IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9737

A SYSTEMATIC APPROACH FOR HEART DISEASE PREDICTION USING PRINCIPAL COMPONENT ANALYSIS ALGORITHMS

Janani S[1], Marikkannan M[2]

M.E Computer Science and Engineering, Government College of Engineering,

Vasavi College Post, Erode, Tamilnadu - 638316, India.[1]

Assistant Professor (Senior), Department of Computer Science and Engineering,

Government College of Engineering, Erode, Tamilnadu - 638 316, India.[2]

Abstract: Data collection mechanism is planned and correlation analysis of those gathered data. Exploratory data analysis is data analysis techniques that can be visually address the information installed in dataset. All the recreations are performed in Spyder (Python 3.9) and other obligatory libraries framework to process the patient data. The risk factors that cause disease are through of and predict and predict using some machine learning calculations. At last, it is intended to predict the future ailment of the most heart patients in light of their ongoing wellbeing status. The most effective way to forestall such clinical blunder is by diminishing the dependability of memory and by further developing the data access. The wellbeing related solo information is utilized for finding stowed away elements that might demonstrate a sickness state in patients. To foresee coronary illness, AI calculation is utilized alongside information examination and perception instrument. In result section, with this dataset and EDA approach clearly shows that prediction is accurate for heart disease patients.

Keyword: Exploratory Data Analysis, PCA Algorithm.

INTRODUCTION

OVERVIEW

Medical industries produce huge amount of data, so it's called big data that accommodates hidden knowledge for decision making. The large volume of data is used to make decision which is more accurate than instinct. Exploratory Data Analysis (EDA) finds mistakes, detects appropriate data, checks assumptions and desides the correlation among the explanatory variables[8]. Analytics is an fundamental technique for any profession as it estimate the future and hidden pattern. Data analytics is considered as a practical technology in the recent past and it plays an fundamental role in medical industries which includes emergency situations and outbreaks of disease.

The use of analytics in healthcare improves care by working with preventive care and EDA is a fundamental step while analysing data. In this paper, the heart disease is considered and predicted using PCA algorithm and the analysis is carried out using a dataset for heart disease[1].

2.1 PROBLEM STATEMENT

The initial deployment of an instantiation of a Data based on Health Analytics as a Service model. The model is used to designed foresee the future health condition of the most heart patients based on their health status. To support the critical care unit and healthcare in hospital management by health analytics a service.

2.2 OBJECTIVE

Critical care units internationally provide acute care for patients in critical conditions involve interdisciplinary teams of healthcare workers.

Solutions that enable the real-time effective use of this healthcare organization.

Heart disease can be predicted based on various symptoms such as age, gender, pulse rate etc.

Data analysis in healthcare assists in predicting diseases, improving diagnosis, analyzing symptoms, providing appropriate medicines, improving the quality of care, minimizing cost, extending the life span and reduces the death rate of heart patients.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9737

3. DATA ANALYTICS

It is used to examining the data to uncover patterns for extract valuable from it. The exploring of data analysis and evaluate the positive and negative patient related heart disease. Data analysis in healthcare helps in predicting diseases, improving diagnosis, examining symptoms, giving suitable medicines, improving the quality of care, limiting cost, expanding the life span and reduces the death rate of heart patients. Exploratory Data Analysis (EDA) is classified into Graphical or non-graphical and Univariate or multivariate Univariate data consider one data column at a time while multivariate method considers more than two variables while analyzing. The figure 3.1 show the working flow of EDA process for health analytics[10].

Distribution of the data: Whether the data is normally distributed or it its skewed (to the left or right). To identify the outliers- extremely low or high values that do not fail near any other data points.

Input: Dataset with selected feature

Techniques used: Correlation matrix, PCA algorithm is used to evaluate the values

Output: Identification and evaluate the data.

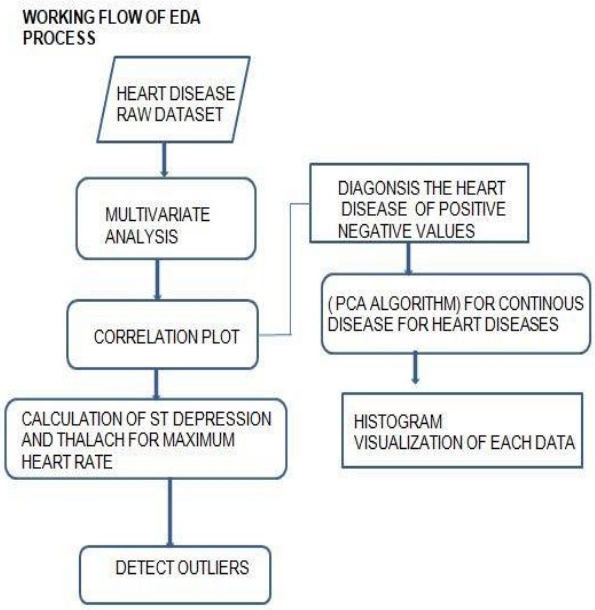

Figure 3.1 Working flow of EDA process for health analytics.

3.1 DATA ANALYTICS (CORRELATION MATRIX)

A correlation matrix is used to calculate the positive and negative variables in the given dataset.

The measure is best used in variables that demonstrate a linear relationship between each other.

Pearson correlation matrix = $\frac{\text{cov}(x, y)}{(\text{Std } V(x) * \text{Std } V(y))}$ ------(1)

From equation (1) shows that the correlation coefficient returns between -1 and 1 represent the limits of correlation from a full negative correlation to full positive correlation. A value zero means no correlation. The value must be interpreted, where often a value below -0.5 or above 0.5. Scipy function can be used to calculate the Pearson correlation coefficient between two data sample with same length.

3.2 PCA ALGORITHM

Principal component analysis is a unsupervised learning algorithm. In predictive modelling PCA is specific useful as a data pre-processing method. PCA serves as a tool for exploratory data analysis and outlier discovery, but also for dimensionality reduction when the number of factors outnumbers the sample size (d>n). Past that PCA is often applied on data sets with highly repetitive variables, or in other words of highly correlated variables.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9737

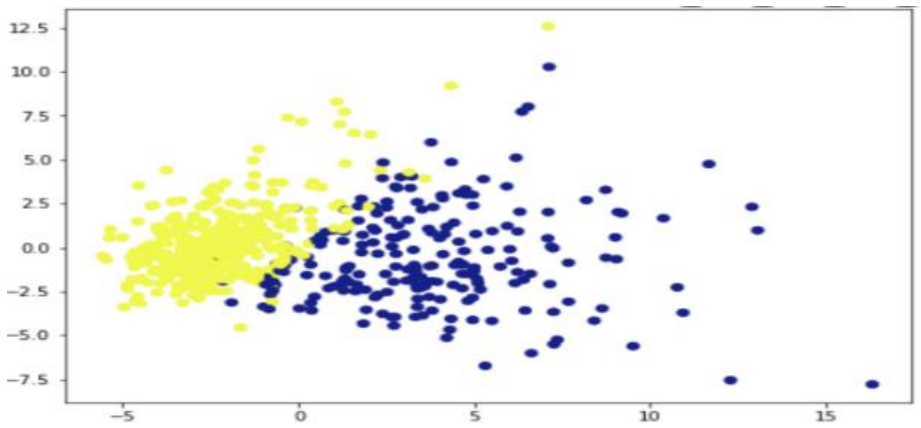

Figure : 3.2 PCA Algorithm

4. RESULTS DISCUSSION AND PERFORMANCE ANALYSIS

The proposed work is implemented in Python 3.9 with libraries sklearn.metrics, Spyder, pandas, matplotlib, scipy.stats, seaborn and other mandatory libraries. Machine learning algorithm is used.

4.1. RESULT DISCUSSION:

Step 1: The file path is read for the given health dataset

Step 2: After the file read next data preparation is process is taken. Here the data is showing the missing values in the given dataset

Step 3: The figure 4.1 shows the correlation matrix for positive and negative patients

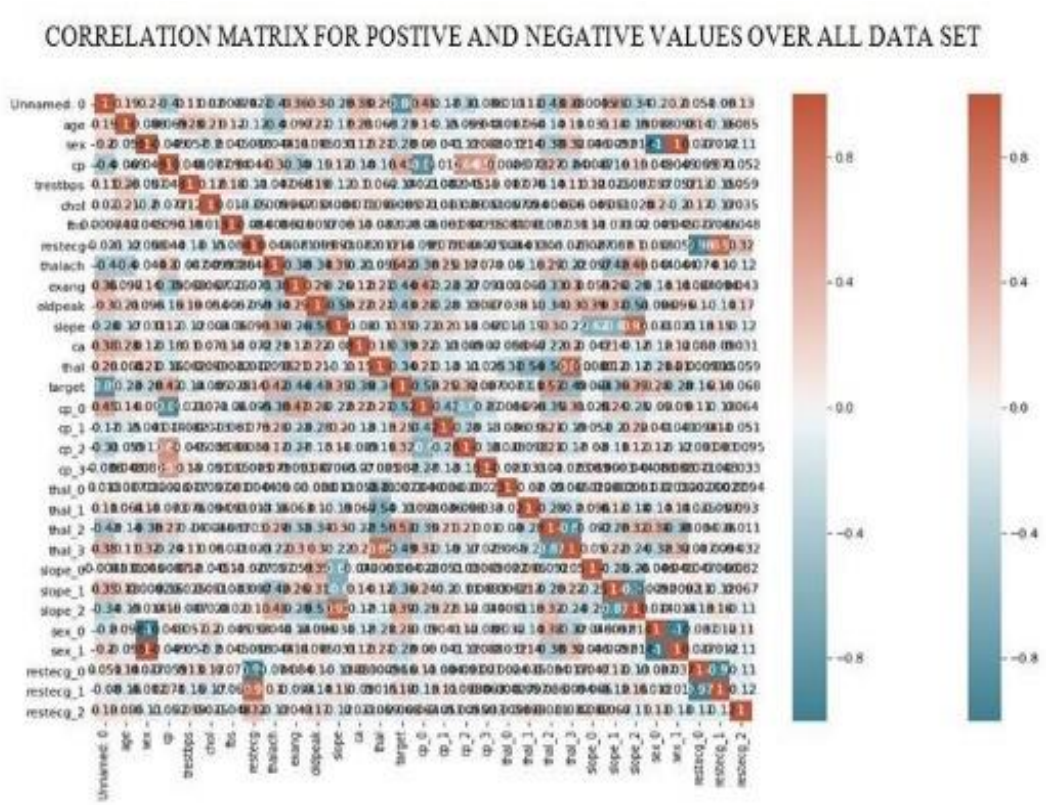

Figure 4.1 Correlation matrix

Step 4: After the correlation, next is to finding the maximum heart rate disease in dataset for positive and negative patients. The figure 4.2 shows the ST depression (segment) slope of heart disease patients range like upslope, down slope and flat in visualization

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9737

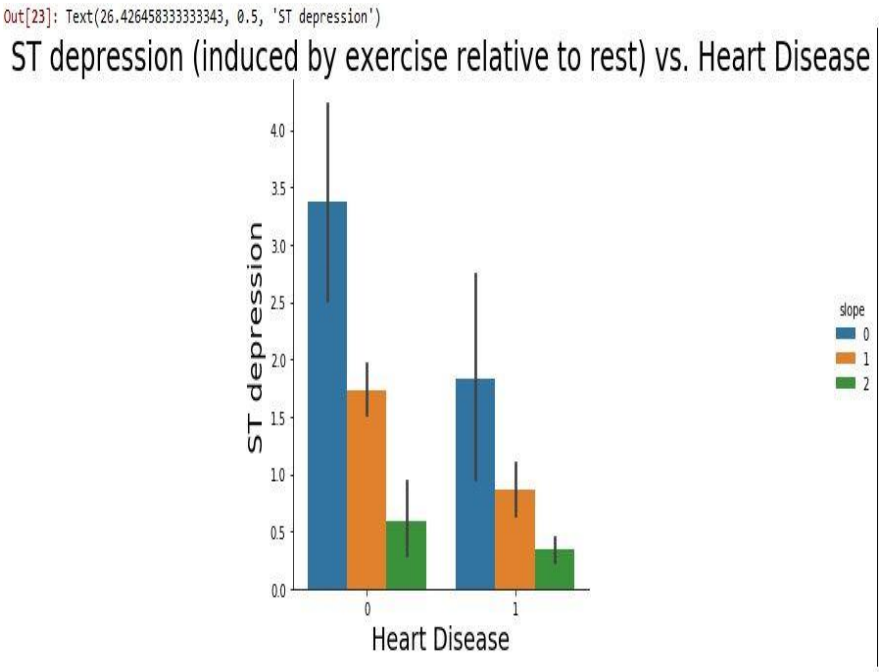

Figure 4.2 ST depression (segment) range for heart disease

Step 5: In given health dataset , Finding the major heart disease like thalach and ST depression by their values in dataset .The ST depression values for both positive and negative range patients. Similarly the thalach values for both positive and negative range patients.

Step 6: After finding the values for major disease then filtering the whole positive and negative patient in over all dataset.

Step 7: Next process for getting more accurate and statistical information for number of patients having with heart disease and number of patients without heart disease.

Step 8: Using PCA algorithm calculate the heart disease using target values 0 and 1. The target indicates 0 means the patients are not having chance of heart disease The target indicates 1 means the patients are having chance of heart disease .The figure shows 4.3 represent in data visualizing technique .

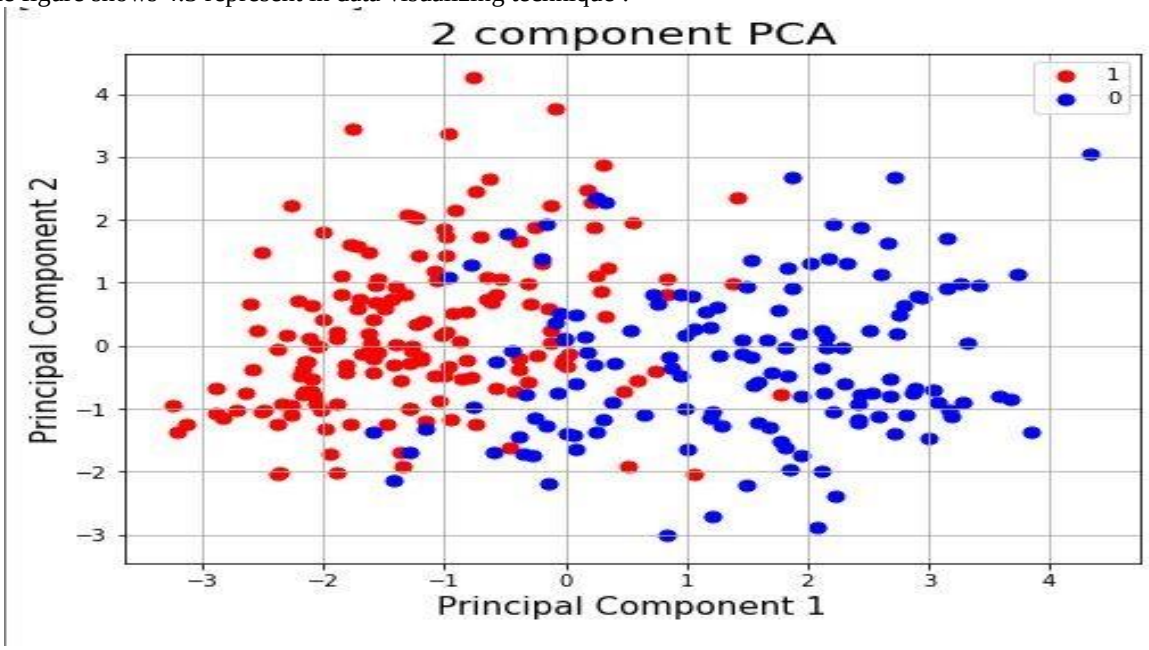

Figure 4.3PCA Components for positive and negative

Step 8: Here the cloud is used to store the data in server as shown in figure 4.4

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 7, July 2022

DOI: 10.17148/IARJSET.2022.9737

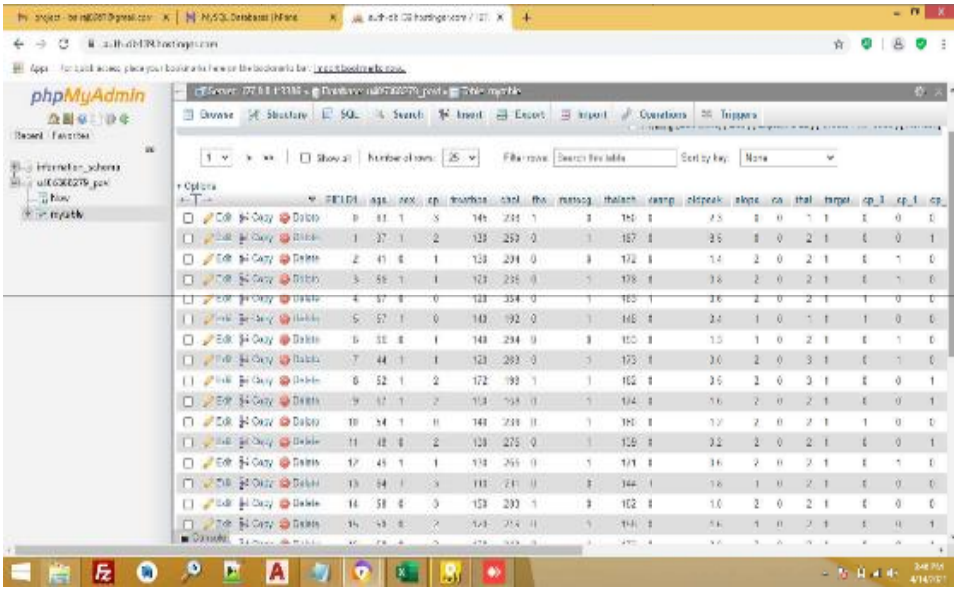

Figure 4.4 Cloud store the data in server

5. CONCLUSION

Heart stroke and vascular disease are the major cause of disability and premature death. Chest pain is the key to perceive the heart disease. In this work heart disease are predicted by considering major factors with four types of chest pain.

The exploratory data analysis in heart disease dataset helped for visualizing various aspect needed to determine whether the patient having heart disease or not.

6. REFERENCES

[1] Prasan Kumar Sahoo and Suvendu Kumar Mohapatra. "Analyzing Healthcare Big Data With Prediction for Future Health Condition", IEEE ACCESS., vol.4, pp.182-243.-2017.

[2] Likewin Thomas and Annappa, Manoj Kumar. "A Healthcare management using clinical decision support system", Institute of Electrical and Electronics Engineers IEEE ACESS,vol.6, pp.459-597- 2018.

[3] Carolyn McGregor and Catherine Inibhunu, Jonah Glas ."Health Analytics as a Service with Artemis Cloud: Service Availability" vol. 154, no. 10, pp. 28–31, 2018,.

[4] Chaitanya Kaul and Ashmin Kaul ,Saurav Verma Mukesh Pate "Comparative Study on Healthcare Prediction systems using Big Data", IEEE Institute of Electrical and Electronics Engineers. Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems IEEE Access, vol. 3, pp. 675–692, 2017.

[5] K. Lin and F. Xia, W. Wang, D. "System design for big data application in emotion-aware healthcare," IEEE Access, vol. 4, pp. 6901–6909, 2016.

[6] L. A. Towable and S. M. R. Islam, D. Kwan, The internet of things for healthcare: A comprehensive survey, IEEE Access, vol. 3, pp. 678–708, 2015.

[7] S. Wang, and X. Chang, X. Li, G. Long Diagnosis code assignment using sparsely -based disease correlation embedding," IEEE Institute of Electrical and Electronics Engineers, IEEE Access.,vol12 ,pp 456-523,2016.

[8] R.Indrakumari, T.poongodi,Soumya Ranjan jena , "Heart Disease Prediction using Exploratory Data Analysis" International Conference on Smart Sustainable Intelligent Computing and Applications under IEEE Access.,vol12 ,pp 452-513,2020

[9] S. Ram and K. Baby and A. Ravikumar, "Big Data : An Ultimate Solution in Health Care," vol. 106, no. 10, pp. 28–31, 2014

[10] Sabeel Ashfaq khan and Senthil velan .S , "Application of Exploratory data analysis to generate inferences on the occurrence of breast cancer using simple dataset" vol. 19, no. 4, pp. 1193–1208, 2020

