ISSN:0975-9646

A Modern Data Mining Method for Assessment of Teaching Assistant in Higher Educational Institutions

Surjeet Kumar
MCA Dept. VBS Purvanchal University, Jaunpur

Abstract- Assessment of teacher's performance in higher education system is very important. Improvement of teacher's performance in the developing countries can be well motivated from these points of views: Our national policies on higher education support evaluation of teachers and system which are at the forefront of the education reforms agenda. India is a country surging ahead in full steam on education reforms. The MHRD has planned massive reforms aimed at bringing flexibility, transparency and quality into the Indian education system. These reforms would also help the country deal with the challenges faced by the sector. The MHRD has also invited the private sector to contribute to the growth of education system in the country while emphasizing upon the fact that "profit" and "surplus" needed to be delineated distinctly. The main aim of the education reforms in the country is to enhance "Access, Affordability and Accountability" among the population. Consequently, the evaluation of instructor's performance is especially relevant for the academic institutions as it helps to formulate efficient plans to guarantee quality of instructors and learning process. An intelligent technique and effort in this work is directed at modeling for evaluation of instructor's performance, propose an optimal techniques and designing a system framework suitable for predicting instructor's performance and as well as recommend necessary action to be taken to aid school administrators in decision making considering the limitations of the classical methodologies. The proposed technique will overcome the limitations of the existing techniques; improve reliability and efficiency of instructor's performance evaluation system, provide basis for performance improvement that will optimize student's academic outcomes and improve standard of education. Consequently, it will contribute to successful achievement of the goals and objectives defined in the vision and mission of the new education reform agenda.

Keywords: teacher's Assessment, MHRD, Teacher Performance, education reform agenda, optimal techniques.

1. INTRODUCTION

Data mining is a discovering pattern for searching in data. The process must be automatic or semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage [1]. Educational data mining also referred to as "EDM" is defined as the area of research centered around the development of methods for making discoveries within the unique kinds of data that come from educational sector, and using those methods to better understand students as well as teachers[2]. In the developing countries the recent national policies on higher education mandating high stakes evaluation of instructors and the learning system coupled with the quest for an optimal algorithm for evaluation of instructor's performance in higher institutions. Most research focused on improving the performance of students and improves the curriculum and what is reflected in the educational process, there are a few researches that have been proposed for teacher performance. The main objective of this paper is to improve teacher performance through the study of their expertise and specialization and the time of the period in the service of the educational process, evaluate and determine courses for needy teachers under improving their performance. By offering précised directed courses to the teacher according to his need and build on what he has from previous knowledge. So the training adds new information and knowledge to the experience and improves his performance in the classroom and in the delivery of scientific material for students, and how to manage time and deal with the modern means. The different techniques and Algorithms like Clustering, Classification, Neural Networks, Regression, Artificial Intelligence, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases [3]. This paper investigates the educational domain of data mining using a case study from the teacher data collected from the UCI Machine Learning Repository Teaching Assistant Evaluation Data Set. How can we obtain from the discovered knowledge it showed how could we preprocess the data, how to apply data mining methods on the data. There are many kinds of knowledge can be discovered from the data. In this work we implemented the most common algorithms IBK, J48 and Bagging. The weka 3.6.9 software is used for applying the methods on the teacher's data set. The rest of this paper is organized as follows: Section 2 presents related works in educational data mining. Section 3 describes the methodology performed. Section 4 reports result discussion and analysis on the educational data. Finally we conclude this paper with a conclusion and an outlook for future work in Section 5.

2. RELATED WORKS

There are many works have been already done in the field of educational data mining and performance of the faculty. For improving the performance of students as well as faculty many researchers have been given their review. Some of the related work is given as follow.

Surjeet et al, [4] perform a research on educational data mining to predict student's retention. They used in this study the machine learning algorithms (ID3, C4.5 and ADT) to analyze and extract information from existing student data. They established predictive models and showed that machine learning algorithm such as Alternating Decision Tree (ADT) can learn predictive models from the student retention data accumulated from previous year.

Bharadwaj and Pal [3] performed classification method to evaluate student's performance. The given decision tree method is used for predicting student performance. By this classification method they extract knowledge that describes student's performance in final semester examination. It also helps earlier in identifying the dropouts and students who need special attention to reduce failure ration and allow the teacher to provide appropriate advising or to provide counseling and taking appropriate action for the next semester examination.

Pal and Chaurasia [5] used four classification methods BFTree, J48, RepTree and Simple Cart for analyzing is alcohol affect higher education students performance during their study for higher education. This is a searching and predicting pattern using Data Mining algorithms. In their proposed work they result that the performance of the students affected if they consume alcohol and find that the BFTree Classification with accuracy of 80.2%.

Ola and Pallaniappan [6] conduct an intelligent technique for evaluation of instructor's performance in higher institutions of learning, and suggest an optimal algorithm and designed a system framework which is suitable for predicting instructor's performance. The proposed system, if fully implemented, will aid school administrators in decision making, provide basis for instructor's performance improvement that will optimize student's academic outcomes and improve standard of education. Consequently, this will contribute to successful achievement of the goals.

Surjeet et al, [7] perform a research using C4.5, ID3 and CART decision tree algorithms on engineering student's data to predict their performance in the final exam. Prediction models that include all personal, social, psychological and other environmental variables are necessary for the effective prediction of the performance of the students. C4.5 technique has highest accuracy of 67.7% compared to other methods ID3 and CART algorithms. From the classifiers accuracy the true positive rate of the model for the FAIL class is 0.786 for ID3 and C4.5 decision trees. They can produce short but accurate prediction list for the student by applying the predictive models to the records of incoming new students.

Ahmadi and abadi [8] analyzed the performance of final Teacher Evaluation of a semester of a college and presented the result which is achieved using WEKA tool. Data used in this study were 104 records on teacher's behaviors in classroom with data mining algorithms such Association Rule and decision trees (j48). At teacher's evaluation, evaluation's score of students is very important factor.

Hemaid and El-Halees [9] a study was carried out by to examine the factors associated with the assessment of teacher's performance. In this study, data was collected for teachers from the Ministry of Education and Higher Education in Gaza City. They proposed a model to evaluate their performance through the use of techniques of data mining like association, classification rules (Decision Tree, Rule Induction, K-NN, Naïve Bayesian (Kernel)) to determine ways that can help them to better serve the educational process and hopefully improve their performance and thus reflect it on the performance of teachers in the classroom. In each tasks, they presented the extracted knowledge and described its importance in teacher performance domain.

Chin-Chia Hsu and Tao Huang [10] conducted a study on the use of data mining technology to evaluate student's academic achievement via multiple channels of enrolment like joint recruitment enrolment, athletic enrolment and application enrolment.

Osofisan and Olamiti [11] where they investigated the academic background in relationship with the performance of students in a computer science programme in a Nigerian university. Their study showed that the grade obtained from senior secondary school examination (SSCE) in mathematics is the highest determinant of student's performance using the C4.5 learning algorithm in building the model of the student's performance.

Pal and Chaurasia [12] perform a study on performance of students who consume alcohol during their higher study. Four classifiers such as Sequential minimal optimization (SMO), Bagging, REP Tree and Decision table (DT) were used for diagnosis of performance of the students. Observation shows that bagging performance is having more accuracy, when compared with other three classification methods. The best algorithm based on the student alcohol data is Bagging Classification with accuracy of 80.25 %.

3. METHODOLOGY

This research paper presented the classification method of Data mining for the prediction of teacher's performance. The prediction model based on the Classification methods of the Data mining technique. The lazy IBK, Decision Trees J48 and Meta Bagging data mining technique is implemented in WEKA and their performances were compared to each other. After comparing each method to each other we conclude that IBK performance is better than other two. The WEKA 3.6.9 Data mining software tool was also used to carry out the prediction processes.

3.1. Data Source

The raw data that is used in this study was collected from UCI Machine Learning Repository Teaching Assistant. The data consist of evaluations of teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were divided into 3 roughly equal-sized categories ("poor", "satisfactory", and "average") to form the class variable. as shown in table 1.

Table 1: Teacher's Data variable

VARIABLE NAME	VARIABLE FORMAT	VARIABLE TYPE
English speaker	binary	1= English-speaker, 2= non-English speaker
Course instructor	categorical	25 categories
Course	categorical	26 categories
Summer or regular	binary	1=Summer, 2=Regular
Class size	numerical	1, 2, 3, 4, 5, 6……
Performance	categorical	1=Poor,2=Average, 3=Satisfactory

3.2. Preprocessing of Data Set And Analysis

As part of the data preparation and preprocessing of the data set and to get better input data for data mining techniques, we did some preprocessing for the collected data before loading the data set to the data mining software, irrelevant attributes should be removed. The attributes marked as selected as seen in Table 1 are processed via the Weka software to apply the data mining methods on them. The attributes such as the Teacher_Name or Teacher_ID, etc. are not selected to be part of the mining process; this is because they do not provide any knowledge for the data set processing and they present personal information of the teacher. Here we take six variables which are directly relevant to the performance of the teaching assistant.

4. Results Discussion And Analysis

The proposed model was developed using WEKA. The model was built with three machine learning algorithms: IBK, J48 and Bagging. A comparative analysis of the performance of the models was carried out. Figure 1 shows the visualization of all six attributes in Weka.

The Weighted averages of the models were compared using different performance measures like:

- TP Rate
- FP Rate
- Precision
- Recall
- F-Measure
- ROC

The best model was then selected using Tables 2, Table 3 and Figure 2. The performances of these models were evaluated based on these criteria:

- Accuracy prediction
- Time taken to build the model and
- Different error rate

These are illustrated in table 2. IBK algorithm predicts better than the J48 and Bagging algorithms since its accuracy is the highest compared to others. The results obtained from the analysis demonstrated a slight higher performance of model. Both IBK and J48 algorithms results show great superiority over Bagging algorithm in terms of performance. IBK algorithm performed better than other algorithms not only in terms of the number of correctly classified instances also in terms of RMSE, MAE, RAE. Time taken to build the model by IBK algorithm is less than by two other. By these results we can say that IBK is the best algorithm.

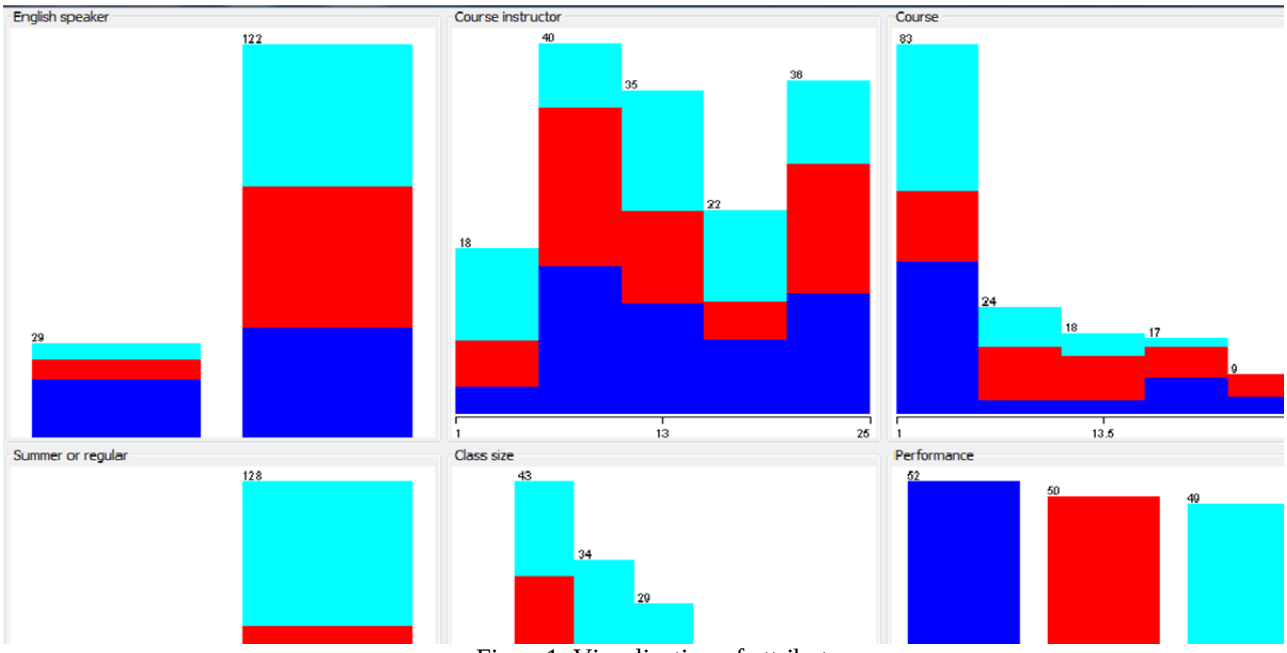

Figure1: Visualization of attributes

Algorithms	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
IBK	0.623	0.188	0.625	0.623	0.622	0.724
J48	0.583	0.209	0.58	0.583	0.581	0.745
Bagging	0.57	0.215	0.568	0.57	0.568	0.732

Table 2: Performance accuracy of the model

Evaluation Criteria	Classifiers		
	IBK	J48	Bagging
Time taken to build model	0	0.03	0.03
Correctly Classified Instances (%)	62.2517	58.2781	56.9536
Incorrectly Classified Instances (%)	37.7483	41.7219	43.0464
Kappa statistic	0.4338	0.3737	0.3538
Mean absolute error	0.2527	0.2929	0.3705
Root mean squared error	0.485	0.4677	0.4329
Relative absolute error (%)	56.8588	65.9158	83.3758
Root relative squared error (%)	102.882	99.2168	91.8207

Table 3: Comparative analysis on the models

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 they can be easily understandable. Fig 4

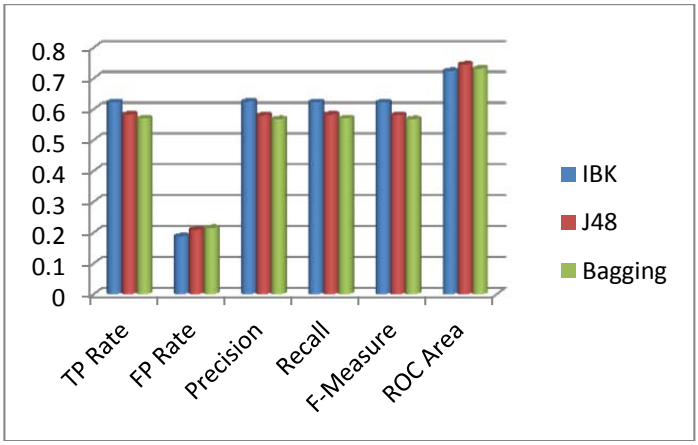

Figure 2: Comparison between performance measure parameters

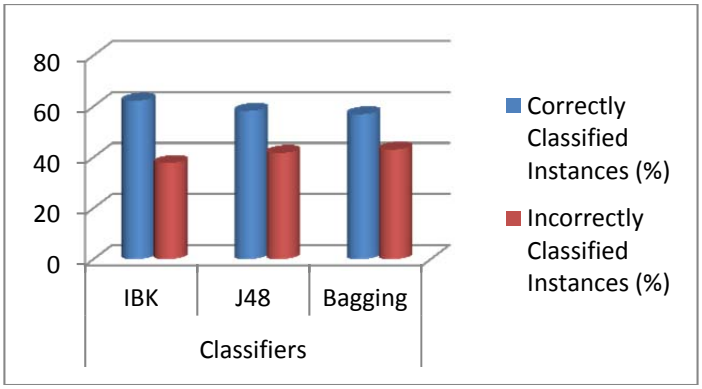

Figure 3: Accuracy prediction of model

.

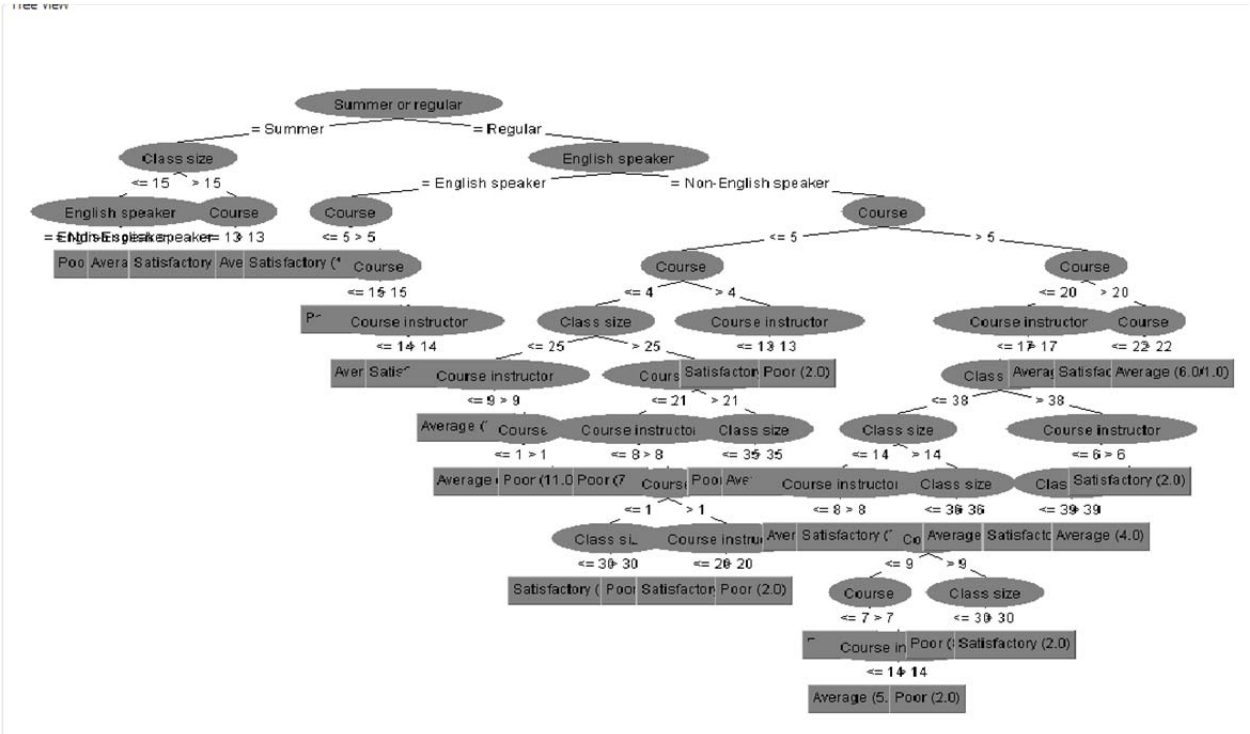

Figure 4: J48 tree

We can summarize the tree as follow:

```
Summer or regular = Summer
|  Class size <= 15
|  |  English speaker = English speaker: Poor (2.0)
|  |  English speaker = Non-English speaker: Average (3.0/1.0)
|  Class size > 15
|  |  Course <= 13: Satisfactory (16.0/2.0)
|  |  Course > 13: Average (2.0)
Summer or regular = Regular
|  English speaker = English speaker
|  |  Course <= 5: Satisfactory (12.0/2.0)
|  |  Course > 5
|  |  |  Course <= 15: Poor (3.0/1.0)
|  |  |  Course > 15
|  |  |  |  Course instructor <= 14: Average (2.0)
|  |  |  |  Course instructor > 14: Satisfactory (3.0/1.0)
|  English speaker = Non-English speaker
|  |  Course <= 5
|  |  |  Course <= 4
|  |  |  |  Class size <= 25
|  |  |  |  |  Course instructor <= 9: Average (3.0)
|  |  |  |  |  Course instructor > 9
|  |  |  |  |  |  Course <= 1: Average (3.0/1.0)
|  |  |  |  |  |  Course > 1: Poor (11.0/2.0)
|  |  |  |  Class size > 25
|  |  |  |  |  Course instructor <= 21
|  |  |  |  |  |  Course instructor <= 8: Poor (7.0)
|  |  |  |  |  |  Course instructor > 8
|  |  |  |  |  |  |  Course <= 1
|  |  |  |  |  |  |  |  Class size <= 30: Satisfactory (3.0/1.0)
|  |  |  |  |  |  |  |  Class size > 30: Poor (5.0)
|  |  |  |  |  |  |  Course > 1
|  |  |  |  |  |  |  |  Course instructor <= 20: Satisfactory (4.0)
|  |  |  |  |  |  |  |  Course instructor > 20: Poor (2.0)
|  |  |  |  |  Course instructor > 21
|  |  |  |  |  |  Class size <= 35: Poor (2.0)
|  |  |  |  |  |  Class size > 35: Average (4.0/1.0)
|  |  |  Course > 4
|  |  |  |  Course instructor <= 13: Satisfactory (3.0)
|  |  |  |  Course instructor > 13: Poor (2.0)
|  |  Course > 5
|  |  |  Course <= 20
|  |  |  |  Course instructor <= 17
|  |  |  |  |  Class size <= 38
|  |  |  |  |  |  Class size <= 14
|  |  |  |  |  |  |  Course instructor <= 8: Average (4.0)
|  |  |  |  |  |  |  Course instructor > 8: Satisfactory (3.0/1.0)
|  |  |  |  |  |  Class size > 14
|  |  |  |  |  |  |  Class size <= 36
|  |  |  |  |  |  |  |  Course <= 9
|  |  |  |  |  |  |  |  |  Course <= 7: Poor (2.0)
|  |  |  |  |  |  |  |  |  Course > 7
|  |  |  |  |  |  |  |  |  |  Course instructor <= 14: Average (5.0/1.0)
|  |  |  |  |  |  |  |  |  |  Course instructor > 14: Poor (2.0)
|  |  |  |  |  |  |  |  Course > 9
|  |  |  |  |  |  |  |  |  Class size <= 30: Poor (8.0/1.0)
|  |  |  |  |  |  |  |  |  Class size > 30: Satisfactory (2.0)
|  |  |  |  |  |  |  Class size > 36: Average (7.0/2.0)
|  |  |  |  |  Class size > 38
|  |  |  |  |  |  Course instructor <= 6
|  |  |  |  |  |  |  Class size <= 39: Satisfactory (2.0)
|  |  |  |  |  |  |  Class size > 39: Average (4.0)
|  |  |  |  |  |  Course instructor > 6: Satisfactory (2.0)
|  |  |  |  Course instructor > 17: Average (6.0)
|  |  |  Course > 20
|  |  |  |  Course <= 22: Satisfactory (6.0)
|  |  |  |  Course > 22: Average (6.0/1.0)
```

5. CONCLUSION

This research paper shows that the performances of classification algorithms used in building a model necessarily indicate that the algorithm that used the least time is the best model to use. IBK used the least time and produce the best result in term of accuracy. Considering the time taken to build the models and performance accuracy level, IBK performance is best than the J48 and Bagging algorithms with good performance of 62.2% accuracy level. This result also shows that the teaching performance over three regular semesters and two summer semesters of 151 teaching assistant (TA) assignments and variable English speaker that contributed mostly to the performance of the teachers in this study. Thus, teacher with good experience of English and experiences with summer and regular semester might likely perform better according to the findings. Another important factor that positively influences teacher's performance is Course instructor, Class size, Course. Finally we can say that data mining techniques plays an important role to judge the performance of teachers by implementing different algorithms. As we implement here three algorithms of data mining likewise another algorithms could be implemented for finding the accuracy in the predicting model.

REFERENCES

1. Witten, I. H. (Ian H.) Data Mining: practical machine learning tools and techniques / Ian H.Witten, Eibe Frank. – 2nd Ed.
2. Baker, R.S.J.d., Corbett, A.T., Aleven, V. (2008) More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing. Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 406-415.
3. Bharadwaj, B. and Pal, S. Mining Educational Data to Analyze Student's Performance, International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
4. Surjeet et al, (2012): Mining Educational Data to Predict Student's Retention: A comparative study, Vol. 10, No 2.
5. S. Pal, Vikas Chauraisa, Is Alcohol Affect Higher Education Students Performance: Searching and Predicting pattern using Data Mining Algorithms, International Journal of Innovations & Advancement in Computer Science, Vol. 6, Issue. 4, april 2017.
6. Ola, A., and Pallaniappan, S., A data mining model for evaluation of instructor's performance in higher institutions of learning using machine learning algorithms, International Journal of Conceptions on Computing and Information Technology Vol. 1, sue 2, Dec' 2013; ISSN: 2345 – 9808.
7. Surjeet et al, (2012): Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 2, 51-56, 2012.
8. Ahmadi,F.,and Abadi, S.(2013): Data Mining in Teacher Evaluation System using WEKA, International Journal of Computer Applications (0975 – 8887) Vol 63 – No.10, February 2013.
9. Hemaid and El-Halees (2015): Improving Teacher Performance using DataMiningInternational Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 2, February 2015.
10. Chin Chia Hsu and Tao Huang (2006): The use of Data Mining Technology to Evaluate Student's Academic Achievement via multiple Channels of Enrolment. An empirical analysis of St. John's University of Technology.
11. Osofisan A.O. and Olamiti A.O. (2009): Academic Background of Students and Performance in Computer Science Programme in a Nigerian University. European Journal of Social Science, Vol. 33 Issues 4. 2009.
12. S. Pal, Vikas Chauraisa, Performance analysis of students consuming alcohol using data mining techniques, International Journal of Advance Researches in Science and Engineering, Vol. 6, Issue. 2, February 2017.