Volume 6, Issue 02, February 2021 ISSN (online): 2456-0006

International Journal of Science Technology Management and Research

Available online at: www.ijstmr.com

Heart disease prediction with customized K-mean and Support vector Machine

Mr. Vijayandra Ashok Yeole
M Tech scholar
Lakshmi Narain College of Technology and Science
Indore, India

Mr. Hemant Gupta
Assistant Professor
Lakshmi Narain College of Technology and Science
Indore, India

Abstract***: In medical sciences prediction of Heart disease is most difficult task. In India, main causes of Death are due to Heart Diseases. The deaths because of coronary illness in numerous nations happen because of work over-burden, mental pressure and numerous different issues. It is found as main reason in adults is due to heart disease. Thus, for detecting heart disease of a patient, there arises a need to develop a decision support system. Data mining arrangement systems, 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 Customized K-mean and SVM.***

In propose work use SVM and K-Mean machine learning techniques, both are get best result depends upon dataset but in our experimental data set SVM get best result compare to Customized K-Mean.

Keywords: SVM,K-Mean, Heart disease, Data Mining;

I. INTRODUCTION

Today Era is vary fast lift style it side effect is increasing many type of disease like Heart disease, Hypertension, Obesity, etc. Heart Disease is top disease causes of depth mostly in many country. The diseases of heart region behind muscles, valves, conduction system, heart attack & others. Heart attack is very big disease rapidly increased in few many year. Heart disease is seen in all type of society people day by day.

Heart disease is even being highlighted as a silent killer that ends up in death of an individual while not obvious symptoms. This nature of the malady is that the reason for growing anxiety concerning the malady & its consequences. thus continuing efforts square measure being done to predict the possibility of this deadly malady in previous. so numerous tools & techniques square measure frequently being experimented to suit the current day health wants. data processing techniques will be a boon during this regard.

Even though heart disease will occur in several forms, there's a standard set of core risk factors that influence whether or not somebody can ultimately be in danger for heart condition or not. By grouping the info from varied sources, classifying them underneath appropriate headings & finally analyzing to extract the required knowledge we will reach a conclusion.

This system are often all right custom-made to the do the prediction of heart condition. Because the documented quote says "Prevention is best than cure", early prediction & its management are often useful to stop & decrease the death rates because of heart condition.

A. Heart Disease

Heart disease is a term used for any type of disorder that affects the heart. Heart disease means the same as cardiac disease. Depending upon the pathology occurred, the heart disease are of various forms.

B. Coronary Heart Disease

Coronary Artery disease (CAD) is also called by the name Ischemic heart disease (IHD). It comes under cardiovascular diseases which involve the valves of the heart. It comprises of a group of diseases, angina, myocardial infarction & sudden cardiac arrest. The symptoms include,

- Chest pain / discomfort which radiates to shoulder, neck or back
- Heartburn
- Shortness of breath

C. Heart Attack

Heart attack is the common term used for Myocardial infarction (MI). It is due to the interruption of blood supply to a part of heart leading to damage to the heart muscle. Chest pain is the common symptom which may pass on to shoulder, neck, back or jaw. It may present in the centre or left side of the chest. D. Arrhythmia Cardiac Arrythmia is also knowncommonly as irregular heart beat is a group of conditions. Here, the heart beat will be too fast or slow or irregular. Sypmtoms may not be present many a times. Symptoms

may include palpitations or feeling of gap between heart beats. E. Heart Failure Heart failure is the inability of the heart to pump the blood for the body funtioning. Coronary artery disease, heart attack, valve diease etc. can lead to heart failure. Sypmtoms are difficulty in breathing, fatigue & leg swelling. Chest pain is not present always in heart failure.

II. PROPOSED CUSTOMIZED K-MEAN ALGORITHM

Today, numerous doctors' facilities oversee social insurance information utilizing medicinal services data framework; as this framework contains gigantic measure of information, and it is utilized to extricate shrouded data for therapeutic finding. The primary target of this framework is to manufacture Heart Disease Prediction System utilizing recorded heart database that gives analysis of coronary illness. To construct this framework, therapeutic terms, for example, circulatory strain, sex, cholesterol, sugar and so forth 13 input qualities are utilized. Information mining systems, for example, bunching, Classification is utilized in separating learning from database.

1. Customized K-mean

The proposed Customized algorithm proves to be a better method to determine the initial centroid and it is easy to implement. By eliminating one of its drawbacks, this Customized K-mean tries to enhance the k mean clustering algorithm. K-mean was used to apply on numerical data only. But, we encounter both numerical and categorical combination data values. This algorithm does not require number of clusters (k) as input is described below. By choosing two initial centroid, two clusters are created initially, which are farthest apart in the datasets. It can create two clusters with the data members at the initial steps, which are most dissimilar ones.

Input:
D: The set of n tuples with attributes Al, A2, and Am. All attributes are numeric, (where m = no. of attributes)
Output:
With n tuples suitable number of clusters distributed properly
Method:

1) To find the points in the data set which are farthest apart, compute sum of the attribute values of each tuple
2) As initial centroid take tuples with maximum and minimum values of the sum.
3) Using Euclidean distance create initial partitions (clusters) between the initial centroids and every tuple
4) From the centroid find distance of every tuple in both the initial partitions. Take other than zero. d=minimum of all distances.
5) For the partitions created in step 3, compute new mean (centroids)
6) From the new mean (cluster centers) compute Euclidean distance of every tuple. And depending on the following objective function, find the outliers: If Distance of the tuple from the cluster mean>d then only it is an Outlier.
7) New centroids of the clusters can be computed
8) From the new cluster centroids, calculate Euclidean distance of every outlier and find the objective function in step6. Outliers is not satisfying
9) Let the set of outliers obtained in step 8 is B= {Yl, Y2, Y p} (Where value of k is depends on number of outliers).
10) Repeat the steps until I (B==<D)
    a) By taking mean value of its members as centroid, create a new cluster for the set B,
    b) Depending on the objective function in step 6, find the outliers of this cluster,
    c) Check if no. of outliers = p then

i) Test every other outlier for the objective function as in step 6 after creation of a new cluster with one of the outliers as its member

ii) If there is any out fliers find it

    d) From the centroid of the existing clusters, calculate the distance of every outlier. If the existing clusters which satisfy the objective function in step 6?Then adjust the outliers
    e) The new set of outliers is B= {ZI, Z2 .... Zq}. (Where value of q is depends on number of outliers).

III. PROPOSED SYSTEM MODEL

Shows Figure 1.1 divided into five parts like Understanding of Research, Understanding of Domain, Data Preparation, Modeling and Evaluation and Deployment. First Understanding of the Data Mining Need, second Identification of attributes pertaining to risk of heart diseases , third collection of data and pre-processing the dataset , Fourth Apply 10-fold cross validation and classifiers (SVM and K-Mean) and Five Obtain Results (Accuracy) and Compare and Conclude.

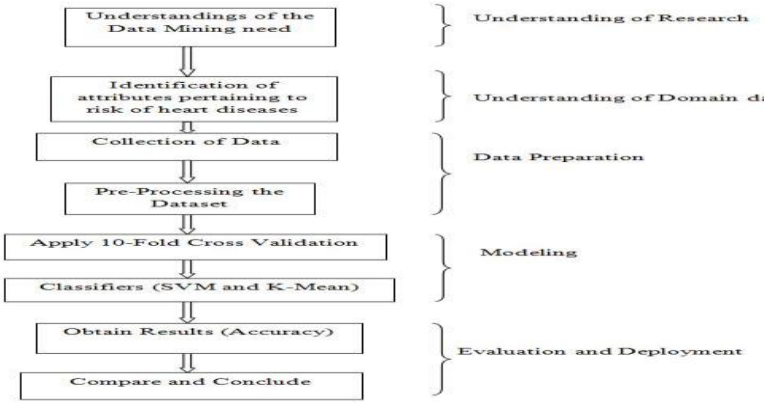

Fig. 1 Proposed System flow chart

IV. RESULT ANALYSIS

Table 1 Represent compression between Existing work and proposed work. In existing worked used to two approach Naïve Bayes and Decision tree but decision tree method get best accuracy compare to Naïve Bayes existing method, In propose work use SVM and K-Mean machine learning techniques, both are get best result depends upon dataset but in our experimental data set SVM get best result compare K-Mean.

TABLE 1
Compression between Existing Work and Proposed Work

Existing Work		Proposed Work	
Methods	**Accuracy**	**Methods**	**Accuracy**
Naïve Bayes	70.00	K-Mean	82.85
Decision tree	90.00	SVM	98.79

In figure 2 draw base on accuracy on existing work verses proposed work.

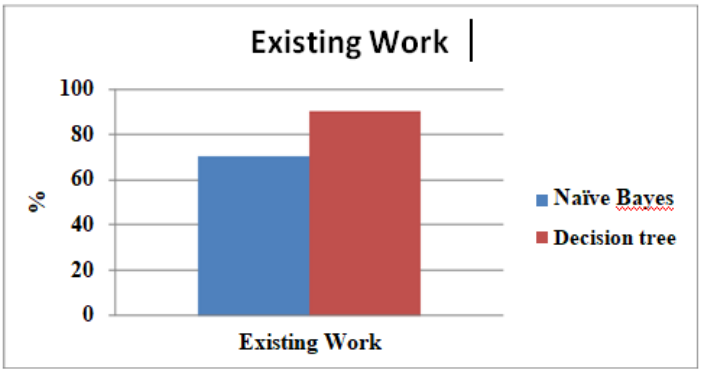

Fig. 2 Compare Accuracy Naïve Bayes Verses Decision Tree

Shows Fig2.and Fig3.Compare Accuracy Naïve Bayes Verses Decision Tree in Existing Work. In this figure shows decision tree better compare to Naïve Bayes algorithm.

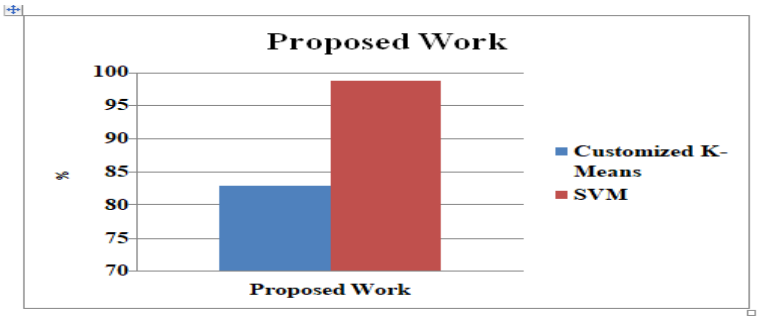

Fig.3 Compare Accuracy SVM verses Customized K-Mean

Shows Fig.3 Compare Accuracy SVM verses Customized K-Mean in Proposed Work. In figure shows Support Vector Machine better compare to Customized K-Mean algorithm.

CONCLUSION

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 Customized K-mean and SVM. In a Customized K-Mean algorithm is proposed which tries to remove one of the major limitations of basic K-Mean algorithm, which requires number of clusters as input. Existing work and proposed work. In existing worked used to two approach Naïve Bayes and Decision tree but decision tree method get best accuracy compare to Naïve Bayes existing method, In propose work use SVM and K-Mean machine learning techniques, both are get best result depends upon dataset but in our experimental data set K-Mean get best result compare SVM.

REFERENCE

1. Umar S., Agarwal K., Rizwan Beg, "Genetic Neural Network based Data Mining in Prediction of Heart Disease using Risk Factors", IEEE, 2013.
2. Chen A.H, Huang S.Y, Hong P.S, Cheng C.H, E J Lin,"HDPS: Heart Disease Prediction System", IEEE, 2011.
3. Jabbar M.A., Deekshatulu B. L.,Chandra P., "Heart Disease Prediction using Lazy Associative Classification", IEEE, 2013.
4. Nag P., Mondal S., Ahmed F., More A. and Raihan M.,"A Simple Acute Myocardial Infarction (Heart Attack) Prediction System Using Clinical Data and Data Mining Techniques", IEEE, 2017.
5. Amutha J,.Padmajavalli R, Prabhakar D., "A Novel approach for the Prediction of Treadmil Test in Cardiology using Data mining Algorithms implemented as a Mobile Application, Science Direct, 2018.
6. Pouriyeh S., Vahid S., Sannino G., Giuseppe De Pietro ,Arabnia H, Gutierrez J., "A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease", IEEE, 2017
7. Priyanga and Dr. Naveen, "Web Analytics Support System for Prediction of Heart Disease Using Naïve Bayes Weighted Approach (NBwa)", IEEE, 2017.
8. Gagandeep Kaur, Anshu Sharma, Anurag Sharma,"Heart Disease Prediction using KNN classification approach"JCSE, Vol.-7, Issue-5, May 2019 E-ISSN:2347-2693

.