www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 11990-11996

An Experimental Study of Data Mining Techniques in Blood Donors Sector

[1]Deepthi Srambical Poulose and [2]Dinesh Kumar Sahu and [3]Anil Rajput
1M.Tech Scholar, Department of CSE, Sri Satya Sai College of Engineering , Bhopal M.P., India
2Ph.D Scholar, Department of Computer Science, Barkatullah University , Bhopal M.P., India
3Professor, Department of Mathematics, CSA, Govt. P.G. College, Sehore M.P., India

[1]Deepsp57@gmail.com
[2] Dinesh_sahu20@yahoo.com

Abstract—In today's computer age data storage has been growing in size to unthinkable ranges that only computerized methods applied to find information among these large repositories of data available to organizations whether it was online or offline. Data mining was conceptualised in the 1990s as a means of addressing the problem of analyzing the vast repositories of data that are available to mankind, and being added to continuously. Data mining is necessary to extract hidden useful information from the large datasets in a given application. This usefulness relates to the user goal, in other words only the user can determine whether the resulting knowledge answers his goal. The growing quality demand in the blood bank sector makes it necessary to exploit the whole potential of stored data efficiently, not only the clinical data and also to improve the behaviours of the blood donors.

Keywords— Data Mining, classification, Decision Tree

INTRODUCTION

Data mining can contribute with important benefits to the blood bank sector, it can be a fundamental tool to analyse the data gathered by blood banks through their information systems. In recent years, along with development of medical informatics and information technology, blood bank information system grows rapidly. With the growth of the blood banks, enormous Blood Banks Information Systems (BBIS) and databases are produced.
It creates a need and challenge for data mining. Data mining is a process of the knowledge discovery in databases and the goal is to find out the hidden and interesting information . Various important steps are involved in knowledge discovery in databases (KDD) which helps to convert raw data into knowledge.

Data mining is just a step in KDD which is used to extract interesting patterns from data that are easy to perceive, interpret, and manipulate. Several major kinds of data mining methods, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization, and meta-rule guided mining will be reviewed. The explosive growth of databases makes the scalability of data mining techniques increasingly important. Data mining algorithms have the ability to rapidly mine vast amount of data.

Data mining is needed in many fields to extract the useful information from the large amount of data. Large amount of data is maintained in every field to keep different records such as medical data, scientific data, educational data, demographic data, financial data, marketing data etc. Therefore, different ways have been found to automatically analyze the data, to summarize it, to discover and characterize trends in it and to automatically flag anomalies. The several data mining techniques are introduced by the different researchers. These techniques are used to do classification, to do clustering, to find interesting patterns. In our future work, the data mining techniques will be implemented on blood donor's data set for predicting the blood donor's behaviour and attitude, which have been collected from the blood bank center.
Data mining is the extracting knowledge from the large amount of data .It is defined find the hidden information in the database. Data mining is used for the many application there are medical, fraud detection, and marketing. Data mining is the four types of classes are used there are classification, clustering, regression, and association rule. Data mining is provides automatic pattern recognition and attempts to uncover patterns in data that are difficult to detect with statistical method. Data mining is called the another names are Knowledge Discovery Database (KDD). Exploratory data analysis, datadiscovery, and deductive learning. Platelets are count with white blood cell and are related to the gravity of blood. Platelets are count with low and normal.

Platelets are very low in our body they are affected in cancer, leukaemia and heavy bleeding. Suppose the blood platelet transfusion are successful we have check the many condition, for example the illness, temperature, age, Decision tree is used to data analysis tool and can easily understand the easily transform the rule, the decision tree algorithm are CART, ID3, C4.5, HUNTS, SLIQ, and SPRINT. The main goal is the decision tree is minimizing the number of tree levels and tree node

2. Brief review of the work already done in the field

In June 2009, [1] work is carried out on management information system that helps managers in providing decision making in any organization. This MIS is basically all about the process of collecting, processing, storing and transmitting the relevant information that is just like the life blood of the organization that uses a data mining approach.

In June 2009, [2] the research work entitled on web based information system for blood donation, all the information regarding blood donation are available on world wide web i.e. online systems that communicate/interconnect all the blood donor societies in a country using LAN Technology.

In July 2009, [3] the research work entitled on Segmentation, Reconstruction and Analysis of Blood thrombus formation in 3D-2 photons microscopy images. In this work, method and platform are applied to differentiate between the thromby formed in wild type and low FV11 mice. The high resolution quantitative structural analysis using some algorithms that provide a new matrix, that is likely to be critical to categorize and understanding bio-medically relevant characteristics of thromby. Also, in this work, composition of different components on the clot surface and number of voxels in each clot component has been computed.

In December 2009, [4] the work proposed on an application to find spatial distribution of blood donors from the blood bank information system, blood donation and transfusion services are carried out and help the patients to access the availability of blood from anywhere.

In the year 2009, [5] the research was based geographical variation in correlates of blood donor turn out rates-an investigation of Canadian metropolitan areas. In this work, Canadian blood services i.e. an organization that aims the collecting and distributing the blood supply across the country. Accessibility variables are introduced and calculated using a 2-step floating catchment area in order to analyse the accessibility of services. The regression technique is also used.

In the year 2010,[6] the research is accomplished on application of CART algorithm in blood donor's classification. In this work, one of the popular data mining technique i.e. Classification is used and through the use of CART application, a model is created that determines the donor's behavior.

In the year 2010, [7] the data mining tool is used to extract information from PPI (Protein-Protein Interaction) systems and developed a PPI search system known as PP Look that is an effective tool 4-tier information extraction systems based on a full sentence parsing approach. In this paper, researchers introduced a useful tool that is PP Look which uses an improved keyword, dictionary pattern matching algorithm to extract protein-protein interaction information from biomedical literature. Some visual methods were adopted to conclude PPI in the form of 3D stereoscopic displays.

In December 2010, [8] the work is proposed on binary classifiers for health care databases i.e. a comparative study of data mining classification algorithms in the diagnosis of breast cancer. This work helps in uncovering the valuable knowledge hidden behind them and also helping the decision makers to improve the health care services. In this work, the presented experiment provides medical doctors and health care planers a tool to help them quickly make sense of vast clinical databases.

In February 2011, [9] the research paper presented on classifying blood donors using data mining techniques, the work is performed on blood group donor data sets using classification technique.

In August 2011, [10] the research work carried out on interactive knowledge discovery in blood transfusion data set, the work is done through conducting data mining experiments that help the health professionals in better management of blood bank facility.

In year 2011, [11] the work presented on rule extraction for blood donors with fuzzy sequential pattern mining, fuzzy sequential pattern algorithm is used to extract rules from blood transfusion service center data set that predicts the behaviour of donor in the future.

In August 2011, [12] the research work proposed on a comparison of blood donor classification data mining models, which uses decision tree to examine the blood donor's classification. In this work, comparison between extended RVD based model and DB2K7 procedures are carried out and it was discovered that RVD classification is better than DB2K7 in aspects of recalling and precision capability.

In the September 2011, [13] the research work carried out on real time blood donor management using dash boards based on data mining models, the data mining techniques

are used to examine the blood donor classification and promote it to development of real time blood donor management using dash boards with blood profile or RVD profile and geo-location data.

In year 2011, [14] the work was proposed on an intelligent system for improving performance of blood donation. In this work, many important techniques i.e. Clustering, K-means classification is adopted that improves the performance of blood donation services.

In January 2012, [15] the data mining techniques are applied on medical databases and clinical databases which are used to store huge amount of information regarding patient's diagnosis, lab test results, patient's treatments etc. which is a way of mining the medical information for doctors and medical researchers. In all developed countries, it is found that the routine health tests are very common among all adults. It is also determined that precautionary measures are less expensive rather than the treatments and also facilitates a better chance for patient's treatment at earlier stages. In this research work, five medical databases are used and experimental results are computed using data mining software tool.

In February 2012, [16] the work is carried out through the Health Care Applications to diagnose different diseases using Red Blood Cells counting. In this work, the researchers presented the automatic process of RBC count from an image and some of the data mining techniques like Segmentation, Equalization and K-means clustering are used for preprocessing of the images. Some diseases which are related to sickness of RBC count also predicted.

3 .Noteworthy contribution in the field of proposed work.

[I] Empirical Study on Applications of Data Mining Techniques in Healthcare

The healthcare environment is generally perceived as being 'information rich' yet 'knowledge poor'. There is a wealth of data available within the healthcare systems. However, there is a lack of effective analysis tools to discover hidden relationships and trends in data. Knowledge discovery and data mining have found numerous applications in business and scientific domain.Valuable knowledge can be discovered from application of data mining techniques in healthcare system. In that study, we briefly examine the potential use of classification based data mining techniques such as Rule based, decision tree and Artificial Neural Network to massive volume of healthcare data. In particular we consider a case study using classification techniques on a medical data set of diabetic patients

[II] **Classifying Blood Donors Using Data Mining Techniques**

Data mining refers to extracting knowledge from large amount ofdata. Real life data mining approaches are interesting because they often present a different set of problems for data miners. The process of designing a model helps to identify the different blood groups withavailable stock in Indian Red Cross Society (IRCS) Blood Bank Hospital Classification techniques for analysis of Blood bank data sets. The availability of blood groups in blood banks is a critical and important aspect in a Blood bank. Blood banks are typically based on a healthy person voluntarily donating blood and used for transfusions or made into medications.

The ability to identify regular blood donors will enable blood bank and voluntary organizations to plan systematically for organizing blood donation camps in an efficient manner. The analysis had been carried out using a standard blood group donor's dataset and using the J48 decision tree algorithm implemented in Weka. The research work is used to classify the blood donors based on the sex, blood group, weight and age. This may be achieved through collecting the data utilizing the data mining technique and choosing the most suitable implementation tool for the domain.

we have described classification techniques for Blood Group Donors datasets. We have used data mining classifiers to generate decision tree. The primary focus of this research is the development of a system that is essential for the timely analysis of huge Blood Group Donors data sets. The traditional manual data analysis has become insufficient and the methods for efficient computer assisted analysis indispensable.

This technique will be applied to the blood group transfusion database maintained in the Indian Red Cross Society (IRCS) Blood Bank Hospital Chennai. This algorithm will be adapted to find conditions under which blood groups are frequently requested during emergency situations.

[III] Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool

The growing demand in the blood bank sector makes it necessary to exploit the whole potential of stored data efficiently. Data mining can contribute with important benefits to the blood bank sector, it can be a fundamental tool to analyze the data gathered by blood banks through their information systems. In this paper an attempt has been made to classify and predict the number of blood donors according to their age and blood group. J48 algorithm and Weka tool have been used for the complete research work.

The predict the number of blood donors of a particular age and blood group. The purpose of this work is to build a data mining model to extract knowledge of blood donor's classification to aid clinical decisions in blood bank centre. This study utilized real world data collected from an EDP

department of a blood bank centre and used J48 algorithm for the classification of donors, which can help the blood bank owner to make proper decisions faster and more accurately. Through training and evaluation, the experimental results showed that the generated classification rules performed well in the classification of blood donors, whose accuracy rate reached 89.9%.

[IV] Application of Data Mining Methods and Techniques for Diabetes Diagnosis

Medical professionals need a reliable prediction methodology to diagnose Diabetes. Data mining is the process of analysing data from different perspectives and summarizing it into useful information. The main goal of data mining is to discover new patterns for the users and to interpret the data patterns to provide meaningful and useful information for the users. Data mining is applied to find useful patterns to help in the important tasks of medical diagnosis and treatment. This project aims for mining the relationship in Diabetes data for efficient classification. The data mining methods and techniques will be explored to identify the suitable methods and techniques for efficient classification of Diabetes dataset and in mining useful patterns We have applied many classification algorithms on Diabetes dataset and the performance of those algorithms have been analysed. A classification rate of 91% was obtained for C4.5 algorithm.

.[V] Discretizing the Preprocessed Automated Blood Cell Counter Data Using Chi Merge Algorithm in Clinical Pathology

The preprocessing phases of the Knowledge Discovery in Data-bases to the automated blood cell counter data and creates discrete ranges of blood cell counter data that can be used in grouping data using classification, clustering and association rule generation. The functions of an automated blood cell counter from a clinical pathology laboratory and the phases in KnowledgeDiscovery in Databases are explained briefly. Twelve thousand records are taken from a clinical laboratory for processing. The preprocessing steps of the KDD process are applied on the blood cell counter data. the Chi Merge algorithm on the blood cell counter data and generates discretized data representing ranges of values for the data.

A brief study of Blood Cell Counter and Blood Cell Counter data is presented in the paper. The blood cell counter data was analyzed and few attributes were se-lected for processing, based on the knowledge given by the Clinical Pathologist. The KDD steps namely Data Cleaning, Integration, Selection, Transformation, and Mining were explained and were applied on the Blood Cell Counter Data to convert the raw data into a transformed data that was used for generating knowledge from the system. The data is discretized using Chi Merge algorithm.

[VI] Classifying Blood Donors Using Data Mining Techniques

Blood donation dataset belongs to blood transfusion organization of Birjand which contains 1998 sample with 6 attributes. This dataset gathered during months of December 2012 and February, April and May 2013. According the aim of this research, we select clustering methods for identifying the blood donor segmentation and analysis them. Clustering methods help discover groups of data records with similar values or patterns. These techniques are used in marketing (customer segmentation) and other business applications [11]. Clementine offers three clustering methods namely Kohonen networks, k-means clustering and two-step clustering that in this paper, we use k-means and two-step clustering methods for blood donor segmentation.

K-means clustering is a relatively quick method for exploring clusters in data] and can be used when we don't know what distinct groups are at the beginning. The k-means method aim is to minimize the sum of squared distances between all points and the cluster center. In this algorithm user sets the number of clusters (k) and each data record is then assigned to the nearest of the k clusters. This procedure is typically run several times because the user must set k and algorithm runs for each k. Therefore, this algorithm is an iterative algorithm .

Two-step clustering will automatically select the number of clusters. The user specifies a range Minimum and Maximum for the number of clusters. This method perform on two steps that in the first step, all records are classified into pre-clusters. In the second step, a hierarchical agglomerative cluster method is used to successively combine the pre-clusters].

After selective clustering methods introduced, we run the k-means clustering algorithm on under review dataset with 2, 3, 4, 5 and 6 values for number of clusters. The dataset variables are age, blood donation status, blood group, gender, highest education background and marital status. The aim of clustering available dataset is identifying the blood donor behavior. After running k-means clustering algorithm, we must calculate the optimal number of clusters which we use the Dunn's index for it. Dunn’s Validity Index attempts to identify those cluster sets that are compact and well separated which equation defines it. The aim of Dunn's index is to maximize the inner cluster distance and minimize the outer cluster distance. If obtained value for Dunn's index is large then it is better.

[VII] Study on Applications of Data Mining Techniques in Blood group

[1] P.Ramachandran& et al is discussed about the blood donors . the analyses the different blood groups data set. They are explaining different blood group type and dataset and preprocessing data. The dataset are attribute, instance, numeric nominal and class. The weka tools are used.

[2] Devchand J.Chaudhari is discussed the blood platelet transfusion is successful or unsuccessful transaction many attributes there are age, sex, illness, temperature, weight, patient HLA type, donor HLA type. The classification rules are used, and the algorithms are rule indication, decision tree, and ANN, and Backpropogation.

[3] Arpita GUPTA is discussed the platelet are count in white blood cells and the leukemia is suffer from the low platelet. The platelet measurement is specific gravity of blood of blood serum.WBC and MEMS .The matlab tool are used.

[4] Harleen kaur is discussed about the healthcare environment is understanding the „information rich" and yet „knowledge poor". the rule indication, decision tree, and artificial neural network in massive volume in healthcare application. the analyze the children diabetes mellitus and diabetes insipidus .The concept of the classification method is used in study in healthcare.

[5] Zhenghao Shi &el is discussed the applicationof neural network in medical pre processing. The neural network is used to the preprocessing, image segmentation, object detection and recognition. the neural network is have several disadvantage is compared to another techniques. First, it ability to express qualitative knowledge and network topological structure. Second collect large number of abnormal cases for training is very difficult to the CAD scheme.

[6] John N.Weinstein is discussed the neural computing is using the cancer drug development .The neural network is used to the various areas are biomedical science. The cancer drug development is discussed about the neural network.

[7] Ankit Bhardwaj is discussed about the data mining and current trends are associated. The classification, clustering and association rule are used blood bank sector.

II PROPOSED METHODOLOGY DURING THE TENURE OF THE RESEARCH WORK..

It is proposed to perform the following methodology for the research work. The basic approach that we are going to follow during the tenure of our research is given as below:

- Literature survey.
- Narrowing down to a specific problem.
- Developing new algorithms for improving efficiency.
- It is proposed to develop a new framework for integrated method for CBIR.
- Validating our approach by implementing our algorithms and testing them on real/ synthetic database images

III IMPLEMNATION ALGORITHMS

data mining algorithms identified by the: C4.5, *k*-Means, SVM,Apriori, EM, PageRank, AdaBoost, *k*NN, Naive Bayes, and CART. These algorithms are among the most influential data mining algorithms in the research community.With each algorithm, we provide a description of the algorithm, discuss the impact of the algorithm, and reviewcurrent and further research on the algorithm. These algorithms cover classification

Artificial Neural Network (ANN) is used the medical image preprocessing and medical image object detection and recognition. The neural network is effectively making the information. and etc.

1.1 CLASSIFICATION

Classification is the predefined groups or classes. Classification is the guessing the data and it is the algorithms are requiring to the class and define to data attribute values. The blood platelet transfusion is using the classification techniques are rule induction, decision tree, neural network and back propagation.

1.2 CLUSTERING

Clustering techniques are using the algorithm are K Nearest neighbor means algorithm, and k medoid algorithm. Clustering is the grouping the data item.

1.3 ASSOCIATION RULE

Association rule is created and analyzing the data .the if /then patterns are used to the two relationship . There are support and confidence .support is how the data items are frequently seem in database. Confidence is how many times the if/then statement is true.

1.4 REGRESSION

Regression is the mathematical function it is find a function and which the models of data with least error.Rule – induction In the rule induction is the using if/then patterns .The rules are used in two parts .There are antecedent (if part)and consequent (then part). Example of the platelet transfusion of successful or unsuccessful in IF-THEN Rule induction.

IV. EXPECTED OUTCOME OF THE PROPOSED WORK.

Expected outcome of our proposed research work will be as follows:

- New algorithms and framework for image retrieval by topological and geometrical models techniques.
- Development of efficient algorithms for image retrieval techniques.
- Comparison of the various existing schemes.
- Efficiency will be in the form of
  - Mathematical models requirement.

- Less number of comparisons among test and query images.
- Less computing complexity.

❖ Experimental results to show validity and efficiency of our proposed methods.

The overview of data mining and its techniques which have been used to extract interesting patterns and to develop significant relationships among variables stored in a huge dataset. Data mining is needed in many fields to extract the useful information from the large amount of data. Large amount of data is maintained in every field to keep different records such as medical data, scientific data, educational data, demographic data, financial data, marketing data etc. Therefore, different ways have been found to automatically analyze the data, to summarize it, to discover and characterize trends in it and to automatically flag anomalies. The several data mining techniques are introduced by the different researchers. These techniques are used to do classification, to do clustering, to find interesting patterns. In our future work, the data mining techniques will be implemented on blood donor's data set for predicting the blood donor's behaviour and attitude, which have been collected from the blood bank center.

Our focus the blood platelet transfusion is used to classification rule. The blood platelets transfusions are transforming to successful or unsuccessful in human body and the cancer patient dataset are used sometimes the dataset used in the diabetes patient.

the main purpose of the system is to guide diabetic patients during the disease. Diabetic patients could benefit from the diabetes expert system by entering their daily glucoses rate and insulin dosages; producing a graph from insulin history; consulting their insulin dosage for next day. It's also tried to determine an estimation method to predict glucose rate in blood which indicates diabetes risk. In that work, the WEKA data mining tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared.

Our focus applied to blood type classification, diagnosing diabetic symptoms, classifying blood donor type and diagnosing cancer patients based on blood cells or predicting the cancer types on the blood groups.

Our main goal the data mining techniques will be implemented on blood donor's data set for predicting the blood donor's behavior and attitude, which have been collected from the blood bank center.

V. REFERENCES

[1] Badgett RG: How to search for and evaluate medical evidence. Seminars in Medical Practice 1999, 2:8-14, 28.

[2] Jaiwei Han and Micheline Kamber, 'Data Mining Concepts and Techniques', Second Edition, Morgan Kaufmann Publishers.

[3] 'Data Mining Introductory and Advanced Topics' by Margaret H. Dunham.

[4] Sunita B Aher, LOBO L.M.R.J., International Conference on Emerging Technology Trends (ICETT)-2011 Proceedings published by International Journal of Computer Applications (IJCA).

[5] B.G. Premasudha et al., 'An Application to find spatial distribution of Blood Donors from Blood Bank Information',Vol. II, Issue No.2, July-December 2009.

[6] Jian Mu et al., 'Segmentation, Reconstruction, and analysis of blood thrombus formation in 3 D-2photon microscopy images', EURASIP Journal on Advances in Signal Processing, 10 July 2009.

[7] Abdur Rashid Khan et and al. 'Web based Information System for Blood Donation', International Jounal of digital content Technology and its applications,Vol. 3, Issue No.2, July 2009.

[8] G. Satyanaryana Reddy et al; 'Management Information System to help Managers for providing decision making in any organization'.

[9] T. Santhanm and Shyam Sunderam: 'Application of Cart Algorithm in Blood donor's classification', Journal of computer Science Vol. 6, Issue 5.

[10] Zhangetal. 'PPLook: an automated data mining tool for protein–protein interaction', BMC Bio-informatics- 2010

[11] Dr. Varun Kumar, Luxmiverma; 'Binary classifiers for Health Care databases - A comparative study of data mining classification algorithms in the diagnosis of Breast cancer', IJCST, Vol. I, Issue 2, December 2010.

[12] Vikram Singh and Sapna Nagpal; 'Interactive Knowledge discovery in Blood Transfusion Data Set'; VSRD International Journal of Computer Science and Information Technology; Vol. I, Issue 8, 2011.

[13] Wen-ChanLee and Bor-Wen Cheng; 'An Intelligent system for improving performance of blood donation', Journal of Quality, Vol. 18, Issue No. II, 2011.

[14] P.Ramachandran et al. 'Classifying blood donors using data mining techniques'; IJCST, Vol. I, Issue1, February 2011.

[15] F. Zabihi et al. 'Rule Extraction for Blood donators with fuzzy sequential pattern mining'; The Journal of mathematics and Computer Science; Vol. II, Issue No. I, 2011.

[16] Shyam Sundaram and Santhanam T: 'A comparison of Blood donor classification data mining models', Journal of Theoretical and Applied Information Technology, Vol.30, No.2, 31 August 2011.

[17] Shyam Sundaram and Santhanam T; 'Real Time Blood donor management using Dash boards based on Data mining models', International Journal of Computing issues, Vol.8, Issue5, No.2, September 2011.

[18] Prof. Dr. P.K. Srimani et al. 'Outlier data mining in medical databases by using statistical methods', Vol.4, No.1, January 2012.

[19] Alaa hamouda et al; 'Automated Red Blood Cells counting', International Journal of Computing Science, Vol.1, No.2, February 2012.

[20] Ivana D. Radojevic et al. 'Total coliforms and data mining as a tool in water quality monitoring', African Journal of Microbiology Research, Vol. 6(10), 16 March 2012.

[21] P.Yasodha, M. Kannan; Analysis of a Population of Diabetic Patients Databases in Weka Tool; International Journal of Scientific & Engineering Research Volume 2, Issue 5, May 2011.