IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 8, No. 4, December 2019, pp. 422~428
ISSN: 2252-8938, DOI: 10.11591/ijai.v8.i4.pp422-428

Hybridisation of RF(Xgb) to improve the tree-based algorithms in learning style prediction

Haziqah Shamsudin, Maziani Sabudin, Umi Kalsom Yusof
School of Computer Science, 11800 Universiti Sains Malaysia, Pulau Pinang, Malaysia

Article Info

Article history:

Received Aug 21, 2019
Revised Nov 10, 2019
Accepted Nov 23, 2019

Keywords:

Hybrid
Hyperparameter Optimization
Learning Style
Online Learning
Random Forest

ABSTRACT

This paper presents hybridization of Random Forest (RF) and Extreme Gradient Boosting (Xgb), named RF(Xgb) to improve the tree-based algorithms in learning style prediction. Learning style of specific users in an online learning system is determined based on their interaction and behavior towards the system. The most common online learning theory used in determining the learning style is the Felder-Silverman's Learning Style Model (FSLSM). Many researchers have proposed machine learning algorithms to establish learning style by using the log file attributes. This helps in determining the learning style automatically. However, current researches still perform poorly, where the range of accuracy is between 58%-89%. Hence, RF(Xgb) is proposed to help in improving the learning style prediction. This hybrid algorithm was further enhanced by optimizing its parameters. From the experiments, RF(Xgb) was proven to be more effective, with accuracy of 96% compared to J48 and LSID-ANN algorithm from previous literature.

Corresponding Author:

Maziani Sabudin,
School of Computer Science,
11800 Universiti Sains Malaysia,
Pulau Pinang.
Email: maziani@usm.my

1. INTRODUCTION

Learning style is known as way of learning or preference by the learner on how materials are presented, how to work with it and how to internalize information [1-2]. Identifying a student's learning style has several benefits, such as making students aware of their strength and weaknesses when it comes to learning. It is also meant to be used in determining the learning preferences of each student either in a traditional classroom or through an online learning based system [3]. An online learning based system can be defined as an online system where there is an interaction between students and system [4]. Initially, in an online learning based system, the learning style of the user is determined by using available learning style questionnaires based on selected learning style model and the most commonly used learning style model is the FSLSM which also incorporates different elements from different learning style models such as [5]. However, when students are asked to fill in the questionnaire, they take longer time to fill it as the questions are long and they refuse spending too much time on the questionnaire which causes them to put in random answers [6]. Therefore, researchers came out with an alternative where they determine the learning style automatically [5]. This is done by collecting log files of the interactive behavior of the user with the system. The content of the log files consists of several related attributes matched to the system such as the number of visits, characteristics and types of objects chosen, sequences of actions and selected search terms, number of visits, time spent and performance. It also includes the activities tracked such as the searching, enroll in exam, quiz, self-assessment test, using forum, sending email and discussion board including reading or downloading of materials from the system [5, 7]. These attributes were then matched with the learning style model. Then, the result is further analyzed using machine learning algorithms until the learning style of the user is determined.

Researchers have applied some widely used techniques such as Artificial Neural Network (ANN), Naïve Bayes (NB) and Decision Tree (DT), and fuzzy logic in predicting student academic performance [8]. Tree based learning algorithms are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map non-linear relationships quite well. Methods like RF and Xgb are used in all kinds of data science problems [9]. From previous research, there are two papers in learning style prediction that use decision tree algorithms [10-11]. Both of these papers manage to increase the percentage of accuracy in learning style prediction compared to previous papers. However, there are still a gap within the usage of the stated algorithm in terms of the accuracy of the result obtained. One of the approaches to enhance the performance of the algorithms, is by performing hyperparameter optimization in the selected algorithms. Hyperparameter optimization is the process of choosing a set of optimal hyperparameters for a learning algorithm. Identifying a good value for hyperparameters, λ where λ = parameter, is called hyperparameter optimization [12]. The critical step in hyperparameter optimization is to choose the set of trials λ^1….λ^s.

Machine learning systems are abounding with hyperparameters. Hyperparameter optimization is the minimization of parameter over a subset of parameter. This function is sometimes called the response surface in the experiment design literature. Different datasets, tasks, and learning algorithm families give rise to different sets of parameters and functions [13]. Choosing the best hyperparameters are both crucial and frustratingly difficult. Hyperparameters are chosen to optimize the validation loss after complete training of the model parameters [14]. The critical step in hyperparameter optimization is to choose the set of trials λ1…λs. The most commonly used technique in hyperparameter optimization is a grid search technique. Grid search requires choosing a set of values for each variable. It is simple to implement and parallelization is trivial. Other than that, it is also is reliable in low dimensional spaces [12]. The other crucial step to further improve the performance of the algorithms is by doing a hybrid. Numerous methods have been suggested for the creation of hybrid of classifiers [15]. Although many methods of hybrid have been proposed, yet there is no clear picture of which method is the best [16]. Thus, an active area of research in supervised learning is the study of methods for the construction of good hybrid algorithms. Hybrid algorithms is obtained by combining a portion of elements from existing elements and composing a meaningful combination. This results in strengthening the techniques combined to provide a stable and accurate results. Selecting the relevant algorithms produced efficient combinations. Many researchers have actively worked on combining multiple algorithms together for mining [17-18]. Although there are many methods proposed for hybrid algorithms, yet there is no clear picture of which method is the best [19].

In this paper, Xgb was chosen to be incorporated in the RF algorithms. Xgb is known to have an ability to help a weak learner grows into a strong learner. The advantage of using Xgb method, is that it improves the trees by increasing the weight of one tree after another [20]. One important hyperparameter in Xgb is the learning rate. Commonly, in Xgb, the lower the learning rate means it is better for testing error, but this will result in increasing more trees. With that, the hybrid between RF and Xgb may result in better performance of accuracy. The organization of the paper is as follows. Section 2 presents the methodology of the hybrid algorithms proposed in this paper. In Section 3, the results of the hybrid algorithm are evaluated and compared with other results reported in the literature. Finally, Section 4 concludes the paper.

2. THE HYBRID OF OPTIMIZED RANDOM FOREST AND EXTREME GRADIENT BOOSTING RF(Xgb)

2.1. Data selection

The datasets used in this research is taken from a research done by [11]. The data is collected from the year 2012 to 2016. It contains a record of 507 students enrolled in the Computer Technology courses which have successfully completed the Computer Programming 1 subject. This dataset consists of 15 different attributes. As mentioned by [11] the attributes selected is based on relevancy and the suitability designed as referred from previous research by [21-22]. Table 1 shows the summary of the dataset.

Table 1. Summary of dataset.

Parameter	Value
Source of Dataset	Computer Technology courses from University of Phillipines
Number of instances	507
Number of attributes	15

2.2. Performance metrices

The performance measures which are considered in this paper is the effectiveness of the proposed algorithm. It is measured by the percentage of accuracy using confusion matrix. Confusion matrix contains information about actual and predicted classifications done by a classification system [23]. In confusion matrix, the evaluation is based on some standards and terms, such as true positive (TP), False negative (FN), false positive (FP), and true negative (TN). From the terms, some equation can be deduced such as the equation for accuracy value. Accuracy is needed to determine how often the classifier is correct. The equation is shown in (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2.3. Incorporating the extreme gradient boosting function in the random forest

Boosting is based on weak learners (high bias, low variance). In terms of decision trees, weak learners are shallow trees, sometimes even as small as decision stumps (trees with two leaves). The boosting continue to update the weights of training set based on previous weaker learner to improve the importance of data which are classified wrongly. The illustration diagram of the parameter is shown in Figure 1. In this paper, the function of nrounds, eta, α, and λ is incorporated in the RF algorithms to form the hybrid algorithm, RF(Xgb). This function is selected, as it has the ability to control the number of iterations needed in build the tree in order to get an optimal tree-based model which results in a better prediction accuracy. The role of this function taken in the Xgb is to improve the majority vote value in RF and at the same time help in improving the formation of individual tree in the process of bagging methods in RF. RF(Xgb) helps in improving the RF model in reducing the OOB error value which eventually increased the accuracy value in the model. This is because RF tends to overfit its model, as it has the problem in deciding the most optimal number of tree. The overall flow of the proposed RF(Xgb) is discussed in Section 2.4.

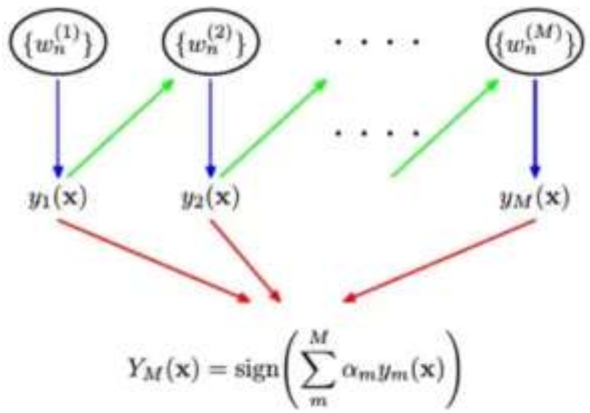


Figure 1. Illustration diagram for parameter in Xgb

2.4. The overall flow of the proposed RF (Xgb)

Figure 2 shows the working flow diagram of RF(Xgb) algorithm. First, the dataset is specified. The RF(Xgb) is adjusted accordingly by using hyperparameter optimization to obtain the most optimal parameter. Next, from the optimal parameter, RF(Xgb) is used to detect the user learning style, based on the FSLSM model and evaluated using different performance measure which are accuracy and ROC curve. In this paper, training dataset D is used to specify the supporting parameter of model t as shown in Algorithm 1. Given a training dataset, D=x_1, x_2….x_n, each training instance is represented as x_i=x_i1, x_i2…x_in and D contains the following attributes k1, k2…kn. First, the tree is specified with 10-fold cross validation and the ntree bootstrap samples is draw. For each bootstrap samples, unpruned tree is grew by choosing the best split based of random samples of mtry prediction at each node. Then, value of t is specified and mtry is optimized to reduce OOB error. The optimized value of Xgb; nround, c, eta, λ, and α is determined and insert in model t along with the specified parameter for RF that was determined earlier. The optimized parameter values is shown in Table 2. Model t is then applied to a test set D_i which contains a subset of training dataset. The algorithm is shown in Algorithm 1.

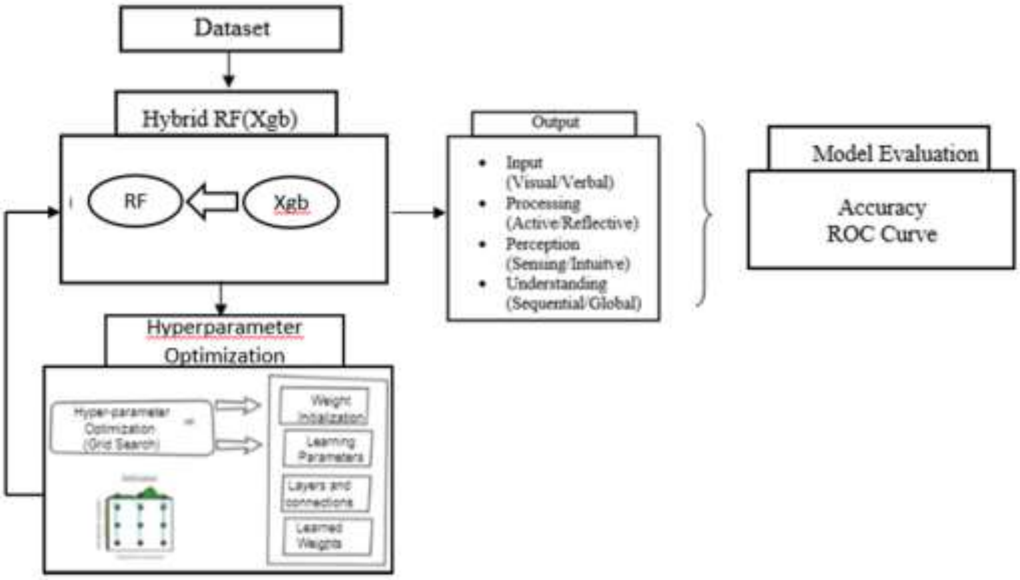


Figure 2. The working flow diagram of RF(Xgb)

Algorithm 1 RF(Xgb)

```
1: procedure RF(Xgb) (Ntree,mtry,nrounds,c,eta,α,λ)
2:     for each class, C_i ∈ D, do
3:        Specify the trControl with 5 fold of cross-validation and grid search
4:     end for
5:     for RF functions do
6:        Draw ntree bootstrap samples
7:        For each bootstrap sample, grow un-pruned tree by choosing best split based of random sample of mtry prediction at each node
8:        t:optimize mtry to reduce OOB error
9:     end for
10:    for t, use (D), do
11:       specify the supporting parameter of model t and ntree
12:       Determine best mtry value
13:    end for
14:    for Xgb Functions do
15:       t:optimize nrounds, c, eta, λ, α to reduce OOB
16:       Create new model t1 combine the function from Step 5 and Step 6
17:    end for
18:    for Function t1 do
19:       Predict the result using testing data based on final model t1
20:    end for
21:    Print the result
22:    return
23: end procedure
```

Table 2. Optimized parameter value for RF(Xgb)

Parameter	Value
ntree	300
mtry	2
nround	500
eta	0.001
λ	1
α	0

3. COMPARATIVE ANALYSIS ON THE PERFORMANCE OF ALGORITHMS

3.1. Accuracy value

This section discusses the overall result on the use of RF(Xgb) algorithm in predicting the learning style of the user. From Table 3, it shows that the result of accuracy is consistent in the range of 0.91% to 0.98%. It is noticed that when doing the hybrid the range of accuracy for all of the learning style dimension is in a constant range. This breaks the gap from previous research in the area of predicting the learning style using an automated approach done by several researchers. In the previous researches, some researchers had problem of not getting a good accuracy for certain dimension while few researches exclude some dimensions as they are not compatible with their models. However, by using RF(Xgb), a better accuracy was obtained in detecting the learning style of the user.

Table 3. Percentage of accuracy for RF(Xgb)

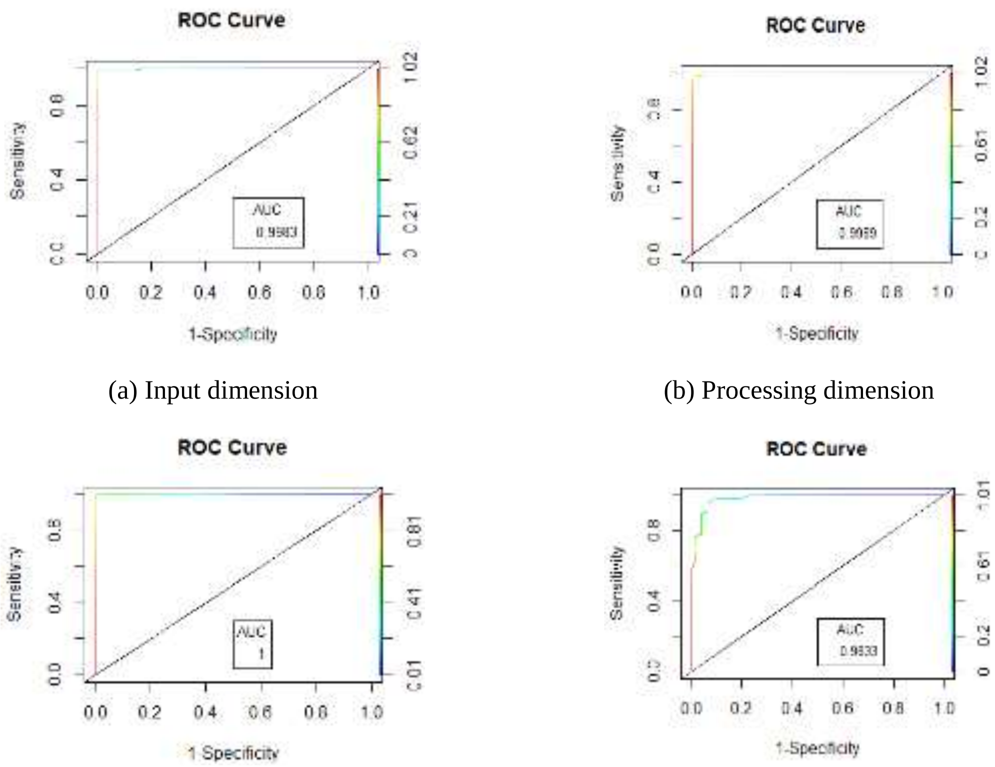
FSLSM Dimension	RF(Xgb)
Input	0.97
Perception	0.97
Processing	0.98
Understanding	0.91

3.2. ROC and AUC value

In order to evaluate further the effectiveness of the model, ROC curve is included in this paper. The concept of an ROC curve is based on the notion of a "separator" (or decision) variable. The plot of TPF (sensitivity) versus FPF (1-specificity) across varying cut-offs generates a curve in the unit square called an ROC curve. ROC curve corresponding to progressively greater discriminant are located progressively closer to the upper left-hand corner in "ROC space". The ROC curve lie on the diagonal line reflects the performance of the prediction test that is no better than chance level, i.e. a test which yields the positive or negative results unrelated to the true class label. The slope of an ROC curve at any point is equal to the ratio of the two density functions describing, respectively, the distribution of the separator variable in the class label. The area under the curve (AUC) summarizes the entire location of the ROC curve rather than depending on a specificity operating point.

The AUC is an effective and combined measure of sensitivity and specificity that describes the inherent validity of determining the class label. If two tests are to be compared, it is desirable to compare the entire ROC curve rather than at a particular point. The maximum AUC = 1 means that the model is perfect in the differentiation between the class. This happens when the distribution of the class label do not overlap. AUC = 0.5 means the chance discrimination that curve located on diagonal line in ROC space. The minimum AUC should be considered a chance level i.e. AUC = 0.5 while AUC = 0 means test incorrectly classify all subjects with class A to class B and class B to class A. Overall, the ROC curve for each of the dimension is shown in Figure 3. From Figure 3(a), the AUC value is 0.9983. The value which shows the prediction is almost perfect as mention in before that the maximum AUC=1, which means that the diagnostic test is perfect in the differentiation between visual and verbal class. The percentage of accuracy for input dimension is 0.97% which is high enough in terms of the classification accuracy.

As for the processing dimension, the percentage of accuracy is 0.98% which contribute in the high value of AUC which is 0.9989. The curve of the processing dimension is almost perfect where the curve located progressively closer to the upper-left hand corner in ROC space. The curve is shown in Figure 3(b). For perception dimension as shown in Figure 3(c), the percentage of accuracy is 0.98% while the AUC value is 1.0. As mention in the previous paragraph, when AUC = 1, it means that the model is perfect in the differentiation between the class where in this dimension it means for the class sensing and intuitive learners. Lastly, the understanding dimension produce a slightly low percentage of accuracy compared to other dimension, but it still manage to produce a better result compared to previous literature [11, 24-25]. The AUC value is 0.9833 as shown in Figure 3(d) where in this case is still in the category of almost perfect value. The ROC curve for understanding dimension is also nearly perfect as the curve located progressively closer to the upper left-hand corner in the "ROC space".



(a) Input dimension (b) Processing dimension

(c) Perception dimension (d) Understanding dimension

Figure 3. ROC value for FSLSM of RF (Xgb)

3.3. Results comparison

Table 4 shows the result comparison of RF, Xgb and RF(Xgb). From the table, RF(Xgb) shows a promising result and consistent for each of the dimension. The consistency is not only based on the percentage of accuracy, but also it can be proved based from the ROC curve as shown in Figure 3. The results shown that proposed RF(Xgb) has higher percentage of accuracy for every dimensions. RF(Xgb) is then compared with previous literature in terms of its average accuracy value and the result is shown in Table 5. From the average accuracy value, RF(Xgb) was slightly improved when compared with previous literature with a range of improvement from 0.03 to 0.1. Higher percentage of accuracy increased the ability to predict a more accurate learning styles. The results achieved is inline with [24], which the higher accuracy in detecting learning style, the higher the adaptiveness of the online learning system which lead to a better enhancement of the online learning system that can suits the user needs.

Table 4. Comparison of accuracy value for RF, Xgb, and RF(Xgb)

FSLSM Dimensiom	RF	Xgb	RF(Xgb)
Input	0.94	0.91	0.97
Perception	0.95	0.83	0.97
Processing	0.95	0.92	0.98
Understanding	0.86	0.79	0.91
Average	0.93	0.86	0.96

Table 5. Comparison of average accuracy of RF(Xgb) with literature

Method	Average Accuracy
RF(Xgb)	0.96
RF	0.93
Xgb	0.86
LSID-ANN [24]	0.81
J48 [11]	0.89

4. CONCLUSION

In conclusion, this paper presents the hybrid of RF and Xgb algorithms with hyperparameter optimization named RF(Xgb). RF(Xgb) gives a promising results in terms of improving the percentage of accuracy in learning style detection. To evaluate the effectiveness of the algorithms, few performance measures were taken into consideration which are accuracy and ROC value. Based on the comparison, RF(Xgb) shows a better accuracy value in the learning style detection. The increasing value of accuracy helps in improving the learning style detection which leads to a better adaptivity of the online learning system.

ACKNOWLEDGEMENTS

The authors wish to thank Universiti Sains Malaysia (USM) for the support it has extended in the completion of the present research through Incentive Grant (304/PKOMP/6316381). The authors also wish to thank Asst. Prof. Dr. Renato Racelis Maaliw III from Southern Luzon State University, Philippines for dataset sharing.

REFERENCES

[1] Felder, R. M., and Silverman, L. K, "Learning and teaching styles in engineering education". *Engineering education*, vol. 78, no. 7, pp. 674-681, 1988.
[2] Litzinger, Thomas A, Sang Ha Lee, and John C Wise, "A study of the reliability and validity of the felder-soloman index of learning styles*." In American Society for Engineering Education*, vol. 1, no.1, 2005.
[3] Romanelli, F., Bird, E., and Ryan, M, "Learning styles: a review of theory, application, and best practices". *American Journal Of Pharmaceutical Education*, vol. 73, pp. 1-5, 2009.
[4] Su, B., Bonk, C. J., Magjuka, R. J., Liu, X., and Lee, S.-h, "The importance of interaction in web-based education: A program level case study of online MBA courses". *Journal of Interactive Online Learning*, vol. 4, pp. 1-19, 2005.
[5] Truong, H. M., "Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities". *Computers in Human Behavior*, vol. 55, pp. 1185-1193, 2016.
[6] Mokhtar, R., Zin, N. A. M., and Abdullah, S. N. H. S, "Rule-based knowledge representation for modality learning style in aiwbes". *In Knowledge management international conference* vol. 1, pp. 614-617, 2010.
[7] Ciloglugil, B, "Adaptivity based on felder-silverman learning styles model in e-learning systems". Vol. 1, pp. 1523-1532, 2016.
[8] Anil et. al., "Academic Performance Prediction Algorithm based on Fuzzy Data Mining". *International Journal of Artificial Intelligence*, vol. 8, no. 1, pp. 26-32, 2019.
[9] Norsyela et. al, "Performance Analysis of Supervised Learning Models for Product Title Classification". *International Journal of Artificial Intelligence*, vol. 8, no. 3, pp. 299-306, 2019.
[10] Özpolat, E., and Akar, G. B, "Automatic detection of learning styles for an e-learning system". *Computers & Education*, vol. 53, pp. 355-367, 2009.
[11] Maaliw III, R. R, "Classification of learning styles in virtual learning environment using data mining: A basis for adaptive course design". *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 7, pp. 56-61, 2016.
[12] Bergstra, J., and Bengio, Y, "Random search for hyper-parameter optimization". *Journal of Machine Learning Research,* vol. 13, pp. 281-305, 2012.
[13] Bergstra, J., Yamins, D., and Cox, D. D, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures". *JMLR: W&CP*, vol. 28, no. 1, pp. 1-9, 2013.
[14] Maclaurin, D., Duvenaud, D., and Adams, R, "Gradient-based hyperparameter optimization through reversible learning". *In International conference on machine learning*, vol. 1, no. 1, pp. 2113-2122, 2015.
[15] Dietterich, T. G, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization". *Machine learning*, vol. 40, pp. 139-157, 2000.
[16] Vilalta, R., and Drissi, Y, "A perspective view and survey of meta-learning". *Artificial Intelligence Review*, vol. 18, pp. 77-95, 2002.
[17] Ahlawat, A., and Suri, B, "Improving classification in data mining using hybrid algorithm". *In Information processing (IICIP), 2016 1st India international conference*, vol. 1, no. 1, pp. 1-4, 2016.
[18] Hung, Yu Hsin, Ray I Chang, and Chun Fu Lin, "Hybrid learning style identification and developing adaptive problem-solving learning activities." *Computers in Human Behavior*, vol. 1, no. 55, pp. 552–561, 2016.
[19] Vilalta, Ricardo and Youssef Drissi , "A perspective view and survey of meta-learning." *Artificial Intelligence Review*, vol. 18, no. 1, pp. 77–95, 2002.
[20] Chen, T., and Guestrin, C, "Xgboost: A scalable tree boosting system". *In Proceedings of the 22nd ACM sigkdd International Conference on Knowledge Discovery and Data Mining*, vol. 1, no. 1, pp. 785-794, 2016.
[21] Cha, Hyun Jin, et. al, "Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system." *In International Conference on Intelligent Tutoring Systems*, vol. 4053, no. 1, pp. 513–524, 2006.
[22] Graf, S., Viola, S. R., Leo, T., and Kinshuk, "In-depth analysis of the felder-silverman learning style dimensions". *Journal of Research on Technology in Education*, vol. 40, pp. 79-93, 2007.
[23] Torgo, Luis, Data Mining with R, Learning with Case Studies. *New York: Chapman and Hall/CRC*, vol. 2, 2nd edition, 2017.
[24] Bernard, J., Chang, T.-W., Popescu, E., and Graf, S., "Learning style identifier: Improving the precision of learning style identification through computational intelligence algorithms". *Expert Systems with Applications*, vol. 75, no. 1, pp. 94-108, 2017.
[25] Graf, L. T.-C., Sabine, et al., "Supporting teachers in identifying students' learning styles in learning management systems: An automatic student modelling approach". *Journal of Educational Technology & Society*, vol. 12, no. 4, pp. 3-14, 2009.