Data Mining Application in Enrollment Management: A Case Study

Surjeet Kumar Yadav
Research scholar, Shri Venkateshwara University,
J. P. Nagar, (U.P.) India

Saurabh pal
Head, Dept. of MCA
VBS Purvanchal University, Jaunpur, India

ABSTRACT

In the last two decades, number of Higher Education Institutions (HEI) grows rapidly in India. This causes a cut throat competition among these institutions while attracting the student to get admission in these institutions. Most of the institutions are opened in self finance mode, so all time they feel short hand in expenditure. Therefore, institutions focused on the strength of students not on the quality of education. Indian education sector has a lot of data that can produce valuable information. Knowledge Discovery and Data Mining (KDD) is a multidisciplinary area focusing upon methodologies for extracting useful knowledge from data and there are several useful KDD tools to extract the knowledge. This knowledge can be used to increase the quality of education. But educational institution does not use any knowledge discovery process approach on these data. Now-a-day a new research community, educational data mining (EDM), is growing which is intersection of data mining and pedagogy.
In this paper we present the data mining method for enrollment management for MCA course.

General Terms

Educational Data Mining

Keywords

Data mining, Knowledge Discovery, Higher Education, Enrollment Management, ID3 Decision Tree.

1. INTRODUCTION

Quality education is one of the most promising responsibilities of any University/ Institutions to his students. Quality education does not mean high level of knowledge produced. But it means that education is produced to students in efficient manner so that they learn without any problem. For this purpose quality education includes features like: methodology of teaching, continuous evaluation, categorization of student into similar type, so that students have similar objectives, demographic, educational background etc.

Advent of computer opens a new era in the field of information because of high storage capability and complex study. Huge number of data can be organized in any order with the help of computer. Now we can explore a new knowledge on these data which was either impossible or a very time consuming process for a person [1], [2], [3].

Education sector has a lot of data in the form of student's information. Application of computer in the education can extract valuable information to provide quality education. Due to this combination of education and computer (data mining) a new research community is growing i.e. educational data mining [13].

Data mining, which is the science of filtering data for information and knowledge retrieval, has recently developed new album of applications and engendered an emerging discipline, called Educational Data Mining (EDM). EDM carries out tasks such as prediction (classification and regression), clustering, relationship mining (association, correlation, sequential mining, and causal data mining), distillation of data for human judgment, and discovery with models [4]. Moreover, EDM can solve many problems based on educational domain. Data mining is non-trivial extraction of implicit, previously unknown and potentially useful information from large amounts of data. It is used to predict the future trends from the knowledge pattern. Remarkable amount of EDM endeavors have been conducted and published in many journals and conference proceedings related to, but not limited to, Artificial Intelligence, Learning Systems, Education, and others.

The main objective of this paper is to use data mining methodologies to select student for enrollment in a particular course (MCA). Data mining provides many tasks that could be used to select good student for a particular course. In this research, the classification task is used to evaluate previous student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information like stream, marks in graduation, students performance etc. were collected from the student's management system, to predict the suitable student for enrollment in a particular course.

2. DECISION TREE

A decision tree is a flow-chart-like tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits, which test the value of an expression of the attributes. Arcs from an internal node to its children are labeled with distinct outcomes of the test. Each leaf node has a class label associated with it.

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

The four widely used decision tree learning algorithms are: ID3, CART, CHAID and C4.5.

ID3 algorithm introduced by J. R. Quinlan [5] is a greedy algorithm that selects the next attributes based on the information gain associated with the attributes. The attribute with the highest information gain or greatest entropy reduction is chosen as the test attribute for the current node.

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is a successor of ID3. C4.5 made a number of improvements to ID3. C4.5 uses Gain ratio [6] as an attribute selection measure. Also C4.5 can handle both discrete and continuous attribute.

CART (Classification And Regression Tree) algorithm, which was proposed by Breiman, is conceptually is same as that of ID3. The impurity measure used in selecting the variable in CART is Gini index [6]. If the target variable is nominal it generates classification tree and for continuous-valued numerical target variable it generates regression tree.

CHAID (CHi-squared Automatic Interaction Detector) uses Chi square contingency test for tree construction in two ways [7]. First, it determines whether levels in the predictor can be merged together. Once all predictor level is compressed to their smallest significant form, it determines most significant predictor in distinguishing among the dependent variable levels.

3. THE ID3 DECISION TREE

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan [5]. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric - information gain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

3.1 Measuring Impurity

Given a data table that contains attributes and class of the attributes, we can measure homogeneity (or heterogeneity) of the table based on the classes. We say a table is pure or homogenous if it contains only a single class. If a data table contains several classes, then we say that the table is impure or heterogeneous. There are several indices to measure degree of impurity quantitatively. Most well known indices to measure degree of impurity are entropy, gini index, and classification error.

$$Entropy = \sum_j - p_j \log_2 p_j$$

Entropy of a pure table (consist of single class) is zero because the probability is 1 and log (1) = 0. Entropy reaches maximum value when all classes in the table have equal probability.

$$Gini\ Index = 1 - \sum_j p_j^2$$

Gini index of a pure table consist of single class is zero because the probability is 1 and 1-12 = 0. Similar to Entropy, Gini index also reaches maximum value when all classes in the table have equal probability.

$$Classification\ Error = 1 - \max\{p_j\}$$

Similar to Entropy and Gini Index, Classification error index of a pure table (consist of single class) is zero because the probability is 1 and 1-max (1) = 0. The value of classification error index is always between 0 and 1. In fact the maximum Gini index for a given number of classes is always equal to the maximum of classification error index because for a number of classes n, we set probability is equal to $p = \dfrac{1}{n}$ and

maximum Gini index happens at $1 - n\dfrac{1}{n^2} = 1 - \dfrac{1}{n}$, while

maximum classification error index also happens at

$$1 - \max\left\{\frac{1}{n}\right\} = 1 - \frac{1}{n}.$$

3.2 Splitting Criteria

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S, is defined as

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where Values (A) is the set of all possible values for attribute A, and S_v is the subset of S for which attribute A has value v (i.e., $S_v = \{s \in S \mid A(s) = v\}$). The first term in the equation for Gain is just the entropy of the original collection S and the second term is the expected value of the entropy after S is partitioned using attribute A. The expected entropy described by this second term is simply the sum of the entropies of each subset, weighted by the fraction of examples $\dfrac{|S_v|}{|S|}$ that belong to Gain (S, A) is therefore, the expected reduction in entropy caused by knowing the value of attribute A.

$$\text{Split Information (S, A)} = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

and

$$\text{Gain Ratio(S, A)} = \frac{Gain(S, A)}{Split\ Information(S, A)}$$

The process of selecting a new attribute and partitioning the training examples is now repeated for each non terminal descendant node. Attributes that have been incorporated higher in the tree are excluded, so that any given attribute can appear at most once along any path through the tree. This process continues for each new leaf node until either of two conditions is met:

1. Every attribute has already been included along this path through the tree, or
2. The training examples associated with this leaf node all have the same target attribute value (i.e., their entropy is zero).

3.3 The ID3Algoritm

ID3 (Examples, Target Attribute, Attributes)

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
  - A = The Attribute that best classifies examples.
  - Decision Tree attribute for Root = A.
  - For each possible value, v_i, of A,

- ▪ Add a new tree branch below Root, corresponding to the test $A = v_i$.
- ▪ Let Examples(v_i) be the subset of examples that have the value v_i for A
- ▪ If Examples(v_i) is empty
  - ➢ Then below this new branch add a leaf node with label = most common target value in the examples
  - ➢ Else below this new branch add the subtree ID3 (Examples(v_i), Target_Attribute, Attributes – {A})
- • End
- • Return Root

4. RELATED WORK

Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students as described by Alaa el-Halees [7].

Oladipupo and oyelade [8] made their study using association rule data mining technique to identify student's failure patterns. They take a total number of 30 courses for 100 level and 200 level. Their study focuses on constructive recommendation, curriculum structure and modification in order to improve student's academic performance and trim down failure rate.

Bharadwaj and Pal [6] obtained the university students data like attendance, class test, seminar and assignment marks from the students' previous database, to predict the performance at the end of the semester.

Hijazi and Naqvi [9] conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students family income, students mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

Khan [10] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general.

Al-Radaideh, et al [11] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Pandey and Pal [12] conducted study on the student performance based by selecting 60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India. By means of association rule they find the interestingness of student in opting class teaching language.

Bray [13], in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socio-economic conditions.

Z. J. Kovacic [14] presented a case study on educational data mining to identify up to what extent the enrolment data can be used to predict student's success. The algorithms CHAID and CART were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and 60.5 respectively.

Bharadwaj and Pal [15] conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attributes, it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

Yadav, Bharadwaj and Pal [17] obtained the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and shows that CART is the best algorithm for classification of data.

5. DATA MINING PROCESS

India is a vast country whose engineering and professional student population outnumbers every other country, except possibly China. India has one of the largest higher education systems in the world. Today we have 545 Universities, with varying statutory bases and mandates. Of these, 41 are Central Universities, 281 are State Universities, 129 are Deemed Universities, 94 are State Private Universities. The number of colleges has increased from 500 in 1947 to around 26000, where most of the enrolment in higher education occurs.

In the spheres of technical education, there were a whooping 1346 engineering colleges in India by the end of the year 2004, all of which were approved by the All India Counsel for Technical Education (AICTE) with a seating capacity of 440,000. In 2008, a total of 2388 engineering colleges were registered, with a total student intake capacity of 8.2 lakhs students and 1231 management institutes, with an annual intake capacity of 1.5 lakhs students. Other than engineering and MBA colleges, there were 1137 MCA, 1001 Pharmacy, 86 HMCT, 11 Applied Arts and Crafts, and 285 PGDM institutes were approved by AICTE. By the end of December 2008, the number of applications pending to seek AICTE

approval was 886 for engineering and technology, 250 for MBA, 834 for MCA, 120 for Pharmacy, 124 for HMCT, and 1 for Applied Arts and Crafts - a total of 2237. The total number of engineering seats has crossed the mark of 1 million in 2009.

University's success highly depends upon students' success in that University. Knowing the reasons of failure of student can help the teachers and administrators to take necessary actions so that the success percentage can be improved. The data is collected from Department of MCA of V. B. S. Purvanchal University, Jaunpur, India. The MCA department has been started in the year 1997 and 12 batches have completed their study. It is very important to improve the student's success rate every year to attract more and better students.

5.1 Data Selection and Preprocessing

Master of Computer Application (MCA) is a Three Year (Six Semester) full time course. The MCA course is open to graduate in any discipline with Mathematics as a subject at 10+2 level. That candidate, who has done BCA are exempted from mathematics at 10 + 2 level.

Data of 432 students of the Department of MCA, VBS Purvanchal University, Jaunpur is collected who get admission from 1997-2000 batch to 2007-2010 batch. The data was collected through the enrolment form filled by the student at the time of admission. The student enter their demographic data (category, gender etc), past performance data (SSC or 10th marks, HSC or 10 + 2 exam marks and Graduation Marks etc.), address and contact number. From these the attributes that possibly influence their result are selected as shown in Table 1. Most of the attributes reveal the performance of the students.

Table 1. Student Related Variables

Variable	Description	Possible Values
Sex	Students Sex	{Male, Female}
Cat	Students category	{General, OBC, SC, ST}
GSS	Students grade in Senior Secondary education	{O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40%}
GMSS	Students grade in Math at Senior Secondary education	{O – 90% -100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40%, Not Applicable}
GS	Graduation Stream	{BA with Math, B.A. without Math, BSc. With Math, B.Sc. without Math, B.Com, BCA, BBA, B.Tech}
GOG	Grade obtained in Graduation	{First ≥ 60% Second ≥45 & <60% Third ≥36 & <45%}
MED	Medium of Teaching in Graduation	{Hindi, English, Regional}

CL	College Location	{Rural, Urban}
ATYPE	Admission Type	{UPSEE, Direct}
GOM	Grade obtained in MCA	{First ≥ 60% Second ≥45 & <60% Third ≥36 & <45% Fail < 36% }

The domain values for some of the variables were defined for the present investigations are as follows:

- Cat – From ancient time Indians are divided in many categories. These factors play a direct and indirect role in the daily lives including the education of young people. Admission process in India also includes different percentage of seats reserved for different categories. In terms of social status, the Indian population is grouped into four categories: General, Other Backward Class (OBC), Scheduled Castes (SC) and Scheduled Tribes (ST). Possible values are General, OBC, SC and ST.

- GSS - Students grade in Senior Secondary education. Students who are in state board appear for five subjects each carry 100 marks. Grade are assigned to all students using following mapping O – 90% to 100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, and F - < 40%.

- GMSS – Student Grade in Mathematics at Senior Secondary education. Grade in mathematics at 10+2 level are assigned to all students using following mapping O – 90% to 100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, and F - < 40%. If student has not the mathematics at 10 + 2 level then assign Not-Applicable.

- GS – Graduation Stream. MCA admission is open for all stream students, therefore, Graduation Stream is split into following classes BA with Math, B.A. without Math, B.Sc. with Math, B.Sc. without Math, B.Com, BCA, BBA, B.Tech.

- GOG – Grade Obtained in Graduation. Marks/Grade obtained in graduation. It is also split into four class values: First – ≥60%, Second – ≥45% and <60%, Third – ≥36% and < 45%.

- Med – This paper study covers only the degree colleges and institutions of Uttar Pradesh state of India. Here, medium of instructions are Hindi or English or Regional.

- ATYPE - The admission type which may be through Uttar Pradesh State Entrance Examination (UPSEE) or direct admission through University procedure.

- GOM – Grade Obtained in MCA. Marks/Grade obtained in Master of Computer Applications. It is also split into five class values: First – ≥60% , Second – ≥45% and <60%, Third – ≥6% and < 45% and Fail < 36%.

5.2 Model Construction

The Weka Knowledge Explorer is an easy to use graphical user interface that harnesses the power of the Weka software. The major Weka packages are Filters, Classifiers, Clusters, Associations, and Attribute Selection is represented in the Explorer along with a Visualization tool, which allows datasets and the predictions of Classifiers and Clusters to be visualized in two dimensions. The workbench contains a collection of visualization tools and algorithms for data

analysis and predictive modeling together with graphical user interfaces for easy access to this functionality. It was primarily designed as a tool for analyzing data from agricultural domains. Now it is used in many different application areas, in particular for educational purposes and research. The main strengths is freely available under the GNU General Public License, very portable because it is fully implemented in the Java programming language and runs on any modern computing platform, contains a comprehensive collection of data preprocessing and modeling techniques. Weka supports several standard data mining tasks like data clustering, classification, regression, preprocessing, visualization and feature selection. These techniques are predicated on the assumption that the data is available as a single flat file or relation. Each data point is described by a fixed number of attributes and an important area is currently not covered by the algorithms included in the Weka distribution is sequence modeling.

From the above data, mca.arff file was created. This file was loaded into WEKA explorer. The classify panel enables the user to apply classification and regression algorithms to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, or the model itself. There are 16 decision tree algorithms like ID3, J48, Simple CART etc. implemented in WEKA. The algorithm used for classification is ID3. Under the "Test options", the 10-fold cross-validation is selected as our evaluation approach. Since there is no separate evaluation data set, this is necessary to get a reasonable idea of accuracy of the generated model. The model is generated in the form of decision tree.

5.3 Results Obtained

The ID3 decision tree generated from mca.arff. The accuracy of the model is 60.46 %. That is out of 430 instances 260 instances are correctly classified. The most important attribute in predicting student's enrollment is found to be GS. The social attributes like category, Medium, College Location and Admission Type are not appearing in the decision tree indicating less relevance of the prediction with such attributes. The knowledge represented by decision tree can be extracted and represented in the form of IF-THEN rules in Table 2.

Table 2: Rule Set generated by ID3

Rules for GOM= 'First'
IF GS= 'B.Sc. with math' and Sex = 'Male' THEN GOM = 'First'
IF GS = 'BCA' and GSS = 'A' and GMSS= 'A' or 'B' THEN GOM = 'First'
IF GS = 'BCA' and GSS = 'A' and GMSS= 'C' and GOG = 'Second' THEN GOM = 'First'
Rules for GOM= 'Second'
IF GS= 'B.A. without math' and GMSS = 'A' THEN GOM = 'Second'
IF GS= 'B.A. without math' and GMSS = 'A' and GMSS = 'B' and Sex = 'Male' THEN GOM = 'Second'
IF GS= 'BCA' and GSS = 'A' and GMSS = 'Not Applicable' THEN GOM = 'Second'
IF GS= 'BCA' and GSS = 'A' and GMSS = 'D' THEN GOM = 'Second'
Rules for GOM= 'Third'
IF GS= 'B.A. without math' and GMSS = 'C' and GSS = 'B' and CL = Urban' THEN GOM = 'Third'
IF GS= 'B.A. without math' and GMSS = 'C' THEN GOM = 'Third'
IF GS= 'B.A. without math' and GMSS = 'D' and GSS =

'D' THEN GOM = 'Third'
IF GS= 'B.A. without math' and GMSS = 'E' or 'F' and GSS = 'B' or 'C' THEN GOM = 'Third'
Rules for GOM= 'Fail'
IF GS= 'B.A. without math' and GMSS = 'B' and Sex = 'Female' THEN GOM = 'Fail'
IF GS= 'B.A. without math' and GMSS = 'C' and GSS = 'D' THEN GOM = 'Fail'
IF GS= 'B.A. without math' and GMSS = 'D' and GSS = 'B' or 'C' THEN GOM = 'Fail'
IF GS= 'B.A. without math' and GMSS = 'E' and GSS = 'E' or 'F' THEN GOM = 'Fail'
IF GS= 'BCA' and GMSS = 'F' and GSS = 'B' or 'C' THEN GOM = 'Fail'

The classification matrix has been presented in Table 3, which compared the actual and predicted classifications. In addition, the classification accuracy for the four-class outcome categories was presented.

Table 3: Classification Matrix-ID3 prediction Model

GOM		Predicted				% of correct Precision
		First	Second	Third	Fail	
Actual	First	40	20		10	57.14%
	Second	20	40	10	10	50.00%
	Third		20	20	20	33.33%
	Fail	10	10	10	30	50.00%

The class wise accuracy is shown in Table 4

Table 4: Class wise accuracy for four class prediction

GOM	TP Rate	FP Rate
First	0.571	0.15
Second	0.5	0.263
Third	.333	0.19
Fail	0.5	0.095

6. CONCLUSION

This study shows that students past academic performance can be used to create the model using ID3 decision tree algorithm that can be used for prediction of student's enrollment in MCA course. From the classification matrix it is clear that the true positive rate for obtaining the Third Class and Fail Class is 33.33% and 50% respectively. The students going to take admission in MCA course can be considered for proper counseling so as to improve their result or choose any other course like MBA. This study concludes that B.Sc. students with mathematics and BCA students are performed better in MCA course but the student of B.A. without mathematics did not perform well in MCA course.

7. REFERENCES

[1] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.

[2] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.

[3] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.

[4] Romero C and Ventura S, "Educational Data Mining: A survey from 1995 to 2005" expert system with Application 33(2007) 135-146.

[5] J. R. Quinlan, "Introduction of decision tree: Machine learn", 1: pp. 86-106, 1986.

[6] B.K. Bharadwaj and S. Pal. "Mining Educational Data to Analyze Students' Performance", International Journal of Advance Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp. 63-69, 2011.

[7] Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.

[8] Oladipupo O O and Oyelade O J, "Knowledge Discovery from student's result repository: Association rule mining Approach", IJCSS vol. (4) Issue(2).

[9] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[10] Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005.

[11] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

[12] U. K. Pandey, and S. Pal, "A Data mining view on class room teaching language", (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.

[13] M. Bray, The shadow education system: private tutoring and its implications for planners, (2nd ed.), UNESCO, PARIS, France, 2007.

[14] Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference 2010.

[15] B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[16] S. K. Yadav, B.K. Bharadwaj and S. Pal, "Data Mining Applications: A comparative study for Predicting Student's Performance", International Journal of Innovative Technology and Creative Engineering (IJITCE), Vol. 1, No. 12, pp. 13-19.

8. AUTHORS PROFILE

Surjeet Kumar Yadav received his M.Sc. (Computer Science) from Dr. Baba Saheb Marathwada University, Aurangabad, Maharastra, India (1998). At present, he is working as Sr. Lecturer at Department of Computer Applications, VBS Purvanchal University, Jaunpur. He is an active member of CSI and National Science Congress. He is currently doing research in Data Mining and Knowledge Discovery.

Saurabh Pal received his M.Sc. (Computer Science) from Allahabad University, UP, India (1996) and obtained his Ph.D. degree from the Dr. R. M. L. Awadh University, Faizabad (2002). He then joined the Dept. of Computer Applications, VBS Purvanchal University, Jaunpur as Lecturer. At present, he is working as Head and Sr. Lecturer at Department of Computer Applications.
Saurabh Pal has authored more than 25 numbers of research papers in international/national Conference/journals and also guides research scholars in Computer Science/Applications. He is an active member of CSI, Society of Statistics and Computer Applications and working as reviewer and member of editorial board for more than 15 international journals. His research interests include Image Processing, Data Mining, Grid Computing and Artificial Intelligence.