algorithms square measure found to look out the produce span of the assorted kind of dataset. it will trim the machine quality and increase the performance and quality of dataset. In the Second future work, there is a planning to vogue a refined code based totally on this technique that may targeted to use in very secure transmission data transmission applications.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Frawley and G. Piatetsky -Shapiro, Knowledge Discovery in Databases: An Overview. Published by the AAAI Press/ The MIT Press, Menlo Park, C.A 1996.
- [2] Yanwei, X.; Wang, J.; Zhao, Z.; Gao, Y., "Combination data mining models with new medical data to predict outcome of coronary heart disease". Proceedings International Conference on Convergence Information Technology 2007, pp. 868 – 872.

- [3] Sellappan Palaniappan, RafiahAwang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008
- [4] Niti Guru, Anil Dahiya, NavinRajpal, "Decision Support System for Heart Disease Diagnosis Using Neural Network", Delhi Business Review, Vol. 8, No. 1 (January - June 2007).
- [5] HeonGyu Lee, Ki Yong Noh, Keun Ho Ryu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.
- [6] ShantakumarB.Patil, Y.S.Kumaraswamy "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network". ISSN 1450-216X Vol.31 No.4 (2009), pp.642-656.
- [7] Carlos Ordonez, "Improving Heart Disease Prediction Using Constrained Association Rules," Seminar Presentation at University of Tokyo, 2004.
- [8] Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer, Vol:345, pp: 721- 727, 2006.
- [9] Franck Le Duff, CristianMunteanb, Marc Cuggiaa, Philippe Mabob, "Predicting Survival Causes After Out of Hospital Cardiac Arrest using Data Mining Method", Studies in health technology and informatics, Vol. 107, No. Pt 2, pp. 1256-9, 2004.
- [10] LathaParthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological, Biomedical and Medical Sciences, Vol. 3, No. 3, 2008.
- [11] Dr. Yashpal Singh, Alok Singh chauhan "Neural Networks in data mining" Journal of Theoretical and Applied Information Technology , 2005 - 2009 JATIT.
- [12] N. AdityaSundar, P. PushpaLatha, M. Rama Chandra, "Performance Analysis of Classification Data Mining Techniques over Heart Disease Data base" [IJESAT] international journal of engineering science & advanced technology ISSN: 2250-3676, Volume-2, Issue-3, 470 – 478
- [13] Blake, C.L., Mertz, and C.J.: "UCI Machine LearningDatabases", Cleveland heart disease dataset"
- [14] E.P. Ephzibah, "A Hybrid Genetic-Fuzzy Expert System for Effective Heart Disease Diagnosis" D.C. Wyld et al. (Eds.): ACITY 2011, CCIS 198, pp. 115–121, 2011. © Springer-Verlag Berlin Heidelberg 2011
- [15] EsraMahsereciKarabulut&Turgayİbrikçi "Effective Diagnosis of Coronary Artery Disease Using The Rotation Forest Ensemble Method" June 2011 / Accepted: 30 August 2011 / Published online: 13 September 2011 # Springer Science+Business Media, LLC 2011
- [16] V.V.Jaya Rama krishniah, D.V.ChandraSekar, Dr.K.Ramchand H Rao, "Predicting the Heart Attack Symptoms using Biomedical Data Mining Techniques" Volume 1, No. 3, May 2012 ISSN – 2278-1080 The International Journal of Computer Science & Applications (TIJCSA)
- [17] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006
- [18] S.Sarumathi, N.S.Nithya, "Effective Heart Disease Prediction System Using Frequent Feature Selection Method" International Journal of Communications and Engineering Volume 01– No.1, Issue: 01 March2012
- [19] Mrs.G.Subbalakshmi, Mr. K. Ramesh, Mr. M. ChinnaRao, "Decision Support in Heart Disease Prediction System using Naïve Bayes" Indian Journal of computer Science and Engineering (IJCSE), ISSN: 0976-5166 Vol. 2 No. 2 Apr-May 2011
- [20] R. Sumathi, E. Kirubakaran "Enhanced Weighted K-Means Clustering Based Risk Level Prediction for Coronary Heart Disease" European Journal of Scientific Research ISSN 1450-216X Vol.71 No.4 (2012), pp. 490-500 © Euro Journals Publishing, Inc. 2012
- [21] Dr. K. Usha Rani "Analysis of Heart Diseases Dataset Using Neural Network Approach" (IJDKP) Vol.1, No.5, September 2011