IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.7, July 2022 E-ISSN: 2347-2693

Comparison and Evaluation of Various Machine Learning Algorithms on Heart Disease Data Set

M. Prameela[1*], M. Kamala Kumari[2]

[1,2]Department of Computer Science and Engineering, Adikavi Nannaya University, Andhra Pradesh, India

[*]*Corresponding Author: prameela.mtech@gmail.com, Ph No.: +918712248122*

DOI: https://doi.org/10.26438/ijcse/v10i7.111 | Available online at: www.ijcseonline.org

Received: 21/Jun/2022, Accepted: 05/Jul/2022, Published: 31/Jul/2022

Abstract— Machine Learning is now one of the thrust areas where computers are trained automatically learn from the given data automatically without any human intervention. It is the study of making machine learn automatically and do the things through algorithms which humans are doing without being explicitly programmed. Decision making is a major problem that effects the entire system under consideration irrespective of commercial databases, transactional databases, e-commerce data, social networking data or any other of that kind. Predicting the future and taking a right decision at right time is a big challenge. Supervised machine learning algorithms are solutions to those kinds of problems that are faced. They have a wide range of applications. Due to the lack of well-defined principles, choosing a suitable ML algorithm for a given problem and data is a big challenge. In this paper it is intended to do a quick and brief review of famous machine learning classification algorithms, their advantages and disadvantages, their area of application and suitable algorithm suggestion for particular type of problems. In this paper evaluation is done on supervised machine learning algorithms. Based on evaluation comparison of supervised algorithms is done.

Keywords—Supervised learning, classification, regression, Naïve Bayes theorem, SVM, Linear Regression, Decision Trees, coronary artery disease (CAD)

I. INTRODUCTION

Machine Learning is a subset of Artificial Intelligence. It is a field of study, which takes the existing data, develops a model, train and test the model and predict the outcome for new data. The ability to acquire its own knowledge by extracting patterns from raw data and infer something from the acquired knowledge is known as Machine Learning. In general machine learning is predicting the future based on the learning made in the past in terms of experience, knowledge or existing data.

Fig1. Shows the relationship between Data Science, Artificial Intelligence, Machine Learning, Deep learning, Data mining, Data Analytics and Data Analysis.

A: Relation between different fields of Data Science

Data Science: It is a broad area which comprises of the artificial intelligence, different domains, analytics, statistics and programming skills which gives a deep and meaningful insights and solutions to the questions in long run which are unseen from raw and unstructured data.

Data Analysis: Data analysis comprises of the complete investigation of the data beginning from gathering till the report generation. Data Analysis has different phases which are also part of data science, and is a complete process to retrieve the knowledge. Analysis gives the answers to what has to be done and analytics gives the justification for technical details of how it has to be done.

Data Analytics: It is a subfield of Data Analysis which comprises of statistics. Data Analytics mainly focuses on the technical components of the process that is decided to adapt. Data Analytics has the solutions for how a task can be completed. Meaningful correlations are obtained by applying algorithms on the data which is analyzed.

Data Mining: Data mining is the *process* of discovering interesting patterns and knowledge from *large* amounts of data for analysis. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically [1]. Data which is gathered for mining is clean, integrated and consistent data. Data Mining is a multifunctionality process. Data can be described, predicted and interesting patterns can be mined. Data mining a subfield of machine learning which comprises of many algorithms based on the functionality required. All algorithms use statistical measures, correlations and inferential phenomenon for predictions.

Looks like data mining is same as data analytics as both the concepts focus on the technical aspects to retrieve information. This is a process of inferring meaningful knowledge from the vast data through classification, clustering concepts, descriptions and outlier detection.

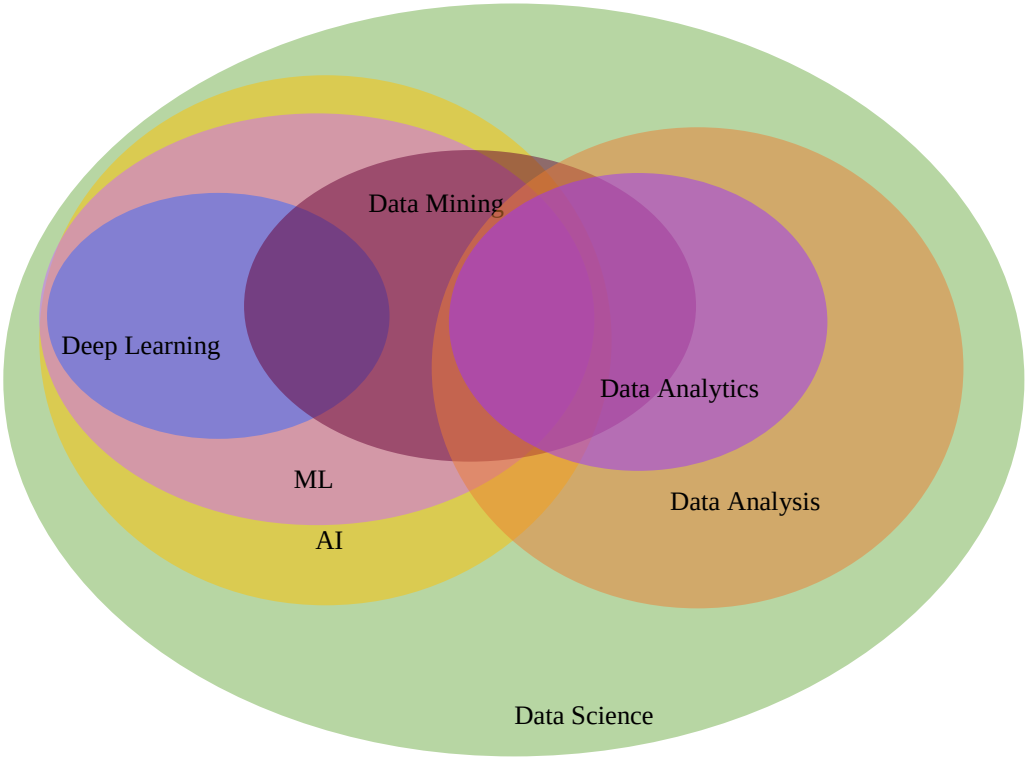

Fig 1: Relationship between Different Concepts

Artificial Intelligence (AI): As defined by Gartner "Artificial intelligence (AI) applies advanced analysis and logic-based techniques, including machine learning, to interpret events, support and automate decisions, and take actions". Artificial intelligence is the field of computer science which is used to build smart machines which performs the jobs and tasks as humans do ie., to think and act as humans. It is making machines which works and thinks as Human brains and do the decision making and justification for any job to be done. Along with Machine Learning Algorithms heuristic algorithms are used in AI to find a solution to a problem.

Machine learning (ML): It is a multi/inter disciplinary field comprising of many fields. Machine learning is subfield of Artificial Intelligence which in turn is a subfield of Computer Science. Over a period of time machine learning has been incorporating multiple fields like statistics, mathematics which makes it multidisciplinary field. The base for any Mining algorithm or any Machine learning algorithms is Statistics and Mathematical Proofs and Heuristics. The New Technologies or concepts like Deep Learning and Machine Learning are all inter-related to one another.

Deep Learning: Deep learning a subfield of Machine Learning which works on the concept of neural network which is inspired by the human brain network. These neural networks have hidden layers through which data is processed and the system learns by itself for getting the desired output

B Need of Machine Learning

Decision making is a complex task now-a-days and it becomes worse when the size of the data is huge and type of the data is different. Machines which act like humans should take dynamic decision as per the environment they live in. Heuristic searches and dynamic adaption of data plays key role in machine Learning. The need of machine Learning will make the machines or objects to take dynamic decisions, predict the future values, recognize the complex identities, classify the data etc.,

The data which is produced today is more of unstructured data, like data from Facebook, twitter, news, etc.

1. Size of data: The data which is produced every day is huge and enormous in size. Companies are loaded with enormous volumes and different types of data which need to be processed. For an individual to analyze this data is a tedious task. Machine Learning models are developed to process massive data, derive conclusions and predict the output from the data.
2. Ability to adapt to changes made: Unlike pre-programmed, or embedded devices, machine learning algorithms adapts to the changes made in the structure, input data or output data.
3. Hidden data: when dealing with large data humans may not identify the important relationships and correlations hidden between them. Using Machine learning algorithms those relationships can be extracted.
4. Machine learning algorithms are best suited for complex problems where there's no solution in traditional approach.

Domains where machine learning can be applied: One cannot think of today's world without machine learning. Some of the areas where machine learning is applied, but not limited to are

- Image Recognition
- Medical and Health Care
- Computer vision
- Natural language processing
- Weather forecasting
- Disaster management
- Cricket

C Motivation

According to WHO 17.9 million people die each year from cardio vascular disease which is estimated to 32% deaths worldwide. So, taking heart disease dataset, we compared different machine learning algorithms. In this paper decision tree, Naïve Bayes, Logistic Regression and Support Vector Machines (SVM) algorithms were evaluated on Heart disease dataset and their accuracies are compared.

The rest of the paper is organized as follows. Section II contains the related study and the literature review done for the machine learning algorithms and heart disease related papers. Section III contains the Related work which discusses the Logistic Regression, Naïve Bayes, Decision trees and SVM. Section IV contains methodology of the work, section V contains the Results and analysis and Section VI contains the Conclusion

II. RELATED STUDIES

A Machine Learning Classification

Fig 2 depicts the family tree for classification of the machine learning algorithms

Supervised Learning: In Supervised learning we have input variables called features and an output variable called class label or target. Supervised machine learning algorithms learn to map the features to class labels. In training phase desired class variable is fed along with input variables to the machine learning algorithm. The main goal of a supervised learning model is to predict the correct class label when new input data is given.

Supervised learning is divided into 2 categories: Classification and Regression.

Classification: As the name signifies, is the process of categorizing something. Here also in machine learning, classification is the process of predicting and categorizing the class label from the given features. Classification models, predict categorical (discrete, unordered) class labels. Example of classification models are spam filtering, image classification etc. This can be represented using a simple equation

$$y = f(x)$$

where y is dependent or class label. x is the feature and $f(x)$ is the classification algorithm.

Regression: Regression models predict continuous class labels from the features given. Examples of Regression models are prediction of sales, marks of an exam, the height of a person etc. This can be represented using a simple continuous equation

$$z = v_0 + v_1x_1 + v_2x_2 + \ldots + v_nx_n$$

in the above equation z generates continuous values.

Unsupervised Learning: In unsupervised learning there are only input variables called features and no corresponding output variable hence the dataset is unlabeled. Unsupervised machine learning algorithms groups the data and forms clusters of similar data based on certain patterns. The model learns itself through training from the data and forms clusters.

Reinforcement Learning: Definition of Reinforcement by Oxford dictionary "the process of encouraging or establishing a belief or pattern of behaviour". In the same way in Machine learning also, Reinforcement learning is learning a pattern of behaviour by taking right decisions and coming to a conclusion. It is the training of machine learning models in a complex situation to take a sequence of decisions on its own, wherein a reward is awarded for such a decision taken. The reward is a positive one when a right decision is taken, and a negative reward which can be considered as a penalty when a wrong decision is taken. The goal of reinforcement learning is to maximize the total reward obtained by summing all the rewards. In Reinforcement learning the learner must explicitly explore the environment in order to highlight the problem of exploration [2]

Semi-Supervised Learning: Semi supervised learning is a combination of Supervised and unsupervised learning. This algorithm performs well when the dataset has small amount of labeled data and huge amount of unlabeled data. In these cases, with the help of labeled data it is possible to derive new patterns or structures from unlabeled data.

Machine Language can be broadly classified as shown in Fig 2.

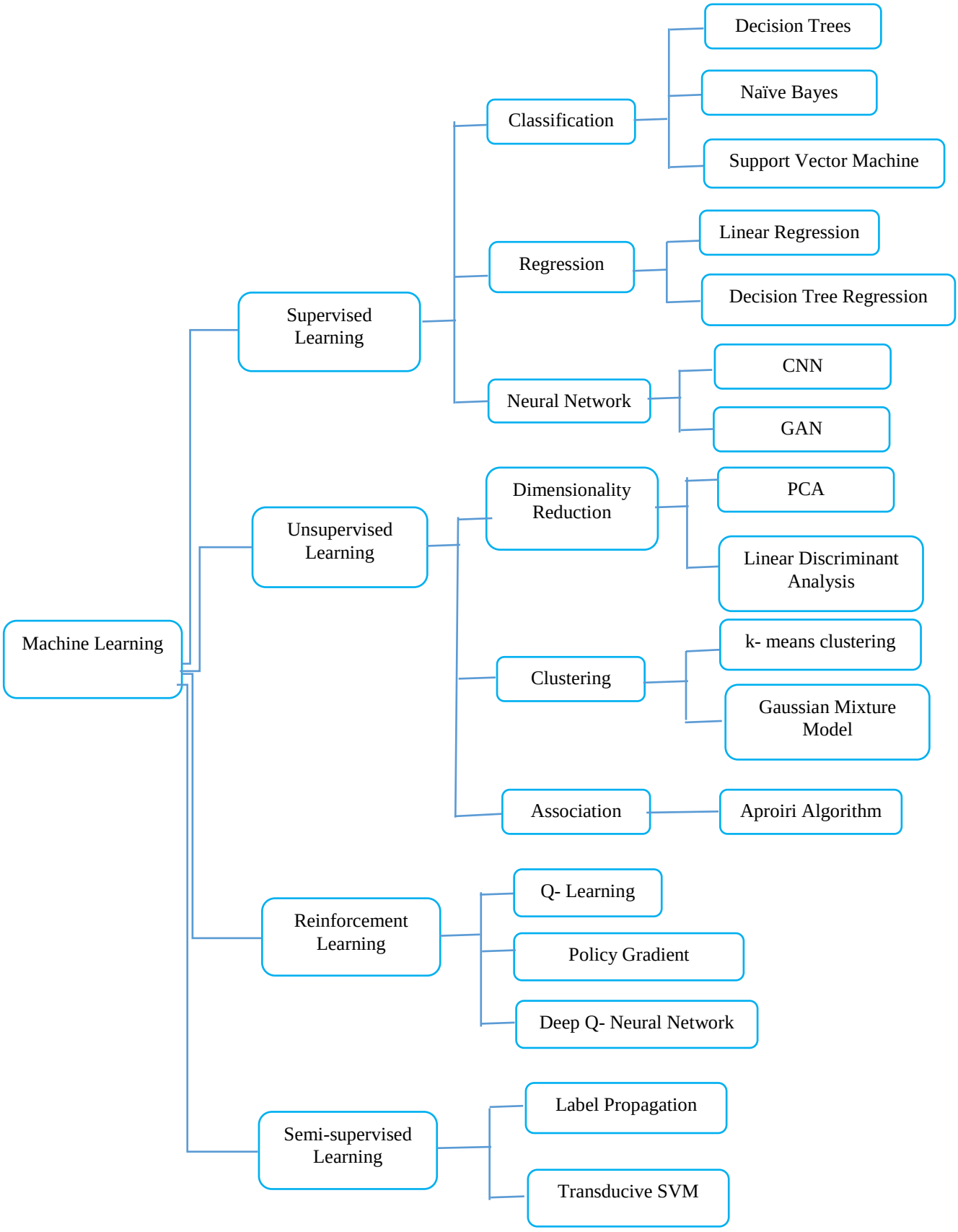

Fig 2 Classification of Machine Learning Algorithms

B Literature Survey

Machine Learning Algorithms

Table 1 Literature Review

Reference	Algorithm	Description
[1][3] [4][5]	Decision Trees	Decision Tree Algorithm, calculating Entropy and Information Gain
[6][7][8]	SVM	Derives an equation for optimal Hyperplane for Support vector machines. Applications of SVM are taken from [7]
[9][10][11]	Naïve Bayes	Shows the dependency distribution of the nodes with mathematical explanation of Naïve Bayes algorithm.
[12][13]	Logistic Regression	Logistic Regression basics and its performance with other algorithms

Heart Disease Literature Review

Mai Shouman et el [5] presented a model which integrates multiple voting classifier with different types of discretization methods and different Decision trees to increase accuracy on heart disease dataset.

Dissanayake, Kaushalya, Md Johar, Md Gapar (2021) [14] Examined different feature selection techniques like chi square, ANOVA on different machine learning algorithms like Decision Trees, SVM, K-nearest Algo on the heart disease data set.

J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor [15] proposed a novel feature selection model. Compared different machine learning algorithms like Logistic Regression, SVM, K-nearest, ANN with different parameters and measured the accuracies of all of them.

Bhuvan Sharma, Spinder Kaur [16] discusses the reasons for silent heart attack.

Jian Ping Li et el [17] proposed a new Feature Selection algorithm Fast Conditional Mutual Information(FCMI

III. RELATED WORK

Decision Trees: It is classification based predictive learning model which is represented in tree structure, where internal node is a test on an attribute, branch is an outcome of a test and leaf nodes are the class labels. Decision Trees are performed on categorical attributes. It is a greedy algorithm which forms a tree structure by splitting in a recursive partition manner of the features. Decision tree is a predictive model that can be viewed as a tree where each branch of the tree is a classification question and leaves represent the partition of the data set with their classification [18]. The branches are formed by selecting the best split from a set of possible splits which is decided by calculating the entropy and information gain. To identify the splitting attribute of the Decision Tree, one must calculate the Information Gain for each attribute and then select the attribute that maximizes the Information Gain [1]. Entropy is the measure of randomness or degree of uncertainty of the features which is denoted by the equation [3,14] is

$$E = -\sum_{j=1}^{m} q_j \log_2 q_j$$

q_j is the probability of the feature j.

Information gain is calculated in ID3 as

$$I = 1\text{-}E.$$

Where E is the entropy. Gini Index is used for CART which is measured as the probability of a specific feature that is classified incorrectly when selected randomly. It is the measure of impurity. Gini index defined by [1] as

$$G = 1\text{-}\sum_{j=1}^{m} q_j^2$$

q_j is the probability of feature j being classified.

An Example Decision Tree for eligible candidates for further tests to be taken for the job of Sub Inspector of police in Andhra Pradesh of India can be seen in Fig 3.

Support Vector Machines: SVM is an algorithm which provide a general approach for different inferential tasks such as classification, regression and clustering based on consistent mathematical theory [8]. SVM is a combination of Support Vector Classification (SVC) and Support Vector Regression (SVR).

The objective of the SVM is to find the optimal separating maximal margin hyperplane. We'll see how to find this.

A vector is an object which has both magnitude and direction.

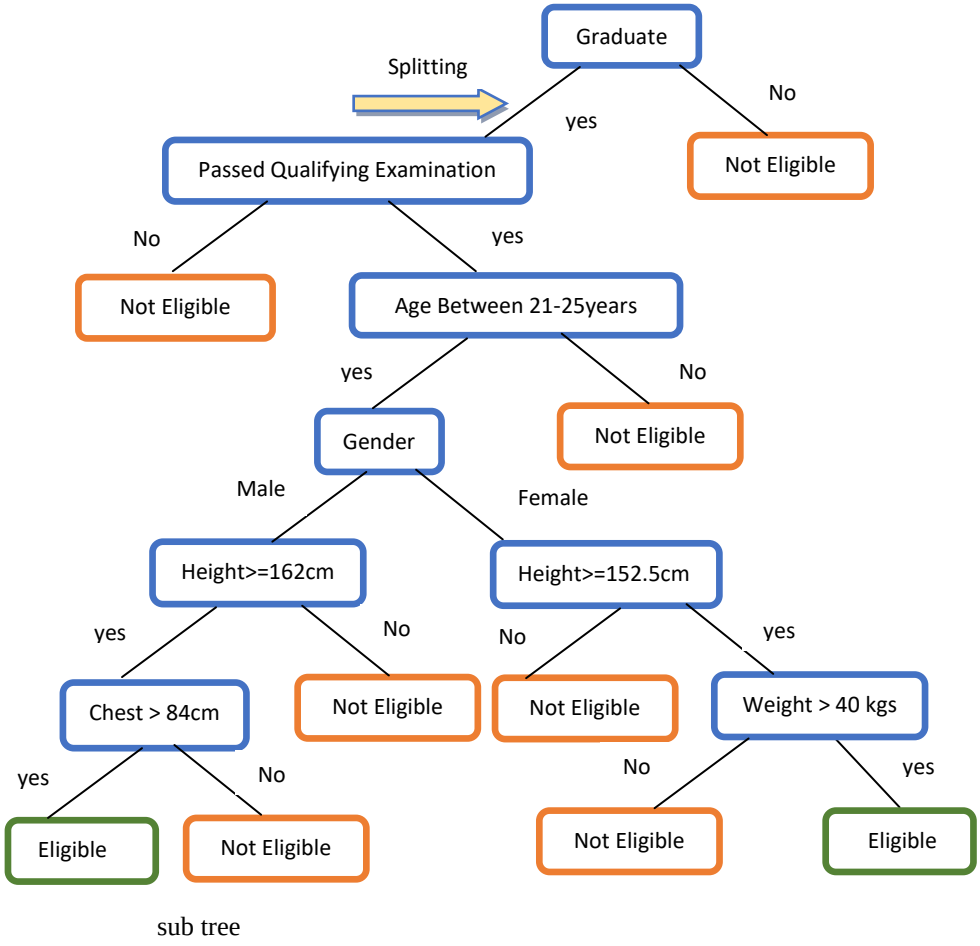

Figure3: Sample Decision Tree

The magnitude or length of a vector is x is written as ||x|| and is called its norm. For a 2D vector **x** (x_1, x_2)

$\|\mathbf{x}\| = \sqrt{(x_1^2+x_2^2)}$

The direction of the vector **x**(x_1,x_2) is the vector **w**($\frac{x_1}{||x||}$, $\frac{x_2}{||x||}$). The norm of **w** is equal to 1 which means a unit vector.

The magnitude of n-dimensional vector

$\|\mathbf{x}\| = \sqrt{(x_1^2+x_2^2+\ldots\ldots\ldots x_n^2)}$

A separating hyperplane can be written as

$$\mathbf{W}.\mathbf{X} + b = 0 \quad (1)$$

Where **W** is the weight vector **W**= $\{(w_1, w_2, \ldots\ldots w_n\}$ n is the number of attributes and b is bias.

Considering an example given below we try to derive an equation for SVM.

Let us consider a 2-D eg (**X**= (x_1, x_2)) where x_1, $and x_2$ are the values of attribute features A1, and A2 respectively for X as shown in table 2.

Table 2: Sample Example

Attribute A1	**Attribute A2**	**……**	**…….**	**Attribute n**
x_1	x_2	……	…….	…… x_n
y_1	y_2	……	……	…….. y_n
z_1	z_1	……	……	…… z_n

If we think b as an additional weight, and represented as w_0 Eq(1) for 2D can be written as

$w_0+w_1x_1+w_2x_2=0$ (2)

Let H_1, H_2 be 2 hyperplanes defining the sides of the margin

Eq(2) can be rewritten as

$H_1=w_0+w_1x_1+w_2x_2 >= 1$ for $y_i= 1$ (3)

$H_2=w_0+w_1x_1+w_2x_2 <= -1$ for $y_i= -1$ (4)

Any tuple falls on or above H_1 belongs to class +1 and any tuple falls on or below H_2 belongs to class -1.

Hyperplanes can be represented diagrammatically as below. Positive hyperplane in Fig 4, and negative hyperplane in Fig 5

The distance from the separating hyperplane to any point on H_1 is $\frac{1}{||w||}$ where ||W|| is Euclidean norm of W that is $\sqrt{W.W}^2$ The maximal margin is $\frac{2}{||w||}$.

After training the SVM, according to Cortes C, Vapnik(1992)[6] and Cortes C, Vapnik (1995) [7] to classify the tuples Based on Lagrange formulation, Maximum Marginal Hyperplane (MMH) can be rewritten for decision boundary as defined in

$$\boldsymbol{d(x^T)=\sum_{i=1}^{l} y_i \alpha_i x_i x^T + b_0} \quad (5)$$

y_i is the class label of SVM
α_i b_0 are the numeric constants that are determined automatically
x_i is support vector
x^T is test tuple
l is the number of support vectors
$d(x^T)$ is Maximal margin hyperplane which is nothing but decision boundary.

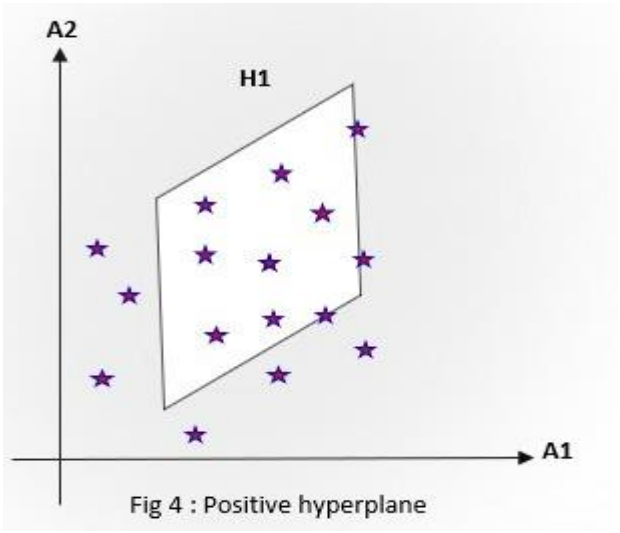

Fig 4 : Positive hyperplane

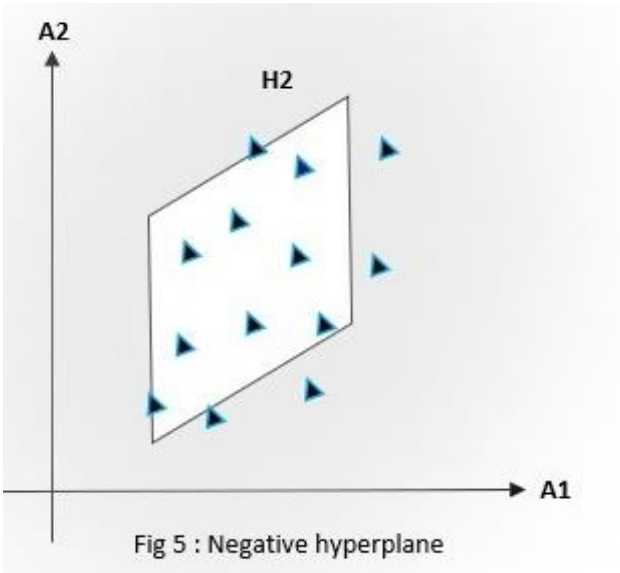

Fig 5 : Negative hyperplane

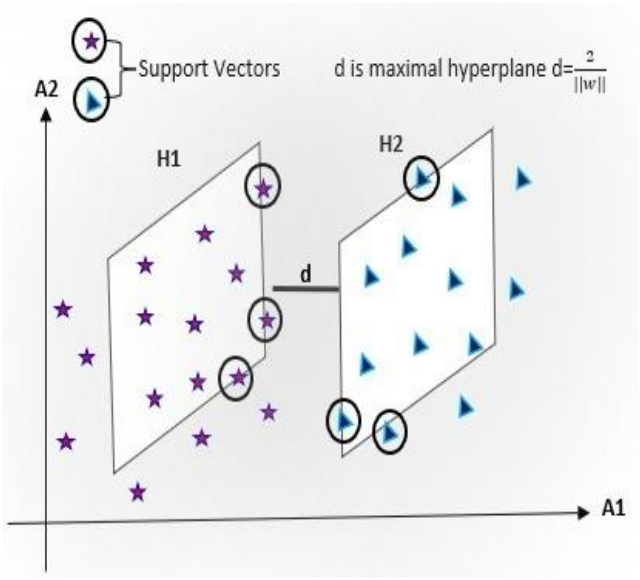

Fig 6: Showing Hyperplanes H1 and H2

Disadvantages: For large datasets the computational cost of SVM is very high because the training matrix grows in quadratic.

More experience is needed to find out which kernel function is to be used in SVM for classification.

Applications of SVM: As discussed in J..Cervantes et al [8] the applications are

- Text categorization
- Image classification
- Bioinformatics (protein and cancer classification)
- Hand written character recognition
- Face detection
- Classification of credit card fraud.

Naïve Bayes Theorem:
H. Zhang [9] states that Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable. Bayes' theorem states the following relationship, given class variable y and dependent feature vector x_1 through x_n, according to Abdulhamit Subasi [6]

$$P(y/x_1,\ldots,x_n) = \frac{P(y)P(x_1,\ldots,x_n|y)}{P(x_1,\ldots,x_n)} \quad (6)$$

Using the naive conditional independence assumption that

$$P(x_i|y, x_1, \ldots x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i y) \quad (7)$$

for all i, this relationship is simplified to

$$P(y|x_1,\ldots,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1,\ldots,x_n)} \quad (8)$$

Since $P(x_{1,\dots\dots}x_n)$ is constant given the input, we can use the following classification rule:

$$P(y|x_{1,\dots\dots}x_n) \propto P(y)\prod_{i=1}^{n}P(x_i|y) \quad (9)$$

and we can use Maximum A Posteriori (MAP) estimation to estimate P(y) and P(x_i|y); the former is then the relative frequency of class y in the training set.

A continuous valued attribute is typically assumed to have gaussian distribution with mean μ and standard deviation σ, defined as

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (10)$$

so that

$$P(x_k|C_i) = g(x_k, \mu_{Ci}, \sigma_{Ci}) \quad (11)$$

Applications:

- Document classification
- spam filtering
- Ranking of Pages

Logistic Regression: Logistic Regression is a supervised learning classification model which is used to classify a categorical variable. Logistic Regression is basically used for binary classification (like 'yes' or 'no'/ 1 or 0) unlike linear regression which is used to classify the continuous data. Classification of data is obtained using Sigmoid function defined as

$$f(x) = \frac{1}{(1+e^{-g(x)})}$$

$$g(x) = x_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_n\beta_n + \varepsilon$$

where $g(x)$ is a function consisting of $x_0, x_1, \dots x_n$ features and their corresponding $\beta_1, \beta_2, \dots \beta_n$ weights and ε is the bias or random error. The values of $f(x)$ lies between 0 and 1 which means sigmoid function which converts linear regression into logistic regression is the probability density function for the data to be predicted. Unlike the graph of linear regression which is a straight line the graph of logistic regression takes 's' shaped curve.

IV. METHODOLOGY

Fig 7 depicts the data flow diagram of the methodology followed in the programming part of this paper. Code is implemented in Python 3 using pandas, numpy, matplotlib, different scalar functions and corresponding functions for different machine learning algorithms for Decision Trees, Naïve Bayes, Logistic Regression and Support Vector Machines.

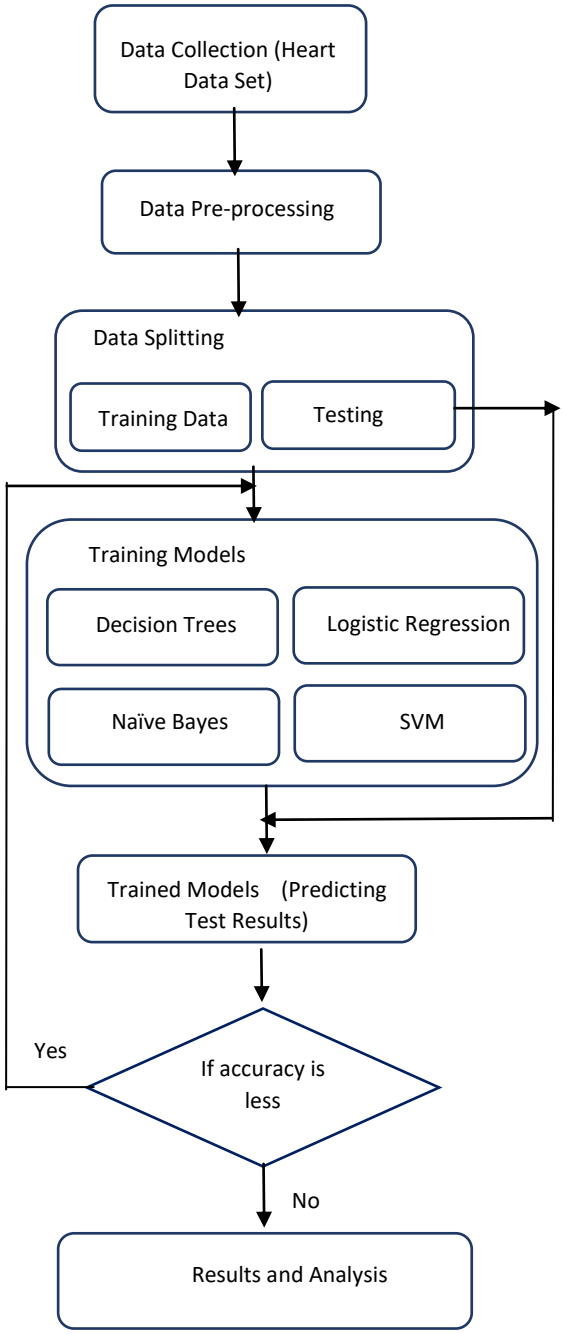

Fig 7: Data Flow Diagram

Data Collection: Dataset used in this case study is from UCI Cleveland. Coronary Artery Disease (CAD), a multivariate classification data is taken and injected in each algorithm discussed above to compare the respective performances. URL is http://archive.ics.uci.edu/ml/datasets/Heart+Disease.. It has a total of 76 raw attributes, but all the published experiments use 14 attributes where the 14th attribute is the target attribute.

UCI Cleveland Heart Disease dataset is described in below Table 3.

Table 3: UCI Cleveland Data set for Heart Disease

Attribute	Description	Type
age	Patient's Age in years	Continuous
sex	Gender of the patient. M: male F: Female	Discrete
cp	This is Categorized in to 4 types 1. typical angina (TA) 2. Atypical angina (ATA) 3. Non- anginal pain (NAP) 4. Asymptomatic (ASY)	Discrete
trestbps	Level of Bloop Pressure in resting mode measure as mm/Hg	Continuous
chol	Serum Cholesterol mg/dl	Continuous
fbs	Fasting Blood Sugar levels 1. > 120 mg/dl 0. <120 mg/dl	Discrete
Restecg	Electrocardiogram results at rest are represented in 3 distinct values. Value 0: Normal State. Value 1: Abnormality state ST-T wave (ST). Value 2: Any Probability or certainty of LV hypertrophy by Estes (LVH)	Discrete
Thalach	Maximum Heart Rate	Continuous
Exang	Angina induced by exercise represented by 2 values. 0:No(N), 1: Yes	Discrete
Oldpeak	Exercise induced ST depression in comparison with state of rest	Continuous
Slope	Slope of ST Segment measured in 3 Values Value 1: ST Elevation Value (Up) 2: Flat Value (Flat) 3: ST- Depression (Down)	Discrete
Ca	Number of major vessels (0-3) coloured by fluoroscopy	Discrete
Thal	thal - 3 = normal; 6 = fixed defect; 7 = reversable defect	Discrete
Diagnosed HD	Heart Disease diagnosis represented in 2 values 0: No 1: Yes	Discrete

Data Pre-processing: Data is cleaned, in the sense that data is checked for any null values This dataset contains a total of 303 rows out of which 6 rows has missing values. StandardScaler() and MinMaxScaler() techniques are used for eliminating any missing values. All the float values are converted to integer values, and checked for any negative values. All the negative values are converted to absolute values.

Data Splitting: The data is split into 80% training set and 20 % test set. Out of 303 rows, 80% training data comes to 242 rows and 20% test data comes to 61 rows.

Training Models: The training data set of 242 rows is given to the training models. In this paper the training models used are Decision Trees, SVM, Naïve Bayes and Logistic Regression.

Trained Models: Testing data set of 61 is given to the trained models to predict the output.

V. RESULTS

Results and Analysis: The results obtained from the predicted test results are evaluated and analyzed. Detailed results with feature selection are shown in table 1.

According to [1] the results for accuracy are calculated as below

Accuracy = $\frac{TP+TN}{P+N}$

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

Sensitivity = $\frac{TP}{P}$

Specificity = $\frac{TN}{N}$

Where

P- is the total number of people with heart disease are **165**.

N- is the total number of people without heart disease are **138**

True Positive (TP) – Number of heart patients predicted correctly as having heart disease

True Negative (TN) - Number of non - heart disease patient predicted correctly not having heart disease

False Positive (FP) – Number of non – heart disease patients predicted incorrectly as heart disease patients

False Negative (FN) – Number of heart disease patients predicted incorrectly as non – heart disease patients.

Table 4: Showing accuracy for different algorithms

	Logistic Regression (LR)	***Naïve Bayes (NV)***	***Decision Tree (DT)***	***SVM***
Accuracy (%)	83.61	**85.25**	75.41	83.61
Precision (%)	81.6	83.8	83.8	80
Recall (%)	91.2	91.2	91.2	94.1
Mean Square Error	0.164	0.148	0.246	0.164
Sensitivity (%)	86.96	87.5	68.75	90.48
Specificity (%)	81.58	83.78	82.76	80

Confusion Matrices for different algorithms for the test data for 61 rows with all the features

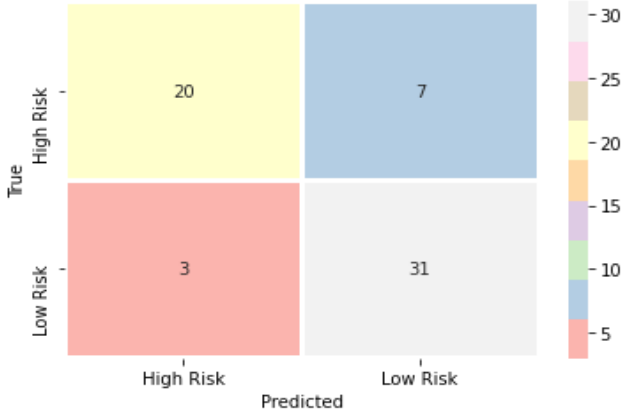

Figure 8: Logistic Regression

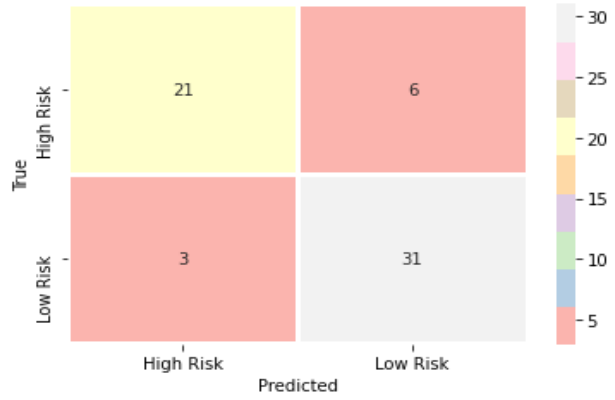

Figure 9: Naïve Bayes

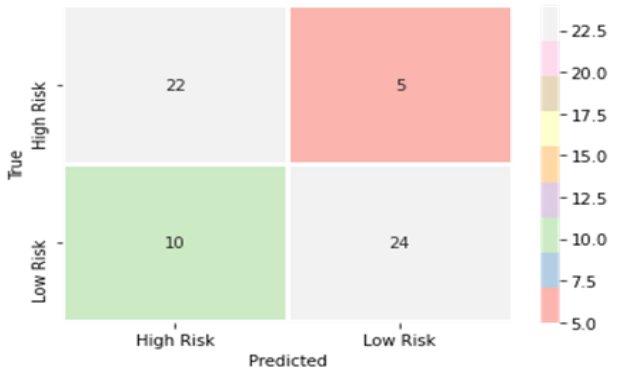

Figure 10: Decision Trees

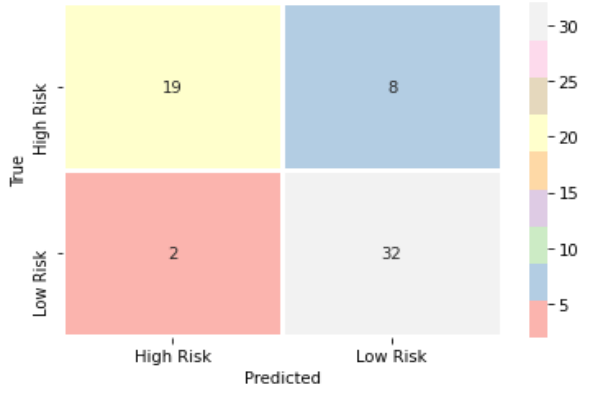

Figure 11: SVM

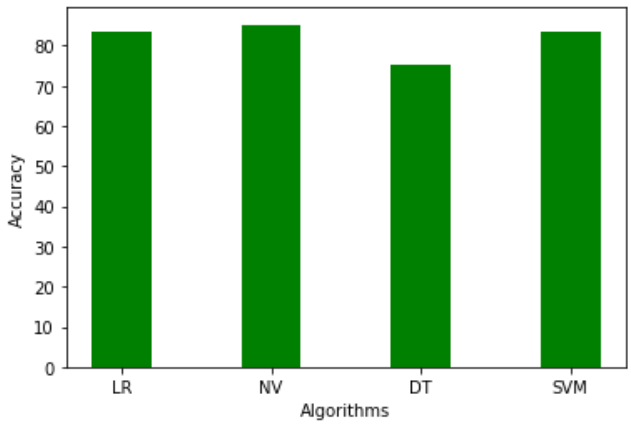

Figure 12 Graph Showing Accuracies of Different Algorithms

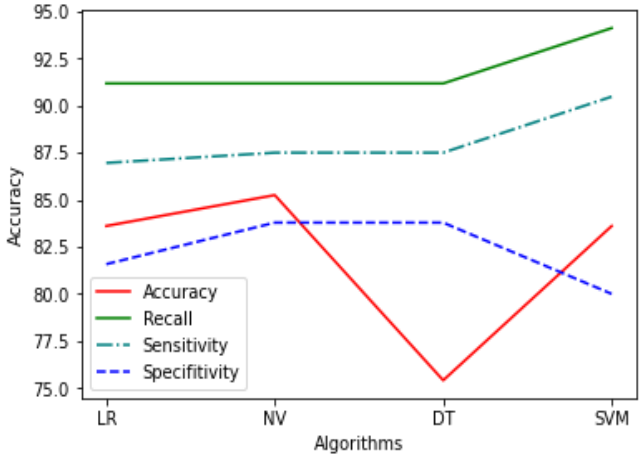

Figure 13: Graph Depicting accuracy, recall, sensitivity and specificity.

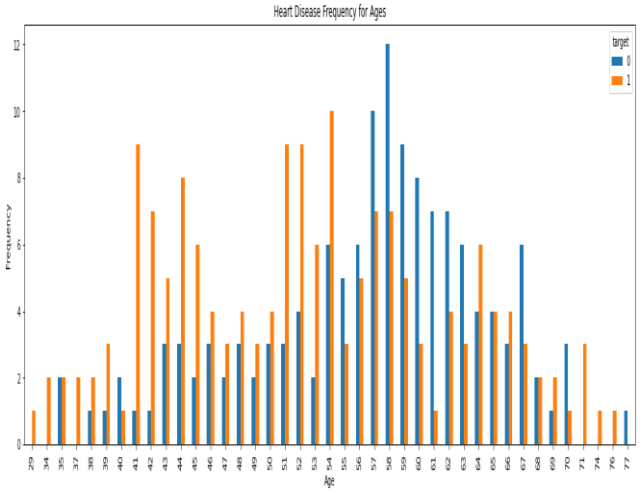

Figure 14: Heart Disease frequency for ages of Male and Female

V. CONCLUSION AND FUTURE SCOPE

Table 4 shows the accuracies, precision, sensitivity, and specificity of different algorithms. Figures 8-11 depicts the confusion matrices of above algorithms. Figure 12 depicts the comparative graph of accuracy for the machine learning algorithms of Logistic Regression, Naïve Bayes, Decision Tree and SVM. Figure 13 depicts the combined graph for accuracy, recall, sensitivity and specificity of the above algorithms. For the heart disease dataset Naïve Bayes algorithm gives the highest accuracy of 85.25% and Decision Tree gives the lowest with 75.41%. According to paper [5] the obtained accuracy was 84.1%. Using the above algorithm, the accuracy is increased to 85.25%. Figure 14 depicts the heart disease frequency for ages of male and female. Males are more prone to heart attack at the age of 58 and females are more prone to heart attack at the age of 54 according to the above dataset. This is a very basic implementation on heart disease dataset. In future similar comparisons using different algorithms, implementing feature selection can be applied to improve the accuracy. This work can be extended to different diseases and on various datasets.

REFERENCES

[1]. Han, j, M. Kamber and J Pei "*Data Mining Concepts and Techniques",* Morgan Kaufmann Publishers USA, **2006.**

[2]. Kaelbling,LeslieP.*;* Littman, Michael L.*;* Moore, Andrew W. . *"Reinforcement Learning: A Survey"*. Journal of Artificial Intelligence Research **vol. 4: pp 237–285.1996.**

[3]. Bramer, M. *Principles of data mining*, Springer, **2007.**

[4]. Quinlan, J. R. *Decision trees and decision-making*. IEEE Transactions on Systems, Man, and Cybernetics, **20(2), pp.339–346, 1990.**

[5]. Mai Shouman, Tim Turner, and Rob Stocker. 2011. *Using decision tree for diagnosing heart disease patients.* In Proceedings of the Ninth Australasian Data Mining Conference - Volume 121(AusDM '11). Australian Computer Society, Inc., **AUS, 23–30, 2011.**

[6]. Cortes, C., Vapnik, V COLT '92: Proceedings of the fifth annual workshop on Computational learning theory **July 1992 Pages 144–152, 1992.**

[7]. Cortes, C., Vapnik, V. *Support-vector networks*. Mach Learn **20, 273–297 1995.**

[8]. J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua et al., A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing,https://doi.org/10.1016/j.neucom.2019.10.118

[9]. H. Zhang. The optimality of Naive Bayes. Proc. FLAIRS 2004 conference, **2004.**

[10]. Russell, Stuart*;* Norvig, Peter *"*Artificial Intelligence: A Modern Approach *" 2nd ed. Prentice Hall.* ISBN 978-0137903955 2003, Chapter 5 **pp.480-502, 2003.**

[11]. McCallum, Andrew; Nigam, Kamal (1998). A comparison of event models for Naive Bayes text classification (PDF). AAAI-98 workshop on learning for text categorization. **Vol. 752, 1998.**

[12]. Daniel Jurafsky, James H. Martin. "Speech and Language Processing" chapter 5

[13]. Feng, J., Wang, Y., Peng, J., Sun, M., Zeng, J., & Jiang, H. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries. Journal of Critical Care, 54, **110–116, 2019.**

[14]. Dissanayake, Kaushalya, Md Johar, Md Gapar (2021) Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms, Hindawi Applied Computational Intelligence and Soft Computing Volume **2021**.

[15]. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in IEEE Access, **vol. 8, pp. 107562-107582, 2020.**

[16]. Bhuvan Sharma, Spinder Kaur, "Analysis and Solutions of Silent Heart Attack Using Python," *International Journal of Computer Sciences and Engineering*, **Vol.10, Issue.1, pp.37-40, 2022.**

[17]. J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, **vol. 8, pp. 107562-107582, 2020.**

[18]. C. Ganesh, E. Kesavulu Reddy, "Overview of the Predictive Data Mining Techniques," *International Journal of Computer Sciences and Engineering*, **Vol.10, Issue.1, pp.28-36, 2022.**

AUTHORS PROFILE

Mrs M. Prameela. pursed Bachelor of Technology in Computer Science and Engineering from JNTU Hyderabad, in 2003 and Master of Technology from Andhra University Visakhapatnam in year 2007. She is currently pursuing Ph.D. in Department of Computer Science and Engineering, University College of Engineering, Adikavi Nannaya University and currently working as Senior Lecturer in Department of Technical Education, State Board of Technical Education and Training Andhra Pradesh since 2012.. Her main research work focuses on Artificial Intelligence, Machine Learning, Deep Learning ,She has 10 years of teaching experience and 5 years of Industry Experience.

Dr M. Kamala Kumari pursed Bachelor of Technology in Computer Science and Engineering from Acharya Nagarjuna University and Master of Technology from JNTU Kakinada. She received her Ph.D in from Adikavi Nannaya University, Rajamahendravaram. She is currently Working as Principal MSN Campus, Kakinada and also Associate Professor in Department of Computer Science and Engineering, University College of Engineering Adikavi Nannaya University. She is a member of IACSIT, CSI, CSTA. She has published more than 10 research papers in reputed international journals and conferences. Her main research work focuses on Artificial Intelligence, Machine Learning, Deep Learning, Data Analysis, Data Science, Big Data Analytics and Data Mining, She has 19 years of teaching experience and 12 years of Research Experience.

