VISUALIZING AND PREDICTING HEART DIESEASES WITH AN   INTERACTIVE DASHBOARD

TEAM ID: PNT2022TMID52562

LITERATURE SURVEY

Team member : Dhakshinya B G

Team member : Dharani S

Team member : Janani K

Team member : Sruthi S

FACULTY MENTOR:

Senthilvel A N

Literature paper-1:

Heart disease prediction using machine learning algorithms

**AUTHORS:** Harshit Jindal ,Sarthak Agrawal, Rishabh Khera, Rachna Jain,

Preeti Nagrath

Day by day the cases of heart diseases are increasing at a rapid rate and it's very Important and concerning to predict any such diseases beforehand. The strength of the proposed model was quiet satisfying and was able to predict evidence of having a heart disease in a particular individual by using KNN and Logistic Regression which showed a good accuracy in comparison to the previously used classifier such as naive bayes etc.

Machine Learning is a very vast and diverse field and its scope and implementation is increasing day by day. Cardiovascular diseases are very common these days, they describe a range of conditions that could affect your heart.

An efficient Cardiovascular disease prediction has been made by using various algorithms some of them include Logistic Regression, KNN, Random Forest Classifier. The supervised network has been advised for diagnosis of heart diseases .

This paper shows the analysis of various machine learning algorithms, the algorithms that are used in this paper are K nearest neighbors (KNN), Logistic Regression and Random Forest Classifiers which can be helpful for practitioners or medical analysts for accurately diagnose Heart Disease. classifier is used to classify the pre-processed data the classifier used in the proposed model are KNN, Logistic Regression, Random Forest Classifier.
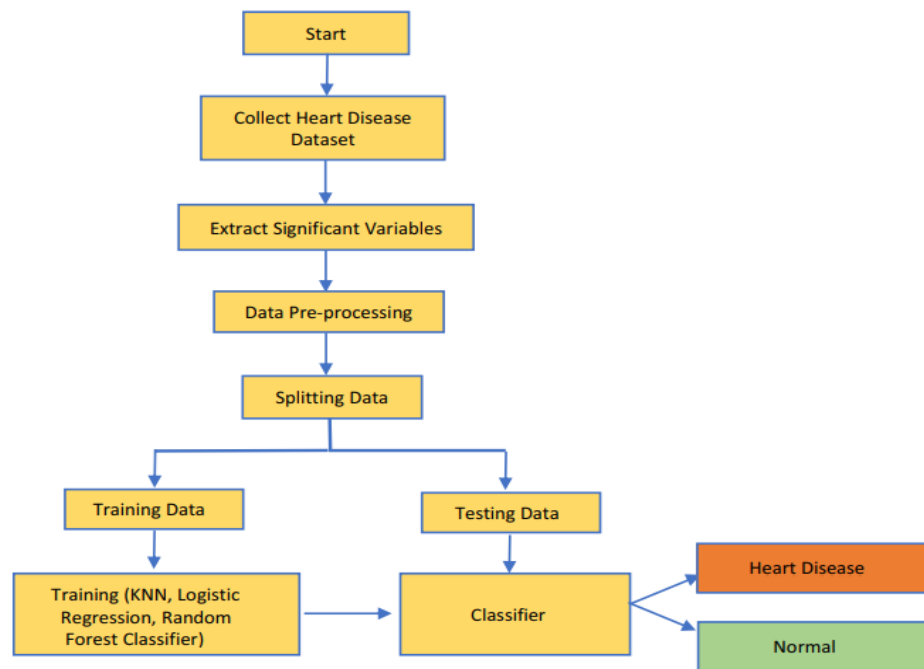
1.1 This diagram have been developed using different classifier. This model uses 13 medical parameters such as chest pain, fasting sugar, blood pressure, cholesterol, age, sex etc. for prediction.

The algorithms that we used are more accurate, saves a lot of money i.e. it is cost efficient and faster than the algorithms that the previous researchers used. Moreover, the maximum accuracy obtained by KNN and Logistic Regression are equal to 88.5% which is greater or almost equal to accuracies obtained from previous researches.

Therefore, in conclusion this project helps us predict the patients who are diagnosed with heart diseases by cleaning the dataset and applying logistic regression and KNN to get an accuracy of an average of 87.5% on our model which is better than the previous models having an accuracy of 85%.
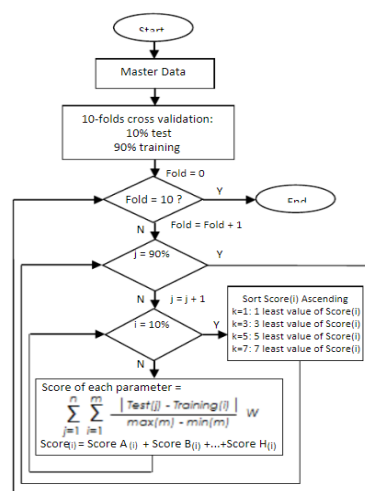
Literature paper-2:

Heart Disease Prediction System using k-Neareast NeighborAlgorithm with   Simplified Patient's Health Parameters

Authors: I Ketut Agung Enriko, Muhammad Suryanegara, Dadang Gunawan

Year of publication : 2016

Heart disease is the primary cause of death nowadays. Treatments of heart disease patients have been advanced, for example with machine-to-machine (M2M) technology to enable remote patient monitoring. These years, data mining has become popular in many fields of industry, thanks to its purpose to convert large to become valuable information . Medical data contains the huge volume of undiscovered data. This data may contain redundant, noisy and irrelevant data. Redundant data may cause classifier to produce less accurate results. PSO as a feature selection measure discards redundant features to improve the accuracy of the classifier. Our proposed method effectively identifies the redundant features compared to other existing features to effectively predict the heart disease. Prediction of heart disease using neural network was proposed by Dangare et al. Feature selection is used to predict the disease. Their method obtained an accuracy of 92.5% for 13 features and 100% accuracy with 15 features. There is a 7.5% improvement after discarding 2 features from 15 to 13.
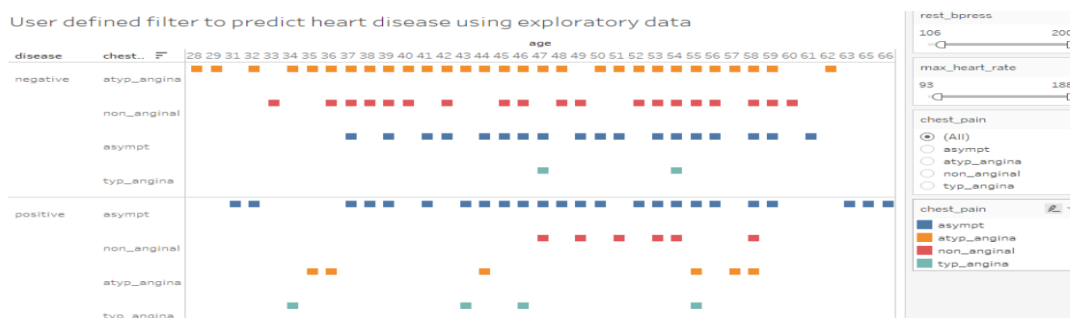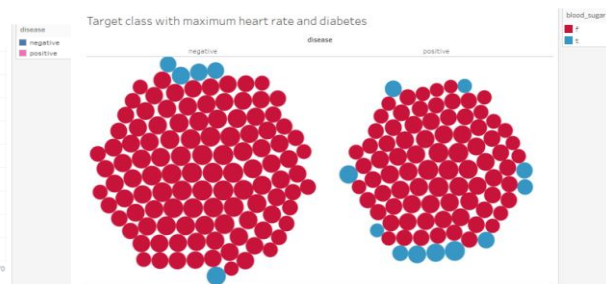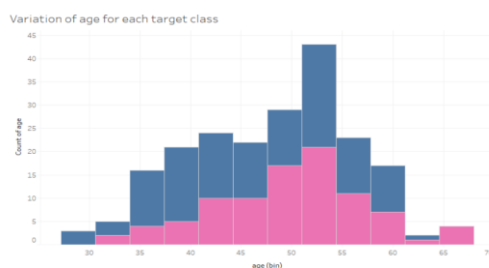
Literature paper-3:

Heart Disease Prediction using Exploratory Data Analysis

Authors name: R Indrakumari, T Poongudi, Soumya Rajan jena

Year of publication: 2022

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. This paper discusses the pre-processing methods, classifier performances and evaluation metrics. 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. The dataset is analysed with visualization tool tableau and K means clustering. 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.

Fig.2 Histogram of variation of age for each target class

Fig.3 Target class with maximum heart rate and diabetics

Fig.5 User defined filter to predict heart disease using exploratory data

Literature paper-4:

Big Data Analytics in Heart Attack Prediction

Authors name: Lidong Wang

Year of publication: 2017

Big Data analytics, known in the corporate world for its valuable use in controlling, contrasting and managing large datasets can be applied with much success to the prediction, prevention, management and treatment of cardiovascular disease. Data mining, visualization and Hadoop are technologies or tools of big data in mining voluminous datasets for information. The aim of this literature review was to identify usage of Big Data analytics in heart attack prediction and prevention, the use of technologies applicable to big data, privacy concerns for the patient, and challenges and future trends as well as suggestions for further use of these technologies. The national and international databases were examined to identify studies conducted about big data analytics in healthcare, heart attack prediction and prevention, technologies used in big data, and privacy concerns. A total of 31 studies that fit these criteria were assessed. As Per the studies analyzed, Big Data analytics is useful in predicting heart attacks, and the technologies used in Big Data are extremely vital to the management and tailoring of treatment for cardiovascular disease. And as the use of Big Data in healthcare increases, more useful personalized medicine will be available to individual patients. This review offers the latest information on Big Data analytics in healthcare, predicting heart attack, and tailoring medical treatment to the individual. The results will guide providers, healthcare organizations, nurses, and other treatment providers in using Big Data technologies to predict and manage heart attacks as well as what privacy concerns face the use of Big Data analytics in healthcare. Effective and tailored medical treatment can be developed using these technologies.

Sections	Main Findings
Big Data analytics in health care	Big Data analytics is a novel method of handling the numerous amounts of healthcare data that is streamed daily. It includes technologies and tools adept at navigating the massive amounts of data in any given healthcare system and mining useful information to treat patients
Patient-customized health care and big data	Big Data tools are now able to predict, prevent, and suggest the best evidence-based treatment plans for the patient based on the data from a variety of sources. Care managers, specially trained registered nurses trained to work with providers to empower and assist patients to make lifestyle changes, are essential in the data-driven society to personalize medicine.
Troponin and implanted sensors	Implanted sensors are now available to detect troponin levels in the blood prior to some heart attack and potentially alert emergency personnel to the problem and prevent mortality and morbidity associated with heart attacks.
The Internet of Things (IoT) and disease prediction	The IoT can be used to predict diseases based on monitoring of the elderly, sensors, and the data which can then be processed using Big Data analytics.
Wireless detection of a heart attack	Smartphones, wristwatches, and other human-based sensors can be used in predicting and preventing heart attacks prior to occurrence by reading the EKG which may show changes prior to heart attack and alert key personnel immediately.
Telecardiology in heart attack prediction	Telemedicine is in a key position to monitor and indicate when a person is having a heart attack in healthcare systems where cardiologists may not be available immediately. The data relayed in these eICU and eER settings are key to preventing morbidity and mortality in small rural healthcare area.
Data mining in big data processing and heart attack prediction	Data mining is a key tool of Big Data used to predict, prevent, and suggest the best treatment plan for heart attacks.
The issue of privacy in big data	Privacy is a key issue in today's data-driven healthcare society as most information is de-identified; however, it can be re-identified under certain circumstances. With the onset of Big Data uses in healthcare, it is imperative that privacy against hackers, identity theft, and the illegal uses of healthcare data is prevented.

Literature paper-5:

Prediction and analysis the occurrence of heart disease using data mining techniques

AUTHOR:POOJA KAMAT

YAER OF PUBLICATION:JAN 2018

Prediction of the occurrence of heart diseases in medical centers is significant to identify if the person has heart disease or not. Data mining is used to retrieve hidden information in medical centers that help to predict different disease. Heart disease is one of the most common diseases that lead to death in this world. Each year 17.5 million of people are dying due to cardiovascular disease according to World Health Organization reports. One of the most common problems in medical centers is that all experts do not have equal knowledge and skill to treat their patients, they give their own decision that may give poor results and lead the patients to death. To overcome such problems prediction the occurrence of heart diseases using data mining techniques and machine learning algorithms are playing vital roles for automatic diagnosis of disease in healthcare centers. Some machine algorithms used for predicting the occurrence of heart diseases are Support Vector Machine, Decision Tree, Naïve Bayes, K-Nearest Neighbour, and Artificial Neural Network.

A decision tree is one of supervised learning algorithm classifier, which is simple to understand and interpret. It deals with both numerical and categorical data sets
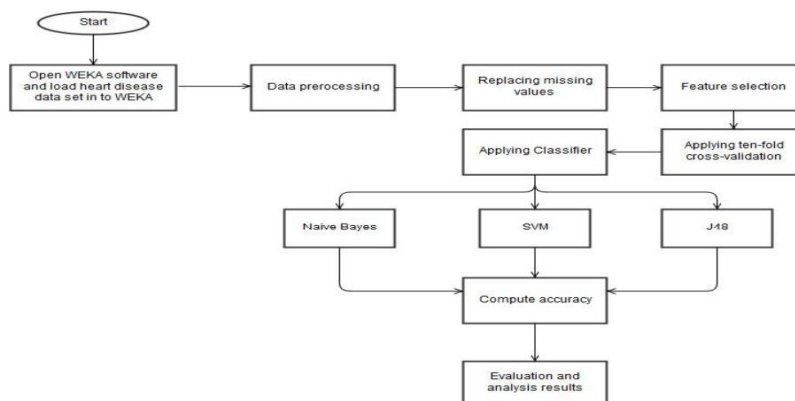

Fig - 1:flow diagram of proposed methodology

Determine the prediction performance of each algorithm and apply the proposed system for the area it needed. Use more relevant feature selection methods to improve the accurate performance of algorithms. Make the stakeholders use this proposed methodology for creating an attractive working condition that helps for better decision-making.

Literature paper-6:

Big data analytics in heart diseases prediction

Author name : Samir Ahmed

Year of publication : 2020

The healthcare data can be employed to develop a health prediction system that can improve in heart disease prevention. Big data on health care, including patient records, clinical notes, diagnosis, parents and family past ailments, hospitals, and scan results can aid in the phase of disease identification and prediction. The emerging machine learning method offers an important framework for forecasting cardiac diseases. The proposed framework used clusters from MapReduce. The system introduced a platform for storing heterogeneous and large data in IoT. The wearable devices send data as time series to IoT hub on Microsoft Azure as seen in fig 5. The streaming data is analyzed then can be stored on the cosmos DB.

The proposed work aims to develop a real-time prediction system for health issues based on big medical data processing on the cloud. In the proposed scalable system, the medical parameters are sent to Apache Spark to extract the attributes from the data and to apply the proposed machine learning algorithm aiming to predict the healthcare risks and send them as alerts and recommendations to the users and the healthcare providers as well.
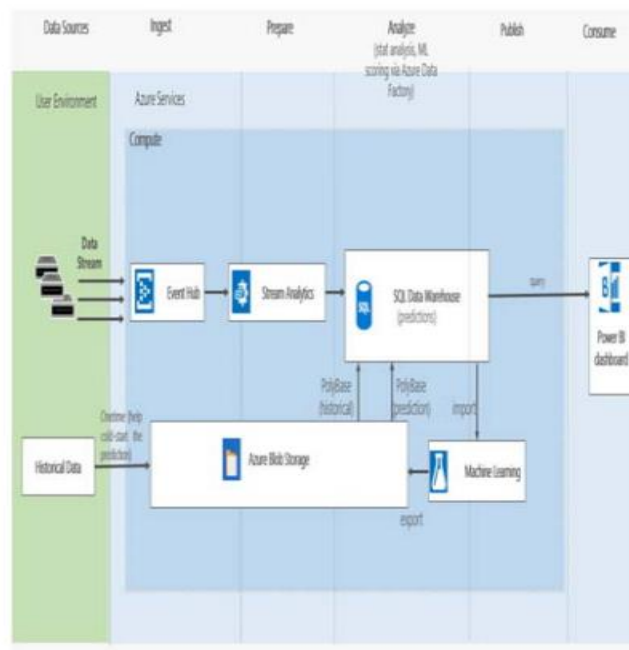

Fig 6: The Proposed System Based On Azure Data Analysis.

The program approach offered a broad computer structure that manages EHR computer focused on streaming data from connected medical devices and patient history for patients. The proposed method relies on the main parameters for the SVM based on tanning for these parameters.

Literature paper-7:

A Literature Review on Heart Disease Prediction Based on

Data Mining Algorithms

Authors name : Apruv Patel, kunjan D. khatri, Smit Kiri, Kathan Patel

Year of publication : 2018

"Heart Disease Prediction using Evolutionary Rule Learning". This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. This will facilitate (help) in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of coronary sickness. d "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection". This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. This methodology is efficient and less complex than conventional random forest model. Comparing to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection

Literature paper-8:

A LITERATURE SURVEY OF PREDICTING
HEART DISEASE

Authors name: M. Preethi, Dr. J. Selvakumar

Waveform analysis, time-frequency analysis, Neuro Fuzzy RBF ANN and Total Least Square-based Prony modeling algorithms are some of the techniques used to identify heart disease in the literature. However, in a study by Marshall et al (Marshall et al 1991), classification accuracy was not good with this technique (up to 79%) and the range of improvements to select the appropriate model was still sufficient. They also demonstrated the efficiency of neural networks in diagnosing heart attacks (acute myocardial infarction) by comparing multiple neural network classifiers, the multilayer perceptron and the Boltzmann perceptron classifier. Most of these approaches relate to diagnosis, not to the understanding of fundamental knowledge. Carlos Ordonez studied the prediction of heart disease with the help of association rules. They used a simple mapping algorithm. This algorithm treats attributes as numeric or categorical. It is used to convert medical records into transaction formats. Enhanced algorithms are used to minify restricted association rules. The mapping table is prepared and the attribute value is mapped to the item. Decision trees are used for data mining because they automatically divide numeric values . Split points selected in the decision tree are rarely used. Clustering is used to gain an overall understanding of the data.