Journal of Theoretical and Applied Information Technology
15th May 2022. Vol.100. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ENSEMBLE LEARNING FOR MINING EDUCATIONAL DATA

ASMAE BENTALEB[1] , JAAFAR ABOUCHABAKA[1]
[1]Department of Computer Science, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
E-mail: [1]asmae.bentaleb@uit.ac.ma

ABSTRACT

Predicting students' academic performance allows giving feedback that helps in making the right decisions. Previous studies proposed using traditional statistical methods; however, they are tedious and not practical especially with the large volume of data. As for the solution, this article proposes a classification model to classify students according to their performance. It combines some of the well-known classification machine learning algorithms with Xgboost. The combination is done thanks to the voting classifier. This model was applied on two datasets and gave good results for the two: 85% and 92%.

Keywords: *Xgboost, Model,Voting Classifier,E-learning,Student Academic Performance, Recommendation*

1. INTRODUCTION

Data mining tools have proven their utility in several domains. They are key factors in the decision making process. E-learning is one of these fields where analyzing and predicting students' performance data have become very crucial [1]. However, in some educational institutions, it is a very challenging task due to the huge volume of data. Machine learning mechanisms allow analyzing this data through building patterns from past data. These models help in achieving performance prediction and classification for students. For the educational field, such tools help in detecting issues in the students' learning path at early stages. Therefore, making the necessary decisions and modifications to overcome any encountered problems and improve the learning outcomes. This can be done by enhancing the teaching approaches and methodologies and making them suitable to students according to their background.

The main contribution of this article is proposing an architecture that models the classification of the students' academic performance dataset. It is an educational dataset that was collected from the University of Jordan. To confirm the results, the proposed approach will be tested on a second dataset which was collected from the Portuguese schools using questionnaires and reports. The article is organized as follows: it starts by citing the works that were already done in the students' performance prediction using variety of approaches. Then, it explains the proposed approach including the architecture and the algorithms that will be used in the experiment. Later, it describes the experiments' setup and finally presents the results and discussion.

Different works and propositions were published to enhance the quality of classifications and predictions of data in the e-learning field. Variety of machine learning algorithms and approaches were used to achieve the best performance. Here are some of the propositions that have shown interesting results and findings: This article [2] has tackled the issue of classifying students according to their performance using some internal features for assessment. The classification in this work was based on Artificial Neural Network which was used widely for this purpose. As to judge the best machine learning algorithm for learning students' outcome, it is required to perform a comparative study including the main ones and apply it to them. In this work [3], the authors have performed an analysis of the students' performance classification and prediction based on some of the well-known machine learning algorithms. They have worked mainly on two datasets: Student

Performance Dataset (SPD) and Students Academic Performance Dataset (SAPD). These datasets were analyzed by considering the impact of social factors on students' outcome. This was mainly to enhance the education's quality for the next generations through improving the impacting factors. The use of machine learning algorithms differs according to the task whether it is a prediction or a classification task. For prediction, three algorithms were used: Backpropagation (BP), Support Vector Regression (SVR) and Long-Short Term Memory (LSTM) wheras for classification, BP and SVR were used in addiction to Gradient Boosting Classifier (GBC). Significant results were achieved, around 80% accuracy.

The authors of this article [4] performed the prediction of students' performance based on small sized dataset. The task of predicting students' performance has become very crucial to educational institutions. Before performing the prediction, the key features were identified using clustering and visualization mechanisms. These features were input to machine learning algorithms to perform prediction and classification tasks. Experiments have shown that learning discriminant analysis algorithms and support vector machine have performed well in training small dataset. They have given good accuracy for classification tasks.

This paper [5] introduced a model for predicting and classifying students' performance by employing artificial neural network. This approach included also the conventional statistical analysis to detect and identify the features that most impact the students' performance. For the configuration of the neural net, it has 11 input variables, two layers of neurons that are hidden and a unique output layer. To measure the neural network's performance, several metrics were used including: error rate, confusion matrix, regression and accuracy that achieved a good percentage of 84.8%.

In this work [6], the authors suggested a model based on neural networks for performing a sentiment analysis concerning the attrition of students with regards to MOOCS. The evaluation of the proposed method was done using several metrics. Notably, it resulted in an accuracy of 72.1%. The authors of this work [7] performed the analysis of the learner's dataset to predict whether a given student will quit a course or not. This was achieved using machine learning techniques. To achieve the task of classification, they chose the logistic regression algorithm. For the purpose of validation, the model was tested on a sample of 100 students. As a remedy to the results of this study, authors have prepared in parallel an action for tutoring the students who were at risk of dropping a course. Such methodology helped in reducing the dropout rate of courses by 14 % which is considered to be very good compared to previous years.

The proposition of this article [8] aims at building prediction models for dropout. The main objective is to personalize adequate interventions for MOOCs users who are academically at risk. Such models were built using deep learning algorithms in order to output dropout probabilities for students. These reports are produced on a weekly basis. Thanks to the deep learning algorithms, the proposed approach improved the accuracy of dropout prediction models. Also, using this method, the interventions are planned, personalized and prioritized based on the probabilities output by the model and according to each student's case. This article [9] proposes a platform that develops models for student intelligent educational systems. This methodology is based on machine learning approaches, and is applied to intelligent navigation tutoring systems. It involves the process of modelling that includes mainly the processing of data, and deployment of model. Models are built based on historical data and trained in experiments with large scaled data.

From the literature works presented earlier, the following gap can be identified: the performance of the previous works needs to be further enhanced and developed. It did not exceed 80% for the e-learning datasets. The following section will present the suggested approach aiming to the enhancement of the performance metrics of the e-learning classification task. The research objectives of this work can be summarized in the following points:

- An ensemble of machine learning algorithms including a novel algorithm which is the XGboost one will be used in classifying the students' learning datasets.
- Experiments have been conducted on two different datasets SPD and SAPD.
- Accuracy, precision, recall and F1-score have been taken as the evaluation criteria for testing the robustness of the proposed methodology.
- A comparison with existing techniques of the proposed methodology reveals superior results with the same number of defined parameters.

- Empirical evaluation of the proposed methodology has been conducted with conventional base classifiers like Support Vector Machine, Random forest, and Naïve Bayes.

2. THE PROPOSED METHOD

Previously, there were many works of predicting and classifying learner's performance using one or several algorithms of machine learning. However, the problem of these works is the accuracy of the classification tasks. It needs to be improved in order to enhance the quality of classifications.

To improve the performance of such models, the use of combinatorial method is proposed. This is because in the literature, this approach has led to interesting results in terms of performance compared to using one algorithm at a time. It belongs to the family of "ensembling methods". According to the task to be performed, classification or prediction, combinatorial methods can be implemented in two ways: by voting or calculating the average of the resulted predicted values. Since, this is about performing the classification of students' performance data, the suggested model will be implemented using a voting algorithm [10].

As far as the machine learning algorithms are concerned, four well-known algorithms were chosen. In addition to them, the model was enriched using an advanced machine learning classifier that is an improvement of the decision trees and has given interesting results in terms of performance. It is the Xgboost algorithm, it was used widely in Kaggle competitions and in financial and risk assessment domains. It resulted in better accuracy in models [11].

The Xgboost was included in our classification model because it has several advantages. According to [12], this algorithm:

- Minimizes overfitting thanks to introducing the regularization step.
- Uses parallel processing in order to enhance the speed of training.
- Increases the flexibility by allowing users to include predefined evaluation criteria and goals.
- Handles missing values.
- Controls the complex nature of decision trees by using special steps in pruning

According to Figure1, the architecture of our approach is as follows:

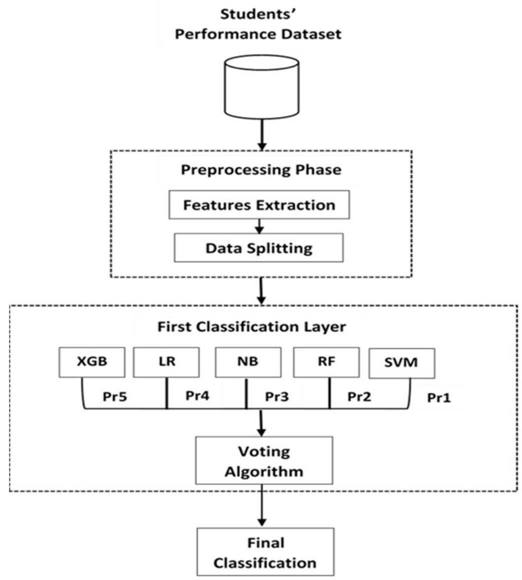

Figure 1. Proposed Model Architecture [13]

It starts by the pre-processing of the students' performance dataset by extracting the features and preparing the training and testing sets that will be used by the machine learning algorithms.

Concerning the choice of machine learning algorithms, five of them were chosen as they were used widely in many domains:

- Support Vector Machine (SVM) [14]: is one of machine learning algorithms used for performing supervised learning for both tasks: classification and regression of datasets. They are based on statistical learning theory. They perform data analysis through using algorithms and kernels like radial basis function, polynomial and quadratic. SVMs were used widely in different domains like clinical decision making support. For the classification task, it is used as a linear classifier functions through building a hyperplane separating the classes. The best one is the one providing the maximal margin. It is adequate for small, nonlinear and high dimensional datasets.
- Random Forest (RF) [15]: it is a well-known machine learning algorithm. It is widely used for classification. It works through building trees on the bootstrap data that are diverse and less correlated. This algorithm picks the most optimal split from all the features chosen

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

randomly for each intermediate node. This algorithm is used widely as it produces higher accuracy. Also, it is noticeably quicker and fast during training and predictions' phase; it is adequate for parallel processing, and can be used for multi-class capabilities. It is used with high dimensional data and delivers higher performance.

- Naive Bayes (NB) [16]: It performs the classification task using the Bayes' theorem. The key point about this algorithm is the ability of adapting events as soon as new data is added. It works by either assuming naïve or strong independence amongst the different attributes of data points. It is applied in many domains such as spam filtering, medical diagnosis and text analysis. Concerning the implementation of this algorithm, it is easy and simple to implement and run compared to other Bayes' algorithms.
- Logistic Regression (LR) [17]: It is used for supervised learning to predict the probability of an outcome which is binary. Such outcome has one or maximum two possible values. This machine learning algorithm performs predictive modelling in order to analyze large datasets given that one or more independent variable is responsible for deciding on the outcome. As this algorithm works by estimating the probability that an instance is part of class A or B, it uses the sigmoid function for mapping the estimated predictions with their probabilities. It compares the estimated probability with the threshold to decide the belonging of an instance to either class A or B.
- XGBOOST (XG) [18]: Is an algorithm that was used widely and successfully in Kaggle competition and in the domain of applied machine learning for structured data. It is a concrete implementation of the gradient boosting decision trees in order to achieve higher performance. Thanks to the contributions of experts, xgboost is an open access library that can be imported and used. It includes new features like regularization under three forms of gradient boosting (the classical gradient boosting, the scholastic gradient boosting and the regularized gradient boosting).

Each algorithm is trained using the same training data, then the models are generated. While performing the testing operation, each algorithm makes its decision of classification. The decisions generated in the previous step are input to the voting algorithm that will work on building a final classification decision. According to [19], such methodology:

- Facilitates the integration of architectures of several types of classifiers.
- Experiments done previously showed the positive impact of such approach on the performance of classification task. It outperforms using classification algorithms individually.

This approach can be implemented in two ways: the hard or the soft way as shown in the following figure:

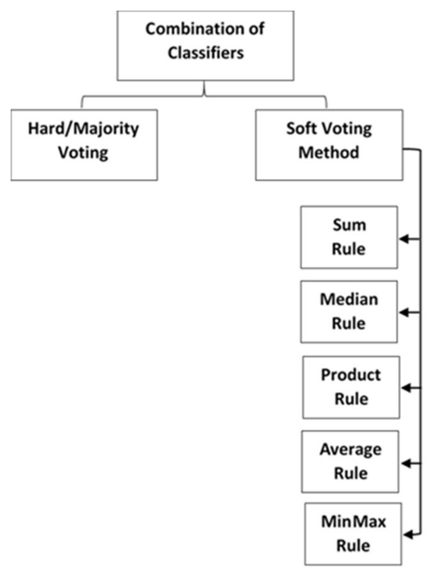

Figure 2.Voting Approaches

As shown in the figure above, the hard approach of combining classifiers is equivalent to majority voting. Each classifier result is considered as a vote for a class, and the majority wins. Whereas for the soft method of combination, it is based on the aposteriori probabilities of each class. According to Figure2, here are the various combination approaches: sum, medium, product, average or min/max values.

3. RESEARCH METHOD

As far as the experimentations of this article are concerned, the effectiveness of applying the XGboost classifier as being one of the most successful classifiers in the literature will be proven. Also, the voting classifier will be applied using the soft method to prove its efficiency in improving the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

accuracy of the classification tasks. These methods will be compared with the four classical algorithms: LR, NB, SVM, RF. For the experiment, two datasets will be used to confirm our results: the students' academic dataset collected from Jordanian University (the first dataset) and the Portuguese student dataset (the second dataset).

3.1. Overview of the first dataset: Students' Academic Performance Dataset

The dataset that will be used in performing the experiment is the Students' Academic Performance Dataset [20]. It is collected from Kalboard 360 which is a learning management system. Such LMS offers, to students, access to educational resources synchronously via any device online. It contains 480 students' records and is made of 16 features. These attributes belong to three main categories: First, academic background, section, grade level and educational stage. Second, the demographic attributes including nationality and gender. Third, the behavioral ones like activity in class, school satisfaction, and parents' answers to surveys. In the dataset, students are classified in three labels: Low-Level, Middle-Level, High-Level.

Gender	Nationality	PlaceOfBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	RaisedHands	VisitedResources	AnnouncementsView
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	15	16	2
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	20	20	3
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	10	7	0
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	30	25	5
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	40	50	12
F	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	42	30	13
M	KW	KuwaIT	MiddleSchool	G-07	A	Math	F	Father	35	12	0

Figure 3: Overview of the first dataset [20]

3.2. Overview of the Student Academic Performance Portuguese Dataset

This dataset [21] was collected from Portuguese schools based the schools' questionnaires and reports. It contains information about student performance in secondary education. The features represented in the dataset can be grouped in the following categories: social, demographic, and academic achievement. The dataset contains 33 attributes. It is about two subjects at secondary school, which are Mathematics and Portuguese Language. The target attribute is G3, which represents the final grade. It can be classified in the following classes: Poor, Fair, Good. It is strongly correlated to the grades of the first and second period of the academic year.

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	4	0	11	11
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	2	9	11	11
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	6	12	13	12
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	0	14	14	14
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	0	11	13	13

Figure 4: Overview of the second dataset [21]

3.3. Setup

These experiments were run on CPU, on a machine with the following characteristics:

- Processor: Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz
- RAM: 8,00 Go
- The machine has windows 64 bits

The implementation of the different classification algorithms was done on Jupyter Notebook using the python language. Several libraries from python were used like pandas, xgboost, seaborn. These dependencies and tools were installed thanks to Anaconda. It is a platform allowing the implementation and management of different machine learning algorithms.

3.4. Evaluation Metrics

For evaluating the results of the experiments, several metrics were used:

- Accuracy: it refers to the measurement of the ratio of correct predictions over the total number of instances evaluated, and is calculated through Equation (1) [22] where:

  tp: true positive
  tn: true negative
  fp:false positive
  fn: false negative

$$\text{Accuracy}=(tp+tn)/(tp+fp+tn+fn) \qquad (1)$$

- Precision: it is called the rate of true positive. It is the portion of relevant retrieved items out of all retrieved items. The higher the value is the better [23].

$$\text{Precision} = tp/(tp + fp) \qquad (2)$$

- Recall: it is the rate of true positive which is the portion of retrieved items that are relevant out of all relevant items. The higher the value is the better [24].

$$\text{Recall} = tp/(tp + fn) \qquad (3)$$

- F1 Score: it combines the two previous metrics. This is because it is the harmonic mean of the precision and recall. It is expressed in the following equation[25]:

$$\text{F1 Score} = tp/(tp + 0.5(fp+fn)) \qquad (4)$$

During the experimentations, confusion matrix was used. It englobes precision, recall, accuracy and f1 score. This is to give an idea about the performance of the classification model.

Journal of Theoretical and Applied Information Technology
15th May 2022. Vol.100. No 9

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.5. Data Preparation and Distribution

Before feeding the data to the classification model, they were preprocessed through the following steps:

- Check if missing values exist or not
- Transform labels to integers
- Dataset is divided into features and target
- Split features into training, testing and validation data

4. RESULTS AND DISCUSSION

4.1. The First Dataset

Table 1 shows the results of executing the first layer of classification. The algorithms SVM, RF, NB, LR and Xgboost were run on the first dataset.

Table 1. Results of the First Dataset

Model	Accuracy
SVM	83%
RF	80%
NB	80%
LR	83%
XB	84%

```
Testing Set
Accuracy: 0.9246575342465754

              precision    recall  f1-score   support

           0     1.0000    1.0000    1.0000         6
           5     1.0000    1.0000    1.0000         1
           6     1.0000    1.0000    1.0000         1
           7     0.8000    1.0000    0.8889         4
           8     1.0000    0.9412    0.9697        17
           9     1.0000    0.8421    0.9143        19
          10     0.9184    0.9574    0.9375        47
          11     0.9333    0.9767    0.9545        43
          12     0.9600    0.8000    0.8727        30
          13     0.9268    0.9744    0.9500        39
          14     0.9062    0.9062    0.9062        32
          15     1.0000    0.8696    0.9302        23
          16     0.9167    0.8462    0.8800        13
          17     0.7000    1.0000    0.8235        14
          18     1.0000    1.0000    1.0000         3

    accuracy                         0.9247       292
   macro avg     0.9374    0.9409    0.9352       292
weighted avg     0.9321    0.9247    0.9251       292
```

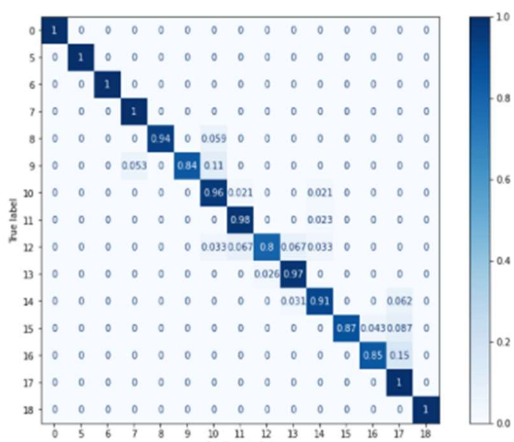

Figure 5. Confusion Matrix Of The First Dataset Using The Voting Classifier On The Testing Dataset

Figure 5. Confusion Matrix of the first dataset using the voting classifier on the testing dataset

After applying the voting classifier, using the soft method, the performance of the classification was improved. Here are the results:

```
Testing Set
Accuracy: 0.8576388888888888

              precision    recall  f1-score   support

           0     0.8256    0.8554    0.8402        83
           1     0.8933    0.9178    0.9054        73
           2     0.8583    0.8258    0.8417       132

    accuracy                         0.8576       288
   macro avg     0.8591    0.8663    0.8624       288
weighted avg     0.8577    0.8576    0.8574       288
```

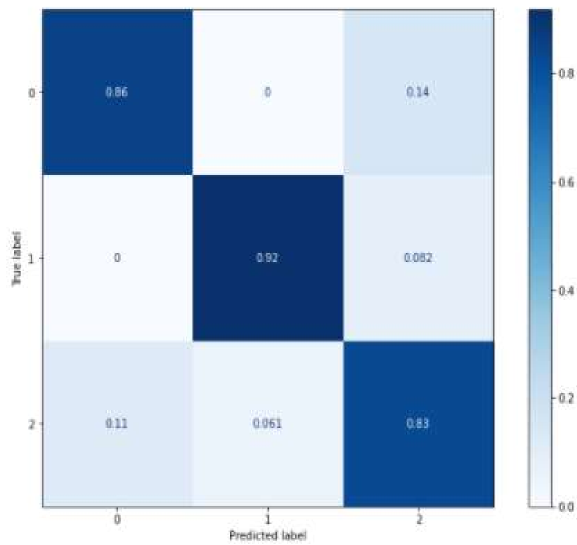

Figure 6. Confusion Matrix Of The Second Dataset Using The Voting Classifier On The Testing Dataset

4.2 The Second Dataset

The same experiment done in the previous step was re-applied on the second dataset. Results are shown in Table 2:

Table 2.Results of the Second Dataset

Model	Accuracy
SVM	90%
RF	90%
NB	50%
LR	91%
XB	87%

Using the voting approach has enhanced the performance of the classification to 85% for the first dataset and 92% for the second one. The soft method was adopted, and the following matrix illustrates the results:

Table 3. Summary of Executing the Voting Classifier

Model	Accuracy
First DS	85%
Second DS	92%

DS: Dataset

Throughout the experimentation, it was remarked that for the first dataset, the Xgboost classifier gave a good result 84%, after applying the voting classifier using the soft method; a better result was achieved, which is 85%. This proves the effectiveness of our suggested approach. In order to further confirm these findings, the same methodology was applied on a second dataset (the Portuguese one). Results have confirmed the findings of the first dataset, 87% accuracy was achieved with Xgboost classifier, and the voting classifier has improved it much better to reach 92%. Hence, using the voting approach and most precisely the soft method of combining contributes in improving the performance of classification.

Compared to the literature, in the work of [3], similar experiments were performed on the same e-learning datasets have resulted in 80% of accuracy. The proposed architecture has enhanced this metric to 85% for the first dataset and 92% for the second one.

5. CONCLUSION

Throughout this article, it started by introducing the issue of e-learning data classification problem, especially, the ones concerning students' academic performance. Then it moved to presenting a state of art of the already achieved works in that sense after the introduction. Later, the proposed approach in tackling the issue in question was presented along with its architecture and modules. In order to prove the efficiency of this article's proposition, a detailed experimentation was performed to compare the classical machine learning algorithms mainly support vector machine, random forest, linear regression and naïve bayes with the Xgboost classifier. This latter is considered as being successful in classification works. Concerning the voting classifier, the soft method of combining the different classifiers was applied. Finally, the findings of the two datasets were discussed. Concerning the future works, in order to achieve even better results, the following propositions can be considered:

- A new layer can be added to the proposed architecture in this article. The role of this layer would be filtering the features in order to use only the ones that would result in the best performance of the task. There are several methods in the literature that have proven their efficiency in features filtering. Such proposition would guarantee an interesting added value to our approach in the future.
- Fine-tuning the actual classification model to achieve higher performance Metrics.
- Applying the classification model on more datasets to confirm the results.
- Using large sized data to improve the learning of the model.
- Using more sophisticated machines for execution that have GPUs allowing the handling of large sized data.

REFERENCES

[1] Livieris, I. E., Drakopoulou, K., Tampakas, V. T., Mikropoulos, T. A., & Pintelas, P. (2019). Predicting secondary school students' performance utilizing a semi-supervised learning approach. Journal of Educational Computing Research, 57 , 448–470. URL: arXiv: https://doi.org/10.1177/0735633117752614.

[2] Hussain, S., Muhsen, Z., Salal, Y., Theodorou, P., Kurtoglu, F., & Hazarika, G. (2019). Prediction model on student performance based on internal assessment using deep learning. International Journal of Emerging Technologies in Learning (iJET), 14 , 4. doi:10.3991/ijet.v14i08.10001.

[3] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. ICEIT 2019: Proceedings of the 2019 8th International Conference on Educational and Information Technology.(pp. 7{11). doi:10.1145/3318396.3318419.

[4] Mahmoud Abu Zohair, L. (2019). Prediction of student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education. doi:10.1186/s41239-019-0160-3.

[5] Lau, E., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial

neural networks. SN Applied Sciences, 1 . doi:10.1007/s42452-019-0884-7.

[6] Chaplot, D., Rhim, E., & Kim, J. (2015). Predicting student attrition in moocs using sentiment analysis and neural networks. AIED 2015 Workshop Proceedings. volume 1432.

[7] Burgos, C., Campanario, M., Peña, D., Lara, J., Lizcano, D., & Martínez, M. (2017). Data mining for modeling students' performance: A tutoring action plan to prevent Academic dropout. Computers Electrical Engineering,66. doi:10.1016/j.compeleceng.2017.03.005

[8] Xing, W., & Du, D. (2019). Dropout prediction in moocs: Using deep learning for personalized intervention. Journal of Educational Computing Research, 57, 547–570.URL:https://doi.org/10.1177/0735633118757015. arXiv:https://doi.org/10.1177/0735633118757015.

[9] Yang, C., Chiang, F.-K., Cheng, Q., & Ji, J. (2021). Machine learning-based student modeling methodology for intelligent tutoring systems. Journal of Educational Computing Research, 0 , 0735633120986256. URL: https:// doi.org/10.1177/0735633120986256. doi:10.1177/0735633120986256. arXiv:https://doi.org/10.1177/0735633120986256.

[10] Alves, A. (2016). Stacking machine learning classifiers to identify higgs bosons at the lhc. Journal of Instrumentation, 12 . doi:10.1088/1748-0221/12/ 05/T05005.

[11] Chang, Y.-C., Chang, K.-H., & Wu, G.-J. (2018). Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. Applied Soft Computing, 73 . doi:10.1016/j.asoc. 2018.09.029.

[12] Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of p2p network loan default based on the machine learning lightgbm and xgboost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications, 31 . doi:10. 1016/j.elerap.2018.08.002.

[13] Saez, J. A., Galar, M., Luengo, J., & Herrera, F. (2015). Inffc: An iterative class noise filter based on the fusion of classifiers with noise sensitivity control. Information Fusion, 27 . doi:10.1016/j.inffus.2015.04.002.

[14] Yao, Lei & Fang, Zhanpeng & Xiao, Yanqiu & Hou, J.-J & Fu, Zhijun. (2021). An Intelligent Fault Diagnosis Method for Lithium Battery Systems Based on Grid Search Support Vector Machine. Energy. 214. 118866. 10.1016/j.energy.2020.118866.

[15] Han, Sunwoo & Kim, Hyunjoong & Lee, Yung-Seop. (2020). Double random forest. Machine Learning. 109. 10.1007/s10994-020-05889-1.

[16] Gao, C.-z., Cheng, Q., He, P., Susilo, W., & Li, J. (2018). Privacy-preserving naive bayes classifiers secure against the substitution then comparison at-tack. Information Sciences, 444 . doi:10.1016/j.ins.2018.02.058.

[17] Os, H., Ramos, L., Hilbert, A., Leeuwen, M., van Walderveen, M., Kruyt, N., Dippel, D., Steyerberg, E., Schaaf, I., Lingsma, H., Schonewille, W., Majoie, C., Olabarriaga, S., Zwinderman, K., Venema, E., Marquering, H., & Wermer, M. (2018). Predicting outcome of endovascular treatment for acute ischemic stroke: Potential value of machine learning algorithms. Frontiers in Neurology, 9 . doi:10.3389/fneur.2018.00784.

[18] Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A datadriven design for fault detection of wind turbines using random forests and xgboost. IEEE Access, 6 , 21020–21031. doi:10.1109/ACCESS.2018. 2818678

[19] Mohandes, M., Deriche, M., & Aliyu, S. (2018). Classifiers combination techniques: A comprehensive review. IEEE Access, PP, 1{1. doi:10.1109/ ACCESS.2018.2813079.

[20] Amrieh, E., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. International Journal of Database Theory and Application, 9 , 119{136. doi:10.14257/ijdta.2016.9.8.13.

[21] Cortez, P., & Silva, A. (2008). Using data mining to predict secondary school student performance. Dep. Information Systems/Algoritmi R&D Centre. EUROSIS.

[22] Silveira, T., Zhang, M., Lin, X. et al. How good your recommender system is? A survey on evaluations in recommendation. Int. J. Mach. Learn. & Cyber. 10, 813–831 (2019). https://doi.org/10.1007/s13042-017-0762-9

[23] Katzman, J., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. (2018). Deepsurv: Personalized treatment recommender system using a cox proportional hazards deep neural network. BMC Medical Research Methodology, 18.

[24] Huang, Z., Yu, C., Ni, J., Liu, H., Zeng, C., and Tang, Y. (2019). An efficient hybrid recommendation model with deep neural networks. IEEE Access, 7:137900–137912.

[25] Dalianis,H.. (2018). Evaluation Metrics and Evaluation. 10.1007/978-3-319-78503-5_6.