RESEARCH ARTICLE OPEN ACCESS

Sorting and Guess of Heart Disease Risk Using Data Mining Techniques and Machine Learning

Mrs J Sarada MCA,M.Tech,M.Phil,(Ph.D) [1], Yeddula Pavani [2]

[1] Associate Professor, Department of Computer Applications

[2] Student, Department of Computer Applications

[1],[2] Chadalawada Ramanamma Engineering College (Autonomous)

ABSTRACT

Nowadays, heart disease is one of the prevailing main causes of morbidity and mortality. It is a hot health topic in our daily life, and heart disease treatment is very complicated. It is one-third of all deaths globally, stroke and heart disease. They both are globally the biggest killer, and their diagnosis availability is infrequent, especially in developing countries. This paper contains a framework based on some machine learning and data mining classification techniques on the heart disease dataset. There is no operational use of the data produced from the hospitals. Some convinced tools are used to extract the facts from the database to recognize the heart. This work is done by using Cleveland heart disease dataset that is sourced from the "UCI Machine Learning (ML) repository" to test and analyze on some various supervised ML and data mining techniques, some different attributes associated with causing of cardiovascular heart disease age, sex, chest pain type, chol, thal, etc. We will use these respective data to a model that will predict whether the patient has heart disease or not. This paper discussed the results of the modern techniques and will be used to predict the results for heart disease by summarizing some current research. The proposed method works best result in 86.89% accuracy by using a logistic regression algorithm.

Keywords: - Machine Learning, Classification Techniques, Prediction, Data Mining, Heart Disease, Python Programming.

I. INTRODUCTION

Data mining is a process that is used for mining information or knowledge from a huge database. It is an essential and significant step for discovering knowledge from existing databases. Data mining's primary task is that extract the hidden information and knowledge from the vast database. It is identified as Knowledge Discovery in Database (KDD). It is an important process where some common data mining techniques are used to extract the data arrangement. Data mining's technique helps to organizations to gain knowledge-based information. It includes understanding the business, data preparation, evaluating the data, and deployment. Its techniques work very rapidly and can find large amounts of data with the short passage of time. More likely, sometimes, it is referred to as knowledge discovery in databases. Suppose we use some professional and proficient computerized systems that are based on data mining and machine-learning algorithms. In that case, they can help us for achieving clinical assessments or diagnoses to minimize heart disease risk.

Machine learning is self-restraint that deals with programming, and it learns automatically and improves with experience. Bayesian and data mining analysis is trending, adding the demand for machine learning. Data mining has four different main techniques: cluster, Regression, Classification, and association rules. Classification is a fundamental technique in data mining. We can get the future outcome and predict the data based on historical data available in a database. The dataset can be classified into two categories through the classification technique, namely Yes and No. This method can achieve relevant and essential information for data and easily classify our data into different classes. "Data mining is the method for determining potentially useful arrangements through huge data sets and a large amount of database or metadata. It comes from different data sources, it may be sorted in various data warehouses and data mining sorting techniques" [2]. Knowledge Discovery in Database (KDD) is used for data integration and cleaning, data discovery patterns, Knowledge Presentation, data selection, and data

transformation. Healthcare association produces broad data to mark the factual decisions.

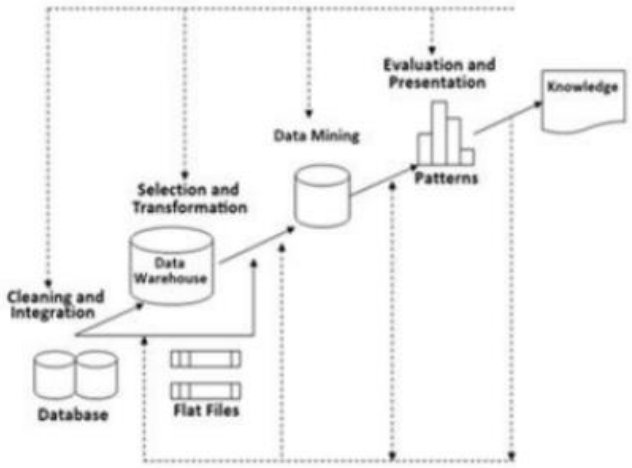

Figure 1. Process of knowledge discovery in data.

Data mining whole process is based on some various steps for Extracting respective knowledge. Data cleaning in data mining is how we can remove noise and corrupt or inaccurate records from data. We can prepare correct and complete data for data analysis by eliminating duplication in data through the data cleaning process. This data is usually not helpful when it comes to data analysis. The data cleaning process helps ensure that respective information is matched with the field and ensures data selection and transformation. The data transformation process is used to transform the data in a proper way required by data mining procedures. The pattern evaluation is used to represents knowledge based on different measures of interest that are given. We can use other heart disease patients' data collected after some diagnosis analysis and utilize the experience and knowledge of several specialists split with the same symptoms of coronary heart diseases. Complete and correct data helps the diagnosis analysis of patients for providing efficient treatment.

II. LITERATURE SURVEY

Shaikh Abdul Hannan et al. [5] used a Radial Basis Function(RBF) to predict the medical prescription for heart disease. About 300 patient's data were collected from the Sahara Hospital, Aurangabad. RBFNN (Radial Basis Function–Neural Network) can be described as a three-layer feed forward structure. The three layers are the input layer, hidden layer and output layer. The hidden layer consists of a number of RBF units (nh) and bias (bk). Each neuron on the hidden layer uses a radial basis function as a nonlinear transfer function to operate on the input data. The most often used RBF is usually a Gaussian function. Designing a RBFNN involves selecting centres, number of hidden layer units, width and weights. The various ways of selecting the centres are random subset selection, k-means clustering and others. The methodology was applied in MATLAB. Obtained results show that radial basis function can be successfully used (with an accuracy of 90 to 97%) for prescribing the medicines for heart disease.

AH Chen et al. [6] presented a heart disease prediction system that can aid doctors in predicting heart disease status based on the clinical data of patients. Thirteen important clinical features such as age, sex, chest pain type were selected. An artificial neural network algorithm was used for classifying heart disease based on these clinical features. Data was collected from machine learning repository of UCI .The artificial neural network model contained three layers i.e. the input layer, the hidden layer and the output layer having 13 neurons, 6 neurons and 2 neurons respectively. Learning Vector Quantization (LVQ) was used in this study. LVQ is a special case of an artificial neural network that applies a prototype-based supervised classification algorithm. C programming language was

used as a tool to implement heart disease classification and prediction trained via artificial neural network.The system was developed in C and C# environment.The accuracy of the proposed method for prediction is near to 80%.

Mrudula Gudadhe et al.[7] presented a decision support system for heart disease classification. Support vector machine (SVM) and artificial neural network (ANN) were the two main methods used in this system. A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease.

Manpreet Singh et al. [8] proposed a heart disease prediction system based on Structural Equation Modelling (SEM) and Fuzzy Cognitive Map (FCM).They used Canadian Community Health Survey (CCHS) 2012 dataset. Here, twenty significant attributes were used. SEM is used to generate the weight matrix for the FCM model which then predicts a possibility of cardiovascular diseases. A SEM model is defined with correlation between CCC 121(a variable which defines whether the respondent has heart disease) along with 20 attributes. To construct FCM a weight matrix representing the strength of the causal relationship between concepts must be constructed first. The SEM defined in the previous section is now used as the FCM though they have achieved the required ingredients (i.e. weight matrix, concepts and causality).80% of the data set was used for training the SEM model and the remaining 20% for testing the FCM model. The accuracy obtained by using this model was 74%.

III. EXISTING SYSTEM

Numerous types of information mining furthermore, machine learning methods have been executed have applied some techniques of data mining and machine learning namely Random Forest, J48 and Logistic Model Tree to build up a framework for precise heart disease prediction. Data mining can be viably used to anticipate sicknesses from the information base. Characterization procedures for example Decision Tree, Neural Network, Naive Bayes are used. Naive Bayes, KNN and ID3 algorithms are selected for prediction. It consists of two phases classifier and prediction. To foresee coronary illness with great exactness outcome. Coronary illness using a weighted fuzzy rule-based choice aid framework. The automated technique to produce the fuzzy rules is the benefit of the proposed framework. The amount of analytical information of coronary illness in the information layer, unseen layer might be changed to get reduce mistakes and huge exactness. The large information for checking the patients to have an upgraded arranging, dynamic, and cure with a decreased time and price.

Disadvantages

- To train the dataset KNN algorithm is used and in the expectation module, information is tried through the ID3 method. Then, risk level was divided into three categories minimum, maximum and average.
- Analytical information of coronary illness in the information layer, unseen layer might be changed to get reduce mistakes and huge exactness.

IV. PROPOSED SYSTEM

In this we introduced to sum up the new examination along relative outcomes on coronary illness expectation furthermore construct scientific ends by using techniques of data mining Clinical science additionally utilized a portion of crucial accessible instruments inside personal computer technology. During a decade ago expert system has acquired its instant in light of development within calculation ability. The clinical information of the patients has unseen patterns that are very important for information investigation in the diagnosis of coronary illness. It is the way of finding hidden information by examining a lot of information stored the heap of cardiovascular infection is quickly expanding. We will save human wealth since we do not require complex detection procedures in medical clinics with correct expectations, we can tackle the unessential issues.

Advantages

- Input dataset of heart disease has been taken have applied some techniques to anticipate coronary illness.
- The automated technique to produce the fuzzy rules is the benefit of the proposed framework.
- Fuzzy standards are applied to make the choice aid framework by applying a fuzzy conclusion framework and the prediction of risk can be completed on a planned fuzzy system

IMPLEMENTATION

- Admin
- Doctor
- Patient

Admin

Admin can add the doctor and schedule as per the doctor available time. Admin can view the patient appointment and he can schedule patient time as per available doctor.

Doctor

Doctor can view the patient request and after the check-up doctor add the patient report like Personal Information,Lab Observation 1, Lab Observation 2, Critical Conditions, Diseases Details, Treatment Detail, and Medicine Detail. As a final result we will showing Actual result, Predicated Result and Attack Stage.

Patient

Patient should register and login patient can book appointment and see the doctor schedule which doctor available that time.

V. DISCUSSION

The whole section describes a portion of new task performed within information mining identified with cardiovascular illnesses. Information mining methods can be successfully applied to convert raw data from the tremendous measure of information available in medical field. Such studies illustrate, instead of implementing a specific data mining method to a particular database, the creators have applied a fusion framework. Outcomes are more beneficial if an assortment of mining strategies are utilized. Weka, MATLAB and so forth are a portion of the other famous instruments utilized for information investigation. Cautious determination of the mixture of data mining strategies with exact usage of these strategies on database returns quick and viable usage of framework for coronary illness. A portion of the task is to differentiate various categorization strategies on database to exactly characterize whether the patients have heart disease or not. Normally utilized classifying methods are Decision tree, Naive Bayes, ANN, fuzzy approach. Other than investigating those generally utilized procedures, a portion of new task have considered mixture of models. The possibility of mixture model is to consolidate few familiar determination methods in solitary model to give better outcomes. It has been noticed that mixture of models has achieved better accuracy than a solitary model.

VI. CONCLUSION

Heart disease is another dangerous and preventable disease that cause millions of deaths in the world today. Each clinical specialist has not equivalent information moreover expertise to obtain a precise choice, where a few specialists provide a weak sensible choice that guide individuals to risky circumstances. So, preventing the occurrences of heart disease is much important to reduce the number of deaths. Different types of techniques have been used to predict heart disease based on risk factors and brief analysis of the accuracy of those techniques. The exactness of algorithms in machine learning relies on the dataset that is used for preparing and testing purposes. One of the significant disadvantages of the task is that the principal centre has been around

the use of categorization procedures for coronary illness expectation, instead of examining different information cleaning and pruning methods that plan and create a dataset reasonable for mining. For the longer-term extension, extra machine learning method will be utilized for the most useful examination of the heart infections as well as for prior expectation of sicknesses with the goal that the pace of the demise cases can be limited by the realization about the illnesses.

REFERENCES

[1] H. B. F. David and S. A. Belcy, "Heart Disease Prediction Using Data Mining Technques" ICTACT Journal On Soft Computing, 2018, Vol: 09, Issue: 01

[2] X. Liu, X. Wang, Q. Su, M. Zhang, Y. Zhu,Q. Wang, and Q. Wang, "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method", Computational and Mathematical Methods in Medicine, Volume 2017, Article ID 8272091, 11 pages, DOI: https://doi.org/10.1155/2017/8272091.

[3] C. B. C. Latha, S. C. Jeeva, "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques", Informatics in Medicine Unlocked 16,2019.

[4] C. Gazeloglu, "Prediction of heart disease by classifying with feature selection and machine learning methods", Progress in Nutrition 2020; Vol. 22, N. 2: 660-670, DOI: 10.23751/pn.v22i2.9830.

[5] Heart Disease UCI, https://www.kaggle.com/ronitf/heartdisease- uci, Accessed Date: 20. June. 2020

[6] N. Guru, A. Dahiya and N. Rajpal, "Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review, Vol. 8, No. 1, pp. 1-6, 2007.

[7] S. Palaniappan and R. Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", International Journal of Computer Science and Network Security, Vol. 8, No. 8, pp. 1-6, 2008.

[8] S. B. Patil and Y.S. Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System using Data Mining and Artificial Neural Network", European Journal of Scientific Research, Vol. 31, No. 4, pp. 642-656, 2009.

[9] N. Singh and D. Singh, "Performance Evaluation of K-Means and Hierarchal Clustering in Terms of Accuracy and Running Time", Ph.D Dissertation, Department of Computer Science and Engineering, Barkatullah University Institute of Technology, 2012

[10] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert Systems with Applications, vol. 35, no. 1-2, pp. 82-89,2000