Department of Computer Engineering

<u>**Date- 13 / 10 /2021**</u>

TE Product Based Seminar

SYNOPSIS

Name of Student: Sayali Padmawar
Roll No: TC143
Guide Name: Prof. Jagruti Wagh

Title of Seminar: Heart Disease Prediction using Machine Learning Algorithms

Introduction and Motivation:

Heart plays a significant role in living organisms. Diagnosis and prediction of heart related diseases requires more precision, perfection and correctness because a little mistake can cause fatigue problems or death of the person. There are numerous death cases related to heart disease and their counting is increasing exponentially day by day. To deal with the problem there is an essential need for a prediction system for awareness about diseases. Machine learning is the branch of Artificial Intelligence(AI), it provides prestigious support in predicting any kind of event which takes training from natural events.

Sometimes heart disease may be "silent" and not diagnosed until a person experiences signs or symptoms of a heart attack, heart failure, or an arrhythmia.

When these events happen, symptoms may include:-

➢ <u>Heart Attack</u>: Chest pain or discomfort, upper back or neck pain, indigestion, heartburn, nausea or vomiting, extreme fatigue, upper body discomfort, dizziness, and shortness of breath.
➢ <u>Arrhythmia</u>**:** Fluttering feelings in the chest (palpitations).
➢ <u>Heart Failure</u>: Shortness of breath, fatigue, or swelling of the feet, ankles, legs, abdomen, or neck veins.

● Heart disease is most common in the present era. The treatment cost of heart disease is not affordable by most of the patients.So we can reduce this problem by a Heart Disease Prediction System (HDPS).It is helpful for earlier diagnosis of heart disease.
● It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors.Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease.
● Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.

With the rampant increase in the heart stroke rates at juvenile ages, I thought there is a need to put a system in place to be able to detect the symptoms of a heart stroke at an early stage

and thus prevent it. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. Thus I propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc.

Literature Review:

The classification of high dimensional data produces wrong results. To get the accurate results the high dimensional data set must be compressed to a lower level in order to improve the accuracy of the results. To do so the redundant and the inconsistent data must be removed. They introduced a method using dimensionality reduction techniques. It used the constraint selection algorithm for eliminating the reduction elimination. The processing time is reduced due to the dimensionality reduction techniques

In Heart Disease Prediction System Using Data Mining Technique,It presents a heart disease prediction system based on naïve bayes algorithm .This system is convenient, effective and gives good prediction of diseases to users. Overall conclusion is that, this is a better prediction system for customers using data mining technology.

In Heart Disease Prediction using Machine Learning it studies & compares the accuracy score of Decision Tree, Logistic Regression, Random Forest and Naive Bayes algorithms for predicting heart disease using UCI machine learning repository dataset. The result of this study indicates that the Random Forest algorithm is the most efficient algorithm with an accuracy score of 90.16% for prediction of heart disease.

In Heart Disease Prediction System using Data Mining Method by Keerthana T K ,In this paper the precision and accuracy of three different classifiers are measured. The result shows Naive Bayesian classification possesses high precision and less error rate. Random Forest classification method produces better results than J48 classification.

Various approaches for the prediction of heart attack risk levels from the heart disease database are done in medical industries. Firstly, the heart disease database is clustered using the K-means clustering algorithm. The ID3 algorithm has been used as the training algorithm to show the level of heart attack with the decision tree.

The dimensionality reduction technique reduces the space and increases the performance. Metaheuristics techniques are used for dimensionality reduction. Dimensionality Reduction technique is more useful to minimize space; information retrieval is faster, efficient image processing, better visualization, accurate classification for domain oriented data sets.

A classification approach which uses Artificial Neural Network (ANN) and PCA for feature subset selection is used to analyze the dataset.Similarly, several techniques have been used to improve the accuracy of the classifiers for better prediction.

Objective:

- The objective of the project is to collect the data and train the data by a trained dataset and predicting the disease using machine learning dimensionality reduction techniques.

- To detect the symptoms of a heart stroke using the Heart Disease Prediction System at an early stage and thus prevent it.

- Disease prediction has the potential to benefit stakeholders such as the government and health insurance companies and hence this can help to identify patients at risk of disease or health conditions.

Problem Statement: Heart Disease Prediction System using Machine Learning Algorithm to save People's life.

Methodology/ Algorithm in Brief:

DIMENSIONALITY REDUCTION:-

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done. These factors are basically variables called features. The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play. Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

There are two components of dimensionality reduction:

Feature selection: In this, we try to find a subset of the original set of variables, or features, to get a smaller subset which can be used to model the problem. It usually involves three ways:
1. Filter
2. Wrapper
3. Embedded

Feature extraction: This reduces the data in a high dimensional space to a lower dimension space, i.e. a space with lesser no. of dimensions.

Methods of Dimensionality Reduction
The various methods used for dimensionality reduction include:

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Generalized Discriminant Analysis (GDA)

    FLOW:-
    1. Collection of Data set for Prediction Purpose.
    2. Apply the Data Processing Techniques.
    3. Apply Machine Learning Algorithms.
    4. Disease Prediction.

The proposed methodology includes steps, referred to as the:-

1. Preprocessing stage where the exhaustive exploration of the data is carried out. It will account for dealing with missing values, balancing data and normalizing attributes depending on algorithms used.
2. After pre-processing of data is performed, predictive modeling of the data is carried out using classification models and ensemble approach.
3. Finally, prescriptive modeling is undertaken, where the predictive model is evaluated in terms of performance and accuracy using various performance metrics. Accordingly, the modifications are either made to the data preparation stage or the predictive modeling strategy is amended

Principal component analysis (PCA) and feature selection using information gain ratio can be used to select the relevant attributes . This is created with the step by step process.

- The first step is to remove the outliers. The outliers are the observed data that deviated much from the observed data. It can also be called noise. The noise can be data noise or attribute noise. The data cleaning is carried out as the initial step in the data mining process to remove the outliers.
- The second step is the feature extraction where the principal component analysis is used to extract the critical feature which is the most relevant feature.

Benefits:-
- It helps in data compressing and reducing the storage space required
- It fastens the time required for performing the same computations. Less dimensions leads to less computing, also less dimensions can allow usage of algorithms unfit for a large number of dimensions
- It takes care of multicollinearity that improves the model performance. It removes redundant features. For example: there is no point in storing a value in two different units (meters and inches).
- Reducing the dimensions of data to 2D or 3D may allow us to plot and visualize it precisely. You can then observe patterns more clearly. Below you can see how a 3D data is converted into 2D. First it has identified the 2D plane then represented the points on these two new axes z1 and z2.

Scope for Social Problem Solution:

The user can select various symptoms and can find the diseases with their probabilistic figures. The project can be improved by implementing medicine suggestion to the patient along with the results. Also I can implement a feedback from the experienced doctors who can give their views and opinions about certain medicines /practices done by the doctor on the patient. I can implement a live chat option where the patient can chat with a doctor available regarding medication for the respective result for their symptoms.
Project could be used as a training tool for Nurses and Doctors who are freshly introduced in the field related to heart diseases. The patient can have a choice in choosing the medicines he/she should take in order to have a healthier life. Moreover, if implemented on a large scale it can be used in medical facilities like hospital, clinics where a patient wouldn't have to wait in long queues for treatment if he is feeling symptoms related to heart disease.
In future an intelligent system may be developed that can lead to selection of proper treatment methods for a patient diagnosed with heart disease. A lot of work has been done already in making

models that can predict whether a patient is likely to develop heart disease or not. There are several treatment methods for a patient once diagnosed with a particular form of heart disease. Data mining can be of very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases.

References:

- Archana Singh,Rakesh Kumar,Heart Disease Prediction Using Machine Learning Algorithms,International Conference on Electrical and Electronics Engineering (ICE3-20201),Feb 2020.
- Gaoshuai Wang,Fabrice Lauri,Amir Hajjam EL Hassani ,A Study of Dimensionality Reduction's Influence on Heart Disease Prediction,IEEE,July 2021.
- R.Kavitha,E.Kannan,An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining,IEEE,October 2016
- Geetha.M, Ganesan.R, Tallam Tharun Sai,Research on the Machine Learning Algorithms on Heart Condition Predictions,International Journal of Innovative Technology and Exploring Engineering (IJITEE),April 2019
- Devendra Ratnaparkhi, Tushar Mahajan, Vishal Jadhav,Heart Disease Prediction System using Data Mining Method,International Research Journal of Engineering and Technology (IRJET),Volume2,Nov 2015.
- Keerthana T K,Heart Disease Prediction System using Data Mining Method,(IJETT) – Volume 47 Number 6,6 May 2017,
- M. Marimuthu, M. Abinaya ,K. S. Hariesh , K. Madhankumar ,V. Pavithra , A Review on Heart Disease Prediction using Machine Learning and Data Analytics , 18, September 2018 Approach,International Journal of Computer Applications (0975 – 8887) Volume 181.
- Khaled Mohamad Almustafa,Prediction of heart disease and classifiers' sensitivity analysis,2020.
- Swati A sonawale & Roshani Ade: Dimensionality Reduction: An Effective Technique for Feature Selection : International Journal of Computer Applications, Volume 117-No 3, May (2015).
- Tarun Kumar Gupta, Chanchal Kumar, Shiv Prakash and Mukesh Prasad: Dimensionality Reduction Techniques and its Applications: Computer Science Systems Biology, (2015).

Prof. Jagruti Wagh

Name and Signature of Guide