A

PROJECT REPORT

ON

HEART DISEASE ANALYSIS
AND PREDICTION

SUBMITTED TO

CHHATTISGARH SWAMI VIVEKANAND TECHNICAL UNIVERSITY
BHILAI (C.G.)

FOR FULFILLMENT OF THE AWARD OF DEGREE
BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

UNDER THE GUIDANCE OF

PROF. (MR./ MS. MADHURI GUPTA)

BY

AKASH LILHARE

(SECTION – B)

DEPARTMENT OF INFORMATION TECHNOLOGY

SHRI SHANKARACHARYA TECHNICAL CAMPUS
P.O. JUNWANI NEHRU NAGAR BHILAI-490020
SESSION: 2021-2022

Email : akashlilhare14@gmail.com

Project Link: https://github.com/akashlilhare/heart-disease-analysis

Abstract -

Heart disease is one of the dangerous diseases in the world where it may cause death and the patient who has this disease may undergo a serious long term disability. Effective tools and different models will be used to discover this system and new skills in e-health data. In Heart Disease Prediction medical diagnosis plays an important role to help and save the patient life so it has to be used or be executed accurately and efficiently. An accurate and appropriate computer-based exact decision provider system is required to reduce the cost of clinical tests. The main aim of this study is to find hidden features by using data mining techniques, which are necessary to find heart diseases and to predict the presence of heart disease in patients.

Keywords -

Prediction, Machine learning, Algorithms, Accuracy, Heart disease, Comparing

I. Introduction

Heart disease is one of the major diseases in the world since many years ago according to the survey at least one person will die due to Heart disease for every one minute in the world. many researcher studies on this by using Data Mining techniques to help the patient diagnosis who are suffering from Heart Disease. Data mining techniques reduces the test number that is required they use quick and effective techniques to reduce the deaths cases from Heart Disease. Here, datasets consist of 303 instances along with 14 attributes. A classification algorithm is used to reduce the size of the data by its optimal potential as all know that the heart will be the second most important organ in the human body. the main work of the heart in the human body is to pumps blood and supplies to other organs in the human body. Predicting heart disease is the most significant work in the medical field. Data analytics plays a useful role in the prediction of more information and it also helps in the prediction of various disease in the medical centre. In the health centre, a large number of datasets related to each patient will be recorded based on monthly records. This recorded data will be helpful in future for the prediction of disease here some of the data mining and also machine learning techniques that have been used to predict heart disease such as Logistic Regression, K-Nearest Neighbor(KNN) ClassifierRandom Forest Classifier This paper provides the discernment of each six existing algorithms and also an overall summary of work. So, nowadays Heart Disease is one of the prevalent problems in human life [8].

II. Review Of Literature

Many kinds of research had been carried out with intention of predicting heart disease at an earlier stage so that doctor can give a treatment earlier the patient can live a longer life. Heart disease can be predicted based on some symptoms like chest pain, difficulty in breathing, fullness, detecting diseases in patient and giving the treatment based on the disease stage may help the patient to live longer. et al. [1] In today world Heart Disease is the major disease causes of mortality in the world. Prediction of heart disease is also one of the difficult challenges in many medical centres. The large number of data produced by

different health centres will use machine learning techniques for decision making and prediction. These days machine learning will be used in different areas of IoT by studying machine learning techniques it just gives only a glimpse into a prediction of heart disease. In this paper, the aim is to solve the heart disease prediction problem with some significant feature by using machine learning technique which improves the accuracy level in the prediction of heart disease. et al. [2] One of the most common types of heart disease is coronary heart disease where it affecting the heart and causes death according to the view of medical science data mining plays an important role in discovering the various metabolic syndromes. Here the data prediction and analysis can be done through classification techniques. To detect and predict the events occurring in CHD can be found by using Decision tree techniques. In this paper, the incidence related the CHD and accuracy can be predicted by using random forest which is developed from data mining. This model helps in the prediction of CHD and shows you that it is related to different segments of the population. et al. [3] According to this paper heart disease is one of the global health extrusion in the medical system. The experiences of human and some of the expertise in heart disease diagnosis will cause inaccurate diagnosis. There are various types of medical equipment which are helpful in collecting information about the illness or less

accurate in the prediction. In this paper particle swarm optimization algorithm is used to develop some set of rules for working in heart disease prediction. Here, the first random rules will the applied after that based on accuracy they going to optimize the dataset. et al. [4] Heart disease is one of the major reason for the increase in the death rate. Healthcare is one of the most important beneficiaries of huge knowledge & analytics. Extracting medical data is progressively becoming more and more necessary for the prediction and treatment of high death rate due to heart attack. Terabytes of data are produced every day. Quality services are needed to avoid poor clinical decisions that lead to disastrous consequences. Hospitals can make use of appropriate decision support systems thus minimizing the cost of clinical tests. et al. [5] Heart Disease can also cause sudden heart failure which will be considered as one of the dangerous disease y prevention of HF (heart failure) risks will help to provide the treatment in earlier stages. Here they used an artificial neural network, which helps in diagnosis for HF and supports in the investigation of HF risks of the attributes. The equal risk assumption along with existing methods would not help in the diagnosis of HF patient. et al. [6] In this paper they introduced a hybrid method to combine the different algorithms and features selection techniques. Here the dataset is collected from UCI and among 76 attributes are selected to perform different matching learning model.
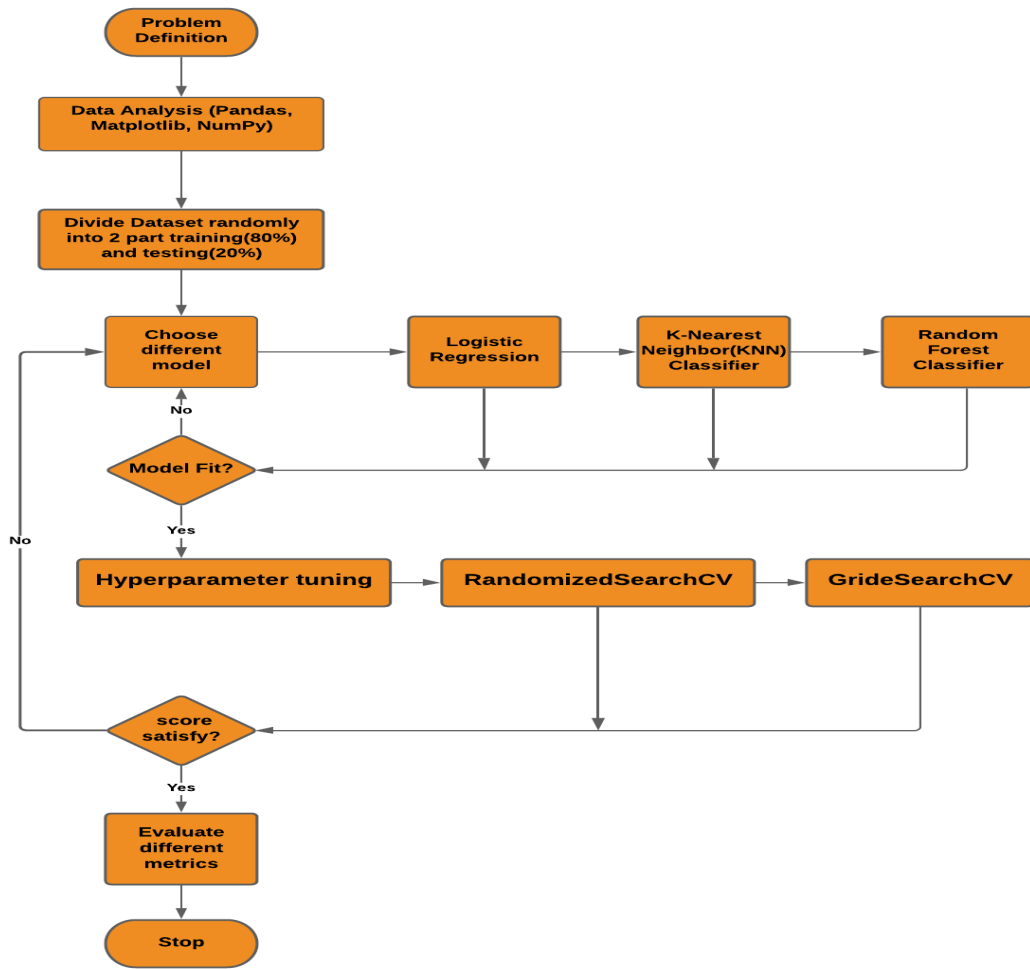
III. Materials and Methods

Fig 1: **Methodology of Heart Disease Prediction**

1. Data Definition

The analysis is carried out using publicly available data for heart disease. The dataset holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain. The dataset is analysed with visualization tool tableau and K means clustering. The dataset to define the proposed algorithm is the Cleveland heart disease raw dataset with 76 features of 303 patients. During the pre-processing method, some samples are removed to eradicate error due to inconsistency of data. The prediction of heart disease is made with 209 samples with seven independent features like age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain and the habitual of physical exercise. Age is considered as the main risk factor for heart diseases as coronary fatty streaks develop in the adolescence stage. Male are at higher risk of coronary diseases than females, hence the data set considered here is for The only male. Angina is the discomfort caused when the muscles of the heart are not supplied with sufficient oxygen-rich blood. High blood pressure is one of the major causes of heart

disease as it damages arteries. Blood pressure combined with diabetes can increase the risk even more. Heart rate with high blood pressure increases the risk of heart diseases. Heartbeat rate is directly proportional to the risk of coronary disease. The symptom of heart disease includes feeling gripping and tight usually on the chest but spread to shoulders up to the stomach. The types of angina are atypical angina, typical angina, asymptomatic and non-anginal pain.

Sl.No	Feature Name	Feature code	Description	Domain of Value
1	Age	Age	Age of the person in years	28 <age< 66
2	Type of chest pain	chest_pain	1. atypical angina 2. typical angina 3. asymptomatic 4. non-anginal pain	1 2 3 4
3	Resting blood pressure	rest_bpress	mm Hg	92 to 200
4	Fasting blood sugar	blood_sugar	Fasting blood sugar >120 mg/dl	t = true f = false

Table 2: **Features information and description of Clevel and heart disease dataset**
(Source:https://archive.ics.uci.edu/ml/datasets/Heart+Disease)

2. Data Analysis

Fig-1 is data visualization for the maximum heart rate of a person with age. With the help of Fig-1, we can determine which age group of person are most valuable to heart disease.
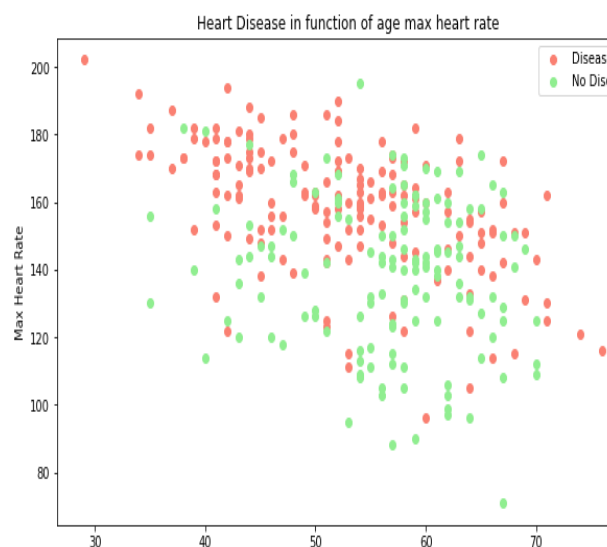

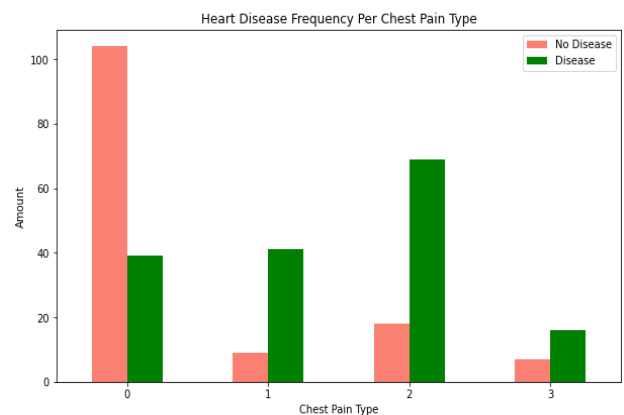**Fig -2 Heart Disease in functionof age and max heart rate**

Fig-3 Heart Disease Frequency Per Chest Pain type

Heart Disease frequency per Chest pain type

- **0: Typical angina**: chest pain related to decreasing blood supply to the heart
- **1: Atypical angina**: chest pain not related to heart
- **2: Non-anginal pain**: typically esophageal spasms (non-heart related)
- **3:Asymptomatic**: chest pain not showing signs of disease

3 Applying machine learning model

After analysing data we will split data into training and test sets, After splitting data into two-part we can apply different machine learning model to training data once we get the accuracy score for the dataset we can test the model for the test dataset for finding accuracy score we are going to try 3 different machine learning model.

  a. **Logistic Regression**
  b. **K-Nearest Neighbours classifier**
  c. **Random Forest classifier**

After applying different machine learning model we get the following accuracy score

Machine Learning model	Accuracy score
Logistic Regression	0.885245901639
K-Nearest Neighbours classifier	0.688524590163
Random Forest classifier	0.836065573770

Table 3: Accuracy Score of different Machine Learning Model

Since the LogisticRegression model provides the best scores so far, we will try to improve them we will use the 2 different hyperparameter tuning methods to improve our score further.

  a. **RandomizedSearchCV**
  b. **GrideSearchCV**

Iv. Conclusion

Since the LogisticRegression model with GrideSearchCV provides the best scores so far. We are going to calculate another matrix for heart disease analysis we are going to calculate the following matrix.

  ● **Confusion matrix**
  ● **Precision**
  ● **Recall**
  ● **F1-score**

The score for the confusion matrix given below

Fig-4 Confusion Matrix

The scores for the Precision, Recall and F1-score given in below table

Attribute	Precision	Recall	F1-Score	Support
0	0.89	0.86	0.88	29
1	0.88	0.91	0.89	32
accuracy			0.89	61
macro avg	0.89	0.88	0.88	61
weighted avg	0.89	0.89	0.89	61

Table 4: Precision, Recall and F1- score

V. Tools used for Analysis and Prediction

1. Pandas

pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license.

2. Numpy

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

3. Matplotlib

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK

4. Scikit-learn

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machines.

Vi. REFERENCES

1. Senthil Kumar Mohan, Chandrasegar Thirumala and Gautham Srivatsan Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques in Proc. Int. Conf. Recent Trends Methods, Controls, IEEE 2019, pp. 322342.
2. A. S. Abdullah and R. R. Rajalaxmi, A data mining model for predicting the coronary heart disease using random forest classifier, in Proc. Int. Conf. Recent Trends Comput. Methods, Commun. Controls, Apr. 2012, pp. 2225.
3. A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, Using PSO algorithm for producing best rules in diagnosis of heart disease, in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306 311.
4. C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, Analysis of neural networks based heart disease prediction system, in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233239.

5. O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, and G. Li, An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction, Expert Syst. Appl., vol. 68, pp. 163172, Feb. 2017.

6. Deepali Chandna, 2014, Diagnosis of Heart Disease Using Data Mining Algorithm, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (2), pp. 1678-1680.

7. Vikas Chaurasia, and Saurabh Pal, 2013, Early Prediction of Heart Diseases Using DataMining Techniques, Caribbean Journal of Science and Technology, ISSN: 0799-3757, Vol.1, pp. 208-218.

8. Parameshachari BD et. al Epileptic Seizure Detection Using Machine Learning, 1st International Conference on Emerging Trends in Engineering, Innovative Science and Management (ICETEISM-2019), 2019.