Prediction of Blood Donors Population Using Data Mining Classification Technique

Aman Paul[1], Daljeet Singh[2]

[1]Assistant Professor, Department of CSE, JCDM College of Engineering, Sirsa, Haryana, India

[2]M.Tech. Scholar, Department of CSE, JCDM College of Engineering, Sirsa, Haryana, India

ABSTRACT

Data mining is a technique that finds relationships and trends in large datasets to promote decision support. Classification is a data mining technique that maps data into predefined classes often referred as supervised learning because classes are determined before examining data. Different classification algorithms have been proposed for the effective classification of data. Among others, Weka is an open-source data mining software with which classification can be achieved. It is also well suited for developing new machine learning schemes. It allows users to quickly compare different machine learning methods on new datasets. It has several graphical user interfaces that enable easy access to the underlying functionality. CBA is a data mining tool which not only produces an accurate classifier for prediction, but it is also able to mine various forms of association rules. It has better classification accuracy and faster mining speed. It can build accurate classifiers from relational data and mine association rules from relational data and transactional data. CBA also has many other features like cross validation for evaluating classifiers and allows the user to view and to query the discovered rules.

Keywords : Data mining, classification algorithms, Weka tool, CBA tool, classification accuracy, error rate.

I. INTRODUCTION

Data mining derives it meaning from the similarities between searching for valuable information in a large database and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material or intelligently probing it to find where the value resides. Basically, Data mining refers to "using various techniques to identify valuable information or decision-making knowledge in large volume of data, and extracting these in such a way that they can be used in the areas like decision making, prediction, forecasting and many others. In simple words, Data mining is the extraction of hidden

predictive information from large databases. Data mining scours databases for hidden patterns, finding predictive information that experts may miss, as it goes beyond their expectations.

Literature review: Sundaram and Santhanam, 2019 Decision trees derived from data mining modelling techniques are used to examine the blood donor's classification. The primary goal of this classification model is provide the capability to determine voluntary blood donor ship based on blood donation patterns. In specific, a comparison is made of two models (one based on a specific donation drive versus the regular voluntary donor patterns) based on a standard data set for blood transfusion.

Lee and Cheng, 2019 Due to advancement in the clinical medicine, the need for blood is growing day by day but the number of voluntary blood donors is steadily decreasing in recent years. The proposed work developed a system employing clustering and classification algorithms to determine the disparities in blood donation behaviour among the present donors and predict their intentions towards donation to understand the problems and to increase voluntary blood donation frequency. This study takes an empirical approach and developed an intelligent system to identify bidding behaviours and shows that data mining techniques may be used to enhance the results.

Bhardwaj *et al.*, 2012 Besides other areas, Blood Bank sector also employs the data mining techniques and their implementation. 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.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 analyse the data, to summarize it, to discover and characterize trends in it and to automatically flag anomalies.

DATASET USED: For the blood donors' prediction I have used blood donors' dataset from the blood bank of IGMC hospital (Shimla) with Weka tool and CBA tool. Table 4.1 shows our selection of dataset and its attributes. The dataset was analyzed at 1000 instances with 7 attributes in the area of Health Science and there are no missing values.

Table 4.1: The dataset used for the experiments and its attributes

Data Set	Instance	Attributes	Area	Missing Value
Blood Donors	1000	7	Health Science	0

For classification algorithms, I have explored blood donors' dataset and tried to figure out which classification technique has the best accuracy rate & least error rate for the prediction of blood donors through their blood group, age & weight. In Weka, results are computed with two applications i.e. Explorer & Experimenter. In Explorer, accuracy rates & errors rates are computed for different algorithms under three Test Modes namely Training Mode, Cross Validation Mode (10 Folds) and Percentage Split Mode (66%). In Experimenter, different algorithms are compared and their accuracy is obtained at once. Error Rates are obtained as higher the accuracy rate; least will be the error rate. By comparing the results of Explorer & Experimenter in Cross Validation Mode (10 Folds), we found out that in some cases their results are slightly different. So, we preferred to consider the results of Experimenter in order to make comparison with CBA (Classification Based Association). A comparison is also done among PART, J48 and CBA algorithm on the basis of varying

confidence factors. The results of CBA are computed directly with the help of CBA tool.

II. REFERENCES

[1]. Amarendra K, Lakshmi K. V. and Ramani K. V. "Research on Data Mining Using Neural Networks",International Journal of Computer Science & Informatics, vol. 2, 1, pp. 1-8.

[2]. Antunes C and Oliveira A. L. "Sequential Pattern Mining with Approximated Constraints", Department of Information Systems and Computer Science, Instituto Superior Tecnico / INESC-ID, Lisboa, Portugal.

[3]. Bhardwaj A, Sharma A and Shrivastava V.K. (2012) "Data Mining Techniques and their Implementation in Blood Bank Sector –A Review", International Journal of Engineering Research and Applications, vol. 2, 4, 2012, pp. 1303-1309.

[4]. Bhargavi Pand Jyothi S (2009) "Applying Naive Bayes Data Mining Technique for Classification of Agricultural Land Soils", International Journal of Computer Science and Network Security, vol. 9, 8, 2009, pp. 117-122.

[5]. Boonyanusith W and Jittamai P (2012) "Blood Donor Classification Using Neural Network and Decision Tree Techniques", Proceedings of the World Congress on Engineering and Computer Science (WCECS- 12), 24-26 October 2012, Member, IAENG San Francisco, USA.

[6]. Bramer M (2007) "Automatic Induction of Classification Rules from Examples Using N-Prism", Faculty of Technology, University of Portsmouth, Portsmouth, UK.

Cite this article as :

Sh