Cardiovascular Disease Prediction Using Machine Learning

Anurag Singh

July 6, 2020

Cardiovascular Disease Prediction Using Machine Learning

(**Mr.R Karthick** ,Assistant professor in department of **computer science and engineering,   Galgotias University ,Greater noida**)

Author Name   Anurag Singh
Department of computer science and engineering
Galgotias university,Greater Noida -201310
Email id**-anuragsinghgu@gmail.com**

ABSTRACT-**In the current era of computer science each and every thing becomes intelligent,self suffiecient and perform task like humans. For these purposes there are various tools, techniques and methods are proposed. Support Vector Machine is a model for statistics and computer science, that has a algorithm to perform supervised learning, methods that are used for the analysis of data and Recognize patterns. SVM (Support Vector Machine) is mostly used for the classifications and the regression analysis and in the same way K- Nearest Neighbor Algorithm   is a classification algorithm that is used to classify data using training examples. In this research paper we use SVM and KNN algorithm to classify data and predicting (find hidden patterns) the target clusters. Here we use Medical patients nominal data from the common patients to classify and discover the data pattern to predict the   future diseases.Here we also Uses data mining which is used to classify the text analysis in future.**

Keywords-Machine learning , support vector machine ,k-nearest neighbour algorithm.

INTRODUCTION

Heart diseases are even being highlighted as the silent killers which ultimately leads to death of a person without the obvious unexpected Symptoms. This nature of these diseases is the cause of growing anxieties about the Diseases & there consequences. Henceforth, continuous efforts are being made to predict the possibilities of these deadly diseases in prior. So, that various Tools & Techniques which uses futuristic science are regularly being experimented to suit the present day Health Needs. Machine Learning techniques can be a boon in this Concern. Even though heart disease have occured in Various forms, even then there are various common matrix of deep and basic risk reasons that influence and determine whether the patient will be at ultimatum of risk for HEART DISEASES. By gathering and combining the data of various sources like different hospital institutions, classifying them under suitable and proper   headings & finally Analyzing to extracted

Then we get the desired data where we can define   the Conclusion. These techniques are very well performed by the doctor for the prediction of Heart and its

related Diseases. As the well known quote states "Prevention is better than cure", timely determination of the diseases can be really helpful to cure the drastic consequences & lower the increased mortality rates caused by the Heart diseases.

Heart diseases are the leading skyrocketing reason of deaths compare to other diseases like cancers. The digits of male & female having heart disease are sky rocketing each year. This prompts us to diagnose and start the treatment at the earliest. Due to the lack of the Resources , Machines and Equipment

In the medical era of now, the accuracy to determine the heart disease at every occasion is not a menace to face but the Utilization of newest and

Has proven to be definitely beneficial to the Medical field and there patient. There problems can be solved by Adapting Machine Learning techniques and utilizing them 100 percent. This research paper intends to adopt SVM (Support Vector Machine) & KNN (K-Nearest Neighbor)Algorithms – two qualities for the effective and accurate determination of Heart diseases. It compare the efficiency & accuracies of the two techniques to tell and choose out of the best.

SYSTEM ARCHITECTURE
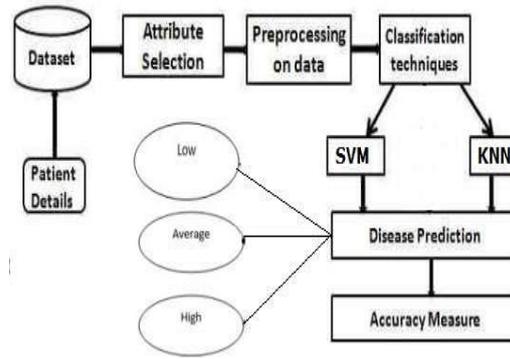
The processing of the System Architecture is explained in the following steps:

1. Data set repository which includes all the Patients disease Details.

2. entity Selected for the process chooses the important attributes to predict the vascular diseases.

3. After determining all the provided data sources, they are later on choosen, cleaned and are transformed into the required form.

4. All the classifiers techniques and tools as described will be used in the pre-processed data to tell the accuracy of cardio diseases.

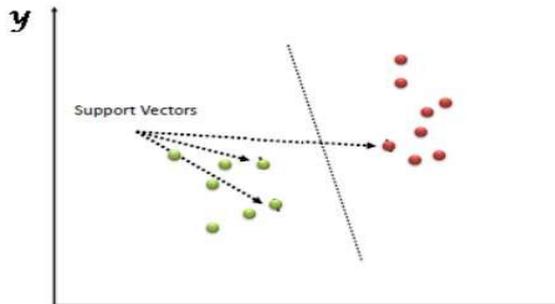5.These Accuracies measures and comparasion of accuracies of all the differing Classifications.

MODULES

Support Vector Machine (SVM):

Support Vector Machine is the most famous supervised type of   machine learning strategy (that has a already set goal of Variables and factors) which will be used for the classification and also for prediction. the classification of dataset, it looks for a hyper-plane in the featured area that differs between all the other datasets. A SV Machine represents all the learning data factors by points in the Featured region, mapping in this manner that the points refering to different classes are kept apart or

separated with a large distance or as far as poss. The Testing data sets are then pointed on that same region of the feature space and the classification is done on which part of the line they are plotted.

$$Eucledian\ Distance = D(x, y) \qquad (1)$$
$$= (x_i - y_i)_{2k_i} = 1$$

K=number of cluster

x , y=co-ordinate sample spaces

The algorithm for K-Nearest Neighbour is defined in the steps below:

1. Here, D represents the samples used for the training datasets and k here used to denote the total no. of Nearest neighbours.

2. Form a Super class for every Sample class.

3. Calculate the Euclidian distance for each and every training or learning sample datasets.

4. On the basis of majority of the neighbour and use classification to separate the samples.

K-**NEAREST NEIGHBOUR ALGORITHM**

KNN is also a supervised learningor training algorithm but slower, it requires more amount of watch to get learned for classification that differs algorithms are separated into 2 steps learning from the dataset and applying it on new instances . The K- Nearest Neighbour algorithm is formed on the test of the weight for every data set that is termed as neighbour. In K Nearest Neighbouring dispalcement is calculated for the learning datasets of each and every of the K __Nearest data clusters and then classifiers are used and the basis of the most of points.There are 3 different varities of distances needed to be calculated in knn algorithm thses are:- Euclidian distance, Manhattan distance and Minkowski distance out of these Euclidian distance is considered for most of the points. This formulae is used to find this distance.
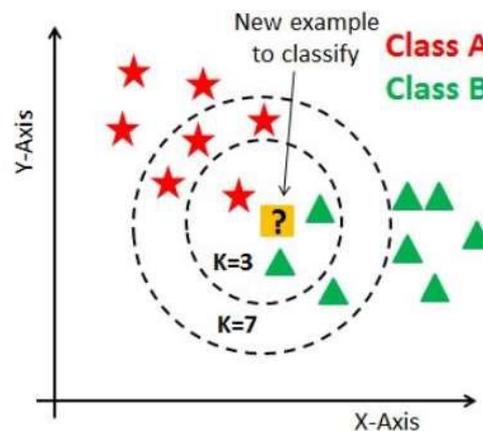

K-NEAREST NEIGHBOUR

CLUSTERING MODULE

• We uses Fuzzy C means Clustering algorithm which is an application

for clustering the medical data.

The data given by the doctor collected from the patient is then Fed into the clustering

algorithm. It is then the job of the clustering

algorithm.

• Data clustering is the form of separating the datasets from the doctors into the

elements of the classes or clusters to differentiate   in the same class can be identified as similar type product and are as same as possible, and objects in different clusters could easily be distinguished from the 1st cluster and are as dissimilar as possible.

SYSTEM CONFIGURATION:

Hardware requirements:

Processer : Any Updated Intel Processor

Ram : Min 4 GBand more.

Hard Disk : Min 250 GB.

Software requirements:

Operating System

: Windows 8 or any Window

family

Technology : Python 3.6 and Java7.

IDE : PyCharm or IntelliJ.

Fuzzy C-Mean Clustering

• Fuzzy C-Means Algorithm includes the following steps-

– firstly Initiate k seeds of prototypes like $p1 , p2 , …, pk.$

– secondly start **Grouping**:

Assign samples to all their nearest prototypes.

Then start Forming the non-overlapping clusters from these samples.

–thirdly start doing Centering:

Form the Centers of clusters which will become the new prototypes

–Finally Repeat the grouping and centering steps, until the Convergence.

ADVANTAGES

• Ensures a higher rate of accuracy and efficiency.

• Computation time is greatly reduced.

CONCLUSION

• Our Proposed System is a Fuzzy C-Means clustering algorithm for predicting the risk or the chances of heart attack in a patient using the attributes like heart rate ,blood pressure which is collected by the doctor.

• Proper adaptation and separation of FCM clusters or the datasets into an optimum

number of clusters helps in   detecting the normal and the critical or abnormal cases efficiently.

We extend our sincere thanks to Dr.JohnA for supporting me at every stage.

We would also like to thanks to our project guide and Associate professor

Mr R Karthick.

REFERENCES

1. Sellan Palapan , Raf Awadh, *"The Intelligent Heart Disease System."*

Its Prediction System Using Data Mining and Clustering Techniques" IEEE

Conference, 2008,pp 1008-1015.

2. Caroluos Ordenez *"Assosiative Rules Discovery of the Train and*

Tests Approaches for the Heart Diseases Predictions" IEEE

Relative Transactions on Information and Technology in Biomedicines, Volume. 100, No. 22, April 2016.

3. Shanti kumar Bhushan.Patil,Kumar swamyrana

"Intelligent and most Effective

Heart Attack Prediction System Using Data Mining techniques and Artificial

Neural Networks and machine learning". European Journals of Scientific Researches Volume. 301,

No. 004, 2010, 6042-6506 .

4. S.Palapan,R. Awadh,   2008. *"* **heart disease prediction system using data mining techniques and machine learning"**. International

Journal of Computer Science and Network Security 8, 105–156.   and website

www.benjaminandjamesbush.com/

fuzzyclustering.**pptx**