A Machine Learning Based Approach for Student Performance Evaluation in Educational Data Mining

Muhammad Sammy Ahmad
Department of Computer Science
FGSSR, Cairo University
dr.muhammad.badran@gmail.com

Ahmed H. Asad
Department of Computer Science
FGSSR, Cairo University
ah_hamza@cu.edu.eg

Ammar Mohammed
Department of Computer Science
FGSSR, Cairo University
ammar@cu.edu.eg

2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) | 978-1-6654-1243-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/MIUCC52538.2021.9447602

***Abstract*—Astonishing progress in machine learning and data mining techniques has been achieved during the last two decades. Education should benefit from these improvements to discover about how people learn in different educational settings. The aim of the research is comparing artificial neural network (ANN) to random forest (RF) machine learning models for predicting performance of students based on their demographic and assessment information. After analyzing the Open University Learning Analytics Dataset (OULAD), we applied feature-engineering techniques, and then the two models were evaluated. Our results showed that the ANN model outperformed the RF model accuracy 91.08% to 81.35 %. ANN performs well on educational data and can be efficiently used for predicting student performance and in early warning systems.**

Keywords— Predicting Student Performance, Machine learning, Artificial Neural Network, Educational Data Mining, Learning Analytics

I. INTRODUCTION

Data mining reveals patterns in massive datasets to extract useful information. A data miner should have knowledge about database systems, statistics, and machine learning, in addition to knowledge about the domain from which the data were collected. The continuous increase in the number of researchers who are affiliated with the education domain and are concerned with applying data mining techniques on data generated by student's activities in learning environments, led to emergence of a relatively new scientific discipline called educational data mining (EDM). EDM converts raw data that comes from different learning environments or educational settings into useful information for research [1]. Predicting student performance (PSP) is considered the oldest application of data mining techniques in education and one of the most popular and challenging problems that has been tackled by EDM through the previous years [2]. The prediction can be for a numerical value or a categorical value, the former would be a regression task, while the latter would be a classification task.

Initial attempts of using machine-learning tools, such as decision trees and artificial neural networks, in EDM for PSP were in early 2000s [3]. Decision tree (DT) is popular machine learning tool that is used for building a predictive model, which maps from observations to target values [4]. There are two main types of DTs, which are classification trees and regression trees. If the predicted values are discrete, they are named classification trees, but if the predicted values are continuous, they are named regression trees [5]. An ensemble method, which is called bootstrap aggregation method, is used for building bootstrapping aggregated DTs, or bagged DTs, that have higher performance than normal DTs because instead of using a single DT to get a prediction, multiple DTs are trained after resampling the data by replacement, and then they vote for the predicted outcome. The chosen predicted value for the output is the one that takes the highest number of votes. Random Forest (RF) uses the bootstrap aggregating ensemble method [6].

Artificial neural network (ANN) is one of the most efficient machine learning tools. It continuously proves itself as a popular and a robust tool for complex modeling in machine learning, which can be used for creating models of nonlinear functions that describe most real-life systems [7]. They are efficient and successful in performing many tasks such as regression, classification and clustering. The common structure of ANN is: (1) an input layer that usually has a large number of input nodes, (2) a number of hidden layers, and (3) an output layer that can perform regression or either binary or multiclass classification. The connections between nodes are weighted, and back-propagation learning algorithm adjusts these weights until they reach their optimum values. Multilayer ANN can overcome the limitations that faced the back-propagation learning, and their models can be trained, not only to classify data, but also to generate sensory data [8].

For solving the PSP problem, deep DT-based models and different types of shallow/deep ANN models can be used on dataset from a virtual learning environment (VLE), and it was found that student interaction with environment, such as the number of clicks on the resources, has a great impact on determining the outcome [9]. Although deep DTs have low bias, they have a high variance, in other words, deep DTs models are prone to overfitting on the training set. This phenomenon is called bias-variance tradeoff [10]. This effect can be mitigated by using RF that averages the output of DTs to decreases the variance on the expense of increasing bias. Using RF model greatly improves the performance but interpretability decreases [11]. On the other hand, ANNs have shown their higher performance on educational data in relatively few recent researches [12]. However, their performances on educational data is still questionable because their accuracies of prediction in most studies until now were average or a little above average [13].

The contributions of this research are: (1) to verify whether ANN can reach the-state-of-art performance on a relatively large amount of educational data as it does on other data that come from different fields, (2) to compare ANN performance with other models performances such RF model. Therefore, we built two machine learning based models and compare their performances. The first model used ANN with TensorFlow as backend and the other one used RF to predict the academic performance of students using just their demographic and assessments data in the Open University Learning Analytics Dataset (OULAD) [14].

This paper is organized as follows: section II gives a background about evaluation metrics which were used to evaluate the models, and surveys some of the most important

978-1-6654-1243-8/21/$31.00 ©2021 IEEE

previous related studies, section III presents the proposed approach to solve the research problem, section IV introduces the experimental work that was carried out during this research, and it includes the following subsections: software and tools and libraries, dataset, results, and discussion. Finally, section V provides the conclusion and some suggested future work.

II. Historical Background

Several evaluation metrics are used to evaluate the performance of machine learning-based classification models, such as accuracy, precision, recall, and F1 score which is calculated from precision and recall.

In this research, we used accuracy, precision, recall, and F1 score metrics. Therefore, this section gives a background about these metrics, which their calculations depend on building a confusion matrix (CM). CM finds true positives (TPs), false positives (FPs), true negatives (TNs) and false negatives (FNs) from the classification result that is outputted by a classifier model. Table I shows where each group is located in the CM for the three classes that were used in this research, namely “Distinction”, “Fail”, and “Pass” classes.

True positives (TPs) are the test cases in which the actual class matches the predicted class. For example, it is a TP when the actual class is “Pass” and the predicted class is “Pass” too. The same rule applies for the “Fail” and “Distinction” classes. The values of TPs can be found on the diagonal of the CM.

False positives (FPs) are the test cases in which the actual classes mismatches the predicted class. An instance is classified as a member of a certain class, but actually it does not belong to it. For example, when the predicted class is “Pass”, but the actual class is either “Fail” or “Distinction”. The values of FPs are the values in the column of each class except the TPs values which are described above.

True negatives (TNs) are test cases that are correctly classified as not being members of a certain class. The values of TNs for each class can be found by getting the sum of all columns and rows in the CM except the column and row of that class. For example, TNs for ‘Distinction’ class is the sum of all values in CM except values in row [0] and column [0].

TABLE I. Positions of TPs, FPs, TNs and FNs in the CM

		Predicted Classes		
		Distinction [0]	***Fail [1]***	***Pass [2]***
Actual Classes	***Distinction [0]***	$TP_{Distinction}$ TN_{Pass} TN_{Fail}	FP_{Fail} $FN_{distinction}$ TN_{Pass}	FP_{Pass} $FN_{distinction}$ TN_{Fail}
	Fail [1]	$FP_{distinction}$ TN_{Pass} FN_{Fail}	TP_{Fail} $TN_{distinction}$ TN_{Pass}	FP_{Pass} $TN_{distinction}$ FN_{Fail}
	Pass [2]	$FP_{distinction}$ FN_{Pass} TN_{Fail}	FP_{Fail} $TN_{distinction}$ FN_{Pass}	TP_{Pass} $TN_{distinction}$ TN_{Fail}

False negatives (FNs) are the test cases that are incorrectly classified as not being members of a certain class. The values of FNs are the values in the row of each class except the TPs values which are described above. After determining the locations of TPs, FPs, TNs, and FNs in CM, accuracy, precision, recall, and F_1 score can be calculated.

Accuracy, or error rate, which is the ratio of the correctly predicted instances (TPs and TNs) to the total data points (TPs, TNs, FPs, and FNs). It can be defined by equation (1):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision, which is also named positive predictive value, is a measure of accuracy of prediction of a specific class. For example, it measures how much the classifier model is accurate in predicting the “Distinction” class correctly. Precision can be calculated using equation (2):

$$Precision = \frac{TPs}{TPs + FPs} \tag{2}$$

Recall, which is also named sensitivity or true positive rate, is the ability of the classifier model to detect relevant cases in the dataset. For example, the classifier model is able to detect “Fail” class from all actual “Fail” instances, including those who were classified as not “Fail”, but actually, they belong to that class. Recall can be calculated using equation (3):

$$Recall = \frac{TPs}{TPs + FNs} \tag{3}$$

Finally, the general formula to measure F score for positive real β is named F-beta score, which can be calculated from equation (4), and it is the harmonic mean of precision and recall:

$$F_\beta = (1 + \beta^2).\frac{precision.recall}{(\beta^2.precision) + recall} \tag{4}$$

The default value of β is 1.0, so in this case it is called F_1 score. F_1 reaches its best value at 1, which means perfect, and its worst value is 0. It can be written as in equation (5):

$$F_1 = \frac{2 \times precision \times\ recall}{precision + recall} \tag{5}$$

Generally, the majority of previous researches regarding PSP problem can be classified into two main approaches: (1) a traditional approach that uses generalized linear models such logistic regression and linear support vector machines (SVMs) [15]; and (2) an emerging approach that uses different types of ANNs in their models. Performance of an ANN based model can be improved by using feature engineering to reduce the number of input dimensions that limit the performances of these models [16].

From the researches of the first category which use traditional statistical methods or simple machine learning algorithms, is a research conducted by Hlosta *et al.* in an attempt to identify students at-risk of failing or dropout their course using legacy data without models. They were able to predict the submission of the first assessment, A1, from demographic information and VLE activities before A1 assessment. They found that A1 assessment is a very important predictor for students who are at-risk of failure because the probability of failure of students who do not submit A1 assessment is 90% [17].

Sherimon and Puliprathu used multiple linear regression model to predict the performance of students of Arab Open University based on their assessment marks in Relational Database course which was taught in a blended learning environment, and their finding was that building a predictive model for predicting performance of students is challenging. They concluded that accurate prediction would be difficult [18].

Some researches focus mainly on tree-based classifiers, such as a research conducted by Khalaf *et al*., they used 161 questionnaires, each contains 60 questions, and then three classifiers were built which were J48, Random Tree, and REP Tree using Weka 3.8 tool. They concluded that J48 algorithm had the best performance compared to the other two algorithms [19]. Furthermore, Doijode and Singh used Decision Trees (DT) on VLE data to identify most important features help predicting student success [20].

Different types of neural networks show their superiority above many other machine learning algorisms and prove their capability of solving difficult and complex problems such image classification, speech recognition, and machine translation. In contrast, Baker and Inventado claimed that ANNs are not a typical and a less common choice in EDM although they are widely used in machine learning community, while other conservative algorisms are more successful than neural networks [21]. Their claim was because results of ANNs on educational data were usually lower than results on data that come from other fields.

From the second category of researches that focus on using different types of ANNs, Gray *et al*. conducted a research for three years to predict students at risk of failure in first year using an online, self-reporting, learner-profiling tool and students' enrollment data, the researchers compared between eight classification algorithms, and the accuracies were between 70% and 72% [22]. ANNs are able to achieve the-state-of-the-art performance in many modeling and prediction tasks, but their performances on educational data was questionable because their accuracies of prediction in most studies that have been covered during this research were average or a little above average [13].

On the other hand, ANNs have shown an acceptable performance on educational data in few relatively recent studies. For instance, Adewale *et al*. used feed-forward ANN of topology (5-5-5-1) to predict academic performance of students of secondary school during submission of their applications in universities based on some cognitive and psychological factors. They concluded that ANN is efficient and achieve high accuracies [16].

Aydoğdu used ANN used for predicting student performance based on data from Learning Management System (LMS) that was used by 3518 university students. The accuracy of prediction was 80.47%. The researcher used different predictors such as gender, content score, homework score, number of times students entered the content, number of times students attended live sessions, the time they spent attending them, the number of times students viewed archived sessions, and the time they spend viewing them. They found that the independent variables that contribute most to predicting student performance are their course attendance and the time they spent viewing the content [23].

Cumulative Grade Point Average (CGPA) can be evaluated and predicted by using social data, economic data, and results of the entrance exam of the undergraduate university students using ANN. The performance of their ANN was evaluated using confusion matrix, error histogram, regression analysis, and Mean Square Error (MSE). The accuracy of the ANN reached 84.8%, but it performed inefficiently when it is used to classify student performance according their gender because they used imbalanced sample which increased the false negative rates [24].

Some researchers give more attention to the factors affecting student performance rather than the type of model which is used for prediction. Some researches use both educational and demographic factors of student to carry out their studies, but other researchers focus on students' behaviors and activities. In 2015, Bainbridge *et al*. conducted a research to analyze students' behaviors such as number of login times, forums participation, and accessing lessons to identity factors or indicators that indicate risk of failure. They found that these indicators can be used for predicting student academic performance and can be used in an early warning system, along with other traditional factors such as age, assignments scores, class size, and grade point average (GPA) of students [25]. It worth mentioning that some researchers used students' facial expressions to evaluate their performances [26-28].

III. Proposed Appraoch

The proposed approach that was used to tackle the research problem will be discussed in this section. The methodology which was used for solving the problem under study started with finding an adequate dataset. This research is the first research paper that uses ANN on OULAD to predict student academic performance. Only relevant features has been selected and engineered. A survey has been carried out to select the best models that can be used. After building the models, they were tested and evaluated.

A. Data Manipulation

Demographic and assessment data were imported from the dataset. Records which contain missing data were dropped from the dataset. The dependent vector, which is the "final_result" column in "studentinfo" table, contains four value which are "Withdrawn", "Fail", "Pass", and "Distinction." Students who withdrew and did not finish their modules were excluded from this study due to their missing and incomplete data. Finally, we ended up with three categories which are "Distinction", "Pass", and "Fail".

B. Feature Engineering

Because artificial neural networks deal only with numerical data, all categorical data should be properly encoded. For example, module codes and regions are encoded using one hot encoder, and dummy variable trap has been avoided by removing one of the dummy variable columns to prevent correlation between input variables. Highest level of education, age bands, and Indices of Multiple Deprivation (IMD) band were encoded in a way that represent their ordinal nature. Finally, binary variables such gender and having disability were encoded using label encoder.

C. Building ANN and RF Models

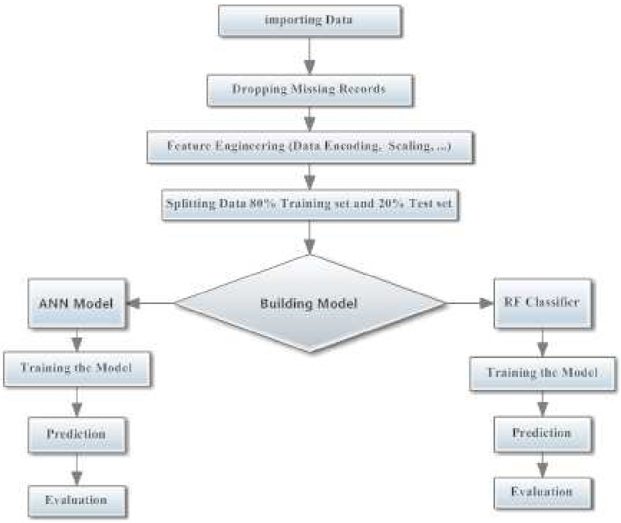

Fig. 1. The ANN and RF Model Appraches

Two machine learning models were built and their performances were compared. The first model uses ANN with TensorFlow as backend and the other one uses RF to predict academic performance of students using the OULAD.

The ANN model in this research consisted of 5 layers (4 hidden layers and 1 output layer). The input layer contained 31 nodes, while each hidden layer contained 128 nodes, and the output layer contained only 3 nodes for “Distinction”, “Pass”, and “Fail” cases. The numbers of input nodes and output node should be equal to the numbers of input features and the output classes respectively, while other parameters such number hidden layers and number of node in each layers were chosen after trying many different combinations of them.

In case of the RF model, five different RFs were trained during this research. The best results were obtained from the RF model which consists of 500 DTs. The criterion parameter was set to “entropy”, while default values were used for other parameters. Both models were trained on 80% of the data, and the rest of data were used as a test set. Fig. 1 below illustrates both ANN and RF approaches.

IV. Experimental Work

A. Software Tools and Libraries

The following software tools were used during this research: Python 3.7.1, Jupyter notebook 5.7.8, Numpy 1.17.4, Pandas 0.24.2, TensorFlow 1.13.1, Keras 2.2.4, and Scikit-learn 0.22.

B. Dataset

OULAD was selected in this research because it is a clean, labeled, well documented, and large benchmark dataset that contains many interesting and relevant features that are required for solving the research problem. The dataset contains seven tables in csv format which are classified into three categories:

1) Demographic Information contains one table, namely “studentInfo” table. It stores information about students’ age, gender, region where they live, Index of Multiple Deprivation (IMD) of their areas, whether they are disabled or not, and their final result.

2) Activities Information: There are three tables in this category. They are “studentRegisteration” table which contains data about students’ registration, “studentVLE” table which contains information about students’ interaction with the virtual learning environment, and “studentAssessment” which contains information such as assessment scores and date of submission.

3) Modules and Presentations Information: The remaining three tables are “assessments” table which contains information about assessments types, “courses” table which contains information about modules, and “vle” table which contains information about students’ interaction with the learning environment.

Only the “studentinfo” and the “studentAssessment” tables from OULAD were used in this research to predict student academic performance. The two tables were merged together using the unique student_id column. Indices of multiple deprivation (imd) of North Region and Ireland are not compatible with the system which is used in other regions of the UK, so observations from these regions were excluded from this research.

The dependent vector, which is the “final_result” column in “studentinfo” table, contains four values. They are “Withdrawn”, “Fail”, “Pass”, and “Distinction”. Students who withdrew and did not finish their modules are excluded from this research due to their missing and incomplete data. Finally, we end up with three categories which are “Distinction”, “Pass”, and “Fail”. Because ANNs deal only with numerical data, all categorical data should be properly encoded.

C. Results

In this section, the results of the experimental work will be presented.

1) Accuracy: It shows an overall correctness of the model as a whole. The prediction accuracy of the ANN model on the test set was 91.06%, while that of RF model was 81.35%. The accuracy of ANN model is higher than that of the RF model by about 10%. Other metrics such as precision, recall, and F_1 score are calculated using scikit-learn library [29].

2) Precision: In case of the ANN model, the precision of prediction of “Distinction”, “Fail”, and “Pass” classes are 0.885, 0.893, and 0.921 respectively. While in case of RF model, the precision of prediction of “Distinction”, “Fail”, and “Pass” classes are 0.713, 0.711, and 0.792 respectively. The precision of prediction using ANN model for all classes is higher than that of the RF model.

3) Recall: The recall of “Distinction” class is 0.850, “Fail” class is 0.819, and “Pass” class is 0.952 in case of the ANN model. The recall of “Distinction” class is 0.677, “Fail” class is 0.594, and “Pass” class is 0.957 in case of RF model.

4) F_1 Score: F_1 score for “Distinction”, “Fail”, and “Pass” classes using ANN model are 0.867, 0.855, and 0.936 respectively. While when RF model is used, they are 0.695, 0.647, and 0.867 respectively.

5) Cross-Validation: “Pass” class has the best results among all the three classes in both models. To check for bias and variance of the ANN model, ten cross-validations were performed. Results show that the ANN model is not biased to any class, and the average accuracy of the ten cross validations was 90.64%, while the variance was 0.0023.

6) ROC Curve: Receiver Operating Characteristic (ROC) is a metric used for evaluating the quality of a classifier. It is usually used with binary classifiers, but its functionality can be extended to include multiclass classifiers.

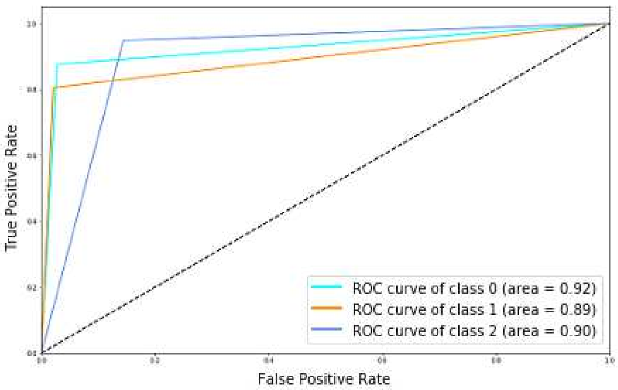

Fig. 2. ROC curve of the ANN model

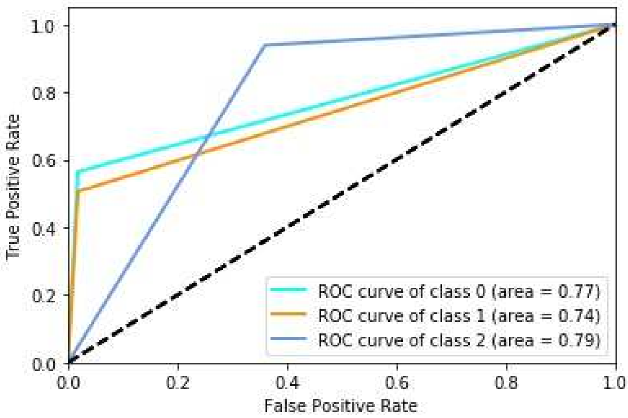

Fig. 3. ROC curve of the RF model

Confusion matrices of the two models summarize their outcomes and they are shown in following tables II and III.

TABLE II. CONFUSION MATRIX OF ANN

		Predicted Classes		
		Distinction	***Fail***	***Pass***
Actual Classes	***Distinction***	85.02%	1.82%	13.15%
	Fail	1.56%	81.97%	16.45%
	Pass	2.32%	2.41%	95.26%

TABLE III. CONFUSION MATRIX OF RF

		Predicted Classes		
		Distinction	***Fail***	***Pass***
Actual Classes	***Distinction***	67.74%	0.77%	31.48%
	Fail	9.52%	50.59%	39.88%
	Pass	4.07%	2.09%	93.38%

D. Discussion

The results show that ANNs are very efficient and can be used with educational data. They perform very well in predicting student academic performance based on their demographic and assessment data, but the following conditions must be met to get high results using ANNs:

- The input features must be relevant to student academic performance such as student highest level of education, their age, deprivation level of regions where they live, assessments scores, dates of assessments submissions, and number of previous attempts to solve the assessments, etc.
- Relatively large dataset with a large number of observations must be used to allow the model to "learn" from various training examples, and hence improves its performance on test set.
- Increasing the size of the neural net (i.e. the number of hidden layers and the number of nodes in them) improves its performance significantly until certain threshold. The size of the ANN depends on the amount of data available.

Although there is no single algorithm that can be the best choice for every problem, but ANNs show their capability in solving many difficult and complex problems. In most of the previous researches, performance of ANNs on educational data was questionable.

In our research, ANN and RF machine learning models were used for predicting academic performance of students of the OULAD dataset. ANN accuracy of prediction reached 91.08% on OULAD, which is considered high compared to accuracies that were obtained from most other researches that dealt with similar problems as shown below in the following table IV that compares between accuracies of prediction of ANN of our research and that of some other researches. The ANN model outperformed RF model in accuracy that was 81.35%, and in other evaluation metrics such as precision, recall, and F_1 score.

V. CONCLUSION FOR FUTURE WORK

The results of our research showed the invalidity of the claim that ANNs are not convenient for educational data. It is confirmed that ANNs perform very well on educational data as they do on data that come from other fields. Giving irrelevant or insufficient inputs to ANNs or using them on small datasets would decrease their performances. This research supports other researches that suggest using ANN on educational data to predict student academic performance and in early warning systems.

TABLE IV. COMPARISON BETWEEN RESULTS OF ANNS

Research	Year	Accuracy
Gray *et al.* [22]	2016	70% - 72%
Aydoğdu [23]	2020	80.47%
Lau *et al.* [24]	2019	84.8%
Adewale *et al.* [16]	2018	90%
Our Research	**2021**	**91.08%**

For the future work, VLE data can be included with demographic data in another research to predict student performance. Moreover, recurrent neural networks (RNNs) can be used for detecting students who are about to withdraw or dropout.

REFERENCES

[1] Romero C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE transactions on systems, man, and cybernetics, part C (applications and reviews), 40(6), 601-618.

[2] Asiah, M., Zulkarnaen, K. N., Safaai, D., Hafzan, M. Y. N. N., Saberi, M. M., & Syuhaida, S. S. (2019). A review on predictive modelling technique for student academic performance monitoring. In proceedings of materials science, engineering and chemistry (MATEC) web of conferences, 255 (03004).

[3] Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: an open source analytics initiative. Journal of Learning Analytics, 1(1), 6-47.

[4] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.

[5] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

[6] Rokach, L., & Maimon, O. Z. (2008). Data mining with decision trees: theory and applications (Vol. 69).

[7] Abiodun, O. I., Jantan, A., Omolara, A. E, Dada, K. V, Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: a survey. Heliyon, 4(11).

[8] Hinton, G. E. (2007). Learning multiple layers of representation. Trends In Cognitive Sciences, 11(10), 428-434.

[9] Aljohani, N. R., Fayoumi, A., & Ul-Hassan, S. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. Journal of Sustainability, 11 (24), 7238.

[10] Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Rodríguez, S., & Corchado, J. M. (2019, April). Students' performance analysis based on machine learning techniques. In International Workshop on Learning Technology for Education in Cloud (pp. 428-438). Springer, Cham.

[11] Al-shabandar, R., Hussain, A. J., Liatsis, P., & Keight, R. (2019). Detecting at-risk students with early interventions using machine learning techniques. IEEE Access, 7, 149464-149478.

[12] Kim, B. H., Vizitei, E., & Ganapathi, V. (2018). GritNet: Student performance prediction with deep learning. arXiv preprint arXiv:1804.07405.

[13] Gardner, J., Yang, Y., Baker, R., & Brooks, C. (2019, July). Modeling and experimental design for MOOC dropout prediction: a replication perspective. The 12th International Conference on Educational Data Mining (EDM2019). Montreal, Canada.

[14] Kuzilek J., Hlosta M., Zdrahal Z. (2017). Open University Learning Analytics Dataset. Sci. Data 4:170171 doi: 10.1038/sdata.2017.171.

[15] Umair, S., & Sharif, M. M. (2018). Predicting students grades using artificial neural networks and support vector machine. In Encyclopedia of Information Science and Technology, 4th Edition (pp. 5169-5182). IGI Global.

[16] Adewale, A. M., Bamidele, A. O., & Lateef, U. O. (2018). Predictive modelling and analysis of academic performance of secondary school students: artificial neural network approach. International Journal of Science and Technology Educational Research, 9(1): 1-8.

[17] Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017, March). Ouroboros: early identification of at-risk students without models based on legacy data. In proceedings of the seventh international learning analytics & knowledge conference (pp. 6-15).

[18] Sherimon V., Puliprathu Cherian S. (2018) Building a Multiple Linear Regression Model to Predict Students' Marks in a Blended Learning Environment. In: Auer M., Tsiatsos T. (eds) Interactive Mobile Communication Technologies and Learning. IMCL 2017. Advances in Intelligent Systems and Computing, vol 725. Springer.

[19] Khalaf, Alaa & Hashim, Ali & Akeel, Wid. (2018). Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis. International Journal of Interactive Multimedia and Artificial Intelligence. 5. 26-31. 10.9781/ijimai.2018.02.004.

[20] Doijode, V. & Singh, N. (2016). Predicting student success based on interaction with virtual learning environment. SAS Conference Proceedings: Southeast SAS Users Group (SESUG 2016). October 16-18, 2016. Bethesda, Maryland.

[21] Baker, R. & Inventado, P. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson, B. White (Eds.), Learning Analytics: From Research to Practice (pp. 61-75). New York, NY: Springer.

[22] Gray, G., McGuinness, C., Owende, P., & Hofmann, M. (2016). Learning Factor Models of Students at Risk of Failing in the Early Stage of Tertiary Education. Journal of Learning Analytics, 3(2): 330-372.

[23] Aydoğdu, Ş (2020). Predicting student final performance using artificial neural networks in online learning environments. Educ. Inf. Technol. 25(1913–1927).

[24] Lau, E.T., Sun, L. & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. SN Appl. Sci. 1(982).

[25] Bainbridge, J., Melitski, J., Zahradnik, A., Lauría, E. J., Jayaprakash, S., & Baron, J. (2015). Using learning analytics to predict at-risk students in online graduate public affairs and administration education. Journal of Public Affairs Education, 21(2): 247-262.

[26] Hesham, H., M. Nabawy, O. Safwat, Y. Khalifa, H. Metawie, and A. Mohammed (2020, June). Detecting education level using facial expressions in e-learning systems. International Conference on Electrical, Communication, and Computer Engineering (ICECCE), pp. 1-6.

[27] Megahed, M., and A. Mohammed (2020). An approach to model adaptive e-learning environment using facial expressions and fuzzy logic. Expert Systems with Applications, 157(113460).

[28] Diaa Salama Abdul.Elminaam, Abdulrhman M. Almansori, Mohamed Taha, Elsayed Badr (2020). A deep facial recognition system using computational intelligent algorithms, PLOS ONE, 15(12): e0242269.

[29] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.