

REVIEW ON PREDICTION SYSTEM FOR HEART DIAGNOSIS USING DATA MINING TECHNIQUES

Divya Kundra

Shaheed Udham Singh College of Engineering & Technology, Tangori, Distt Mohali.

Er. Navpreet Kaur

Shaheed Udham Singh College of Engineering & Technology, Tangori, Distt Mohali.

ABSTRACT: Data mining is the process of analyzing large sets of data and then extracting the meaning of the data. It helps in predicting future trends and patterns, allowing business in decision making. Data mining applications can answer business questions that take much time to resolve traditionally. Large amount of data which is generated for the prediction of heart disease is analyzed traditionally and is too complicated and voluminous to be processed. Data mining provides the techniques and methods for the transformation of data into useful information for decision making. These techniques make the process fast and it takes less time for the prediction system to predict the heart disease with more accuracy. In this paper we survey different papers in which one or more algorithms of data mining used for the prediction of heart disease. Result from using neural networks is 99.62% in one paper [6]. By Applying data mining techniques to heart disease data which needs to be processed, we can get effective results and achieve reliable performance which will help in decision making in healthcare industry. It will help the medical practitioners to diagnose the disease in less time and predict the probable complications well in advance. Identifying the major risk factors of Heart Disease categorizing the risk factors in an order which causes damages to the heart such as high blood cholesterol, diabetes, smoking, poor diet, obesity, hyper tension, stress, etc. Data mining functions and techniques are used to identify the level of risk factors to help the patients in taking precautions in advance to save their life.

Keywords: heart disease ,data mining, data mining techniques, data mining tools, data mining applications, methodology

INTRODUCTION

Data mining is the analytical process to explore specific data from large volume of data. It is a process that finds previously unknown patterns and trends in databases. This information is further used to build predictive models. The main objective of our paper is to learn the different data mining techniques which are used in the prediction of heart diseases using any data mining tool. Heart is the most vital part of the human body as life is dependent on efficient working of heart. A Heart disease is caused due to narrowing or blockage of coronary arteries. This is caused by the deposition of fat on the inner walls of the arteries and also due to build up cholesterol. Heart diseases can be caused due to number of factors:

High blood pressure: when the heart pumps blood, the force of the blood pushes against the walls of the arteries causing pressure. If the pressure rises and stays high over the time it is called high blood pressure or hypertension which can harm the body in many ways i.e. Increasing the risk of heart stroke or developing heart failure, kidney failure etc.

High cholesterol: cholesterol is a waxy substance found in the fatty deposits in the blood vessels. Increase in the fatty deposits (high cholesterol) does not allow sufficient blood to flow in through the arteries causing heart attacks.

Unhealthy diet: eating too much fast food increases blood pressure and cholesterol level causing the risk of heart attacks.

Smoking: it damages the lining of arteries and builds up a fatty material called atheroma which narrows the arteries causing heart attacks.

Lack of physical activity: lack of exercise increases cholesterol level in blood vessels which further increases the risk of heart attacks.

Obesity: obese people are more likely to have high blood pressure, high cholesterol level and diabetes (increase in blood sugar level) which increases the risk of heart strokes in human body. Nowadays, data mining is gaining popularity in health care industry as this industry generates large amount of complex data about hospital resources, medicines, medical devices, patients, disease diagnosis etc. This complex data needs to be processed and analysed for knowledge extraction which will further help in decision making and is also cost effective.

REVIEW ON PREDICTION SYSTEM FOR HEART DIAGNOSIS USING DATA MINING

World health organisation has estimated 17.5 million people died from cardio vascular diseases in 2012, representing 31 percent of all global deaths. Out of these, an estimated 7.4 million were due to coronary heart disease and 6.7 million were due to stroke. WHO estimated by 2030, almost 23.6 million people will die due to heart disease as written in [1].

Thus, a beneficial way to predict heart diseases in health care industry is an effective and efficient heart disease prediction system. This system will find human interpretable patterns and will determine trends in patient records to improve health care.

LITERATURE SURVEY

Over the years, numerous works have been done related to heart disease prediction system using different data mining algorithms by different authors. They tried to achieve efficient methods and accuracy in finding out diseases related to heart by their work including datasets and different algorithms along with the experimental results and future work that can be done on the system to achieve more efficient results. This paper aims at analyzing different data mining techniques that has been introduced in recent years for heart disease prediction system by different authors.

M.A.Nishara Banu and B.Gomathy[2] used C4.5 algorithm, MAFIA and K means clustering in the year 2014 using 13 attributes in the dataset achieving 89 percent accuracy.

Aqueel Ahmed et al. [3] show the classification techniques in data mining and show the performance of classification among them. In this classification accuracy among these data mining has discussed. In this decision tree and SVM perform classification more accurately than the other methods and was able to achieve 91% accuracy

Ms.Ishtake et al. [4] developed a prediction system for heart diagnosis using decision tree, Neural Network and Naive Bayes techniques using 15 attributes in the year 2013.

Chitra R.et al. [5] developed the computer aided heart disease prediction system that helps the physician as a tool for heart disease diagnosis. From the analysis it is concluded that neural network with offline training is good for disease prediction in early stage and good performance can be obtained by pre-processed and normalized dataset.

Nidhi Bhatla et al. [6] projected the study of different data mining techniques that can be employed in automated heart disease prediction systems. The analysis shows that Neural network with 15 attributes has shown the highest accuracy. On the other hand, Decision tree has also performed well with 99.62% accuracy by using 15 attributes.

Shadab et al. [7] used Naive Bayes technique in the year 2012 using 15 attributes in the dataset for the heart diagnosis in heart prediction system.

COMPARATIVE STUDY OF DATA MINING TRENDS FROM PAST TO FUTURE

PAST:- in the previous years, statistical and Machine learning techniques were used on numerical data stored in traditional databases and the computing resources were 4G PL and various related techniques[8].

PRESENT: - these days, along with the statistical and Machine learning techniques, artificial intelligence and pattern reorganisation techniques are also used [8].

FUTURE:-in future, for complex data objects which includes high dimensional, high speed data streams, sequence, noise in the time series and for multi instance objects, soft computing techniques like fuzzy logic, neural networks and genetic programming is used. Computing resources used would be multi-agent technologies and cloud computing [8].

DIFFERENT DATA MINING TOOLS USED IN HEART DISEASE PREDICTION SYSTEM WITH ACCURACY

Abhishek et al in the year 2013 used data mining tool Weka 3.6.4 in heart disease prediction system using J48 technique achieved 95.56% accuracy and using Naive Bayes achieved 92.42%. [9]

Rashedur et al in the year 2013 used Neural network technique using Weka data mining tool and achieved 79.19% and to compare various classification techniques, he used another technique fuzzy logic using TANGRA data mining tool and achieved 83.85% accuracy. [10]

Nidhi et al in the year 2012 used data mining tool Weka 3.6.6 in the analysis of heart disease prediction system and achieved 99.52% using Naive Bayes. She also used TANGRA data mining tool but could achieve up to 52.33% only using decision trees. She also tried .NET data mining tool and achieved up to 96.5% using neural networks. [6]

Resul et al in the year 2009 used SAS base software 9.1.3 achieving 97.4% using neural networks. [11]

DATA MINING TECHNIQUES ALSO USED IN DIAGNOSIS OF OTHER DISEASES

Humar et al in the year 2008 used classification, Backpropagation, Fuzzy neural network techniques for diabetes and heart diseases. [12]

Marcel et al in the year 2007 used in Bayesian classification for Characinoid heart disease. [13]

Mohammad et al in the year 2012 used C4.5 and C5.0 algorithm for heart disease and breast cancer diagnosis. [14]

M.Akhil et al in the year 2012 used associative classification and genetic algorithm for the diagnosis of breast cancer, Pima Indian Diabetes and heart disease. [15]

DATA MINING AND ITS TECHNIQUES

DATA MINING

It is main concerned with extracting useful information from large amount of databases. Data mining techniques and tools are used to find unknown patterns and trends from the data set. Its main objective is to automatically find the patterns in the dataset with minimal user effort and input. Data mining's main contribution is in decision making and in forecasting future trends of market. Many organisations use data mining as a tool these days for data analysis as it easily evaluates patterns and trends of market and produce effective results.

DATA MINING TECHNIQUES

ASSOCIATION: it is the best known and well researched method for data mining. Association is also called relation technique because patterns which are discovered from the dataset are based on the relationship between the items. For example, when association technique is used in heart disease prediction system, it tells us the relationship between all the attributes and sort out all the patients with all the risk factors which are required for the heart disease predictions.

CLASSIFICATION: it is a data mining technique which is used to classify each item in a data set into one of predefined set of classes or groups. It is a classic data mining technique which is based on machine learning.

CLUSTERING: It is a data mining technique which creates useful cluster of objects that have similar characteristics using automatic technique. There is a slight difference between clustering and classification. Clustering defines classes and put objects in them while classification assigns objects into predefined classes. Clustering helps to make clusters or list of patients having same risk factor.

PREDICTION: it is a data mining technique which discovers relationship between independent variables and relationship between dependent and independent variables.

SEQUENTIAL PATTERNS: it is a data mining technique that discovers similar patterns or regular events in transaction data over a business period.

DECISION TREE: it is the most used data mining techniques and its model is easily understandable. The root of the decision tree is a simple question or condition that has multiple answers. Each answer leads to a set of questions or conditions which helps to determine the data so that we can take a final decision based on it.

DATA MINING APPLICATIONS

Data mining is used in various fields such as retail industry, telecommunication industry, healthcare industry, financial data analysis, intrusion detection, sports and also in analyzing student's performance.

RETAIL INDUSTRY: data mining is a great application in retail industry as it collects large amount of data which includes transportation, sales and consumption of goods and services. This data expand rapidly due to increase in purchase and sales in business. Data mining helps to identify customer's buying patterns and trends that lead to improved quality of customer service and customer's satisfaction.

TELECOMMUNICATION INDUSTRY:

Telecommunication industry is the most growing industry as it provides various services such as fax, pager, cellular phones and e-mails. With the development of computer, telecommunication services have integrated with the communication technologies and works more effectively. Data mining helps to identify telecommunication patterns, fraud activities, make better use of resources and improve quality of service.

REVIEW ON PREDICTION SYSTEM FOR HEART DIAGNOSIS USING DATA MINING

HEALTHCARE INDUSTRY: Data mining is very useful in healthcare industry in diagnosis of heart diseases, breast cancer and diabetes. It helps in identifying patterns and trends in patient's records having same risk factor and helps in decision making.

FINANCIAL DATA ANALYSIS: financial data in banking is reliable and of high quality which facilitates systematic data analysis in financial industry. It helps in loan payment prediction and customer credit policy analysis. It also helps in clustering of customers for target marketing.

INTRUSION DETECTION: intrusion is any kind of action that threatens the confidentiality or integrity of network resources from any outside party. With the increased usage of internet and availability of the tools and tricks of intrusion and attacking network, intrusion detection has become an important issue for network administration. Data mining helps in the development of data mining algorithm for intrusion detection and analysis of stream data so that intrusion threats can be avoided.

SPORTS: in sports, vast amount of statistics are gathered for each player, team, game and season. Data mining is used in the prediction of performance of players, selection of players and forecast of future events.

STUDENT'S PERFORMANCE: data mining is used to evaluate student's performance using classification technique for data classification. Attendance, class test, seminar and assignment marks are collected from the student record to predict the performance of the student at the end of the semester.

DATA MINING TOOLS:

There are various data mining tools used for data mining purpose. These are WEKA, TANAGRA, MATLAB and .NET FRAMEWORK.

WEKA: it is a data mining tool which was developed in New Zealand by the University of Waikato that implements data mining algorithms using JAVA language. WEKA is a collection of machine learning algorithms and their application to the data mining problems. These algorithms are directly applied to the dataset. WEKA supports data file in ARFF format. WEKA is open source software and hence, it is not dependent on any platform. It includes algorithms for data processing, classification, regression, clustering, association and also visualization tools.

TANAGRA: it is open source software as researchers can access to the source code and add their own algorithms and compare their performances, if it conforms to the software distribution license. It includes several data mining algorithms from statistical learning, machine learning, data analysis and database area.

MATLAB: It is a data mining tool built in high level language. It provides interactive environment for visualization, numerical computation and programming. The built in math functions, language and tool explore various approaches and helps to reach a solution faster than with the spreadsheet of traditional programming languages like C,C++ and JAVA. It analyse data, develop algorithms, and create models and applications.

.NET FRAMEWORK: it is a software framework developed by Microsoft which runs primarily on Microsoft windows. It provides secure communication and consistent applications. it provides language interoperability(each language can code written in other languages) across several programming languages.

DATA MINING METHODOLOGY:

data mining is the fundamental part of knowledge discovery database. Knowledge discovery database is the process of discovering hidden knowledge from massive amounts of data that we are technically capable of gathering and storing. It is a process which contains sequence of following steps:

PROBLEM DEFINITION: in this, we define the problem that is to be solved and the noise or the irrelevant data is removed.

DATA INTEGRATION: here multiple data sources are combined to gather and prepare information.

MODEL BUILDING: here the data is retrieved from the database which is relevant to the analysis task and is transformed into the form which is appropriate for mining by performing certain operations.

DATA MINING: in this, different algorithms are applied to extract patterns.

KNOWLEDGE DEPLOYMENT: knowledge representation techniques are used to present mined knowledge to the user.

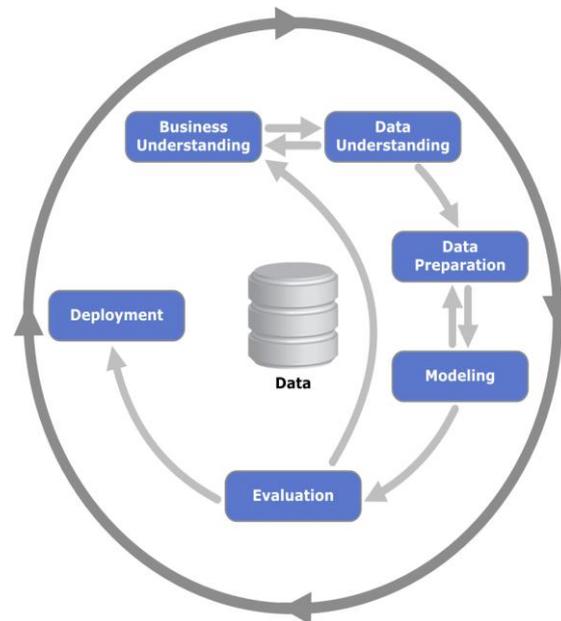


Fig 1: concept of data mining

CONCLUSION

Our goal is to analyse different data mining techniques in a way that they help us to predict and employ the heart disease prediction system which predicts the possibility of heart disease risk of patients for next ten years. To spread awareness amongst patients to take precautionary actions and hence live longer; and to assist doctors to diagnose and predict the probable complications well in advance. As the heart disease patients are increasing world-wide each year and huge amounts of data is available for research, researchers are using data mining techniques in the diagnosis of heart disease. Analysis presented by different researcher's shows that different data mining techniques and classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that using different techniques and taking different number of attributes we get different accuracies for predicting heart diseases. Taking analysis from some papers published by the researcher's it is shown that techniques such as neural networks give an accuracy of 100% in predicting heart diseases where as some may show that using the technique of Decision Tree we can get an accuracy of 99.62%. So, different techniques used have shown different accuracies depending upon number of attributes taken and tool used for implementation.

REFERENCES

- [1]. Ms. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology", 2012
- [2]. M.A.Nishara Banu and B.Gomathy, "Disease Forecasting System Using Data Mining Methods", 2014
- [3]. Aqueel Ahmed, Shaikh Abdul Hannan, "Data Mining Techniques to Find Out Heart Diseases", International Journal of Innovative Technology and Exploring Engineering(IJITEE)ISSN: 2278-3075, Volume-1, Issue-4, September 2012.
- [4]. Ms. Ishtake S.H, Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research, 2013
- [5]. Chitra R and Seenivasagam V, "REVIEW OF HEART DISEASE PREDICTION SYSTEM USING DATA MINING AND HYBRID INTELLIGENT TECHNIQUES", ISSN: 2229-6956(ONLINE) ICTACT JOURNAL ON SOFT COMPUTING, JULY 2013, VOLUME: 03, ISSUE: 04, 2013
- [6]. Nidhi Bhatla and Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 1 Issue 8, October – 2012
- [7]. Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical Sciences, 2012
- [8]. Venkatadri.M, Dr. Lokanatha C. Reddy a review on data mining from past to the future. International Journal of Computer Applications, 2011.

REVIEW ON PREDICTION SYSTEM FOR HEART DIAGNOSIS USING DATA MINING

- [9]. Abhishek taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
- [10]. Rashedur M. Rahman, Farhana Afroz, Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications, 2013.
- [11]. Nidhi Bhatla Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2012.
- [12]. Humar Kahramanli, Novruz Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Elsevier, 2008.
- [13]. Marcel A.J. van Gerven, Predicting carcinoid heart disease with the noisy-threshold classifier, Elsevier, 2007.
- [14]. Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012
- [15]. M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, Heart Disease Prediction System using Associative Classification and Genetic Algorithm, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012