International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 04 | April-2021 ISSN: 2582-3930

FORECASTING OF MORTALITY IN PATIENTS WITH CARDIOVASCULAR DISEASE USING DATA MINING TECHNIQUES

TEJASWI M[1], SOWMYA K[2], RANJANI K[2], SANGEETHA R[2], ANITHA M[3]

[1]UG Scholar, Department of CSE, Kingston Engineering College, Vellore-59
[2]UG Scholar, Department of CSE, Kingston Engineering College, Vellore-59
[3]Asst.Professor, Department of CSE, Kingston Engineering College, Vellore-59

Abstract -Healthcare information systems store a huge amount of patient data, so the trend of the use of data mining in healthcare is on the rise. Heart and blood vessel diseases are a leading cause of mortality in patient's worldwide. Hence we decided to built a prediction system that will analyse the health status of patient i.e.,whether the patient suffering from heart disease will be alive or died. Here we collect the atmost 72 attributes of the patients and provide that data to trained data mining algorithms for prediction of health status. The aim of this project is to compare the effectiveness of these algorithms and choose the best efficient method among Decision Tree and Random Forest to predict mortality in patients with cardiovascular disease.

Key Words:Data mining, Decision tree, Random forest, prediction, Heart disease,

1.INTRODUCTION

Today's era is marked by data, and what happens is that every day, generating more and more datainevery aspectimaginable.AccordingtoGoogle'sChiefExecutiveOfficerEricSchmidt1 today for only two days, generating as much information as is generated from the dawn of civilization to 2003. That's somewhere around five Exabyte of data. Nowadays, it is almost inconceivable that there is an information system that does not have database in which data is stored. Because healthcare systems use information systems that have their own databases that generate huge amounts of patient data, it is also possible to apply data mining techniques to discover useful information that could be used to address some important issues in health care to help identify the symptoms of certain diseases to support diagnostics, to determine the extent to which particular medications and therapies affect patients, and generally improve the quality of life. Dataminingisaprocessofdiscoveringvariousmodels,summaries,andderivedvalues from a given collection ofdata.

According to the World Health Organization (WHO) 3, 17.9 million people die each year asaresultofheartdiseaseandbloodvessels,anditisestimatedthatby2030thatnumberwill increaseto23million.Fordecades,cardiovasculardiseases havebeenasignificantpartofthe leading causes of disease in the population of the Federation of Bosnia and Herzegovina, accounting for 17.6% of the disease structure in2018.

There are many factors that lead to the development of cardiovascular disease, the most important are: age, gender, inheritance, smoking, obesity, high blood pressure, and lack of physical activity. One of the problems that occur in medicine is to identify patients who are at certain risk of developing a disease on time. By identifying high-risk patients, steps can be taken to hinder or delay the onset of the disease or prevent it altogether. Cardiovascular diseasesareverypopularwithpredictionresearchbecausethey causehighmortalityandthere are great benefits of early detection to reduce risk. This project presents the application of a decision tree, neural network, and logistic regression to create models in predicting death withinaperiodof12monthsinpatientswithcardiovasculardisease.Thepurposeofthemodel istopredictoutcomesoffuturepatients,suchasmortality,basedoninformationfromprevious events.

2.LITERATURE SURVEY

Mangesh Limbitote, Dnyaneshwari Mahajan, Kedar Damkondwar, pushkar patil proposed a survey on prediction techniques of heart disease using machine learning in 2020.There objective is Heart is one of the most important part of the body. It helps to purify and circulate blood to all parts of the body. Most number of deaths in the world are due to Heart Diseases. Some symptoms like chest pain, faster heartbeat, discomfort in breathing are recorded. This data is analysed on regular basis. In this review, an overview of the heart disease and its current procedures is firstly introduced. Furthermore, an in-depth analysis of the most relevant machine learning techniques available on the literature for heart disease prediction is briefly elaborated. The discussed machine learning algorithms are Decision Tree, SVM, ANN, Naive Bayes, Random Forest, KNN. The algorithms are compared on the basis of features. We are

working on the algorithm with best accuracy. This will help the doctors to assist the heart problem easily. Then Sarath Babu, E M Vivek, K P Famina, K Fida, P Aswathi,M Shanid,M Hena proposed heart disease diagnosis using data mining technique in 2017.There objective is Data mining is an advanced technology, which is the process of discovering actionable information from large set of data, which is used to analyze large volumes of data and extracts patterns that can be converted to useful knowledge. Medical data mining has a great potential for exploring the hidden patterns in the data sets of medical domain. These patterns can be utilized to do clinical diagnosis. These data need to be collected in a standardized form. From the medical profiles fourteen attributes are extracted such as age, sex, blood pressure and blood sugar etc. can predict the likelihood of patient getting heart disease. These attributes are fed in to K-means algorithms, MAFIA algorithm and Decision tree classification in heart disease prediction, applying the data mining technique to heart disease treatment; it can provide as reliable performance as that achieved in diagnosing heart disease. By this medical industries could offer better diagnosis and treatment of the patient to attain a good quality of services. The main advantages of this paper are: early detection of heart disease and its diagnosis correctly on time and providing treatment with affordable cost.Then sellappan palaniappan, Rafah awang proposed intelligent heart disease prediction system using data mining techniques in 2008. There objective is The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not ";mined"; to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naive Bayes and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the defined mining goals. IHDPS can answer complex ";what if"; queries which traditional decision support systems cannot. Using medical profiles such as age, sex, blood pressure and blood sugar it can predict the likelihood of patients getting a heart disease. It enables significant knowledge, e.g. patterns, relationships between medical factors related to heart disease, to be established. IHDPS is Web-based, user-friendly, scalable, reliable and expandable. It is implemented on the .NET platform.

3.PROPOSED SYSTEM

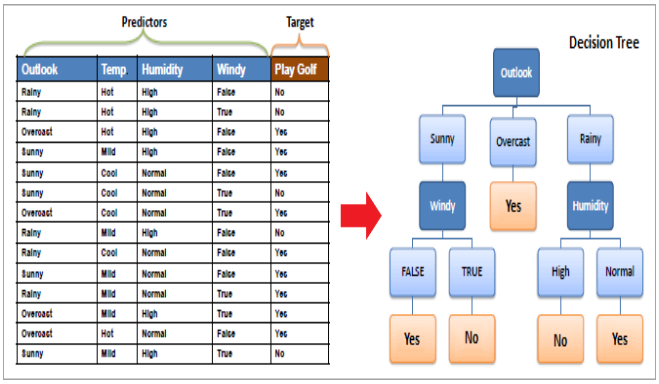
The proposed system presents the application of decision tree and random forest to create models in predicting death within a period of 12 months in patients with cardiovascular disease.Intially these algorthims gets trained from data set upto 70% then we provide 30% of test data to make its prediction based on training it got and by applying its knowledge. The purpose of this model is to predict the outcomes of future patients. Such as mortality, based on information from previous events. Cardiovascular diseases are very popular with prediction research because they cause high mortality and there are great benefits of early detection to reduce risk.The results showed that the random forest model provides slightly higher accuracy in both the classification of positive and negative samples than decision tree.Although all models showed great results when it comes to the prediction of positive cases, samples that patients will survive, slightly worse results were in the classification of negative cases, patients who will not survive. It can be concluded that using these methods one can build a model for predicting mortality in patients with cardiovascular disease in efficient way.

3.1 ALGORITHMS USED

We used two efficient machine learning data mining algorithms like Decision Tree, Random Forest.

Decision Tree Mining is a type of data mining technique that is used to build Classification Models. It builds classification models in the form of a tree-like structure, just like its name. This type of mining belongs to supervised class learning. In supervised learning, the target result is already known. Decision trees can be used for both categorical and numerical data. The categorical data represent gender, marital status, etc. while the numerical data represent age, temperature, etc. Decision Tree is used to build classification and regression models. It is used to create data models that will predict class labels or values for the decision-making process. The models are built from the training dataset fed to the system (supervised learning). Using a decision tree, we can visualize the decisions that make it easyto understand and thus it is a popular data mining technique.

An example of a decision tree with the dataset is shownbelow.



Randomforestisasupervisedlearningalgorithmwhich isusedforbothclassificationaswell as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of themandfinallyselectsthebestsolutionbymeansofvoting. Itisanensemblemethodwhich isbetterthanasingledecisiontreebecauseitreducestheover-fittingbyaveragingtheresult.

Working of Random Forest Algorithm

We can understand the working of Random Forest algorithm with the help of following steps

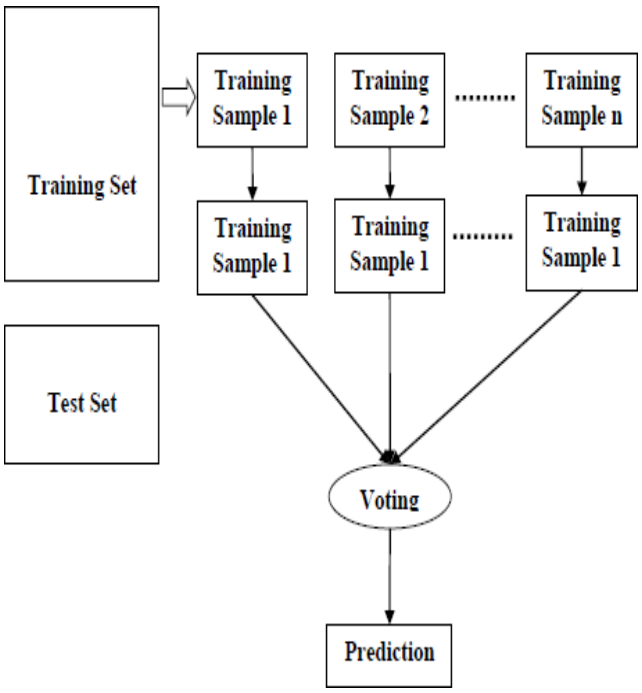
Step 1 − First, start with the selection of random samples from a given dataset.

Step 2 − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.

Step 3 − In this step, voting will be performed for every predicted result.

Step 4 − At last, select the most voted prediction result as the final prediction result.

The following diagram will illustrate its working –



3. CONCLUSIONS

- There are many researchers to under go this project to provide an efficient Machine learning algorithm to predict heart disease more efficiently which is main source of death worldwide. According to use Random Forest algorithm has anticipate the mortality rate in patients more accurately.
- It reduces the manal effort of doctor by storing the proper data about the patients in database and providing it to algorthims for anticipation.
- It will provide clarity to doctor what type of treatment ,he can provide to the patient.
- It will also ensure patients that what he/she has to take in daily routine like diet, exercise, type of food to eat etc so that patient can be healthly and may recover sooner.

REFERENCES

1. Mangesh Limbitote , Dnyaneshwari Mahajan , Kedar Damkondwar , Pushkar Patil, 2020, A Survey on Prediction Techniques of Heart Disease using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 06 (June 2020).
2. Heart disease diagnosis using data mining techniqueBabu, Sarath / Vivek, E M / Famina, K P / Fida, K / Aswathi, P / Shanid, M / Hena, M In:2017 International conference of Electronics, Communication and Aerospace Technology (ICECA) ; 1 ; 750-753 ; 2017ISBN:978-1-5090-5686-6, 978-1-5090-5685-9, 978-1-5090-5684-2.
3. Palaniappan, S. and R. Awang. “Intelligent heart disease prediction system using data mining techniques.” 2008 IEEE/ACS International Conference on Computer Systems and Applications (2008): 108-115.