Comparative Analysis of Heart Diseases Prediction using Machine Learning Algorithms

M. Sanjana[1], P.L.V.Sai Sudheer[2], M. Lakshmi Vaishnavi[3], S. Jahnavi[4], Dr. R. Vijaya Kumar Reddy[5]

[1-4] Students, Department of IT, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, A.P, India
[5]Assistant Professor, Department of IT, Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada, A.P, India

Abstract: In today's world cardiac arrest is seen commonly in many people. This attacks a person instantly so that there would be no sufficient time for diagnosing the patients. So, diagnosing the patients in timely basis has become the most challenging task. In this paper we can predict the occurrence of heart attack based on the given input details. It allows users to get an instant guidance on their heart disease through an intelligent system. We used machine learning algorithms (Logistic Regression, Decision Tree, Naïve Bayes, SVM) for database facilitating decision support system. In this paper various algorithms are used to extract the hidden knowledge from heart disease historical dataset. We used the medical profiles such as age, gender, chest pain, and some more profiles to predict the likelihood of patients getting a heart disease. Based on the predicted result, one can consult the doctor for the further medication.

KEYWORDS: Machine Learning, Heart Attack Disease Prediction, Support Vector Machine (SVM)

DOI of the Article: https://doi.org/10.46501/IJMTST0706038

Available online at: http://www.ijmtst.com/vol7issue06.html

As per UGC guidelines an electronic bar code is provided to seure your paper

To Cite this Article: M. Sanjana; P.L.V.Sai Sudheer; M. Lakshmi Vaishnavi; S. Jahnavi and Dr. R. Vijaya Kumar Reddy. Comparative Analysis of Heart Diseases Prediction using Machine Learning Algorithms. *International Journal for Modern Trends in Science and Technology* 2021, *7*, 0706175, pp. 218-224. https://doi.org/10.46501/IJMTST0706038

Article Info. Received: 16 May 2021; Accepted: 11 June 2021; Published: 17 June 2021

INTRODUCTION

'Health is Wealth' is a World-famous proverb concerning health. A healthy body is defined as the overall ability of the body to function well. Heart is an important part of the human body. Efficient working of heart plays a major role. If the working of heart is not efficient then it leads to affect the other parts of the body as well. Heart functions as an operating system of a human body. Now-a-days, Heart diseases are the major reason for human deaths. According to the recent survey, one in three adults worldwide has raised blood pressure which causes around half of the deaths from strokes and heart diseases. Heart disease is also known as cardiovascular disease (CVD). Heart diseases mean not just heart attacks but also includes the functional problems of the heart such as heart-valve abnormalities or irregular heart rhythms. All this different kind of problems also leads to heart failure.

Heart disease prediction is a very difficult task in the medical industry. In the medical industry there will be the huge amount of patient's data, disease diagnosis, electronic patient's records and medical devices. These patient's data is the key source for the extraction of the hidden knowledge. So, all these existing practice takes more time and there will be no prior knowledge to the patient. But coming to our work, it will predict the heart disease by using different machine learning algorithms. As the disease is predicted early there will prior knowledge about the patient condition and have the good treatment. Here, early detection is the main key. In the present days, there is no particular application for prediction of the heart disease. The only way for the patients is to visit the hospitals for their cardinal issues. In this system, clinical decisions are made by the doctor and it also depends upon their experience. This may lead to many unwanted bias, errors, excessive medical costs which affects the quality of service provided to the patient's and also there can be many ways that the diagnosis may go wrong. According to the research 42% of medical patient's feel they have had experienced a medical error or missed diagnosis. Whatever the approach we use for diagnosis, the main priority must be the efficient results which is found not be accurate in the existing system. We can put an end to medical misdiagnosis by knowledge rich data hidden in the database rather than doctor's intuition.

Here, we develop an intelligent system to predict the heart diseases in an accurate way without any misdiagnosis by using machine learning algorithms. Here, we trained the dataset with already available Cleveland data. If the patient wants to predict their heart disease, they have to register in the application by providing the required details. Once the user is successfully registered, there will be a medical form which has to be filled with appropriate details to predict their heart disease. This form consists of 13 different medical profiles which helps us to predict the chance of getting a heart disease. Here we trained our model using different machine learning algorithms to obtain the most accurate outcomes. Many hospitals have vast amount of data about the patient details but it is not accessed by them. In this proposed system, we utilize the data from the hospitals to train the models for the prediction of heart disease in a more précised and accurate way. This paper aims to use machine learning algorithms namely Logistic Regression, Decision Tree, Naïve Bayes, Support Vector Machine (SVM) to predict the heart disease.

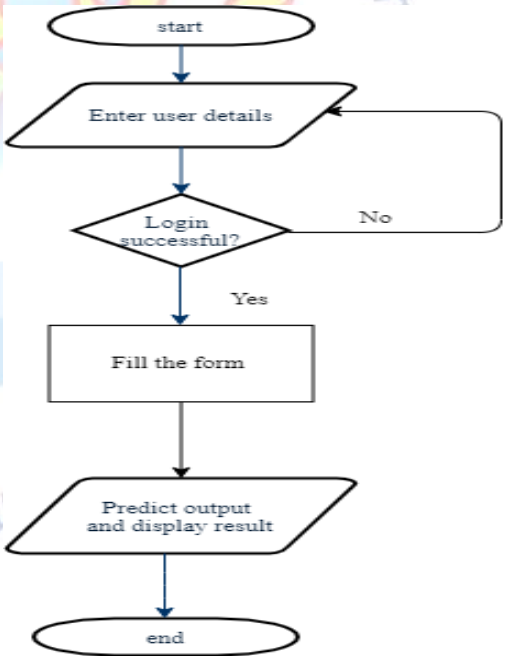


Figure 1: Workflow of Heart Attack Prediction.

METHODOLOGY

In today's tech revolution machine learning is seen in the prior list of technologies. As it made a benchmark in the industry. Usage of machine learning is also

becoming more based on the requirements of adding intelligence to an application. Unlike humans the decision-making process for a machine (computer) is enormous. Machine learning is adding the flavor to this decision-making process by its formative approach. Making a machine more intelligent and training it according to act like humans thinking in terms of prediction by using effective algorithms. To present these ideas practically there are many ways, but the major recognized categories are classified into three parts in machine learning.

Supervised Learning

The most commonly used path in machine learning is supervised learning, in general it is used to predict the next values in other terms it is popularly known as 'task driven mechanism'. It is popular because it is an easiest approach to understand and less complex to implement.

In this, we provide a dataset for training and predicting it based upon the given data, the algorithm predicts and comes out with an output every time . We need to train it back by giving the feedback as an input to get the desirable predicted output in overtime.

When this process comes to an end after many iterations happened, in final the supervised learning algorithm will able to predict efficiently than before and gives the most accurate predicted outputs.

Unsupervised Learning

The learning mechanism of an unsupervised learning is quite opposite to supervised learning Based upon the data and its properties, it is well known as a 'data-driven mechanism'. It is an intelligent algorithm because if a data is fed to the algorithm which is not in a featured way and it lacks in data labels, but the algorithm works with a predefined tool which adopts the feed from the given algorithm. It is called an intelligent learning process because it understands the properties of the data and organize them in a patterned way. Due to its intelligent learning capabilities in the machine learning it's also became the most adoptable for productivity.

Reinforcement Learning

The Reinforcement learning is built upon the term 'Mark of mistakes', in simple words reinforcement learning system is nothing but it can learn over a period of time from its own mistakes. It is not a biased mechanism of supervised and unsupervised learning.To understand reinforcement learning briefly the algorithms work in a flow of positive and negative behavioral aspects which means an algorithm makes lot of mistakes in the initial stages, we can reinforce the algorithm to associate with both positive and negative behavior methods by providing the suggestion dialogs to the algorithms with the help of this cyclic process the algorithm prone to less mistakes than at the beginning state.

Logistic Regression

It is one of the most implemented ML algorithms under supervised learning category. It is used to predict the various probabilities of a required variable or a target variable. The characteristics of a target would have only two possible classes.

In general words, logistic regression is binary in nature its data is coded as 1 or 0 for which '1' stands as a Success and '0' stands as a failure. Mathematically this model predicts Predict(Y=1) as a function of X. Due to its simplistic nature of predicting a target variable

It is widely used in resolving the problems like Cancer detection , lungs failure detection etc.

The binary logistic regression having the dependent variables, the predication can be done by two more categories those are Binomial, Multinomial and ordinal. It also specifies some rules to get the desired predicted values as a truthful outcome like choosing a large sample sized data for training the model and also implementing a valid or meaningful variable

Decision Tree

It is a decision predictive modelling tool that can be implemented across different areas. It can be applied by a decision building approach which easily divides the dataset in different logical ways to predict on different situations or conditions that can be applied. It is also called as "one of the most powerful algorithms" that falls under the category of supervised learning models.

Decision tree algorithm is further classified into two sun-branches Classification tasks & Regression tasks. These entities of a decision tree are named as decision nodes in the modelling process, where the data is leaving and splitting into different ways based on conditions.

Implementing a decision tree algorithm by the Gini index, split creation.

Classification decision tree: The target variable is a categorical where splitting is based on the given decision variable non – regular

Regression decision tree: The target variable is a continuous where the decision is Exercised regularly

Naive Bayes (NB):

It is well known as a classification technique which is dependent on the bayes theorem a strong assumption for this theorem that all the predictors or the algorithms are independent to each other.

The most popular Bayesian classification, is to find the posterior probabilities. Which is stated that a probability of a label given with some observed nature, with the help of Naïve bayes algorithm we can show this feature in a quantitative structure

Predict(L | f) = predict(L)(f | L) / Predict(f)

Building a model using naïve bayes in python is became more flexible by importing scikit learn library

The classic applications offered by the naïve bayes are Real-time prediction, Multi-class prediction, text Classification, Recommendation System related to some of the complex structures with the reinforcement platform

Support Vector Machine (SVM)

It is also one of the famous supervised learning algorithms, it is also used for the classification as well as regression problems. In most of the cases SVM is used as a classification problem solver in machine learning.

The formative of the SVM is to present the best line or decision boundary which is segregated into n-dimensional space into classes. So, the accuracy is time lined we can put the new data point in the correct sector in the future. And this decision boundary is called a 'hyperplane'

This algorithm chooses the extreme vectors/data points in order to create a hyperplane. These classes are known as support vectors, Support vector machine is readily used for face detection, image classification, text categorization

The modeling of this algorithm is classified into two types: Linear and non-linear SVM

1. Linear SVM is used for linearly separable data, if a dataset can be determined into two classes by using a straight line, then it is coined as a linearly separable data (Linear SVM classifier)

2. Non-linear separable data , which states if a dataset cannot be classified by using a straight line, such data is termed as non-linear data (Non-linear SVM classifier)

ARCHITECTURE

The below figure is the architecture of the heart attack prediction system. It is the blueprint for developing the paper. Here we can see the flow of various components that are used in the development of the paper.

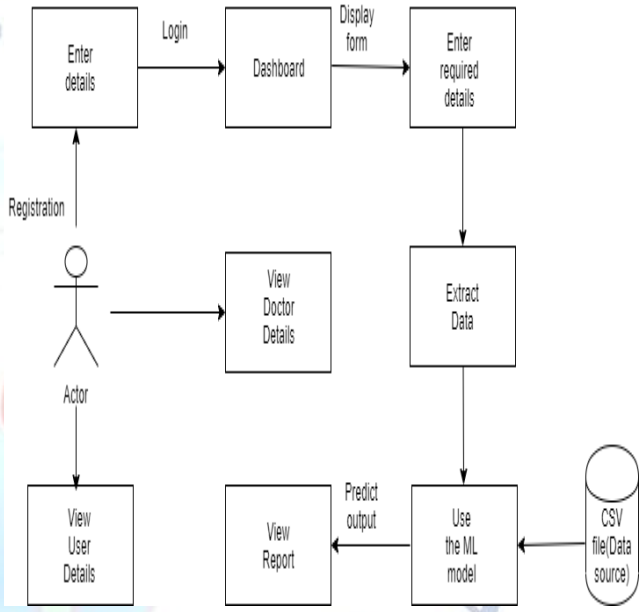


Figure 2: Architecture of Heart disease prediction system.

Classification Algorithms

Enter details: The user should register himself/herself by entering the details for heart disease prediction.

Dashboard: After creation of an account, the user needs to login himself/herself with their respective login credentials so that it will directly redirect to the dashboard.

Enter required details: Once the user has successfully logged in, the user should enter his/her heart health details.

Extract data: The hidden data will be extracted from the details provided by the user.

Use the ML model: By using ML model, the hidden data will be extracted.

View report: By using the ML model, it predicts the output of heart disease and the user can view his/her prediction which states the possibility of heart attack.

CSV file (data source): The details entered by the user is stored in this CSV file.

View user details: The user can view, edit his/her profile anytime and anywhere by logging-in to the application.

The algorithms that used in the paper are Logistic Regression, Decision Tree, Naïve Bayes, Support Vector Machine (SVM).

Implementation of Algorithms

Input: Details given in the form.

Output: Predicting the class label.

Steps for Implementation of algorithm:

Initialize the classifier to be used.

Train the classifier: All classifiers in scikit-learn uses a fit(X, y) method to fit the model (training) for the given train data X and train label y.

- Predict the target: Given a non-label observation X, predict (X) returns the predicted label y.
- Evaluate the classifier model.

EXPERIMENTAL RESULT

The following steps describes the working functionality of the application. They are:

1. First we need a system with python software, Jupyter notebook, and anaconda, django installations in it.

2. We need a dataset with different attributes related to the heart diseases in it.

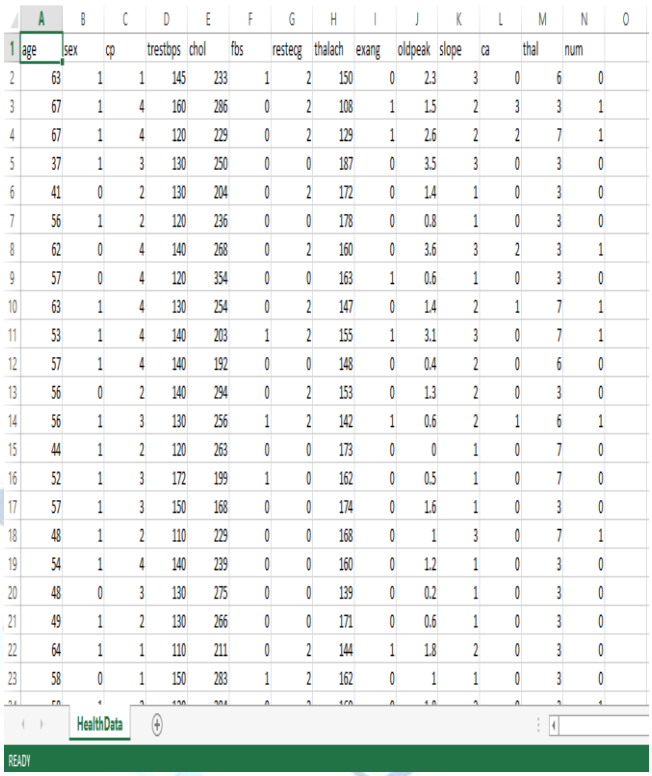
3. We should train and test the dataset with machine learning algorithms like Logistic Regression, Decision Tree, Naïve Bayes and SVM and save them with logistic_regression.py, decision_tree.py, naïve_bayes.py and scv.py.

4. After that we need to apply the migrations and run the server in the anaconda prompt and open the application.

5. If he/she is a valid user then they will be directed to their profile from the login page.

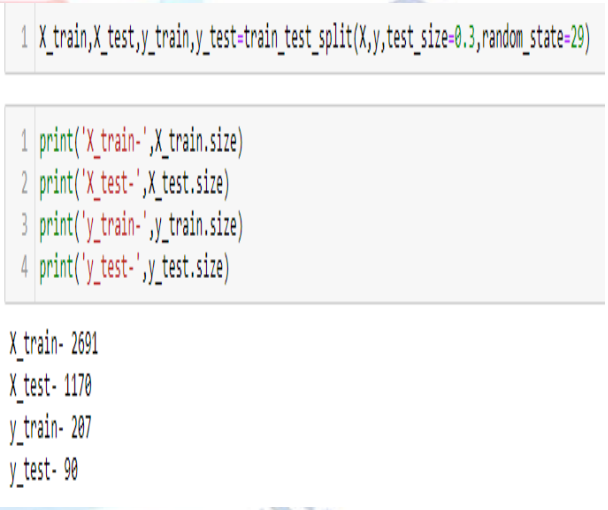
6. Thereafter, user can fill the medical information form and click on the predict button to view their result of heart disease.

7. Finally the user can view the prediction of heart disease.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
2	63	1	1	145	233	1	2	150	0	2.3	3	0	6	0
3	67	1	4	160	286	0	2	108	1	1.5	2	3	3	1
4	67	1	4	120	229	0	2	129	1	2.6	2	2	7	1
5	37	1	3	130	250	0	0	187	0	3.5	3	0	3	0
6	41	0	2	130	204	0	2	172	0	1.4	1	0	3	0
7	56	1	2	120	236	0	0	178	0	0.8	1	0	3	0
8	62	0	4	140	268	0	2	160	0	3.6	3	2	3	1
9	57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
10	63	1	4	130	254	0	2	147	0	1.4	2	1	7	1
11	53	1	4	140	203	1	2	155	1	3.1	3	0	7	1
12	57	1	4	140	192	0	0	148	0	0.4	2	0	6	0
13	56	0	2	140	294	0	2	153	0	1.3	2	0	3	0
14	56	1	3	130	256	1	2	142	1	0.6	2	1	6	1
15	44	1	2	120	263	0	0	173	0	0	1	0	7	0
16	52	1	3	172	199	1	0	162	0	0.5	1	0	7	0
17	57	1	3	150	168	0	0	174	0	1.6	1	0	3	0
18	48	1	2	110	229	0	0	168	0	1	3	0	7	1
19	54	1	4	140	239	0	0	160	0	1.2	1	0	3	0
20	48	0	3	130	275	0	0	139	0	0.2	1	0	3	0
21	49	1	2	130	266	0	0	171	0	0.6	1	0	3	0
22	64	1	1	110	211	0	2	144	1	1.8	2	0	3	0
23	58	0	1	150	283	1	2	162	0	1	1	0	3	0

Figure 3: Prediction dataset (HealthData.csv)

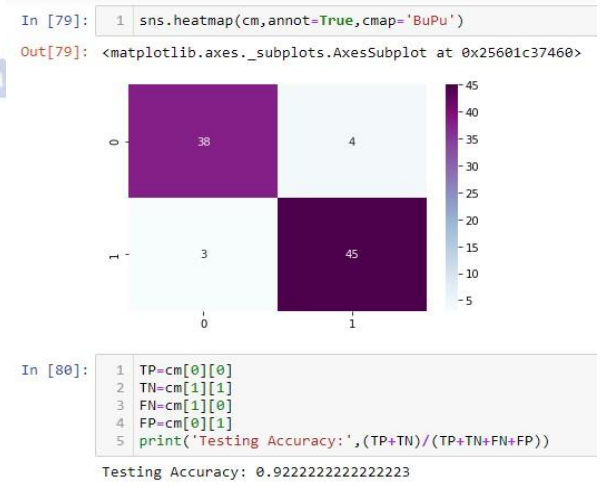


```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3,random_state=29)
```

```
print('X_train-',X_train.size)
print('X_test-',X_test.size)
print('y_train-',y_train.size)
print('y_test-',y_test.size)
```

```
X_train- 2691
X_test- 1170
y_train- 207
y_test- 90
```

Figure 4: Training and Testing size of dataset.



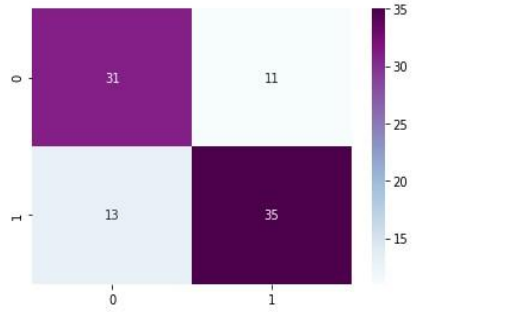
```
In [79]: sns.heatmap(cm,annot=True,cmap='BuPu')
Out[79]: <matplotlib.axes._subplots.AxesSubplot at 0x25601c37460>
```



```
In [80]: TP=cm[0][0]
         TN=cm[1][1]
         FN=cm[1][0]
         FP=cm[0][1]
         print('Testing Accuracy:',(TP+TN)/(TP+TN+FN+FP))
Testing Accuracy: 0.9222222222222223
```

Figure 5: Testing Accuracy for Logistic Regression

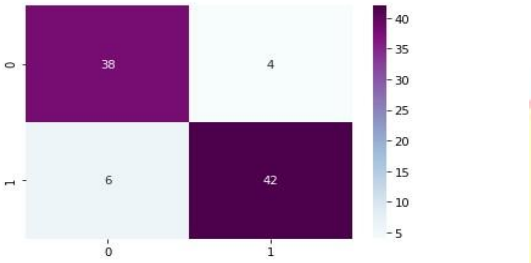

```
In [85]: 1 sns.heatmap(cm,annot=True,cmap='BuPu')
Out[85]: <matplotlib.axes._subplots.AxesSubplot at 0x256037fc640>
```




```
In [86]: 1 TP=cm[0][0]
         2 TN=cm[1][1]
         3 FN=cm[1][0]
         4 FP=cm[0][1]
         5 print('Testing Accuracy:',(TP+TN)/(TP+TN+FN+FP))
Testing Accuracy: 0.7333333333333333
```

Figure 6: Testing Accuracy for Decision Tree

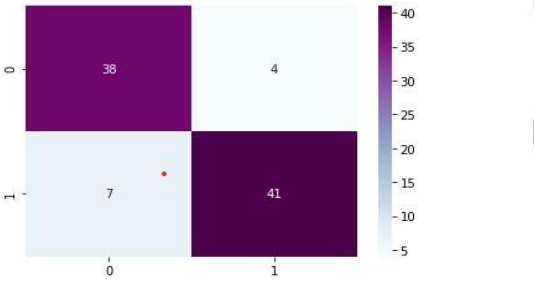
```
In [91]: 1 sns.heatmap(cm,annot=True,cmap='BuPu')
Out[91]: <matplotlib.axes._subplots.AxesSubplot at 0x256040d6280>
```




```
In [92]: 1 TP=cm[0][0]
         2 TN=cm[1][1]
         3 FN=cm[1][0]
         4 FP=cm[0][1]
         5 print('Testing Accuracy:',(TP+TN)/(TP+TN+FN+FP))
Testing Accuracy: 0.8888888888888888
```

Figure 7: Testing Accuracy for Naïve Bayes

```
In [97]: 1 sns.heatmap(cm,annot=True,cmap='BuPu')
Out[97]: <matplotlib.axes._subplots.AxesSubplot at 0x25601ba2430>
```




```
In [98]: 1 TP=cm[0][0]
         2 TN=cm[1][1]
         3 FN=cm[1][0]
         4 FP=cm[0][1]
         5 print('Testing Accuracy:',(TP+TN)/(TP+TN+FN+FP))
Testing Accuracy: 0.8777777777777778
```

Figure 8: Testing Accuracy for SVM

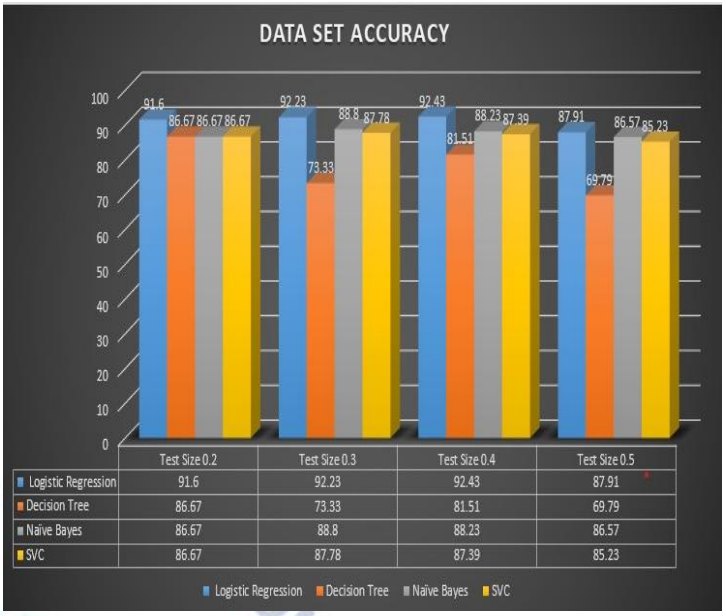


Figure 9: Accuracy of Machine learning algorithms.

CONCLUSION

The research undertook an experiment on application of various Machine learning algorithms to predict the heart attacks and to compare the best method of prediction. The research results do present a dramatic difference in the prediction when using different classification algorithms in ML. The experiment can serve as an important tool for physicians to predict risky cases in the practice and advise accordingly. The model from the classification will be able to answer more complex queries in the prediction of heart attack diseases. The predictive accuracy determined by regression, Naïve-bayes, SVM algorithms suggests that parameters used are reliable indicators to predict the presence of heart diseases. This paper mainly focuses on the progressive and Accuracy build algorithms for heart disease prediction, the dataset provides much better accuracy than an untouched one with missing values. So, the selection of profound and effective techniques for data polishing along with standard and précised machine learning algorithms will helps to the development of prediction systems that produces the efficient and stable Results.

REFERENCES

1. Andrea D'Souza, "Heart Disease Prediction" , IJRES, Vol 3, No 3, April-2015.
2. Dhanashree S. Medhekar, Mayur P. Bote, Shruti D. Deshmukh, "Heart Disease Prediction System using Naïve Bayes", IJERSTE, Vol 2, No 3, March-2013.

3. Jyoti Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal Computer Applications, Vol17, No 8, March-2011.
4. G. Purusothaman and P. Krishnakumari, "A Survey on Risk Prediction: Heart Disease", Indian Journal of Science, Vol 8, No 12, 2015.
5. https://researchdesignlab.com/smart-health-monitoring-system.html
6. https://pdfs.semanticscholar.org/8610/4041dd610f78b6a5718e7ebe3cf6affb6f28.pdf
7. https://www.ijrte.org/wpcontent/uploads/papers/v8i6/F9241038620.pdf
8. https://pdfs.semanticscholar.org/caf8/b4d1b1cabe259aa84d12e44c6cdbdf908c81.pdf
9. https://www.ijana.in/Special%20Issue/TPID22.pdf
10. https://towardsdatascience.com/what-are-the-types-of-machine-learning e2b9e5d1756f
11. https://www.ripublication.com/acst17/acstv10n7_13.pdf
12. ShantakumarB.Patil,Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research, ISSN 1450-216X, Vol 31, No 4, 2009.
13. Syed Immamul Ansarullah, "Heart Disease Prediction System", IRJET.
14. S.B.Bhalerao1, Dr. B.L.Gunjal2, "Survey Of Heart Disease Prediction Based On Algorithms", IJARIIE vol-2 Issue-02, 2016.