International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2019): 7.583

Naive Buyers Theorem for Heart Disease Prediction System

Indu Kumari[1], Dinesh Kumar Bahgel[2]

[1]Professor, School of Computing Science and Engineering, Galgotias University, Uttar Pradesh 203201, India

[2]UG Scholar, School of Computing Science and Engineering, Galgotias University, Uttar Pradesh 203201, India

Abstract: ***The health care industry produces a huge amount of data. This data is not always made use to the full extent and is often underutilized. Using this huge amount of data, a disease can be detected, predicted or even cured. A huge threat to human kind is caused by diseases like heart disease, cancer, tumor and Alzheimer's disease. In this paper, we try to concentrate on heart disease prediction. Using machine learning techniques, the heart disease can be predicted. The medical data such as Blood pressure, hypertension, diabetes, cigarette smoked per day and so on is taken as input and then these features are modelled for prediction. This model can then be used to predict future medical data. The algorithms like K- nearest neighbour, Naïve Bayes, support vector machine and decision tree are used. The accuracy of the model using each of the algorithms is calculated. Then the one with a good accuracy is taken as the model for predicting the heart disease.***

1. Introduction

In our day to day life, people are undergoing a routine and busy schedule which leads to stress and anxiety. In addition to this, the percentage of people who are obese and addicted to cigarette goes up drastically. This leads to diseases like heart disease, cancer, etc. The challenge behind these diseases is its prediction. Each person has different values of pulse rate and blood pressure. But medically proven, the pulse rate must be 60 to 100beats per minute and the blood pressure must be in the range of 120/80 to 140/90. Heart disease is one of the major cause of death in the world. [11] The number of people affected by heart disease increases irrespective of age in both men and to heart disease. They are smoking which raises the risk of developing heart disease, high blood pressure makes the heart work harder to pump blood and it can strain heart and damage blood vessels, abnormal cholesterol levels also contribute to heart disease and obesity. Also, family history of heart disease can be a cause of having heart disease. But this history is not considered in this paper for prediction of heart disease. The other risk factors include age, gender, stress and unhealthy diet. Chance of having a heart disease increases when a person is getting older. Men have a greater risk of heart disease. However, women also have the same risk after menopause. Leading a stressed life can also damage the arteries and increase the chance of coronary heart disease. So, in this paper based on the factors mentioned above we try to predict the risk of heart disease. A large amount of work has been done related to heart prediction system by using various technique sand algorithms by many authors. These techniques may be based on deep-learning, machine-learning, data mining and so on. The aim of all those papers is to achieve better accuracy and to make the system more efficient so that it can predict the chances of heart attack.

2. Purposed Model

In this paper, comparison of various machine learning methods is done for predicting the 10 year risk of coronary heart disease of the patients from the medical data.

In this project, we propose a system which can be used for both heart disease monitoring and diagnosis. In this project, the proposed system can notify if an emergency situation occurs. The pulse rate sensor (AMPED) and Bluetooth 4.0 fitted to it and sends the data to the mobile application. Now the sensor data is pushed to the cloud for further analyzing of Heart Rate Variability (HRV). The graph of the heart rate and the prediction of heart disease can be seen through the mobile app. Based on the increase or decrease of the HRV the model will be predicting the chances of the disease to happen and the notification will be sent to the user's phone.

3. Implementation

The aim of this project is to find the capabilities of machine learning tools available in the market, to come up with the advantages and disadvantages of each tool based on an experimental analysis with data sources available from the open source network. In this project, building predictive data model in the machine learning tool will be identified from major IT giants like Google, Microsoft and Oracle. Check the level of easiness offered to both the developers and as well to the end users (business users). The advantages and disadvantages of each of these tools will also be explored in terms of different features, different traditional ML algorithms and capability of integrating programming languages like R and Python. With multiple tools and technologies related to machine learning accessible in the market it is always a challenge to the developer or the architect team in IT industry to decide on the most efficient one. The best tool helps in building modules at high degree of accuracy in multiple domains (marketing, sales, HR, logistics, insurance, life sciences). On the other hand, the tools which are available in the market for machine learning will in turn help the business owner to use them with much ease without having a total

dependency on the development tea. The following are steps to execute this project.

Tools installation and configuration:
1) DVD-Oracle
2) Azure ML-Power BI
3) Google Analytics
   - Dataset exploration and explanation (EDA)
   - ML model building.
   - Tools performance evaluation.

4. Result

The *machine learning models is evaluated using thaub-ROC metric. This can be used to understand the model performance.*

5. Conclusion

This paper discusses the various machine learning algorithms such as support vector machine, Naïve Bayes, decision tree and k- nearest neighbour which were applied to the data set. It utilizes the data such as blood pressure, cholesterol, diabetes 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 intern may prevent the heart disease.

References

[1] https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset
[2] Monika Gandhi, Shailendra Narayanan SinghPredictions in heart disease using techniques ofdata mining
[3] *Benjamin EJ et.al Heart Disease and StrokeStatistics 2018 At-a-Glance*