NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTION, AFFILIATED TO VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELGAUM,

APPROVED BY AICTE & GOVT.OF KARNATAKA

Learning Activity Project Proposal

on

HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES

Submitted in partial fulfilment of the requirement for the award of Degree of

Bachelor of Engineering

in

Computer Science and Engineering

Faculty In-charge: Dr. VANI VASUDEVAN

Submitted by:

D S SAI JAHNAVI	1NT19CS062
JALPA C J	1NT19CS088
LEENA JENNIFER EDWIN	1NT19CS100
MAHATHI D N	1NT19CS104

INTRODUCTION:

Data mining is defined as **a process used to extract usable data from a larger set of any raw data**. It implies analysing data patterns in large batches of data using one or more software. Data mining has applications in multiple fields, like science and research.

A dataset is a structured collection of data generally associated with a unique body of work.

The heart is the most crucial & critical organ of the human body. Life is completely dependent on the efficient working & functioning of our heart. It is one of the major causes of mortality in today's world. Heart disease remains one of the most serious health issues of our day. It is said to be the primary motive in death globally. Many times it's difficult for medical professionals to expect a heart disease on time.

Nowadays, the health sector contains a lot of precious hidden facts & information which could prove to be very helpful in making predictive decisions especially in the field of medicine.

Data mining is a method or technique used to analyze vast datasets and then derive significant and useful results with the use of extraordinary AI-based techniques. This article attempts to use three of these AI-based methods namely Decision Tree, Naïve Bayes, & Neural Network for forecasting cardiovascular or heart disease. All of these methods will be evaluated based on different unique & parameters with optimizations for better accuracy. The accuracy of each method will then be compared depending on accuracy based on various parameters. The best & accurate technique is then implemented for predicting whether or not a man or a woman will have coronary heart disease. This technique can be used by medical practitioners for early prediction of the disease so that timely care can be taken by the patient.

DATA MINING TASKS:

- Classification techniques – used widely in healthcare because of their capabilities of processing very large data sets.
- Prediction – It uses regression analysis and detect the missing values of data
- Outlier analysis–To discard noise
- Correlation analysis -To reduce dataset ,to get more accuracy
- Data discrimination-compares common features of class which is under study. Output can be represented in many forms like graphs, pie chart, curves

DATA SET:

We took the dataset from Kaggle website, and it has the attributes Age ,Sex, Chest pain type, BP, Cholesterol ,FBS over 120, EKG results, max HR ,Exercise angina ,ST depression.

METHODS AND MODELS:

To come out with a good output we use the following methods:

1. Classification
2. Decision Tree
3. Naïve Bayes
4. Neural Networks

ASSESSMENT:

The **Naive Bayes** classification algorithm is a probabilistic classifier. It is based on probability models that incorporate strong independence assumptions.

The independence assumptions often do not have an impact on reality. Therefore they are considered as naive

A Naive Bayes model consists of a large cube that includes the following dimensions:

- Input field name
- Input field value for discrete fields, or input field value range for continuous fields.

  Continuous fields are divided into discrete bins by the Naive Bayes algorithm

- Target field value

A **Decision tree** is a graph that uses a branching method to illustrate every possible outcome of a decision. Decision trees can be drawn by hand or created with a graphics program or specialized software. decision trees are useful for focusing discussion when a group must make a decision.

PRESENTATION AND VISUALIZATION:

AI-based methods namely Decision Tree, Naïve Bayes, & Neural Network for forecasting cardiovascular or heart disease. All of these methods will be evaluated based on different unique & parameters with optimizations for better accuracy. The accuracy of each method will then be compared depending on accuracy based on various parameters. The best & accurate technique is then implemented for predicting whether or not a man or a woman will have coronary heart disease. This technique can be used by medical practitioners for early prediction of the disease so that timely care can be taken by the patient.
For visualization, the graphs would be a good fit. Using matplot in python the graphs are created.

ROLES:

D S Sai Jahnavi	Working on the data visualization, presentation, Training the model, Drafting the project proposal, Report
Jalpa C J	Working on the aggregation of the results Training the model, Drafting the project proposal, Report
Leena Jennifer Edwin	Implementing the algorithm Training the model, Drafting the project proposal, Report
Mahathi D N	Data collection, Drafting the project proposal Training the model, Report

SCHEDULE:

Date	Task to be completed
03/01/2022	Proposal uploaded in github
15/01/2022	Completion of implementation
17/01/2022	Final submission of report

BIBLIOGRAPHY

1. www.kaggle.com
2. **https://www.kaggle.com/rishidamarla/heart-disease-prediction**
3. **https://www.tutorialspoint.com/data_mining/index.htm**
4. **https://youtu.be/D3DHNDD-hbY**