

# Data Classification Using Data Mining Techniques

S S Lakshmi Lavanya M, Gilla Keerthi Kumar, Sireesha Balineni, G Sruthi  
Assistant Professors, Sree Dattha College of engineering and science, Hyderabad, India.

**Abstract** – Data Classification Using Data mining techniques is used for classify the data. Classification is performed on the input data and returns a classifiers tree as output. It is a two step process-In first step; a classifier is built describing a pre determined set of data classes or concepts. The model is constructed by analyzing data mining tuples described by attributes. In second step, the classifier is used for classification. The predictive accuracy of the classifier is estimated. It has been tested on different samples and was observed that the tuples are successfully classified. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information.

Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions and categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining contains the following procedure-Extract, transform, and load transaction data onto the data warehouse system. Store and manage the data in a multidimensional database system. Provide data access to business analysts and information technology professionals. Analyze the data by application software. Present the data in a useful format, such as a tree or table. Data classification technique is capable of processing a wider variety of data and is growing in popularity. Data Classification involves the following procedure that is user gives the trained set data to system, then system takes this data and create entropy set and then convert to data set and then select root node and then analyze all remaining nodes, then generate a decision tree. This process involves the following modules.

**Index Terms** – Data Mining, classification, tuples, multidimensional data base.

## 1. INTRODUCTION

Data Classification Using Data mining techniques is used for classify the data. Classification is performed on the input data and returns a classifiers tree as output. It is a two step process-In first step; a classifier is built describing a pre determined set of data classes or concepts. The model is constructed by analyzing data mining tuples described by attributes. In second step, the classifier is used for classification. The predictive accuracy of the classifier is estimated. It has been tested on different samples and was observed that the tuples are successfully classified. Data classification technique is capable of processing a wider variety of data and is growing in popularity. You'll also find output that is much easier to

interpret. By using this technique you'll receive a decision tree that requires a series of binary decisions.

Data Classification involves the following procedure that is user gives the trained set data to system, then system takes this data and create entropy set and then convert to data set and then select root node and then analyze all remaining nodes, then generate a decision tree. This process involves the following modules.

As we are growing in terms of population, technology advancements, literary rate and globalization, education systems are also taking a new shape as business systems. Now Data Mining has become universal tool for strategic management in business organizations as well as social system and organizations. Today out of all infrastructural support for the development of society, education is considered as one of the key inputs for social development. This paper analyzes the data available on student's academic record and student likelihood in terms of placement may be predicted on the basis of entrance examination marks, quantitative ability marks and verbal ability marks by using.

In existing systems jtrees are not in use for building classifier trees .They are using different structures to build classifier trees. It is not in such a way to understand easily for common people. Good security is not provided by existing system.

## 2. DATA MINING

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions and categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.Data mining contains the following procedure-Extract, transform, and load transaction data onto the data warehouse system. Store and manage the data in a multidimensional database system. Provide data access to business analysts and information technology professionals. Analyze the data by application software. Present the data in a useful format, such as a tree or table.

To overcome this present existing system we are implementing jtrees for easy understanding. Using this system any one can understand at any time easily.

### 3. DECISION TREE

Decision tree is one of the models for classifying the given set of data. This model can be design using a technique called Decision Tree Induction. A tree is a collection of nodes for selecting an attribute at a node. We have to calculate the information gain for every attribute and the attribute with the more information gain is placed at that node. For information gain, we subtract the value of entropy from the value of expected information.

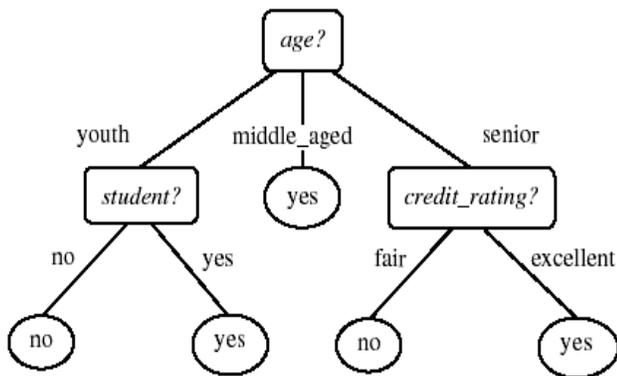


Fig1.2. A decision tree for the concept buys computer, indicating whether a customer at All Electronics is likely to purchase a computer. Each internal (nonleaf) node represents a test on an attribute. Each leaf node represents a class (either buys computer = yes or buys computer = no).

A tree-shaped structure that represents a set of decisions. These decisions generate rules for the classification of a dataset. Tree where internal nodes are simple decision rules on one or more attributes and leaf nodes are predicted class labels. This technique is capable of processing a wider variety of data and is growing in popularity. You'll also find output that is much easier to interpret. By using this technique you'll receive a decision tree that requires a series of binary decisions.

### 4. SYSTEM SPECIFICATIONS

The system specification is a document that serves as the foundation for hardware engineering, software engineering, database engineering and human engineering .It describes the function and performance of a computer based system and the constants that will govern its development. The specification bounds are each allocated system elements.

The software requirement specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing an indication of performance requirements and design constraints, appropriate validation criteria and other data patient to requirements.

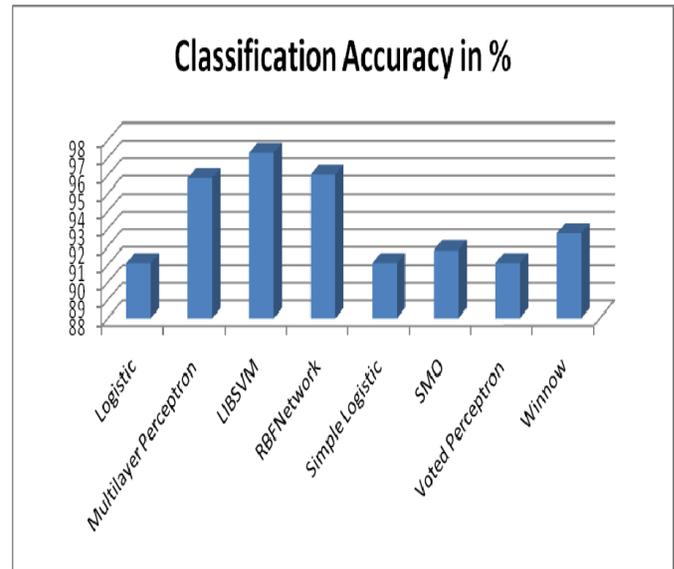


Fig.1. Classification methods

### 5. IMPLEMENTATION

The most crucial phase of any project is the implementation. This includes all those activities that take place to convert from the old system to the new system. It involves setting up of the system for use by the concerned end user. A successful implementation involves a high level of interaction between the analyst, programmers and the end user. The most common method of implementation is the phased approach, which involves installation of the system concurrently with the existing system. This has its advantage in that the normal activity carried out, as part of the existing system is anyway hampered. The end users are provided with sufficient documentation and adequate training in the form of demonstration/presentation in order to familiarize with the system.

The working of the system was under observation for a period of two days after implementation and it was found to the hassle free. The feedbacks from the users are awaited.

### 6. CONCLUSION

“Data Classification Using Data Mining Techniques” has been developed after a detailed study of the existing manual system. After the completion of the coding of the projects, the developed application was tested on different kinds of tables with different inputs to check the accuracy of the data manipulation from single window and their associated outputs. This software Tool is advantageous and highly useful, overcoming the conventional techniques, which were not easy.

Moreover, it takes lot of time to learn how to use a console to each and every database and start working on large amount of data. The software developed eradicates all those hurdles and easy to use and satisfies all the necessary formalities. It has

been developed with systematic design principles confirming to the step of the system development life cycle. The system has been modularly developed and implemented interrelation between the modules and thereby overcoming the drawbacks of the manual system. The software is user friendly for both the users and the administrator.

In future we can also perform classification on images, we can implement web application for the classification so, multiple users can access at a time. Classifier tree can also be visualized using advanced graphics. We can also display the total no of sub nodes that are occurring under each node.

### REFERENCES

- [1] Allen, M.W., Armstrong, D.J., Reid, M.F., Riemenschneider, C.K. (2009). "IT Employee Retention: Employee Expectations and Workplace Environments", SIGMIS-CPR'09, May 2009, Limerick, Ireland.
- [2] Al-Radaideh, Q. A., Al-Shawakfa, E.M., Al-Najjar, M.I. (2006). "Mining Student Data Using Decision Trees", International Arab Conference on Information Technology (ACIT 2006), Dec 2006, Jordan.
- [3] Chein, C., Chen, L. (2006) "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", Expert Systems with Applications, In Press.
- [4] Cho, S., Johanson, M.M., Guchait, P. (2009). "Employees intent to leave: A comparison of determinants of intent to leave versus intent to stay", International Journal of Hospitality Management, 28, pp374-381.
- [5] CRISP-DM, (2007). Cross Industry Standard Process for Data Mining: Process Model. <http://www.crisp-dm.org/process/index.htm>. Accessed 10th May 2007.
- [6] Delavari, N., PHON-AMNUAISUK S., (2008). Data Mining Application in Higher Learning
- [7] Dreher, G.F. (1982) "The Role of Performance in the Turnover Process", Academy of Management Journal, 25(1), pp. 137-147.
- [8] Han, J., Kamber, M., Jian P. (2011). Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers.
- [9] Ibrahim, M.E., Al-Sejini, S., Al-Qassimi, O.A. (2004). "Job Satisfaction and Performance of Government Employees in UAE", Journal of Management Research, 4(1), pp. 1-12.
- [10] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010a) "Knowledge Discovery Techniques for Talent Forecasting in Human Resource Application", International Journal of Humanities and Social Science, 5(11), pp. 694-702.
- [11] Jantan, H., Hamdan, A.R. and Othman, Z.A. (2010b). "Human Talent Prediction in HRM using C4.5 Classification Algorithm", International

Journal on Computer Science and Engineering, 2(08-2010), pp. 2526-2534.

- [12] Karatepe, O.M., Uludag, O., Menevis, I., Hadzimehmedagic, L., Baddar, L. (2006). "The Effects of Selected Individual Characteristics on Frontline Employee Performance and Job Satisfaction", Tourism Management, 27 (2006), pp. 547-560.
- [13] Kayha, E. (2007). "The Effects of Job Characteristics and Working Conditions on Job Performance", International Journal of Industrial Ergonomics, In Press. [14] Khilji, S., Wang, X. (2006). "New evidence in an old debate: Investigating the relationship between HR satisfaction and turnover", International Business Review, In Press.
- [14] Lavrac, N. (1999). "Selected Techniques for Data Mining in Medicine", Artificial Intelligence in Medicine, 16(1999), pp. 3-23.

### Authors



Ms. S S Lakshmi Lavanya M received Master Degree in Computer Science and Engineering (CSE) form Jawaharlal Nehru Technological University, Hyderabad (JNTUH). Her research interest includes Big Data and Data Mining. Presently working as Assistant Professor in CSE Department, Sree Dattha College of engineering and science, Hyderabad., India.



Mr. Gilla Keerthi kumar received Master Degree in Computer Science and Engineering (CSE) form Jawaharlal Nehru Technological University, Hyderabad (JNTUH). His research interest includes ruby and rails. Presently working as Assistant Professor in CSE Department, Sree Dattha College of engineering and science, Hyderabad., India.



Ms. Sireesha Balineni received Master Degree in Computer Science and Engineering (CSE) form Jawaharlal Nehru Technological University, Kakinada(JNTUK). Her research interest includes Network Security, Cloud Computing. Presently working as Assistant Professor in CSE Department, Sree Dattha College of engineering and science, Hyderabad., India.



Ms. G Sruthi received Master Degree in Computer Science and Engineering (CSE) form Jawaharlal Nehru Technological University, Hyderabad (JNTUH). Her research interest includes Data Mining and Network Security. Presently working as Assistant Professor in CSE Department, Sree Dattha College of engineering and science, Hyderabad., India.