Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/3824
DOI URL: http://dx.doi.org/10.21474/IJAR01/3824
ISSN NO. 2320-5407

RESEARCH ARTICLE

INTEGRATION BETWEEN KEYWORD SPOTTING AND TEXT MINING TECHNIQUES FOR INSTRUCTOR EVALUATION.

Doaa Mohammed Abdella El-bourhamy.
Dept. of Computer Teacher Preparation, Faculty of Specific Education, Kafr El-sheikh University, Egypt.

Manuscript Info

Manuscript History

Received: 05 February 2017
Final Accepted: 08 March 2017
Published: April 2017

Key words:-
Instructor Evaluation; text mining; Decision Tree; Predictive Model.

Abstract

Instructor evaluation is an important field in the educational process because it develops the level of instructor which can improve the educational level of students consequently. In this work, integration between keyword spotting and text mining for prediction with total instructor evaluation. The proposed system is design predictive model for total instructor evaluation by decision tree algorithm. Experimental results demonstrate the effectiveness of the proposed system to predict total instructor evaluation. The proposed system can improve reliability and efficiency of instructors' performance; provide the basis for performance improvement that will affect students' academic outcomes.

Introduction:-

While most previous researches focused on improving the performance of students and developing the curriculum, in addition to all the elements that affect the educational process, but there are few researches that have been proposed for instructor evaluation.

Instructor evaluation is an important field in educational process because it develops the level of instructor which can improve the educational level of students and in the educational process in general [1].

Text data mining methods have been connected in many application domains, for example, Banking, Fraud discovery, Instruction identification and Communication, Marketing, real estate, client relationship administration, designing and web mining[2],[3].

Recently, there are increasing research interests in using data mining and text mining in education, this new important field is called "educational data mining". This sort of data worries with developing methods that discover knowledge from data originating from educational environments and make type of community [4].

This community is assist for the most part with the improvement of exploring data coming from educational settings, and utilizing those techniques to accomplish better comprehension for instructor, students and learning processes [4].

Keyword spotting system (KWS) is a technologically pertinent issue, which plays an important role in sound indexing and speech data mining applications. KWS is also used for locating occurrences of keyword in speech

Corresponding Author:- Doaa Mohammed Abdella El-bourhamy.
Address:- Dept. of Computer Teacher Preparation, Faculty of Specific Education, Kafr El-sheikh University, Egypt.

signal [5]. This problem is like speech recognition, but the additional signals around the words of interest must be disregarded.

Keyword spotting system is to recognize the presence of a small set of pre-determined words in a continuous stream of speech. The process includes recognizing chose keywords in speech utterances containing extraneous (out of vocabulary) speech and noise [6].

So, in this work used keyword spotting system with text mining for instructor evaluation and protection with total evaluation by designing predictive model.

There are many works in this area for improving the performance of instructor and integrating with text mning, sample of these works will be discuss in the following part.

Ola and Pallaniappan [1] used directed modeling and intelligent technique for an assessment of educators' execution in higher establishments of learning, and proposed an ideal calculation and composed a framework system which is suitable for foreseeing instructors' performance. The proposed system, if completely executed, will help school executives in decision making, provide basis for instructors' performance improvement that will optimize students' academic results and enhance standard of education. Consequently, this will contribute to successful accomplishment of the objectives.

In addition, Ahmadi and abadi [7] analyzed the performance of final Teacher Evaluation of a semester of a college and presented the outcome which is accomplished utilizing WEKA tool.
Data utilized as a part of this study were 104 records on instructor's practices in classroom with data mining algorithms such Association Rule and decision trees (j48). At teacher's evaluation, the evaluation's score of students is a very important factor.
Rightness of rules depends on a variety of data sets and statistical instances which can vary. But data mining tools such as WEKA can conclude variety results that help education managers in universities. These results will be utilized by supervisors as a part of decision-making to submit new instructors and proceed with chose old educators.

Ajay and Saurabh [8] discussed the instructors' performance evaluation using data mining techniques at university instructors. The used techniques are Naive Bayes, ID3, CART and LAD tree. Bayes classifier has more precision of 80.35% followed by LAD tree with a percentage of 75.00% and subsequently CART with 65.17%. Content arrangement effects output the most impact. 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.

Mardikyan and Badur [9] conducted a study to understand the key factors affecting the teaching performance of the instructors and identifying the factors associated with the teaching performance; during the period 2004-2009.They used two different data mining techniques; stepwise regression and decision trees. They concluded that a small average relationship exists between learning and the evaluations, but not applicable to all teachers. The employment status of the instructor that is not included in the questionnaire is found to be significant. The instructor attitudes are the most important factor to explain the instructors' teaching performance, which are basically measured by the assessment process. In addition, the attendance of the student is another important factor that impacts positively the performance of the instructor. Hence, the instructors that attract more students to the classes are assessed all the more effectively.

Hemaid and El-Halees [10] examined the factors associated with the assessment of teacher's performance. To improve the instructor performance, great prediction of training course that will be acquired by teacher is a way to reach the highest level of quality in Teacher performance. The real data is collected for instructors from the Ministry of Education and Higher Education in Gaza City, during the period from 2010 to 2013. Teacher data set consists of 813 records and 46 attribute after combining the training, administrative and questionnaire information for those teachers who passed the training successfully.

They used data mining techniques like association, classification rules (Decision Tree, Rule Induction, K-NN, Naïve Bayesian (Kernel)). By applying the K-NN classifier, the model has an precision of 79.92% which is acceptable

accuracy; But By applying the Naïve Bayesian Kernel classifier, the model has an accuracy of 77.46% which is acceptable accuracy.

They show that a small average relationship exists between learning and the assessments but not applicable to all instructors. The most imperative element to clarify the instructors' teaching performance is the instructor attitudes, which are essentially measured by the assessment process. The strategy of instructor in lecture is an imperative variable that impacts positively the performance improve of the student and attract more students to the classes.

So in the present study, the researcher introduces an intelligent system which integrates keyword spotting as method of Speech Recognition with text mining for prediction with total instructor evaluation by designing a predictive model. Used results of instructor evaluation system to design predictive model [11].

Methodology:-

Modern educational organizations start developing and enhancing the educational system. They increase their capability to help the decision makers obtain the right knowledge and make the best decisions, using the new techniques such as data and text mining methods [28]**.** In this section, the researcher used text mining techniques to predict total evaluation of instructor. Fig. 1, shows the predictive model by decision tree technique:

Fig (1):- The Predictive Model

Text mining aims at analyzing text in order to identify the consolidated occurrence of events and use the previous criteria, with data and text mining techniques. Such knowledge can be extracted and accessed via transforming the databases tasks from storing and retrieval into learn and extract knowledge. So, decision tree learning is a common method used in data mining. It is an efficient method for producing classifiers from data and a tree-structured plan of a set of attributes prepared to test in order to predict the output. Furthermore, it is a type of tree-diagram used in determining the optimum course of action, in situations having several possible alternatives with uncertain outcomes [12].

Decision Tree dataset is used to the definition of possible values (total evaluation: Excellent, Good, and Weak). Table 1, shows the list of dependent variables and values.

Table 1:- list of Dependent Variables and Values Used in this Study

Variable name	Data type	Values
Positive concept	Nominal	{yes, no}
Example	Nominal	{EX,G,W}
Communication	Nominal	{high, normal, low}
Joke	Nominal	{EX,G,W}
Content	Nominal	{EX,G,W}

Evaluation score of instructor which are used in the proposed system are shown in tables 2, 3 and 4.

Table 2:- (Evaluation score) for Positive Concept

Raw-Score	*Nominal Representation*
Score<=100	Yes
Score<50	No

Table 3:- (Evaluation Score) for Example, Joke and Content

Raw-Score	*Nominal Representation*
Score<60	Weak
60<=Score<75	Good
75<=Score<=100	Excellent

Table 4:- (Evaluation Score) for Communication

Raw-Score	*Nominal Representation*
Score<60	Low
60<=Score<75	Normal
75<=Score<=100	High

Knowledge Discovery to Predict Instructor Total Evaluation:-

Knowledge Discovery (KD) is an active and important research area with the promise for a high payoff in many business and scientific applications such as instructor performance evaluation. One of the main tasks in KD is classification. A particular efficient method for classification is decision tree. Decision trees have been found very effective for classification especially in Text Mining and Comparing with others. A decision tree is a faster and more accurate. As a very important and widely used technology in data mining, data classification is currently used in many fields. The purpose of data classification is to construct a classification model, which can be mapped to a particular subclass through the data list in the databank. The decision tree algorithm is a more general data classification function approximation algorithm based on machine learning [12].

Decision trees are a classic method of inductive deduction, that is still very famous. They are not just simple to execute and use for classification and relapse tasks, but also good predictive performance, computational efficiency [13].

Decision tree uses information gain measure to choose the splitting attribute. It only accepts categorical attributes in building a tree model. To build a decision tree, information gain is calculated for each and every attribute and selected the attribute with the highest information gain to assign as a root node. Names of the attributes as a root node and the conceivable estimations of the attribute are represented as arcs. Then, all conceivable result occurrences 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. Generally, the splitting attribute is chosen to classify the instances [14].

Let p be the size of the dataset D and p_j the number of samples in class j. assuming that there are K class labels. The entropy theory states that the average amount of information needed to classify a sample is as follows [12]:

$$\text{info}(D) = \sum_{j=1}^{k} \frac{pj}{p} \log_k \left[\frac{p_j}{p}\right] \quad (1)$$

When the dataset D is split into several subsetsD_1, $D_2 \ldots D_n$ according to the outcomes of attributes X, the information gain is defined as[12]:

$$\text{Gain}(X, D) = \text{Entropy}(X) - \text{Entropy}(X, D) \quad (2)$$

In the present study, the researcher designed the decision tree to extract knowledge for instructor evaluation by using Microsoft excel. Table (5) shows the sample of instructor's dataset.

Table 5:- The Sample of Instructor's Dataset [11]

Attributes **Target**

ID	POSITIVE CONCEPT	EXAMPLE	COMMUNICATION	JOKE	CONTENT	TOTAL EVALUATION
1	Yes	Excellent	High	Excellent	Excellent	Excellent
2	Yes	Excellent	High	Excellent	Good	Excellent
3	Yes	Excellent	High	Excellent	Weak	Good
4	Yes	Excellent	High	Good	Excellent	Excellent
5	Yes	Excellent	High	Good	Good	Good
6	No	Good	High	Weak	Weak	Weak
7	No	Weak	High	Excellent	Good	Good
8	No	Good	Normal	Excellent	Good	Good
9	No	Good	Normal	Excellent	Weak	Good
10	No	Good	Low	Excellent	Excellent	Good

Table (6):- Information Gain values

Gain	*value*
Gain(D, positive concept)	0.320614168
Gain(D, example)	0.221191
Gain(D, communication)	0.147974
Gain(D, joke)	0.082622
Gain(D, content)	0.1261

Positive concept had the highest gain; therefore it was used as the root.

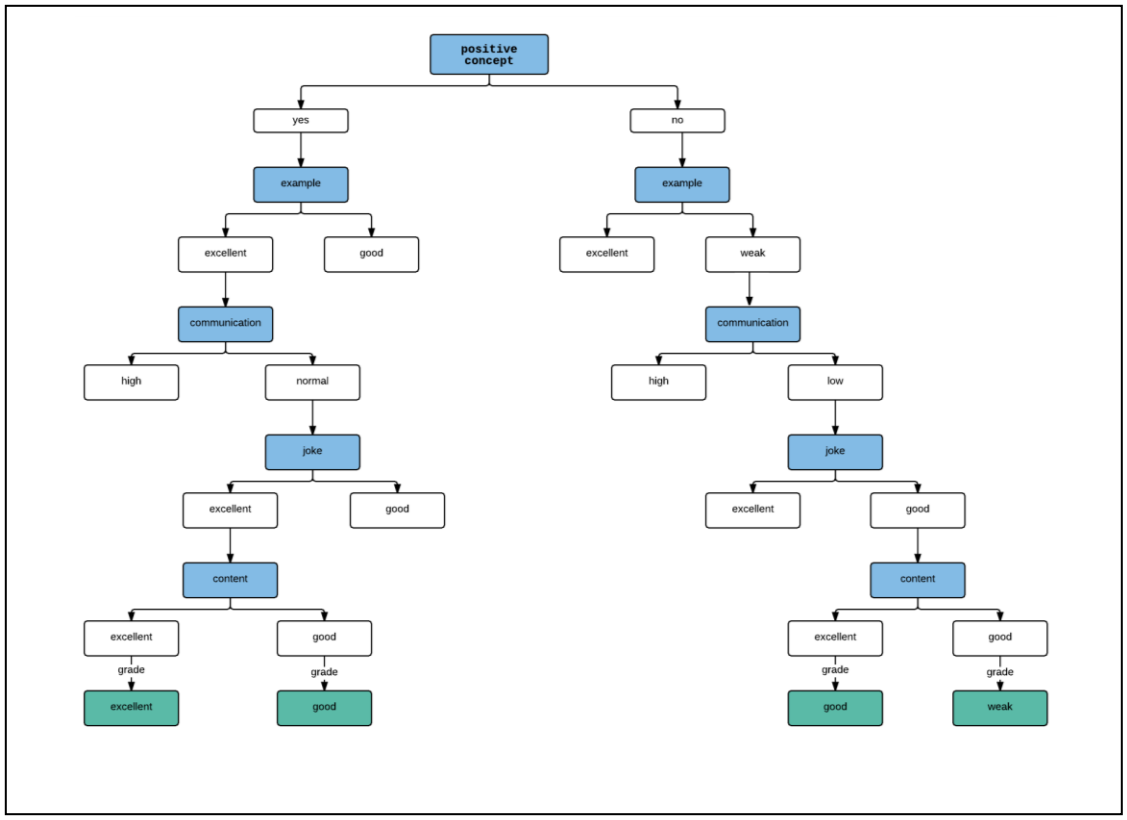

Fig (2):- Sample of Decision Tree

A decision tree can easily be transformed to a set of rules by mapping from the root node to the leaf nodes one by one. Sample of the decision rules is given in the following:

R1:IF(PositiveConcept=YES)and(example=EX)and(comm. =H)and(joke =EX) and (content =EX) -→ THEN target attribute (evaluation)=EX.
R2:IF(PositiveConcept=NO) and (example=EX)and (comm. =H) and (joke =EX) and (content =G) -→ THEN target attribute (evaluation)=G.
R3: IF (PositiveConcept =NO) and (example =W) and (comm. =L) and (joke =G) and (content =G) -→ THEN target attribute (evaluation)=W.
R4: IF (PositiveConcept =NO) and (example =EX) and (comm. =H) and (joke =EX) and (content =EX) -→ THEN target attribute (evaluation)=G.
R5: IF (PositiveConcept =NO) and (example =G) and (comm. =N) and (joke =EX) and (content =W) -→ THEN target attribute (evaluation)=G.
R6: IF (PositiveConcept =YES) and (example =EX) and (comm. =H) and (joke =G) and (content =EX) -→ THEN target attribute (evaluation)=EX.

Experimental Work:-
The Experiments includes 485 sounds from two instructors (speaker dependent system), these sounds consist of 97 keywords, and every word has 5 utterances from 5 different speakers for getting dataset [11] and 50 other instructor data that used in predictive model.

The proposed system is implemented using: MATLAB R2013a for speech processing and designs the proposal system Graphical User Interface (GUI), Microsoft SQL for creating the database, and Microsoft Excel for computing entropy and information gaining to create decision tree.

Predictive Model:-
In the proposed system of the present study, the researcher designed a predictive model to predict total evaluation of instructor through the selection of its degrees in each criterion by pressing "prediction" button as shown in Figure (3):

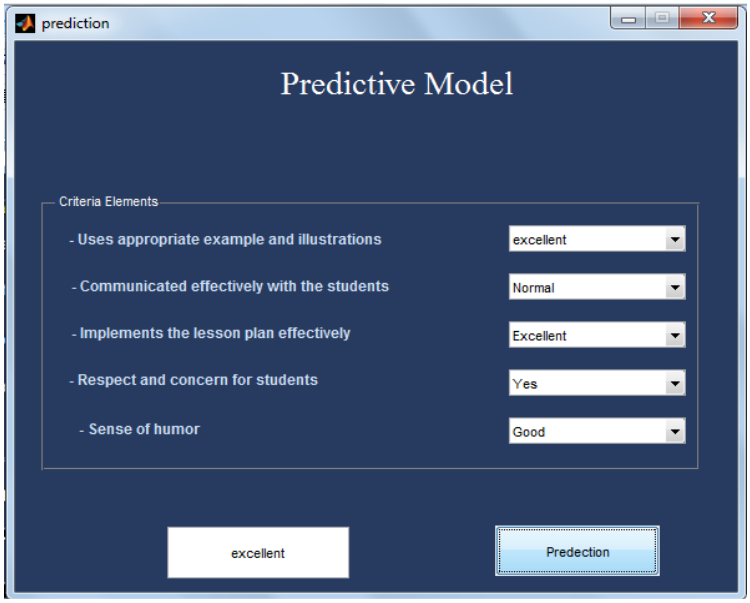

Fig (3):- GUI of Predictive Model

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. That is because of their simplicity and comprehensibility which a professor can easily understand and interpret.

Experimental result, the researcher used built-in function "classregtree" as shown in the following part.

```
load matlab.mat
t = classregtree(Predictor,Response,...
'Names',{'PC' 'EX' 'CM' 'JK' 'CN'})
```

Where "PC"= positive concept, "EX"=example, "CM"= communication, "JK"=joke and "CN"= content.

In this function, the researcher built two tables, the first table was "predictor" that contained the dataset attributes as shown in figure (4), and second table was "response" that contained total evaluation (target) as shown in figure (5):

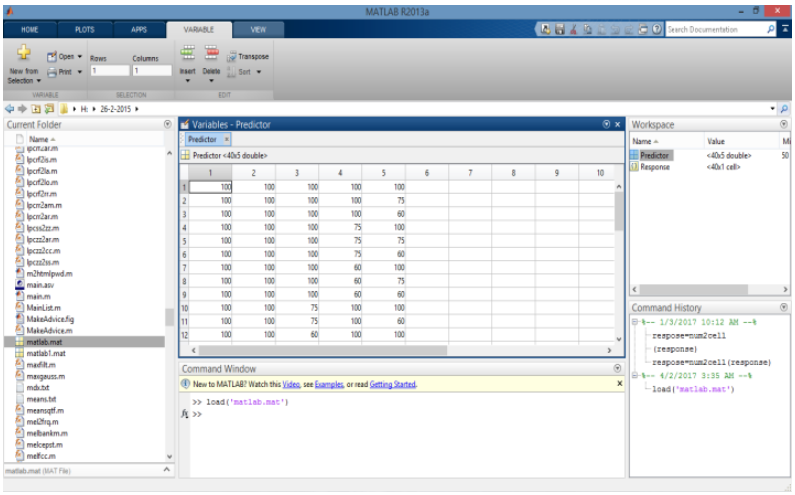

Fig (4):- Table of Dataset

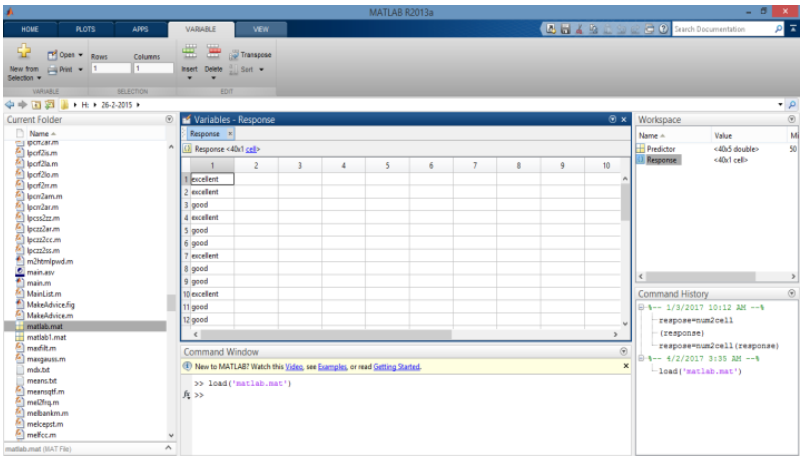

Fig (5):- Table of Total Evaluation (Target) In addition

Figure (6) shows the implementation of previous function:

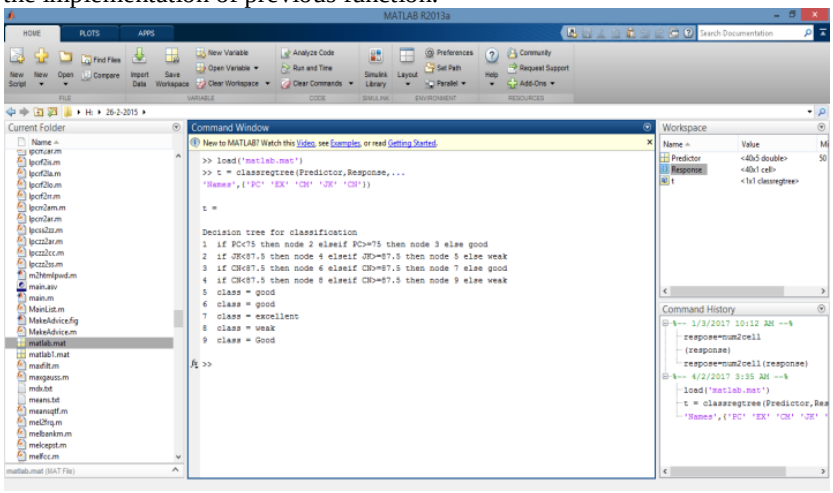

Fig (6):- Implementation Results of "Classregtree" Function

Figure (7) shows the output decision tree:

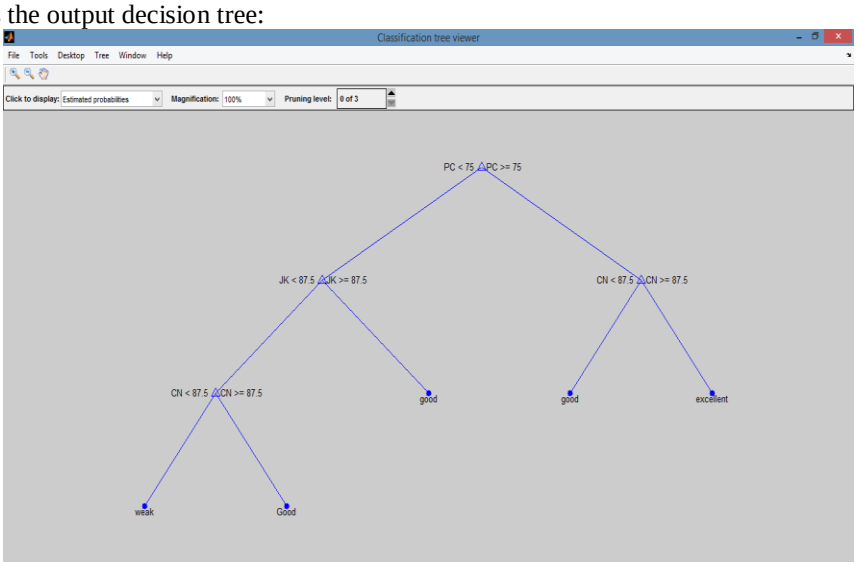

Fig (7):- Decision Tree for Prediction Instructor Total Evaluation

Conclusion:-

In this paper, the proposed system for prediction with total instructor performance evaluation was produced. the researcher used decision tree as a technique of text mining to design the predictive model for predicting instructor total evaluation also, the researcher used built-in function "classregtree" to ensure results, and then extracting knowledge for instructor performance evaluation.

It is worthy to say that evaluation will optimize students' academic outcomes and improve the standard of education. Consequently, this will contribute to the achievement of the goals and objectives which are defined in the vision and mission of the new education reform agenda.

References:-

1. Ola, A., Pallaniappan, S. (2013): "A data mining model for evaluation of instructors' performance in higher institutions of learning using machine learning algorithms", International Journal of Conceptions on Computing and Information Technology, Vol. 1, due 2; ISSN: 2345 – 9808.
2. Naeimeh, D. , et al. (2005): "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System", In Proceedings of the ITHET 6th Annual International Conference, IEEE.
3. Yadav, S., Pal, S. (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.
4. Reategui, E. ,et al. (2011):"Sobek: a Text Mining Tool for Educational Applications. International Conference on Data Mining", Las Vegas, Estados Unidos. Anais do DMIN '11, p. 59-64.
5. Jansen, A., Niyogi, P. (2009):"Point process models for spotting keywords in continuous speech". IEEE Transactions on Audio, Speech, and Language Processing 17(8), p.1457–1470.
6. Gopalan, K., et al. (2009): "An Utterance Recognition Technique for Keyword Spotting by Fusion of Bark Energy and MFCC Features". the Air Force Research Laboratory, Rome, NY, U.S.A.
7. Ahmadi,F., Abadi, S. (2013): "Data Mining in Teacher Evaluation System using WEKA", International Journal of Computer Applications (0975 – 8887)Volume 63 – No.10.
8. Pal, A., Pal, S. (2013): "Evaluation of Teacher's Performance: A Data Mining Approach", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 12, PP.359 – 369.
9. Mardikyan, S. Badur, B. (2011): "Analyzing Teaching Performance of Instructors Using Data Mining Techniques", Informatics in Education, Vol. 10, No. 2, 245–257, Vilnius University.
10. Hemaid, R., El-Halees, A. (2015): "Improving Teacher Performance using Data Mining", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 2.

11. Elgamal, A. Fawzy, M. and Elourhamy, D. (2016):" Intelligent Techniques for Instructor Performance Evaluation", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol.6, No1.
12. Baradwaj, B. K., Pal, S. (2011): "Mining Educational Data to Analyze Students' Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6.
13. Nowozin, S. (2012): "Improved Information Gain Estimates for Decision Tree Induction", Appearing in Proceedings of the 29th International Conference on Machine Learning, Edinburgh, Scotland, UK.
14. Ravindra C., et al. (2012): "Classification by Decision Tree Induction Algorithm to Learn Decision Trees from The Class labeled Training Tuples", International Journal of Advanced Research in Computer Science and Software Engineering 2, No. 4: p.427-434.