

Visualizing and Predicting Heart Diseases with an Interactive Dash Board

Domain: Data Analytics

Team Id: PNT2022TMID40418

Batch no: B2-2M4E

By

Srinivasan S - 513419104043

Vasanthavasan G - 513419104053

Saravanan G G -513419104037

Faayiz khan N - 513419104014

Literature Survey

PAPER-1

TITLE: Heart Attack Disease Data Analytics and Machine Learning.

Author: Muhammad Nabeel et al...

Summary

Machine learning is the branch of Artificial Intelligence that have an important role in the field of the prediction and analysis of the diseases [1-7]. The heart failure or cardiac failure and we see this very commonly exclusively in the aging population and unembellished forms of heart failure have underprivileged prognosis and these patients who can have a relatively low 2 year and 5-year probability [8]. The patients are brought into the world with the condition and in some cases because of the disease of the heart muscle or all the more regularly harm to the heart valves as can occur in rheumatic heart disease so it's a condition that we deal with relatively commonly. Unfortunately, and it's a syndrome describes a group of clinical features so these patients are very often short of breath and breathless. The patients can be quite uncomfortable and even distressed at rest because of the breathlessness and they get very fatigued again the fatigue tends to start within the earlier stage [9]. But as the condition worsens they become fatigued for more of their day and they also develop swelling or edema the collection of excess fluids in the tissues of the body and this is gravity dependent so very often. We first notice it in the in the feet and the ankles so heart failure is an incapability of the heart to happen the metabolic strains of body. The bodies all the time have a metabolic demand there is biochemistry going on in all the cells of our body and this requires energy there's a metabolic demand. So all the cells in my body need a supply of oxygen they need nutrients, metabolic

substrates such as glucose fatty acids and of course all the other nutrients are also required and as well as that we need an adequate circulation to wash away .

CONCLUSION

Heart attack disease is one of the most common problems that results in the loss of a huge number of people. This project is the project that is very helpful to the people from the people, losses and also it is a great point for the savage of the human. First of all, we build models via the machine learning algorithms such as the decision tree, logistic regression, random forest and K neighbors. These are supervised learning algorithms in machine learning. In this project we find out the heart attack disease and normal results of disease on the basis of the different attributes like few of them are age of patient, sex of patient, chest pain type of patient, number of major vessels that are 0 to 3, Resting blood pressure in mm Hg, cholesterol in the mg, fasting blood sugar, and rest ECG. We find out the disease with 90.16% accuracy.

PAPER -2

TITLE: Heart Disease Prediction using Machine Learning and Data Analytics Approach

Author: Sanath Kapoor, Lekhraj Kasar*, Ashutosh Mandole and Dr Jayant Mahajan

Summary

Machine learning is an AI method that enables software applications to improve their ability to predict outcome without explicitly programming. Data mining provides many techniques which discover hidden patterns or similarities from data [8]. The algorithms used in Machine Learning rely on historical data to predict new output values. It does so by studying large sets of data. ML

applications are tools that are used to perform specific tasks. They collect data independently and without human intervention. These algorithms are used to improve prediction accuracy. However, they require significant effort and time to develop and operate correctly. A deep dive into machine learning reveals three critical components of algorithms: representation, evaluation, and optimisation. Representation means that data must be classified in a form and language that a computer can handle. The ML algorithms determine the most suitable model for the most effective and accurate outputs for the optimisation process. For example, a machine learning-based model trained on the most critical factors leading to heart failure can be helpful to predict it.

CONCLUSION

The data collected for the study includes only 303 individuals. In the above-discussed research papers, different algorithms were used such as Random Forest, Decision Trees, Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes, Logistic Regression and Feature Subset Selection etc. Out of all the mentioned algorithms, Random Forest, SVM and Naive Bayes presented more accurate results than others. This paper proposes a heart disease algorithm based on Ensemble Learning Technique, known as Stacking Classifier.

PAPER-3

TITLE: Forecasting of Heart Diseases in Early Stages Using Machine Learning Approaches

Author: Khushi Kumari Jha¹ et al...

Summary

Starting with the core notion of Machine Learning is an excellent place to start for this paper. In Machine Learning, a computer programme is given a set of tasks to complete, and it is claimed that the machine has learned from its experience if its measured performance in these tasks improves as it obtains more experience. As a result, the machine makes decisions and forecasts based on facts. In fact, machine learning is a subset of artificial intelligence. Data mining is a technique of discovering different kinds of patterns which are inherent in the data, they are accurate, new as well as helpful. It is a process of extracting knowledge from a big amount of data. Business analytics is a subset of data mining, which is akin to experimental research. Databases and statistics are the foundations of data mining. Implementing data mining techniques necessitates the use of two components: a database and machine learning. Data mining entails data comprehension, data preparation, and data modeling, whereas machine learning uses processed data as input to make predictions using algorithms. As a result, data mining necessitates the engagement of humans to clean and prepare the data as well as to comprehend the patterns. Human work is required just to define an algorithm in machine learning, after which the algorithm takes over operations. Many years ago, academicians attempted to create intelligent programmes using pre-defined rules, similar to how a regular programme works[2]. This method, however, did not work since there were too many special instances to examine. We can define rules to discover the shortest path between two points, for example. Making criteria for programmes like photo tagging, identifying emails as spam or not spam, and web page ranking, on the other hand, is extremely tough. To complete these objectives, the only option was to develop a programme that could generate its own rules by evaluating some samples (also called training data). Machine Learning was the name given to this method. From 2010 to 2015, the

accuracy of intelligence software based on ML, such as image recognition systems, increased from 72 percent to 96 percent.

CONCLUSION

ML, new techniques in making ML more effective in the promising area of healthcare may emerge soon. Some of the comparative approaches used are the confusion matrix, recall, F1-measure, precision, and accuracy. For the 13 attributes in the dataset, the K-NN classifier and Support Vector classifier have performed better as compared to decision tree classifier. When the dataset size is increased in the future, deep learning (DL) and many other recent optimizations can be used to create more promising results. To improve the evaluation findings, machine learning and a variety of other optimization methodologies can be used.

PAPER-4

TITLE: Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics

Author: D.R. Krithika

Summary

Cholesterol and unhealthy diet in saturated fat leads to high cholesterol level worsening the situation. Smoking is one of the major causes of damaging heart functions. CO₂ in cigarette smoke will increase in heart rate and blood pressure. Prevention is avoid using tobacco and smoking and dieting vegetables and whole grains, less sugar and less salt in food it helps to protect heart. Start to exercise regularly. The blood viscosity abnormality related closely to pathogenesis. Very large amount of data to making analytical process is bigdata analytics.

CONCLUSION

Fourteen attribute data applied machine learning algorithms like Extreme Gradient Boost, DT, KNN, SVM, Logistic Regression, Naïve bayes, Random forest, ANN, HPTRF (Hyper parameter tuned random forest classifier). HPTRF is my tuned algorithm and got better accuracy in this algorithm. Second part of the work hundred master checkup data is applied to predict cardiovascular patients. In this data attributes taken like age, gender, family history, smoke, alcohol, height, weight, BMI, SPO2, systolic,

diastolic, heartrate, ECG, ECO, TMT, HBA1C, HCT, TCHOL, LDL. These 100 data was applied in HPTRF, and I got 100% accuracy of training and testing and checked cardio affected patients counts based on these attributes and predicted cardio patients above HCT limit patients. So, hematocrit is important role plays to predicting cardiovascular disease. In future more factors to predict various disease.

PAPER-5

TITLE: Cardiovascular Disease Prediction using Deep Learning

Author: Paranthaman M et al...

Summary

New medical data mining approaches and machine learning techniques are being researched and developed. The major goal of this research is to use classification algorithms to find essential patterns and features in medical data, and next select the most important characteristics for detecting a silent heart attack. The results will be even more accurate thanks to the implementation of an Artificial Neural Network. While the development of such a system is not unique, current systems have flaws and are not designed to detect the possibility of silent

heart attacks. The goal of this study is to solve these issues and propose the installation of unique features to create a

more comprehensive system. The current technology for forecasting heart attacks does not produce results that are accurate enough. The machine learning approaches deployed, as observed in the literature survey, are pushing the accuracy to a certain limit. Furthermore, the utilization of attributes is a problem with the current heart attack prediction algorithm. The characteristics that should be chosen for heart attack prediction are conventional, and as a result, the findings frequently produce incorrect output. The suggested approach seeks to extract the appropriate properties from datasets to improve prediction precision. It will also provide a correct diagnosis to consumers so that they can grasp the problem without trouble.

The project distinguishes itself by merging deep learning and data mining capabilities. The study provides a system that includes a powerful classification module and a complete report generating module, as well as a strong prediction algorithm. This research proposes to build a self-learning

procedure in which past inputs of disease results influence future heart disease possibilities for a specific user. The proposed model makes extensive use of preprocessing methods to ensure that the classification and prediction of the dataset are error-free. To make the prediction more accurate, a large

number of training sets will be used.

CONCLUSION

The difficulty of restricting and summarizing diverse data mining techniques used in the field of medical prediction is examined in this work. For intelligent and successful heart disease prediction using data mining, the focus is on

combining various methodologies and combinations of several target criteria. Data mining is a useful tool for deriving meaningful

medical rules from medical data, and it plays a significant role in illness prediction and clinical diagnosis. There is a growing interest in utilizing categorization to determine whether or not a disease is present. The current study used a huge sample of hospitalized patients to demonstrate classification. The classification technique is extremely sensitive to noisy data. If there is any noisy data present, it presents major challenges in

terms of classification processing capacity. It not only slows down but also impairs the performance of the classification algorithm. As a result, before using a classification method, all attributes that will subsequently act as noisy attributes must be removed from datasets. We can apply preprocessing processes and classification rule algorithms, such as Multi-layer perceptron, in this research effort for classifying datasets that are supplied by users. The Multi-layer perceptron technique produces superior outcomes than other strategies, according to the testing data. Other data mining techniques and algorithms are likely to be used in the future to improve performance efficiency.