HEART DISEASE PREDICTION WITH INTERACTIVE DASHBOARD

PROPOSED SOLUTION

TEAMID:PNT2022TMID44038

Faculty mentor:

Dhrisya.S

Teamleader:Roshini.K

Teammember:Sinthiya.M

Teammember:Sugantharani.F

Teammember:Swetha.K

PROPOSED SOLUTION:

NOVELTY:

The result of the data analysis to identify the necessary hidden patterns for predicting heart diseases are presented in this section. Here the variables considered to predict the heart disease are age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain and exercise angina. The heart disease dataset is effectively pre-processed by eliminating unrelated records and given values to missing tuples. The pre-processed heart disease data set [10] is then composed by K-means algorithm. Here, four types of heart diseases are discussed namely asymptomatic pain, atypical angina pain, non-anginal pain and non-anginal pain. The results are computed using all the four types of chest pain with other deciding variables.

FEASIBILITY OF IDEA:

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.

BUSINESS MODEL:

Predicting and diagnosing heart disease is the biggest challenge in the medical industry and it is based on factors like physical examination, symptoms and signs of the patient . Factors which influence heart diseases are cholesterol level of the body, smoking habit, and obesity, family history of diseases, blood pressure and working environment. Machine learning algorithms play a vital and accurate role in predicting heart disease . The advancement of technologies allows machine language to pair with big data tools to handle unstructured and exponentially growing data . In the paper, K means clustering method is proposed in big data environment and the visualization is made with the tableau dashboard.

SOCIAL IMPACT:

Recent advances in molecular genetics are making it increasingly feasible to construct individual genetic profiles predicting susceptibility to heart disease, cancer and respiratory disorders. This paper reviews current knowledge about the social and cultural impact of providing people with

information relating to their risk for future disease, focusing not only on currently available genetic testing but also on hypertension, hyperlipidaemia and cancer screening. We highlight the importance of issues of probability and uncertainty, and the tension between collective and individual goals in the assessment of medical risk. We conclude with a proposed research agenda for studies of the social and cultural impact of predictive genetic testing, and argue that there is a pressing need for rigorous, empirical, social research in this area.

SCALABILITY OF THE SOLUTION:

As wearable medical sensors continuously generate enormous data, it is difficult to process and analyse. This paper focuses on developing scalable sensor data processing architecture in cloud computing to store and process body sensor data for healthcare applications. Proposed architecture uses big data technologies such as Apache Flume, Apache Pig and Apache HBase to collect and store huge sensor data in the Amazon web service. Apache Mahout implementation of MapReduce-based online stochastic gradient descent algorithm is used in the logistic regression to develop the scalable diagnosis model. Cleveland heart disease database (CHDD) is used to train the logistic regression model. Wearable body sensors are used to get the blood pressure, blood sugar level and heart rate of the patient to predict the heart disease status. Proposed prediction model efficiently classifies the heart disease with the accuracy of training and validation sample is 81.99% and 81.52%, respectively.