www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May-2022, pp : 62-64

e-ISSN : 2583-1062

Impact Factor : 2.205

DATA SCIENCE: HEART PREDICTION AND VISUALIZATION

Aniket Kadam[1], Vinay Kadam[2], Manmanth Mengji[3], Saurabh Jagtap[4]

[1,2,3,4]Computer Science, S.V.P.M College Of Engineering, Malegaon, Maharashtra, India.

ABSTRACT

Healthcare is among the most important fields of medical study and right now new technology and data are improving. It is very hard to keep track of all the information and data that the patients given to us because it is so much and increase. To deal with this data, you can use Big Data to do that. There are a lot of ways to treat a lot of different illnesses around the world. Machine Learning is a new method that helps predict and diagnose a disease and with the help of data visualization we improve it more. Here, we see the prediction of disease based on symptoms with Machine learning algorithms, like Naive Bayes, Decision Trees, SVM, Random Forest and ANN are used to look at the data and with the help of data mining and visualization identify which diseases are likely to show up. It is written in the python programming language, which is used to write it. The research shows that the best algorithm is the one that is the most accurate. Algorithms performance is depend on the datasets which are provided for training and testing

Keywords: Heart Disease, Random Forest, Machine Learning, Prepossessing, Classification, SVM, Decision Tree

1. INTRODUCTION

At present, when one hurt from particular disease, then the person has to visit to physician which is time consuming and costly too. Also if the user is out of reach of physician and hospitals it may be difficult for the user as the disease can not be identified. So, if the above process can be completed using a computerized program which can save time as well as money, it could be easier to the patient which can make the process easier. There are other Heart related problem Prediction System using data mining techniques that identify the risk level of the patient. Heart Disease Predictor is a Desktop based application that predicts the heart disease of the user with respect to the symptoms given by the user.

Heart Disease Prediction system has data sets collected from various health related sites. With the help of heart Disease Predictor the user will be able to know the probability of the disease with the given symptoms. In proposed system, A considerable amount of raw healthcare data is transformed into information that may be used to make better decisions and predictions using data mining. Machine learning algorithms such as SVM, Decision Tree, NB, Random Forest , and ANN are used in the proposed system. In the proposed system, we must upload the Heart Disease Dataset and train the model using a machine learning technique. We know that we are performing data processing functions on the system, therefore we use three data processing modules: preprocessing, feature extraction, and classification, all of which utilizes different algorithms. Then, construct a model and evaluate its effectiveness. With the help of this model, you can predict heart disease.

2. METHODOLOGY

For the prediction of heart disease , we use the machine learning algorithms and .data mining and data visualization tools.

After the data mining and extracting the useful information from the data we train and test the data and after that apply machine learning algorithms on that and predict the heart disease .

The performance of the algorithm is depend on the dataset which is given to Machine learning algorithms like svm ,random forest ,decision tree,ann is used into the prediction of the heart disease

2.1 Support Vector Machine:-

Support Vector Machine(SVM) The Support Vector Machine (SVM) is a well-controlled machine learning approach that excels at both classification and regression. It is, however, mostly employed in classification problems. Each feature value represents the value of a single coordinate in the SVM algorithm, and each data item is specified as an n-dimensional space point (where n is the number of features). Support Vectors, which are essentially individual observation coordinates, are then performed.

2.2 Decision Tree:-

In general, decision tree analysis is a flexible technique that may be used in a wide range predictive modelling scenarios. An algorithmic strategy that splits the data set in a variety of ways according to specified conditions can be used to create decision trees. Decision settings are the most powerful algorithms in the supervised algorithm category. They can be used to tackle problems with regression as well as classification. The two main parts of the tree are decision-making nodes, which divide the data and leave the outcomes.

3. MODELING AND ANALYSIS

MODULES

Enter Symptoms: Symptoms of a heart attack include cp, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, thal etc

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May-2022, pp : 62-64

e-ISSN : 2583-1062

Impact Factor : 2.205

Train Dataset: Identify Your Goal. The initial step is to pinpoint the set of impartial that you want toachieve through a machine learning application. Select suitable Algorithms. Different algorithms are suitable for training artificial neural networks. Develop Your Dataset

Analyze Symptoms: In this module we analysis the parameters of the symptoms value like cp, chol etc.

Preprocessing: Pre-processing is a term used to describe processes on datasets where the input and output are both intensity text at the lowest level of abstraction. Which was used to enhance text data before by suppressing undesired distortions or enhancing certain text features that are relevant for later processing.

Feature Extraction: Feature extraction is a step in the dimensionality cutting process, which divides and reduces a large set of raw data into smaller groupings. As a result, processing will be simpler.

Cluster Cluster : A group of objects. For example data component into different groups of similarity between in a single group cluster partitions the data set in to cluster classes. Each and every near object is neighborhood object. There are two goals of cluster. First one is a bury class second is an intra class. Inter class cluster means cluster distance is maximized. Intra cluster means cluster distances are minimized.

Classification: Classification is a action of categorizing a given set of data into classes. The classes are often referred to as target, label or categories. Classification predictive modeling involves assigning a class label to input examples. then in garding where we use our CNN algorithm to classify and prediction ,we pass this geometry based features of heart disease to the classification to for classification and prediction , then on that basis it predict heart disease. Result: Predict the Heart Disease detected or Not.

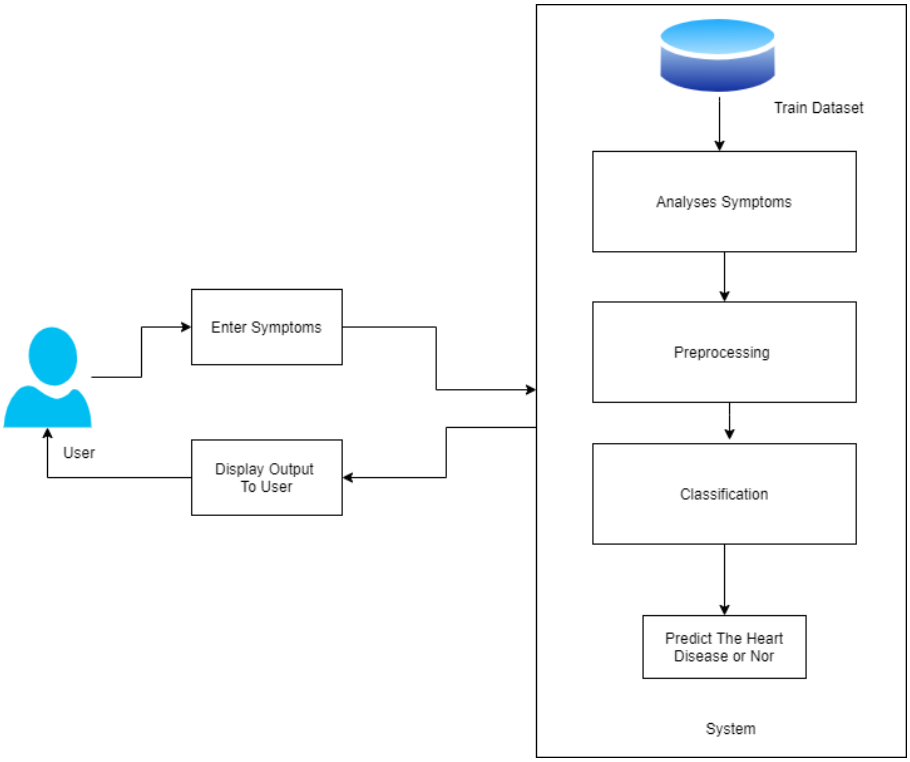

Figure 1: System Architecture of the Heart Prediction System

4. RESULTS AND DISCUSSION

The Heart Disease Dataset consists of 13 input attributes including age, gender, type of chest pain, blood pressure, cholesterol, blood sugar level, electrocardiograph result, maximum heart rate, exercise-induced angina, old peak, Slope, number of vessels colored, thal. The dataset hold 303 instances, from which 2 instances for a count of vessels colored attribute and 4 instances for thal attribute are not found , which are filled by their mean value for the dataset respectively. The prediction attribute make of 5 classes ranging from integer value 0 - 4 where 0 shows absence and the integer value from 1 - 4 indicate the availability of heart disease. We code the prediction attribute to class 0 and 1 to indicate not presence or presence of heart disease respectively. The feature election from 13 input parameter by backward elimination resulted in a total of 11 significant insert parameters which include gender, type of chest ache , blood pressure, blood sugar level, electrocardiograph result, greatest heart rate, exercise induced angina, old peak, Slope, number of vessels colored, thal. Fig. 4 shows the comparison in the middle of the accuracy of the models obtained for heart dataset. Support vector machine was found to have the highest accuracy among all of them

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
Vol. 02, Issue 05, May-2022, pp : 62-64

e-ISSN : 2583-1062

Impact Factor : 2.205

Table 1. Sample Comparison

SN.	Sample	Quantity (Liter)
1	SVM	98.98
2	Decision Tree	97.37
3	Random forest	98.25
4	Naïve Bayes	89.45
5	ANN	98.00

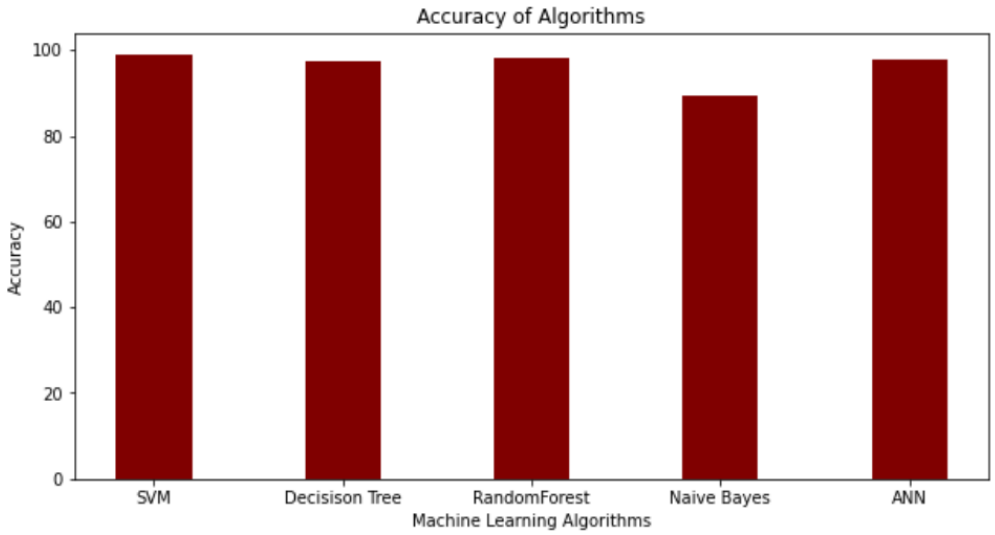

Figure 2: Accuracy of algorithms

5. CONCLUSION

We have Proposed Heart Disease Predication applied different machine learning algorithms like SVM, ANN, Random forest, Decision tree, Naıve Bayes for diagnosis of heart disease. The goal of this study is to predict heart disease based on symptoms. The project is set up in such a way that the system takes the user's symptoms as input and outputs a prediction of heart disease.

6. REFERENCES

[1]. Chittampali Sai Prakash,Myneni Madhu Bala,Attluri Rudra "Data Science Framework-Prediction ,Varient Models and Visualization ",2020 4rd International Conference on Trends in Electronics and Informatics (ICOEI).

[2]. Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, " Design And Implementing Heart Disease Prediction Using Naives Bayesian", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).

[3]. Aditi Gavhane , Gouthami Kokkula , Isha Pandya , Prof. Kailas Devadkar , "Prediction of Heart Disease Using Machine Learning", 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)

[4]. Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 2018 4th International Conference on Computational Intelligence Communication Technology (CICT).

[5]. C. Sowmiya, P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques",2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).

[6]. R Latha, P Vetrivelan, "Blood Viscosity based Heart Disease Risk Prediction Model in Edge/Fog Computing",2019 11th International Conference on Communication Systems Networks (COMSNETS).

[7]. Gaurav Meena, Pradeep Singh Chauhan, Ravi Raj Choudhary, "Empirical Study on Classification of Heart Disease Dataset-its Prediction and Mining",2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)..