Clarify of the Random Forest Algorithm in an Educational Field

Nawzat Sadiq Ahmed
Directorate of General Registration
Duhok Polytechnic University
Duhok, Iraq
nawzat.ahmed@dpu.edu.krd

Mohammed Hikmat Sadiq
Computer Science Department
University of Zakho
Duhok, Iraq
hamamezory@gmail.com

***Abstract*—Many supportive decision systems using classification algorithms have been built as a black box in the last years. Such systems were hiding its inner operations to users. Lack of explanation of these algorithms leads to a practical problem. The education field is one of the areas that needs more clarification in such systems to help users in order to get more information for a right decision. In this paper, the Random Forest algorithm has been clarified and used in analyzing the students' performance, as a dataset. The result showed that the clarified method of the aforementioned algorithm can give an accuracy of 83.56%. On the other hand, WEKA tool gives an accuracy of 80.82% for the same algorithm and dataset. Also, the proposed method of the Random Forest algorithm has been tested using another previous study's dataset. The comparison results showed that the proposed method can give an accuracy of 92.65%, which is in turn better than the accuracy of 91.2% that obtained by another study done. Furthermore, to make the Random Forest algorithm work as a white box, Rules have been extracted from the Random Forest black box algorithm in order to make it more interpretable and helpful in predicting student's performance.**

***Keywords*—rules extraction, classification, random forest, black-box, white-box**

I. INTRODUCTION

The most active area in Computer Science is Data Mining (DM), which is a promising and young field [1]. Data Mining is the procedure of extracting useful information from a large database. Techniques of data mining were used to clean the large databases so as to get the most useful information that might remain unknown. Therefore, prediction of the outcome of future observations are also provided by these techniques [2].

The DM has many methods and techniques. These methods and techniques can be applied on numerous areas, such as marketing, real states, engineering, sales, trades, web mining, etc. [3]. The most DM methods used are classification, clustering and association rules. This research focuses only on the classification method.

Classification is a supervised learning method. It is one of the data mining tools which assigns items in a group to aim at categories or classes. A classification technique has two models: descriptive model and predictive model. Descriptive model is an explanatory tool used to distinguish between the objects of different classes. On the other hand, a predictive model is to predict the class of unseen records[2].

Classification aims at predicting the target class for each example in the data accurately, and also, classification is discrete and does not indicate order. A predictive model containing a numerical target utilizes a regression algorithm, rather than a classification algorithm [4]. A lot of classification algorithms had been considered by many researchers in statistics, machine learning, and pattern recognition. These algorithms split into two categories depending on its understandability. "Techniques that produce interpretable classification models are known as white-box approaches, whereas those that do not are known as black-box approaches". These techniques have many advantages such as errors could be detected in the model or in the data, prediction confidence increases for the user, and also a new solution for the classification problems is provided. Decision tree and classification rules are examples of white box algorithms [5], while Random Forest and Support Vector Machine are examples of black box algorithm [6]. A new prominent area of DM, which is called educational Data Mining, can be used on many education related data. This new prominent creates methods that extract knowledge from data of an educational environment [3].

In this study, the Random Forest classification algorithm has been used in order to extract useful information from the students' dataset. When such algorithm give a high accuracy in classification, it is not enough. This algorithm should also be more understandable for users and help them to make the right decision early. For this purpose, the Random Forest classification algorithm has clarified. To obtain predictive students' performance, the new data has been tested on all trees in the forest and the most voted label can be assigned to that record. In order to clarify the Random Forest algorithm, the trees have extracted from the Random Forest using the functions of tree creation.

II. LITERATURE REVIEW

Many of the machine learning algorithms are hard to understand. To overcome this problem, researchers, by using different techniques for rule extraction, converted these algorithms from black box to white box algorithms.

A. Rule Extraction Works

Support Vector Machine (SVM) is one of the black box algorithms in which its process is not visible to users. Thus, researchers tried to use many techniques to extract rules from SVM and make it more obvious for users.

Moghimi and Varjani [7] illustrated a new method that detected the attacks of phishing in internet banking which is called rule based method. In the aforementioned research, two novel feature sets were used to determine the webpage identity.

978-1-5386-6696-8/18/$31.00 ©2018 IEEE

The page resource was evaluated by features which contains four feature sets, and four features for identifying the access of protocols of page resource elements. The relationship was done among the URL of a page and content in first feature set. Additionally, Support Vector Machine (SVM) was employed to classify webpages. The result showed that the model could predict the phishing attack with an accuracy of 99.14% true positive and 0.86% false negative. The hidden knowledge was extracted from the SVM model and was included in bowser extension named Phish Detector in order to make the method easier to use and more functional. Park, Jo, and Kim [8] showed two different models, which are linear and non-linear. These models were presented and are based on two different types of blended learning classes. In case of blended learning course that is based on a lecture which contains main online activities such as downloading material and submitting students, the linear multiple regression was not the suitable technique for prediction. On the other hand, in case of blended learning course that is based on a discussion class containing active learner's participation in online forum, linear multiple regression could explain the achievement of the students. The results illustrated that the consideration of the divers in blended learning course should be used and the prediction models should depend on those considerations. Furthermore, in the aforementioned study by using Random Forest, they found that different important variables were indicated by the two cases in which the attributes of both cases were reflected. Depending on the results of the study, both cases should be combined in order to make a model that predicts the learning of the students.

Bart and David [6] said that in order to minimize the cost of software projects, an important operation should be used which is software fault and effort prediction. Also, different data mining algorithms were used, which considered that the predictive model not only be accurate but also should be understood by the user. Understanding many classification algorithms is very hard, in order to overcome this problem, the researchers extracted tree from the performing Support Vector Machine for regression (SVR) and Random Forest (RF). The rules were extracted from both algorithms by using ALPA (Rule Extraction Algorithm), this method builds trees by using c4.5 and REPTree which could explain the two black box algorithms as closely as possible. The result showed that the extracted rules were very obvious and could explain the prediction models very well. Han et al. [9] showed one of the world wide public challenge which is a chronic disease called Diabetes mellitus. Support Vector Machine (SVM) was used to classify the ones that have diabetes and the ones that do not, and added an ensemble learning module that converts the black box of SVM to something understandable and which has transparent rules, it also helps in solving the imbalance problems. The result illustrated that the proposed model could generate rule sets with an average precision of 94.2% and average recall with 93.9% for all classes. Additionally, this mixed model could provide a tool for the diabetes's diagnosis and also help in second opinion for lay users.

Gao et al. [10] developed a fuzzy-based SVMs multi classifier in order to tackle a 3-class classification trouble of the hot metal silicon substance from the control viewpoint. A fuzzy-based SVMs classification algorithm has been suggested for, including feature selection from extensive candidate input and the determination of controllable bounds from the real blast furnace data. After that, some practicable rules for blast furnace are drawn out using CART Algorithm from the built SMVs model to direct the silicon substance in low, proper or high range. The agreements between the values which were predicted and the real ones show that the extracted rules can work well for the blast furnace system. If the extracted rules are compared with SMVs black-box model, they can yield the label of classifying the silicon substance as well as the reason of producing it. The advantages of white-box models and black-box models have been effectively incorporated in the current blast furnace applications. The rules which were extracted can aid in making control decision on the blast furnace process. Shinde and Kulkarni [11] was not like the previous version of Fuzzy min-max neural network which could only deal with continuous attributes. A Modified Fuzzy min-max neural network classifier was proposed. Both Continuous and discrete attributes could be processed by the model, and also rules could be extracted in order to justify the decision of its classification. Hyper boxes that have minimum confidence were removed by pruning the trained network for finding the most accurate and concise rule sets. The result showed that the proposed model gives better accuracy rate and a lesser number of hyper boxes when compared with the old version. Also, the new model gives better quality rules in terms of accuracy.

B. Random Forest in Education

Hussain et al. [1] presented that data mining techniques are used in higher education institutes for improving the students' academic performance to prevent them from dropping out. In the aforementioned work, data is collected from 3 different colleges, the data consists of demographic, socio-economic as well as academic information of three hundred students. Four Classification algorithm J48, Bayes Net, PART, and random forest were used in order to predict the student performance. The result showed that Random Forest Outperform better than others according to accuracy and Classification error. Furthermore, a prior algorithm was utilized to find the association mining rules among all attributes and also best rules were shown.

Mahboob, Irfan, and Karamat [12] demonstrated that one of the biggest efforts in educational data mining is predicting student academic performance. The research could indicate that by using machine learning algorithms such as J48, Naïve Bayes, and Random Forest. It is possible to predict the students' success rate. The goal of this research was to find the factors that affect the performance of students during their academic years which makes them drop out, this research also tried to help students avoid dropping out and evaluate themselves depending on previous instances. The results showed that the model was very helpful for predicting the students' performance. Random Forest algorithm depicts that 100% instances are correctly classified and J48 shows that 93.3% instances are correctly classified while Naive Bayes presents that 86.6% instances are classified as correct.

Mishra, Kumar, and Gupta conducted in [13] that the education sector became similar to an industry, and like other industries faces many challenges. The major challenge that education sector face is that the rate of drop out students increases or they leave the course before completion. Two data mining techniques J48 (implementation of c4.5) and Random Trees were utilized to build the classification made for prediction. The model was applied to students' records of MCA students of colleges affiliated to Guru Gobind Singh Indraprastha University to predict third-semester performance. The model depended on students' academic integration, social integration, and many emotional skills. The result showed that the Random trees outperformed better than j48.

In the previous works, many techniques were used in order to extract rules from some classification algorithms that are called black boxes such as Support Vector Machine and others, also such algorithms are used in many areas. On the other hand, another study proposed black boxes such as Random Forest and other algorithms in higher education for preventing the students from dropping out and also enhance the students' performance. Furthermore, these studies do not show the black box algorithms in terms of what the rules are that the classification depends on. In the next section, how Random Forest algorithm works and how it generates rules are detailed.

III. RESEARCH METHOD

A. Random Forest Algorithm

A machine learning method which has the ability to perform both regression and classification tasks. A Random Forest classifier grows a number of decision trees which are trained on different parts of the same training set in order to improve the classification rate and to overcome the problem of over fitting [12]. Random Forest chooses the attributes randomly for creating K number of trees each time with different attributes without pruning. In decision tree, the test data will be tested on the only constructed tree, unlike Random Forest while the test data will be tested on all the constructed trees and then the most frequent output will be assigned to that instance[13]. Generally, if the forest has more trees then it will have more robust. Random Forest Classifier has the same idea, best accuracy will be given by Random Forest techniques if it has a higher number of trees in the forest. Furthermore, missing values can be handled by Random Forest, Random Forest never cares about number of trees in the forest, it never over fits, and Random Forest can also deal with categorical values. According to measuring the purity and impurity of attributes , Random Forest uses Gini index indicator.

1) Gini Index: It is a selection measure of the attribute that measures the impurity of an attribute related to classes. Gini index is similar to entropy, it gets maximum value when all classes in the dataset have equal probability and is defined as [14]:

$$\mathrm{Gini}(t) = 1 - \sum_{i=0}^{c-1}[p(i|t)]^2,$$

2) Random Forest Flow Chart: The following Figure shows the Random Forest process as a flow chart.

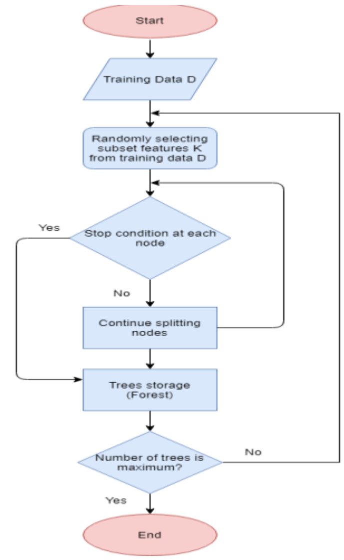

Fig. 1. Random Forest Algorithm

B. Proposed Method of Rules Extraction in Random Forest

Quinlan, J R [15] presented that the production rules could be generated from decision trees. Based on the same training data that was used to build the tree, a set of production rules could be extracted. The process of generating rules starts with rules extraction from the decision tree individually. And each rule is constructed from the root node of the tree to a leaf node. The initial form of the rule is:

If Cond1 & Cond2 & Cond3 Then class C1 were Condi is the conditions and Ci is the class labeled by leaf. Then all rules are grouped according to their class label.

In this study, Accord.Net library has been used. Accord.Net is a machine learning library in visual studio C# constructed initially in 2010 and was stabled on October 2017. Functions in visual studio library can be used to generate trees for random forest algorithms and also for extracting rules from the algorithm. Random Forest can create trees by each time choosing different subset of dataset and store it in a variable called forest. Therefore, the trees will be extracted one by one using a property called (.tree). After that the extracted tree from the forest will be stored in a decision set variable, then by using toRules function the tree will be converted to a set of rules for that tree (see Fig. 2).

C. Data Collection

Students' dataset are used in this research which contain records of 73 students of the computer science department consisting their Gender, G1 (Exam one), G2 (Exam 2), Reports, Assignment, Dormitory, Absence, Practical, OnlineRes., and participation in the class as an input attribute. Class attribute is used as output (see Table I). All attributes have a relation with each other except the Class attribute which depends on all others. From the literature reviews, the most common features used by them were extracted. However, we did not stop there, we dug deeper by doing interviews with teachers of governmental institutes. From the literature reviews and interviews, our features were selected for this research. All attributes in the research have the same importance at the beginning, but when the algorithm runs, attributes will be selected based on Gini index measurement. The attribute with the highest Gini index will be the root. It is the most important feature, and the process will continue to the end. This process shows us the most important information in the dataset. The attributes that can be seen in the tree, are the most effective factor for students' performance.

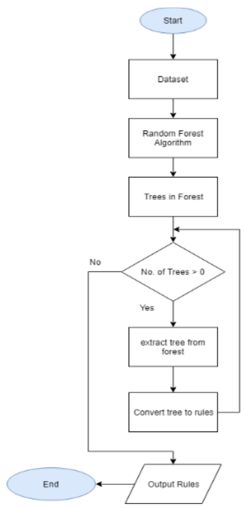

Fig. 2. Rules Extraction Process

TABLE I. STUDENTS' DATASET

Attribute Name	*Value*	*Description*
G1	Numeric	The grades of exam 1
G2	Numeric	The grades of exam 2
Gender	Male, Female	Gender of students
Reports	Numeric	Grades of report of students
Assignment	Numeric	Grades of assignment given to students
Dormitory	Yes, No	If the student live in dormitory or not
Absence	High, Low, Average	How much the student be absents
OnlinRes.	Yes, No	If the student use the online resource or not
Practical	Numeric	Lab work marks of the students
Participation	Yes, No	If the student participate during the class
Class	Pass, Fail	The goal attribute to check if the student will pass the final exam or not in a specific subject.

D. Evaluation method

Evaluation is one of the most important steps in building classification models because it helps in checking the performance of a model. There are many types of evaluation such as bootstrap, leave-one-out cross-validation, two-fold cross-validation, k-fold cross-validation, and others.

a) Cross-Validation

Tenfold cross-validation is a common method to measure the error rate of a model on an appropriate dataset. This method splits data into ten equal size parts, each time the model will train on nine parts and the one remaining part will be used for testing then the error rate will be calculated. This process is repeated ten times but each time it chooses a different fold from the remaining nine training and replaces it with the testing one part without it choosing the testing one again. Then the average error rate for all tenfold will be calculated [16].

b) Confusion Matrix

It is a table that explains the performance of the classification models. The class values are separated into Positive P and Negative N. The class that is actually positive and correctly predicted positive is named true positive (TP), while the actual class incorrectly predicted as negative which is named false negative (FN). The same goes for negative, the class that is actually negative and correctly predicted as negative is named true negative (TN), while the actual class incorrectly predicted as positive is named false positive (FP) [17]. These information are given in Table II.

In confusion matrix, Accuracy (1), Precision (2), Recall (3), and F-score (4) will be calculated.

Accuracy will measure the total correct predictions to all predictions. Precision will measure the total of all predictions marked as positive by the classifier. The recall will measure the total of positive classes that were correctly predicted as positive [17]. F-score is a harmonic mean of precision and recall [1].

TABLE II. CONFUSION MATRIX

		Predicted Class		
Actual Class	Positive	TP	FN	P
	Negative	FP	TN	N
	Total	P'	N'	P+N

$$Accuracy = (TP+TN) / (TP+TN+FN+FP) = (TP+TN) / (P+N) \quad (1)$$
$$P = TP / (TP + FP) \quad (2)$$
$$R = TP / (TP + FN) = TP / P \quad (3)$$
$$F = 2PR / (P+R) \quad (4)$$

E. Weka Tool

WEKA is defined as a workbench for machine learning and it aims to help in the use of machine learning techniques to a number of different real-world problems, especially those resulting from agricultural and horticultural areas. Unlike other machine learning projects, the focus is on offering a working environment for the specialist in the field rather than the expert in machine learning. Lessons learned to incorporate the necessity of providing a variety of interactive tools for the data manipulation, result, visualization, and database linkage. Furthermore, in order to complement the tools for basic machine learning a cross-validation and comparison rule sets used [18].

IV. EXPERIMENTAL AND RESULTS

In this study, the researchers took the students' dataset that contains records of 73 students. The dataset included attributes mentioned in Table I. The classification model Random Forest was applied on a dataset in order to analyze and predict the students' performance in the final exam. All attributes in the dataset are independent of each other except the class attribute (output) depends on all of them. Therefore, the dataset will be trained on the Random Forest model two times. The first time our proposed model will be used which is programmed by a machine learning library Accord.net using C# language in visual studio. In the second time, a Weka tool Random Forest model will be used. In the next sections, the results of both operations are shown and compared.

A. WEKA Tool Results

The dataset inserted into WEKA tool and Random Forest algorithm were applied on datasets and the result shows in Table III and also Fig. 3.

TABLE III. WEKA RESULTS

Class	*precision*	*Recall*	*F-Score*
Pass	0.826	0.864	0.844
Fail	0.778	0.724	0.750
Accuracy	80.8219 %		

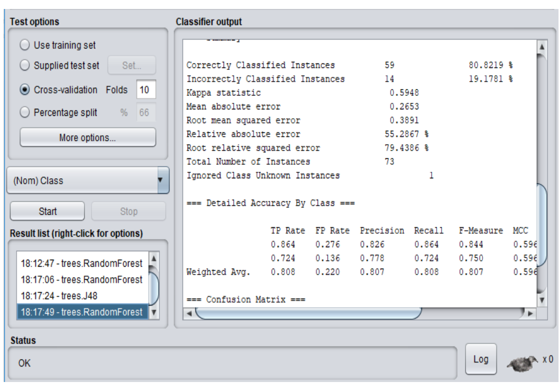

Fig. 3. WEKA Tool

B. Proposed Method Results

Random Forest algorithm applied on dataset and the result shows in Table IV and also Fig. 4.

TABLE IV. RANDOM FOREST RESULTS

Class	*precision*	*Recall*	*F-Score*
Pass	0.886	0.847	0.866
Fail	0.758	0.814	0.785
Accuracy	83.56 %		

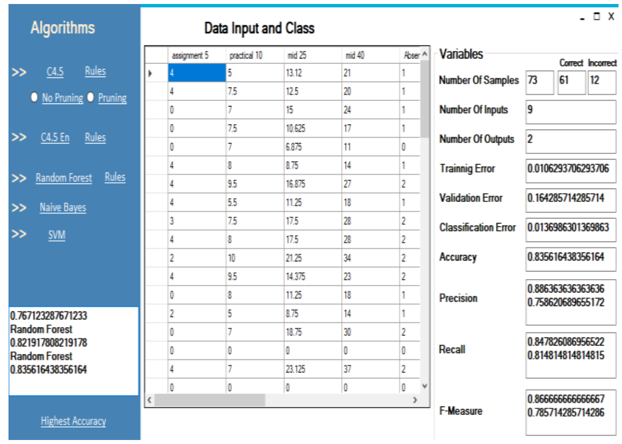

Fig. 4. Random Forest Model

V. DISCUSSION

A. Comparing the Proposed Method with WEKA Tool

Student dataset of this research, as case study, was applied on both WEKA tool and the proposed method. The results (see Table III and Table IV) showed that the accuracy in the proposed method is 83.56%. On the other hand, WEKA tool gave accuracy with 80.82%. Furthermore, the proposed model allows to run the process many times, and each time gives a different accuracy because Random Forest starts choosing a subset from dataset randomly each time and continues to the end of the process. This helps in selecting the best model.

The most important idea of this study is converting the Random Forest algorithm from the black-box algorithm to the white-box algorithm like decision tree. By showing how the algorithm classifies the dataset and how the prediction is done in this algorithm. This makes the Random Forest algorithm more interpretable and clearer to the users. Thus, the proposed model used a machine learning library tool in order to extract rules from the Random Forest trees (see Fig. 5 as an example). But, WEKA tool cannot extract rules from this algorithm.

pass =: (assignment 5 <= 1.5) && (practical 10 > 8.5)
pass =: (assignment 5 > 1.5) && (practical 10 > 4.5) && (mid 25 > 10.9375)
pass =: (assignment 5 <= 1.5) && (practical 10 <= 8.5) && (mid 25 > 10.9375) && (Part. == no)
pass =: (assignment 5 > 1.5) && (practical 10 > 4.5) && (mid 25 <= 10.9375) && (Gender == f)
fail =: (assignment 5 > 1.5) && (practical 10 <= 4.5)
fail =: (assignment 5 <= 1.5) && (practical 10 <= 8.5) && (mid 25 <= 10.9375)
fail =: (assignment 5 <= 1.5) && (practical 10 <= 8.5) && (mid 25 > 10.9375) && (Part. == yes)
fail =: (assignment 5 > 1.5) && (practical 10 > 4.5) && (mid 25 <= 10.9375) && (Gender == m)

Fig. 5. Extracted Rules(Students' dataset)

B. Comparing the Proposed Method with another study

Cortez and Silva [19] used Decision tree, Random Forest, Naïve Bayes, Neural Network, and Support Vector Machine in order to improve the quality of education. The main purpose of the study was to overcome the hardly success in the Classes of the Portuguese language and Mathematics in secondary schools by improving and developing the effective prediction model for students. The aforementioned researchers used two datasets (One for mathematic students and the other for Portuguese language). The results showed that Naïve Bayes classifier performed better than others for first case A with an accuracy of 91.9% (see Table V).

TABLE V. STUDY RESULTS [19]

Input Setup	Mathematics				
	NV	NN	SVM	DT	RF
A	**91.9**$^{\dagger}_{\pm 0.0}$	88.3$_{\pm 0.7}$	86.3$_{\pm 0.6}$	90.7$_{\pm 0.3}$	91.2$_{\pm 0.2}$
B	**83.8**$^{\dagger}_{\pm 0.0}$	81.3$_{\pm 0.5}$	80.5$_{\pm 0.5}$	83.1$_{\pm 0.5}$	83.0$_{\pm 0.4}$
C	67.1$_{\pm 0.0}$	66.3$_{\pm 1.0}$	**70.6**$^{\bullet}_{\pm 0.4}$	65.3$_{\pm 0.8}$	70.5$_{\pm 0.5}$

Moreover, the mathematics students' dataset was downloaded from UCI website. This dataset has been used on the proposed model to check the performance of the model, the results showed that the model can perform better with an accuracy of 92.65% (see Fig. 6). And also, all possible rules were extracted from this dataset in order to explain what the main effective factors on students' performance are (see Fig. 7 as an example).

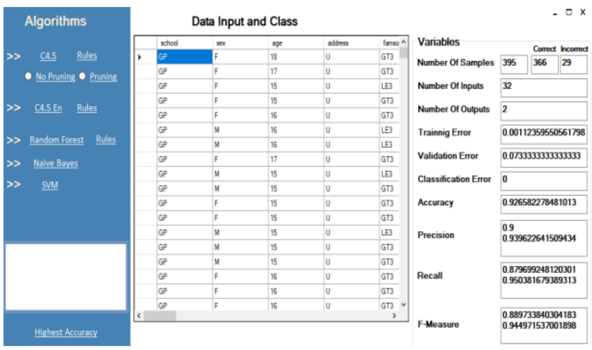

Fig. 6. Proposed Model Results

FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == course)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == other) && (paid == no)
FAIL =: (G2 > 9.5) && (G1 <= 10.5) && (nursery == yes) && (absences > 10.5)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == reputation) && (freetime <= 3.5)
FAIL =: (G2 <= 9.5) && (traveltime > 2.5) && (sex == M) && (Pstatus == T) && (Mjob == other)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == reputation) && (freetime > 3.5) && (Walc <= 1.5)
FAIL =: (G2 <= 9.5) && (traveltime > 2.5) && (sex == M) && (Pstatus == T) && (Mjob == at_home)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == other) && (paid == yes) && (sex == M)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == home) && (address == U) && (Fedu > 1.5)
FAIL =: (G2 <= 9.5) && (traveltime <= 2.5) && (reason == home) && (address == U) && (Fedu <= 1.5) && (famsize == LE3)
FAIL =: (G2 > 9.5) && (G1 <= 10.5) && (nursery == yes) && (absences <= 10.5) && (famrel > 3.5) && (failures > 0.5)
FAIL =: (G2 > 9.5) && (G1 <= 10.5) && (nursery == yes) && (absences <= 10.5) && (famrel > 3.5) && (failures <= 0.5) && (health > 4.5) && (Fedu <= 3.5)
FAIL =: (G2 > 9.5) && (G1 <= 10.5) && (nursery == yes) && (absences <= 10.5) && (famrel > 3.5) && (failures <= 0.5) && (health <= 4.5) && (Walc <= 1.5) && (famsize == LE3)
FAIL =: (G2 > 9.5) && (G1 <= 10.5) && (nursery == yes) && (absences <= 10.5) && (famrel > 3.5) && (failures <= 0.5) && (health <= 4.5) && (Walc <= 1.5) && (famsize == GT3) && (age
...

Fig. 7. Rule Extraction from Dataset of the Study [19]

VI. CONCLUSION

Classification algorithms split into two categories, which depend on its understandability. These categories are black-box and white-box approaches. The Random Forest algorithm is an example of black box algorithm, which is the bedrock of this research. Random Forest classification algorithm has been used in order to extract useful information from the students' dataset. Rules extraction model of Random Forest Algorithm were proposed in order to be understandable for users and helpful to make the right decision based on students' dataset. After experimental testing and evaluation of the proposed model, it has been found that the accuracy of predicting students' performance is higher compared to the WEKA tool and other models proposed in the literature of this research. This reaches that the proposed method of clarifying the Random Forest algorithm can be selected as a white-box algorithm and provided more accuracy in the classification process for any dataset.

REFERENCES

[1] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018.

[2] Pang-Ning, "Introduction to Data Mining," 2005.

[3] K. Parmar, D. Vaghela, and P. Sharma, "Performance prediction of students using distributed Data mining," *2015 Int. Conf. Innov. Information, Embed. Commun. Syst.*, no. Ddm, pp. 1–5, 2015.

[4] N. V. K. Rao, N. Mangathayaru, and M. S. Rao, "Evolution and Prediction of Radical MultiDimensional E-Learning System with Cluster based Data Mining Techniques," *Int. Conf. Trends Electron. Informatics*, pp. 701–707, 2017.

[5] R. C. Barros and A. A. Freitas, *Automatic Design of Decision-Tree Induction Algorithms*. 2015.

[6] D. Bart and B. David, "Comprehensible Software Fault and Effort Prediction: a Data Mining Approach," 2014.

[7] M. Moghimi and A. Y. Varjani, "New rule-based phishing detection method," *Expert Syst. Appl.*, vol. 53, pp. 231–242, 2016.

[8] Y. Park, I. Jo, J. H. Kim, and I. Jo, "Predicting Students â€™ Learning Performance by Using Online Behavior Patterns in Blended Learning Environments : .... Predicting Students â€™ Learning Performance by Using Online Behavior Patterns in Blended Learning Environments : Comparison of Two Cases on," no. October, 2014.

[9] L. Han, B. Ins, S. Luo, J. Yu, and L. Pan, "Rule Extraction from Support Vector Machines Using Ensemble Learning Approach : An Application for Diagnosis of Diabetes," vol. 2194, no. 5, 2014.

[10] C. Gao, S. Member, Q. Ge, and L. Jian, "Rules Extraction from Fuzzy-based Blast Furnace," no. c, pp. 1–12, 2013.

[11] S. Shinde and U. Kulkarni, "Extracting classification rules from modified fuzzy min–max neural network for data with mixed attributes," *Appl. Soft Comput. J.*, pp. 1–15, 2015.

[12] T. Mahboob, S. Irfan, and A. Karamat, "A machine learning approach for Student Assessment in E-Learning Using Quinlan ' s C4 . 5 , Naïve Bayes and Random Forest Algorithms," 2016.

[13] T. Mishra, D. Kumar, and S. Gupta, "Mining students' data for prediction performance," *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, pp. 255–262, 2014.

[14] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.

[15] J. R. Quinlan, "J. Ross Quinlan_C4.5_ Programs for Machine Learning.pdf," *Morgan Kaufmann*, vol. 5, no. 3. p. 302, 1993.

[16] M. Kaufmann, *Practical machine learning tools and techniques*, Second Edi. Elsevier, 2005.

[17] M. Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," *IEEE Access*, vol. 4, no. c, pp. 2379–2387, 2016.

[18] G. Holmes, A. Donkin, and I. H. Witten, "WEKA: a machine learning workbench," *Proc. ANZIIS '94 - Aust. New Zealnd Intell. Inf. Syst. Conf.*, pp. 357–361, 1994.

[19] P. Cortez and A. Silva, "Using Data Mining To Predict Secondary School Student Performance," *5th Annu. Futur. Bus. Technol. Conf.*, vol. 2003, no. 2000, pp. 5–12, 2008.