Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7923038

CHEST PAIN BASED HEARTDISEASEDIAGNOSIS AND PREDICTIONUSINGSTACKING CLASSIFIER.

MRS. JALPA SHAH[1], DR. CHINTAN SHAH[2],DR. PREMAL PATEL[3]

PhD Scholar[1] , Associate Professor[2,3],
Department of Computer Engineering[1,2,3]
College of Technology[1,2,3]
Shahjalpa.comp@gmail.com[1], chintan.shah84@gmail.com[2,]
premalpatel.ce@socet.edu.in[3]
Silver Oak University, Ahmedabad, Gujarat[1,2,3].

ABSTRACT

Heart Disease is one of the foremost common diseases nowadays, and for people that provide health care, it'svery necessary to figure out with them to require care of their patients' health and save their life. during thispaper, different classifiers were analyzed by performance comparison to classify the guts Disease dataset toclassifyitcorrectlyandortoPredictheartconditioncaseswithminimalattributes.Largeamountsofknowledgethatcontainsomesecretinformationwerecollectedbythehealthcareindustries.Thisdatacollection is beneficial for creating effective decisions. during this case, aHeart Disease Prediction System(HDPS) is developed using Logistic Regression, K Nearest Neighbor, Decision Tree, Random Forest Classifier,andSupport VectorMachine algorithmsto predict the gutsdisease risklevel.

Keywords: HeartDisease,Support VectorMachine,Accuracy,DecisionTree.

I. INTRODUCTION

The most vital and essential part (organ) of the mortal body is the Heart. There are numerous conditions thatare linked to the heart so the analysis of vaticination of the heart must be accurate. To resolve this, virtual studyabout this field is obligatory. Typically, these conditions are prognosticated at the end stage and this is the mainreason for the death of heart cases due to insufficiency of correctness because of this there's a need to identifycomplete algorithms for conditions vaticination. One of the complete able and effective technologies in MachineLiteracy that have established on specific training and testing with the support of python and python libraries.The system acquires training directly from data and skill, grounded on this training, testing should be done oncolorful types of need as per needful algorithms. For Testing and Training, Machine literacy can be used in aneffectivefashion.Disease-relatedtotheguts,alsoappertainedtoascomplaint(CVD),discussescolorfulconditionsthataffectthegutsnotjustthecomplaint.Heartcomplaintsofthecoronaryhighways,cardiomyopathy, and cardiovascular health issues are certain services where the blood is pumped and itsrotation is formed throughout the body. The opinion is an important task that has to be performed efficiently.This is substantially done under a croaker's guidance. This causes wrong results & inordinate medical costs oftreatments handed to cases. So, we conclude that an automatic

opinion and vaticination system would proveextremelyfavorable.

Nowadays, the lifespan of a human being is reduced because of heart diseases. So, World Health Organization(WHO) developed targets for prevention of non-communicable diseases (NCDs) in 2013, in which, 25% ofrelative reduction is from cardiovascular diseases and it is being ensured that at least 50% of patients withcardiovascular diseases have access to relevant drugs and medical counseling by 2025. Around 17.9 millionpeople died just because of cardiovascular diseases in 2016, which is 31% of deaths around the world. A majorchallenge in heart disease is its detection. It is difficult to predict whether a person has heart disease or not.There are instruments available that can predict heart diseases but either they are expensive or are not efficienttocalculatethechanceof heartdiseaseinhumans.

A check of the World Health Organization (WHO) says that medical professionals can prognosticate just 67 ofheart complaints, so there's a vast compass of exploration in this field. In the case of India, access to goodcroakersandhospitalsinpastoralareasisveritablylow.2016WHOreportsaysthatjust58ofthe croakershave a medical degree in civic areas and 19 in pastoral areas. Heart conditions are a major challenge in medicalwisdom, Machine Literacy could be a good choice for prognosticating any heart complaint in humans. Heartconditionscan beprognosticatedusingNeuralNetwork,DecisionTree,KNN,etc.

There are many types of Heart disease types based on CP values.

1. Typical Angina is a type of chest pain caused by reduced blood flow to the heart.

2.Atypical Angina pectoris which does not have associated classical symptoms of chestpain.

3.Non Angina or Noncardiac chest pain is often described as feeling like angina, the chest pain caused by heart disease. It feels like a painful squeezing or tightness in your chest, or like pressure or heaviness, particularly behind your sternum. You may feel it on the right side or the left side or in the middle

4. Asymptomatic if a patient tests as a carrier for a disease or infection but experiences no symptoms.

II.METHODOLOGY

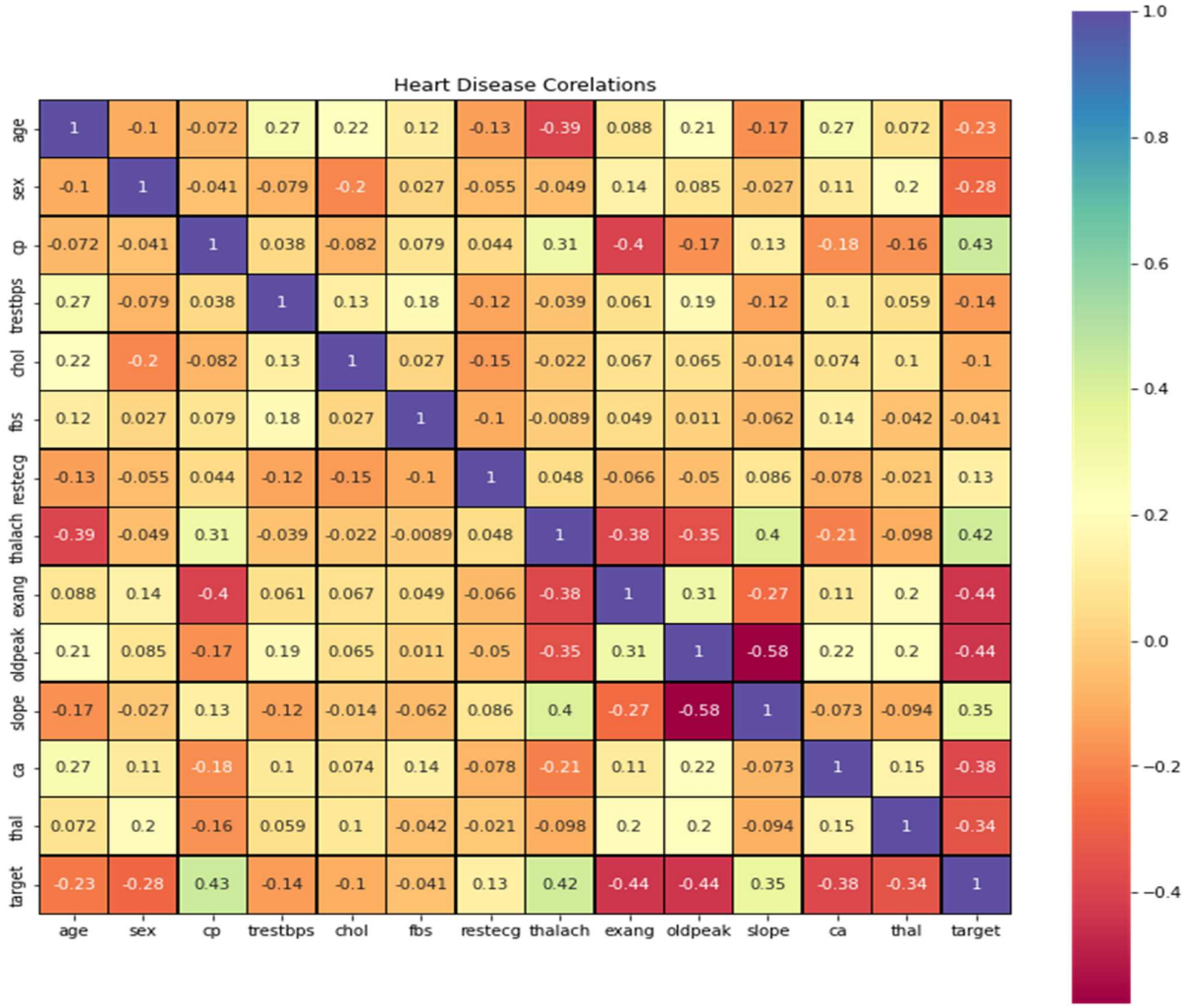
Preface World Health Organization has estimated 12 million deaths do worldwide; every time due to Heartconditions.HalfthedeathsintheUnitedStatesandotheradvancedcountriesareduetocardiovascularconditions. The early prognostic of cardiovascular conditions can prop in making opinions on life changes inhigh threat cases and turn reduce the complications. This exploration intends to pinpoint the most applicable/threat factors of heart complaint as well as prognosticate the overall threat using logistic retrogression DataPreparation.Stacking Classifierisatypeofretrogressionanalysisinstatisticsusedforvaticinationofoutgrowthofacategoricaldependentvariablefromasetofthepredictor.InStacking Classifier,thedependent variable isalwaysdouble. Logisticretrogression issubstantiallyusedfor vaticination and alsocalculatingthe probabilityofsuccess.

	age	sex	cp	trestbps	chol	fbs	...	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	...	0	2.3	0	0	1	1
1	37	1	2	130	250	0	...	0	3.5	0	0	2	1
2	41	0	1	130	204	0	...	0	1.4	2	0	2	1
3	56	1	1	120	236	0	...	0	0.8	2	0	2	1
4	57	0	0	120	354	0	...	1	0.6	2	0	2	1
5	57	1	0	140	192	0	...	0	0.4	1	0	1	1
6	56	0	1	140	294	0	...	0	1.3	1	0	2	1
7	44	1	1	120	263	0	...	0	0.0	2	0	3	1
8	52	1	2	172	199	1	...	0	0.5	2	0	3	1
9	57	1	2	150	168	0	...	0	1.6	2	0	2	1
10	54	1	0	140	239	0	...	0	1.2	2	0	2	1
11	48	0	2	130	275	0	...	0	0.2	2	0	2	1
12	49	1	1	130	266	0	...	0	0.6	2	0	2	1
13	64	1	3	110	211	0	...	1	1.8	1	0	2	1
14	58	0	3	150	283	1	...	0	1.0	2	0	2	1
15	50	0	2	120	219	0	...	0	1.6	1	0	2	1
16	58	0	2	120	340	0	...	0	0.0	2	0	2	1
17	66	0	3	150	226	0	...	0	2.6	0	0	2	1
18	43	1	0	150	247	0	...	0	1.5	2	0	2	1
19	69	0	3	140	239	0	...	0	1.8	2	2	2	1

Figure1: DatasetDistribution

The dataset as shown in Figure. 1 is from an ongoing cardiovascular study on resides of the city of Framingham,Massachusetts. The bracket thing is to prognosticate whether the casehas a 10- a time threat ofunborncoronary heart complaint (CHD). The issue has come less clear over recent times, in part, because of misgivings in the attention of heart failure,the lack of methodical recordings of arterial pressure previous to the onset of, and treatment for heart failure,andtheabsenceofmethodicalvisualization.ofepicardialcoronaryhighways,that'seasilydepicted.

Smoking damages the heart and bloodvessels veritably snappily, but the damage is repairedsnappily for utmost smokers who stop smoking. Indeed, many cigarettes now also damage the heart, so the onlyproven strategy is to keep your heart safe. From the goods of smoking is to quit. Researchers factors that highbloodsugar(glucose)causesstrongercontractionofbloodvesselsandalsolinkedaproteinassociatedwiththis increased contraction. The results could lead to new treatments to ameliorate issues after a heart attack orstroke is shown in Figure 5. Age is the most important threat factor in developing cardiovascular or heartconditions, with roughly a tripling of threat with each decade. life. Coronary adipose stripes can begin to form innonage. It's approximated that 82 percent of people who die of coronary heart complaints are 65 and age.Contemporaneously,thethreatofstrokedoubleseverydecade afterage55.



Correlation Analysis: Correlations have three important characteristics. They can tell us about the direction ofthe relationship, the form (shape) of the relationship, and the degree (strength) of the relationship between twovariables**.**

Figure 2: CorrelationMatrixoftheModel

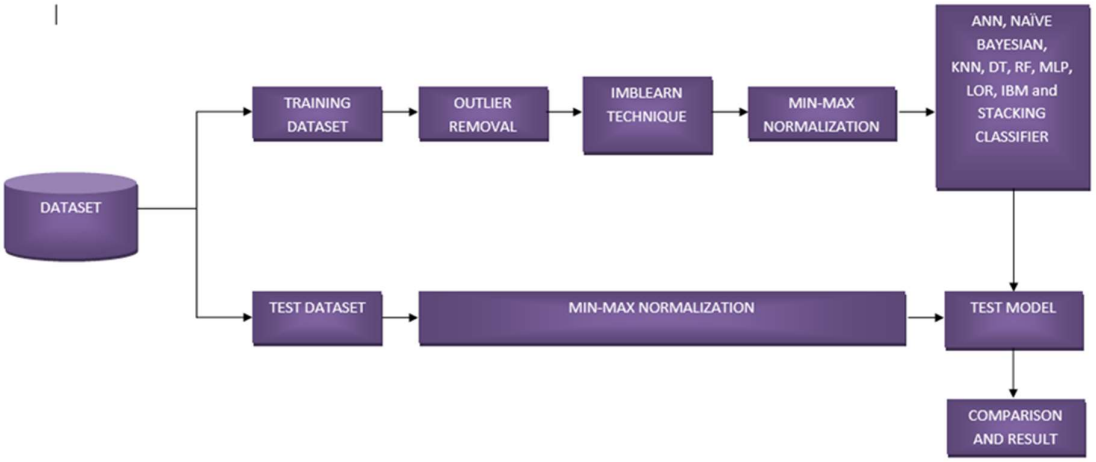


Figure3:BlockDiagram

Figure6 represents the block diagram of proposed system of the heart disease prediction. The block diagramconsistsofmainlytwosteps, the first oneisthetrainingdatasetand theotherone istestdata.

1. NaïveBayesClassification

NaiveBayesclassificationalgorithmtendstobeabaselinesolutionforsentimentanalysistasks.The basicidea
oftheNaiveBayestechniqueistofindtheprobabilitiesofclassesassignedtotextsbyusingthejointpr obabilitiesofwordsandclasses.Let'shaveabrieflookatmathematics.

GiventheBayestheorem: P(A|B) = P(B|A)
P(A)P(B)

For given elements, A and B and their probability of circumstance P (X) are calculated, where P (A) is theprobability of circumstance of element A, P (B) is the probability of circumstance of element A and P (A| B)isthe tentative probability of element A given element B occurs, and similar theorem will be used to perform thebracket.So,forindependentfeatures,thementionedtheoremwouldperformadirectadditionoft heprobability of eachpointpassing.

2. DataSetInformation

The name of the dataset isheart.csv. There are 303 cases in this dataset, where the cases are either peoplehaving heart complaints or they're healthy. Among 303, 165 (54.45) cases are people with a heart complaintand 138 (45.54) are people without heart complaint. The number of attributes is 14. There will be no missingvaluesinthe datasetnoranynull values.

Features include age, coitus, casket- pain type, rest BP, cholesterol, blood sugar position, ECG result, maximumheartrateachieved,exercise-convincedangina,STdepression,thepitchofpeakexerciseSTmember,numberofmajorvessels,an ddisfigurementinheartasof3-normal,6-fixeddisfigurementand7-reversibledisfigurement.

3. KNNMethods

K-Nearest Neighbor (K-NN) In the K-NN algorithm a data point is taken whose classification is not available,then the number of neighbors, k is defined. After that k neighbors are selected according to the lowest Euclidiandistance between the selected data points and their neighbors. The selected data point is then classified into acategory,whichisthesameasthecategorywhichhasthemajority
ofneighborsamongtheKneighbors.

4. Random Forest Methods

Random Forest works by building multiple decision trees of the training data. each of the trees predicts aclass as an affair and the class, which is the affair of the topmost number of decision trees is taken as the finalresult, in the case of the bracket. In this algorithm, we need to define the number of trees we want to produce.Random Forest is a bootstrap aggregating or bagging fashion. This fashion is used to drop the friction in theresults.

All classification and regression are achieved utilizing Random Forest algorithms. The information is categorized into a tree, and predictions are based on that tree. With a significant

number of record values missing, the Random Forest algorithm still can produce the same results when deployed to huge datasets. The decision tree's generated samples can be preserved and used for different data sets. In a random forest, there are two stages: first, generate a random forest, and then, using a classifier produced in the first stage, make a prediction.

5. Decision Tree

The Decision Tree algorithm's central node symbolizes the dataset traits, while the outlying branches will accomplish a specific objective. Decision trees are used because they are quick, reliable, simple to understand, and require very little data preparation. The decision tree's root determines the class label prediction. The core attribute's value, as well as the record's attribute, are evaluated.

6. Stacking Classifier

Stacking is a way of ensembling classification or regression models it consists of two-layer estimators. The first layer consists of all the baseline models that are used to predict the outputs on the test datasets. The second layer consists of Meta-Classifier or Regressor which takes all the predictions of baseline models as an input and generate new predictions.
Specifically, we will evaluate the following five algorithms:

- Multi-Layer Perceptron
- Adaboost Classifier
- Decision Tree.
- Extra Tree Classifier

III. RESULTS

Our main goal going into this research was to predict Heart disease using various machine learningtechniques.We predicted using K-Nearest Neighbor (K-NN) , ANN, DT, RF, IBM, MLP,LOR and Stacking Classifier. Stacking classifier gives 99.96 % of accuracy with better results. With Each algorithm, we have observed Accuracy, Precision,Sensitivity andSpecificityasfollows:

```
STACKING CLASSIFIER Accuracy is :91.00%
```

```
from sklearn.metrics import classification_report
STK_Pred=STK.predict(x_test)
STKreport = classification_report(y_test, STK_Pred)
print(STKreport)
```

```
              precision    recall  f1-score   support

           0       0.87      0.97      0.92       117
           1       0.92      1.00      0.96        33
           2       1.00      1.00      1.00        66
           3       0.00      0.00      0.00        17

    accuracy                           0.91       233
   macro avg       0.70      0.74      0.72       233
weighted avg       0.85      0.91      0.88       233
```

Figure 4: Without SMOTEand CP Based STACKING CLASSIFIER

```
STACKING CLASSIFIER Accuracy is :99.00%
```

```
from sklearn.metrics import classification_report
STK_Pred=STK.predict(x_test)
STKreport = classification_report(y_test, STK_Pred)
print(STKreport)
```

```
              precision    recall  f1-score   support

           0       0.98      1.00      0.99       117
           1       1.00      1.00      1.00        33
           2       0.98      0.97      0.98        66
           3       1.00      0.94      0.97        17

    accuracy                           0.99       233
   macro avg       0.99      0.98      0.98       233
weighted avg       0.99      0.99      0.99       233
```

Figure 5: With SMOTE and CP Based STACKING CLASSIFIER

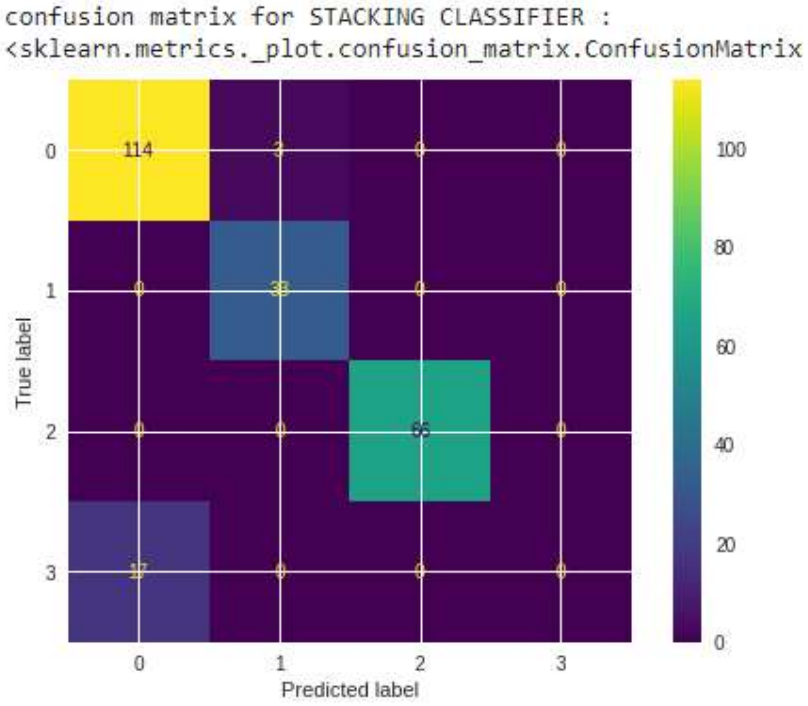


Figure 6: Without SMOTE and CP Based Confusion Matrix

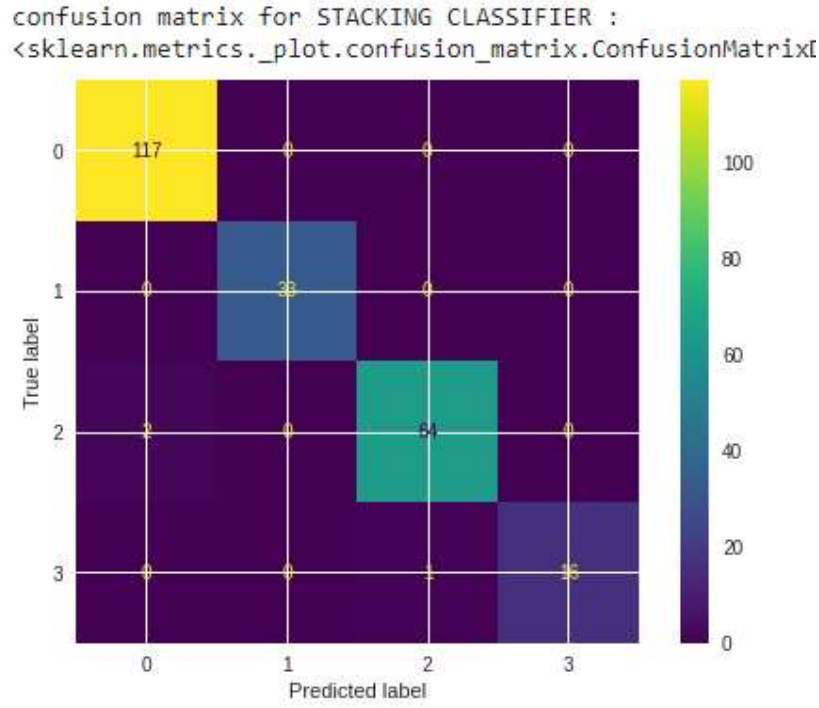


Figure 6: With SMOTE and CP Based Confusion Matrix

IV.COMPARISON CHART

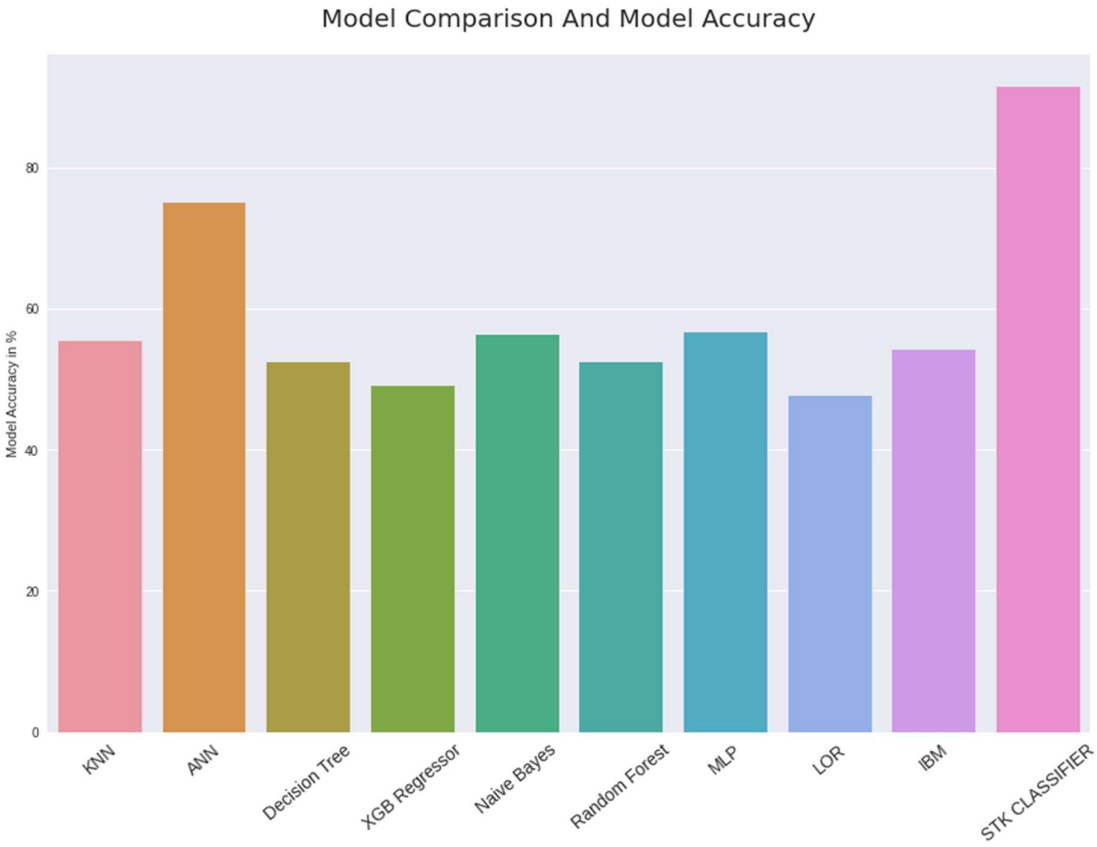


Figure 7: Without SMOTE and CP Based Comparison Chart

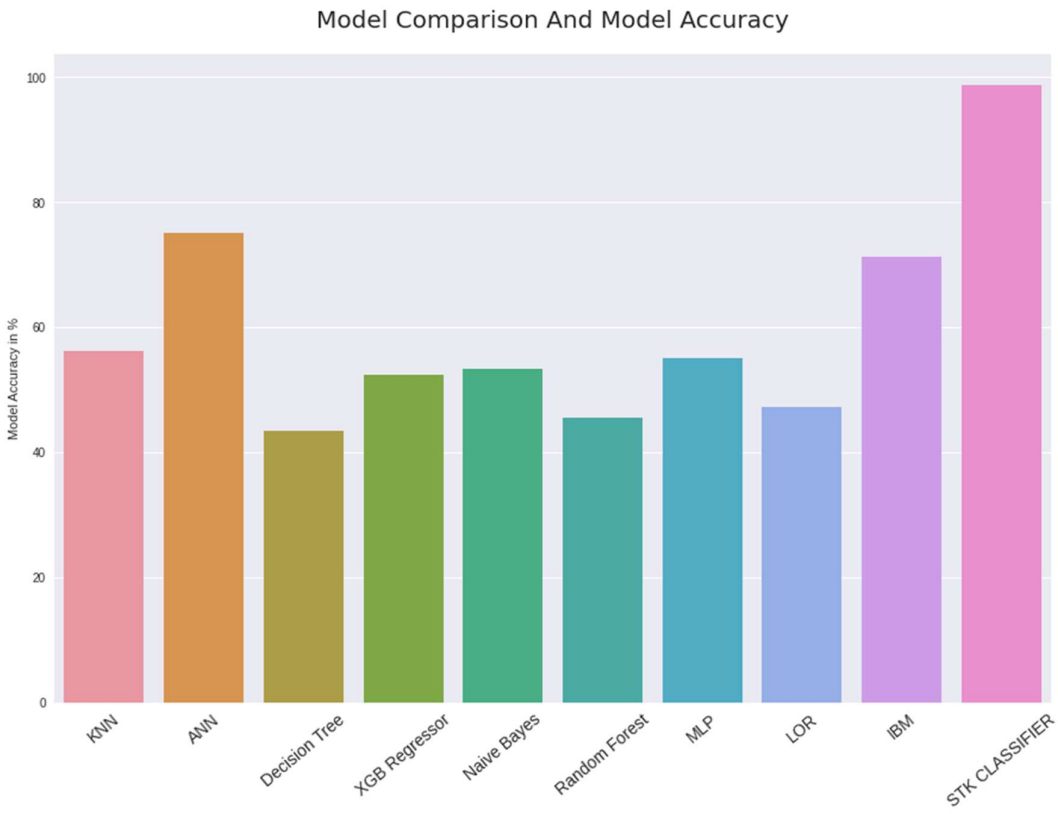


Figure 8: With SMOTE and CP Based Comparison Chart

V. CONCLUSION

The number of Heart conditions can exceed the current script to reach the maximum point. Heart complaint Iscomplicated and each and every time lots of people are dying from this complaint. It's delicate to Manuallydetermine the odds of getting a heart complaint grounded on threat factors preliminarily shown. By Using thissystem one ofthemajor downsides of thisworkis that its mainfocus is aimed only attheoperationofclassifying ways and algorithms for heart complaint vaticination, by studying Colorful data cleaning and miningways that prepare and make a dataset applicable for data Mining so that we can use this Machine Learning inthat logistic retrogression algorithms by prognosticating If the case has heart complaint or not. The any-medicalhand can use this software and prognosticate heart complaints and reduce the time complexity of the croakers.It's still an open sphere staying to get Enforced in heart complaint vaticination and increase the delicacy. Afterapplyingcolorfulalgorithms,itcanbesaidthatmachineliteracyisprovingtobeextremelypreciousinprognosticating heart complaints which is one of the most prominent problems of society in the moment'sworld. As further and further work is being done in the field of machine literacy, soon there may be new stylesto make machine learning further helpful in the field of healthcare. The algorithms used in this trial haveperformedwellusingtheavailableattributes.Theconclusioncanbeeventuallydrawnthatmachineliteracycanreducethedamagedonetoapersonphysicallyand mentally,byprognosticating

heartcomplaints.so in this research using Stacking Classifier model with SMOTE and CP value thatresultedinhighestaccuracywithanaccuracyof99% predict the accuracy and gives 91% of accuracy using Stacking Classifier model without SMOTE and CP value to identify the Heart Disease risk based on CP value.

VI.REFERENCES

[1] AvinashGolande,PavanKumarT.Heartcomplaintvaticinationusing effectivemachine Learningways.

[2] TheLancetGlobalHealth.Thechangingpatternsofcardiovascularconditionsandtheirthreatfactorsinthe countriesof IndiaTheglobal burdenofcomplaintstudy1990-2016.

[3] HimanshuSharma,MARizvi.VaticinationofheartcomplaintsusingmachineliteracyalgorithmsAcheck.Worldhealthranking.

[4] SanaBharti,Analyticalstudyof heart complaint vaticination comparedwithdifferentalgorithms;Internationalconferenceoncomputing,communication,androbotization(ICCA2022).

[5] MonikaGandhi,Vaticinationinheartcomplaintusingwaysofdatamining,internationalconferenceonfuturistic trend incomputational analysisandknowledge operation(ABLAZE-2022)

[6] SarathBabu,2017.heartcomplaintopinionusingdataminingfashion,internationalnationalconferenceonelectronics,communicationandaerospacetechnology (ICECA2022)

[7] AHChen,2011.HDPSheartcomplaintvaticinationsystem;

[8] GroverSA,LowensteynI,EsreyKL,etal.Docroakersdirectlyassessthecoronarythreatintheircases.Primary resultsof theCoronaryHealthAssessmentStudy.

[9] FriedmannPD,BrettAS,Mayo-SmithMF.Differencesingeneralistsandcardiologists'comprehensionsofcardiovasculartheratandtheissuesofpreventativeremedy inthecardiovascularcomplaint.

[10] MelandE,LaerumE,LehmannEH.Assessmentofcoronaryheartcomplaintthreat,IIaclinicalmulti-centerstudyofgeneralinterpretersthreatassessment.

[11] Mohan.S,Thirumalai.C,&Srivastava.G,(2021).Effectiveheartcomplaintvaticinationusingmodernmachine-learningways.