Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 2, Issue. 12, December 2013, pg.359 – 369

RESEARCH ARTICLE

Evaluation of Teacher's Performance: A Data Mining Approach

Ajay Kumar Pal[1], Saurabh Pal[2]

[1]Research Scholor, Sai Nath University, Ranchi, Jharkhand, India
[2]Head Dept. of MCA, VBS Purvanchal University, Jaunpur (U.P.), India

Abstract— Most Organizations use performance appraisal system to evaluate the teacher's performance. The teacher's performance is very important to the students and as well as college management, in which usually involves crisp and uncertain values to evaluate teacher's performance. In this paper we proposed to evaluate teacher's performance on the basis of different factors, using data mining techniques, Data mining is the powerful technology for analyzing important information from the data warehouse. It is data analysis methodology used to identify hidden patterns in a large data set. We can consider some of the most relevant factors, and develop rules using data mining techniques.

The idea proposed in this paper is to perform an analysis considering number of parameters for the derivation of performance prediction indicator s needed for teachers performance assessment, monitoring and evaluation. The aim is to predict the quality, productivity and potential of faculty across various disciplines which will enable higher level authorities to take decisions and understand certain patterns of teacher's motivation, satisfaction, growth and decline. The analysis depends on many factors, encompassing student's feedback, organizational feedback, institutional support in terms of finance, administration, research activity etc. The data mining methodology used for extracting useful patterns from the institutional database is able to extract certain unidentified trends in teacher's performance when assessed across several parameters.

Keywords— Performance Appraisal; Teacher; KDD; ID3; CART and LAD

I. INTRODUCTION

Data mining is a promising field of information and knowledge discovery (Han et al., [1]). It started to be an interest target for information industry, because of the existence of huge data containing large amounts of hidden knowledge. With data mining techniques, such knowledge can be extracted and accessed transforming the databases tasks from storing and retrieval to learning and extracting knowledge.

Data miming consists of a set of techniques that can be used to extract relevant and interesting knowledge from data. Data mining has several tasks such as association rule mining, classification and prediction, and clustering. Classification techniques are supervised learning techniques that classify data item into predefined class label. It is one of the most useful techniques in data mining to build classification models from an input data set. The

used classification techniques commonly build models that are used to predict future data trends. There are several algorithms for data classification such as decision tree and Naïve Bayes classifiers. With classification, the generated model will be able to predict a class for given data depending on previously learned information from historical data.

In student point of view, we find teacher's performance in different form, for example suppose some a teacher beginning is good but his presentation is not good or some other teacher's voice is slow but his explanation and presentation power is good. Therefore evaluation of teacher's performance is a critical task.

Evaluating the performance of a teacher is necessary due to many reasons for betterment of students and teachers -

- Improvement of the student's performance
- Monitoring of the students
- Betterment of the students
- Increase motivation to further improve performance
- Increase self respect and ambition

Teachers' performance evaluation incentives as the basis for school reforms have recently attracted considerable attention and support among researchers and policy makers. Evaluating teachers' performance is a difficult task, but it is made for years through pedagogical surveys. The most usual kind of pedagogical survey is just a set of closed-ended questions, with multiple choices that follow some order (as in the Likert scale (Likert1932 [2])). In order to assess teacher's performance, students individually fill a pedagogical survey, and the mean of the answers corresponds to the teacher evaluation. Despite the spread of such tools, to our knowledge there is no work on automatically predict teachers' performance based on surveys.

In this paper we discuss the teachers' performance evaluation using data mining techniques at university teachers. The proposed model consider the various aspects of performance measures of teachers, like Students' Feedback (voice modulation, speed of delivery, content arrangement, presentation, communication, overall impression, content delivery, explanation power, overall teaching and regularity, Results, Students attendance, that have deep influence on the teachers' performance in university. Proposed model is designed to combine the knowledge and expertise of human experts with reasoning capabilities that will provide a great support to the head of the department for decision-making in educational institutions.

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website.

II. DATA MINING

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data. The amount of data stored in databases is increasing quickly but the rate of increasing resources analyzing data is too lower than rate of data increase. There is an essential and necessary need for automatic and intelligent tools and methods. This need directs us to a new area named data mining and knowledge discovery [1]. Data mining is: Discovering the methods and patterns in large databases to guide decisions about future activities. It is expected that data mining tools to get the model with minimal input from the user to recognize. The model presented can be useful to understand the unexpected and provide an analysis of data followed by other tools to put decision-making are examined and it ultimately leads to strategic decisions and business intelligence. The simplest word for knowledge extraction and exploration of volume

data is very high and the more appropriate name for this term is "Exploring the knowledge of database". A database is knowledge of discovery process. This process includes the preparation and interpretation of results.

A KDD process includes the following stages:

1. Developing some detection from the applied program that is related to previous knowledge and final goal.
2. Making a target data collection to be used in knowledge discovery.
3. Data refinement and pre-processing (including controlling invalid data values, disorder in data, calculations of time series and recognized changes)
4. Reducing number of variables and finding similar samples of data, if possible.
5. Selection of data mining process (classification, regression, clustering, etc)
6. Selection of data mining algorithm
7. Exploring favorite patterns (this stage is actual data mining)
8. Interpretation of the discovered pattern; If needed, stages one to seven are repeated.
9. Integration of discovered information and preparation of a report [1]

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Data classification is a two-step process.

In the first step, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of the attributes, called the class label attribute. The data tuples analyzed to build the model collectively form the training data set. The individual tuples making up the training set are referred to as training samples and are randomly selected from the sample population. The learned model is represented in the form of classification rules, decision tree or mathematical formulae.

In the second step, the model is used for classification. First the predictive accuracy of the model is estimated. If the accuracy of the model is considered acceptable, the model can be used to classify future data tuples or objects for which the class label is not known. The basic techniques used in this paper are Naive Bayes, ID3, CART and LAD tree.

A. Naïve Bayes Classification

The Naïve Bayes Classifier technique is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods. Naïve Bayes model identifies the characteristics of dropout students. It shows the probability of each input attribute for the predictable state.

A Naive Bayesian classifier is a simple probabilistic classifier based on applying Bayesian theorem (from Bayesian statistics) with strong (naive) independence assumptions. By the use of Bayesian theorem we can write.

$$P(Ci \mid X) = \frac{P(X \mid Ci)P(Ci)}{P(X)}$$

We preferred Naive Bayes implementation because:

- Simple and trained on whole (weighted) training data
- Over-fitting (small subsets of training data) protection
- Claim that boosting "never over-fits" could not be maintained.
- Complex resulting classifier can be determined reliably from limited amount of data

B. ID3 (Iterative Dichotomise 3)

This is a decision tree algorithm introduced in 1986 by Quinlan Ross [3]. It is based on Hunts algorithm. The tree is constructed in two phases. The two phases are tree building and pruning.

ID3 uses information gain measure to choose the splitting attribute. It only accepts categorical attributes in building a tree model. It does not give accurate result when there is noise. To remove the noise pre-processing technique has to be used.

To build decision tree, information gain is calculated for each and every attribute and select the attribute with the highest information gain to designate as a root node. Label the attribute as a root node and the possible values of the attribute are represented as arcs. Then all possible outcome instances are tested to check whether they are falling under the same class or not. If all the instances are falling under the same class, the node is represented with single class name, otherwise choose the splitting attribute to classify the instances.

Continuous attributes can be handled using the ID3 algorithm by discretizing or directly, by considering the values to find the best split point by taking a threshold on the attribute values. ID3 does not support pruning.

C. CART

CART stands for Classification and Regression Trees introduced by Breiman [4]. It is also based on Hunt's algorithm. CART handles both categorical and continuous attributes to build a decision tree. It handles missing values.

CART uses Gini Index as an attribute selection measure to build a decision tree .Unlike ID3 and C4.5 algorithms, CART produces binary splits. Hence, it produces binary trees. Gini Index measure does not use probabilistic assumptions like ID3, C4.5. CART uses cost complexity pruning to remove the unreliable branches from the decision tree to improve the accuracy.

D. LAD Tree

Logical Analysis of Data is the method for classification proposed in optimization literature. It builds a classifier for binary target variable based on learning a logical expression that can distinguish between positive and negative samples in a data set. The basic assumption of LAD model is that a binary point covered by some positive patterns, but not covered by any negative pattern is positive, and similarly, a binary point covered by some negative patterns, but not covered by positive pattern is negative. The construction of Lad model for a given data set typically involves the generation of large set patterns and the selection of a subset of them that satisfies the above assumption such that each pattern in the model satisfies certain requirements in terms of prevalence and homogeneity.

III. BACKGROUND AND RELATED WORK

Performance appraisal system is basically a formal interaction between an employee and the supervisor or management conducted periodically to identify the areas of strength and weakness of the employee. The objective is to be consistent about the strengths and work on the weak areas to improve performance of the individual and thus achieve optimum process quality [5].

Chein and Chen [6] used several attributes to predict the employee performance. They specified age, gender, marital status, experience, education, major subjects and school tires as potential factors that might affect the performance. Then they excluded age, gender and marital status, so that no discrimination would exist in the process of personal selection. As a result for their study, they found that employee performance is highly affected by education degree, the school tire, and the job experience.

Sanders [7] observed that once teachers are given feedback pertaining to classroom-level instructional outcomes, they start to modify their instruction to address their weak areas. It is important, however, that when administrators make decisions and provide feedback to teachers on their performance, that the information is a valid measure of their actual job performance, which means it should include a teachers' responsibilities both in-class and out-of-class

Han and Kamber [1] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Pal and Pal [8] conducted study on the student performance based by selecting 200 students from BCA course. By means of ID3, c4.5 and Bagging they find that SSG, HSG, Focc, Fqual and FAIn were highly correlated with the student academic performance.

Tongshan Chang, & Ed.D [9] introduces a real project to assist higher education institutions in achieving enrollment goals using data mining techniques Furthermore, the results also provide evidence that data mining is an effective technology for college recruitment. It can help higher education institutions mange enrollment more effectively.

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 socioeconomic status had relatively higher academic achievement in general.

Pal and Pal [11] presented a study for evaluating and investigating placement of student after doing MCA by the three selected classification algorithms based on Weka. The best algorithm based on the placement data is Naïve Bayes Classification with an accuracy of 86.15%.

Z. J. Kovacic [12] 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.

Pal [13] conducted study on the student dropout rate by selecting 1650 students from different branches of engineering college. In their study, it was found that student's dropout rate in engineering exam, high school grade; senior secondary exam grade, family annual income and mother's occupation were highly correlated with the student academic performance.

Galit [14] gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Yadav, Bhardwaj and Pal [15] conducted study on the student retention based by selecting 398 students from MCA course of VBS Purvanchal University, Jaunpur, India. By means of classification they show that student's graduation stream and grade in graduation play important role in retention.

Al-Radaideh, et al [16] 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.

Sudheep Elayidom , Sumam Mary Idikkula & Joseph Alexander [17] proved that the technology named data mining can be very effectively applied to the domain called employment prediction, which helps the students to choose a good branch that may fetch them placement. A generalized framework for similar problems has been proposed.

Baradwaj and Pal [18] 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.

Ayesha, Mustafa, Sattar and Khan [19] describe the use of k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

Bray [20], 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 socioeconomic conditions.

Yadav, Bhardwaj and Pal [21] 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.

IV. **DATA MINING PROCESS**

"Evaluation of Teachers performance uses data mining techniques" in this research teacher's performance is evaluated. First a survey of the teachers' requirements and students' requirements is made. Then we interact with the teachers and got some knowledge about their methods. We should meet different teachers that have been given some ideas about the finding of the teacher's performance.

A. Data Preparations

For this study data were collected from post graduate studies at department in the college of engineering over three years period of the same students for the purpose of investigating if there is any changes in their evaluation of lecturers over that period. The demographics data are not collected during the normal survey process but the university has a record of this data and they are used for the paper. That means they are not exactly the same for the sample of students who completed the performance test but are an average representation of the students enrolled for the unit. The profile of students is defined based on the academic and cultural backgrounds of students. The academic background is measured using the entry requirements to be fulfilled by the students to get entry into the university.

Also it is investigated whether there is any correlation between the student learning outcomes and teacher's performance. In all instances the teaching evaluations were collected before the end of the semester and at that time students did not know their final grade for the unit. Therefore it is difficult to gather this information in class, from the same set of students. Instead, the learning outcomes are measured using the end of semester exam results for the students enrolled in the course.

B. Data selection and Transformation

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table I for reference.

TABLE I

STUDENT RELATED VARIABLES

Variables	Description	Possible Values
Name	Teachers Name	Text
SD	speed of delivery	{1,2,3,4,5}
CA	content arrangement	{1,2,3,4,5}
PR	presentation	{1,2,3,4,5}
CO	communication	{1,2,3,4,5}
OI	knowledge	{1,2,3,4,5}
CD	content delivery	{1,2,3,4,5}
EP	explanation power	{1,2,3,4,5}
DC	Doubts Clearing	{1,2,3,4,5}
DP	Discussion of Problems	{1,2,3,4,5}
OCR	overall completion of course and regularity	{1,2,3,4,5}
SA	Students attendance	{1-below,2- average,3- high}
RE	Result	{1-Pass, 2-Fail, 3 - Promoted}
PT	Performance of Teacher	{1,2,3,4,5}

The survey uses a Likert-type scale, and the operationalization of the survey for items is as follows:

5 = Strongly Agree,
4 = Agree,
3 = Disagree, and
2 = Strongly Disagree.
1 = Neither Agree nor Disagree,

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

- **SA –** Attendance of Student. Minimum 70% attendance is compulsory to participate in End Semester Examination. But even though in special cases low attendance students also participate in End Semester Examination on genuine reason basis. Attendance is divided into three classes: *Poor - <60%, Average - ≥ 60% and < 80%, Good - ≥ 80%.*
- **Result –** Students result in of Engineering. It is split into three classes Pass, Fail or Promoted. If a student passes all the paper is awarded pass class. If students fail in up to three theory and two practical subjects of an academic year or vice versa, he/she is promoted to next class, otherwise he/she is fail.

C. Implementation of Mining Model

Weka is open source software that implements a large collection of machine leaning algorithms and is widely used in data mining applications. From the above data, teacher.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. The algorithm used for classification is Naive Bayes, ID3, CART and LAD. 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.

D. Results

To better understand the importance of the input variables, it is customary to analyse the impact of input variables on the performance of teacher's evaluation, in which the impact of input variable of the model on the output variable has been analysed. Tests were conducted using four tests for the assessment of input variables: Chi-square test, Info Gain test and Gain Ratio test. Different algorithms were provide different results, i.e. each of them accounts the

relevance of variables in a different way, therefore the average value of all the algorithms is taken in the final result of variables ranking, instead of selecting one algorithm and evaluating it. The results obtained with these values are shown in Table II.

TABLE III

RESULT OF TESTS AND AVERAGE RANK

CA	0.025	4.74986	0.0373	4.7872933
RE	0.0251	5.15798	0.035	5.1947467
CO	0.0646	10.62739	0.0973	10.724423
CD	0.0574	18.01863	0.0979	18.108663
OI	0.0717	18.55387	0.112	18.662903
PR	0.0588	19.63136	0.1111	19.727193
Name	0.0624	25.00844	0.1555	25.122673
SA	0.1447	31.33758	0.2093	31.552047
DP	0.3496	141.7417	0.593	142.28897
DC	0.4115	188.10537	0.6611	188.73724
OCR	0.3874	202.70396	0.6284	203.30083
EP	0.3494	238.60901	0.515	239.13008
SD	0.442	251.93558	0.6043	252.57901

The aim of this analysis is to determine the importance of each variable individually. Table II shows that attribute *CA(Content Arrangement)* impacts output the most, and that it showed the best performances in all of the three tests. Then these attributes follow: RE (Result), CO (Communication), and CD (content delivery).

Now, we have carried out some experiments in order to evaluate the performance and usefulness of different classification algorithms for predicting students' placement. The results of the experiments are shown in table III.

TABLE III

PERFORMANCE OF THE CLASSIFIERS

Evaluation Criteria		**Classifiers**		
	NB	**ID3**	**CART**	**LAD**
Timing to build model (in Sec)	0	0	0.11	0.08
Correctly classified instances	90	73	81	84
Incorrectly classified instances	22	34	31	28
Accuracy (%)	80.35%	65.17%	72.32%	75.00%

The percentage of correctly classified instances is often called accuracy or sample accuracy of a model. So Naïve Bayes classifier has more accuracy than other three classifiers.

E. Discussion

Based on the above Table IV, we can clearly see that the highest accuracy is 80.35% and the lowest is 65.17%. In fact, the highest accuracy belongs to the Naïve Bayes Classifier followed by LAD tree with a percentage of 75.00% and subsequently CART.

Decision trees are considered easily understood models because a reasoning process can be given for each conclusion. Knowledge models under this paradigm can be directly transformed into a set of IF-THEN rules that are one of the most popular forms of knowledge

representation, due to their simplicity and comprehensibility which professor can easy understand and interpret Fig 1.

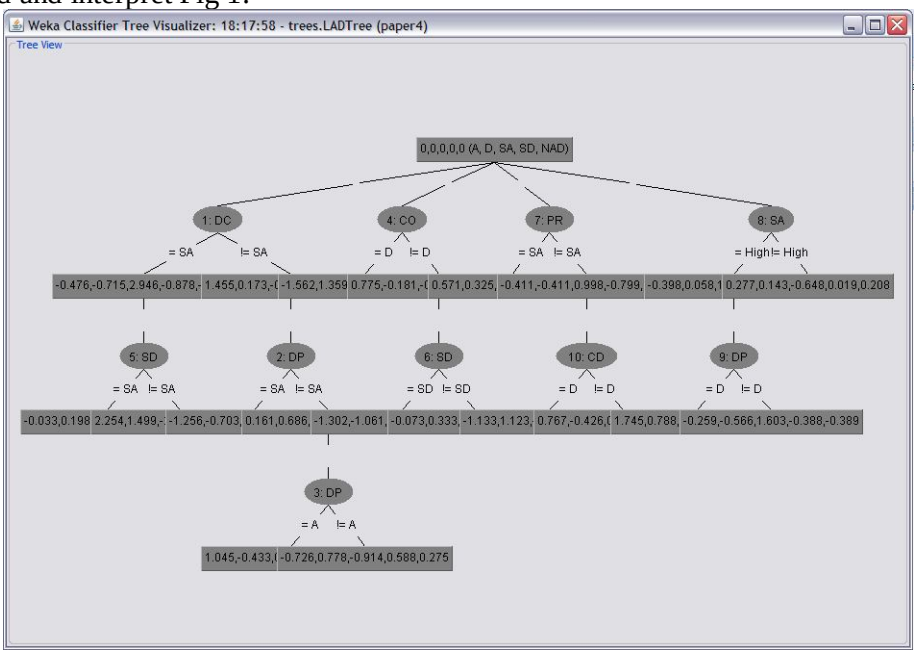

Fig 1: LAD Tree

After examining the classification tree, we can summarize the following results:

weka.classifiers.trees.LADTree:

: 0,0,0,0,0
| (1)DC = SA: -0.476,-0.715,2.946,-0.878,-0.878
| | (5)SD = SA: -0.033,0.198,0.812,-0.49,-0.488
| | (5)SD != SA: 2.254,1.499,-2.107,-0.824,-0.822
| (1)DC != SA: 1.455,0.173,-0.024,-0.748,-0.856
| | (2)DP = SA: -1.256,-0.703,3.16,-0.604,-0.598
| | (2)DP != SA: 0.161,0.686,-0.446,-0.062,-0.34
| | | (3)DP = A: 1.045,-0.433,0.509,-0.568,-0.554
| | | (3)DP != A: -0.726,0.778,-0.914,0.588,0.275
| (4)CO = D: -1.562,1.359,1.41,-0.614,-0.593
| (4)CO != D: 0.775,-0.181,-0.175,-0.136,-0.283
| | (6)SD = SD: -1.302,-1.061,-0.549,3.467,-0.556
| | (6)SD != SD: -0.073,0.333,0.272,-0.747,0.215
| (7)PR = SA: 0.571,0.325,-0.342,0.05,-0.603
| (7)PR != SA: -0.411,-0.411,0.998,-0.799,0.624
| | (10)CD = D: -1.133,1.123,-0.573,-0.755,1.338
| | (10)CD != D: 0.767,-0.426,0.713,-0.514,-0.539
| (8)SA = High: -0.398,0.058,1.352,-0.493,-0.519
| | (9)DP = D: 1.745,0.788,-1.455,-0.527,-0.55
| | (9)DP != D: -0.259,-0.566,1.603,-0.388,-0.389
| (8)SA != High: 0.277,0.143,-0.648,0.019,0.208

V. CONCLUSIONS

Teachers' regular assessment is suggested to maintain quality in higher education. As a conclusion, we have met our objective which is to evaluate and investigate teacher's performance by the four selected classification algorithms based on Weka. The best algorithm based on the data is Naïve Bayes Classification. Naïve Bayes classifier has the lowest average error as compared to others. These results suggest that among the machine learning algorithm tested, Naïve Bayes classifier has the potential to significantly improve the conventional classification methods.

On working on performance, many attributes have been tested, and some of them are found effective on the performance prediction. The content arrangement was the strongest attribute, and then the result plays an important role in the performance of teachers.

The speed of delivery attribute did not show any clear effect while the overall completion of course and regularity attribute has shown some effect in some of the experiments for predicting the performance. Other attributes had a degree of effect on predicting the performance.

REFERENCES

[1] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000.
[2] R. LIKERT, A Technique for the Measurement of Attitudes. Archives of Psychology, 22(140), 1-55, 1932.
[3] J. R. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco, 1993.
[4] Breiman, Leo, Jerome Friedman, R. Olshen and C. Stone (1984). Classification and Regression Trees. Belmont, California: Wadsworth.
[5] Archer-North and Associates, "Performance Appraisal", http://www.performance-appraisal.com, 2006, Accessed Dec, 2012.
[6] Chein, C., Chen, L., "Data mining to improve personnel selection and enhance human capital: A case study in high technology industry", Expert Systems with Applications, In Press (2006).
[7] Sanders, W. L., & Horn, S. P., Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. Journal of Personnel Evaluation in Education, 12, 247-256, 1998.
[8] K. Pal, and S. Pal, "Analysis and Mining of Educational Data for Predicting the Performance of Students", (IJECCE) International Journal of Electronics Communication and Computer Engineering, Vol. 4, Issue 5, pp. 1560-1565, ISSN:2278-4209, 2013.
[9] Tongshan Chang, Ed.D., Data Mining: A Magic Technology for College Recruitment´s, Paper of Overseas Chinese Association for Institutional Research (www.ocair.org), 2008.
[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] K. Pal, and S. Pal, "Classification model of prediction for placement of students", IJMECS, vol. 5, Issue 11, 49-56, 2013.
[12] Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference, 2010.
[13] Pal S., "Mining Educational Data to Reduce Dropout Rates of Engineering Students", I.J. Information Engineering and Electronic Business (IJIEEB), Vol. 4, No. 2, 2012, pp. 1-7.
[14] Galit.et.al, "Examining online learning processes based on log files analysis: a case study". Research, Reflection and Innovations in Integrating ICT in Education, 2007
[15] S. K. Yadav, B.K. Bharadwaj and S. Pal, "Mining Educational Data to Predict Student's Retention :A Comparative Study", International Journal of Computer Science and Information Security (IJCSIS), Vol. 10, No. 2, 2012
[16] 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.
[17] Sudheep Elayidom , Sumam Mary Idikkula & Joseph Alexander "A Generalized Data mining Framework for Placement Chance Prediction Problems" International Journal of Computer Applications (0975– 8887) Volume 31– No.3, October 2011.
[18] 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.
[19] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data mining model for higher education system", Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010

[20] M. Bray, The shadow education system: private tutoring and its implications for planners, (2nd ed.), UNESCO, PARIS, France, 2007.
[21] 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, 2011.

Authors' Profile

Ajay Kumar Pal received his MCA. (Master of Computer Applications) from VBS Purvanchal University, Jaunpur, UP, India. Currently he is working as Head Of Department of Computer Application, Shri Vishwanath P.G. College Kalan Sultanpur(U.P.) At present, he is doing research in Data Mining and Knowledge Discovery. He is an active member of CSI and National Science Congress.

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 40 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.

