Prediction of Students' Academic Success Using Data Mining Methods

Vahide Nida Uzel
Department of Computer Engineering
Adana Science and Technology University
Adana, Turkey
vnuzel@adnabtu.edu.tr

Sultan Sevgi Turgut
Department of Computer Engineering
Cukurova University
Adana, Turkey
ssturgut@student.cu.edu.tr

Selma Ayşe Özel
Department of Computer Engineering
Cukurova University
Adana, Turkey
saozel@cu.edu.tr

***Abstract*—Success is very important for all of us. Most people wants prosperity, reputation, and richness that can only be achieved with the success. A society that wants to be successful should pay attention to their new generation because they are the future of the world. If we want to invest to our future, we must contribute to success of our new generations. Therefore, in this study, the academic performance of the students that belong to different levels of education like primary, secondary, and high school levels is tried to be determined by applying various classification methods such as Multilayer Perceptron (MLP), Random Forest (RF), Naïve Bayes (NB), Decision Tree (J48), and Voting classifiers. It is also observed which characteristics are more related to the improvement of academic performance of the students. Features like absence of student, parent's school satisfaction, raising hands on class, and parent who is responsible for the student can affect the success of the student. A comparison is made with other study that previously worked on the same data set. As a result, better classification accuracy is achieved. We observe the best classification accuracy as 80.6% by Voting classifier, while the previous study has the highest accuracy as 79.1% by applying Artificial Neural Network (ANN) classifier. Also, in our study, Apriori algorithm is applied to detect relationships between features.**

Keywords— Data mining, student evaluation, classification, apriori, machine learning.

I. Introduction

Every student wants to be successful. Also, governments try to increase the success of students for guaranteeing a bright future for their country. So, improving the academic performance of the students is very important by bringing the education system to a better level. How can the academic performance of students increase? The success of students changes according to several conditions. If these conditions can be determined, the achievement of the students can be increased. Which conditions can affect the success of students? Conditions like "attitudes of students and their parents", "grades of students", "difficulty of the courses", etc. may be important. To improve the students' academic performance, first of all we need to determine which factors are the most effective ones over the academic success. For this purpose, various data mining methods have been applied over several datasets that were prepared in the past. Today, new datasets continue to be generated, and new data mining techniques are applied to have better results.

Data mining is a procedure for extracting meaningful information from datasets by using various machine learning techniques. Recently, educational data mining (EDM) which is a field that uses data mining techniques to predict student performance, has emerged. EDM also determines the factors that affect achievement of students, makes inferences by using these factors, and provides improvements for the education system [1].

In this study, an educational dataset (xAPI) which is generated from an e-learning system by [2] is used. The dataset includes information about 480 students from various levels like primary, secondary, and high school levels; and contains 16 attributes. These attributes are divided into three categories namely as "demographic", "academic", and "behavioral information". 305 male and 175 female students who are from 14 different countries (e.g Kuwait, Jordan, Iraq etc.) exist in the dataset. According to 16 attributes, previous study [2] classifies the success level of students into three categories as "low", "medium", or "high". To examine this dataset there are a wide number of data mining methods. But, this dataset is actually created for classification. Thus, most researchers handle this dataset with using classification methods. However, in this study we use both classification, and association rule mining to make further analysis over this dataset.

Association rule mining (ARM) was first revealed by Agrawal et al. in 1993 [3]. ARM tries to find frequent itemsets to reveal interesting relationships from the dataset. In order to deduce association rules, the first algorithm proposed is Apriori algorithm which is developed by Agrawal and Srikant in 1994 [4]. Apriori algorithm explores the dataset at multiple passes; and finds many candidate frequent itemsets. After iterations, longest frequent itemsets are accepted, and they are used to generate interesting association rules [5].

In this study, we process this dataset by applying both classification methods and association rule mining methods. We aim to improve classification performance and reveal interesting relationship among attributes to draw conclusions about the factors that affect students' academic successes. Therefore, four classifiers are used separately, and some of them are selected for voting method to improve classification performance of the previous studies. Classifiers applied are Decision Tree (J48), Random Forest (RF), Naive Bayes (NB), and Multilayer Perceptron (MLP). MLP, J48, and RF are also used together as a classifier in the voting method.

The rest of the paper is organized as follows: in the second section a brief overview of the related work is presented. The dataset used, and the applied methods are described in the third section. The fourth section presents and discusses the experimental results. Finally, section five concludes our study.

II. Related Work

Amrieh et al. [6] have presented a student performance model by examining xAPI dataset. Firstly, the dataset has been pre-processed, and it has been used to examine the effects of student behavior as new attributes such as "following the lesson", "participating in the exercises", and "individual/group studies" on the academic performance.

978-1-5386-7786-5/18/$31.00

Then, the dataset is classified by using MLP, NB, and J48 methods. It has been observed that the classification results improve by including behavioral information, for example, the success rate in J48 method increase from 57.0% to 75.8% when behavioral attributes of the students are considered. Thus, they decided that there is a correlation between students' behavior and their success. Also, Bagging, Boosting, and RF ensemble methods were applied to get better classification accuracy which is improved up to 79.1%.

Acharya et al. [7] used classification methods for predicting student performance. Their aim is to define successful or insufficient students by applying classification algorithms with the most proper features. They used a dataset that consists of 309 students from some of the colleges in Kolkata. Five classification methods that are C4.5, Naive Bayesian Network (NBN), MLP, 1-Nearest Neighborhood (1-NN), and Sequential Minimal Optimization (SMO) have been applied to the dataset that includes student's information such as "gender", "family size", "attendance to a lesson", and "family income" etc. According to the experimental evaluation, average F-measure for training and testing sets are 0.79 and 0.66, respectively. They decided that the most appropriate algorithm is C4.5 decision tree method, and "family income" attribute has great influence on the success of the student.

According to Daud et al. [9] previous studies have examined academic performance in relation to family existence and family income. Unlike other studies, they also included "expenditure of the family", "their personal views", and "personal information" into consideration. By using these additional features, they can decide more correctly whether a student could complete his/her degree or not. Support Vector Machine (SVM), C4.5, Classification and Regression Tree (CART), Bayes Network (BN), and NB classifiers were used for classification. As a result, they have found that including "expenditure of the family", "their personal views", and "personal information" features gives better classification results than other studies that have not employed these features.

Oyedotun et al. [10] used the questionnaire made by the students at Ankara Gazi University. 28 questions about student satisfaction for their enrolled courses were asked, and in addition the difficulty of the course and the frequency of attendance to the course were added as features. According to the answers given by the students, they tried to estimate how many times the student repeated the lecture. The number of repetitions is divided into 3 categories and students can repeat at most 3 times. In order to estimate the number of repetitions, two neural network methods were used. These are back propagation neural networks, and radial basis function networks. As a result, backpropagation network outperforms the radial basis network.

There exists a lot of works done for student performance appraisal, but most of them use classification methods. And it has been noticed that, the classification performances obtained in the previous studies may be further increased. Moreover, finding the relationships between the factors influencing the performance of the students is thought to be a major contribution to improve the achievement of the students. Therefore, in this study we focus on these issues.

III. Materials and Methods

In this study our first aim is to improve the classification accuracy to determine the success levels of the students. Therefore, we apply and compare several classifiers that are MLP, RF, NB, J48, and voting method. Our second aim is to investigate the relationships between features on the success levels of students. So, we apply Apriori algorithm and generate association rules to reach our second goal. In this section, the used dataset and the applied methods are explained in more detail.

A. Dataset

An educational dataset (xAPI) [2] which is generated from an e-learning system is used in this study. This dataset includes 480 students (i.e., instances) from different levels such as primary, secondary, and high school; and 16 attributes that are "nationality", "gender", "place of birth", "parent responsible for the student", "school level", "grade level", "classroom that student belongs to", "semester", "course topic", "student absence days (under 7, above 7)", "parent answering survey", "parent-school satisfaction", "discussion groups", "visited resources", "raised hands on class", and "viewing announcements". According to these attributes, success levels of students are classified into three categories as "low", "medium", and "high".

In this study, ten-fold cross validation is applied for training and testing.

B. Association Rule Mining

Association rule mining is a data mining technique that tries to find frequent patterns, correlations, associations, or causal structures between objects in the dataset. It can be used to find objects that together make sense. Generally, the objects are frequently purchased items by a customer in one transaction. Therefore, by applying association rule mining, we can recommend an item to buy for a customer when he buys (or examines) one of the frequently purchased items to increase revenue of our market. For example, a mouse is suggested to a customer who is looking for a computer on a website. However, every frequent itemset may not form an interesting association rule. Generating association rules depends on several measurements such as confidence, support, lift, etc. These measurements are used as thresholds, such that itemsets that have lower confidence, support, or lift values than the predefined minimum thresholds are not accepted as frequent. The support and confidence measures are sometimes insufficient to eliminate most of the non-interesting rules. Therefore, it may be necessary to examine the correlation among the items. There are various correlation measures, one of which is lift. If the lift value is less than 1, it means that there is a negative correlation, or if it is greater than 1 there is a positive correlation between the items, if lift is equal to 1, this means there is no correlation [11], [12]. Support, confidence, and lift formulas are given in equations 1, 2, and 3, respectively; where X and Y are considered as frequent itemsets, and $X \rightarrow Y$ is an association rule.

$$\text{Support(X)} = \frac{\text{frequency of } (X \cap Y)}{\text{total number of transactions}} \tag{1}$$

$$\text{Confidence(X}\rightarrow\text{Y)} = \frac{\text{frequency of } (X \cup Y)}{\text{frequency of X}} \tag{2}$$

$$\text{Lift } (X \rightarrow Y) = \frac{\text{frequency of } (X \cup Y)}{(\text{frequency of } X)*(\text{frequency of } Y)} \quad (3)$$

C. Apriori

The most popular frequent itemset mining algorithm is the Apriori algorithm which is proposed by Agrawal and Srikant in 1994. Apriori algorithm scans the transactional database at multiple iterations and creates candidate lists that satisfying the predefined minimum support value. Purpose of generating a candidate list is to reach the longest frequent itemsets. Firstly, a database is explored to identify frequent 1-itemsets which have support greater than the user defined minimum support. Then, frequent 1-itemsets are joined to create the candidate 2-itemsets. After that support of candidate 2-itemsets are computed to generate frequent 2-itemsets. By using frequent k-itemsets, candidate (k+1)-itemsets are generated and their support is computed to find frequent (k+1)-itemsets. Iterations continue until no new frequent itemset found from the dataset [5], [12].

In this study, Apriori is used for discovering relations between features which may affect the success of students. For example, "school satisfaction of the student" and "the number of days that the student does not come to the school" together may influence success of the student. This result can be obtained by association rule mining.

D. Classifiers

For this study, four classifiers and one voting method are used. Classifiers are J48, RF, NB, and MLP from the Weka data mining tool. MLP, J48, and RF are also used together as classifiers in the voting method to improve classification accuracy.

Decision tree [13] is a tree structure-based classifier so that, each internal node represents a feature whereas the leaves represent class labels, and the branches that go to the internal nodes or leaves show values of the features. By using a training dataset, a decision tree is learned which is then used for assigning class labels to new instances. To assign class labels to new instances, process starts from the root node and each path to a leaf node from the root forms a classification rule. Instances satisfying the rule are labeled as the class of the leaf node. If more than one rule are satisfied for the same test instance, conflict resolution mechanisms are applied.

J48 [13] is a Weka implementation of C4.5 decision tree induction algorithm. J48 can handle missing values in the dataset and it is very popular method. RF [14] is also another decision tree induction algorithm. Instead of learning one decision tree, RF aims to increase the classification accuracy rate by learning and using more than one decision tree.

NB [15] is a supervised probabilistic classifier which is based on Bayes' theorem. It assumes that the features are independent from each other and computes conditional probabilities of each feature value for each class by using the training dataset. To find out which class a given data instance belongs to, the conditional probability values for the features of the data instance are multiplied for each class; and the data object is assigned to a class which has the highest probability.

Artificial Neural Network (ANN) [16] is a model that is inspired by the human brain and the nervous system. ANN is a structure established as layers. The first layer is called as the input layer, and the last layer is called as the output layer. Layers in the middle are called as hidden layers. Each layer contains a certain number of neurons which are linked to each other by synapses. A synapse contains a coefficient which tells that how important information in the neurons is. MLP [16] is a type of ANN and it uses backpropagation for training the network. After the training process, data instance to be classified is given to the input layer, and the output layer gives the class label.

Voting method [17] is a procedure that combines multiple classifiers. This approach is a highly preferred method to increase the performance of a single classifier and reduce incorrect classifications. Voting method can be applied as weighted and unweighted ways. For unweighted method, all classifiers have equal importance; and a data instance is classified according to the highest vote value. For weighted method, classifiers have different significance. There are different weighting techniques, such as simple weighted vote, re-scaled weighted vote, and best-worst weighted vote, etc. The goal of weighted vote methods is to give more weight to good-performing classifiers and prevent the effectiveness of good classifiers to be reduced by other classifiers.

Classification methods are used for predicting success level and absence days of students. These classification results help us to evaluate students' academic performances.

IV. Experimental Results and Discussions

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Environment

Weka [18] data mining tool is used for classifiers, Apriori algorithm, and voting method.

B. Evaluation Measures

Four measures are used for computing classification quality. These are Accuracy, Precision, Recall, and F-measure. Confusion matrix that is shown in Table I is used to compute these four measures.

Accuracy is the percentage of correctly classified instances in the test dataset. Precision is the ratio of true positives to all instances that are labelled as positive. Recall is the ratio of true positives to all positive instances in the test set. F-measure is the harmonic mean of precision and recall. Equations 4, 5, 6, and 7 describes how to compute these four measures [19].

TABLE I. CONFUSION MATRIX

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (4)$$

$$Presicion = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$F - measure = 2\frac{Presicion * Recall}{Presicion + Recall} \quad (7)$$

C. *Experimental Evaluation and Results*

Amrieh et al. [6] classified the same dataset that we use in the experiments. Their aim is to predict success levels of students by applying ANN, J48, and NB classifiers. Also, boosting method is applied to obtain better accuracy with these classifiers. The best classification accuracy achieved in [6] is equal to 79.1%.

In our study, we also try to classify students according to their success levels as "low", "medium", and "high". In addition to this we apply classifiers to predict "absences of students" as "under 7 days", and "above 7 days". Our aim in this study is to get better classification accuracies with respect to the previous study; and find relationships between features which may affect the success levels of students. The experimental results are presented in Table II, III, and IV where accuracy, recall, precision, and F-measure values are represented with A, R, P, and F respectively.

We apply MLP, RF, NB, J48, and Voting (MLP+J48+RF) methods as classifiers. In Table II, our results and the results that are obtained by Amrieh et al. [6] are listed. According to Table II, the best classification accuracy obtained by [6] is 79.1% which is measured by using ANN classifier. However, in our study, we observe slightly better classification accuracy when we apply Voting classifier which is equal to 80.6%. For other methods we have almost the same results.

We also apply boosting [20] which helps to increase or improve performance of a classifier. Boosting reduces bias, and also variance in the supervised learning method, and converts weak classifiers to strong ones. The results obtained after applying boosting to the classifiers to predict students' success levels are presented in Table III. According to Table III, after boosting, classification performances of J48 and NB classifiers increase; however classification performances of Voting and RF decrease. The results for J48 classifier are different in our study and in [6]. This difference may occur due to the fact that J48 is a greedy approach and it selects the current best feature to be used in the decision tree. When boosting is applied, data distribution in the training set may change which may cause different feature to be selected as the root node. When the root node changes, J48 will generate completely different decision tree.

We also perform another experiment to predict students' absence level by applying classifiers. According to features in the dataset, we try to classify students as whether their absence levels are under 7 days or not. The classification results for this experiment are presented in Table IV. As it can be easily seen from Table IV, we can predict students' absence level with high accuracy such that MLP, RF, NB, and J48 classifiers have accuracies 84.5%, 87.2%, 77.1%, and 84.1%, respectively. The best classifier is RF for predicting students' absence level.

TABLE II. CLASSIFICATION PERFORMANCE TO PREDICT THE SUCCESS LEVELS OF STUDENTS

	Our results					Results of [6]		
	MLP	**RF**	**NB**	**J48**	**Voting**	**ANN**	**J48**	**NB**
A	78.3	76.6	67.7	75.8	**80.6**	79.1	75.8	67.7
R	78.3	76.7	67.7	75.8	**80.6**	79.2	75.8	67.7
P	78.3	76.6	67.5	76.0	**80.6**	79.1	76.0	67.5
F	78.3	76.6	67.1	75.9	**80.6**	79.1	75.9	67.1

TABLE III. CLASSIFICATION PERFORMANCE AFTER APPLYING BOOSTING TO PREDICT THE SUCCESS LEVELS OF STUDENTS

	Our results					Results of [6]		
	MLP	**RF**	**NB**	**J48**	**Voting**	**ANN**	**J48**	**NB**
A	78.3	76.2	**72.2**	**77.9**	78.3	79.1	77.7	72.2
R	78.3	76.3	**72.3**	**77.9**	78.3	79.2	77.7	72.3
P	78.3	76.3	**72.4**	**77.9**	78.4	79.1	77.8	72.4
F	78.3	76.2	**71.8**	**77.9**	78.3	79.1	77.7	71.8

TABLE IV. CLASSIFICATION PERFORMANCE TO PREDICT ABSENCE OF STUDENTS

	MLP	**RF**	**NB**	**DT(J48)**
Accuracy	84.5	**87.2**	77.1	84.1
Recall	84.6	**87.3**	77.1	84.2
Precision	84.5	**87.3**	77.1	84.1
F-measure	84.5	**87.2**	77.1	84.0

We also apply Apriori to discover relationships between features and success levels of students. Dataset format is changed for applying Apriori algorithm. Minimum support value is determined as 0.2 experimentally to discover interesting rules.

Table V lists the top 15 association rules which have the highest confidence values. When we examine the rules in the table, we can see the relationships among the features. For example, according to the first rule in the table, "IF student gender is male AND father is the relative who is interested in student AND parent does not answer school survey AND parent does not satisfy with the school AND student is absent more than 7 times, THEN student success level is low" with 88% of confidence. In Table V, gender values M and F represent Male and Female, respectively; and success values L, M, and H are used for Low, Medium, and High, respectively.

Generally, it can be concluded from Table V that, students who are not dealt with by their families and who have above-7 days absence have low success level. On the contrary, it is seen that, the families of successful students are interested in them, and these students have under-7 days absence. The importance of attendance to the courses has become apparent. It is also observed that even though the exact conclusions cannot be drawn from gender (i.e., Male or Female), and relation (i.e., Father, and Mother) attributes, it is argued that female students are successful with their participation in the classes, and male students have more rules that show their unsuccessfulness than female students.

TABLE V. ASSOCIATION RULES GENERATED BY APRIORI ALGORITHM

Order	RULES	CONF
1	gender=M, Relation=Father, ParentAnsweringSurvey=No, ParentschoolSatisfaction=Bad, StudentAbsenceDays=Above-7 ==> Success=L	0.88
2	gender=F, Relation=Mother, ParentschoolSatisfaction=Good, StudentAbsenceDays=Under-7 ==> Success=H	0.85
3	StudentAbsenceDays=Under-7, Success=H, ParentschoolSatisfaction=Good ==> ParentAnsweringSurvey=Yes	0.84
4	Relation=Father, ParentschoolSatisfaction=Bad ==> ParentAnsweringSurvey=No	0.81
5	SectionID=A, ParentschoolSatisfaction=Bad ==> ParentAnsweringSurvey=No	0.85
6	SectionID=A, ParentAnsweringSurvey=Yes, Relation=Mother ==> ParentschoolSatisfaction=Good	0.97
7	ParentAnsweringSurvey=No, Relation=Father, Success=L ==> ParentschoolSatisfaction=Bad	0.91
8	SectionID=A, ParentAnsweringSurvey=Yes, StudentAbsenceDays=Under-7 ==> ParentschoolSatisfaction=Good	0.93
9	SectionID=A, StudentAbsenceDays=Under-7, Success=H ==> ParentschoolSatisfaction=Good	0.9
10	SectionID=A, ParentschoolSatisfaction=Good, Success=H ==> StudentAbsenceDays=Under-7	1
11	Relation=Mother, ParentschoolSatisfaction=Good, Success=H ==> StudentAbsenceDays=Under-7	0.99
12	gender=F, Success=H ==> StudentAbsenceDays=Under-7	0.97
13	ParentschoolSatisfaction=Bad, Success=L ==> StudentAbsenceDays=Above-7	0.89
14	gender=M, Relation=Father, Success=L ==> StudentAbsenceDays=Above-7	0.88
15	gender=F, ParentschoolSatisfaction=Good ==> StudentAbsenceDays=Under-7	0.83

V. CONCLUSIONS AND FUTURE WORK

In this study, we show that the success level of the students can be determined according to the demographic, academic, and behavioral information-based features of the students, and at the same time, absence level of the students can be predicted by using the same feature list. Classification methods can be used for this purpose. In this study we apply five classification methods that are MLP, RF, NB, J48, and Voting; and obtain better classification accuracy to predict the academic success of students with respect to previous study using the same dataset. We also apply the Apriori algorithm to determine the relationships between the features and academic success level of the students. As a result of our experimental study, we can conclude that "absence of students", "parents' school satisfaction", "amount of raising hands on class" and "parent who is responsible for the student" can affect the academic success level of the students.

In future work, new features that may affect the performance of the students can be included into the dataset. Also, other classification methods can be applied to improve the classification accuracy.

REFERENCES

[1] C. Anuradha, T. Velmurugan, "A Data Mining based Survey on Student Performance Evaluation System," IEEE International Conference on Computational Intelligence and Computing Research Coimbatore India, pp. 18-20, Decdember 2014.

[2] Ibrahim Aljarah, "Students' Academic Performance Dataset, xAPI-Educational Mining Dataset". [Online]. Available: https://www.kaggle.com/aljarah/xAPI-Edu-Data [Accessed May 20, 2018].

[3] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," Acm Sigmod International Conference on Management of Data, Washington DC/USA, 1993.

[4] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," Proc. 20th Int. Conf. Very Large Data Bases, Santiago, Chile, pp. 487–499, 1994.

[5] Q. Zhao and S. S. Bhowmick, "Association Rule Mining: A Survey," Technical Report, CAIS, Nanyang Technological University/ Singapore, 2003.

[6] E. A. Amrieh, T. Hamtini and I. Aljarah, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods," International Journal of Database Theory and Application, vol. 9, pp. 119-136, 2016.

[7] A. Acharya, D. Sinha, "Early Prediction of Students Performance using Machine Learning Techniques," International Journal of Computer Applications, vol. 107, pp. 37-43, 2014.

[8] S. B. Wadkar, S. C. Dharmadhikari, S. K. Dwivedi, "A Survey on Performance Evaluation and Prediction of Teachers', Students' Performance Using Data Mining Techniques," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 6, pp. 281-285, 2017.

[9] A.Daud, N.R. Aljohani, R.A. Abbasi, M.D. Lytras, F. Abbas and J.S. Alowibdi, "Predicting Student Performance using Advanced Learning Analytics," International World Wide Web Conference Committee (IW3C2), Perth/Australia, pp. 3-7, April 2017.

[10] O.K. Oyedotun, S.N. Tackie, E.O. Olaniyi, A. Khashman,"Data Mining of Students' Performance: Turkish Students as a Case Study", IJISA, vol. 7, pp. 20-27, 2015.

[11] T. A. Kumbhare, S. V. Chobe, "An Overview of Association Rule Mining Algorithms," International Journal of Computer Science and Information Technologies (IJCSIT), vol. 5, pp. 927-930, 2014.

[12] J. Han, M. Kamber, J. Pei, "Mining Frequent Patterns, Associations, and Correlations," in Data Mining Concepts and Techniques, Elsevier Inc., 2012, pp. 243-278.

[13] I. Jenhani, N. B. Amor, Z. Elouedi, "Decision trees as possibilistic classifiers ," International Journal of Approximate Reasoning, vol. 48, pp. 784-807, 2008.

[14] L. Breiman, " Random Forests ", *Machine Learning*, vol. 45, pp. 5–32, 2001.

[15] D.D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval", Machine Learning*: ECML-98*, vol. 1398, pp. 4-15, 1998.

[16] M.W. Gardnera, S.R. Dorlinga, "Artificial neural networks (the multilayer perceptron) a review of applications in the atmospheric sciences," Atmospheric Environment, vol. 32, pp. 2627-2636, 1998.

[17] F. Moreno-Seco, J.M. Iñesta, P.J.P. de León, L. Micó, "Comparison of Classifier Fusion Methods for Classification in Pattern Recognition Tasks," Structural, Syntactic, and Statistical Pattern Recognition Lecture Notes in Computer Science, vol. 4109, pp. 705–713, 2006.

[18] The WEKA Workbench. [Online]. Available: https://www.cs.waikato.ac.nz/ml/weka/

[19] D.M.W. Powers, "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, vol 2, pp. 37-63, 2011.

[20] R.E. Schapire, "The Boosting Approach to Machine Learning: An Overview," Nonlinear Estimation and Classification, vol. 171, pp. 149-171, 2003.

[21] E. A. Amrieh, T. Hamtini and I. Aljarah, "Preprocessing and analyzing educational data set using X-API for improving student's performance," Applied Electrical Engineering and Computing Technologies (AEECT) IEEE Jordan Conference, pp. 1-5, November 2015.