A picture containing text, clipart

Description automatically generated**PROJECT REPORT**

**Machine Learning-DS2231**

Detection of Heart Disease

Group Members:

Harsh Kumar-219309034

Aditi Chandorkar-219309056

Guided by: Mr. Venkatesh Shankar

Date of Submission:17-04-2023

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sno.** | **Section** | **Page-No.** |
| 1. | Introduction and background | 3 |
| 2. | Methodology and Experiment Setup | 4 |
| 3. | Result and Discussion | 6 |
| 4. | Conclusion and references | 8 |

**Introduction and Background**

“Machine Learning is a way of Manipulating and extraction of implicit, previously unknown/known and potential useful information about data”

Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. Machine learning Incorporates various classifiers of Supervised, Unsupervised and Ensemble Learning which are used to predict and Find the Accuracy of the given dataset. We can use that knowledge in our project of HDPS as it will help a lot of people.

Cardiovascular diseases are very common these days, they describe a range of conditions that could affect your heart. World health organization estimates that 17.9 million global deaths from (Cardiovascular diseases) CVDs. It is the primary reason of deaths in adults. Our project can help predict the people who are likely to diagnose with a heart disease by help of their medical history. It recognizes who all are having any symptoms of heart disease such as chest pain or high blood pressure and can help in diagnosing disease with less medical tests and effective treatments, so that they can be cured accordingly.

In this project our main focus is to run machine learning algorithms on a heart disease symptoms dataset and draw out appropriate inferences accordingly. We have used classification and clustering algorithms including Principal Component Analysis, Naïve Bayes algorithm, Hierarchical Agglomerative Clustering, etc. for the same.

**Methodology and Experiment Setup**

This report shows the analysis of various machine learning algorithms, the algorithms that are used in this paper are K nearest neighbours (KNN), Principal Component Analysis, Naïve Bayes Algorithm, Hierarchical Agglomerative Clustering , K-Means algorithm and Logistic Regression which can be helpful for practitioners or medical analysts for accurately diagnose Heart Disease.

The methodology is a process which includes steps that transform given data into recognized data patterns for the knowledge of the users. The proposed methodology (Figure 1.) includes steps, where first step is referred as the collection of the data than in second stage it extracts significant values than the 3rd is the pre-processing stage where we explore the data.

Data pre-processing deals with the missing values, cleaning of data and normalization depending on algorithms used.

Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics. Here in this model, an effective Heart Disease Prediction System has been developed using different classifiers.

**Training (KNN, Logistic**

**Regression, PCA)**

**Normal**

**Heart Disease**

**Classifier**

**Testing Data**

**Training Data**

**Splitting Data**

**Data Pre-processing**

**Extract Significant Variables**

**Collect Heart Disease**

**Dataset**

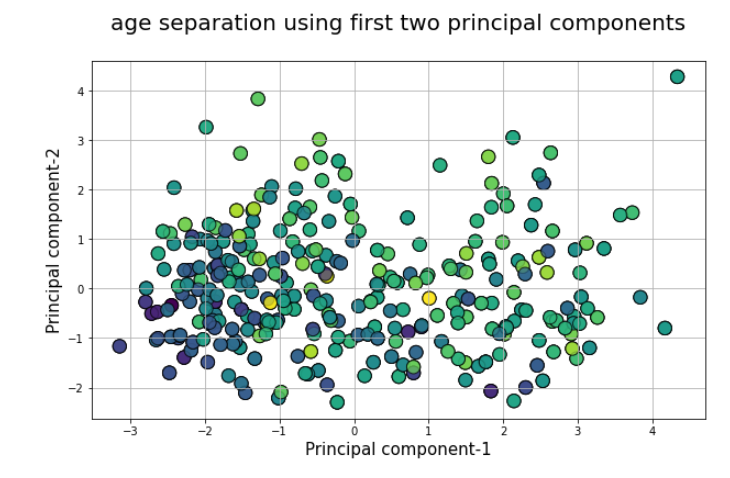
**Start**

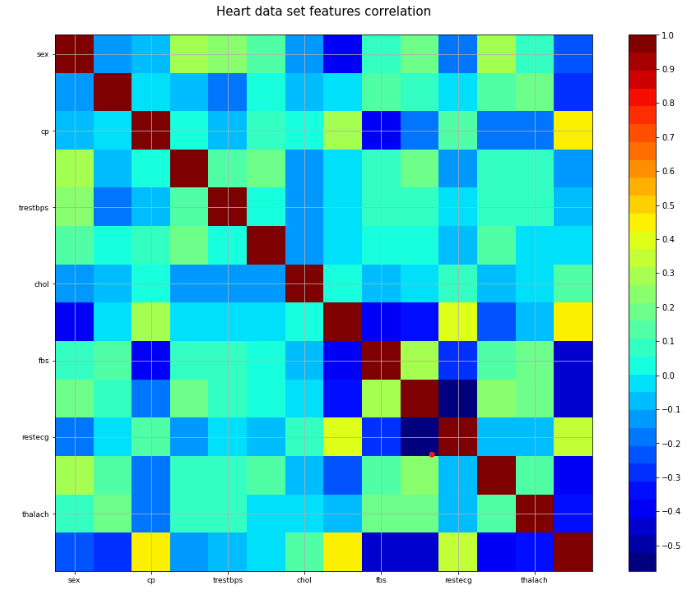
**Result & Discussion**

The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from previous researches. So, we summarize that our accuracy is improved due to the increased medical attributes that we used from the dataset we took.



The figure above shows patients having or not having heart disease on the basis of sex.





These figures show possibility of heart disease in patients based on their age using the first two principal components.

**Conclusions**

A cardiovascular disease detection model has been developed using ML classification and clustering modelling techniques. This project predicts people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients’ medical history such as chest pain, sugar level, blood pressure, etc.

This Heart Disease detection system assists a patient based on his/her clinical information of them been diagnosed with a previous heart disease. The algorithms used in building the given model are Logistic regression, Principal Component Analysis and KNN . The accuracy of our model is 87.5%. Use of more training data ensures the higher chances of the model to accurately predict whether the given person has a heart disease or not .

By using these, computer aided techniques we can predict the patient fast and better and the cost can be reduced very much. There are a number of medical databases that we can work on as these Machine learning techniques are better and they can predict better than a human being which helps the patient as well as the doctors.

Therefore, in conclusion this project helps us predict the patients who are diagnosed with heart diseases by cleaning the dataset and applying logistic regression and KNN to get an accuracy of an average of 87.5% on our model which is better than the previous models having an accuracy of 85%. Also, it is concluded that accuracy of KNN is highest between the three algorithms that we have used.

References:

Opi-sciences- open source research paper