HEART ATTACK POSSIBILITY PREDICTION

Md. Tamim Ahmed Opi
Daffodil International University
tamim15-13773@diu.edu.bd

S. M. Mahamudul Haque
Daffodil International University
mahamudul15-13707@diu.edu.bd

Md. Merajul Islam Siddiqui
Daffodil International University
merajul15-13980@diu.edu.bd

Abstract- The heart appears to be a very complex organ. And even if some of it might have suffered severe damage, the majority of the heart is still beating. However, the injury may weaken the heart and make it less able to pump blood than usual. An further injury can be minimized or prevented with prompt identification of a variety of potential hamstring problems, appropriate therapy, and dietary adjustments following a heart attack. In this study, the likelihood of a heart attack is estimated using various machine learning technique. There are linear regression, logistic regression, decision tree classifier, SVC, naive bayes, KNN classifier, random forest classifier and gradient boosting. Finding the best algorithm with this paper also shows the correction metrics visualizing the feature. In this paper, random forest classifier is the best model with the accuracy of 85.25%.

I. INTRODUCTION

Numerous heart conditions are categorized as cardiovascular illnesses. Heart rhythm abnormalities, coronary artery disease, congenital heart defects, and disorders of the blood vessels all fall under the category of cardiovascular disease (congenital heart defects). The most crucial organ in our body is the heart. The lifeline of our body is the heart. A common disease is heart disease. A heart attack is one of the prevalent ailments. When a heart attack occurs, the blood supply does not function perfectly. After a heart attack, fat, cholesterol, and other substances tend to collect more frequently. An attack or myocardial infarction also happens when the heart muscle does not receive enough oxygen. Heart condition signs: abdominal pressure or discomfort. Most heart attacks result in emotional or left chest discomfort that lasts for more than a few minutes or comes and goes. It may be uncomfortable to experience straining, friction, tightness, or soreness. a feeling of weakness or faintness. You can start to perspire as well. pressure or irritation on the back, neck, or mask. Pain or discomfort in one or both shoulders or arms. shorter breathing. Heart pain can occasionally result from this, but an oxygen shortage can also occur before heart attacks. Heart disease statistics: In America, there were 735,000 heart attacks in 2016. Cardiovascular disease is the leading cause of death in the US. Only 0.3% of men and women between the ages of 20 and 39 have cardiac disease. For men, the average age at first heart attack is 65, while for women, it is 72.

A risk factor is: The following are the main risk factors for preventing a heart attack: Alter. Alter. Heart attacks are more common in men and women over the ages of 45 and 55 than in those under those ages. Smoking. Smoking. Both smoking and inhaling smoke from others were mentioned. the blood pressure being too high. Over time, high blood pressure can harm the arteries leading to your heart. Obesity, high cholesterol, and diabetes all increase the risk of high blood pressure. increased serum or triglyceride cholesterol levels. Low density lipoprotein may be linked to high levels of "bad" cholesterol (LDL). Your chance of having a heart attack is also increased if your blood triglycerides, a kind of blood fat connected to

your diet, are high. A would reduce the risk by substantial amount of high-density lipoprotein ('good' cholesterol) (HDL). Abesthood. Diabetes, high blood pressure, raised serum cholesterol, and high triglyceride levels are all associated with obesity. This risk can be decreased by simply 10% body weight loss. Diabetes. Diabetes. The body can boost blood sugar levels due to insufficient insulin or substances generated by insulin, which increases your chance of having a heart attack. The amount of blood sugar you receive is unaffected by insulin's effect. significant metabolism. You smoke, have asthma, and have high blood sugar, which is why this is. You will experience heart disease twice as frequently as people without metabolic syndrome.

Past familial traumatizes the heart. Because your siblings, parents, or relatives had heart attacks at a young age (55 for men, 65 for women), you may be more susceptible to danger. failure to exercise Obesity and inactive serum cholesterol raise blood levels. Daily exercisers have better cardiovascular health, including reduced blood pressure. Stress: You should react to stress in order to enhance your risk of having a heart attack. illegal drug use. The use of sedative narcotics like cocaine or amphetamines will result in a coronary artery spasm that will result in a heart attack. a history of pre-eclampsia. By increasing the chance of lifelong heart failure, this illness boosts blood pressure levels during pregnancy. an autoimmune condition. The risk will rise in the presence of conditions like lupus or rheumatoid arthritis. Therefore, the probability of a heart attack justifies more research into them.

II. LITERATURE REVIEW

Heart attack has supplanted other causes of death in the United States by 1940. the time when after the death of Franklin D. Roosevelt from hypertensive heart disease in 1945, Americans started to take an interest in heart disease research. To study heart disease and its numerous risk factors, The Framingham Heart Study was first founded in the United States in 1948. Since then, research on heart failure and its many risk factors has been ongoing in an effort to better understand how to prevent it with contemporary technologies. In his article, Fizar Ahmed described the architecture of heart attack rates and utilized the IoT idea to forecast patients who will experience heart attacks in the future. KNN (k Nearest Neighbor), one of the most well-liked machine learning algorithms, was also employed by him to finish his job, improving accuracy. Prince Kansal et all's main focus was on how to predict heart disease early using several data mining techniques, and they employed about 4 machine learning algorithms. Age, sex, blood pressure, and blood sugar are some of the parameters they used in their dataset. The decision tree helped them have improved accuracy. Asha manually diagnosing Rajkumar's heart ailment took a lot of time and required specialists. They concentrated on her research, which used data mining to quickly diagnose heart problems. They employed tanagra tools and three algorithms. The Naive Bayes algorithm took 609ms to diagnose the heart condition. Salman and Issam used various machine learning techniques to primarily concentrate on mortality prediction due to a heart attack. He used real-time data to do his assignment. He used a variety of classifiers, such as NB-tree and Bayesian networks, to improve accuracy. Jesse Davis and others, eMedical has transformed the way we currently handle patients and made it more simple to get patient data online. Today, we cure various diseases online thanks to eMedical. We frequently don't know if this medication will have any other side effects because of us. The risk of a heart attack in eMedical is the main topic of this essay. The work of Himanshu Sharma et al. summarized the most recent methods and technologies for heart attack illness prediction. They have essentially concentrated on comparative study of heart disease identification and have employed both machine learning and deep learning ideas. Beant Kaur et al. Present assessments of numerous studies on the prediction of cardiac disease, and one publication shows that neural networks produce better results almost 100% of the time. Using soft computing technology and data mining techniques, Gayathri et al. Concentrated on the diagnosis of many forms of cardiac disorders (cardiovascular disease, coronary artery disease, heart failure, ischemic heart disease, etc.). Their biggest drawback is that they don't consistently produce decent results as the dataset grows. We concentrated on how to evaluate heart attack possibility using new machine learning algorithms, despite the fact that there has been a lot of prior research on heart attack risk prediction.

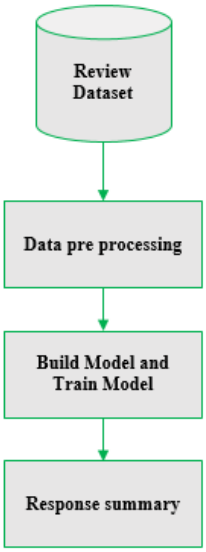
III. METHODOLOGY

1. Input Data: 302 data were used in this study. 14 attributes are observable in this situation, and they are all floating data. A decision class/class variable is also present. This information was gathered via Kaggle.
2. Correction Metrics: Correction metrics are displayed in fig. 2.
3. Split Data: In this study, 70% of the data were used to train the model and 30% were used for testing.

4. Classifier: Decision tree, Naive Bayes, Gradient Boosting, Random Forest, and Base Classifier random forest are employed.
5. Evaluation Results We were able to see the confusion matrix results for all algorithms in the classification report.
6. Best Algorithms: The best algorithms were used in this study, which rely in part on the research. However, the best algorithms in this analysis are those that provide the best true positive, false positive, true negative, and false negative results.

IV. PROPOSED MODEL

This figure indicates the each segment of our proposed model.



V. ALGORITHMS DESCRIPTION

The algorithms used for this research are listed below.

- Linear regression: Linear regression analysis is utilized to anticipate the esteem of a variable based on the esteem of another variable. The variable you need to foresee is called the subordinate variable. The variable you're utilizing to foresee the other variable's esteem is called the free variable. This shape of investigation gauges the coefficients of the direct condition, including one or more free factors that best foresee the esteem of the subordinate variable. Linear regression fits a straight line or surface that minimizes the errors between anticipated and genuine yield values. There are simple linear regression calculators that use a "least squares" method to find the best-fit line for a set of combined data. Then we can estimate the value of X (dependent variable) from Y (independent variable).

- Logistic regression: Logistic regression is a well-known machine learning algorithm that falls under the procedure of supervision learning. Since Logistic regression and linear regression are directed by nature, these algorithms utilize labeled datasets to create expectations. Factors are controlled within the regression analysis to isolated the relationship. When utilizing linear regression and logistic regression within the case of nonlinear regression ships, as it were linear regression can be utilized. In arrange to unravel regression issues, logistic regression is utilized, whereas linear regression is utilized to illuminate classification issues.

- Decision tree: A shaped diagram is utilized to assess a course of activity is a Decision Tree. Each department of the tree could be a potential alternative, event or response. We consider the full preparing set as the root at the starting. For data advantage, traits are accepted to be categorical and traits are accepted to be nonstop for the gini file. Records are recursively disseminated on the premise of property values. As a root or internal node, we utilize statistical methods to arrange attributes.

- SVM classifier: An SVM classifier or support vector machine classifier, is a sort of machine learning algorithm that can be utilized to analyze and classify data. A support vector machine may be an administered machine learning algorithm that can be utilized for both classification and regression assignments. The Support vector machine classifier works by finding the hyperplane that maximizes the edge between the two classes. The Support vector machine algorithm is additionally known as a max-margin classifier. Support

vector machine may be a capable apparatus for machine learning and has been broadly utilized in numerous assignments such as hand-written digit acknowledgment, facial expression acknowledgment, and content classification. Bolster vector machine has numerous preferences over other machine learning algorithms, such as strength to commotion and the capacity to handle huge datasets.

- Naive Bayes: Naive Bayes is a machine learning demonstrate that's utilized for expansive volumes of data, indeed on the off chance that you're working with data that has millions of data records the suggested approach is Naive Bayes. It gives exceptionally good results when it comes to NLP assignments such as sentimental analysis. It may be a quick and uncomplicated classification algorithm.

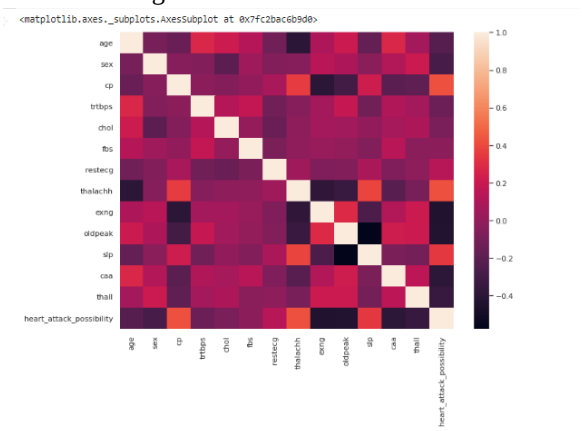
- The k-nearest neighbors: The k-nearest neighbours calculation, moreover known as KNN or k-NN, is a non-parametric, supervised learning classifier, which employments nearness to form classifications or expectations approximately the gathering of an person information point. Whereas it can be utilized for either regression or classification issues, it is ordinarily utilized as a classification algorithm, working off the presumption that comparable focuses can be found close one another.

- Random forest: A technique known as random forest or forest random decision uses numerous decision trees to provide results during training. The majority opinion of the trees is chosen for the ultimate choice by the random forest.

- Gradient Boosting: In Gradient Boosting, each indicator tries to make strides on its forerunner by lessening the mistakes. But the interesting thought behind Gradient Boosting is that rather than fitting a predictor on the data at each emphasis, it actually fits a new predictor to the remaining blunders made by the previous predictor.

VI. IMPLEMENTATION OF PROJECT

Feature scaling:



Dataset:

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	heart_attack_possibility
0	62	Male	atypical	130.0	231	NaN	1	NaN	0	1.8	1	3	3	more
1	54	Male	typical	108.0	309	False	1	156.0	0	0.0	2	0	3	more
2	57	NaN	asymptomatic	130.0	131	False	1	115.0	1	1.2	1	1	3	less
3	34	Female	typical	118.0	210	False	1	192.0	0	0.7	2	0	2	more
4	61	Male	non-anginal	134.0	234	NaN	1	NaN	0	2.6	1	2	2	less
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	58	Male	asymptomatic	125.0	300	False	0	171.0	0	0.0	2	2	3	less
299	56	Female	typical	140.0	294	NaN	0	NaN	0	1.3	1	0	2	more
300	58	Male	asymptomatic	100.0	234	False	1	156.0	0	0.1	2	1	3	less
301	55	Female	NaN	128.0	205	NaN	2	130.0	1	2.0	1	1	3	less
302	62	Female	asymptomatic	160.0	164	False	0	145.0	0	6.2	0	3	3	less

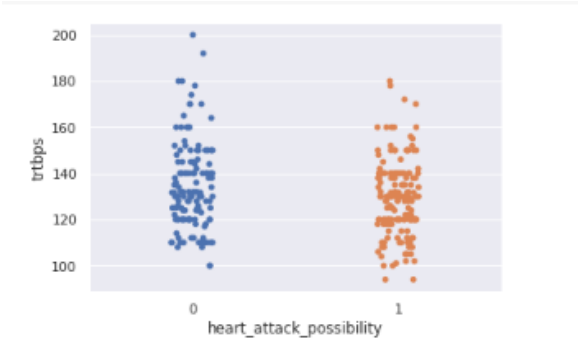
303 rows × 14 columns

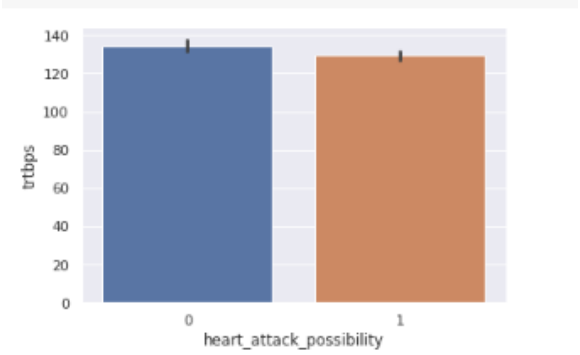
Removing null values:

```
df.isnull().sum()
```

```
age                       0
sex                       0
cp                        0
trtbps                    0
chol                      0
fbs                       0
restecg                   0
thalachh                  0
exng                      0
oldpeak                   0
slp                       0
caa                       0
thall                     0
heart_attack_possibility  0
dtype: int64
```

Plots:





Algorithms:

linear regression

```
[ ] from sklearn.linear_model import LinearRegression
    lrg=LinearRegression()
    lrg.fit(x_train,y_train)
```

```
LinearRegression()
```

```
[ ] lrg.score(x_test,y_test)*100
```

```
47.63366419403454
```

Logistic Regression

```
from sklearn.linear_model import LogisticRegression
lgr=LogisticRegression()
lgr.fit(x_train,y_train)
```

```
[ ] lgr.score(x_test,y_test)*100
```

```
83.60655737704919
```

Decision tree classifier

```
[ ] from sklearn.tree import DecisionTreeClassifier
    dtc=DecisionTreeClassifier()
    dtc.fit(x_train,y_train)
    dtc.score(x_test,y_test)*100
```

```
78.68852459016394
```

SVC

```
[ ] from sklearn.svm import SVC
    svc=SVC(gamma=1,C=1)
    svc.fit(x_train, y_train)
    svc.score(x_test, y_test)*100
```

```
40.98360655737705
```

Naive Bayes

```
[ ] from sklearn.naive_bayes import MultinomialNB
    mnb = MultinomialNB()
    mnb.fit(x_train, y_train)
    mnb.score(x_test, y_test)*100
```

```
80.32786885245902
```

KNNeighbors Classifier

```
[ ] from sklearn.neighbors import KNeighborsClassifier, KNeighborsRegressor
    knnC = KNeighborsClassifier(n_neighbors=10)
    knnR = KNeighborsRegressor()
```

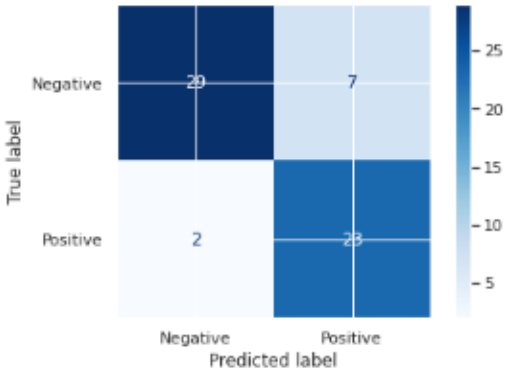
```
[ ] knnC.fit(x_train,y_train)
    knnC.score(x_test,y_test)*100
```

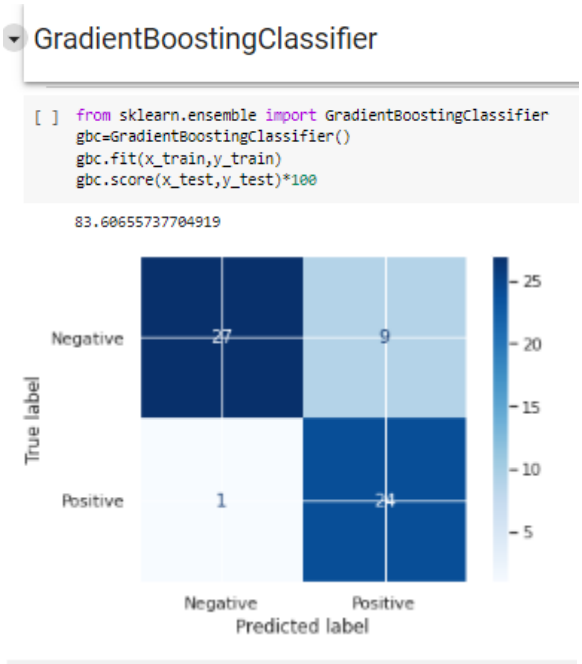
```
62.295081967213115
```

RandomForestClassifier

```
[ ] from sklearn.ensemble import RandomForestClassifier
    rfc=RandomForestClassifier()
    rfc.fit(x_train,y_train)
    rfc.score(x_test,y_test)*100
```

```
85.24590163934425
```





VII. RESULT

The researchers are using different algorithms such as SVC, Decision tree for the detection of patients diagnosed with Heart disease, Support vector, Decision tree, KNN, Naïve bayes, Random Forest Classifier and Logistic regression yield a better result to out rule them. The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by Random Forest is 85.25% which is greater accuracies obtained from other algorithms. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took. Our project also tells us that Logistic Regression, KNN, Support vector, Naïve bayes Classifier, Decision tree and Random Forest Classifier in the prediction of the patient diagnosed with a heart Disease. This proves that Random Forest Classifier is best in diagnosis of a heart disease.

VIII. CONCLUSION

Heart attack prediction model has been developed using eight ML classification modeling techniques. This project predicts people with heart attack by extracting the patient medical history leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, cholestoral level, blood pressure, resting electrocardiographic results, maximum heart rate achieved etc. This Heart attack prediction system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in building the given model are linear regression, logistic regression, decision tree classifier, SVC, naive bayes, KNN classifier, random forest classifier and gradient boosting. The accuracy of our model is 85.25%. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not. By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much. There are a number of medical databases that we can work on as these Machine learning techniques are better and they can predict better a human being which helps the patient as well as the doctors. Therefore, in conclusion this project helps predict the patients who are diagnosed with heart diseases by cleaning the dataset. It is concluded that accuracy of Random Forest is highest between eight algorithms that we have used i.e. 85.25%.

REFERENCES

Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-8

Christo Ananth, S.Shafiqa Shalaysha, M.Vaishnavi, J.Sasi Rabiyathul Sabena, A.P.L.Sangeetha, M.Santhi, "Realtime Monitoring Of Cardiac Patients At Distance Using Tarang Communication", International Journal of Innovative Research in Engineering & Science (IJIRES), Volume 9, Issue 3,September 2014,pp-15-20.

Dangare C S & Apte S S (2012).Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-8.

Christo Ananth, G.Poncelina, M.Poolammal, S.Priyanka, M.Rakshana, Praghash.K., "GSM Based AMR", International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Volume 1,Issue 4,July 2015, pp:26-28

Salman, Issam. "Heart attack mortality prediction: an application of machine learning methods." Turkish Journal of Electrical Engineering Computer Sciences 27.6 (2019): 4378-4389.

Sharma, Himanshu, and M. A. Rizvi. "Prediction of heart disease using machine learning algorithms: A

survey." International Journal on Recent and Innovation Trends in Computing and Communication 5.8 (2017): 99-104.

Kaur, Beant, and Williamjeet Singh. "Review on heart disease prediction system using data mining techniques." International journal on recent and innovation trends in computing and communication 2.10 (2014): 3003-3008.

Gayathri, P., and N. Jaisankar. "Comprehensive study of heart disease diagnosis using data mining and soft computing techniques." (2013).