

Survival Analysis of Heart Failure Patients Using Machine Learning

By

S.M Rakibul Huda

181-15-1925

Nabid Hasan

181-15-1832

Fahad Ahsan

181-15-1914

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Md. Sabab Zulfiker

Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

S.M. Aminul Haque

Associate professor & Associate Head

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

October, 2021

APPROVAL

“Survival Analysis of Heart Failure Patients Using Machine Learning” project is submitted by *S.M.Rakibul Islam Asif* and *Nabid Hasan* and *Fahad Ahsan* to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfilment of the requirements for the degree of B.Sc. In Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *13/01/2022* .

BOARD OF EXAMINERS



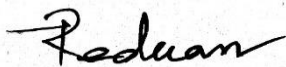
Dr. Touhid Bhuiyan
Professor & Head
Department of CSE
Faculty of Science and Information Technology
Daffodil International University

Chairman



Mr. Shah Md Tanvir Siddiquee
Assistant Professor
Department of CSE
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner



Md. Reduanul Haque (MRH)
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner




Dr. Dewan Md. Farid
Professor
Department of CSE
United International University

External Examiner

DECLARATION

We announce that, this project has been completed by under the supervision of **Md. Sabab Zulfiker**, **Lecturer, Department of CSE** Daffodil International University .We are declaring that, our project are completely done by us and any part of project are not submitted from any where else.

Supervised by:



Md. Sabab Zulfiker

Lecturer

Department of CSE

Daffodil International University

Co-Supervised by:



S.M. Aminul Islam

Associate professor & Associate Head

Department of CSE

Daffodil International University

Submitted by:

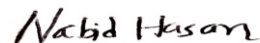


S.M. Rakibul Islam

ID :181-15-1925

Department of CSE

Daffodil International University



Nabid Hasan

ID :181-15-1832

Department of CSE

Daffodil International University



Fahad Ahsan

ID :181-15-1914

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

Firstly we are grateful to almighty Allah for giving us a opportunities to making possible our final year project auspiciously.

We really grateful to our Project Supervisor **Md Sabab Zulfiqer** (Lecturer) Department of Computer Science and Engineering (Daffodil International University) and Co- Supervisor **S.M. Aminul Haque** (Associate professor & Associate Head) Department of Computer Science and Engineering (Daffodil International University). For there Knowledge, interest, friendly behavior and innovative suggestion of our supervisor on the field of “*Artificial Intelligence*” to completely done the project.

He is a man with great attitude, much patience and his friendly behavior makes us more confident to do this project effectively.

We also want to mention some of our classmates of Daffodil International University who share us some important knowledge for completing this project.

I won't miss the appreciation of my parents by showing the interesting about our project. We are also grateful to our honorable faculty members for their deep knowledge, valuable suggestion, and contribution during the our project development.

ABSTRACT

In modern days heart attack has become a major disease all over the world which is causing huge number of death. Heart attack at early age is a greater risk factor for the people of south Asia than the other regions. It is really hard to predict the stage of a heart attack patient quickly and successfully, because it need a long lime experience and passionate knowledge. Medical industries has a large amount of data which can be used to make effective decision with all the concealed information. With the help of effective decision making and few excellent data mining technique like Logistic Regression, Decision Tree, we will able to predict heart attack patients situation or stage quickly. We used four algorithm in our system those are Random forest classifier, Decision tree, Logistic regression, Support vector machine. Our final model accuracy is 92% where the algorithm is Support Vector Machine(SVM).

TABLE OF CONTENT

Approval	i
Board of Examiners	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
Table of Content	v
List of Figures	vii
List of Tables	viii

CHAPTER

CHAPTER 1:INTRODUCTION

1-4

1. Introduction	1-2
2. Motivation	2
3. Rationale of theStudy	2
4. Research Questions	3
5. Expected Output	3
6. Report Layout	4

CHAPTER 2:BACKGROUND

5-10

1. Introduction	5
2. HeartFailuer	5,6
1. Definition	5
2. Causes and RiskFactors	5,6,7
3. Related Works	7
4. Research Summary	8
5. Scope of theProblem	9
6. Challenges	9,10

CHAPTER 3: RESEARCH METHODOLOGY 11-18

1. Introduction	11
2. Data Collection Procedure	11,12
3. Statistical Analysis	13
4. Research Subject and Instrumentation	14
1. Logistic Regression	14
2. Decision Tree (CART)	15
3. Support Vector Machine (SVM)	16
3.5 Selection of Algorithm	17
3.6 Proposed Algorithm	18

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION 19-21

4.1 Introduction	19
4.2 Experimental Results	19,20,21

**CHAPTER 5:
RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH SUMMARY,
CONCLUSION, AND RECOMMENDATION 22-24**

1. Summary of the Study	22
2. Conclusions	22
3. Future Work	23

References 24

CHAPTER 1

INTRODUCTION

1.1 Introduction

The major cause of heart attack death is a loss of blood flow to the heart muscle. A blood clot prevents the heart from receiving adequate oxygen. There are three different types of heart attacks. NSTEMI (non-ST segment elevation myocardial infarction), SEMI (ST segment elevation myocardial infarction), and unstable angina (coronary spasm).

Now a days heart failure disease is a great concerning thing for Bangladesh. Heart disease is the leading cause of hospitalization (National Health News). Heart disease has surpassed cancer as the leading cause of mortality worldwide. Heart attacks claim the lives of more individuals each year than any other disease. In the year 2019, 17.9 million people died as a result of cardiovascular disease. That accounted for 32 percent of all deaths. "Heart failure and stroke cover 85 percent of deaths among them." According to the World Health Organization WHO's country profiles for 2018, heart attack is the leading cause of non-contagious illness mortality in Bangladesh. or angina pectoris.

According to (WHO) World Health Organization dangerously increasing the number of deaths caused by heart disease in Bangladesh in past few years. Identifying heart failure has to be sophisticated. It is necessary that it must be done with good accuracy. In regular method doctors identify the condition of patient. This produce unwanted and huge medical bill. So automated software that can predict condition of heart patient accurately would be extremely helpful.

1.2 Motivation

After the heart attack occur , doctors need to take some initiative about patient from there condition. Doctors are not always available 24hr in one area. But heart attack can be occur anytime. If we has an system in every hospital like an “Survival Analysis of Heart Failure Patients”, Which will get reports and will give an result as output quickly about the situation of patient. Than it will be very helpful to take next steps by acknowledge from the system. Any one can check their heart status any time by this system. And people can be more aware about their heart disease.

1.3 The rationale of thestudy

Today Heart attack is a serious matter, and take the place of serious concern. It is really dangerous sickness because it can bring ending time of life at any time with out any symptom before. We have picked heart failure as our exploration point. To see the number of dead people because of heart attack our exploration point is chosen. Our goal is reduce the dead people number. Medical center will be able to use our system to predict heart failure condition of patients. We hope that it will help people to understand the extent of heart attack.

4. Questions about research

- a). Does that system gives the accurate result of survival prediction of a patient ?
- b). Does it classify survival status by machine learning algorithm?

Human beings is prone to many diseases. Every disease can be prevented because people can take steps to prevent them. But a lot of people are too unconsciousness. No one wants to live a dreadful life, where this disease is bring a big dread for all the time of life. Our system is able to predict almost accurate value for survival status.

We have made a prediction system where it will identify patients survival status. Among the leading disease heart failure is pretty high up. Lot of people including our relatives died of heart attack. Our system is able to classify survival status by machine learning algorithm. Heart failure kills tons of men and women worldwide. There are don't have a good or well known system for quickly detecting of heart failure situation In Bangladesh . That's why we select this topic as our project bass research topic to build and bring a good system of heart failure survival analysis .

5 Expected Result/Output

The Survival Analysis of Heart Failure Patient system will help to generate and result according to the given data .By using this system we can understand about situation or status about the patient quickly . In our system , we used 70% of the training in this system to generate better accurate predictions .The output result accuracy depends on the training dataset . If our data set are rich then the system will produce more accurate data from the system ,it also mattes on the relativity of data set .If our data set will be related to our prediction .We applied a lots of strategies to achieve a better accuracy and desired output.

6. Report's layout

- Chapter 1. Chapter 1 provided an overview of the study, including its rationale, research objectives, and planned outcomes.
- Chapter 2. The "Background" section of Chapter 2 will show the introduction, relevant works, research summary, and problems.
- Chapter 3. Research Methodology will be covered in Chapter 3.
- Chapter 4. Experimental Results and Discussion will be covered in Chapter 4.
- Chapter 5. Summary and conclusion will be included in Chapter 5.

CHAPTER 2

Background Study

1. Introduction

In this part we will discuss about heart failure disease deeply. Some works which are relative to us, Summary of research and the challenges of our system. In related works part we discussed about some other projects which is similar to our project. We also described the methodologies of our project in this section, We had faced a lot of challenges in our project which we discussed in Challenges section.

1. Heart failure

In this Part, we will discuss about some related issues of Heart disease.

2. Definition

When heart failure occurs heart will be incapable to pump blood around the body rightly . When heart has become to weak or inflexible the usually heart failure take place .There is a name of heart failure which is congestive heart failure , now a days it is not used broadly .When heart failure happens it doesn't mean necessarily that your heart completely stop working

3. Causes and Risk Factors

There are a lots of cause of heart attack. Numerous factors to contribute to heart disease. One measure cause is stress and being in an environment that cause stress can lead to heart attack a poor nutrient diet and not enough exercise can also be contributing factor to heart attack .

- High fat diet
- Nutrient deficiencies
- Lack of exercise
- Stress
- Infection
- Drug abuse
- Smoking

There are various risk factors for the disorder .Which are given below:

- Shortness of breath with physical activity or lying down
- Weakness
- Lump in the legs , feet etc.
- Speedy or rambling heart beat
- Decreased capability to exercise
- Increasing weight from fluid buildup
- Inconvenience decreased alertness
- If heart attack is the reason for heart failure then there can be some chest pain.
- Associate unhealthy diet & smoking.

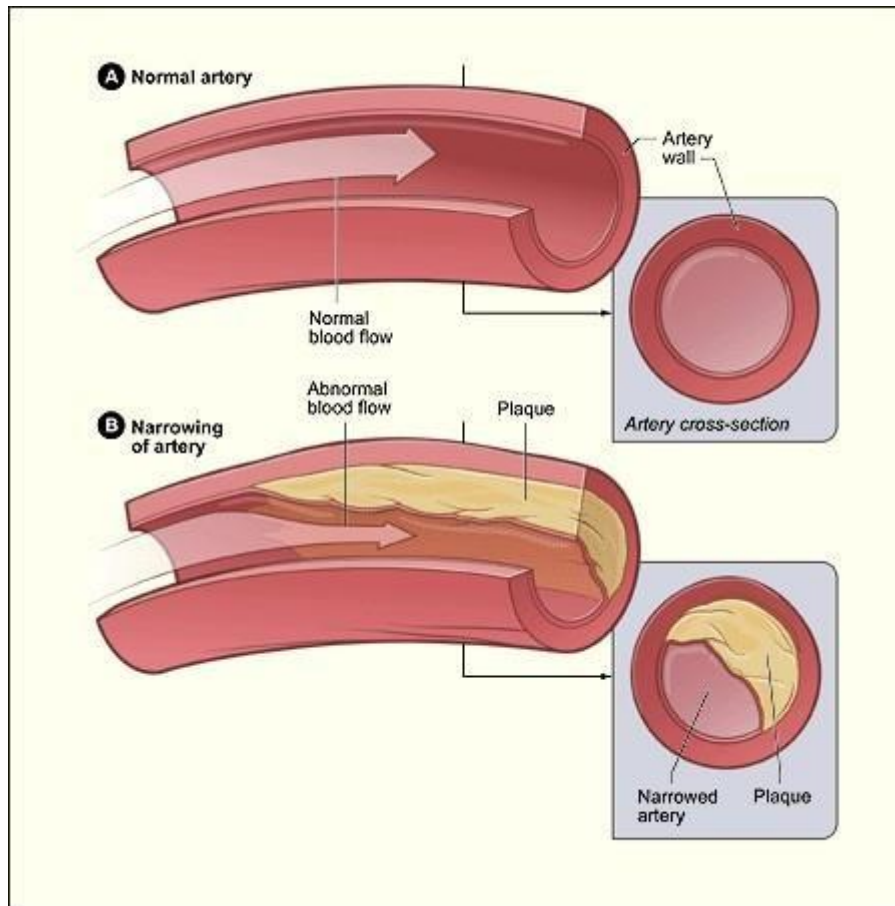


Figure 2.2.2: Plaque on blood vessel

2.2 Related Work

Tanvir et al. worked in Survival analysis of heart failure patients , Based on their thesis (96)32 percent patients died due toCHD.[1]

Davide et al. Worked on Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. They split the dataset into 80% (239 randomly selected patients) for the training set, and 20% (the remaining 60 patients) for the test set.[2]

Asha et al. used Naive Bayes, Decision List, KNN algorithm for diagnosis of patient with heart disease and they got best accuracy rate for Naive Bayes algorithm which was 52.33%[3].

GIOLO et al. worked in Survival analysis of heart failure

patients , Based on their thesis, they summarize the information that (37 percent) the patients have died at the end of the follow-up period. From them, 62.8 percent were men and 37.2 percent women .[4]

Ketut et al. used KNN, Naive Bayes and simple CART algorithm in their research and they got best result for KNN with the accuracy rate of 81.85% [5].

Asif et al. worked in Survival prediction of heart failure patients using machine learning techniques, The dataset was originally collected from the Faisalabad Institute of Cardiology, It contains the medical records of 299 heart failure patients. From their study they got finally 76.25% accuracy. [6]

4. Scope for the problem

We used some algorithms in our System, and we always focused in our accuracy but when we expected a better accuracy then we faced some problem in our algorithms .

Some algorithm gives us better accuracy but there were some drawbacks also some were time swallowing. So we were focused to find out some better algorithms which has better accuracy with shortage of time.

Some problem occurs in the time of searching for a best algorithm,

- Although we are inputting various types of data, But some algorithm was giving us same accuracy in the results.
- There was some problem in data implementation and algorithm etc.

5. Challenges

Mainly Data recruitment is best challenges for getting predicting . Data collection is main part because without data we can't predict anything. We tried to contact with, Bangladesh Heart Failure Foundation but didn't respond , after that we collect our data from kaggle.com. Data Processing and Cloning is the next step , After that null or empty value was reduced and data was clean. With respect value all of feature used by help of Feature scaling. Always we tried to reduce the number of false value for get a better accurate result.

There are some more challenges ,

- Finding the best algorithm
- Work on study shareholder.
- Working with our data
- Data Makin in excel file for input
- On the algorithms have to use dataset properly.

CHAPTER 3

RESEARCH METHODOLOGY

1. Introduction

Data from kaggle.com is collected in our system. In our project investigation, we attempted to find a unique process and produce accurate predictions. We used a total of 299 data points and 12 characteristics attributes. Then we discovered several incorrect feature values, which we had to correct. We've already completed the data cleaning and feature scaling processes in order to obtain accurate predictions. We utilized datasets for training and testing, and we employed algorithms including Logistic Regression (LR), Classification and Regression Trees (CART), support vector machines (SVM), Random Forest Classifier, and Naive Byes. These algorithms were utilized to create our model, which is based on a workingprocess.

2. Data Collection Procedure

We collected our data set from kaggle.com. Make our work unparalleled research. Also analysis number of research paper and found number of paper on the basis heart attack. Our research is for analysis of survival of heart attack in early uniquely. Now a days heart disease increases day by day. That's why we want to make a system for our people to reduce heart diseases death.

Features reference value of the dataset 3.1 –

Table3.1: Feature range of Dataset

Feature	Reference value	Feature	Reference value
High_Blood_Presure	<120/80	Serum_Sodium	1.3-1.5
Anemia	<200mg/dL	Sex	
Creatinine_Phosphokina se	<120	Smoking	
diabetes	120-200ms	Time	< .5
Ejection_fraction	170bpm	Age	>40
Plateletes	True		
Serum_creatinine	Absence		

We attempted to display 12 dataset characteristics attribute reference value in dataset table 3.2. High Blood Pressure (BP), Anemia, Creatinine Phosphokinase, Diabetes, Ejection fraction, Plateletes, Serum Creatinine, Serum Sodium, Sex, Age, Smoking, Time are among the dataset properties shown in table 3.1.

3.3 Statistical Analysis

In our dataset, we have 299 datasets where 12 features and target attribute.

DEATH_EVENT is a target attribute where difference ratio of data are high cause 200 data are survive event and 99 are not service. We had to look on it how to make it sample. We have to try to find out how many null and invalid value are there. But fortunately we had not get that kind of value. Age feature are more effective feature than any other. All of the feature data type are float and int type. How many time we get our a patient go to doctor also very effective feature to our model. 80% percent data a are use for train rest of are test. We are shown below how to collect and processed data through flowchart.

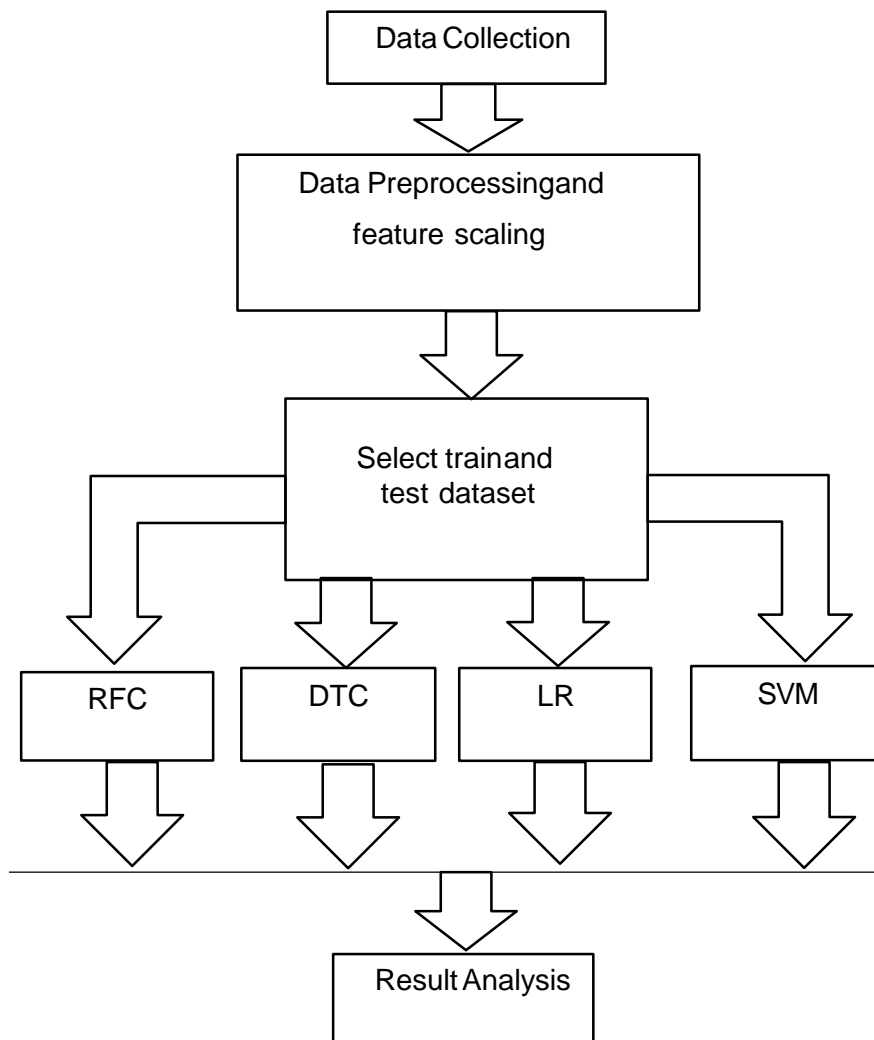


Figure 3.2: Architecture of our proposed model.

The figure have been shown the research procedure shortly in details. We can see, step by step procedure through this .

4. Research Subject and Instrumentation

The popularity of ML are increasing day by day. Measurable methodologies helps PCs to give the capacity of gain information using Machine Learning Algorithms. The most important part is discover all information example and produce a option or prescient learning ,without the help of coding by Machine Learning . To follow this, a same calculation can be relate the dataset of different areas without the change of the whole structures . There have different kind of machine learning of calculation, We have used some of them to our system. Algorithm details are given below:

1. Logistic Regression

Logistic Regression, is one of the Supervised Machine Learning. Classification are the main issues to solve by this. Twofold result is related with this for give a arrangement of free factor. The target variable is discrete. Insight the field of machine learning logistic backslide is a system to produce. Matched Portrayal issues is the technique. Depiction is used to calculated backslide uses of depiction. Specially like straight backslide.

Loads or coefficient regards to of a part a yield regard on regards of information to join straight. Key differentiates to one straight backslide is that the yield regard being exhibited is match of quality (0 or 1) rather than number regard.

Model key backslide condition is : $y = e^{(b_0 + b_1 * x)} / (1 + e^{(b_0 + b_1 * x)})$
.....(1)

From the equation, here predict yield is y , inclination is b_0 or capture term and the coefficient was for single data esteem (x) . b coefficient is related to a gained preparation information.

3.4.2 Decision tree (CART)

The main computation for predicting machine learning is the Decision Tree Classifier. Traditional decision tree calculations have been around for a long time, and the arbitrary timberland in day varieties is one of the most record-breaking techniques. CART stands for Classification and Regression Trees, and it is the name given to the decision tree. The decision tree approach is the most important asset for a machine since it produces persuasive results as quickly as time allows. Many distinct types of computations are distinguished by the choice tree: Cart, ID3, C 4.5, CHH, and H48. J48 is one of them, and it is becoming a popular algorithm. J48 constructs a tree using the pruning strategy. The recursive method is kept in place until the expected results are obtained. This provides correctness and versatility. This formula can be found in the equations below.

$$E = \sum_{i=1}^K P_i \log_2 P_i \dots \dots \dots (2)$$

From equation 2,
 K denotes the number of target characteristics,
 P_i denotes the number of existing classes, and I denotes the total number of instances.
 This count is commonly referred to as "Decision Trees," however on a pair of stages such as R , it is referred to as CART.

3.4.3 Support Vector Machine(SVM)

Support Vector Machine properly describes a detaching hyperplane as a differentiative classifier. A perfect plane is created as a result of the computation, which arranges new points of reference. This hyperplane is a plane in two-dimensional space that is right a plane in a couple of locations where each class is one side of the number of data elements in data shape an n-dimensional place. If you had two data components, for example, this would outline a two-dimensional space.

A line that informs the variable space is called a hyperplane. In SVM, class 0 or 1 is the best way to split the attention on the right information variable space. You can see a picture of the line in two dimensions, and we can expect the bulk of our data points to be completely separated by the line.. For instance:

$$B_0 + (B_1 * X_1) + (B_2 * X_2) = 0 \dots\dots\dots(3)$$

The learning method finds the coefficients (B1 and B2) that pick the inclination of the line and the capture (B0) from the equation, and the two information are X1 and X2.

Selection of Algorithm:

In this section we use variant algorithm to get higher accuracy. We are showing our all implemented algorithms accuracy below:

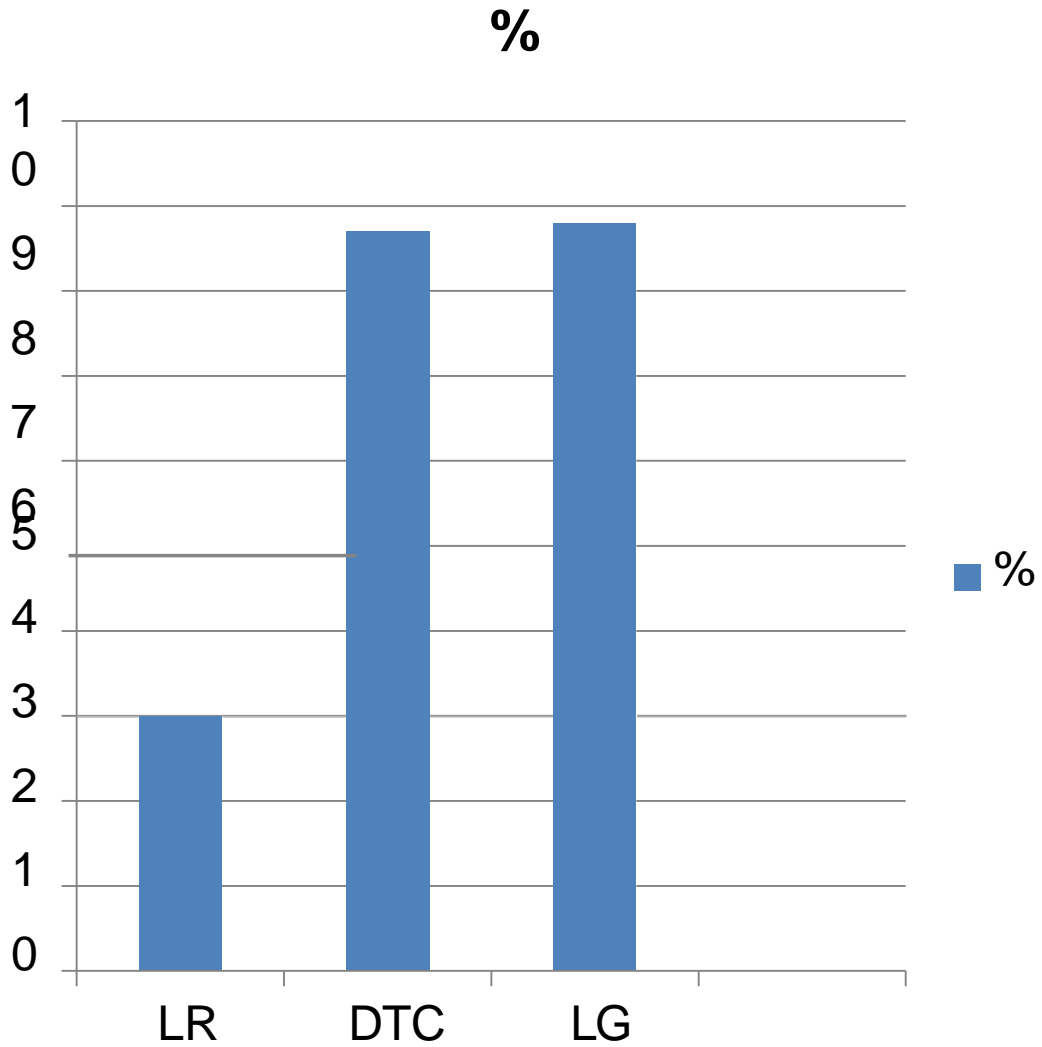


Figure 4.3: Proposed method of applied algorithms

3.6 Proposed Algorithm

In our declare method we utilize Logistic Regression algorithm . We try to create model which estimate Survival Analysis of heart failure patient .The execution of logistic regression our proposed technique to improve . Implementation is the main to thing of a system . Jupyter Notebook is an environment that is user for making our model.

Many Kind Library are used to make our model. In our method -

Step1. Our dataset is selected which contain Heart Failure ClinicalRecords.

Step2.Classification of dataset who survive and who can't af.

Step3: Input the dataset.

Step4: Using ML algorithm in python.

Step5: Highest accuracy is found from different machine learning algorithm.

Step6: Get highest accuracy using Logistic Regression Algorithm.

Step7: Calculate the performance of the model.

LG takes the dataset and classify who can survive who can't

At first we used our LG Algorithm when data is not processed. That time I got 75%

Accuracy .After that I processed data . I used Normalization to make our all the range 0 to 1 . We checked null and empty data from dataset . But Luckily I don't get any. Then I Try to solve target data imbalance using PCA(Principal component analysis) . Then I using our algorithm again . We used Linear Regression get 30% accuracy . After that we used Decision Tree Classifier

And get a good accuracy result 87% accuracy . Then I used Logistic Regression and get 88% accuracy .

Chapter 4

Experimental Results and Discussion

1. Introduction

In this part, we explained the results and conducted the experiment. In this part, we'll examine and contrast different classifier performance and accuracy. Here is the data from 299 people that we used from kaggle.com to predict a heart patient's survival status. We gather some statistical data using different algorithms and generate several charts in our data collection based on data percentage. We evaluated three approaches to determine our system's rate in our data set. We'll display the results as a graph as well as tables.

2. Experimental Results

In our dataset there are 299 data which is survival analysis of heart failure patient afterward we are getting the accuracy 88%. We are getting high precision from our cause real data are used . We use normalization to make our data range between 0 to 1. After normalization , it makes our accuracy higher than before . We also use voting classifier which helps us to apply multiple algorithm in a single model . We also used SMOTE(synthetic minority over sampling) which helps us to find more important feature in our dataset. After implementing this we again fit our dataset into algorithm ,But we didn't any effective improvement to implement this technic. After doing all this thing our logistic regression algorithm give better result than anyother.

We additionally use four types of predictions which are rate of Ture Negative, Rate of True Positive ,Rate of False Positive and Rate of False Negative. Confusion matrix can explain of a classification problem on a data set . Confusion matrix show four terms in its table .

True Positive (TP): we estimated result as who can survive is actually could be survive.

True Negative (TN): we estimated result as who cant survive is actually couldn't survive.

False Positive (FP): we estimated who can survive , actually can't survive.

False Negative (FN): we estimated cant survive, but actually Survive..

Precision : Precision is defined as the number of true positives divided by the total number of positive predictions.

$$\text{Precision} = \frac{tp}{tp+fp} = \frac{45}{45+1} = \frac{45}{46} = .98$$

Recall: The recall is the proportion of relevant examples recovered compared to the total number of relevant occurrences. A high recall algorithm returns the majority of the relevant results.

$$\text{Recall} = \frac{t}{tp+fn} = \frac{45}{45+4} = \frac{45}{49} = .91$$

F-measure: The f-score is a test accuracy metric that considers both precision and recall. It's a symbiotic combination of accuracy and recall.

$$\text{F-score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * .98 * .91}{.98 + .91} = \frac{1.78}{1.89} = .94$$

Accuracy: The familiarity of the measured value to a known value is referred to as accuracy.

$$\text{Accuracy} = \frac{fp+tn}{tp+tn+fp+fn} = \frac{45}{45+10+1+4} = \frac{.55}{.59} = .92$$

True Positive Rate: The term "false positive rate" refers to the rate at which our suggested technique incorrectly predicts that a brain stroke is not occurring while it is. Calculate the false positive rate using the following formula:

$$\text{True Positive Rate} = \frac{TP}{TP+FP} = \frac{45}{10+1} = 4.09$$

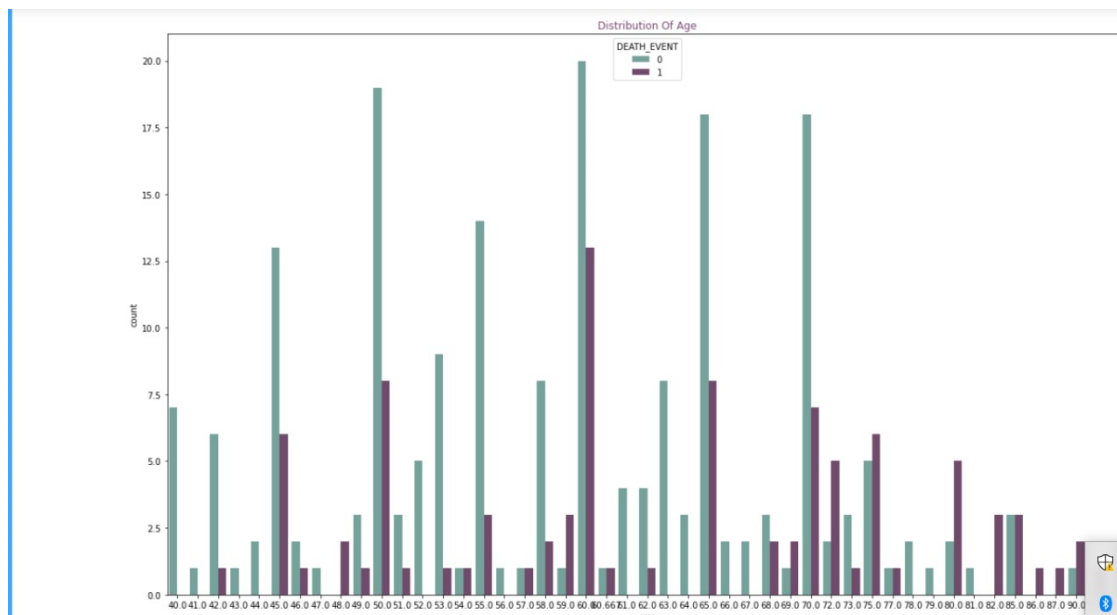
Specificity: The term "specificity" refers to the fact that our suggested technique correctly predicts a brain stroke when it is indeed a brain stroke [18]. Determine the specificity of the following equation:

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{10}{10+1} = .91$$

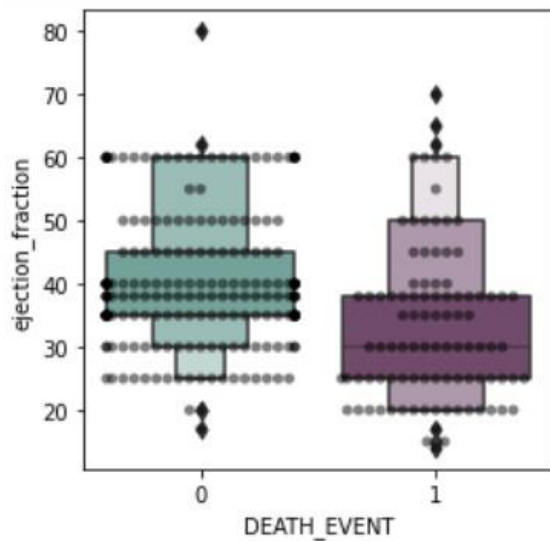
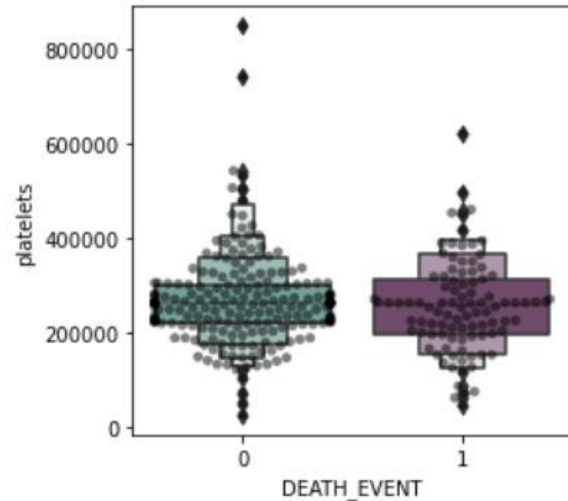
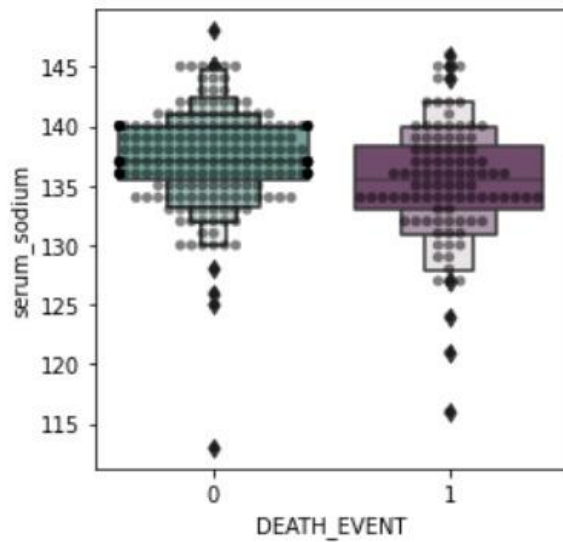
The confusion matrix known as, which may assist us in determining more advanced arrangement measures like as accuracy, recall, specificity, and sensitivity for our classifier. These are described in table 4.1

Experimental result

Precision	.98
Recall	.91
F-Score	.94
Accuracy	.92
True Positive Rate	4.09
Specificity	.91



From this barchat we can see the relation between age and death. Depends on age we can understand how many people will survive or not in that particular age.



From this swarm plot and boxplot we can see serum sodium, ejection and fraction feature effect on target attribute.

Those people who can survive more where their serum sodium rate below 135 . People can survive more when this is 135-140 . When platelets rate 20k-35k they survive more When ejection fraction under 30 , people survive more.

0	0.98	0.022
1	0.29	0.71
	0	1

Predicted label

Confusion matrix for our Random forest classifier

0	0.98	0.022
1	0.29	0.71
	0	1

Predicted label

Confusion matrix for our Support Vector classifier

0	0.85	0.15
1	0.36	0.64
	0	1

Predicted label

Confusion matrix for our Scession tree classifier

Table 4.1: Confusion matrix

	No-Event	Event
No-Event	True negative 10	False positive 1
Event	False negative 4	True positive 45

Here this table 4.1 we see that true negative and positive and false negative and positive in confusion matrix.

Our getting accuracy after implement all algorithm:

Algorithm	Accuracy
Random forest classifier	38%
Decision Tree	87%
Logistic Regression	88%
SVM	92%

Our final model accuracy is 92% where the algorithm is Support Vector Machine(SVM) .

CHAPTER 5

Conclusion and Futureworks

5.1 Summary of theStudy

We are described the Logistic regression approach to the obtained dataset in this study for prediction. We analyzed the estimation and improved the total prediction rate from other used algorithms using an 80% train set and 20% testset.

5.2 Conclusion

“survival status” might be frightening.. Its very glade that medical treatment and procedure are improved day by day , people who are meet with heart attack but can live productive life. It really important to find out which physical condition are responsible heart attack death. If general people can easily estimate their health condition it also helpful to reduce heart attack death ratio. We have discussed logistic regression algorithms in this study. We have estimate depends on some features of a patient which help a patient or some general people to see their body condition after heart attack or not. Using our project A doctor can check his patient condition easily. It also helps a doctor to get correct decision to during rush time. Our system help people take exact and correct decision for a heart disease patient. treatment. It helps people to use reduce death of people causing of heart attack.

Implications for Future Research

As we mentioned earlier we collected our data from kaggle.com. Our accuracy is 88% , we also will try to improve our accuracy rate. Our result is not perfect without giving all the features by user , we also work on that.

In Future will try to collect data from varies cardiac hospital. And try to find more structure and effective feature data. We also will try to make it easier to access for all kind of people

REFERENCE:

1. Chaurasia, V., & Pal, S. (2013). Early prediction of heart diseases using data mining techniques. *Caribbean Journal of Science and Technology*, 1, 208-217.
2. Rajkumar, A., & Reena, G. S. (2010). Diagnosis of heart disease using datamining algorithm. *Global journal of computer science and technology*, 10(10), 38-43.
3. Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), 43-48.
4. Wisaeng, K. (2014). Predict the diagnosis of heart disease using feature selection and k-nearest neighbor algorithm. *Applied Mathematical Sciences*, 8(83), 4103-4113.
5. Enriko, I. K. A., Suryanegara, M., & Gunawan, D. (2016). Heart disease prediction system using k-Nearest neighbor algorithm with simplified patient's health parameters. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(12), 59-65.
- 6." << <https://core.ac.uk/download/pdf/229279242.pdf> >> last accessed on : 23, October 2021 07:00pm
7. Donnan GA, Fisher M, Macleod M, Davis SM (May 2008). "Stroke". *Lancet*. 371 (9624):1612–23.
8. Rymarczyk, T., Kozłowski, E., Kłosowski, G., & Niderla, K. (2019). Logistic regression for machine learning in process tomography. *Sensors*, 19(15), 3400.
9. Loh, W. Y. (2011). Classification and regression trees. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(1), 14-23.
10. Provost, F. (2000, July). Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI'2000 workshop on imbalanced data sets* (Vol. 68, No. 2000, pp. 1-3). AAAI Press.
11. Kevin Zakka's Blog , << <https://kevinzakka.github.io/2016/07/13/k-nearest-neighbor/> >>. last accessed on: 23 , October 2021 11:00pm
11. PMC << https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4916348/?fbclid=IwAR2Zc0VMkyLeFKzYw_xHg-eA8ilj1Cfo_x5FFd0I34hbg7zGmzVNGWfmVM >> last accessed on: 23 , October 2021 08:00pm