



# Banking Data Mining Using Enhances Classification Technique to Improve Correctly Classified Instances

Dr. Gurpreet Singh<sup>1</sup>, Er. Rajwinder Kaur<sup>2</sup>

Professor & Head<sup>1</sup>, M. Tech Scholar<sup>2</sup>

Department of Computer Science & Engineering

St. Soldier Institute of Engineering & Technology, Jalandhar, Punjab, India

## Abstract:

This research delineates a comprehensive and successful application of classification to large banking data set of different banking services obtained by numbers of customers. Complex interaction effects among banking services that lead to increased policy variability have been detected. The extracted information has been confirmed by the database managers, and used to improve the decision process. The research suggests that classification may be particularly useful when data is multidimensional, and the various process parameters and highly complex interactions. In order to classify and identify effective and beneficial saving service and design the appropriate criteria for selecting the right scheme for different persons having different taste, this study developed a data mining framework for analyzing banks data, in which suitable technique is employed to extract rules between present saving schemes. In other words, the objectives of this thesis are

1. To implement self-developed programme, CHAID AND ID3 for mining large-scale Banking datasets
2. To enhance the efficiency with a new enhanced self-programmed algorithm that compares the ID3 and CHAID distance based algorithm.
3. To eliminate error rate.
4. To classify the available saving services of banks and post office to good, medium and bad level.
5. To select the best saving service according to the investor's choice and its preference.
6. To guide the potential investor to invest his money in the particular scheme so as to get more benefits.
7. To help to take the right decision for investment.
8. To reduce the time to take particular decision as there will be no need to analyse each and every available investment scheme thoroughly.

## I. INTRODUCTION

In field of Information technology we have huge amount of data available that need to be turned into useful information. With the enormous amount of data stored in files, databases, and other repositories, it is increasingly important, if not necessary, to develop powerful means for analysis and interpretation of data and for the extraction of interesting knowledge that could help in decision-making. Data mining is a process of identifying useful, valid, novel and understandable pattern in data. It is defined as extracting the information from the huge set of data or we can say that data mining is mining the knowledge from data [6]. Data mining is used for a variety of purposes in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales. For example, the insurance and banking industries use data mining applications to detect fraud and assist in risk assessment (e.g., credit scoring). Using customer data collected over several years, companies can develop models that predict whether a customer is a good credit risk, or whether an accident claim may be fraudulent and should be investigated more closely.

## II. DECISION TREES

There is assortment of algorithms presence used in grouping procedure. One if these are the DT attitude. To characterize

together the worsening prototypes and classifiers result tree in the ceremonial of predicative exemplary is charity. Verdict tree really us the hierarchal prototypical of resolutions and their magnitudes.

The arrangement of DT excludes subdivision, root knob and leaf swelling. Aspects test is designated on each wait node, the test result is signified by subdivision and course labels are exposed by leaf node. The uppermost node is the root node of the sapling. The tree scholarship is completed by dividing the spring into agreed which are commonly created on an assessment of feature worth. The top depressed method of DT sets an example of acquisitive algorithm. Apart from this bottom-up tactic is also mutual these days.

### Many decision tree algorithms are as follows:

- I. ID3 (Iterated dichotomies 3)
- II. C4.5 algorithm
- III. CART
- IV. CHAID (Chi-squared automatic interaction detector)
- V. MARS
- VI. J48
- VII. ID3 and CART follow same approach for decision tree learning from training tuples but both are invented independently.

### III. PROBLEM FORMULATION

The future knowledge is very well matched for several reasons

1. Improved decision tree algorithm which will effort on big scale high dimensional dataset- here is a difficult of records mining in the classification of big datasets. There is no such algorithm stated that performs well in this problem. An algorithm can be ended with definite splitting collection approaches elaborate from the literature which contains algorithms like C4.5 and CART.
2. Enhancement in the efficiency of decision tree construction-many clipping methods is proposed which can support in the enhancement of decision tree structure.
3. Analysis between the computation times- the computation time can be reduced by making alterations in the number of node and leaves. Lesser the no of nodes lesser will be the computation time of the algorithm.
4. Reducing present error rate- the errors rates produced by a predictive model can be reduced by the algorithms. Basically error rate is one minus accuracy.

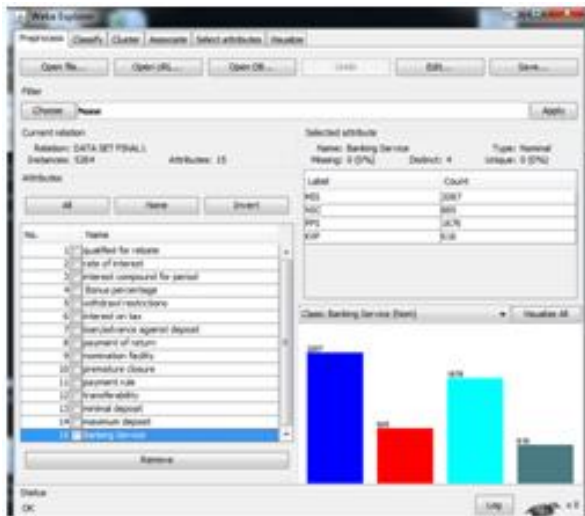
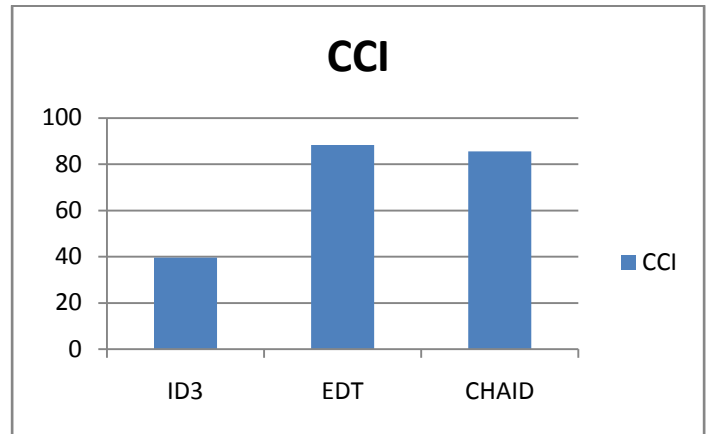
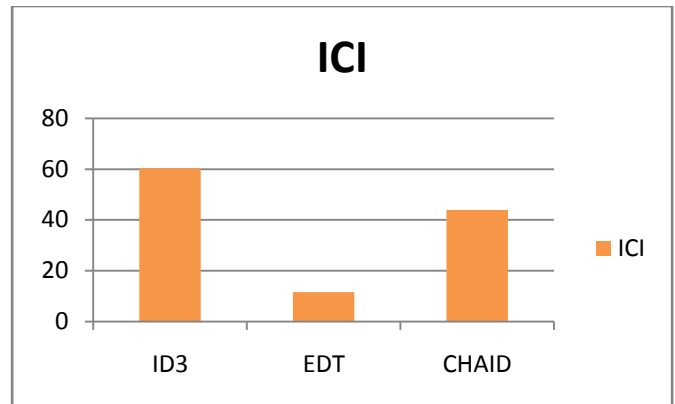


Figure.1. shows the main WEKA Explorer interface with the dataset loaded. The last attribute Banking service is taken as a class attribute by the WEKA. This attribute contains four bank services. The count of number of instances under each service in the dataset is shown numerically as well as graphically.

	ID3	EDT	CHAID
CCI	39.64	88.35	85.6



	ID3	EDT	CHAID
ICI	60.35	11.64	43.95



	ID3	EDT	CHAID
ERROR RATE	98.98	43.95	54.91

### IV. CONCLUSION

In this explore; learn is being accomplished on DT algorithms. The forms of CHAID and ID3 DT algorithms are common and a new algorithm EDT is projected. The evaluation of projected algorithm is completed with the active algorithms ID3 and CHAID on banking dataset using WEKA measurements mining tool. The concern by altering the closely classified instances, incorrect classified instances and error rate doctrines identify that the projected process gives improved result than ID3 and CHAID by falling the error rate which signifies that EDT have elevated intra comparison and is additional precise. Also the expected algorithm can touch big datasets more successfully.

### V. REFERENCES

[1]. V.S.Jagadeeswaran, P.uma, "Hierarchical Algorithm for big Datasets" 2013 International periodical Of Advanced Explore in Computer and message Engineering

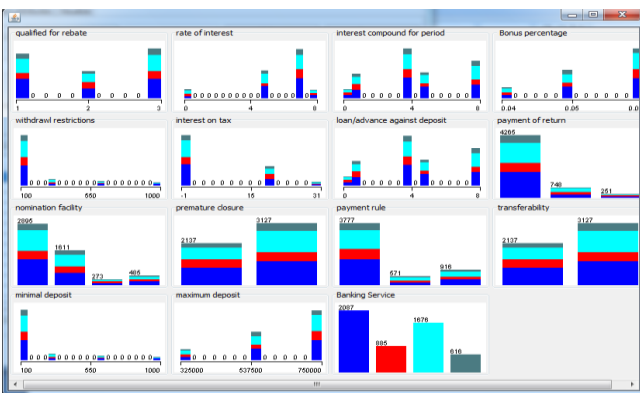


Figure.2. shows the graphical representation of all the 15 attributes of the banking dataset. Each bar shows the distribution of four service of class attribute on that particular value.

[2]. Suwimon Vongsingthong, Naw Aporn Wisitpongphan, "categorization Of University Students' Behaviours in contribution in order on Face book" *2014 IEEE 2014 11th International Joint meeting on Computer Science and Software Engineering*

[3]. Nidal Ismael, Mahmoud Alzaidan, Wesa mAshour, "better Multi Threshold Birch Clustering Algorithm" *2014 worldwide Journal of Artificial Intelligence and request for Smart Devices*.

[4]. K.Kameshwaran, K.Mal arvizhi, "review on Clustering method in Data Mining" *2014 International Journal of Computer discipline and in order Technologies*.

[5]. T. Zhang, R. Ramakrishnan, M. Li, "BIRCH: a well-organized data clustering technique for very big databases" (1996) *ACM SIGMOD International Conference on organization of Data*.

[6]. Du Hai Zhou, Li Yong Bin, "An enhanced BIRCH Clustering Algorithm and request in Thermal Power" *2010 IEEE Web information scheme and Mining*.

[7]. Jiang Sheng Yi, Li Xia, "enhanced BIRCH clustering algorithm", *2009 Journal of Computer function*.

[8]. D. Napoleon, P. G. Laxmi, "A competent K-Means Clustering Algorithm for plummeting Time difficulty using consistent Distribution statistics Points", *2011 IEEE Trends in order science and compute*.