VISUALIZING AND PREDICTING HEART DISEASES WITH AN INTERACTIVE DASHBOARD

Abstract:

Improving the precision of heart diseases detection has been investigated by many researchers in the literature. Such improvement induced by the overwhelming health care expenditures and erroneous diagnosis. As a result, various methodologies have been proposed to analyse the disease factors aiming to decrease the physicians practice variation and reduce medical costs and errors. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of clinical data analysis. The amount of data in the healthcare industry is huge. Data mining turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions.

Heart Disease Prediction using Exploratory Data Analysis

Healthcare industries generate enormous amount of data, so called big data that accommodates hidden knowledge or pattern for decision making. The huge volume of data is used to make decision which is more accurate than intuition. Exploratory Data Analysis (EDA) detects mistakes, finds appropriate data, checks assumptions and determines the correlation among the explanatory variables. In the context, EDA is considered as analysing data that excludes inferences and statistical modelling. Analytics is an essential technique for any profession as it forecast the future and hidden pattern. Data analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare which includes new research findings, emergency situations and outbreaks of disease. The use of analytics in healthcare improves care by facilitating preventive care and EDA is a vital step while analysing data. In this paper, the risk factors that causes heart disease is considered and predicted using K-means algorithm and the analysis is carried out using a 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. To predict the heart disease, K-means clustering algorithm is used along with data analytics and visualization tool. The paper discusses the pre-processing methods, classifier performances and evaluation metrics. In the result section, the visualized data shows that the prediction is accurate.

The experiment was carried out on a publicly available database for heart disease. The dataset contains a total of 303 records that were divided into two sets, training set (40%) and testing set (60%). A data mining tool named Weka 3.6.11 was used for the experiment. Additionally,

multilayer perceptron neural network (MLPNN) with backpropagation (BP) was used as the training algorithm.

Effective heart disease prediction system using data mining techniques

The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some advanced data mining techniques are used. In this study, an effective heart disease prediction system (EHDPS) is developed using neural network for predicting the risk level of heart disease. The system uses 15 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction. The EHDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge, eg, relationships between medical factors related to heart disease and patterns, to be established. We have employed the multilayer perceptron neural network with backpropagation as the training algorithm. The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases.

In order to predict the probability of patients having heart disease, a confusion matrix was created, where A denotes patients with heart disease, and B denotes patients with no heart disease.The confusion matrix contains the following four entries:

*TP (true positive): The number of records classified as true while they were actually true.
*FP (false positive): The number of records classified as true while they were actually false.
*FN (false negative): The number of records classified as false while they were actually true.
*TN (true negative): The number of records classified as false while they were actually false.

The overall process of effective heart disease prediction system (EHDPS) is based on the following three steps:

1. Data collection
2. Data pre-processing and
3. The classification of data.

The data are collected from a standard dataset that contains 303 records. The 15 parameters, such as age, sex, chest pain type (CP), and cholesterol (chol), with some domain values associated with them, considered to predict the probability of heart disease are shown.

In this study, an EHDPS has been presented using data mining techniques. From ANN, an MLPNN together with BP algorithm is used to develop the system. The MLPNN model proves the better results and assists the domain experts and even the person related to the medical field to plan for a better and early diagnosis for the patient. This system performs realistically well even without retraining. Furthermore, the experimental results show that the system predicts heart disease with ~100% accuracy by using neural networks.

Predicting Heart Diseases In Logistic Regression Of Machine Learning Algorithms

Machine learning is one of the trending technologies which used in many spheres around the world including healthcare industry for predicting diseases. The aim of this study is to identify the most significant predicators of heart diseases and predicting the overall risks by using logistic regression. Thus, binary logistic
model which is one of the classification algorithms in machine learning is used in this study to identify the predicators. Further, data analysis is carried out in Python using JupyterLab in order to validate the logistic regression.

The dataset which used for the logistic regression analysis is available on the Kaggle website(https://www.kaggle.com), from an ongoing cardiovascular study of Framingham, Massachusetts. The classification goal of this study is to
predict whether the patient has 10-year risk of future heart diseases. The Framingham dataset consists with 4238 records of patients'dataand15 attributes. The data analysis is carried out in Python programming by using JupyterLab which is more flexible and powerful data science application software.

The aim of this study is to evaluate the risk of 10-year CHD using 14 IVs. The attributes are selected after the backward elimination process considering the P values which are lower than 5%. Therefore, the logistic regression model is derived through P values of the variables <0.05 (sex, age,cigsPerDay, totChol, sysBP, glucose). According to the logistic regression outcome, men are more suspectable to heart disease than women. Age, number of cigarettes per day and systolic blood pressure are the odds of CHD. However, There is no significance change in the total cholesterol level and the glucose level. But, the level of
glucose has a negligible change in odds. The model is more specific than sensitive. Further, the accuracy of the model is 0.87. The value under the ROC curve is 73.5 which is somewhat satisfactory. Moreover, the model could be improved by using more data

Evolutionary algorithm-based convolutional neural network for predicting heart diseases

Convolutional neural networks (CNNs) have been commonly used in medical decision support systems to predict and diagnose different diseases with good precision. CNNs are extremely successful in developing health support systems because of their ability to identify relationships and hidden patterns in healthcare data. One of the most important and useful applications of such systems is in the prediction of heart diseases by observing cardiac anomalies. Fundamentally, CNNs have multiple hyperparameters and various specific architectures, which are costly and impose challenges in selecting the best value among possible hyperparameters. In addition, CNNs are sensitive to their hyperparameter values which have a significant impact on the efficiency and behavior of CNN architectures. Thus, selecting the right set of parameters is of particular concern among practitioners. Consequently, this paper proposes a CNN-jSO approach for the prediction of heart (cardiac) diseases, in which the jSO optimization algorithm is employed to tune those CNN hyperparameters. The performance of the designed system is tested on the PhysioNet heart sound and Kaggle heartbeat sounds datasets. The proposed CNN-jSO is compared with other algorithms and shown to be better than them. The CNN-jSO system was implemented in Python and yielded 97.76% training accuracy and 94.12% testing accuracy.

References:

- V. Manikantan & S.Latha,"Predicting the Analysis of Heart Disease Symptoms Using Medicinal Data Mining Methods", International Journal on Advanced Computer Theory and Engineering, Volume-2, Issue-2, pp.5-10, 2013.

- Dr.A.V.Senthil Kumar, "Heart Disease Prediction Using Data Mining preprocessing and Hierarchical Clustering", International Journal of Advanced Trends in Computer Science and Engineering, Volume-4, No.6, pp.07-18, 2015.

- Uma.K, M.Hanumathappa, "Heart Disease Prediction Using Classification Techniques with Feature Selection Method", Adarsh Journal of Information Technology, Volume-5, Issue-2, pp.22-29, 2016 Himanshu Sharma, M.A.Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms:A Survey",International Journal on Recent and Innovation Trends in Computing and Communication,Volume5,Issue-8,pp.99-104, 2017.

- S.Suguna, Sakthi Sakunthala.N ,S.Sanjana, S.S.Sanjhana, "A Survey on Prediction of Heart Disease using Big data Algorithms", International Journal of Advanced Research in Computer Engineering & Technology,Volume-6,Issue-3,pp.371-378,2017.

- A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," Nature Reviews Cardiology, vol. 8, no. 1, pp. 30–41, 2011.

- J.Mourão-Miranda,A.L.W.Bokde,C.Born,H.Hampel,and M. Stetter, “Classifying brain states and determining the discriminating activation patterns:supportvectormachineon functionalMRIdata,”NeuroImage,vol.28,no.4,pp.980–995, 2005.
- H. Ahmed, E.M. Younis, A. Hendawi, A.A. Ali

  Heart disease identification from patients social posts, machine learning solution

- F. Ali, S. El-Sappagh, S.R. Islam, D. Kwak, A. Ali, M. Imran, K.-S. Kwak

  A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion.

- Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. Int J Comput Sci Net Secur. 2008;8:343–350. [Google Scholar]
- Sayad AT, Halkarnikar PP. Diagnosis of heart disease using neural network approach. Int J Adv Sci Eng Technol. 2014;2:88–92. [Google Scholar]
- Gudadhe M, Wankhade K, Dongre S. Decision support system for heart disease based on support vector machine and Artificial Neural Network. Computer and Communication Technology (ICCCT), 2010 International Conference on; 2010. pp. 741–745. [Google Scholar]
- Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating error. Nature. 1986;323:533–536. [Google Scholar]