

Forecasting of Heart Disease using Hybrid Algorithm

M.Tech. Scholar Neha Soni

Dept. of Computer Science and Engg .
Sushila devi bansal college of engineering
Indore,MP, India
nehasoni034@gmail.com

Asst. Prof. Mrs. Namrata Bhatt

Dept. of Computer Science and Engg.
Sushila devi bansal college of engineering
Indore, MP, India
nsharma@sdbce.ac.in

Abstract - In medicinal sciences forecast of Heart sickness is most troublesome undertaking. In India, fundamental driver of Death are because of Heart Diseases. The passings because of coronary illness in numerous nations happen because of work over-burden, mental pressure and numerous different issues. It is found as fundamental reason in grown-ups is because of coronary illness. Along these lines, for distinguishing coronary illness of a patient, there emerges a need to build up a choice emotionally supportive network. Information mining order systems, to be specific Modified K-means and SVM are broke down on Heart Disease is proposed in this Paper.

Keywords- Data Mining, Heart Disease, Machine Learning, Decision Support, Modified K-Means, SVM.

I. INTRODUCTION

In this world individuals need to carry on with a rich life so they work like a machine so as to acquire part of riches. At extremely youthful age, this kind of way of life doesn't take rest for themselves, which results in diabetics and pulse. Heart is the most fundamental part in human body on the off chance that that heart gets in fluenced, at that point it likewise influences different pieces of the body. Accordingly it is basic for individuals to go for a coronary illness analysis.

Individuals go to medicinal services checkup however the expectation made by them isn't 100% precise. Today, medicinal services industry produces enormous measure of information about patients, malady analysis and so on. Finding is significant undertaking and convoluted that should be executed precisely and effectively. In light of specialist's experience and Knowledge, the conclusion is frequently made. This prompts undesirable outcomes and unnecessary therapeutic expenses of medicines Provided to patients. Nature of administration is a noteworthy test confronting Healthcare industry. Nature of administration assurance Diagnosing ailment accurately and to gives powerful medicines to patients.

II. RELATED WORK

To have center around determination of coronary illness various investigations have been finished. Various information mining strategies has been utilized by them for conclusion and accomplished various probabilities for various techniques. Utilizing information mining procedures an Intelligent Heart Disease Prediction System (IHDPS) is created. Sellappan Palaniappan et al [14] proposed Naïve Bayes, Neural Network, and Decision Trees. For suitable outcomes every strategy has its very own quality. Shrouded examples and connection between

them is utilized to construct this framework. It is easy to understand, expandable and electronic. Niti Guru et al [7] proposed the forecast of Blood Pressure, Sugar and Heart ailment with the guide of neural systems. The record of 13 properties in every wa utilized in the dataset. For preparing and testing of information, the regulated systems for example Neural Network with back proliferation calculation is utilized. Heon Gyu Lee et al. [5] proposed a novel procedure, to build up the multi-parametric element with direct and nonlinear qualities of HRV (Heart Rate Variability) a few classifiers for example Bayesian Classifiers, CMAR (Classification dependent on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine) has been utilized by them. To gauge the polluting influence of a segment or set of preparing tuples [2], CART utilizes Gini record. High dimensional clear cut information can deal with by it.

III. HEART DISEASE

The heart is significant piece of our body. Our life is absolutely subject to proficient working of heart. In the event that task of heart isn't legitimate, it will influence different pieces of body, for example, kidney, cerebrum, and so forth. It is just a siphon, which siphons blood through the body. Demise happens inside minutes, if flow of blood is wasteful. The term Heart infection alludes to vein framework inside it and malady of heart.

There are number of components which increment the danger of Heart illness [4].

- Smoking
- Family history of heart disease
- Poor diet
- High Blood Pressure
- Physical inactivity
- Hyper tension

- Obesity
- Cholesterol

Elements like these are utilized to investigate the Heart illness. Much of the time, finding is commonly founded on specialist's involvement and patient's present test outcomes. In this way the determination is a perplexing errand that requires much experience and high expertise.

III. DATA SOURCE

Dataset with info characteristics is gotten from Cleveland Heart Disease database. With the assistance of recordset, the heart assault forecasts with huge examples are removed. The characteristic "Determination" with worth "1" is distinguished as Heart Disease expectation and worth "0" is recognized as no Heart malady forecast for patients. Here key property is "PatientId" and different properties are utilized as information.

1. Predictable attribute

1. Diagnosis (value 0: <50% diameter narrowing (no heart disease); value 1: >50% diameter narrowing (has heart disease))

2. Key attribute

- Patient Id – Patient's identification number

3. Input attributes

- Sex (value 1: Male; value 0: Female)
- Age in Year
- Oldpeak – ST depression induced by exercise
- Restecg – resting electrographic results (value 0: normal; value 1: having ST-T wave abnormality; value 2: showing probable or
- definite left ventricular hypertrophy)
- Fasting Blood Sugar (value 1: >120 mg/dl; value 0: <120 mg/dl)
- Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: downsloping)
- Exang - exercise induced angina (value 1: yes; value 0: no)
- Serum Cholesterol (mg/dl)
- Trest Blood Pressure (mm Hg on admission to the hospital)
- Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
- CA – number of major vessels colored by fluoroscopy (value 0-3)
- Thalach – maximum heart rate achieved
- Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)

IV. PROPOSED ALGORITHM

Today, numerous emergency clinics oversee medicinal services information utilizing social insurance data framework; as this framework contains colossal measure of information, and it is utilized to concentrate shrouded data for therapeutic determination. The fundamental

target of this framework is to fabricate Heart Disease Prediction System utilizing verifiable heart database that gives determination of coronary illness. To fabricate this framework, restorative terms, for example, pulse, sex, cholesterol, sugar and so on 13 info characteristics are utilized. Information mining procedures, for example, bunching, Classification is utilized in removing learning from database.

1. Modified K-means- The proposed changed calculation demonstrates to be a superior technique to decide the underlying centroids and it is anything but difficult to execute. By dispensing with one of its disadvantages, this altered K-implies attempts to upgrade the k means grouping Algorithm. K-implies was utilized to apply on numerical information as it were. However, we experience both numerical and all out mix information esteems. This calculation does not require number of cluster (k) as information is depicted beneath. By picking two introductory centroids, two groups are made at first, which are most remote separated in the datasets. It can make two groups with the information individuals at the underlying advances, which are most different ones.

2. Input- D: The set of n tuples with attributes A₁, A₂, ..., A_m. All attributes are numeric, (where m = no. of attributes)

3. Output- With n tuples suitable number of clusters distributed properly

V. METHOD

- To discover the focuses in the informational collection which are most distant separated, process total of the characteristic estimations of each tuple
- As introductory centroids take tuples with greatest and least estimations of the whole.
- Using Euclidean separation make introductory segments (cluster) between the underlying centroids and each tuple
- From the centroid discover separation of each tuple in both the underlying allotments. Take other than zero. d=minimum all things considered.
-) For the allotments made in step 3, compute new methods (centroids)
- From the new methods (group focuses) register Euclidean separation of each tuple. furthermore, contingent upon the accompanying target work, discover the exceptions: If Distance of the tuple from the bunch mean > d then just it is an Outlier.
- New centroids of the groups can be processed
- From the new group centroids, compute Euclidean separation of each exception and locate the target work in stage 6. Exceptions isn't fulfilling
- Let the arrangement of anomalies got in stage 8 is B = {Y₁, Y₂, ..., Y_p} (Where estimation of k is relies upon number of exceptions).
- Repeat the means until I (B == < D)

- By taking mean estimation of its individuals as centroid, make another group for the set B,
- Depending on the target work in stage 6, discover the exceptions of this bunch,
- Check assuming no. of exceptions = p at that point
- Test each other exception for the target work as in stage 6 after making of another bunch with one of the anomalies as its part
- If there is any out fliers discover it
- From the centroid of the current cluster, compute the separation of each anomaly. On the off chance that the current groups which fulfill the target work in stage 6. At that point change the anomalies
- The new arrangement of exceptions be $B = \{Z_1, Z_2, \dots, Z_q\}$. (Where estimation of q is relies upon number of anomalies).

1. Flow chart of K- Mean proposed work

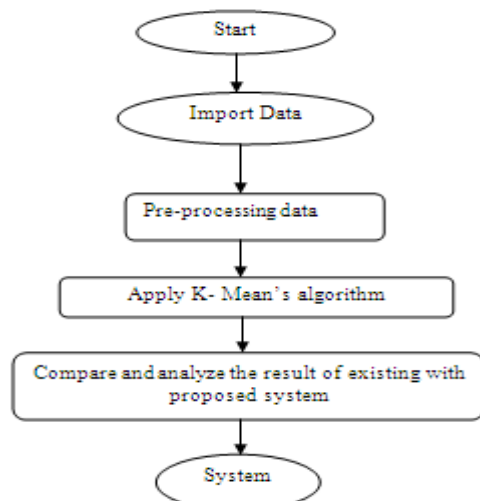


Fig.1 Flow chart of proposed work.

2. Flow Chart of SVM Working Process

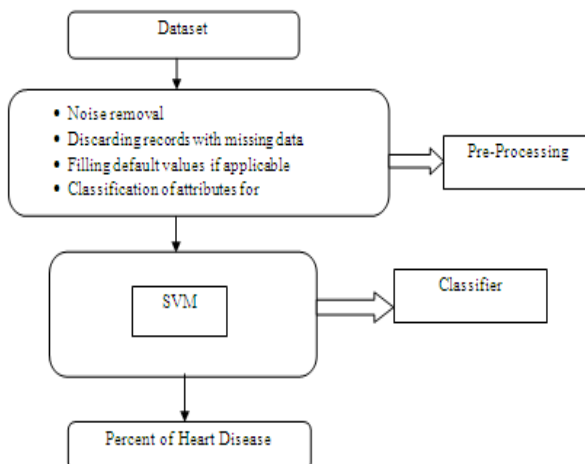


Fig. 2 Flow Chart of SVM Working Process.

VI. EXPERIMENT RESULTS

A sum of 308 records with 14 qualities was utilized from the Cleveland Heart database [1]. User enter values in restorative properties like sex, age, and so on. This model predicts that patient is having coronary illness or not relying upon this worth, specialists would prescribe to go for further heart examination. Fig 1. are used to stack UCI archive dataset for testing reason utilizing K-Mean grouping calculation.

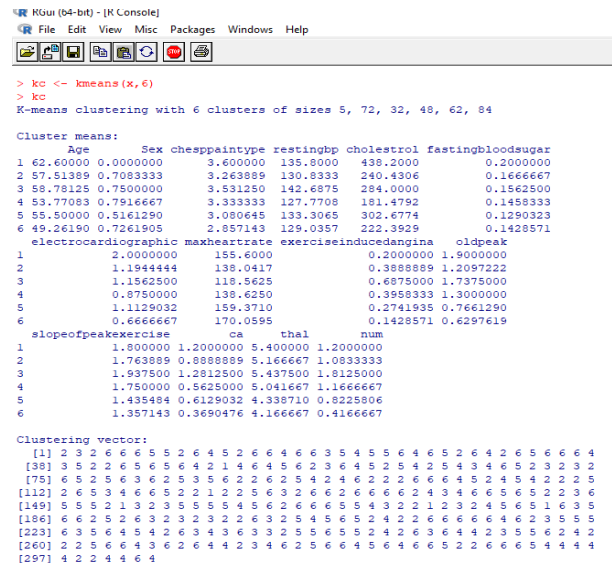


Fig.3 K-Mean Algorithm divided into 6 Cluster in whole Data Set.

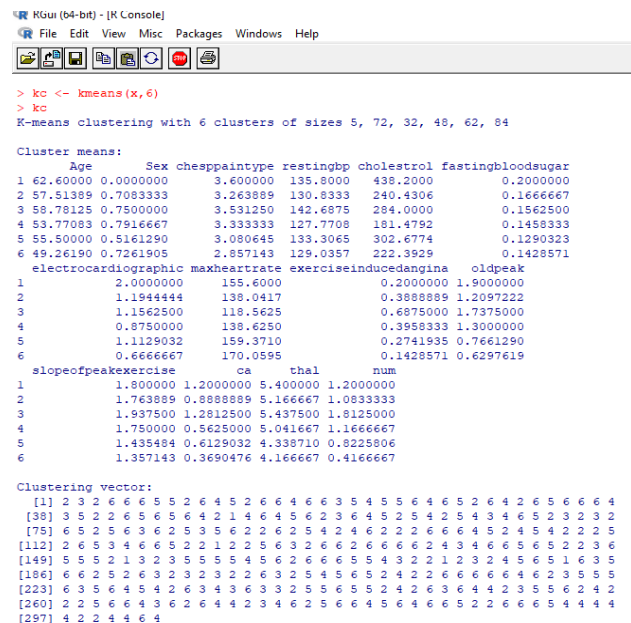


Fig. 4 K-Mean Algorithm plot Graph Age verses Cholesterol.

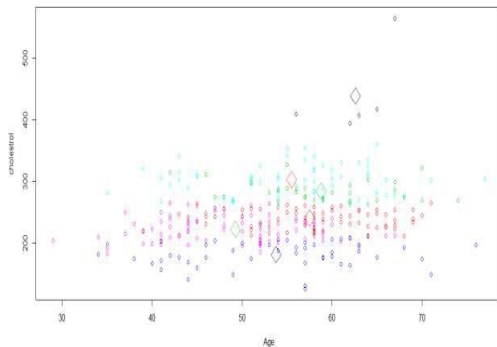


Fig.5 K-Mean Algorithm plot Graph Age verses Cholesterol.

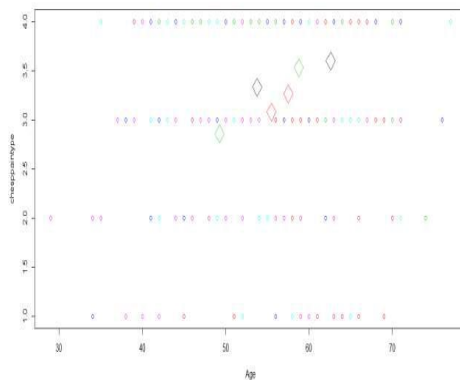


Fig. 6 K-Mean Algorithm plot Graph Age verses Chest Pain Type.

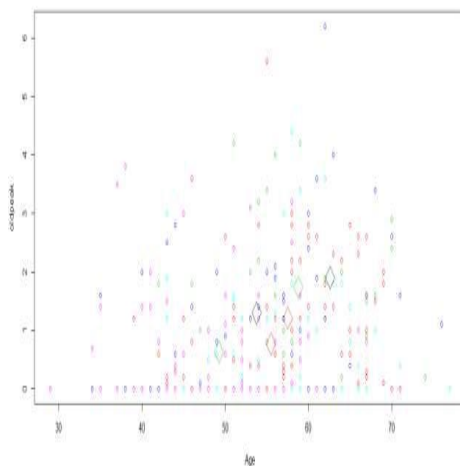


Fig.7 K-Mean Algorithm plot Graph Age verses Old Peak.

```
> summary(svm_tune)

Parameter tuning of 'svm':

- sampling method: 10-fold cross validation

- best parameters:
  cost gamma
  10 0.5

- best performance: 0.01515146

- Detailed performance results:
  cost gamma error dispersion
1 0.1 0.5 0.08012920 0.013799536
2 1.0 0.5 0.01694806 0.004710721
3 10.0 0.5 0.01515146 0.003805655
4 100.0 0.5 0.01515146 0.003805655
5 0.1 1.0 0.14509692 0.018620535
6 1.0 1.0 0.04002291 0.007456359
7 10.0 1.0 0.03605732 0.006336708
8 100.0 1.0 0.03605732 0.006336708
9 0.1 2.0 0.17822141 0.022018390
10 1.0 2.0 0.09888459 0.013073237
11 10.0 2.0 0.08787965 0.010434120
```

Fig. 8 SVM Algorithms.

3. Graph represents proposed work

Shows graph 4.2 Compare Accuracy SVM verses Customized K-Mean in Proposed Work. In figure shows Support Vector Machine better compare to Customized K-Mean algorithm.

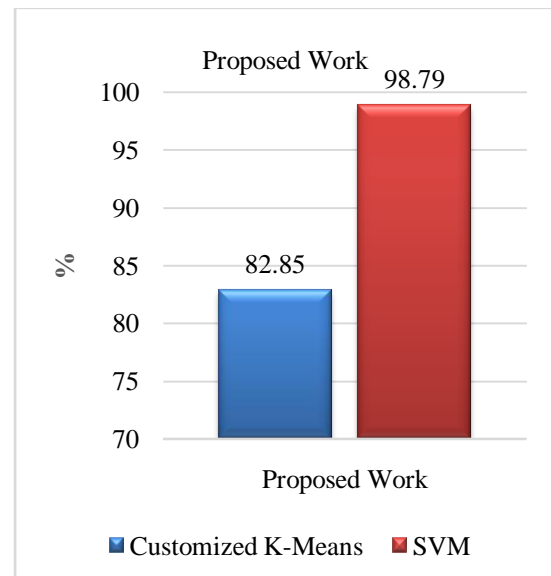


Fig. 9 Compare Accuracy SVM verses Customized K-Mean.

VII.CONCLUSION AND FUTURE WORK

The main aim of our project is to predict more accurately the presence of heart disease. With less number of attributes is a challenging task in Data Mining. Instead of going for a number of tests. Two data classification techniques were applied namely modified K-means and

SVM. In this paper a modified K means algorithm is proposed which tries to remove one of the major limitations of basic K-means algorithm, which requires number of clusters as input. This system can be further used in Future work as, for eg. It can incorporate other medical attributes besides the above list. To mine large amount of unstructured data, the Text mining can be used, available in healthcare industry database.

REFERENCES

- 1 Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", <http://mllearn.ics.uci.edu/databases/heartdisease/>, 2004.
- 2 Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, 2006.
- 3 Mrs.G.Subbalakshmi, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering
- 4 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.
- 5 Heon Gyu 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 Statlog database: <http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart>
- 7 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).
- 8 B M Ahamed Shafeeq, K S Hareesha, " Dynamic Clustering of Data with Modified K-Means Algorithm", International Conference on Information and Computer Networks (ICICN 2012), IPCSIT, vol. 27,pages 221-225,2012
- 9 Mohamed Abubaker, Wesam Ashour, "Efficient Data Clustering Algorithms: Improvements over K-means", International Journal of Intelligent Systems and Applications, vol. 5,issue 3, pages 37-49, 2013
- 10 Mohammed EI Agha, Wesam M. Ashour, " Efficient and Fast Initialization Algorithm for K-means Clustering" ,I.I. Intelligent Systems and Applications, vol. 4, issue 1, pages 21-31, 2012
- 11 Kaur, H., Wasan, S. K.: "Empirical Study on Applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 194-200, 2006.
- 12 MA.Jabbar,B.L Deekshatulu,Priti chandra,"Prediction of Risk Score for Heart Disease using Associative classification and Hybrid Feature Subset Selection",In .Conf ISDA,pp 628-634,IEEE(2013)
- 13 MA. Jabbar, B.L Deekshatulu, Priti chandra,"Knowledge discovery from mining association rules for heart disease prediction" pp45-53,vol 41,no 2 ,JATIT(2013)
- 14 Sellappan Palaniappan, Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security, Vol.8 No.8, August 2008