A Recommender System to Distinguish between Students' Levels and Evaluate their Attitudes

Shaymaa E. Sorour
Dep. of Computer Science
Kafrelsheikh University Egypt

Hanan E. Abdelkader
Dept. of Computer Science
Mansoura University, Egypt

ABSTRACT
Education is reinforced by identifying new students' attitudes to understand the learning process. Learning analysis is one of the most useful tools to achieve this purpose. Therefore, the present study aims to provide a Recommender System (RS) to distinguish between different students' attitudes (cognitive, emotional, and practical) by creating a dictionary to automatically group common features. In this study, freestyle comments data were collected after each lecture then were analyzed to extract words and sentences' parts (noun, verb, adjective and adverb) for extracting the most common and frequently words and phrases. Thus, a predictive and understandable model was created for the students' estimates. In this paper two types of machine learning techniques were used which are: Support Vector Machine (SVM) and Random Forest (SVM-RF). These techniques are used to extract general rules that distinguish each class of students and providing appropriate automatic feedback that helps student level performance enhancement. Precision, Recall, F-Measure and Accuracy were calculated after each lecture to verify the validity of the results. The experimental results indicated the validity of the automatic feature dictionary; SVM-RF exceeded other techniques to extract general rules.

Keywords
Freestyle Comment Data, Rule Extraction, Recommender System, Machine Techniques

1. INTRODUCTION

The fundamental goal of Recommender Systems (RS) is assisting users in their decision making and providing perfect recommendations in different cases (**Ayub, M., and et al.,2014**). **Recommender systems** find information of concerns that punctually customized according to the users' own priorities (**Bobadilla, and et al., 2013**). Recommender systems are effective to identify subjects that could interest student or predict evaluations that could be given to students (**Fábio, O., and Garrido, C.,2014**).

In this context, clustering can be used to improve predictions or to reduce computational time (**Sebastien, F., and Fabian, L., 2017**). Recommender systems are part of personalized information filtering technologies that aim to give beneficial suggestions which might be of interest to a specific user (**Dietmar, J., and Gerhard, F., 2013**). These recommendations can be made by using three essential steps:

a) Acquire preferences (obtaining user's preferences from their input data).

b) Identify recommendation (recommendations are calculated using proper methods).

c) Present recommendation (Where recommendations are presented to the user) [6].

The most essential role of academic institutions is to provide sufficient support for students to face their learning difficulties and increase their achievement (**Yasmeen, A., Wejdan, A., Al-Turaiki,J., and Muna, A.2016**). Predicting academic performance of students is one of the most important recommender systems application (**Kaklauskas,A. 2015**). Early prediction of student performance can be very useful to guide student learning. It must provide tools to detect trends and behavior patterns for a prediction model to be really useful as an effective aid for learning(**Carlos J., Villagr, a., Francisco J., Gallego, F., and Llorens L. 2017**). Improving student performance, knowing their actual progress and trying to predict their attributes at the earliest stages of the learning process can be extremely important to act early and cut off the problems at the root (**Gerard J., Baars,A., TheoStijnen, Ted, A., and Splinter,W. 2017**).

There is a large amount of techniques to fulfillment prediction models. (**Kotsiantis, S. B. ,2012**). These techniques which contain (classification, association rules, regression algorithms, sequential patterns analysis, clustering and web mining) used for educational purposes. In this study we apply two machine learning techniques which are: Support Vector Machine (SVM) and Random Forest (RF). General rules are extracted to build an understandable prediction model from freestyle comments data. Our goal is to build a recommender system by identifying the factors or attributes from student comments that affect and enhance their achievement. Then classifying Egyptian comment data into main six types of attributes that express students' attitudes in each lecturer in order to give an automatic feedback to students and improve their performance.

In this study, the capability of comment data mining was demonstrated to provide pivotal data of students' situations and attitudes and to improve the learning process quality and prediction results.

The rest of this study is organized as follows: Section (2) illustrates an overview of the related work. Section (3) explains the background of the study. Section (4) present the methodology of the proposed methods. Sections (5,6) discuss experimental results. Al last, Section (7) concludes our research.

2. RELATED WORK

In the classroom, teachers cannot pay attention for all students equally the same and it is important to estimate their learning attitudes at an early stage of their class. And mainly teach more advanced contents to students with high level or others whose results are being improved. Average students have to be analyzed so that teachers can detect problems or difficulties they have. In this section, we deal in more details with the application of Educational Data Mining (EDM) and Learning Analysis (LA) technologies in predicting students' performance.

As discussed in (**Evandro B. Costa and et al., 2017**) comparative study was presented to evaluate Educational Data Mining(EDM) techniques effectiveness to predict students likely to fail in programming courses early. The results showed that, the techniques analyzed in our study were able to early identify students likely to fail, the effectiveness of some of these techniques is improved after applying the data preprocessing and/or algorithms fine-tuning, and the support vector machine technique outperforms the other ones in a statistically significant way.

(**Raheela Asif and et al., 2017**) used data mining methods to study the undergraduate student's learning. It's focused upon predicting students' academic achievement and studying typical progressions and combining them with prediction results. The results indicate that it is possible to provide timely warning and support to low achieving students and advice and opportunities to high performing students.

(**Maria Gogaa and et al., 2014**) determine machine learning techniques as a paradigm in students' performance modeling in higher education. It would be a step to students' performance prediction improvement. This study also aimed at structuring intelligent recommender system, Depends on surroundings tools, which can expect the academic performance of students' at first year and Suggest substantial procedures for their improvement.

(**Yadav, Bhardwaj and Pal, 2012**) created predictive models by the alternative decision tree (ADT) as one of machine learning algorithms using students' records and examined the quality of these models. By These predictive models create short, accurate prediction lists which can be produced for identifying students who need support. The results show that some of machine learning algorithms are capable of generating efficient predictive models from existent retention data of student.

Schalk and et al. (2011) a predictive system was constructed to identify students who have risk of failing in Physics and Mathematics courses. They also choose RF to model a prediction system that uses the data from previous Standard Administration Tests (SAT). However, it is very complex to explain the relationship between input and output.

Amjad Abu Saa (2016) study and analyze students' performance is the most important role Educational Data Mining (EDM). This study assists in Finding out knowledge and desirable patterns in educational institutions. This study examines multiple actions supposed to impact students' performance in higher education. Also the study using personal and social factors to find a specific model to classify and predict the students' performance depends on related person factors.

M.Sindhuja and et.al (2013) Reviews the searching in the attributes of student's higher education like attitude, behavior and relationship. The used data were combined related to the referred attributes. The proposed work explores Behavior Attitude Relationship Clustering (BARC) Algorithm, which proved the students' performance improvement by predicting better behavior, reasonable attitude and Improved in relationship with Tutors and staff of faculty. WEKA tool was used to experiment the analysis where Hierarchical clusters were grouped with related similarities. The resulted analysis from the input parameters description was found optimal. and desirable patterns in educational institutions. This study examines multiple actions supposed to impact students' performance in higher education. Also the study using personal and social factors to find a specific model to classify and predict the students' performance depends on related person factors.

M.Sindhuja et.al (2013) Reviews the searching in the attributes of students' higher education like attitude, behavior and relationship. The used data were combined related to the referred attributes. The proposed work explores Behavior Attitude Relationship Clustering (BARC) Algorithm, which proved the students' performance improvement by predicting better behavior, reasonable attitude and Improved in relationship with Tutors and staff of faculty. WEKA tool was used to experiment the analysis where Hierarchical clusters were grouped with related similarities. The resulted analysis from the input parameters description was found optimal.

3. BACKGROUND
Collecting comment data that reflected students' activities, attitudes and tendencies for each lesson is a hard task. To this end, a teacher continuously tells students the importance of writing comments in order to help them think back and improve their learning. He should give some pointers that lead students to describe their comments well.

In this study, the teacher presented some instructions interventions to students to improve their writing and extract

the necessity assignments from all students and provide general instructions to all students.

Comments data were collected by asking students 5 questions as follows wherefore students could express their cases obviously.

(Question1) Did you review the lesson? How long did you study for this lesson?

(Question 2) What did you find and realize in this lesson?

(Question 3) What did you understand for today's lesson?

(Question 4) Did you cooperate with your partners, get answers or solve your problem?

(Question 5) What is your plan for the next lecture?

For Question (1) students express their efforts for the previous lecturer. Question (2) students describe whether found and realized the new points in the lesson after doing their best and how much they made all possible efforts. Question (3) related to understanding the current lesson and the student can describe the easy and difficult points for the lecturer. In Question (4) students are writing any collaborative activities sharing their partners by solving a problem, teaching or discussing the lesson subject. Question (5) students describe their next plan

4. RECOMMENDER SYSTEM PROCEDURES
Figure.1 presents the procedures of the proposed method to build a recommender system.
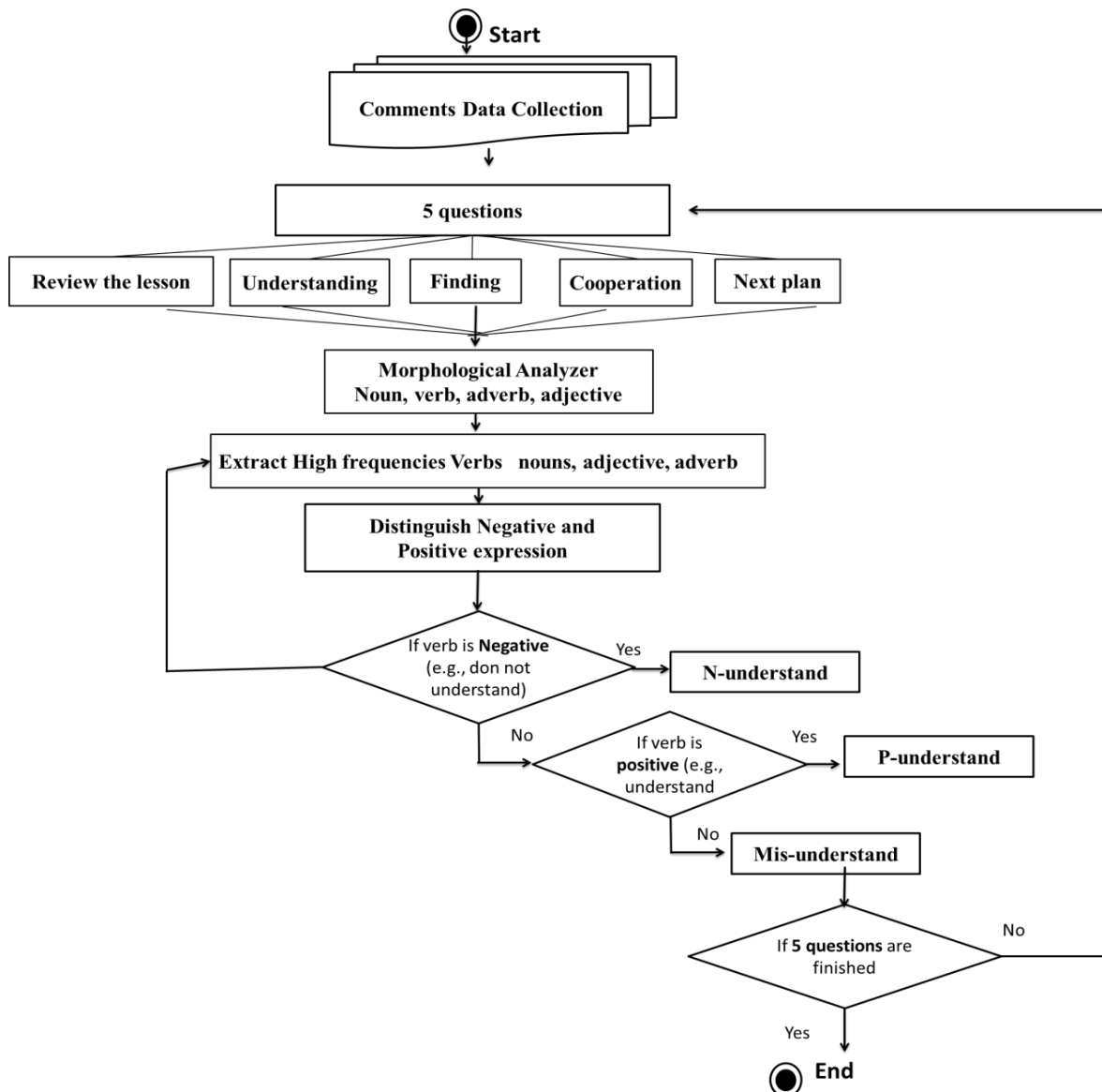
Fig.1 The Procedures of the Proposed Methods

4.1. Comment Data Collection

In the current study comments data were collected form Egyptian students for the first time to predict student's grade. **Table (1)** displays examples comments written by students. Comments data were collected from 103 Egyptian students in one class. They studied programming course that consisted of 12 lessons.

TABLE1: Students Comments

Questions	Example
Question 1:	I reviewed the lesson to be able to solve my problems and check the reference book.
Question 2:	Today, I understood the program well
Question 3:	I found the new idea to solve my problems.
Question 4:	I discussed with my friends to avert the error messages,
Question 5	I will do all my effort to understand the subject and will prepare the next lesson and

4.2. Students' Grades

Students' marks were classified to 5 grades categories are used to classify in order to predict their performance from free style comments. The prediction process is considered correct if only one estimated level with 5-grades category is the actual level of a student. **Table (2)** shows Compatibility between the levels and the mark.

We considered the average mark of student's report to evaluate each student of the three assignments, which contain the topics in lessons (1-4), (5-9) and (10-12), and their attendance rate.

TABLE 2: Students' Levels

Mark	Grade	Total # students
100-85	E	30
84-75	VG	28
74-70	G	30
69-50	A	10
49-0	F	5

The student's number of grade (A) and grade (F) is lower than other grades to enable us predicting their effectiveness. Thus comment data in grade (A) and grade (F) were combined into comments in grade (D) because the number of students is smaller than other levels.

4.3 Data Preparation

The first stage to build a recommender system is to analyze student comments through the parts extracted from speech like (noun, verb, adverb and adjective) of words. Then the method (TF- IDF) was applied by (NLTK) from Python program. Where TF: indicate to a term frequency in a text comment and IDF: indicate to the division result of comments number in the corpus and the comments number that the term contains terms with higher.

4.4 Building a recommender system

To build a recommender system an Attribute Dictionary (AD) was applied automatically which express positive and negative expressions from students' comments. First, we concentrated on negative auxiliary verbs and verbs. In English, not is commonly utilized to display present tense negative verbs, not past tense.
The whole phrases that contain the high frequency words were extracted after negative auxiliary verb stemming. Then, the phrases before verbs till subjects were extracted, furthermore, the whole phrases and the positive verbs that contain higher frequencies of positive expressions were extracted also; which contain words repeated more than 55 times. The number 55 was determined experientially.

To build an Automatic Dictionary (AD) for each question, we determine attribute value dedicated to the extracted words for every student. The value is one of Positive (Pos), Missing (Miss), and Negative (Neg).
Where
Pos: indicates (Positive attribute),
Non: indicates (Missing students who not describe the attribute),
Neg: indicates (Negative attribute).

- Examples of positive attitude are (able to, can do, interesting, - manage to do the best,)
- Examples of negative attitude (cannot do, unable to, confused, frustrated, and afraid).

So, words with high frequencies were collected from the lessons that refer to student's attitudes of the lecturer. Then one of the three values (Pos, Neg or Non) were specified while constructing the Attribute Dictionary (AD).

In experimental results, we concentrate on the negative or positive verbs to define the attribute. Their comments were translated into positive or negative attributes. Such as, understand or not understand. Some of students depict the comments without writing the verbs. Thus, other extracted words were checked to allocate the right attributes. These steps were repeated with (5) questions from lessons (1 to 12).
 Figure 2 displays the six extracted attributes from student comments.
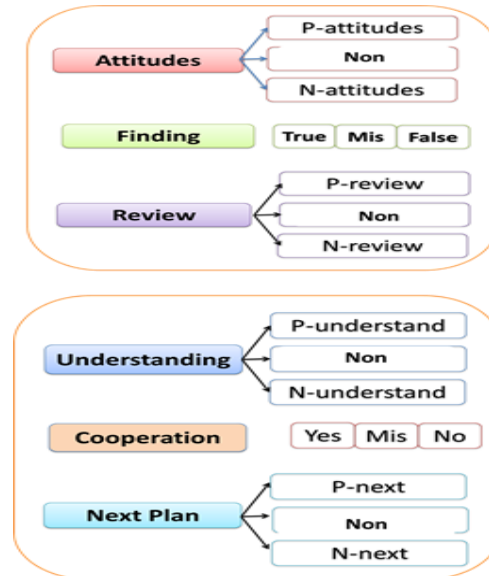

Fig.2 The six attributes

4.5 Prediction Models

In this paper, SVM and RF models are applied to build a recommender system by predicting performance of student and extracting general rules from predictive results. The attribute vectors were generated from students' comments to develop the prediction. After that, the generated vectors of attribute are utilized for training a RF model; therefore, the SVM generates rules representing the concepts learned. WEKA 3.7 was employed as a learning tool.

4.5.1 Random Forests (RF)
(RF): is an effective learning tool for prediction and classification that execute by structure a plenty of decision trees at training data.

The RF model is used to classify a test comment by integrating all results of each trees in the forest.
In this study, to form RF for all lectures (8) trees were used.
A method contains the following steps for every pair of trees was employed to integrate rules from RF, (Breiman, L. (2001):

1-Remove all recurrence rules.
2-Remove all rules with different consequences but the same conditions and with conflict rules and.
3-Remove all rules with a condition set which cnotains another rule's condition.

4.5.2 Support Vector Machine (SVM)
SVM is a supervised model and mostly used in classification problems, but we can't build an understandable prediction model. To control this restriction, trained SVM extracts rules and keep as much of the precision of the SVM as possible. Rule extraction can create understandable descriptions, symbolic with almost the same predictive power as the SVM model itself. The SVM treats the inputs data as a group, while a RF measures the inputs one at a time as grown tree. The issue is to utilize the vectors of attribute generated by the SVM. Then the vectors of attribute from SVM are used by RF with explication ability, which learn what this SVM has learned. (Barakat, N, 2010; Barakat, N., Diederich, J.: 2004). In addition, the system prunes rules that do not increase the accuracy. After that, the test data set is utilized to examine the

extracted rules quality in form of four grades A, B, C and D from (lessons 1 to lesson 12).

5. EVALUATION METRICS

To predict student grades, we used A10-fold cross-validation approach. Let (L) as set of student grades which are {E; VG; G or F}. We measured the effect of text classification by F-measure (F1) where F-measure equal the mean of Precision (P) and Recall (R).

Where

TP: refer to True Positives indicate to predicted positive values which means the value of actual class is yes and the value of predicted class is also yes

Table3: Confusion matrix

	Actual Positive	Actual Negative
Positive Prediction	**TP**	**FP**
Negative Prediction	**FN**	**TN**

The overall F-measure result of the complete classification problem can be calculated by macro-average F_1(Ma.) First, each lesson is calculated partially over each student level (L) by F-measure. Then the average F-measure of overall levels in (Ls) is taken.

The accuracy score (Acc_l) Indicates the ability of the results to correspond the true value of measured quantity

(Metz, 1978). By macro-average $(Acc_l{}^{Ma})$ the accuracy score of the overall classification problem can be measured as follows:

FP: refer to False Positives where actual class is no and predicted
class is yes.

TN: refer to true negative, indicate to predict negative values which
means the value of actual class is no and value of predicted class is
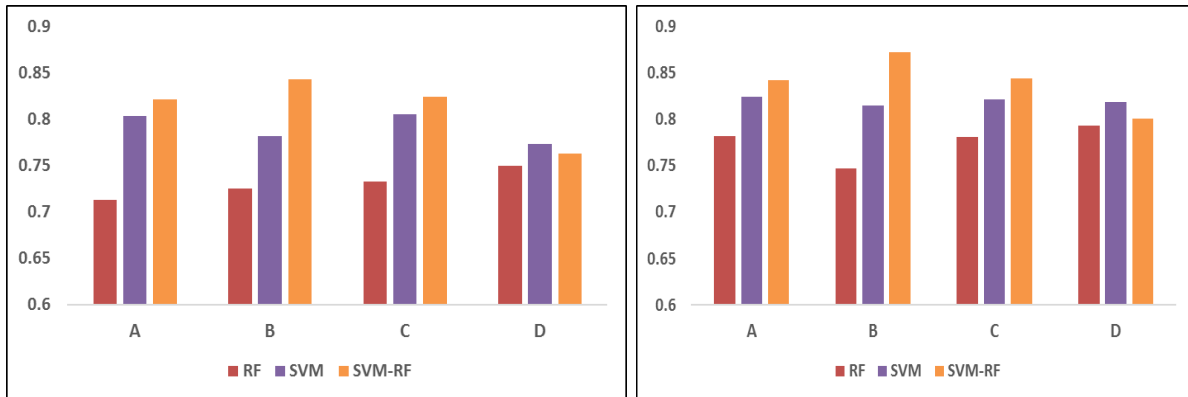also no.

FN: refer to false negative when actual class is yes but predicted
class is no as shown in **Table (3).**

$$Pl = \frac{TPl}{TPl + FPl} \quad , \quad Rl = \frac{TPl}{TPl + FNl} \quad (1)$$

$$F_1 = 2 * \frac{Pl * Rl}{Rl + Pl} \quad , \quad F_1^{Ma} = \frac{\sum_{l \in Ls} F_{1l}}{|Ls|} \quad (2)$$

$$Accl = \frac{TPl + TNl}{TPl + FPl + TNl + FNl} \quad (3)$$

$$Accl^{Ma} = \frac{\sum_{l \in G} Accl}{|Ls|} \quad (4)$$

(a) Before extracting rules

(b) After extracting rules

Fig.3 Overall $F_1^{(Ma)}$ results for each level

6. EXPERIMENTAL RESULTS

In this part, we display the results of SVM, RF and SVM-RF models in students' grades prediction from lessons 1 to lesson 12, and compare these results. Also discuss the effectiveness of the extracted rules for predicting performance of students at each lesson.

6.1 The characteristics of data for each grade

To show the prediction performance between students in each grade, **Figure (3)** show the overall average prediction results with SVM, RF, SVM-RF models by extracted rules and without extracted rules. By using extracted rules, the

predictive results became better. grade VG has the highest prediction results than other grades.

Figure (4) displays the results of $F_1^{(Ma)}$ between RF and SVM-RF from lessons1 to lesson 12 with extracted rules and without extracted rules. As can be seen obviously, the predictive results for SVM RF and RF became higher by using extracted rules and the SVM-RF model had the best results. The predictive results for the two models were higher in the last lessons; students started writing their comments but faced the difficulty of expressing their learning situations. From around the lesson 10, the higher score was achieved by the prediction results. The average F1Ma 1 results without extracted rules were 78.3% and 85.8% for RF and SVM-RF without extracted rules, 82.6 % and 88.8% and the results

using extracted rules became 82.6% and 88.8% for RF and RF-SVM, respectively.

**TABLE4: Prediction results for RF and SVM**

Model	R	P	F	Acc
RF	78.26	78.28	78.26	90.2
RF-rules	77.15	83.07	79.31	91.3
SVM-RF	92.00	85.18	86.95	92.2
SVM-RF(Rules)	86.67	89.66	88.14	90.28
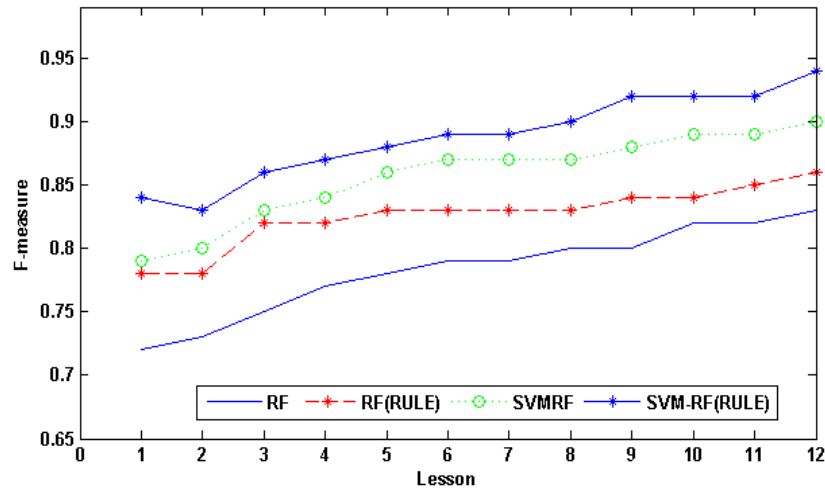
Fig 4: Overall results of prediction for RF and SVM-RF

Figure 5 summarizes the characteristics of patterns attribute of each level from lessons 1 to 12. Using: RF and SVM-RF. The figure summarizes the extracted rules. Using: RF and SVM-RF. Which each level can be distinguished by the same attributes as follows:

- Grade A: students understood most of lessons and present positive attitudes. They cooperated together in most lessons. They are interesting in depicting the outlines that they recognized in each lesson and reviewed the lesson.

- Grade B: almost all students understood the lesson. They achieved positive attitudes toward the lesson.

- Grade C: students did not cooperate with their friends and did not survey the lesson. They had negative attitudes toward the lesson they did not write about finding attribute.

- Grade D: Students did not survey the lessons and understand did not because they had negative attitudes toward these lessons.
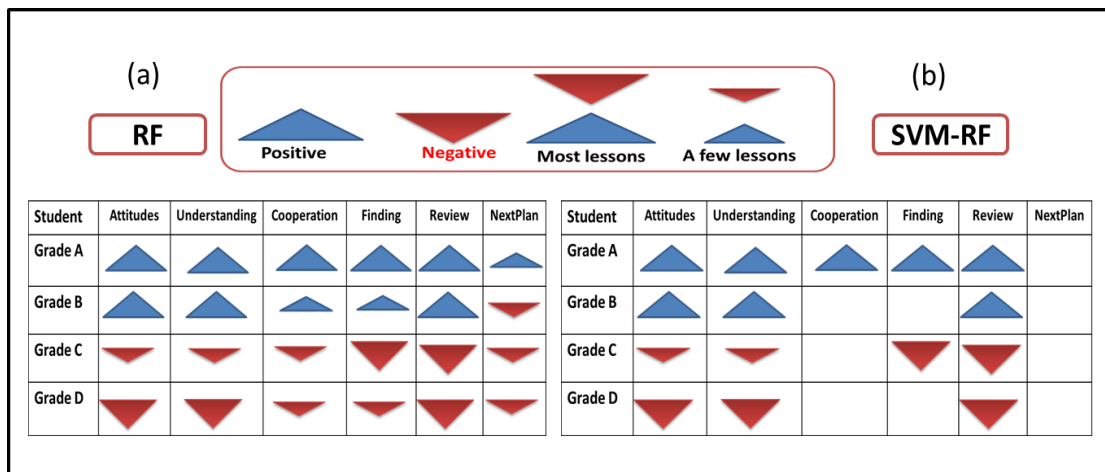
Fig.5 Summarizes the Characteristics of Patterns Attribute

7. CONCLUSION

The main goal of this study was to build a recommender system; to understand students' cases and resolve the predicting students' performance problem. In this study, free-style comments formats were analyzed through multiple methods to totally quantify student's behavior, activities and attitudes in the learning environment.

Although the difficulty of collecting comments data from students, the empirical results display that accuracy / F-measure can detect essential differences between students'

levels and lessons. Furthermore, students' comments informed by detailed indicators concerning the comment format were able to generate the best prediction performance. Otherwise, we can deduce that the extracted rules improved the results of prediction and the prediction results based on attribute vectors have tightly various decisions. This return helps an instructor to accommodate his lesson and focus on each group of student's attribute to provide vital feedback and improve students' performance.

Although expressions (positive or negative) were extracted, some distinctive words were only used to focus on the main six attributes. For future work the recommendations are to extract another attributes from comments of students to process deeper semantic expressions and provide automatic feedback to all students.

8. REFERENCES

[1] Amjad, A. S. (2016). Educational Data Mining & Students' Performance Prediction, (IJACSA) International Journal of Advanced Computer Science and Applications, 7(5).. .

[2] Ayub, M., Cian A., Caliusco, M., and Reynares, E. (2014). Developing an ontology-based team recommender system using EDON method: an experience report, SADIO: Electron. J. of Inform. Operat. Res. 13 1–13.

[3] Barakat, N., Bradley, (2010): A.P.: Rule extraction from support vector machines: a review. Neurocomputing 74(1), 178{190.

[4] . Barakat, N., Diederich, J. (2004) : Learning-based rule-extraction from support vector machines. In: The 14th International Conference on Computer Theory and applications ICCTA'2004

[5] Bobadilla, J. Ortega, F., Hernando, A., and Gutièrrez, A. (2013). Recommender systems survey, Knowl.-Based Syst. 46, 109–132.

[6] Breiman, L. (2001): Decision-tree forests. Machine Learning 45(1), 5-32

[7] Carlos J., Villagr, a., Francisco J., Gallego, F., and Llorens L. (2017) Improving the expressiveness of black-box models for predicting student performance, Computers in Human Behavior, 72, 621-631.

[8] Dietmar, J., and Gerhard, F. (2013).Tutorial: Recommender Systems, International Joint Conference on Artificial Intelligence Beijing, August 4.

[9] Evandro, B., C., B.F., and Marcelo, A. S.(2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Computers in Human Behavior 73 247-56.

[10] Fábio, O., and Garrido, C. (2014).Masters' Courses Recommendation: Exploring Collaborative Filtering and Singular Value Decomposition with Student Profiling, Thesis to obtain the Master of Science Degree in Information Systems and Computer Engineering.

[11] Gerard J., Baars,A., TheoStijnen, Ted, A., and Splinter,W. (2017), A Model to Predict Student Failure in the First Year of the Undergraduate Medical Curriculum, Health Professions Education 3(1), , Pp 5-14.

[12] Kaklauskas,A.(2015). Biometric and Intelligent Decision Making Support, Intelligent Systems Reference Library, Springer International Publishing Switzerland , 81, pp.195-197.

[13] Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades .Artificial Intelligence Review, 37, 331e344. Artif Intell Rev (2012) 37:331–344. DOI 10.1007/s10462-011-9234x. http://dx.doi.org/10.1007/s10462-011-9234-x.

[14] Maria, G., Shade K., and Nicolae, G. (2015) A recommender for improving the student academic performance, The 6th International Conference Edu World 2014 "Education Facing Contemporary World Issues", 7th - 9th November 2014, Procedia - Social and Behavioral Sciences, 180, 1481 – 1488.

[15] Nguyen, T., Lucas, D., Artus, K., Lars, S.(2010). Recommender System for Predicting Student Performance, 1st Workshop on Recommender Systems for Technology Enhanced Learning, Procedia Computer Science 1, 2811–2819.

[16] Raheela, A., Agathe, M., Syed, A., Najmi,. G.(2017) Analyzing undergraduate students' performance using educational data mining, Computers & Education 113, 177-194.

[17] Rahimpour, C.B., Hamid, H., Hoda, M. (2017). User trends modeling for a content-based recommender system, Expert Systems With Applications 87,209–219.

[18] Rodriguez, A., Gago, J., Rifo, L., and Rodriguez, R. (2015). A recommender system for non-traditional educational resources: a semantic approach, J. Univ. Comput. Sci. 21 306–325.

[19] Schalk, P. D., Wick, D. P., Turner, P. R., and Ramsdell, M. W. (2011). Predictive assessment of student performance for early strategic guidance. In Frontiers in education, conference (FIE), 2011. http://dx.doi.org/10.1109/FIE.2011.6143086, S2H-S2H-5.

[20] Sebastien, F., and Fabian, L. (2017). Weighting strategies for a recommender system using item clustering based on genres, Expert Systems With Applications, 77, 105–113.

[21] Sindhuja, M, Rajalakshmi,S., Nandagopal,S.M.(2013). Prediction and Analysis of students Behavior using BARC Algorithm, International Journal on Computer Science and Engineering (IJCSE) ISSN : 0975-3397, 5(6), pp. 474-480.

[22] Yadav S. K., Bhardwaj B. K. and Pal S. (2012). Mining Education Data to Predict Student's Retention: A comparative study, International Journal of Computer Science and Information Security (IJCSIS), 10 (2), 113-117.

[23] Yasmeen, A., Wejdan, A., Al-Turaiki,J., and Muna, A.(2016). Predicting Critical Courses Affecting Students Performance: A Case Study, Symposium on Data Mining Applications, SDMA, 30 March, Riyadh, Saudi Arabi.