Recognize the Most Effective Exploratory Data Analysis and Machine Learning Methods for Predicting the Chances of Having a Heart Attack

Sm Mahamudul Hasan, Md Forhad Rabbi and Mohammad Rubbyat Akram

August 31, 2021

Recognize the Most Effective Exploratory Data Analysis and Machine Learning Methods for Predicting the Chances of Having a Heart Attack.

***Abstract*—Every year, around 17.9 million people die as a result of cardiovascular illnesses, the majority of which are heart attacks or strokes. As a result, it is critical to keep track of the most prevalent symptoms and health behaviors related with cardiovascular disease (CVD). These tests can take a long time, especially if a patient's health is urgent and they need to start taking medication right away, therefore they must be prioritized. A variety of hazardous behaviors contribute to the development of heart disease. As a result, determining which risk factors for CVD exist is essential. Exploratory Data Analysis and Machine Learning enable the extraction of information from enormous amounts of data that would be hard to process manually. This article will review diagnostic testing and go through a variety of cardiovascular risk factors. The essay is chock-full of cutting-edge data analysis and machine learning techniques. We demonstrate the use of exploratory data analysis (EDA) and other techniques such as logistic regression, KNN, decision trees, Random Forest, support vector machines, gradient boosting, XG Boost, MLP classifier, and AdaBoost classifier. The collection has 303 samples, each with 14 distinct features. The chance of contracting the disease increases significantly, and a number of diagnostic criteria are used to assess the disease's diagnostic accuracy. The XG boost classifier outperforms all other models in the data set, with an accuracy of 95.08 percent. It is capable of properly identifying 96.55 percent of those who are at danger. With a total accuracy of 93.44 percent, Adaboost and the MLP classifier performed admirably in our classification model. This study highlights the most promising machine learning approaches for predicting the likelihood of heart attacks. We demonstrated that adopting some approaches might be useful for implementing preventative measures for heart disease patients.**

***Index Terms*—cardiovascular disease (CVD), heart attack prediction, healthcare, exploratory data analysis, machine learning, classification algorithms.**

I. Introduction

According to the World Health Organization, 17.9 million people will die from cardiovascular diseases in 2019 [1] , with heart attacks and strokes accounting for the bulk of those deaths [2]. As a consequence, it's crucial to keep track of the most common CVD symptoms and healthy habits. These tests may take a long time, particularly if a patient's health is critical and they need to start taking medication right away [3], so they must be prioritized. Heart disease is caused by a variety of harmful habits [4]. As a result, it's critical to know which health habits lead to CVD. As a consequence of the abundance of data, machine learning and data science are rapidly expanding areas of study. Humans can now extract information from huge quantities of data via machine learning [5] [6] [7] ,which was previously inconceivable. According to 2013 research, 2.5 exabytes of data is generated every day [8]. 90% of the data was created during the past two years (2.5x1018 bytes) [9]. Exploratory data analysis (EDA) uses advanced techniques to better comprehend a dataset, identify anomalies, and build sparse models to test underlying assumptions [10]. Physical examinations, patient symptoms, and signs are all used to predict and diagnose heart disease. Heart disease is influenced by cholesterol, smoking, obesity, family history, blood pressure, and the work environment [11] [12] [13]. Machine learning methods, on the other hand, are essential for accurate cardiac disease prognosis, and this approach may now be utilized in combination with big data technologies [14] [6] to handle unstructured data.

This article ranks diagnostic tests and covers several cardiovascular disease risk factors. Furthermore, the different machine learning algorithms are contrasted using sophisticated optimization techniques. This article utilizes data that has been manually classified. 70% of the data in this article is evaluated using supervised learning, while 30% is evaluated using classification. This study employs K-NN, SVM, Random Forest, logistic regression, gradient boosting, MLP classifier, adaboost, XGBoost, and decision tree classifiers. For prediction and classification of heart disease datasets, the XGBoost classifier beats the Adaboost and MLP classifiers.

The remainder of this article is organized as follows. The second section covers current research. Section III goes through some of the finer points of our strategy. The description of datasets is covered in Section IV. Our results from exploratory data analysis are discussed in Section V. The description of the prediction models is found in Section VI. Experiments and results are discussed in Section VII. The conclusion of our model is discussed in Section VIII, and data and code are discussed in Section IX.

II. RELATED WORK

Heart attacks are the leading cause of mortality in the United States, accounting for 4% to 10% of all fatalities in those under the age of 45 [15]. Poor circulation or pump failure may cause heart failure in newborns, infants, toddlers, and adolescents [16]. Academics have long been fascinated by the notion of using machine learning and data analysis to identify heart illnesses. This chapter compiles all of the current research that has been conducted by authors and researchers.

Karthick et al. [18] enhanced the study technique for predicting the probability of getting Cardio Vascular Disease (CVD) in individuals under the age of fifty (50). This technique may aid in the early detection of cardiac issues or attacks in individuals over 50, perhaps lowering or avoiding mortality. Based on age, gender, blood pressure, cholesterol, and pulse rate, Thomas et al [19] utilized data mining to estimate a person's risk of heart disease. Data mining techniques such as Nave Bayes, K-Nearest Neighbors, Decision Tree Algorithm, and Neural Networks were used to do this. The disease was classified using the decision tree technique, and the probability was forecasted using the Gaussian algorithm.

Using automated learning techniques, Parthiban et al [20]. detected cardiac problems in diabetic individuals. WEKA is used to apply the Nave Bayes and SVM algorithms. The Chennai Research Institute provided a data collection of 500 patients for this study. There are 142 individuals who have the illness and 358 people who do not. The Naive Bayes algorithm has a 74 percent accuracy rate. SVM has the greatest accuracy, with a rate of 94.60 percent. The Gaussian Nave Bayes, Restreet, and Bayes Network models for classification and regression were suggested by Hlaudi Daniel Masethe et al [20]. The information comes from hospitals and doctors in South Africa. Patients' names, ages, genders, and symptoms such as chest pain, a cardiogram (blood pressure measurement), heartbeat, and cholesterol levels were all recorded. It's a tool for predicting heart attacks.

The Hidden Nave Bayes (HNB) classifier was created by Jabbar et al [21]. to identify heart illnesses. The dataset from the University of California, Irvine repository was used 270 times. WEKA 6.4 was used to conduct the experiment. The classifier, Nave Bayes, was hidden. Accuracy, specificity, sensitivity, and positive predictive value were used to evaluate its performance. The sensitivity, specificity, and positive predictive value of the Hidden Nave Bayes model are all 100 percent. M Wang, X Yao, Y Chen [23] developed a method of lowering the high death rate by anticipating heart attack and stroke patients early using data analysis. Because stroke patients having heart attacks are in the minority of stroke patients, data on stroke patients in the Intensive Care Unit is skewed. SVM and ANN were utilized by S Radhimeenakshi [22] to predict the risk of heart disease. The stat log databases at Cleveland Heart and UCI Machine Learning provided the data. It contained 14 distinct features, 13 of which were connected to the repository and prediction. The Stat log dataset was forecasted and classified using the SVM model. ANN has an accuracy and precision of 81.8 percent and 83.3 percent, respectively, compared to 84.7 percent and 85.6 percent for ANN. As a result, SVM outperforms ANN. ANNs offer higher specificity and flexibility for extended training since they can handle non-linear data and a large number of variables.

III. OUR APPROACH

Machine learning and exploratory Data analysis is used to make heart attack predictions feasible. To select the most probable candidates, our model employs a number of different methods, including logistic regression, KNN, decision trees, Random Forest, support vector machines, gradient boosting, XG Boost, MLP classifier, and AdaBoost classifier. The dataset was produced using Kaggle [24].

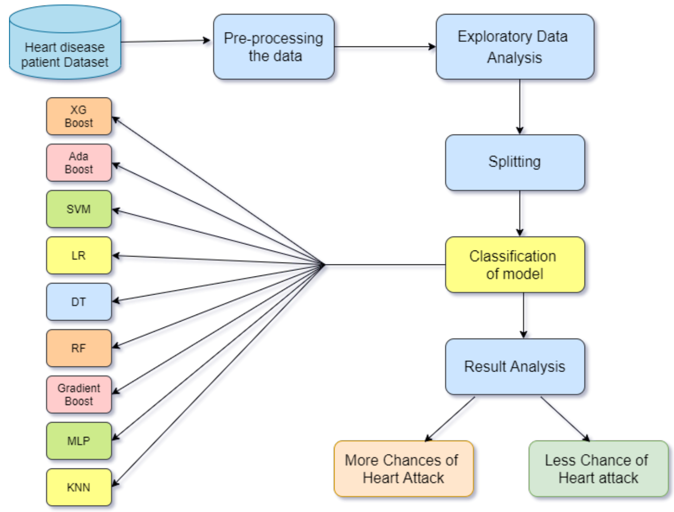

Fig. 1. Diagram of the Whole Process

The collection contains 303 patient records and information on 14 risk factors for heart attacks and strokes. Following data collection, we preprocess the data and substitute new values for any missing values in the dataset. We trained the model on 70% of the dataset and tested it on the remaining 30%. With an accuracy of about 95.05%, our prediction system successfully predict heart disease. This technique can correctly identify 96.55% of heart attack victims. Adaboost and the MLP classifier worked well in our classification model, with a total accuracy of 93.44%. Using exploratory data analysis and a machine learning classification system, we are able to predict cardiac attacks with high accuracy.

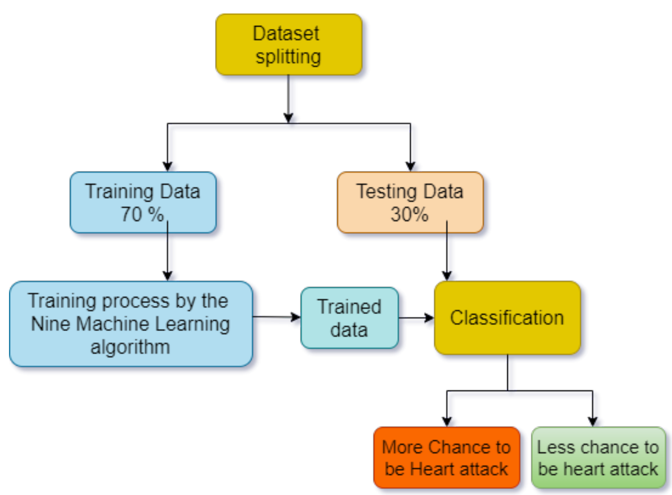

Fig. 2. Diagram of the Classification Process

To determine if a patient is at risk of suffering a heart attack, we used the data from the preceding phase to train nine machine learning classification models. We produced an age chart, dataset corellation heatmap, plotting all the categorical and numeric feature values, a KDE plot, an age vs CP ratio plot and a target plot during the exploratory data analysis phase, in addition to a target plot.

IV. DATASET

Our dataset includes 303 patient records with 14 distinct patient variables. 41 unique age counts, 2 unique sex counts, 4 unique chest pain types, 49 unique blood pressure types, 2 fbs, 3 unique restecg types, 91 unique thalachh and 2 unique exngs. All of the features and their descriptions are included in Table 1. The number of unique values is given below.

TABLE I
DATASET DESCRIPTION

Attribute	Description	Type	Correlation value
Age	Patient age in year	Nominal	0.225439
Sex	Patient Gender (Male =1 and female = 0)	Nominal	0.280937
CP	Chest pain in 4 types: 1. Typical angina 2. atypical angina 3. Non-anginal pain 4. asymptomatic	Nominal	0.433798
Trtbps	Blood pressure level (in mm/Hg)	Nominal	0.144931
Chol	Cholesterol in mg/dl	Nominal	0.085239
Fbs	Blood sugar level on fasting>120 mg/dl. Level 1= true, level 0=false.	Nominal	0.028046
Resting	Result of electrocardiogram while rest. Value 0 = normal, value 1= abnormal, value2= hypertrophy.	Nominal	0.137230
Thalachh	Max rate of heart	Nominal	0.421741
Exng	Angina by exercise 0= no, 1= yes.	Nominal	0.436757
Oldpeak	Exercise included by ST depression.	Nominal	0.430696
Slp	ST segment measured. Value 1= unsloping, 2= flat, 3= down sloping.	Nominal	0.345877
Caa	Fluoroscopy colored from 0 to 1	Nominal	0.391724
Thall	Status of heart illustrate. Normal =3, fixed defect= 6, reversible =7	Nominal	0.344029
output	Heart diagnostic. 0= total absent, 1-4 = percentage of different degree.	Nominal	0.280937

From our dataset, we built an SNS diverging palette and computed the corellation values. All of our correlation coefficients are shown in Figure 3. As can be seen from the correlation matrix above, the correlation between characteristics is low. Chest Pain Type with Condition and Max. Heart Rate with Condition show a strong correlation coefficient of 0.43 and 0.42, respectively, in our dataset. Our characteristics have a high number of negative correlation coefficients, suggesting

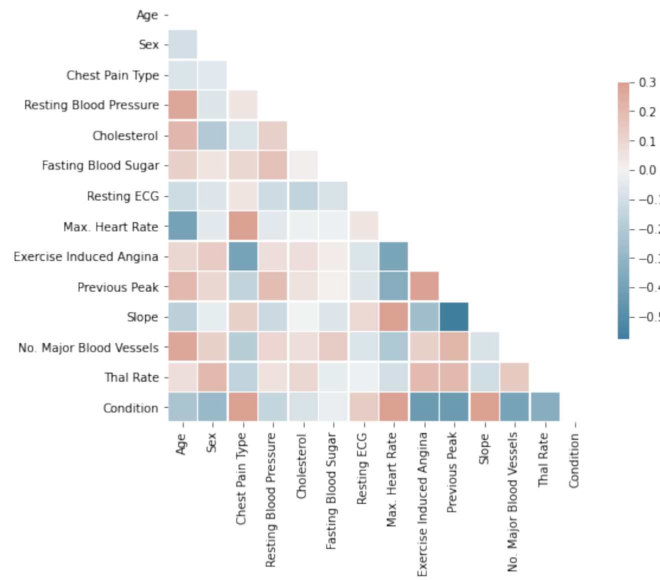

Fig. 3. Correlation Between the Features

that two distinct variables have a statistical connection in which they usually move in opposing directions.

We'll see that no column is exceedingly related to the abdicate, with a most extraordinary relationship of 0.436757 for "exng" and a least relationship of 0.028046 for "fbs". However, we are going to go ahead and utilize them all for our demonstration. Restecg, Thall, Caa, and Slp are likely to have an effect on a few models that are sensitive to data distribution. Fbs and Cp may have an effect on models that are sensitive to data distribution, since their value counts are not optimal. Trtbps and Chol are unlikely to be associated. Thalachh and age may have a modest connection. Oldpeak is almost certainly linked. Chol, rtbps, and Oldpeak all include a smattering of outliers.

V. EXPLORATORY DATA ANALYSIS

The study of data in order to draw conclusions is known as data analysis [25]. Analytics in healthcare analyzes individuals' diseases and hospital resource use. Data analysis may help in the creation of new research, disaster assistance, and disease outbreak management. It helps physicians, hospitals, and other health-care professionals [28] [29].

Blood pressure should be between 90/60mmHg and 120/80mmHg in a healthy person. Hypertension is defined as a blood pressure of 140/90mmHg or above [26]. Many individuals have cholesterol levels between 200 and 300 mg/dL. A total cholesterol level of less than 200 milligrams per deciliter (mg/dL) is recommended for adults [27]. The maximum heart rate of many individuals is between 140 and 160 beats per minute. The majority of this sample's respondents are between the ages of 40 and 70. (fig:4). Blood pressure values of 120 to 150 are common in many people. There is no obvious linear connection between continuous variables, as demonstrated by the heatmap.

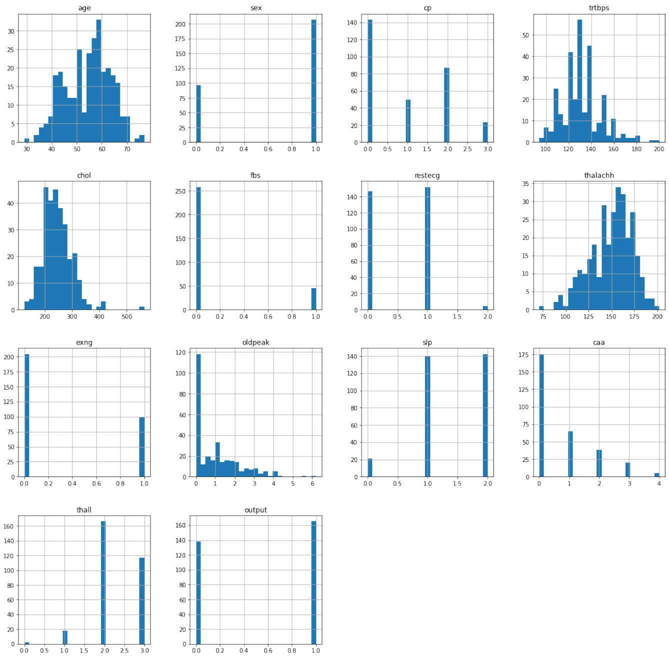

Fig. 4. plotting all the features of dataset

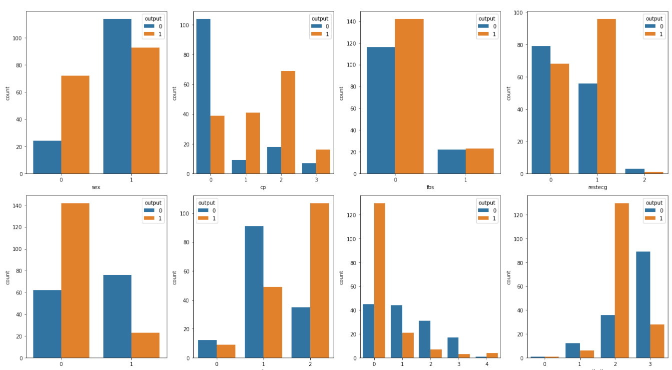

Fig. 5. Target Distribution of the Categorical Features

The risk of heart disease is greatest between the ages of 50 and 60. Because rtbps (resting blood pressure) is evenly distributed, it has a negligible impact on health risks. It is hazardous to have a heart rate of 140 to 180 beats per minute. The oldpeak values of high-risk patients ranged from 0 to 1. Chol, rtbps, and oldpeak all have a lot of outliers, which may cause problems for models that are sensitive to them.

The above-mentioned KDE graphs have been interpreted. The majority of individuals have heart attacks between the ages of 40 and 60. Following a heart attack, many people have Non-Anginal Pain or Type-2 Chest Pain. The maximum heart rates of several heart attack patients were from 140 to 180 beats per minute. A person with a cholesterol level of greater than 200 mg/dL has a higher chance of having a heart attack. There are a lot of outliers in Chol, rtbps, and Oldpeak. This might have an impact on models that are sensitive to them. The distribution of oldpeak and chol (moderately) is not

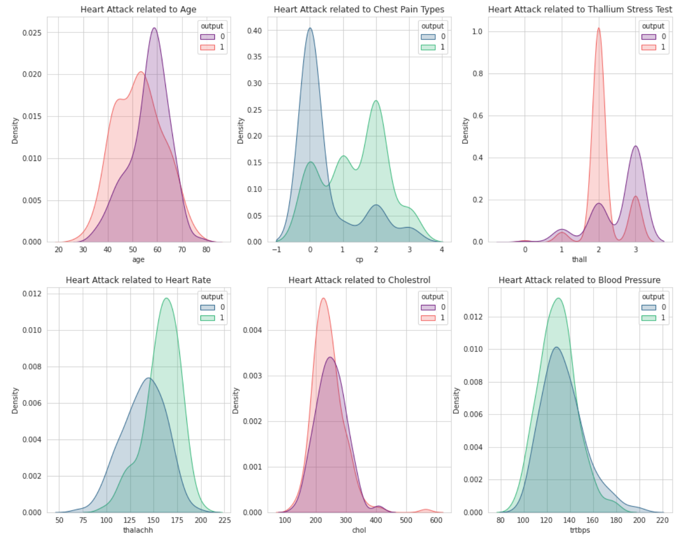

Fig. 6. KDE plot for Heart Attack in various circumstances

uniform. This might have an impact on models or analyses that need uniform distribution. Because value counts for certain values are extremely low, Restecg, Thall, Caa, and Slp are likely to have an influence on a few models sensitive to data distribution.

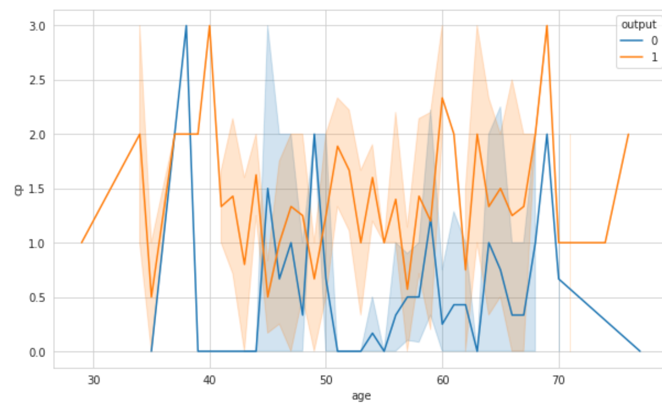

Fig. 7. Target Distribution of the Categorical Features

Females had a greater percentile of high-risk patients than men, while having a larger overall number of patients. There are four distinct kinds of chest pain, abbreviated as CP. The number of patients in CP 0 is the highest, whereas the number of high-risk patients is the highest in CP 2. Fasting blood sugar level (fps ¿ 120 mg/dl; a value of 1 indicates true; a value of 0 indicates false). The fps of high-risk people was 120 mg/dl. A resting electrocardiographic value of 1 indicates the greatest risk. Exercise-induced angina has a higher patient population (a value of 1 indicates yes, a value of 0 indicates no) (1 indicates yes, 0 indicates no). The most dangerous is slp, which has a value of 2. The Caa = 0 genotype has the greatest risk of developing heart disease. Those with a value of two have a higher number of patients.

VI. EXPERIMENTS AND RESULTS

For our study, we utilized the Google Coolab cloud runtime environment. For data visualization and analysis, we utilize pandas, numpy, matplotlib, and seaborn. For data preparation, use the standardscaller and MinMaxScaler packages, as well as the train test split model. We utilize logistic regression, K nearest neighbor, SVM, Decission Tree, and Random Forest Classifiers, The GradientBoostClassifier, Adaboost classifier, and XGboost classifier, and the MLP classification method. With the help of Sklearn, we calculated precision, recall, f1score, confussion matrix, accuracy score, and mean squared error. We use minmax scaller and use 30% of our data for testing and 70% for training.

TABLE II
PERFOMANCES OF THE MODELS

Classifiers	Prec.	Rec.	F1 Score	Ex. time	Mean square error	Acc.
LR	0.882	.938	.909	.01548	.098	90.164
KNN	0.879	.906	.892	.00746	.115	88.525
SVM	0.882	.938	.909	.00584	.098	90.164
Decision Tree	0.818	.844	.831	.00371	.18	81.967
Random Forest	0.935	.906	.921	.48786	.082	91.803
AdaBoost	0.967	.906	.935	.04117	.066	93.443
Gradient Boost	0.966	.875	.918	.02359	.082	91.803
XGBoost	0.968	.928	.952	.01476	.049	95.082
MLP	0.938	.938	.938	.96808	.066	93.443

For scalling the data, we utilize random state 65. For logistic regression, we utilized random state 10. After that, we computed up to four fractional digits. We use a k-neighbor value of =15 for the KNeighbour classifier. We utilize the ”rbf” kernel for SVM. For decision tree cl; assifier, we employ a random state value of 10 and the ’gini’ criteria. In our approach, the decision tree has a maximum depth of 100 levels. We utilize a random state value of 5 and a n estimator value of 300 for the random forest classifier. In the random forest classifier, the maximum depth is 100. The Adaboost classifier employs a learning rate of 0.15 and n estimator values of 25. The random state value is 10, the n estimator value is 20, the learning rate value is 0.29, and the ’deviance’ loss value is utilized in the Gradient Boosting method. The XGboost classifier uses a learning rate of 0.1, a max depth of 1, a n-estimator value of 50, and a colsample bytree value of 0.5 to get the goal as a binary logistic value. We utilize random state values of 48, hidden layer sizes of (150,100,50), and a max iteration value of 150 for the neural network-based MLP classifier model. In our MLP classification method, we utilize the’relu’ activation and the ’adam’ solver.

Table 2 shows the precision, recall, execution time, f1 score, mean square error, and accuracy of the proposed model. The accuracy score for each model is shown in Figure 9. In our machine learning technique, the XG boost algorithm surpasses all others, identifying 96.55 percent of genuine high-risk heart disease. The Adaboost and MLP classifiers worked well on our sample. However, Adaboost detects it more accurately than the MLP classifier. In our model, the decision tree does not work correctly.We calculated our process by this equations:

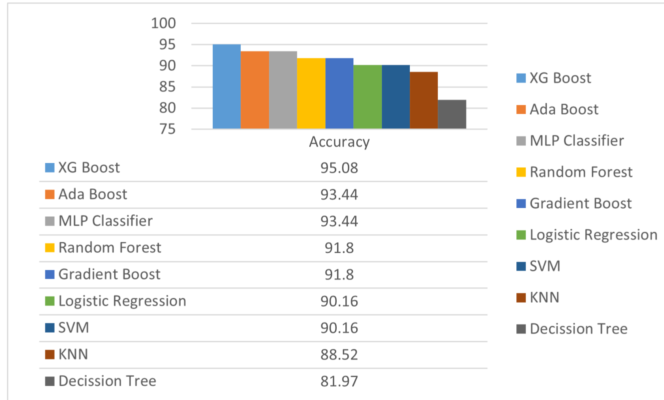

Fig. 8. Accuracies of the Models

Precision is defined as (TP)/(TP + FP).
Recall = (TP)/(TP+FN)
F1 score = (2 * Precision * Recall)/(Precision + Recall).

During the data processing phase, we counted various numeric factors to create a target count, a KDE plot, a scatterplot, and an age vs chest pain graph, among other things. The data analysis section has yielded some findings. The data has no null values. There are more people with sex = 1 in the data set than people with sex = 0. The heatmap shows that there is no clear linear relationship between continuous data points. It seems from the target distribution plot in fig. 6 that older people have a higher risk of heart attack. The age vs. output distribution graph, on the other hand, shows that this is not the case.

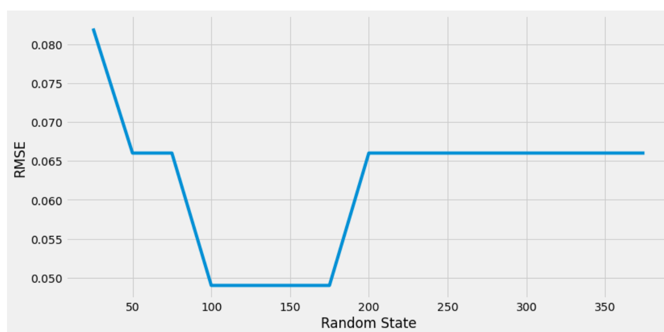

Fig. 9. RMSE Curve of the Models

Figure 9 has been shortened as RMSE (prediction errors). In our model, the root mean square error (RMSE) curve is between 100 and 200 for values ranging from 0 to 360.

A higher maximum heart rate increases the risk of suffering a heart attack, according to the thalachh vs. output distribution plot. Those with lower prior peaks had a higher risk of heart attack, according to the plot of oldpeak against output. Caa = 0 people have a much higher chance of having a heart attack. As sex equals one, the risk of a heart attack increases. People with thall = 2 have a higher chance of having a heart attack. People who have exng = 0 (no exercise-induced angina) are more likely to have a heart attack. As demonstrated in Fig7, non-anginal chest pain (cp = 2) increases the risk of a heart attack.

VII. CONCLUSION

For our research, we utilized the Google Coolab cloud runtime environment. We utilize logistic regression, K nearest neighbor, SVM, Decission Tree, and Random Forest Classifier as machine learning fundamental models, as well as boosting algorithms like XGboost, adaboost, and gradient boosting algorithms, and the MLP classifier neural network model. From the machine learning models, the XG boost classifier outperforms all other models in the data set. This method accurately identifies 96.55 percent of victims of heart attacks. With an overall accuracy of 93.44percent, Adaboost and the MLP classifier performed well in our classification model. The decision tree does not function properly in our model. The XG Boost outperforms all of our models due to its flexible and versatile nature.

The data analysis phase produced plotting all the features, terget destribution of the categorical features, a KDE plot and an age vs chest pain graph. Those with low blood pressure (rtbps) have a very high risk of heart attacks. The presence of exng (Exercised Angina induced angina) is associated with a lower chance of an attack. Restecg, Thall, Caa, and Slp are likely to have a few models that are data distribution sensitive due to the very low value counts for specific variables.The maximum heart rates of several heart attack patients were from 140 to 180 beats per minute. A person with a cholesterol level greater than 200 mg/dL has a higher chance of having a heartattack. The highest chance of having a heart attack is if you are over 50 years old. Exercise-Induced Angina affects more than 32% of the population. 47% of people have Typical Angina or Type 1 Chest Pain, whereas 28% experience Non-Anginal Chest Pain. Individuals with a normal resting ECG account for 48% of the population, whereas 50% have an ST-T wave abnormality.

REFERENCES

[1] Kolukısa, Burak, et al. ”Diagnosis of coronary heart disease via classification algorithms and a new feature selection methodology.” International Journal of Data Mining Science 1.1 (2019): 8-15.

[2] Amiri, Parisa, et al. ”Smoking habits and incidence of cardiovascular diseases in men and women: findings of a 12 year follow up among an urban Eastern-Mediterranean population.” BMC public health 19.1 (2019): 1-10.

[3] Ford, Joan S. ”Living with a history of a heart attack: a human science investigation.” Journal of Advanced Nursing 14.3 (1989): 173-179.

[4] World Health Organization. Technical package for cardiovascular disease management in primary health care: healthy-lifestyle counselling. No. WHO/NMH/NVI/18.1. World Health Organization, 2018.

[5] Sun, Alexander Y., and Bridget R. Scanlon. ”How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions.” Environmental Research Letters 14.7 (2019): 073001.

[6] Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. ”Machine Learning for Social Science: An Agnostic Approach.” Annual Review of Political Science 24 (2021): 395-419.

[7] Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. ”Machine Learning for Social Science: An Agnostic Approach.” Annual Review of Political Science 24 (2021): 395-419.

[8] Ansari, Seema, et al. ”Importance of big data.” Handbook of research on trends and future directions in big data and web intelligence. IGI Global, 2015. 1-19.

[9] Boleng, Jeff, and Colin Clarke. ”Big data challenges failed cities and the rise of the new net.” Matthew B. Ridgeway Journal for International Security Studies (2016).

[10] Hullman, Jessica, and Andrew Gelman. ”To design interfaces for exploratory data analysis, we need theories of graphical inference.” arXiv preprint arXiv:2104.02015 (2021).

[11] Sai, Xiao Yong, et al. ”Combined Effect of Smoking and Obesity on Coronary Heart Disease Mortality in Male Veterans: A 30-year Cohort Study.” Biomedical and Environmental Sciences 34.3 (2021): 184-191.

[12] Leander, Karin, et al. ”Family history of coronary heart disease, a strong risk factor for myocardial infarction interacting with other cardiovascular risk factors: results from the Stockholm Heart Epidemiology Program (SHEEP).” Epidemiology (2001): 215-221.

[13] Jensen, G., et al. ”Risk factors for acute myocardial infarction in Copenhagen II: Smoking, alcohol intake, physical activity, obesity, oral contraception, diabetes, lipids, and blood pressure.” European heart journal 12.3 (1991): 298-308.

[14] Mirza, Bilal, et al. ”Machine learning and integrative analysis of biomedical big data.” Genes 10.2 (2019): 87.

[15] Khan, Safi U., et al. ”Stroke-Related Mortality in the United States–Mexico Border Area of the United States, 1999 to 2018.” Journal of the American Heart Association 10 (2021): e019993.

[16] Hedlund, Eva, and Bo Lundell. ”Fontan circulation has improved life expectancy for infants born with complex heart disease over the last 50 years but has also resulted in significant morbidity.” Acta Paediatrica (2021).

[17] Obasi, Thankgod, and M. Omair Shafiq. ”Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases.” 2019 IEEE international conference on big data (big data). IEEE, 2019.

[18] Alsmadi, Tibra, Nour Alqudah, and Hassan Najadat. ”Prediction of Covid-19 patients states using Data mining techniques.” 2021 International Conference on Information Technology (ICIT). IEEE, 2021.

[19] Khourdifi, Youness, and Mohamed Bahaj. ”Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization.” International Journal of Intelligent Engineering and Systems 12.1 (2019): 242-252.

[20] Banu, NK Salma, and Suma Swamy. ”Prediction of heart disease at early stage using data mining and big data analytics: A survey.” 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT). IEEE, 2016.

[21] Jabbar, Ansam S. ”Hybrid PSO-RBFNN and Proposed Algorithms of DDDWT for the Heart Disease Classification.” Engineering and Technology Journal 39.4A (2021): 520-527.

[22] Wang, Meng, Xinghua Yao, and Yixiang Chen. ”An Imbalanced-Data Processing Algorithm for the Prediction of Heart Attack in Stroke Patients.” IEEE Access 9 (2021): 25394-25404.

[23] Radhimeenakshi, S. ”Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network.” 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2016.

[24] @misckeggle, author = Rashik Rahmen”, title = Heart Attack Analysis Prediction Dataset, url = https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset, year = 2021-03-22,

[25] Rakhra, Manik, and Ramandeep Singh. ”Smart data in innovative farming.” Materials Today: Proceedings (2021).

[26] https://www.nhs.uk/conditions/high-blood-pressure-hypertension/

[27] https://www.medicalnewstoday.com/articles/315900recommended-levels

[28] Awotunde, Joseph Bamidele, et al. ”Big Data and Data Analytics for an Enhanced COVID-19 Epidemic Management.” Artificial Intelligence for COVID-19. Springer, Cham, 2021. 11-29.

[29] Shah, Kesha, and Ana Tomljenovic-Berube. ”A New Dimension of Health Care: The Benefits, Limitations and Implications of Virtual Medicine.” Journal of Undergraduate Life Sciences 15.1 (2021): 10-10.