International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT206421

Heart Disease Prediction Using Machine Learning Algorithms

Abhay Agrahara
Department of Computer Science & Engineering College, Ghaziabad, Uttar Pradesh, India

Article Info
Volume 6, Issue 4
Page Number: 137-149
Publication Issue:
July-August-2020

Article History
Accepted: 10 July 2020
Published : 20 July 2020

ABSTRACT

Heart disease is one of the most fatal problems in the whole world, which cannot be seen with a naked eye and comes instantly when its limitations are reached. Therefore, it needs accurate diagnosis at accurate time. Health care industry produced huge amount of data every day related to patients and diseases. However, this data is not used efficiently by the researchers and practitioners. Today healthcare industry is rich in data however poor in knowledge. There are various data mining and machine learning techniques and tools available to extract effective knowledge from databases and to use this knowledge for more accurate diagnosis and decision making. Increasing research on heart disease predicting systems, it become significant to summarize the completely incomplete research on it. The main objective of this research paper is to summarize the recent research with comparative results that has been done on heart disease prediction and also make analytical conclusions. From the study, it is observed Naive Bayes with Genetic algorithm; Decision Trees and Artificial Neural Networks techniques improve the accuracy of the heart disease prediction system in different scenarios. In this paper commonly used data mining and machine learning techniques and their complexities are summarized.

Keywords: Machine learning, Heart disease, Classification, Naive Bayes, Support Vector Machine, Decision Trees, Associative Rule, Logistics Regression, Random Forest, K-Nearest Neighbours.

I. INTRODUCTION

Heart disease describes a range of conditions that affect your heart. Heart disease term includes a number of diseases such as blood vessel diseases, such as coronary artery disease; heart rhythm problems (arrhythmias); and heart defects you're born with (congenital heart defects), among others. The term heart disease is sometimes used interchangeably with the term cardiovascular disease. Cardiovascular disease (CVD) generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack (Myocardial infarctions), chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease. 17.9 million People die each year from CVDs, an estimated 31% of all deaths worldwide. Nowadays healthcare sector produces large amount of information about patients, disease diagnosis etc.

however this data is not used efficiently by the researchers and practitioners. Today a major challenge faced by Healthcare industry is quality of service (QoS). QoS implies diagnosing disease correctly & provides effective treatments to patients. Poor diagnosis can lead to disastrous consequences which are unacceptable . There are various heart disease risk factors. Family history, Increasing age, Ethnicity and being male are some risk factors that cannot be controlled. But Smoking, Diabetes, High cholesterol, High blood pressure, not being physically active, being overweight or obese are those factors that can be controlled or prevented . Data mining is the process of discovering unknown hidden patterns (knowledge) from large pre- existing data sets with the involvement of data mining and machine learning techniques, statistics, and database systems. The discovered knowledge can be used to build intelligent predictive decision systems in different fields like health care for accurate diagnosis at accurate time to provide affordable services and save precious lives. Machine learning provides computer programs the ability to learn from predetermined data and improve performance from experiences without human intervention and then apply what have learned to make an informed decision. At every successful decision machine learning program improves its performance. . Given below figure depicts the knowledge discovery from data (KDD) process.

II. Prior Knowledge

In every field of education we need prior knowledge to understand and analyze that field very well, prior knowledge become base for successful understanding and analyses of any study. So before we start to study the actual content of this paper we have to study and understand the basic concepts related to the paper that will help us to understand and comprehend the paper very well.

2.1) Classification: Classification is a supervised data mining and machine learning technique. It is a two step process, first step is called learning step where the model is constructed and trained by a predetermined dataset with class labels (training set) and second step is classification (testing) step where the model is used to predict class labels for given data (test data) to estimate the accuracy of classifier model.

2.2) Associative rule: Associative rule miming is a data mining technique which is used to find associative rules or patterns in data. In association rule mining, a pattern is discovered based on a relationship of a particular item to other items in the same transaction. It finds frequent item sets in data by using predefined support and confidence values. The association rule technique is used for heart disease diagnosis to discover the relationship of different attributes used for analysis and sort out the patient with all the risk factor which are required for prediction of disease.

2.3) Decision Tree: Decision tree is a technique that is used as a decision support tool that uses a tree-like graph or model of decisions . It takes as input a record or object described by a set of attributes and returns a "decision with predicted output value for the input". The input attributes can be discrete or continuous. After performing a sequence of tests decision tree reaches its decision. Each non leaf node of a decision tree corresponds to a test for the relevant attribute value, and the branches from the node are labeled with the possible outcomes of the test. Each leaf node in the tree specifies the value (decision) to be returned if that leaf is reached . J48, Random Forest (RF) and Logistic Tree Model (LTM) are Decision tree implementation algorithms.

example of decision tree are:

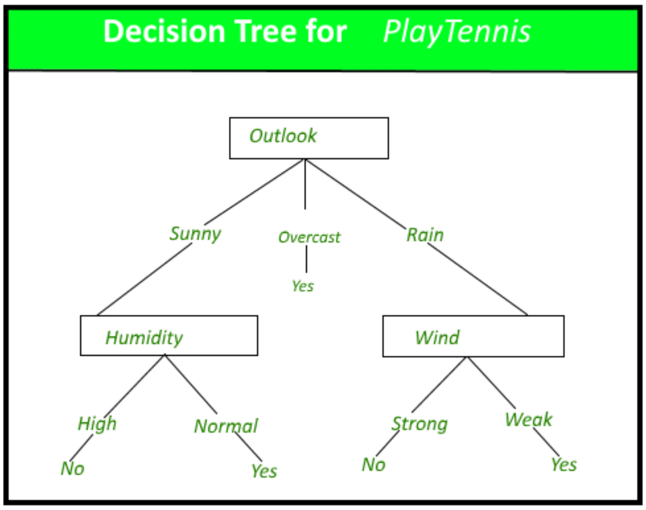


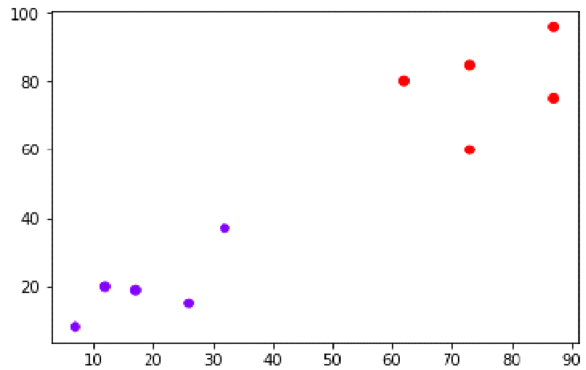
Fig 1 : Decision tree for Table tennis game

2.4) K-Nearest Neighborhood :The k-nearest neighbors (KNN) algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems.The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

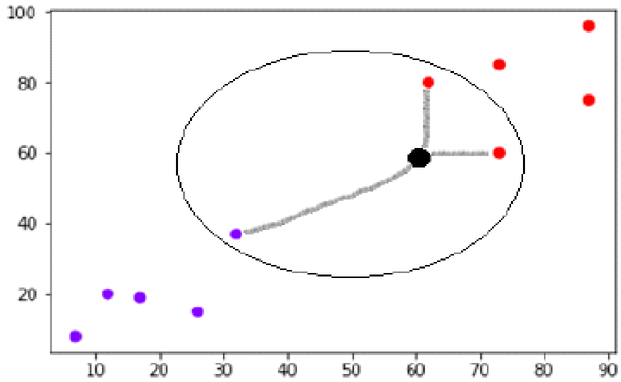
To select the K that's right for your data, we run the KNN algorithm several times with different values of K and choose the K that reduces the number of errors we encounter while maintaining the algorithm's ability to accurately make predictions when it's given data it hasn't seen before.

Example:

Suppose we have a dataset which can be plotted as follows :



Now, we need to classify new data point with black dot (at point 60,60) into blue or red class. We are assuming K = 3 i.e. it would find three nearest data points. It is shown in the next diagram –



We can see in the above diagram the three nearest neighbors of the data point with black dot. Among those three, two of them lies in Red class hence the black dot will also be assigned in red class.

2.5) Cross Validation: Cross-validation is a technique to evaluate predictive models by dividing the original dataset into a training set to train the model, and a test set to evaluate it. In k-fold cross-validation, the original sample is randomly divided into k equal size subsets. Of the k subsets, a single subset is taken as the validation data for testing the model, and the remaining k-1 subsets are used for training the model. The cross-validation process is then repeated k times (the folds), with each of the k subsets used exactly once as the validation data and average accuracy of k-folds is taken as final accuracy. In most experiments 10-fold cross validation technique is used. In 10-fold cross validation all the instances of the data set are used and are divided into 10 disjoint groups, where nine groups are used for training and the remaining one is used for testing. The algorithm runs for 10 times and average accuracy of all folds is calculated.

2.6) Random Forest : Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks). In this post we'll learn how the random forest algorithm works, how it differs from other algorithms and how to use it.

One big advantage of random forest is that it can be used for both classification and regression problems, which form the majority of current machine learning systems. Let's look at random forest in classification, classification is sometimes considered the building block of machine learning. Below you can see how a random forest would look like with two trees:

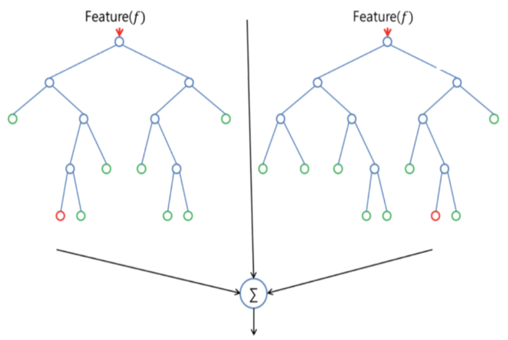


Fig 2 : Two Tree Random Forest

2.7) Support Vector Machine : The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space(N — the number of features) that distinctly classifies the data points.

Support vectors are data points that are closer to the hyperplane and influence the position and orientation of the hyperplane. Using these support vectors, we maximize the margin of the classifier. Deleting the support vectors will change the position of the hyperplane. These are the points that help us build our SVM.

In the SVM algorithm, we are looking to maximize the margin between the data points and the hyperplane. The loss function that helps maximize the margin is hinge loss.

2.8) Logistics Regression : Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is categorical.

For example

- To predict whether an email is spam (1) or (0)
- Whether the tumor is malignant (1) or not (0)
  - *Sigmoid Function:*

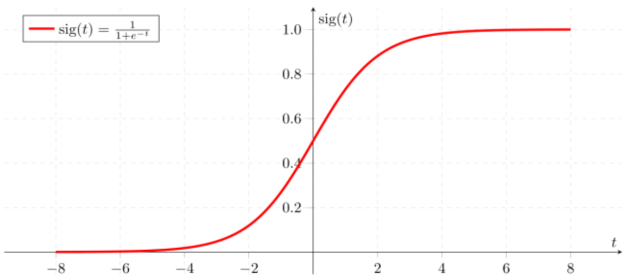
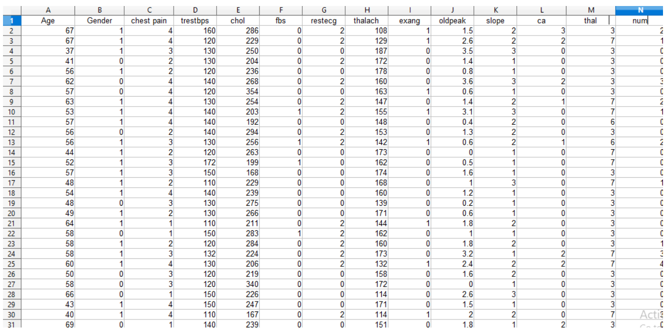


Fig 3 : Sigmoid Function

III. METHODS AND MATERIAL

3.1 : Description about dataset

3.1.1) Dataset is taken from Kaggle website and UCI repository ,this dataset contain 14 colmns with 1004 rows ,in which 5 attributes are numeriacal and other 9 attributes are classification type.Sample of dataset given in fig 4:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Age	Gender	chest pain	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
2	67	1	4	160	286	0	2	108	1	1.5	2	3	3	2
3	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
4	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
5	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
6	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
7	62	0	4	140	268	0	2	160	0	3.6	3	2	3	3
8	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
9	63	1	4	130	254	0	2	147	0	1.4	2	1	7	2
10	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
11	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
12	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
13	56	1	3	130	256	1	2	142	1	0.6	2	1	6	2
14	44	1	2	120	263	0	0	173	0	0	1	0	7	0
15	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
16	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
17	48	1	2	110	229	0	0	168	0	1	3	0	7	1
18	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
19	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
20	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
21	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
22	58	0	1	150	283	1	2	162	0	1	1	0	3	0
23	58	1	2	120	284	0	2	160	0	1.8	2	0	3	1
24	58	1	3	132	224	0	2	173	0	3.2	1	2	7	3
25	60	1	4	130	206	0	2	132	1	2.4	2	2	7	4
26	50	0	3	120	219	0	0	158	0	1.6	2	0	3	0
27	58	0	3	120	340	0	0	172	0	0	1	0	3	0
28	66	0	1	150	226	0	0	114	0	2.6	3	0	3	0
29	43	1	4	150	247	0	0	171	0	1.5	1	0	3	0
30	40	1	4	110	167	0	2	114	1	2	2	0	7	3
31	69	0	1	140	239	0	0	151	0	1.8	1	2	3	0

Fig 4 : sample of dataset

3.1.2) Attributes of dataset are as:

1 - Age
2 - Gender : male-1,female-0
3 - CP : chest pain
cp: chest pain type
-- Value 1: typical angina(Chest portion)
-- Value 2: atypical angina
-- Value 3: non-anginal pain
-- Value 4: asymptomatic
4 - trestbps: resting blood pressure (in mm Hg on admission to the hospital)
5 - chol: serum cholestoral in mg/dl
6 - fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
7 - restecg: resting electrocardiographic results
-- Value 0: normal
-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
-- Value 2: showing probable or definite left ventricular hypertroph by Estes' criteria
8 - thalach: maximum heart rate achieved
9 - exang: exercise induced angina (1 = yes; 0 = no)
10 - oldpeak = ST depression induced by exercise relative to rest
11 - slope: the slope of the peak exercise ST segment
-- Value 1: upsloping
-- Value 2: flat
-- Value 3: downsloping
12 - ca: number of major vessels (0-3) colored by flourosopy
13 - thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
14 - target: diagnosis of heart disease (angiographic disease status)
-- Value 0: < 50% diameter narrowing
-- Value 1: > 50% diameter narrowing
(in any major vessel: attributes 59 through 68 are vessels)

3.1.3) Graphical representation of relation between attributes:

There is relation between no. of disease occured(represented by 1) or not occured(represented by 0).

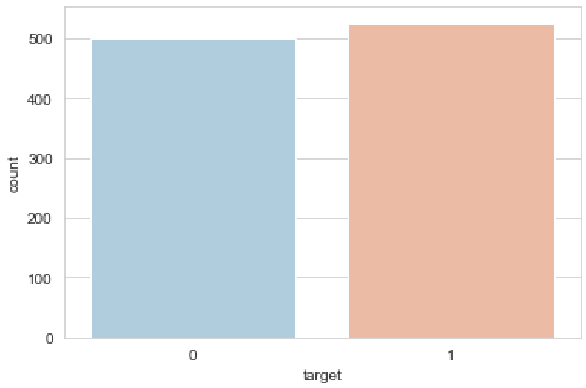


Fig 5 : Represention of disease occurence

There is some relation of attributes with target features as:

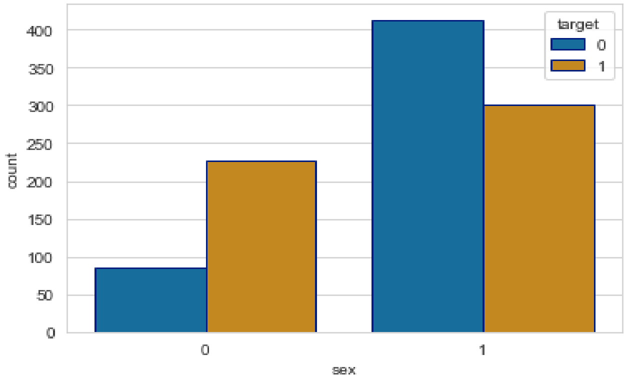


Fig 6 : Gender Vs Target

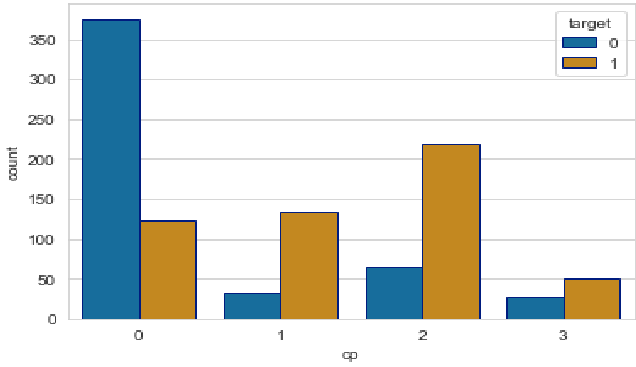


Fig 7 : Chest pain vs Target

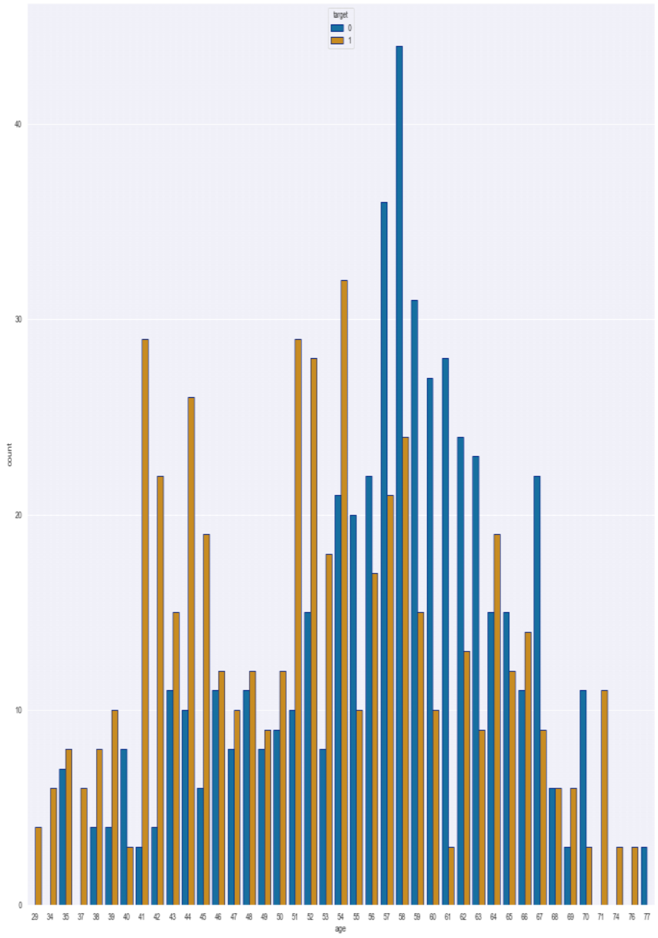


Fig 8 : Age vs Target

Now there is a histogram of all attributes with target as :

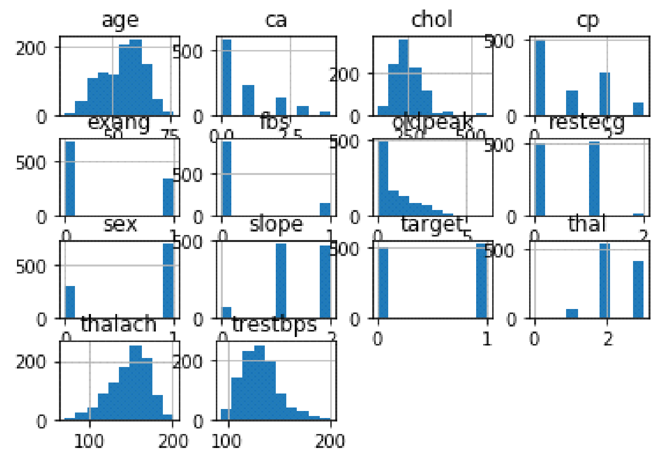


Fig 9 : histogram of All attributes

And there is another relation of target with all attributes

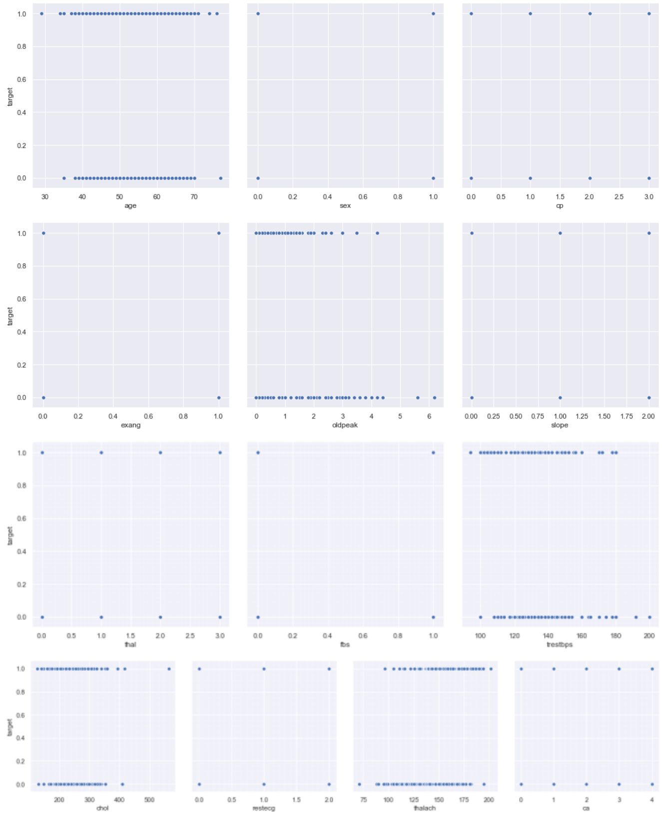


Fig 10 : Represention of relation of target with others attributes

3.2) Proposed model

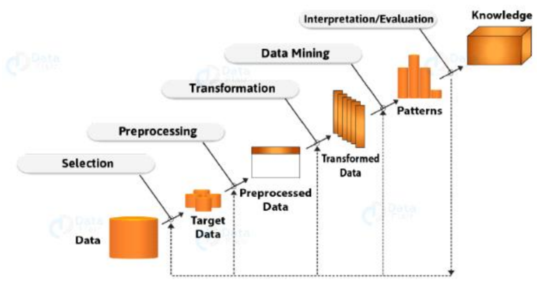


Fig 11 : Steps in Knowledge Discovery Process

IV. Literature Survey

There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centres.

K. Polaraju et al, proposed Prediction of Heart Disease using Multiple Regression Model and it proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms.

Marjia et al, developed heart disease prediction using KStar, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than KStar, Multilayer perception and J48 techniques using k-fold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease.

S. Seema et al, focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

Ashok Kumar Dwivedi et al, recommended different algorithms like Naive Bayes, Classification Tree, KNN, Logistic Regression, SVM and ANN. The Logistic Regression gives better accuracy compared to other algorithms.

MeghaShahi et al, suggested Heart Disease Prediction System using Data Mining Techniques. WEKA software used for automatic diagnosis of disease and to give qualities of services in healthcare centres. The paper used various algorithms like SVM, Naïve Bayes, Association rule, KNN, ANN, and Decision Tree. The paper recommended SVM is effective and provides more accuracy as compared with other data mining algorithms.

Chala Beyene et al, recommended Prediction and Analysis the occurrence of Heart Disease Using Data Mining Techniques. The main objective is to predict the occurrence of heart disease for early automatic diagnosis of the disease within result in short time. The proposed methodology is also critical in healthcare organisation with experts that have no more knowledge and skill. It uses different medical attributes such as blood sugar and heart rate, age, sex are some of the attributes are included to identify if the person has heart disease or not. Analyses of dataset are computed using WEKA software.

V. Observation

5.1 Correlation matrix : A correlation matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. A correlation matrix is used to summarize data, as an input into a more advanced analysis, and as a diagnostic for advanced analyses.

Fig x represents correlation of features with positive or negative values as:

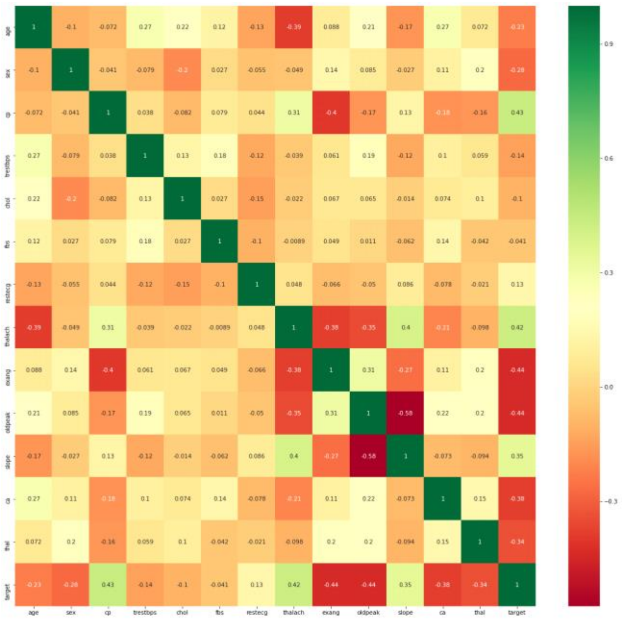


Fig 12 : Heatmap of Features

So, From above correlation matrix we conlude ,there is no features having very high corelation with target.

5.2 Confusion Matrix : a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix). Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa).

Table x represent graphical form of confusion matrix :
where , TP : True Positive, TN : True Negativity ,
FP : False Positive, FN : False Negativity.

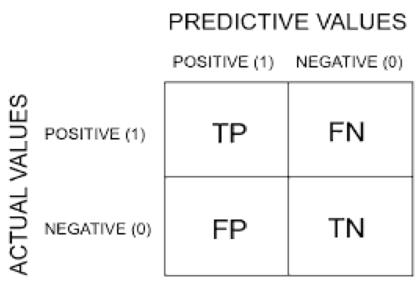


Table 1 : Graphical Representation of Confusion Matrix

5.2.1) Confusion matrix for Logistics regression:

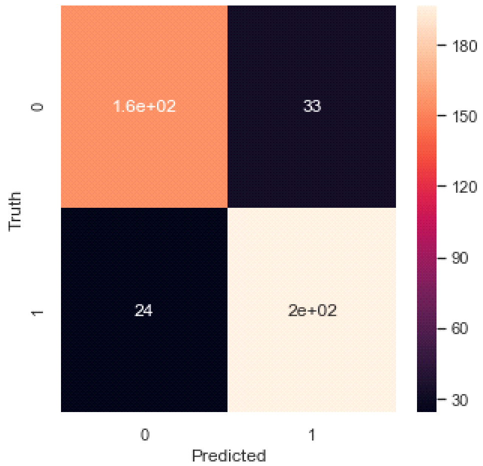


Table 2 : Confusion matrix for Logistics Regression

5.2.2) Confusion matrix for Decision Tree Classifier :

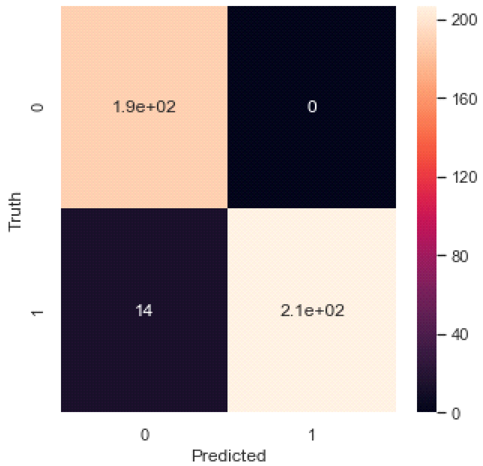


Table 3 : Confusion matrix for Decision Tree Classifier

5.2.3) Confusion matrix for Support Vector Machine:

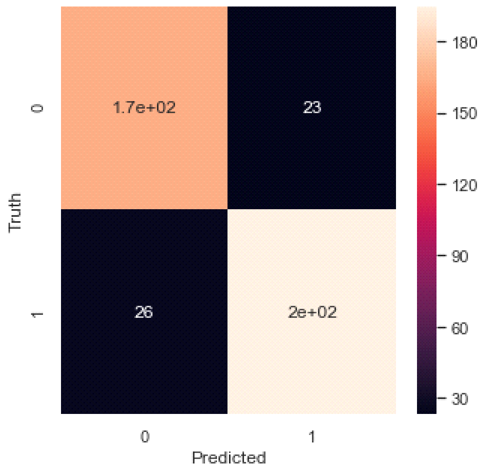


Table 4 : Confusion matrix for Support Vector Machine

5.2.4) Confusion matrix for Random Forest Classifier:

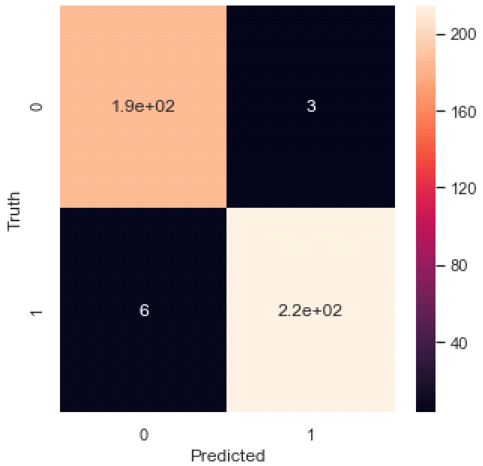


Table 5 : Confusion matrix for Random Forest Classifier

5.2.5) Confusion matrix for K-nearest Neighbors:

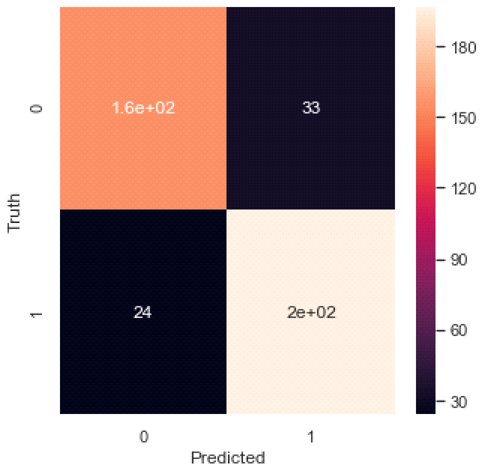


Table 6 : Confusion matrix for K-nearest neighbors

5.3 Accuracy Comparision : Accuracy is the ratio of total no. of correct prediction to the total no. of input samples. It can be the following equation :

Accuracy = (TP+TN)/(TP+TN+FP+FN)

Table 7 below shows the accuracy comparison of applied Algoritms :

Algorithms	Accuracy
K-nearest neighbors	86.09
Support Vector Machine	88.04
Logistics Regresion	86.09
Random Forest	96.82
Decision Tree	98.29

Table 7 : Accuracy Comparison

From the above table we conclude that Decision Tree has maximum accuracy and KNN & Logistics,Regression have minimum accuracy for this dataset.

5.4 Mean Squared Error : Mean Squared Error (MSE) of an estimator measures the average of the squares of the error i.e. the average squared difference between the estimated values and the actual value Table 8 below shows the mean squared error comparison of applied Algoritms :

Algorithms	Mean Sqaured Error
K-nearest neighbors	0.14
Support Vector Machine	0.12
Logistics Regresion	0.14
Random Forest	0.03
Decision Tree	0.02

Table 8 : Mean squared Error Comparison

From the above table we conclude that Logistics and KNN has maximum MSE and Decision Tree have minimum accuracy for this dataset.

5.5 Precision : precision is the fraction of relevant instances among the retrieved instances ,it is also called as positive predicted values.This shows how well the classifier handles the positive observation but not say much about negative ones. It is calculated as :

Precision = TP/(TP+FP)

Table 9 below shows the Precision comparison of applied Algoritms :

Algorithms	Not Caused(0)	Caused(1)	Average
K-nearest neighbors	0.87	0.86	0.86
Support Vector Machine	0.86	0.89	0.88
Logistics Regresion	0.87	0.87	0.86
Random Forest	0.96	0.97	0.97
Decision Tree	0.98	1	0.98

Table 9 : Precision Comparison

From the above table we conclude that Decision Tree Classifier has maximum Precision and KNN and Logistics Regression have minimum Precision for this dataset.

5.6 Recall : Recall is the fraction of the total amount of relevant instances that were actually retrieved.It is also known as sensitivity. It calculates how many of the actual positives or model capture through labeling it as positives. It is calculated as :

Recall = TP/(TP+FN)

Table 10 below shows the Recall comparison of applied Algoritms :

Algorithms	Not Caused(0)	Caused(1)	Average
K-nearest neighbors	0.83	0.89	0.86
Support Vector Machine	0.88	0.88	0.88
Logistics Regresion	0.83	0.89	0.86
Random Forest	0.97	0.97	0.97
Decision Tree	1	0.97	0.98

Table 10 : Recall Comparison

From the above table we conclude that Decision Tree Classifier has maximum Recall value and KNN and Logistics Regression have minimum Recall value for this dataset.

5.7 F-1 Score : F1 score is a measure of a test's accuracy. It considers both the precision p and the recall r of the test to compute the score.It is also known as F-score or F-measure. It is calculated as :

F-1 Score = 2*(Recall*Precision)/(Recall+Precision)

Table 11 below shows the F-1 Score comparison of applied Algoritms:

Algorithms	Not Caused(0)	Caused(1)	Average
K-nearest neighbors	0.85	0.87	0.86
Support Vector Machine	0.87	0.89	0.88
Logistics Regresion	0.85	0.87	0.86
Random Forest	0.97	0.97	0.97
Decision Tree	0.98	0.98	0.98

Table 11 : Recall Comparison

From the above table we conclude that Decision Tree Classifier has maximum F-1 Score and KNN and Logistics Regression have minimum F-1 Score for this dataset.

VI. RESULTS AND DISCUSSION

After implimenting the various model of machine learning classification algorithm ,it shows that the classification Accuracy ,Recall, Precision, F-1 Score. Table x shows that the Decision Tree Classifier algoritm is most precise and significant result comparison to the others algorithm.

Table 12 represents the summary of all observations :

Algorithms	Accuracy	Precision	Recall	F-1 Score
K-nearest neighbors	86.09	0.86	0.86	0.86
Support Vector Machine	88.04	0.88	0.88	0.88
Logistics Regresion	86.09	0.86	0.86	0.86
Random Forest	96.82	0.97	0.97	0.97
Decision Tree	98.29	0.98	0.98	0.98

Table 12 : Comparitive analysis

Graphical represntion of all agorithm is given as:
Figure 1 represents bar plot of Accuracy Comparision, and Decision Tree has maximum accuracy with 98.29 % .

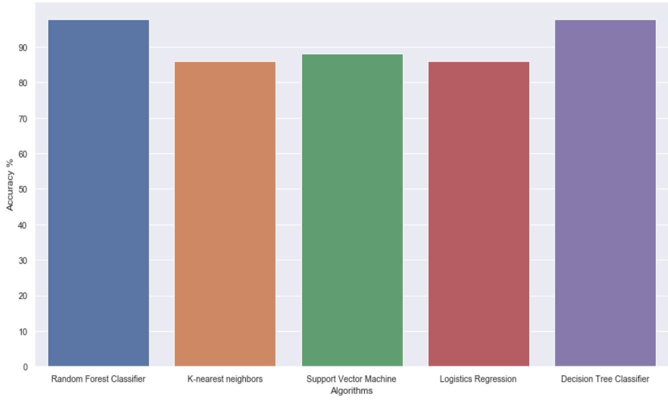


Fig 1 : Graphical represention of Accuracy

Figure 2 represents bar plot of Mean Squared Error Comparision,and Decision Tree ha minimum MSE.

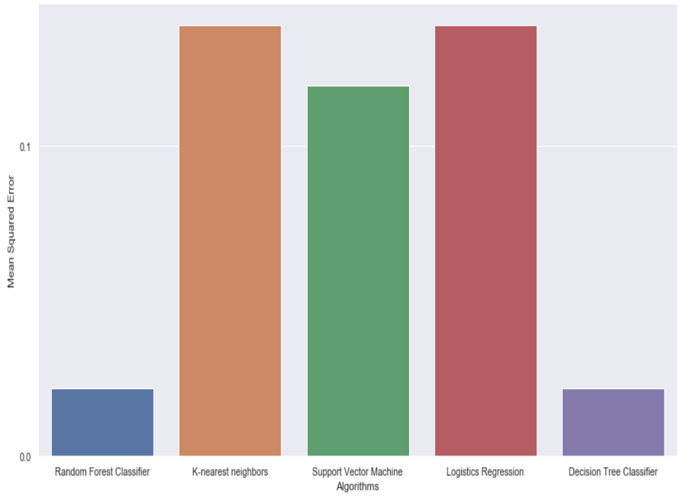


Fig 2 : Graphical represention of MSE

Figure 3 represents bar plot of Recall Comparision,

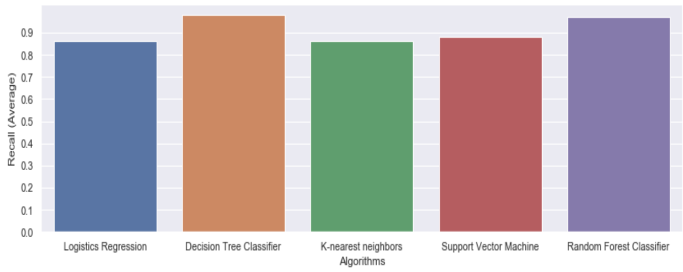


Fig 3 : Graphical represention of Recall

Figure 4 represents bar plot of F-1 Score Comparision,

Fig 5 : Graphical represention of F-1 Score

Hence from above all comparision we can conclude Decision tree is best because of high accuracy, high precision ,low MSE and KNN & Logistics Regression has less accuracy,less Precision and high MSE as comparison to others classification algoritms

VII. CONCLUSION

Heart Disease is one of the major concerns for society today.

It is difficult to manually determine the odds of getting heart disease based on risk factors. However, machine learning techniques are useful to predict the output from existing data. So,this is most critical prediction that's why prediction should carried out with larger datasetto make better accuracy.

VIII. LIMITATION AND FUTURE WORK

We present a heart disease prediction use case showing how synthetic data can be used to address privacy concerns and overcome constraints inherent in small medical research data sets. While advanced machine learning algorithms, such as neural networks models, can be implemented to improve prediction accuracy, these require very large data sets which are often not available in medical or clinical research. We examine the use of surrogate data sets comprised of synthetic observations for modeling heart disease prediction. We generate surrogate data, based on the characteristics of original observations, and compare prediction accuracy results achieved from traditional machine learning models using both the original observations and the synthetic data. We also use a large surrogate data set to build a neural network model (Perceptron) and compare the prediction results to the traditional machine learning algorithms (Logistic Regression, Decision Tree and Random Forest).

The use of surrogate data to train machine and deep learning models in this study appears to be an effective step towards improving heart disease prediction, but with limitations. The origin of the patient data used in this study was from the Cleveland, Ohio area and is not geographically diverse. From a global perspective, each geographical region has its own characteristic diet, lifestyle and availability of healthcare resources. According to the World Health Organization, in 2017, Turkmenistan had the highest death rate due to cardiovascular disease at 411.1 deaths per 100,000 while South Korea has the lowest at 30.76. Application of surrogate data in heart disease prediction using machine or deep learning techniques with patient data from geographically diverse sources would be interesting to explore in the future. By doing so, we can potentially discover data patterns that are not easily seen by using a geographically focused patient data set such as the Cleveland data set

IX. Acknowledgement

I have completed this work under the guidance of Dr. Pankaj Agrawal(Professor & head) & Ms. Sapna Yadav (Assistant Professor), Department of Computer Science and Engineering at IMS Engineering College,Ghaziabad. I am doing online summer internship on Machine Learning where I have learnt various machine learning algorithm from both of my mentors as a course instructors. This work has been assigned as a project assignment to us.

I would like to express my special thanks both of my mentors for inspiring us to complete work and write paper. Without their active guidance,help cooperation & encouragement, I would not leadway in writing the paper. I am extremely thankful for their valuable guidance and support on completetion of this paper.

I extend by gratitude to "IMS Engineering College" for giving me this oppurtunity. I also acknowledge with deep sense of reverence, my gratitude towards my friends, parents and member of my family, who always supported me morally, mentally as well as economically.

Any omission in this brief acknowledgement does not mean lack of gratitude.

X. REFERENCES

[1]. Nidhi Bhatla, and Kiran Jyoti, “An Analysis of Heart Disease Prediction using Different Data Mining Techniques” , International Journal of Engineering Research & Technology (IJERT), Vol. 1, Oct.2012.

[2]. Chaitrali S.Dangare, and Sulabha S.Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications (0975 – 888), Vol. 47, No.10, June.2102.

[3]. Heart disease webpage on MAYO CLINIC [Online]. Available: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118 , 2019.

[4]. Cardiovascular disease webpage on WHO [Online]. Available: https://www.who.int/cardiovascular_diseases/en/ , 2019.

[5]. Dr. T. Karthikeyan, and V.A.Kanimozhi, “Deep Learning Approach for Prediction of Heart Disease Using Data mining Classification Algorithm Deep Belief Network”,International Journal of Advanced Research in Science, Engineering and Technology, Vol. 4, Issue 1, January 2017.

[6]. Hlaudi Daniel Masethe, and Mosima Anna Masethe, “Prediction of Heart Disease Using Classification Algorithms” , in Proceedings of the World Congress on Engineering and Computer Science 2014 Vol. II WCECS 2014, 22-24 Oct. 2014, San Francisco, USA.

[7]. Asha Rajkumar, and Mrs. G. SophiaReena, “Diagnosis of Heart Disease Using Data Mining Algorithm”, Global Journal of Computer Science and Technology, Vol. 10, pp. 38-43, Sept. 2010.

[8]. Sarangam Kodati, and Dr. R Vivekanandam, “A Comparative Study on Open Source Data

Mining Tool for Heart Disease" , International Journal of Innovations & Advancement in Computer Science, Vol. 7, Issue 3, March-2018.

[9]. B.Venkatalakshmi, and M.V Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3, Special Issue 3, March-2014.

[10]. Abhishek Taneja, " Heart Disease Prediction System Using Data Mining Techniques", Oriental Journal Of Computer Science and Technology, Vol. 6, pp. 457-466, Dec.2013.

About Author

Abhay Agrahary is B.Tech 2nd year student in Department of Computer Science & Engineering College, Ghaziabad ,U.P., India. He is done analyst intern from Edulyt india with successful work on three project which are Website blocker, Interactive Dictionory, Credit Card approval prediction. His area of intrest are Web Devlopment, Competitive Programming, Machine Learning, Deep Learning.

Cite this article as:

Abhay Agrahary, "Heart Disease Prediction Using Machine Learning Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 4, pp.137-149, July-August-2020. Available at doi : https://doi.org/10.32628/CSEIT206421

Journal URL : http://ijsrcseit.com/CSEIT206421