

Published studies that used this dataset (heart.csv):

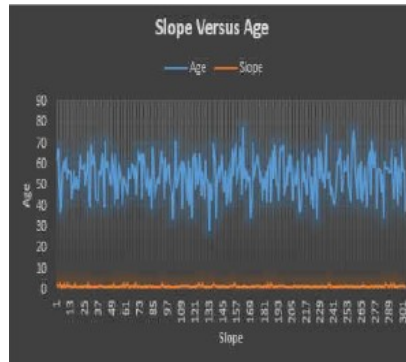
1st study

DEMONSTRATING ELECTROCARDIOGRAM (ECG) SIGNAL PROCESSING TO SHOW NONLINEAR ABNORMALITIES OF THE HEART

Date: 02/2019

Author: Eva Obinna

From Ignatius Ajuru University of Education, Iwofe, Rivers State, Nigeria



Conclusion

The heart is an important and sensitive organ of the human body and should be prevented from any form of hazardous agent. Every individual should do a routine check of their body system so as to know the health status of their body. That way they can also be able to know the condition of their heart from time to time. The ECG is an important signal among all bioelectrical signals used in the diagnosis of many cardiac disorders and can be recorded from the wave passage of the depolarization and repolarization processes in the heart. The voltage in the heart tissues is conducted to the body surface where it is measured using electrodes. This work demonstrated an ECG Signal Processing of an abnormal heart Using Microsoft Excel. In future this demonstration can be done using Mat Lab or Python programming language. Analysis can also be conducted using any of the machine language model

2nd study

HEART HEALTH PREDICTION IN PARTIAL FULFILMENT FOR THE AWARD OF THE DEGREE OF BACHELOR OF TECHNOLOGY

Date:05/2020

Author: Shanu Kumar

From Galgotias University



Conclusion

They used this type of graphic to show heart disease frequency for sex and we used pie graphics but it's all the same.

This project discusses the various machine learning algorithms such as Naïve Bayes, decision tree and k- nearest neighbour which were applied to the data set(heart.csv). It utilizes the data such as cholesterol, diabetes, bp and then tries to predict the possible coronary heart disease patient in next 10 years. Family history of heart disease can also be a reason for developing a heart disease as mentioned earlier. So, this data of the patient can also be included for further increasing the accuracy of the model. This work will be useful in identifying the possible patients who may suffer from heart disease in the next 10 years. This may help in taking preventive measures and hence try to avoid the possibility of heart disease for the patient. So when a patient is predicted as positive for heart disease, then the medical data for the patient can be closely analysed by the doctors. An example would be – suppose the patient has diabetes which may be the cause for heart disease in future and then the patient can be given treatment to have diabetes in control which in turn may prevent the heart disease.

3rd study

PERFORMANCE EVALUATION OF CLASSIFICATION METHODS FOR PREDICTING HEART DISEASE

Date: 29/06/2021

Author: S.Pononguzhali, P.sujatha, P.Sripriya, V.Deepa and K.Mahalakshmi

From: EasyChair Preprint

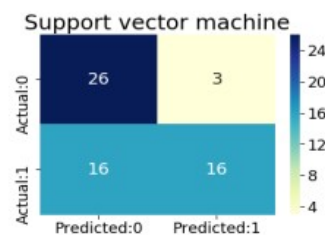


Figure 4 Confusion matrix of SVM

Figure 4 describes that, from 61 testing records, the decision tree classifier predicted 19 times “Yes” and predicted “No” 42 times. Twenty Six people were correctly predicted as “No” and they do not have the heart disease in reality. Sixteen people were correctly predicted as “Yes” and also they have heart disease actually. Sixteen people were incorrectly predicted as “No” even if they have heart problem. Three people were incorrectly predicted as “Yes” though they do not have heart problem.

The following table 2 & Figure 4 evident that decision tree predicts the disease with maximum accuracy of 83.6% compared to SVM that provides the accuracy of 68.8%.

Conclusion

Heart disease covers any kind of disorder of the heart. It emerged as foremost source of death. Heart disease prediction and prevention is necessary. Data mining algorithms help in predicting the heart diseases, and predictions made by algorithms are quite accurate. Decision Tree and Support vector machine are employed on heart disease dataset from kaggle website. The classification algorithms are implemented using python. The decision tree algorithm achieved the maximum accuracy of 83.6%. Comparing with decision tree, SVM produces lowest accuracy of 68.85%. From the experimental results we can conclude that decision tree algorithm performs better when compared to SVM. Our future work is to combine classification algorithm to develop hybrid model to predict the heart ailment.

4th study

ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION AND PREDICTION OF HEART DISEASE

Date: 19/09/2021

Author: Nataliya Boyko, Iryna Dosiak

From: Lviv Polytechnic National University, Profesorska Street 1, Lviv, 79013, Ukraine



Conclusion

Like us, they already used the correlation matrix but with different colors.

The paper considered the relevance of the topic: the use of data mining methods for diagnosing the disease in a patient on a set of indicators, such as symptoms, test results, and other indicators. We used the Heart data set for the study, which we cleared of emissions, Null values, and normalized. We also performed a search and analysis of significant features and patterns between different factors influencing heart disease. In addition, we used two algorithms in this work, which objectively showed the classification results on the selected dataset. The parameters used for the analysis were the selection and deletion of the function. We first tested a classifier with all the features and then gradually reduced the set to determine which algorithm best classifies with fewer features. The simulation results show that the Multinomial Naive Bayes classifier has better accuracy than the Gaussian method with the same data set and parameters. In addition, it reduces training time, which is very important because the annual growth of data in medicine is increasing very rapidly. In future work, it is worth considering two aspects. Namely, we can compare more algorithms to achieve better results and potentially introduce a better algorithm in Naive Bayes. Moreover, we can try to evaluate the effectiveness of their work to justify their use in the health care system.