Prediction of Heart Disease patients by using KNN Algorithm

Shalini B¹, Mr.G.Ananthnath²
¹Student, Dept. of MCA, KMM Institute of Post Graduate Studies, Tirupati
²Assistant Professor, Dept. of MCA, KMM Institute of Post Graduate Studies, Tirupati

Abstract—Now a day’s disease is the primary cause of death. Recently identified that one of the main reason for cause of death is heart diseases. It is necessary to predict the how the heart diseases will be in the future and we have to predict it. As the medical departments have large amount of data for analysis and prediction is necessary. And we have identified the reasons for them and we have to take the initial remedies for not spreading of these heart diseases. In data mining we have so many algorithms for prediction like Naive Bayes, Decision Tree and K-nearest neighbors¹, on by practical results shows that k-Nearest Neighbors (KNN) algorithm will provide the better accuracy and efficiency.

Index Terms—Data mining, Naive Bayes, Decision Tree, K-nearest Neighbors.

I. INTRODUCTION

Medicinal information digging has incredible potential for investigating the concealed examples in the informational collections of the restorative area. These examples can be used for clinical conclusion. Notwithstanding, the accessible crude medicinal information are generally disseminated, heterogeneous in nature, and voluminous. These information should be gathered in a composed shape. This gathered information can be then incorporated to frame a healing facility data framework. Information mining innovation gives a client arranged way to deal with novel and shrouded designs in the information[1].

The World Health Organization has evaluated that 12 million passings happen around the world, consistently because of the Heart sicknesses. A large portion of the passings in the United States and other created nations happen due to cardio vascular maladies. It is likewise the central reason of passings in various creating nations. All in all, it is viewed as the essential purpose for passings in grown-ups. The term Heart sickness incorporates the different maladies that influence the heart. Coronary illness was the real reason for losses in the diverse nations including India. Coronary illness kills one individual like clockwork in the United States[2].

Coronary illness, Cardiomyopathy and Cardiovascular sickness are a few classifications of heart infections. The expression "cardiovascular sickness" incorporates an extensive variety of conditions that influence the heart and the veins and the way in which blood is pumped and coursed through the body. Cardiovascular sickness (CVD) brings about a few ailment, handicap, and demise.

The analysis of illnesses is an essential and mind boggling work in medication.

Therapeutic finding is viewed as a critical yet entangled assignment that should be executed precisely and proficiently. The robotization of this framework would be greatly worthwhile. Lamentably all specialists don't have skill in each sub forte and additionally there is a deficiency of asset people at specific spots. Hence, a programmed medicinal conclusion framework would most likely be exceedingly valuable by uniting every one of them.

Fitting PC based data as well as choice emotionally supportive networks can help in accomplishing clinical tests at a decreased cost. Effective and exact usage of mechanized framework needs a near investigation of different systems accessible. This paper means to investigate the distinctive prescient/graphic information mining procedures proposed lately for the determination of coronary illness[3].

Life itself is totally reliant on the productive task of the heart. Cardiovascular sickness isn't infectious; you can't get it like you can this season's cold virus or an icy. Rather, there are sure things that expansion a man's odds of getting cardiovascular sickness. Cardiovascular sickness (CVD) alludes to any condition that influences the heart. Numerous CVD

© March 2018 | IJIRT | Volume 4 Issue 10 | ISSN: 2349-6002
patients have manifestations, for example, chest torment (angina) and weariness, which happen when the heart isn't accepting sufficient oxygen. According to a study about 50 percent of patients, in any case, have no indications until the point that a heart assault happens. Various variables have been appeared to expand the danger of creating CVD.[4].

Some of these are:
1) Family history of cardiovascular sickness
2) High levels of LDL (terrible) cholesterol
3) Low level of HDL (great) cholesterol.

Coronary illness is the term that incorporates a wide range of infections that influence different segments of the heart. All heart illnesses has a place with the class of CVD. Normal reason for all heart ailments is the lacking pumping of oxygen and blood from the heart to whatever is left of the body and the other way around. As indicated by world wellbeing report 2010, essential hazard variables of coronary illness are
1) Use of tobacco
2) Use of liquor
3) Hypertension
4) Physical idleness
5) Cholesterol
6) Heftiness
7) Undesirable eating routine
8) Raised blood glucose [PHA11][5].

Heart ailments are named
1) Coronary illness: coronary illness otherwise called coronary corridor ailment. It is a condition in which some plaque stores obstruct the veins prompting diminished supply of oxygen and blood to the heart.
2) Angina pectoris: This happens because of lacking supply of blood to the heart.
3) Congestive heart disappointment: This happens when heart can't sufficiently direct blood to whatever remains of the body.
4) Cardio myopathy: This is the debilitating of the heart muscle because of lacking heart pumping.
5) Congenital coronary illness: This is otherwise called innate heart deformity. This alludes to the development of an anomalous heart because of deformity in the structure of the heart.
6) Arrhythmias: This is related with a confusion in the cadenced development of the heart beat.

7) Myocarditis: Myocarditis is an aggravation of the heart muscle typically caused by parasitic and bacterial diseases and viral influencing the heart.

II. RELATED WORK

Boshra Bahrami, et al The objective of this paper is to evaluate K Nearest Neighbors (KNN) is used to classify dataset. After classification, some performance evaluation measures like accuracy, precision, sensitivity, specificity, F-measure and area under ROC curve are evaluated and compared. In this study four different classification algorithms applied on existing heart disease dataset. The Gain Ratio evaluation technique used as feature selection technique and four features extracted from dataset. Then preprocessed datasets, used to test the four classifiers using 10-folds cross validation. Six different performance measures considered for classifiers. Hence, it is a suitable candidate for testing in a clinical environment and implementing in decision support systems for helping physicians and healthcare professionals in diagnosis of heart disease[6].

M.Akhil jabber et al Nearest neighbor (KNN) is very simple, most popular, highly efficient and effective algorithm for pattern recognition. KNN is a straightforward classifier, where samples are classified based on the class of their nearest neighbor. Medical data bases are high volume in nature. If the data set contains redundant and irrelevant attributes, classification may produce less accurate result. Heart disease is the leading cause of death in INDIA. In Andhra Pradesh heart disease was the leading cause of mortality accounting for 32% of all deaths, a rate as high as Canada (35%) and USA. Hence there is a need to define a decision support system that helps clinicians decide to take precautionary steps. In this paper we propose a new algorithm which combines KNN with genetic algorithm for effective classification. In this paper we have presented a novel approach for classifying heart disease. As a way to validate the proposed method. This prediction model helps the doctors in efficient heart disease diagnosis process with fewer attributes. Heart disease is the most common contributor of mortality in India and in Andhra Pradesh. Identification of major risk factors and developing decision support system, and
effective control measures and health education programs will decline in the heart disease mortality[7].

III. PROPOSED TECHNIQUE
A. K-NN Algorithm:-
This strategy for characterization is a standout amongst the most essential and straightforward grouping techniques and ought to be utilized for an order ponder when there is next to zero earlier information about the dispersion of the information. This technique was created from the need to perform separate investigation when dependable parametric assessments of likelihood densities are obscure or hard to decide [8].

The calculation for this strategy is

- The k closest neighbor must be found utilizing the preparation dataset. The Euclidean separation measure is utilized to figure how close every individual from the preparation set is to the objective column that is being analyzed.
- Analyze the k-closest neighbor, which grouping or classification do the vast majority of them have a place with Appoint this characterization or class to the line being inspected.
- Rehash this system for the rest of the lines in the objective set.
- In this product a greatest incentive for k can be chosen, at that point the product constructs models parallel on all estimations of k up to the most extreme indicated esteem and scoring is done on the best of these models.

B. K-NN calculation for coronary illness forecast:-
In this segment we propose a technique for coronary illness characterization utilizing K-closest neighbor classifier with ideal element determination. We utilized symmetrical vulnerability as a decency measure to rank the qualities and in light of the positioning we prune minimum positioning traits. These assessed ascribes are given to KNN calculation which helps in characterization of coronary illness. Symmetrical vulnerability of property is a measure which repays data pick up

SU(X, Y) = 2[IG(X/Y)/H(X) H(Y)]

Comparability among the highlights is figured by: (X, y) = \(-\sqrt{\sum_{i=1}^{n} f(x_i, y_i)}\)

Where n is no. of characteristics .Our approach Comprises of 2 sections. Initial segment manages choosing prevail highlights and second part manages grouping informational collections and estimating the exactness of the informational collection.

Different strides in our proposed calculation:
Step1: pick the comparing informational index.
Step2: Apply SU Measure on the informational index.
Step3: recognize all the prevail highlights to the class idea and evacuate rest of highlights. (Rank the properties in view of SU prune the minimum rank characteristics from the informational collection.)
Step4: Now apply KNN on the dataset.
Step5: discover the exactness of the classifier. Precision measures the capacity of the classifier to accurately foresee the class mark.

IV. CONCLUSION

We utilized KNN algorithmic program with feature set choice to see the options that contributes a lot of towards the malady prediction. This technique indirectly reduces no. of tests to be taken by patients. This prediction model even helps the doctors in economical deciding process with fewer attributes to diagnose the center malady. The results which are obtained by applying the KNN algorithmic program for predicting the center diseases we've got determined that KNN algorithmic program will give high potency and accuracy than compared with the opposite algorithms accessible in data processing.

REFERENCES

[4] L. Shahwan-Akl, "Cardiovascular Disease Risk Factors among Adult Australian-


