

# Data Mining Classification Algorithm for E-Testing and Analysis Processing

**M.Archana**

*School of Computing Sciences, Vel's University,  
Pallavaram, Chennai-600117*

**Veeraraghavan**

*Assistant Professor,  
School of Computing Sciences, Vel's University,  
Pallavaram, Chennai-600117*

**Abstract-**This paper focuses on developing to bring up various features for the distance learning system. There are various modules in this system, to provide extensive evaluation for distance learners . Questions are randomly chosen from the given set of questions in a random manner.

The questions are taken from the database which has extensive student information system where the queries are formulated.

The many number of students are attended test, here we use classification algorithm (C4.5) to classify the Students based on mark .the C4.5 algorithm which improvement from the predecessor ID3 algorithm .C4.5 Algorithm for test. It has to be both exact and comprehensible for instructor is to be used to make decision.

Data mining, the extraction of hidden predictive collection from large databases. It is a powerful technology to help companies with great potential and it is focus on the important information in their data warehouses.

**Keywords:** Decision, predecessors, accurate, comprehensible, classification.

## INTRODUCTION

The ability to predict classify a student's performance is very important in web-based educational environments. Data Mining is used to attain this objective, In fact, the classification of E-learning is one of the important tasks in DM. Educational objectives is based on the classification and to find out the applications is misused by the students. Classification methods are several types. Artificial intelligent algorithms is to be applied to t student outcome, marks or scores.E learning and online application for the students are popular for the interaction and to cope up easily with the system.

Teaching challenges are increasing day by day with the growing enthusiasm shown by students towards online learning. One such challenge is an effective student evaluation mechanism for those taking these online examinations.

Testing is an important component of learner evaluation in the entire process of instruction. In the field of foreign language instruction, for example, placement test is often used to determine the language proficiency of a student at the beginning of his her language study.

Using this Online Exam system, we can conduct any type of objective type examinations on line. Even though I implemented most of the conventional things, the reader is highly appreciated to incorporate all other features whatever so.

The duration and no of questions per exam are set as 10 minutes and 10 questions respectively. But we can set them to any possible values.

Once user starts the exam, the question ids list will come from the server. For every request to the server for a question, a question id in the ills list stored in the client is fetched and sent to the server. The server will return the question content, corresponding to the question id. from the file.

**Case-based reasoning** (CBR) is a problem solving technique based on previous experience knowledge. The problem-solving life cycle in a ("BR system consists essentially of the following four parts. 1. Retrieving similar previously experienced cases (e.g. problem-solution-outcome triples) which problem is judged to be similar 2. Reusing the cases by copying or integrating the solutions from the eases retrieved 3. Revising or adapting the solution(s) retrieved in an attempt to solve the new problem 4. Retaining the new solution once it has been confirmed or validated.

Data mining techniques are the result of a long process of research and product development.

- Massive data collection
- Powerful multiprocessor computers
- Data mining algorithms

Commercial databases are growing at unprecedented rates. A recent META Group survey of data warehouse projects found that 19% of respondents are beyond the 50 gigabyte level, while 59% expect to be there by second quarter of 1996. In some industries, such as retail, these numbers can be much larger. The accompanying need for improved computational engines can now be met in a cost-effective manner with parallel multiprocessor computer technology. Data mining algorithms embody techniques that have existed for at least 10 years, but have only recently been implemented as mature, reliable, understandable tools that consistently outperform older statistical methods.

Support Vector Machines (SVMs). It is a supervised learning methods. A classification method, SVM is a global classification model and it generates non-overlapping partitions. Support Vector Machines is based on the attributes, but it does not consider the dependencies among attributes [2].

The approach of Support Vector Machines classifies to select the hyperplane .That has the margin.

The classification technique is a supervised learning technique. One example of classification would be Support

Vector Machines. Support Vector Machines are a method for creating functions from a set of labeled training data. The function can be a classification function (the output is binary: is the input in a category) or the function can be a general regression function.

For classification, SVMs operate by finding a hyper-surface in the space of possible inputs. This hyper-surface will attempt to split the positive examples from the negative examples.

Technical software like learning environment, e-mail-program, and browser are basic instruments for online education. Picture collections and information databases deliver material for developing learning materials and learning software with information can be used instead of self developed software. Software can be checked for abilities, disadvantages, ease of use, quality and so on.

Interesting is to check services like to hire a server or an online learning environment. Using ready materials or analyzing an existing course needs a view on quality, usability, relevant, combination and so on. Evaluating the learner means looking at learning type, quality of participation, success of learning... Looking at a teacher the way of teaching and the reaction to students can be tested. Instead of testing parts of a seminar like material or people, you can evaluate a seminar at whole. Different from a seminar together with other people a self-learning course needs other points of view.

The split will be chosen to have the largest distance from the hyper-surface to the nearest of the positive and negative examples, such that the margins are maximized. Intuitively, this makes the classification correct for testing data that is near, but not identical to the training data.

In order to successfully construct a classification model, there must exist a set of class identities or labels. This method of analysis is useful in applications such as customer profiling, credit rating, medical diagnostics, and chemical classification. This method is not particularly useful when no preexisting identity or labels are available.

**K-means** clustering is a very simple form of analysis that assigns each object from a collection to exactly one group. The technique begins with k cluster centers, where k is specified when the technique is applied.

For each object, the cluster center that's nearest to it is found. The object then is placed in the cluster represented by this cluster center. As objects are added to and removed from each cluster, the cluster centers are recomputed. The second step is repeated to reassign objects to its nearest cluster center. The cluster centers are recomputed. This iteration is repeated.

The algorithm terminates when no more clusters are altered. **Hierarchical** clustering also generates groups of objects. But each resulting group of size greater than one is in turn composed of smaller groups.

There are two approaches: bottom-up and top-down. The bottom-up approach involves several steps. We first start with n clusters given n objects, each containing one object. Then the two most similar clusters are merged into one

cluster and the previous Step is repeated until there is only one cluster. The top-down approach is the reverse of the bottom-up approach.

The technique of **Self-Organizing Maps** is somewhat related to k-means clustering. It is based on neural networks. This method produces ordered low - dimensional representations of an input data space. Typically, such input data is complex and high-dimensional with data elements being related to each other in a nonlinear fashion.

The basis of the method derives from biophysical studies of the brain and brain maps which form internal representations of external stimuli in a spatial manner. The mechanism of the technique is somewhat complex and will not be discussed here. Clustering is useful as a stand-alone tool for data analysis. It can provide useful insight in the data exploratory process. Clustering can also be used as a preprocessing step for data mining as seen in our study.

### CONCLUSION:

Thus the dissertation gives a clear description about the online test conducted by the College and also it helps the user to understand clearly about each module in the system. The user can easily understand the system and easy to take decision (using c4.5) and they can able to do their process. The system provide a way through which user can make the process and they can generate the reports based on their requirements.

### REFERENCES

- (1) Armengol. E.. Onta. S.. dan Pla/a. E.. *Explaining similarity in CBR*. Eva Armengol, Artificial Intelligence Research Institute (IIIA-CSIC). Campus UAB. 08193 Bellaterra. Catalonia
- (2) Badriyah. T.. Rahmawati. R.. Alat Bantu Klasifikasi dengan Pohon Keputusan unik Sistem Pendukung Keputusan. Proceedings: Seminar Nasional Aplikasi Teknologi Informasi 2006, Jurusan Teknik Informatika, Universitas Islam Indonesia Yogyakarta (2006)
- (3) Berry. Michael J.A.. Linoff. Gordon S.. *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management Second Edition*. Wiley Publishing, Inc.. Indianapolis. Indiana (2004)
- (4) Romero. C. Ventura. S. Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with Applications. 2007, 33(1), pp. 135-146.
- (5) Chen. G.. Liu. C.. Ou. K.. Liu. B. Discovering Decision Knowledge from Web Log Portfolio for Managing Classroom Processes by Applying Decision Tree and Data Cube Technology. Journal of Educational Computing Research 2000. 23(3). pp.305-332.
- (6) Clark. P.. Niblett. T. The CN2 Induction Algorithm. Machine Learning 1989. 3(4). pp.261-283 [9] Cocea. M.. Weibel/ahl. S. Can Log Files Analysis Estimate Learners' Level of
- (7) Yudelson. M.V.. Medvedeva. O.. Legowski. E.. Castine. M.. Jukic. D.. Rebecca. C. Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS. AAAI Workshop on Educational Data Mining. 2006. pp. 1-8.
- (8) Cocea. M.. Weibel/ahl. S. Can Log Files Analysis Estimate Learners' Level of Motivation? Workshop on Adaptivity and User Modeling in Interactive Systems. Hildesheim. 2006. pp.32-35.
- (9) Hamalainen. W.. Vinni. M. Comparison of machine learning methods for intelligent tutoring systems. Conference Intelligent Tutoring Systems. Taiwan. 2006. pp. 525-534.
- (10) Fausett. L.. Elwasif. W. Predicting Performance from Test Scores using Backpropagation and Counterpropagation. IEEE Congress on Computational Intelligence. 1994. pp.3398-3402.