International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Topic-Machine Learning Bases Cardiac Arrest Detection

Omkar Sawant[1], Siddhi Thakur[2], Dr Jayant Nandwalkar[3]

[1,2]Student, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India
[3]Professor, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

Abstract
Coronary heart disease is the principal reason for deaths internationally. To present treatment for coronary heart ailment, lots of advanced technology is used. In scientific center it is the most common place hassle that a lot of medical persons do now not have same information and information to treat their patient so they deduce their personal decision and as an end result it shows poor final results and sometimes results in death. To triumph over those problems predictions of heart disorder using device gaining knowledge of algorithms and facts mining strategies, it emerges as smooth to automatic prognosis in hospitals as they are gambling important role in this regard. Heart disorders can be expected by appearing analysis on a patient's extraordinary health parameters. There are one of a kind algorithm to are expecting heart disorder like naïve Bayes, ok Nearest Neighbor (KNN), selection tree, artificial Neural community (ANN). We've used exceptional parameters to expect heart disorder. Those parameters are Age, Gender, Cerebral palsy (CP), Gender, Cerebral palsy (CP), Blood stress (bp), Fasting blood sugar test (fbs)and so on. In our research paper, we used to build in dataset. We have implemented the five unique techniques with equal dataset to expect heart disorder these applied algorithm are Naive Bayes, k Nearest Neighbor (KNN), decision tree, artificial Neural network (ANN), Random woodland. This paper investigates that which approach gives extra accuracy in predicting coronary heart disease based on fitness parameters.
Experiment shows accuracy of Logistic Regression 85%. K. Nearest Neighbour 84

Keywords: Naive Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN), Random Forest, Heart Disease

Introduction: Coronary heart disease, Cardiomyopathy and Cardiovascular disease are the categories of heart disease. The word "heart disease" includes a variety of conditions that affect the heart and blood vessels and how the fluid gets into the bloodstream and circulates there in the body. Cardiovascular disease (CVD) causes many diseases, disability and death. Diagnosis of the disease is important and complex work in medicine.

Medical diagnosis is considered as a crucial but difficult task to be done efficiently and effectively. The automation of this task is very helpful. Unfortunately, all physicians are not experts in any subject specialists and beyond the scarcity of resources there some places. Data mining can be used to find hidden patterns and knowledge that may contribute to successful decision making. This plays a key role for healthcare professionals in making accurate decisions and providing quality services to the public.

The approach provided by the healthcare organization to professionals who do not have more knowledge and skills is also very important. One of the main limitations of existing methods is the ability to draw accurate conclusions as needed. In our approach, we are using different data mining techniques and machine learning algorithms , Naïve Bayes , k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN). Random Forest to predict the heart disease based on some health parameters

Related Work: In Paper [1], makes use of the statistics from UCI records repository. advise heart disease prediction using KStar, J48, SMO, and Bayes net and Multilayer perceptron using WEKA software program. based on overall performance from distinct issue SMO (89% of accuracy) and Bayes net (87% of accuracy) achieves optimal performance than KStar, Multilayer perceptron and J48 techniques using k-fold pass validation. The accuracy performance achieved by those algorithms is still not great. In Paper [2] they use facts from Kaggle endorse software of know-how coming across procedure on prediction of stroke patients based on artificial Neural network (ANN) and help Vector gadget (SVM), which give accuracy of eighty-one.82% and 80.38% for ANN and SVM respectively for training statistics set and eighty-five. 9% and 84.26% for synthetic Neural network (ANN) and assist Vector gadget (SVM) in take a look at dataset respectively. Paper [3] uses records from UCI repository and compares overall performance of different device learning algorithms using Naive Bayes, KNN, selection Tree, ANN. amongst them ANN gave the very best accuracy of eighty-five.3%. whilst Naïve Bayes and KNN gave almost 78% and Selection Tree gave eighty %. Paper [4] use WEKA tool for measuring performance of different gadget mastering algorithms. ANN with PCA changed into used to hurry the overall performance.

Methodology: The principal reason of the proposed technique is to expect the incidence of heart ailment for early detection of the sickness in a short time. In our approach, we're the use of one-of-a-kind facts mining techniques and machine gaining knowledge of algorithms, Naïve Bayes, okay Nearest Neighbor (KNN), decision tree, artificial Neural network (ANN), Random woodland to predict the coronary heart disease based on a few health parameters.

Data is analyzed using Google Colab. It is an open-source software where we can implement multiple machine learning algorithms by importing libraries. We can also download the needed libraries by anaconda prompt. It allows us to create live code, perform visualizations, process data and plot graphs.

Dataset for implementation: We have used a built-in dataset from Kaggle Machine learning repository for predicting heart disease. We have used 10 parameters

1)Age

2)Sex

3)Cerebral palsy/chest pain (CP)

4)Blood Pressure(bps) in mmHG

5)Cholesterol in mg

6)Fasting blood sugar test (fbs)

7) resting electrocardiographic results

8) thalach (Maximum heart rate achieved)

9) exang (Exercise induced Angina)

10) oldpeak (ST depression induced by exercise relative to rest)

11) ca(Number of major vessels)

Data splitting: Data is splitted into training and testing data. 25 % data is used for testing purposes while 75 % data is used for training purpose. We performed data normalization for removing Nan values.

Data Visualization:

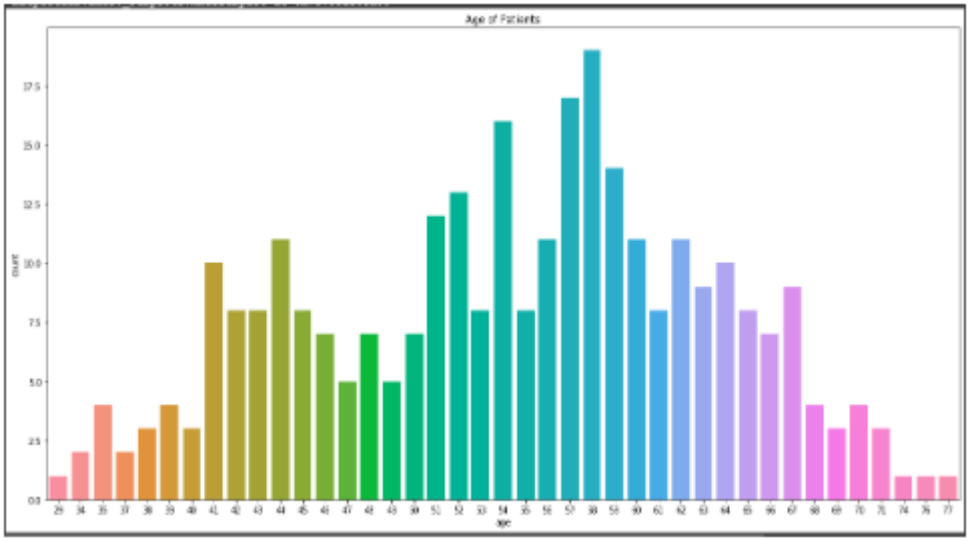

Fig-1

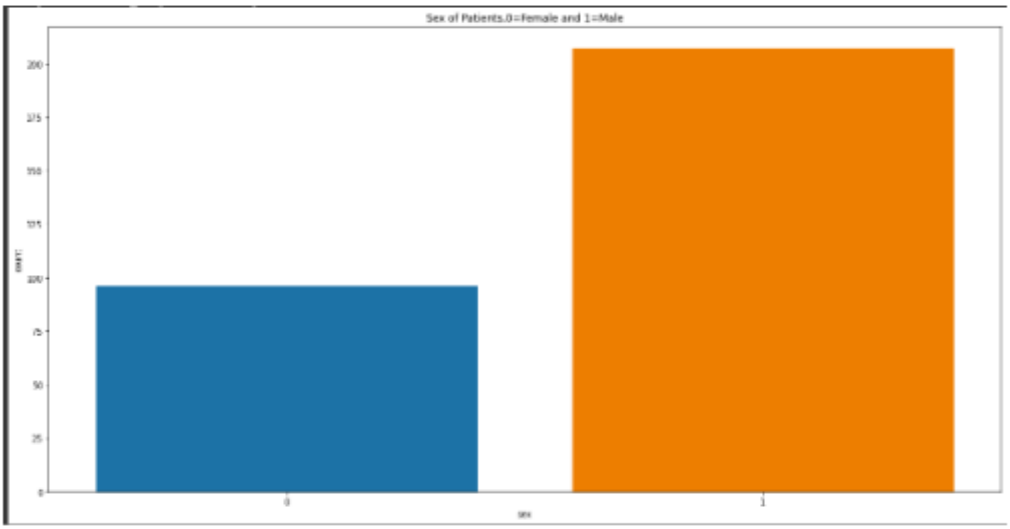

Fig-2

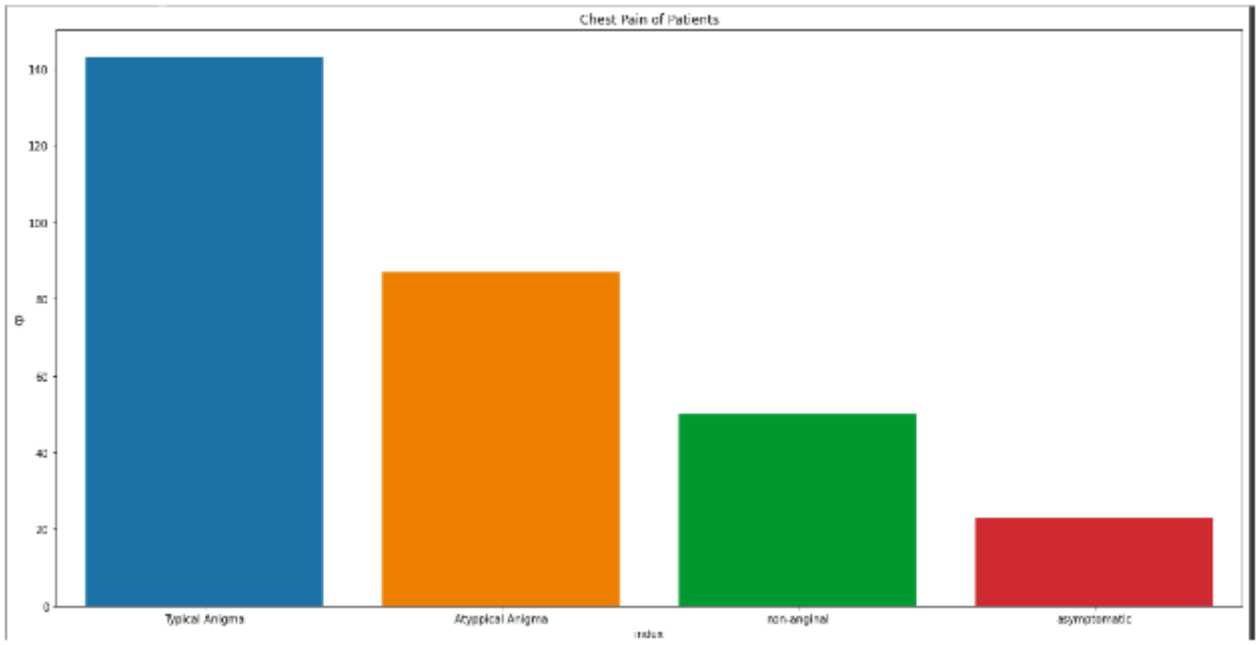

Fig-3

The performance and the accuracy of each experiment is evaluated by standard metrics such as TP rate, TN rate, precision, recall and F-measure which are calculated by Confusion Matrix which is known as a predictive classification table. All these measures will be used to compare the performance of these selected and implemented algorithms.

Heat Map Classification: On this method, the correlation between variables will be tested, that allows you to be used as a basis for evaluation to be expecting cardiovascular ailment. Based totally on the matrix it is located that the variables caused angina (exang), chest pain type (cp), ST depression brought about by way of exercise relative to relaxation (oldpeak), maximal heart charge (thalac) had a robust correlation with the target variable. In the meantime, blood sugar (fbs) and ldl cholesterol (chol) levels do no longer correlate with the target variable. Meanwhile, among the impartial variables, there is a robust correlation between the slope and oldpeak variables. except, thalac, exhang, oldpeak, and slope variables also are strongly correlated. strong correlation also applies to variables Exang, cp, and thalac. It proves that there is no multicollinearity in the courting between variables where every impartial variable does no longer correlate with each other. records that have been imported will be taken as many as 293 random facts as a basis for analysis. The records are split into educated information and take a look at facts. The facts on the next web page are statistics from train_data and test_data so that it will be used in this. Have a look at. The schooling statistics is used to build.a logistic regression version using the glm () function due to the fact logistic regression is protected in the generalized linear version with binomial kind families. Primarily based on the effects of using the logistic regression method, it's far expected that the sex, cp, trestbps, restecg, ca and that variables have an effect on the target variable at an alpha cost of five% extensively. The chosen variables are the variables that appreciably affect the goal variable. In logistic regression, the impact of each variable at the target variable can be seen from the chances ratio cost. as an instance, for the intercourse variable having a coefficient fee of -1.547601 with a reference class with a male fee, the odds ratio value is four.2655 because of this that for male patients, the percentages of having coronary heart disorder are four.2655 times the woman odds or it

could be said the tendency of guys to heart sickness is better than women. For the trestbps variable with a coefficient value of -0.029713, it's miles located that the percentages ratio price is zero.0822 which means that that for the trestbps variable there can be a big boom whilst trestbps enters the price 0.0822 mmHg. However, the thalach variable with a coefficient of 0.032028 may have an extraordinary of 0.08856 which means that at that fee there can be a full-size change inside the overall performance of the coronary heart price or cardiovascular rate. The exang1 variable is exercising-precipitated angina with an estimated coefficient of -1.05855 so that the exang variable with a reference fee of 1 will have an unusual of 2.92710 because of this that if the cost is achieved then cardiovascular overall performance will lower. subsequent is the variable ca with reference ca values 1, 2, and 3. Ca1 with an predicted coefficient of - 1.430110 will have odds of 3.955, even as ca2 with an envisioned ratio of - three.329874 can have odds of 9.1777 and ca3 with an predicted aspect of -zero.553711 could have odds in the amount of 1.5261. It proves that once the number of fluoroscopy vessels reaches its price odds, this may have an impact on reducing cardiac performance with a purpose to affect the expanded ability for cardiovascular disease. except, the composition of value 0 and cost 1 on variable target is ninety seven:116, which is still pretty balance, so the end result could be reliable and

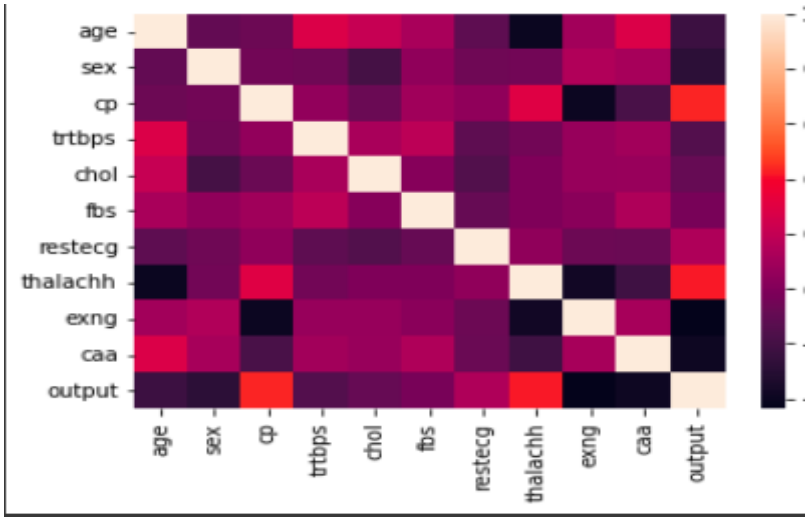

Fig-4

ALGORITHMS USED FOR EXPERIMENTS

1)Logistic Regression: Logistic regression is a method of modeling the chance of a discrete final result given an input variable. The maximum not unusual logistic regression fashions binary final results; something which can take values including actual/fake, sure/no, and so on. Multinomial logistic regression can version scenarios in which there are more than possible discrete results. Logistic regression is a beneficial analysis method for classification issues, wherein you are attempting to decide if a new pattern suits satisfactory into a category. As aspects of cyber safety are class troubles, inclusive of attack detection, logistic regression is a useful analytic approach.

Logistic regression, in spite of its name, is a classification version in preference to the regression model. Logistic regression is a simple and more efficient technique for binary and linear class troubles. it's far from a type version, which could be very smooth to realize and achieves very good performance with

linearly separable training. It's miles and a considerably employed set of rules for classification in enterprise. The logistic regression model, like the Adaline and perceptron, is a statistical technique for binary categories that can be generalized to multiclass type. Scikit-learn has a fantastically optimized model of logistic regression implementation, which supports multiclass category assignment.

	output
0	0.914529
1	0.914529
2	0.914529
3	0.914529
4	0.914529
...	...
298	-1.093459
299	-1.093459
300	-1.093459
301	-1.093459
302	-1.093459

303 rows × 1 columns

Table-1

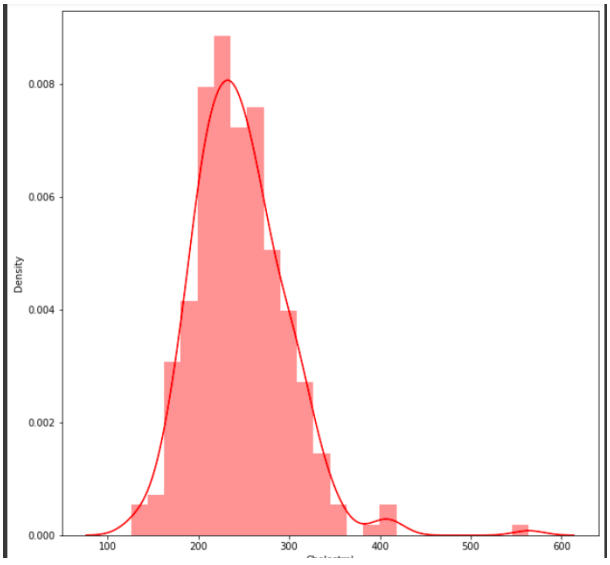

Fig-5

```
Y_pred1 = logreg.predict(x_test)
lr_conf_matrix = confusion_matrix(encoded_ytest,Y_pred1 )
lr_acc_score = accuracy_score(encoded_ytest, Y_pred1)

lr_conf_matrix

array([[35,  9],
       [ 4, 43]])

print(lr_acc_score*100,"%")

85.71428571428571 %
```

Fig-6

This matrix shows that there are 35 true negative rates,9 false positive,4 false negative while 43 true positive

2)K Nearest Mean:KNN is the device studying a set of rules and is the maximum usually used set of rules .It is preferred while parameters are non-stop. In KNN, classification is accomplished through predicting the closest neighbor .It is favored over other category algorithms because of its simplicity and excessive speed .It can be used to resolve each type and regression problem. The set of rules takes the heart disease facts set and classifies whether a person has heart sickness or not. KNN captures the idea with the aid of calculating the gap among points on a graph. We used KNN to categorize and predict humans with coronary heart sickness based totally on parameters such as age, intercourse and so forth. It does not want education data for model technology because the training.information is used in trying out levels. It shops all of the cases and then classifies new records consistent with the nearest neighbor.
KNN has two stages:
1. locate the k variety of instances inside the dataset
2. Use the k instances to discover the nearest neighbor.

```
plt.figure(figsize=(10,6))
plt.plot(range(1,40),error_rate,color='blue', linestyle='dashed', marker='o',
         markerfacecolor='red', markersize=10)
plt.xlabel('K Vlaue')
plt.ylabel('Error rate')
plt.title('To check the correct value of k')
plt.show()
```

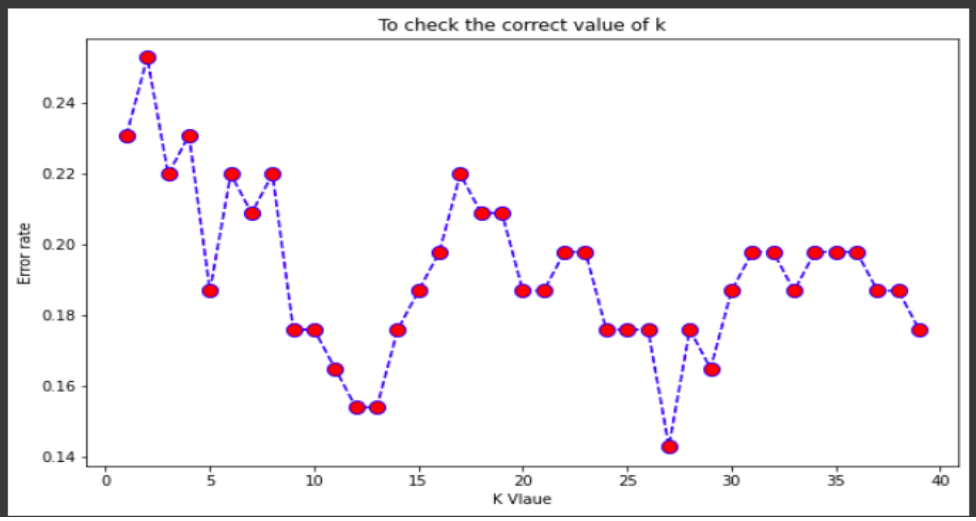

Fig-7

```
knn_conf_matrix

array([[35,  9],
       [ 5, 42]])

print(knn_acc_score*100,"%")

84.61538461538461 %
```

Fig-8

The matrix shows there are 35 true negative rates,9 false positive,5 false negative while 42 true positive cases

3)Random Forest: Random forest is likewise a form of supervised gaining knowledge of. it can be used for type and regression. It's also the maximum bendy and consumer friendly algorithm. A forest is made up of bushes. It is said that the more bushes it has, the stronger a wooded area is. Random forests create choice bushes on randomly selected information samples, acquire predictions from each tree, and select the first-class solution by vote casting.

```
svm_conf_matrix

array([[34, 10],
       [ 8, 39]])

print(svm_acc_score*100,"%")

80.21978021978022 %
```

Fig-9

The matrix shows that it has 42 true negative rates,10 false positive,8 false negative while true positive are 39

4) Support Vector Machine: Coronary heart failure (HF) is a modern syndrome that marks the cease-degree of coronary heart illnesses, and it has an excessive mortality rate and massive value burden. In particular, non-adherence of medication in HF patients might also result in severe outcomes including hospital readmission and death. This study objectives to pick out predictors of drugs adherence in HF sufferers. in this work, we implemented a aid Vector device (SVM), a machine-learning approach useful for facts category.SVM modeling is a promising class method for predicting medicinal drug adherence in HF sufferers. This predictive version allows stratification of the patients in order that evidence-based decisions may be made and sufferers managed accurately. further, this technique should be similarly explored in different complex illnesses the use of different common variables.

```
[ ] svm_conf_matrix = confusion_matrix(encoded_ytest,ypred5)
    svm_acc_score = accuracy_score(encoded_ytest, ypred5)

    svm_conf_matrix

    array([[34, 10],
           [ 8, 39]])

[ ] print(svm_acc_score*100,"%")

    80.21978021978022 %
```

Fig-10

The matrix shows that it has 34 true negative rates,10 false positive,8 false negative while true positive are 39.

5)Decision Tree:A decision tree is a type of a supervised studying algorithm classifier, which is simple to understand. They address numerical and categorical statistics. The decision tree resembles the tree structure consisting of internal nodes, branches, and leaf nodes in which every department represents the values of a given records set, internal nodes checks on a given attribute and the Leaf nodes show the class to be expecting or indicate the consequences of the result. The type rule begins from the root node to the leaf nodes, depending on the predictive characteristic and the given policies.

```
tree_conf_matrix

array([[27, 17],
       [10, 37]])

print(tree_acc_score*100,"%")

70.32967032967034 %
```

Fig-11

The matrix shows that it has 27 true negative rates,17 false positive,10 false negative while true positive are 37.

Comparison of implemented algorithm: The purpose of our have a look at became to use gadget learning algorithms for coronary heart disorder in healthcare. So for this we executed a test via the use of distinctive algorithms on heart sickness sufferers. Through implementation we will realize which type of set of rules is excellent for predicting heart sickness.
After the implementation of different algorithms the second step is the assessment among unique gadget learning algorithms used in these experiments and pick the first-class one that gives the most accuracy.
In order to do comparison of these experiments different performance measures are used for example, Accuracy True Positive, False Positive, False Negative,

Discussion: The point of interest of our take a look at changed into on the usage of records mining strategies in healthcare for heart ailment.

We accomplished a few experiments on our records set of coronary heart disorder by applying five data mining algorithms. via implementation of various type algorithms we try to discover which algorithm is high-quality in predicting coronary heart ailment .And which one offers high-quality accuracy. There are five experiments we achieved and these experiments are designed for the same reason, the cause is to evaluate the consequences of KNN, Neural Networks, Decision Tree, Naive Bayes and Random Forest

ALGORITHM	ACCURACY	TN	FP	FN	TP
LOGISTIC REGR	86%	35	11	4	43
KNN	85%	35	9	5	42
RANDOM FOREST	80%	42	10	8	39
SVM	80%	34	10	8	39
DECISION TREE	70%	27	17	10	37

Table-2

The table shows accuracy on the given dataset is 86% and lowest accuracy of 70%. The Logistic Regression has the highest accuracy while the Decision Tree has the lowest accuracy.By observing other performance measures that are used for result too. TP rate of KNN is 35, Logistic is 35, SVM is 34, Decision tree is 27, and Random forest is 42. This shows that the Logistic has the highest TP rate and Decision Tree has the lowest TP rate. Similarly Decision tree has the highest FP rate of 17 and the KNN has the lowest FP rate of 9.Based on the above comparison it can be seen that Logistic Regression andKNN are good as they have nearly same accuracy and they have the best TP rate and KNN has the FP rate of 9 followed by Logistic Regression with FP rate of 9 .As we all know that Heart Disease is a sensitive and critical disease which causes millions of death. Due to this we need to keep TP rate high and FP rate less. If the diagnosis of disease is correct and also done earlier then it is best to cure patients suffering from that disease. So it is expected from algorithms to perform well. Accuracy also matters to identify heart patients.

Conclusion: Our study mainly focused on the use of data mining techniques in healthcare especially in detection of heart disease. Heart disease is a fatal disease which may cause death. Data mining techniques were implemented using the following algorithm, KNN, Neural Networks, Decision Tree, and Naive Bayes and Random Forest. We measured performance on the basis of Accuracy, TN, FP, FN and TP rate and in some algorithms.

We conducted five experiments with the same data set to predict heart disease. The result of all the implemented algorithm are shown in tabular form for better understanding and comparisons .The experiment shows that Logistic Regression gives the highest accuracy which is 86% followed by KNN with accuracy of 85% . Our findings indicate that data mining can be used and applied in the healthcare industry to predict and diagnose the disease at early stages,

Future Works: The purpose of our have a look at became to use gadget learning algorithms for coronary heart disorder in healthcare. So for this we executed a test via the use of distinctive algorithms on heart sickness sufferers. Through implementation we will realize which type of set of rules is excellent for predicting heart sickness.
After the implementation of different algorithms the second step is the assessment among unique gadget learning algorithms used in these experiments and pick the first-class one that gives the most accuracy.

Further research has to be carried out to grow category accuracy through using superior algorithms consisting of Bagging, ANN or table decision. determine the performance of the predictions per set of rules and apply the proposed device to the region of interest. we will upload more functions to enhance accuracy implementation of algorithms. Stakeholders need to use it as a committed device to make better choices. We did not trade parameters in our implementation. In future, it may be stepped forward and adjusted by way of converting the parameters for the test.
inside the destiny, more work may be finished with the aid of the use of more data associated with coronary heart sickness and by way of the usage of extraordinary facts discount strategies. For higher effects and predictions of coronary heart sickness, high quality orientated datasets can be used that are unfastened from inconsistencies.

Reference:

[1] Ujma Ansari, Jyoti Soni, Dipesh Sharma, Sunita Soni. “Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction”, 258493784_Predictive_Data_Mining_for_Medical_Diagnosis_An_Overview_of _Heart_Disease_Prediction. March 2011Data Mining in Healthcare for Heart Diseases

[2] C. Beyene, P. Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques”, https://www.researchgate.net/publication/323277772_Survey_on_prediction_and _analysis_the_occurrence_of_heart_disease_using_data_mining_techniques, 118(8):165-173 · January 2018

[3] Muhammad Usama Riaz, SHAHID MEHMOOD AWAN, ABDUL GHAFFAR KHAN, “PREDICTION OF HEART DISEASE USING ARTIFICIAL NEURAL NETWORK”,

https://www.researchgate.net/publication /328630348_PREDICTION_OF_HEART_DISEASE_USING_ARTIFICIAL_N EURAL_NETWORK. October 2018

[4] Umair Shafique, Irfan Ul Mustafa, Haseeb Qaiser, Fiaz Majeed, "Data Mining in Healthcare for Heart Diseases",https://www.researchgate.net/ publication/274718934_Data_Mining_in_Healthcare_for_Heart_Diseases. March 2015.

[5] Hossam Meshref, "Cardiovascular Disease Diagnosis: A Machine Learning Interpretation Approach",https://www.researchgate.net /publication/338428682_Cardiovascular_Disease_Diagnosis_A_Machine_Learni ng_Interpretation_Approach, January 2019

[7] Jabbar Akhi,Shirina Samreen
"Heart disease prediction system basedon hidden naïve Bayes classifier ",
https://www.researchgate.net/publication/309735105_Heart_disease_prediction_system_based_on_hidden_naive_Bayes_classifier, October 2016.