.

ABSTRACT:

According to recent survey by UN agency (World health organization) seventeen.9 million individuals die annually owing to heart connected diseases and it's increasing chop-chop. With the increasing population and illness, it's become a challenge to diagnosis illness and providing the suitable treatment at the proper time. however, there'sa light-weight of hope that recent advances in technologyhave accelerated the general public health sector by developing advanced useful medical specialty solutions. This paper aims at analyzing the assorted datamining techniques particularly Naive Thomas Bayes, Random Forest Classification, call tree and Support Vector Machine by employing a qualified dataset for cardiopathy prediction that is include varied attributes like gender, age, hurting sort, pressure level, blood glucose etc. The analysis includes finding the correlations between the assorted attributes of the dataset by utilizing the quality data processing techniques and thus mistreatment the attributes befittingly topredict the possibilities of a cardiopathy. These machine learning techniques take less time for the prediction of the illness with a lot of accuracy which can cut back the get rid of valuable lives everywhere the planet.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

Chapter 1
INTRODUCTION

1.1 Overview:

Health is one in every of the planet challenges for humanity. World health organization (WHO) has mentioned that for a personal correct health is that the elementary right. thus to stay individuals match and healthy correct health care services ought to be provided. thirty-one proportion of all deaths worldwide square measure due to heart connected problems. identification and treatment of cardiovascular disease is incredibly complicated, significantly in developing countries, because of the shortage of diagnostic devices and a shortage of physicians and alternative resources poignant correct prediction and treatment of internal organ patients. With this concern within the recent times engineering and machine learning techniques square measure being employed to develop code to help doctors in creating call of cardiovascular disease within the preliminary stage. Early stage detection of the malady and predicting the likelihood of an individual to be in danger of cardiovascular disease will scale back the death rate. Medical data processing techniques square measure employed in medical knowledge to extract substantive patterns and data. Medical data has redundancy, multi- attribution, unity and an in depth relationship with time. downside the matter} of mistreatment the large volumes of information effectively becomes a serious problem for the health sector. data processing provides the methodology and technology to convert these knowledge mounds into helpful decision-making data. This postulation system for cardiovascular disease would facilitate Cardiologists intaking faster choices in order that a lot of patients will receive treatments inside a shorter amount of your time.

1.2 Scope Of the Project:

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using several classification algorithms used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy.

1.3 Domain Overview:

1.3.1 MACHINE LEARNING

Machine Learning is a system that can learn from example through self-improvement and without being explicitly coded by programmer. The breakthrough comes with the idea that a machine can singularly learn from the data(i.e., example) to produce accurate results.

Machine learning combines data with statistical tools to predict an output. This output is then used by corporate to makes actionable insights. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input, use an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

Machine learning is also used for a variety of task like fraud detection, predictive maintenance, portfolio optimization, automatize task and so on.

1.4 Machine Learning vs. Traditional Programming

Traditional programming differs significantly from machine learning. In traditional programming, a programmer code all the rules in consultation with an expert in the industry for which software is being developed. Each rule is based on a logical foundation; the machine will execute an output following the logical statement. When the system grows complex, more rules need to be written. It can quickly become unsustainable to maintain.

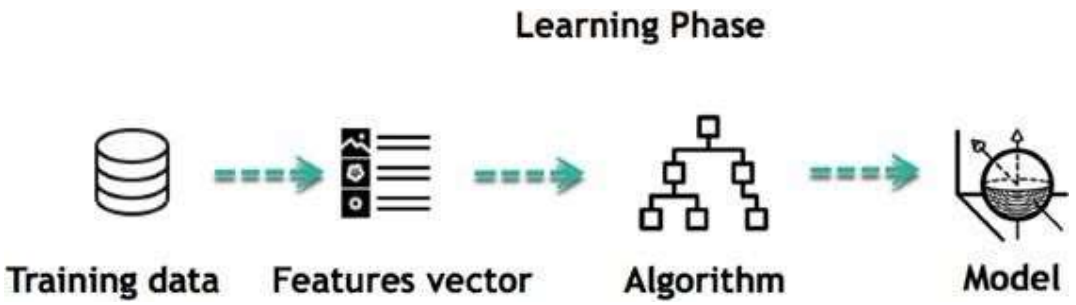
fig 1.1 Machine Learning

1.4.1 How does Machine learning work?

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict.

The core objective of machine learning is the **learning** and **inference**. First of all, the machine learns through the discovery of patterns. This discovery is made thanks to the **data**. One crucial part of the data scientist is to choose carefully which data to provide to the machine. The list of attributes used to solve a problem is called a **feature vector.** You can think of a feature vector as a subset of data thatis used to tackle a problem.

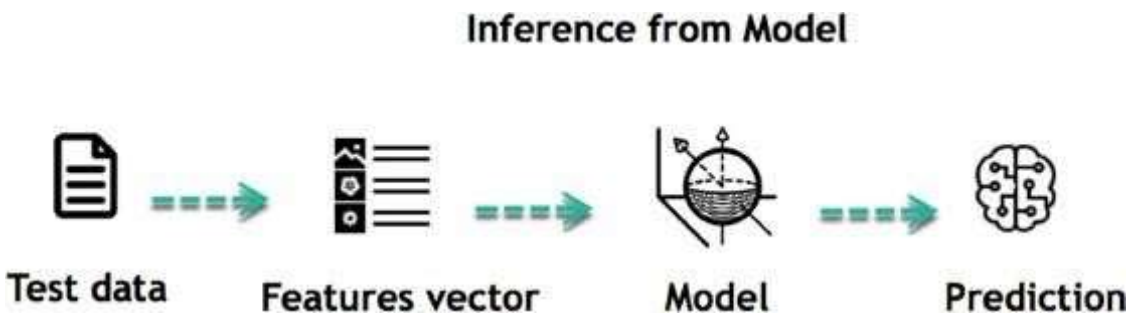
The machine uses some fancy algorithms to simplify the reality and transform this discovery into a **model**. Therefore, the learning stage is used to describe the data and summarize it into a model.



For instance, the machine is trying to understand the relationship between the wage of an individual and the likelihood to go to a fancy restaurant. It turns out the machine finds a positive relationship between wage and going to a high-end restaurant: This is the model

1.5 Inferring

When the model is built, it is possible to test how powerful it is on never-seen- before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the modelpreviously trained to make inference on new data.



The life of Machine Learning programs is straightforward and can be summarized in the following points:

1. Define a question
2. Collect data
3. Visualize data
4. Train algorithm
5. Test the Algorithm
6. Collect feedback
7. Refine the algorithm
8. Loop 4-7 until the results are satisfying
9. Use the model to make a prediction

Once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data.