Global Scientific JOURNALS

GSJ: Volume 12, Issue 1, January 2024, Online: ISSN 2320-9186

www.globalscientificjournal.com

An Improved heart disease prediction system comparing machine learning Algorithm

Anazodo William , Okengwu U. A. , Odia K. M.

CENTER FOR INFORMATION AND TELECOMMUNICATIONS ENGINEERING.

UNIVERSITY OF PORT HARCOURT

ABSTRACT

This project is an intelligent decision system for heart disease prediction using machine learning algorithm. The problems associated with the current process of heart disease diagnosis for healthcare shows the clinical decisions are often made based on the doctor's intuition and experience rather than utilizing the knowledge filled data hidden in the healthcare database, There is also inconsistency in dataset arrangements and this gives rise to potential inaccuracy in healthcare decisions. This system is designed to process the open source data stored in healthcare centers by importing the datasets and loading the datasets to the python environment to evaluate them for diagnosis purposes and machine learning analysis. The software methodology used in this system is the agile methodology with constant Iteration throughout the development cycle. This system was created with the use of python language as the machine learning language. This is a system designed to give function of efficiency and accuracy to the prediction process of heart disease by developing an improved Naïve bayes Algorithm. This machine learning system helps even the healthcare centers in prediction analysis process by using evaluation metrics like box plot, ROC curve, Confusion matrix, F1 score, precision and recall graphs to compare the accuracy of the evaluation metrics with other machine learning algorithms to specify the best accuracy and identify the most accurate prediction for heart disease patients, The modified naïve bayes algorithm model provides the best classification report with f1score 0.97, precision 0.95, recall 1.0, AUC 0.96 and overall accuracy of 97% in machine learning analysis compared to the other models making it very easy for the probability of heart disease to be observed and analyzed properly for diagnosis purposes to the benefit of the efficient Healthcare System.

Keywords: Data mining, Machine learning, Heart disease, Classification, Naive Bayes, Artificial Neural Networks, Decision Trees, Associative Rule

INTRODUCTION

Data mining and Machine Learning deals with finding the relationships and global patterns from large databases which are unseen among large amounts of data. Using medical datasets such as age, gender, chest pain type, fasting blood sugar, resting blood pressure, resting electrographic results, cholesterol, maximum heart rate achieved, exercise induced angina, old peak, slope, number of vessels coloured, defect type, obesity and smoking we can predict the probability of patients getting a heart disease. The Quality of healthcare service is the serious challenge that the healthcare industry has face because of quality of service deals with diagnosing disease correctly & providing efficient treatments to patients. Heavy loss can happen because of poor diagnosis. Diagnosis is an important task that must be executed correctly and efficiently. Correct predictions are critical for diagnosis and treatment. Day by day, researchers continue to develop effective decision support systems. Diagnosis of heart disease remains a challenge and Prediction relies heavily on classification techniques. The diagnosis is always based on doctor's experience and knowledge. This leads to unwanted results and excessive medical costs of treatment provided to patients. Therefore an automatic medical diagnosis system needs to be designed that can take benefit of heart disease database which is publicly available. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data.

PROBLEM STATEMENT

i. Inconsistency in the Datasets arrangements gives rise to error.

ii. There is low scalability with several data points and predictors.

iii. Difficulty in finding the optimal parameters for a specific Classification model.

iv. Potential Inaccuracy in Healthcare Decisions.

AIM AND OBJECTIVES

The aim of this research is to develop an Intelligent Decision system for Heart Disease Prediction using machine learning Algorithm.

The Objectives are:

i. To develop a modified Naïve Bayes Algorithm Model.

ii. To train and test different classification algorithms with the Heart disease Datasets.

iii. To implement the modified Naïve Bayes Model.

iv. To validate the performance of the modified Naïve Bayes Algorithm model using specified evaluation Metrics.

v. To compare the best accuracy for prediction evaluating the different Algorithms.

LITERATURE REVIEW

There have been numerous studies done related to predicting the disease using different machine learning techniques and algorithms which can be used by medical institutions. This paper reviews some of those studies done in research papers using the techniques and results used by them. Using a proposed model a Researcher proposed a disease prediction system in his paper where he used machine learning algorithms. In the prediction of disease, he used techniques like Convolutional Neural Network based Unimodal Disease Risk Prediction CNN-UDRP algorithm, Convolutional Neural Network based multi modal Disease Risk Prediction CNN-MDRP algorithm, Naive Bayes, K-Nearest Neighbor(KNN), and Decision Tree. An Internet of things (IOT) structure for evaluating heart diseases in an accurate manner through Modified Deep Convolutional Neural Network (MDCNN), where heart monitoring device and smart watch were fixed to the patient for monitoring the Electro cardio gram (ECG) and blood pressure of patients. The classification of gathered sensor data was performed using MDCNN to get the classes as abnormal and normal. The designed model was analyzed with other conventional models like logistic regression and deep learning neural networks. The experimental results have revealed that the designed MDCNN attained superior prediction performance for heart diseases regarding accuracy.

Many researchers try to use machine learning and data mining Algorithm to predict complex psychological problems of individuals, such as predicting stress disorders and anxiety disorders.

Early heart disease prediction is an emerging research area in artificial intelligence. Another study focused on feature filtering techniques to predict breast cancer early. Frequent item-set mining is used to select the essential features in patients' datasets. The decision tree, Naive Bayes (NB), k-Nearest Neighbors (k-NN), and Support Vector Machine (SVM) are compared, and it is found that SVM outperforms other models. A research paper focused on reducing erroneous prediction results that is false positive and false negative. Information gain with the Genetic Algorithm approach is utilized to rank highly significant features in the

dataset. Classification of positive and negative results is done by SVM. This approach applied to two independent datasets increased the prediction accuracy and reduced prediction cost.

Two breast cancer datasets are used to differentiate cancer patients from healthy people. A genetic algorithm is used to select the most significant features in datasets.

Three classification algorithms, multilayer perceptron (MLP), probabilistic neural network (PNN) and radial based function (RBF) are used for the classification of breast cancer. MLP requires more processing time for the training model and assigning weights to its neurons than RBF and PNN. The highest accuracies of 97%, 98%, and 100% are achieved with MLP, RBF, and PNN. The probabilistic procedure with sensors is used for censorious procedures. Some patients have diabetes with hypertension; therefore, a research study is conducted to predict diabetes type II and hypertension in both individuals. Synthetic minority oversampling is used to solve data distribution problems and the ensemble method is used to predict hypertension and type II diabetes. This study showed that pre-processing of data prior to model building enhances prediction accuracy. Machine learning methods are also used for the detection of some other medical diseases detection, text mining and network security. Another analysis is conducted to predict the risks associated with diabetes. Logistic regression, Decision Tree, ID3, C4.5, k-NN, and Naive Bayes are used for classification. Irrelevant features are detected and reduced using PCA and PSO algorithms. A comparison of both feature selection techniques is performed in terms of improved accuracy and processing time. It shows that feature reduction is a powerful technique to enhance predictive model accuracy.

In recent decades, heart attacks have been one of the leading causes of death. Random forest is used in conjunction with a linear model to predict heart disease based on patient data in a study. The most significant features are classified using an a priori method, which showed an increase in accuracy. When compared to existing classification techniques, the suggested ensemble model produces the best results. Optimization techniques are used to deal with complex data by reducing irrelevant and functional data attributes. A hybrid approach combines different AI techniques and evaluates them as a single algorithm. A research study applied various AI algorithms for heart disease prediction to find the best classifier. The best performing algorithms are then combined, and the results of the hybrid approach are compared with the results of individual techniques. The proposed hybrid algorithm gave more accurate results as all algorithms functions are combined to calculate a majority vote for a positive class. A drawback of using a hybrid approach is that more processing time is required. Another study applied a Fuzzy Rule-based classification model for heart attack risk prediction. Rough set theory identifies the most influencing features, and features that do not

affect prediction results are eliminated. The Genetic Algorithm is used to reduce time complexity and optimize prediction results. Research proved that the proposed model is efficient with a large number of features. A comparative analysis is presented in for the diagnosis of cardiovascular disease. The experimental setup comprises classification algorithms, namely Logistic Regression, Decision Tree, and SVM. Different subsets of features are used to evaluate classification results. The highest accuracy of 82.97% is achieved by using a decision tree with forward and backward selection approaches.

Junaid Rashid et al, (2022) proposed a machine learning-based predictive model to diagnose three primary chronic diseases, including Diabetes, Kidney, and heart diseases. The feature selection technique based on adaptive probabilistic divergence is used to select the most valuable features. The study concluded that the proposed approach gave the highest accuracy by optimizing the most significant features for disease detection. A feature selection-based machine learning algorithm is proposed to predict three chronic diseases, namely, diabetes, heart attack, and cancer. The incremental feature selection approach with Convolutional Neural Network (CNN) is applied to anticipate disease presence. The proposed method showed 93% classification accuracy in less computation time. Another study is conducted to explore the essential features of often chronic diseases. Feature selection approaches, including Information Gain, Gain Ratio, and correlation-based approaches, are applied. Several subsets of top-ranked features are then used in building the Random Forest prediction model. It showed that exploring the most significant features is significant for the medical diagnosis process. In another research study for chronic disease prediction, the Stacked Generalization approach is used to enhance the performance of classification algorithms. Five classification algorithms are compared: Decision Tree (DT), k-NN, SVM, Logistic Regression (LR), and Naive Bayes, to outperform five chronic disease prediction models. It is found that the Stacked Ensemble approach enhances the model performance and achieves the highest accuracy of 90%.

A study is proposed to predict whether a patient would survive or die because of hepatitis. This study implemented SVM with multilayer perceptron for the binary classification of patients. Principal component analysis (PCA), a feature selection approach, is applied before prediction and three factors of zero correlation with the predictive class are eliminated. The proposed hybrid algorithm gave a higher accuracy than state-of-the-art classifiers. A combination of several prediction techniques results better than an individual model. Another study suggested a model for hepatitis diagnosis using ensemble learning techniques. This

hybrid method manifested improved prediction results achieved in less computation time. Esophageal variants occur most commonly due to the side effects of hepatitis. The esophageal prediction model is proposed in a research study, with six feature selection approaches. Information Gain, PCA, and correlation are used to select the most significant features. Three feature ranking approaches, namely Relief, information gain, and gain ratio, explore the most significant attributes.

DM Algorithm	Type	Formulation
C4.5	decision tree	Classification
k-means	distance-based	Clustering
SVM (Support Vector Machine)	geometric	Classification
Apriori	rule-based	Association rules
EM (Expectation Maximation)	statistical	Clustering
PageRank	network graph	Ranking
AdaBoost	boosting	Ensemble
kNN (k Nearest Neighbor)	distance-based	Classification
Naïve Bayes	statistical	Classification
CART (Classification and Regression Tree)	decision tree	Classification; Regression

Figure 1.0 Machine Learning Algorithms based on Formulation.

RESEARCH DESIGN AND METHODOLOGY

The research design for the development to implement an intelligent heart disease prediction system for the Health Facilities using a web application and Machine learning algorithm. The new modified Algorithm model will help to simplify the complexity of handling enormous amounts of data in terms of Dataset training and testing.

The Improved Heart disease prediction system for Health analysis is made up of the evaluation of different types of machine learning algorithm for Prediction Analysis.

Forecasting of expected prediction by evaluation of the variables or values .These datasets being used are already registered by the hospitals and healthcare centres to help non-specialized research experts to find a solution for system inefficiency for healthcare problems.

The architecture of the proposed system will now be used in the Flow process as follows Data Collection, Patient Database, Reading the Dataset attributes, Modify Naïve bayes Algorithm,

Algorithm analysis of the classification algorithms, Training the Models, Testing the Model, Initializing the Models, classification report and identifying the accurate prediction.

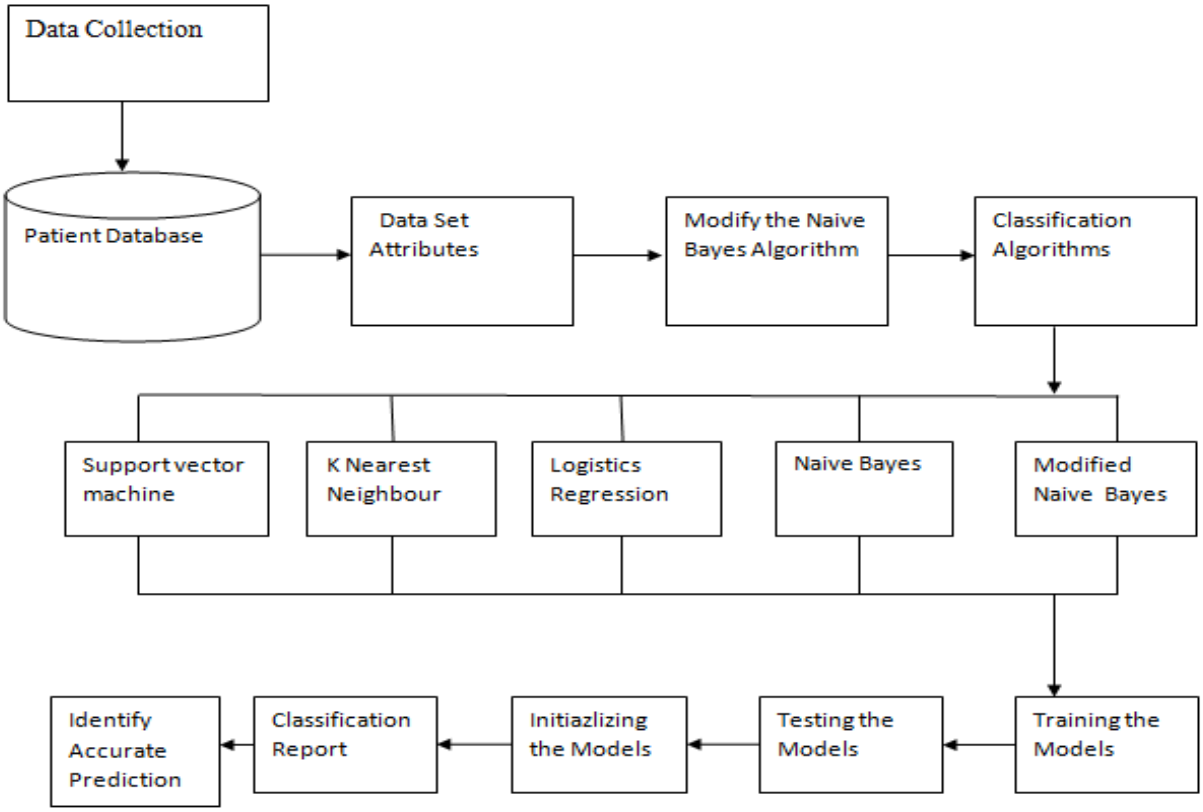


Figure 2.0 Architecture of the proposed heart disease prediction system

Methods of Data Collection

Data collection sources play a crucial role in the research process as they determine the quality and accuracy of the data collected. The Datasets for Training the Model of the system is mainly specified in some open source data, International Health Research Networks like

https://cite-idsapp.online/control/datasets.csv

Here is some major importance of data collection methods.

i. Determines the quality and accuracy of collected data.

ii. Ensures that the data is relevant, valid, and reliable.

iii. Helps reduce bias and increase the representativeness of the sample.

iv. Essential for making informed decisions and accurate conclusions.

v. Facilitates achievement of research objectives by providing accurate data.

vi. Supports the validity and reliability of research findings.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	patientid	age	gender	chestpain	restingBP	serumcho	fastingblo	restingrel	maxheartı	exerciseai	oldpeak	slope	noofmajo	target
2	103368	53	1	2	171	0	0	1	147	0	5.3	3	3	1
3	119250	40	1	0	94	229	0	1	115	0	3.7	1	1	0
4	119372	49	1	2	133	142	0	0	202	1	5	1	0	0
5	132514	43	1	0	138	295	1	1	153	0	3.2	2	2	1
6	146211	31	1	1	199	0	0	2	136	0	5.3	3	2	1
7	148462	24	1	1	173	0	0	0	161	0	4.7	3	2	1
8	168686	79	1	2	130	240	0	2	157	0	2.5	2	1	1
9	170498	52	1	0	127	345	0	0	192	1	4.9	1	0	0
10	188225	62	1	0	121	357	0	1	138	0	2.8	0	0	0
11	192523	61	0	0	190	181	0	1	150	0	2.9	2	0	1
12	201030	59	0	1	190	529	1	1	151	1	3.2	2	2	1
13	208877	58	1	2	192	409	1	0	138	0	2.3	3	1	1
14	223295	27	1	0	129	135	0	1	192	1	1	0	0	0
15	226481	59	1	0	98	209	0	0	117	1	5.6	1	0	0
16	229445	58	1	0	170	354	0	0	170	0	5.6	1	0	0
17	235344	32	1	2	188	0	0	0	134	1	4.5	2	3	1
18	236763	42	0	3	137	350	0	1	110	0	3.2	2	2	1
19	240461	65	1	0	200	247	1	1	194	1	3.7	1	1	0
20	247055	59	1	2	182	177	0	1	168	0	2.1	2	1	1
21	260870	35	1	0	127	269	0	0	87	1	3.8	0	1	0
22	266839	39	1	3	196	253	1	2	140	1	3.5	2	1	1
23	322287	72	1	1	177	397	0	2	124	0	5.2	3	1	1
24	327110	24	0	0	136	164	0	0	91	1	1.8	1	1	0
25	335359	59	1	2	156	223	0	2	184	1	2.4	2	2	1

Figure 3.0 Diagram showing the Attributes of the Datasets

```
Powered by trinket
Python: 3.10.9 (main, Jan 23 2023, 22:32:48) [GCC 10.2.1 20210110]
scipy: 1.9.0
numpy: 1.23.2
matplotlib: 3.5.3
pandas: 1.4.3
sklearn: 1.1.2
<bound method NDFrame.head of       patientid  age  gender  ...  slope  noofmajorvessels  target
0     patientid  age  gender  ...  slope  noofmajorvessels  target
1        103368   53       1  ...      3                 3       1
2        119250   40       1  ...      1                 1       0
3        119372   49       1  ...      1                 0       0
4        132514   43       1  ...      2                 2       1
...         ...  ...     ...  ...    ...               ...     ...
69997    9949544   48       1  ...      2                 2       1
69998    9953423   47       1  ...      1                 0       0
69999    9965859   69       1  ...      3                 1       1
70000    9988507   45       1  ...      3                 2       1
70001   9990855   25       1  ...      0                 0       0

[70001 rows x 14 columns]>
```

Figure 4.0 Python Environment showing the data frame of rows and columns

- **Modified Naïve Bayes** : The New algorithm is developed to handle the Prediction Accuracy Most Effectively, This New Algorithm is an Improvement on the already existing Naïve Bayes Algorithm used for Machine Learning. So this original Naïve Bayes model is Integrated with the Formula

$$P(C\backslash A) = \frac{P(A\backslash C)\,P(C)}{P(A)}$$

The classifier is especially affected by values not being represented in the training set. If the model comes across a categorical feature that isn't present in the training set, the probability of 0 is assigned to that new category. This is very dangerous, as multiplying 0 with other features' probabilities will result in 0. So we add a parameter to both numerator and denominator when calculating the class probabilities. We added a smoothing parameter of (+1) when calculating the class probabilities:

$$P(C\backslash A) = \frac{P(A\backslash C)\,P(C) + 1}{P(A) + 1}$$

The parameter ensures that the probability value is never zero. This will assign a very small probability estimate to such zero frequency occurrences, hence, regularize the Naive Bayes classifier.

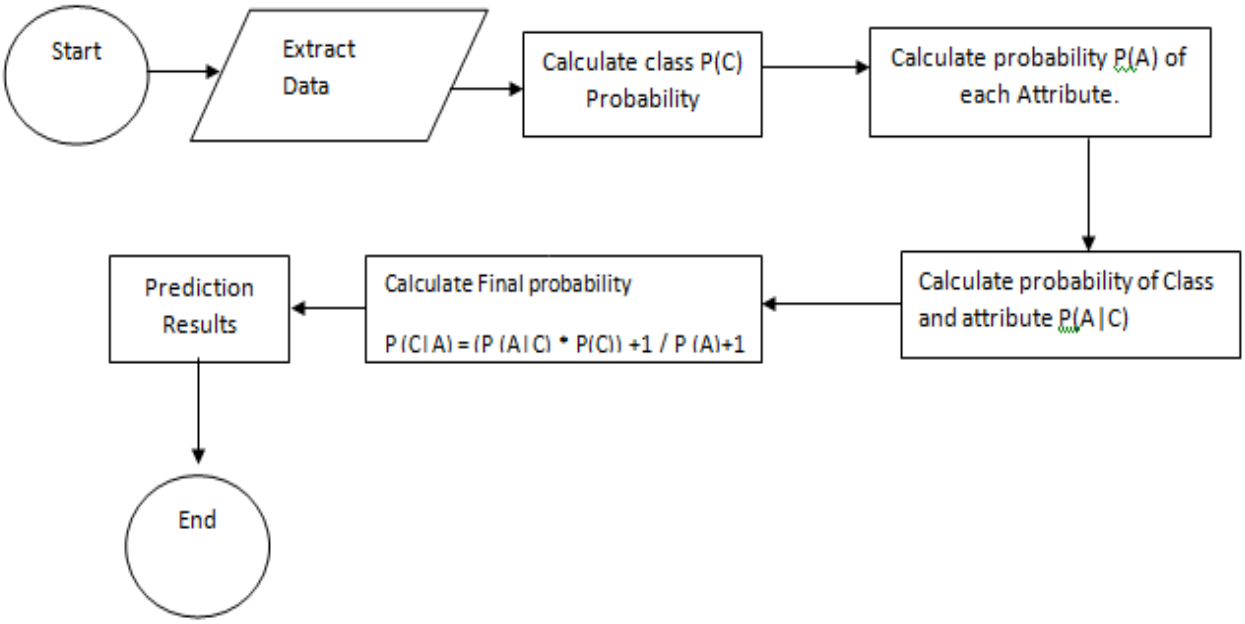


Figure 5.0 Flowchart of the modified naïve bayes Algorithm

Here are some advantages of the modified Algorithm:

- It doesn’t require larger amounts of training data.
- It is straightforward to implement.
- Convergence is quicker than other models, which are discriminative.
- It is highly scalable with several data points and predictors.
- It can handle both continuous and categorical data.

RESULT

The key to a fair comparison of machine learning algorithms is ensuring that each algorithm is evaluated in the same way on the same data. You can achieve this by forcing each

algorithm to be evaluated on a consistent test harness. So now the newly modified naive bayes algorithm will be compared along with 6 other existing algorithm models.

In the Illustration below the 5 different algorithms are compared:

- Logistic Regression
- K-Nearest Neighbors
- Naive Bayes
- Support Vector Machines
- Modified Naïve Bayes

The problem is a standard binary classification dataset called the Pima Indians onset of diabetes problem. The problem has two classes and eight numeric input variables of varying scales.

The 14-fold cross attributes procedure is used to evaluate each algorithm, importantly configured with the same random datasets to ensure that the same splits to the training data are performed and that each algorithm is evaluated in precisely the same way.

The main Algorithm which we developed is Given a short name **MNB** which means

Modified Naïve Bayes, This Algorithm stands as a Modified Naïve Bayesian model.

- models = []
- models.append(('MNB', NewGaussianNB()))
- models.append(('NB', GaussianNB()))
- models.append(('KNN', KNeighborsClassifier()))
- models.append(('LR', LogisticRegression()))
- models.append(('SVM', SVC()))

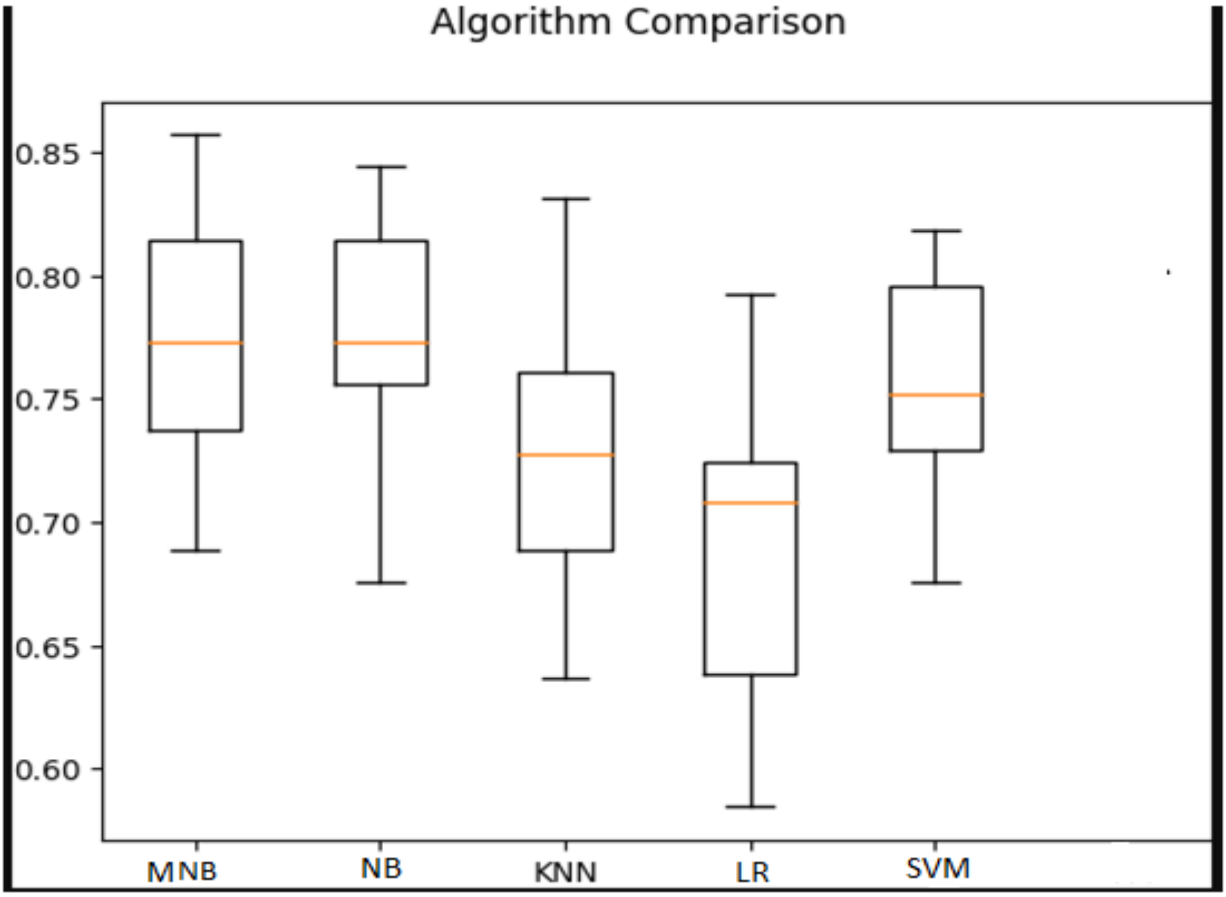


Figure 5.0 Box plot Comparing Mean Accuracy

Each algorithm is given a short name, useful for understanding the Graphical Results and running the example provides a list of each algorithm short name, the mean accuracy and the standard deviation accuracy. The improved Naïve bayes Algorithm holds the highest mean accuracy of 0.861(86.1%)

Modified Naïve Bayes classification report

The modified Naïve Bayes algorithm is developed to handle the Prediction Accuracy Most effectively; This New Algorithm is an Improvement on the already existing Naïve bayes Algorithm used for Machine Learning. So this modified naïve bayes Algorithm Model is integrated into the Software. The classifier is especially affected by values not being represented in the training set. If the model comes across a categorical feature that isn't present in the training set, the probability of 0 is assigned to that new category. This is very dangerous,

as multiplying 0 with other features' probabilities will result in 0. So a constant of +1 is added to both numerator and denominator when calculating the class probabilities. So the new algorithm equation is initialized into the prediction system. The f1 score is 0.97, recall score is 1.0, precision score is 0.95. The confusion matrix below shows 75 true cases of patients with heart disease and 119 false cases of patients with no heart disease. The area under the curve(AUC) is 0.96 and the accuracy of the modified naïve bayes model is 97%.

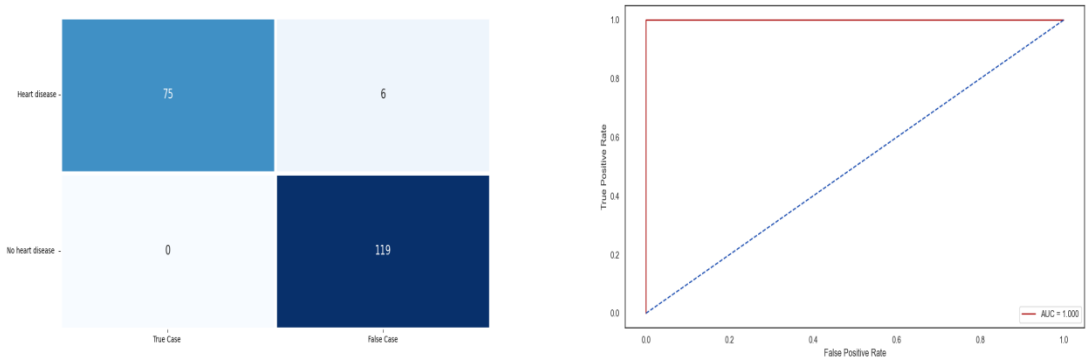
```
Precision: 0.952
Recall: 1.0
F1 Score -->  0.97540983606557377
Classification Report  -->
              precision    recall  f1-score   support

           0       0.93      1.00      0.96        75
           1       1.00      0.95      0.98       125

    accuracy                           0.97       200
   macro avg       0.96      0.98      0.97       200
weighted avg       0.97      0.97      0.97       200

ROC AUC score: 0.9629629629629629
Accuracy Score:  0.97
```

Figure 6.0 Modified Naïve Bayes classification report.



Confusion Matrix ROC Curve

Figure 7.0 Confusion matrixes and ROC Curve for the Modified Naïve Bayes

CONCLUSION

The machine learning system works as proposed and also worked based on the aim and objectives of the project. Loading of the datasets was done successfully and analysis of the to handle their condition effectively. This study managed to provide a significant contribution in computing the strength scores with significant predictors in heart disease prediction. From the evaluation results, we obtained important results and achieved highest confidence score by utilizing the computed strength scores of significant predictors on Classification Models in predicting heart disease.

The Improved Heart disease prediction system for Health analysis is made up of the evaluation of different types of machine learning algorithm for Prediction Analysis.

Forecasting of expected prediction by evaluation of the variables or values .There are open source by the hospitals and healthcare centres to help non-specialized doctors to make correct decision about the heart disease risk level..

So there is need to process that data and instead of just storing that data it is needed to extract some meaningful information or knowledge from that data applying some classification techniques of machine learning as we know is very useful for the helping in prediction, basis

on the some parameters like chest pain, cholesterol, age, resting Bp and many more. Due to analysis of the different algorithms now this system clinical decision making accuracy will be improved as. It will also impact on the improving the treatment process. In such way it will be very useful in the prediction of the heart disease.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to the Centre that has contributed to the successful completion of this research paper.

Center for Information and Telecommunication Engineering (CITE)

Your Platform has been truly instrumental, and we are sincerely grateful.

REFERENCES

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

2) Alty, S. R., Millasseau, S. C., Chowienczyk, P. J., & Jakobsson, A. (2003). Cardiovascular disease prediction using support Vector Machines. 2003 46th Midwest Symposium on Circuits and Systems. https://doi.org/10.1109/mwscas.2003.1562297

3) Asadi, S., Roshan, S. E., & Kattan, M. W. (2021). Random forest swarm optimization-based for heart diseases diagnosis. Journal of Biomedical Informatics, 115, 103690. https://doi.org/10.1016/j.jbi.2021.103690

4) Asha Rajkumar, and Mrs G. Sophia Reena, 2010, "Diagnosis of Heart Disease using Data Mining Algorithms",Global Journal of Computer Science and Technology,Vol. 10,Issue 10, pp.38-43, September

5) Ashish Chhabbi,Lakhan Ahuja,Sahil Ahir, and Y. K. Sharma,19 March 2016,"Heart Disease Prediction Using Data Mining Techniques", International Journal of Research in Advent Technology,E-ISSN:2321- 9637,Special Issue National Conference "NCPC-2016", pp. 104-106.

6) Bambrick, N. (2022). Support Vector Machines: A simple explanation. KDnuggets. Retrieved November 3, 2022, from https://www.kdnuggets.com/2016/07/support-vector-machines-simpleexplanation.html

7) Bhardwaj, R., Nambiar, A. R., & Dutta, D. (2017). A study of machine learning in Healthcare. 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). https://doi.org/10.1109/compsac.2017.164

8) Boshra Bahrami, and Mirsaeid Hosseini Shirvani,February 2015,"Prediction and Diagnosis of Heart Disease by Data Mining Techniques",Journal of Multidisciplinary Engineering Science and Technology(JMEST), ISSN:3159- 0040, Vol. 2, Issue 2, pp. 164-168.

9) Camm, J. D., Cochran, J. J., Fry, M. J., & Ohlmann, J. W. (2021). Business analytics: Descriptive, predictive, prescriptive. Cengage. Centers for Disease Control and Prevention. (2022, September 8). Heart disease and stroke. Centers for Disease Control and Prevention. Retrieved September 28, 2022, from 37 https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart -disease-stroke.htm

10) Junaid Rashid (2022) An Augmented Artificial Intelligence Approach for Chronic Diseases Prediction. National Library of Medicine. 2022 Mar 31

11) Chauhan, N. S. (2022, April). Naïve Bayes Algorithm: Everything you need to know. KDnuggets. Retrieved October 29, 2022, from https://www.kdnuggets.com/2020/06/naive-bayes-algorithmeverything.html?hss_channel=tw-1318985240

12) Chen, L. (2019). Support Vector Machine — simply explained - towards data science. Support Vector Machine — Simply Explained. Retrieved November 4, 2022, from https://towardsdatascience.com/support-vector-machinesimply-explained-fee28eba5496

13) Dadgostar, P. (2019). antimicrobial resistance: Implications and costs. Infection and Drug Resistance, Volume 12, 3903–3910. https://doi.org/10.2147/idr.s234610

14) Davari Dolatabadi, A., Khadem, S. E., & Asl, B. M. (2017). Automated diagnosis of coronary artery disease (CAD) patients using optimized SVM. Computer Methods and Programs in Biomedicine, 138, 117–126. https://doi.org/10.1016/j.cmpb.2016.10.011

15) Gupta, D., Khare, S., & Aggarwal, A. (2021). A method to predict diagnostic codes for chronic diseases using Machine Learning Techniques. IEEE Xplore. Retrieved September 16, 2022, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=781373