Educational Stream Data Analysis: A Case Study

Conference Paper · June 2020

DOI: 10.1109/MELECON48756.2020.9140510

4 authors, including:

Gabriella Casalino
Università degli Studi di Bari Aldo Moro
62 PUBLICATIONS **361** CITATIONS

Giovanna Castellano
Università degli Studi di Bari Aldo Moro
270 PUBLICATIONS **2,287** CITATIONS

Gennaro Vessio
Università degli Studi di Bari Aldo Moro
76 PUBLICATIONS **368** CITATIONS

Educational Stream Data Analysis: A Case Study

Gabriella Casalino
Computer Science Dept.
University of Bari
Bari, Italy
gabriella.casalino@uniba.it
https://orcid.org/0000-0003-0713-2260

Giovanna Castellano
Computer Science Dept.
University of Bari
Bari, Italy
giovanna.castellano@uniba.it
https://orcid.org/0000-0002-6489-8628

Andrea Mannavola
Computer Science Dept.
University of Bari
Bari, Italy
a.mannavola@studenti.uniba.it

Gennaro Vessio
Computer Science Dept.
University of Bari
Bari, Italy
gennaro.vessio@uniba.it
https://orcid.org/0000-0002-0883-2691

***Abstract*—Virtual Learning Environments (VLEs) are Web-based platforms where educational contents, together with study support tools, are provided. Logs recording the interactions between students and VLEs are collected on a daily basis, thus automatic techniques are needed to manage and analyze such huge quantities of data. Students, teachers, managers, and in general all stakeholders involved in the VLEs' learning activities, can take advantage of the insights coming from educational data and useful information can be extracted by using machine learning techniques. Traditionally, educational data have been studied as stationary data by using conventional machine learning methods. However, educational data are non-stationary by nature and they can be better treated as data streams. In this paper, we show the results of a classification study where the random forest algorithm, applied both in batch and adaptive mode, is used to develop a model for predicting the failure/success of students' exams. Moreover, a feature importance analysis is carried out to detect the most discriminant attributes for the predictive task. Experiments were performed on the Open University Learning Analytics Dataset (OULAD) showing the reliability of adaptive random forest in creating accurate classification models from evolving educational data.**

***Index Terms*—machine learning, big data analytics, educational data mining, adaptive learning, stream data mining, online courses, OULAD**

I. Introduction

Distance education tools have grown drastically in the last years. They are Web-based platforms that are thought to remove physical distances between students and schools, thus enhancing the learning process. This has lead to a change in the form of transferring and processing information [1]. In this context, Virtual Learning Environments (VLEs) are interactive platforms where educational contents are provided. Students' activities and their interactions with the platform, such as reading, downloading or uploading material, taking tests, etc., are continuously monitored and collected through logs. These data form a big and valuable source of information about the students' learning and behaviors [2], which could be used to optimize the learning process and the whole environment [3]. All stakeholders involved in VLEs, including students, teachers, tutors, administrators and system designers, can take advantage of the information hidden in these data. Students could be interested in improving their grades. Teachers may be supported when evaluating the effectiveness of their courses, and for adapting them to the students' needs. Administrators can identify potential students "at risk" so as to limit student retention and increase graduation rates. Decisions regarding recruitment policies, course planning and hiring needs, could also be taken [4].

Automatic techniques are needed to manage and extract useful information from these data. To this aim, in the last years, a new research branch, called Educational Data Mining (EDM), has gained more and more attention. EDM refers to the use of machine learning algorithms to model the students' behaviors, so that predictive models can be used for monitoring, analysis, prediction, intervention, tutoring, customization, feedback, adaptation and recommendation purposes [4]. Several studies proved the effectiveness of this approach in the educational field [5], [6].

In [7] and [8], hidden learner skills are extracted from questionnaire results and used to group students in accordance with their preparation level. Process mining methodologies are used to evaluate coding behavior in software development processes [9]. Computer vision techniques are used for measuring and monitoring students' engagement [10]. Natural language processing and generation techniques are used to enhance the interaction between students and VLEs [11], [12]. In the educational field, where the interaction between humans and automatic systems is a main concern, systems that are able to communicate, in a human understandable way, are necessary [13], as well as visualization tools [14], [15]. In [16]–[19], learning analytic platforms aimed at supporting the big quantity of data coming from student-VLE interactions have been proposed. Finally, the main application of machine learning techniques to educational data is devoted to predict the students' outcomes [20]–[22].

However, traditional solutions do not exploit the intrinsic streaming nature of educational data: a continuous, possibly unlimited, flow of data that evolves during the time. Stream data mining algorithms are able to incrementally analyze this kind of data, by creating models that adapt their structure with the new coming data [23]. Moreover, these techniques are able to summarize previously seen data, so that they can process, in nearly real-time, only the actual data [24].

In this work, we propose to apply *adaptive* stream data analysis techniques to educational data for the purpose of predicting the student's outcome, based on their interaction with the VLE platform. Specifically, we used the adaptive random forest classification algorithm and we fairly compared its performance to those provided by the batch version of the algorithm. To this aim, the Open University Learning Analytics Dataset (OULAD) [25] has been used. Furthermore, a feature importance analysis has been carried out in order to detect the most discriminant attributes for the predictive task.

The rest of this paper is organized as follows. Section II details the OULAD dataset and the adopted methods, the outcomes of which are discussed in Section III. The last section draws some concluding remarks and outlines future work.

II. Materials and Methods

A. Data

For the present study, we used the Open University Learning Analytics Dataset [25]. The Open University (OU)[1] is one of the largest distance learning universities worldwide providing free open data related to the on-line courses.[2] It contains demographic data together with aggregated click-stream data of students' interactions in the Virtual Learning Environment, collected on a daily basis. The dataset is student oriented in the sense that the collected information are focused on students.

At the OU, courses can be accessed multiple times during the year; for this reason, the year and the semester (*J* for January and *B* for February) of the specific academic year are specified. In particular, data represent the students' interactions with the educational materials for the academic years 2013 and 2014. Seven different courses, four for STEM (Science, Technology, Engineering and Mathematics) and three for Social Sciences, are considered.

Since the goal of the analysis is to predict the students' outcome, a subset of the available data, focusing only on the student's information, has been considered. Thus, the dataset describes $25,819$ students through 20 attributes grouped in:

- Demographic data: gender, highest education, IMD band, age band and disability;
- Performance: code module, code presentation, number of previous attempts, studied credits and number of assessments;
- Learning behaviour: quiz, forum, glossary, homepage, out collaboration, out content, resource, subpage, url.

[1]Open University (OU) Website: http://www.open.ac.uk/

[2]OU Open Data: https://analyse.kmi.open.ac.uk/open_dataset#data

The target attribute, i.e. the final result, presents four different outcomes: *fail*, *pass*, *distinction* and *withdrawn*. Moreover, for further investigation, the classes *fail* and *withdrawn* have been merged into one single class named *fail*; likewise, the original classes *pass* and *distinction* have been merged into the *pass* class.

All the ordinal attributes have been converted into numerical values to be automatically processed.

B. Classification Models

Both batch and incremental classifiers have been considered to analyze the educational data. Specifically, random forest (RF) and adaptive random forest (ARF) have been used.

Random forest is a tree-based supervised classification algorithm which relies on the so-called *bagging* method to build a "forest" of decision trees at training time and outputting the mode of the classes predicted by the individual decision trees at test time [26]. Bagging consists in iteratively selecting a random sample with replacement from the training set and fitting a decision tree to this set. In contrast to ordinary bagging, RF does not consider the overall set of features to build a decision tree, but chooses a random subset of them. This serves to avoid growing highly correlated trees. As splitting criterion for the trees' construction, the popular Gini index or Gini impurity is typically used. This is a measure of the probability of a particular variable to be wrongly classified when it is randomly chosen. Its formula is: $Gini = \sum_{k=1}^{K} \hat{p}_{mk}(1-\hat{p}_{mk})$, where $k = 1, \dots, K$ are the different classes and $\hat{p}_{mk}$ is the proportion of examples labeled with class k at node m. RFs have proven very effective in a number of classification tasks (e.g., [27]–[30]).

Adaptive Random Forest is an adaption of the original formulation of the RF method [31]. It was conceived to extend the original formulation of the algorithm by including an effective re-sampling method in the streaming setting and adaptive operators that can cope with different types of concept drifts, without further hyper-parameter tuning. *Concept drift* refers to the change in the relationships between the input and output data over time. In particular, the adaptive strategy implemented by ARF is based on using a drift monitor per tree to track warnings and drifts, and to train new trees in the background. The training of these trees starts when a warning is detected and the active tree is replaced only when a drift occurs. One of the key advantage of the algorithm is that it was designed to be embarrassingly parallel, since there is no dependency between couples of trees.

C. Feature Selection

A feature importance analysis was conducted in order to identify the most discriminant features for the prediction task. Since different feature selection algorithms are likely to provide different selections of features, we employed three techniques and we evaluated their overlaps. Specifically, we used: recursive feature elimination (RFE); a ℓ_1-based method (from now onward L1) and a selection based on the RF algorithm itself (from now onward RFS).

RFE removes the least significant features over iterations [32]. The process is computed iteratively until all features are removed from the feature set, thus the final output is a ranked feature list. RFE uses criteria based on the performance of a linear support vector machine (SVM) [33] classifier to evaluate the importance of subsets of features.

Concerning the L1 method, it is worth noting that linear models, such as linear SVMs, penalized with a regularization term, such as the ℓ_1 norm, have sparse solutions, as many of their coefficients are zeroed or reduced to very small values. When the goal is feature selection, these models can be used to select the non-zero or the higher coefficients.

Finally, concerning the tree-based method, for each tree, the feature importance was calculated as the decrease of Gini impurity weighted by the expected fraction of the samples reaching that node. For the overall forest, the normalized feature importance were simply summed.

Overlapping the obtained feature rankings allows one to obtain feature sets that represent a more robust selection of the most important features describing the examples in the dataset.

III. Experimental Results

This section is devoted to discuss the obtained experimental results. Two different sets of experiments have been conducted.

First, we compared the prediction capabilities of both batch and incremental classifiers. To this aim, standard classification measures, i.e. accuracy, precision and recall, have been computed. Moreover, as previously detailed, two different prediction models have been developed for both the binary and the multi-class classification tasks. More precisely, the aggregated values *fail*/*pass* have been firstly considered as a final result; then, all the four possible outcomes *fail*/*pass*/*distinction*/*withdrawn* have been taken into account. Please note that, in the following, classes will be referred to as *F*, *P*, *D* and *W*, respectively, for the sake of brevity.

Then, a semantic analysis has been conducted in order to highlight the hidden relationships behind the data. To this end, a feature importance analysis has been carried out. Indeed, as previously mentioned, detecting the most relevant features could help in better designing courses and supporting students, to avoid their failures.

A. Classification Performance

The RF classifier was trained on the data coming from 2013 and it was then tested on data belonging to the following year, 2014. The aim of this analysis was to evaluate the system capability of predicting the students' outcome based on the stored history. Note that, in this case the algorithm needs all data to be stored. Instead, ARF was trained and tested incrementally. The four available semesters were used as temporal units. Each model was trained on the actual semester and it was tested on the subsequent one, i.e. 2013J as training set and 2013B as test set, 2013B and 2014J and finally 2014J and 2014B. In this way, four chunks, i.e. subsets of data, were considered (from #1 to #4), and at each time the classification model was trained on the current chunk and tested on the subsequent one. Data history does not need to be stored anymore, since it is "summarized" in the model parameters.

TABLE I
Statistics of the Binary Classification.

Test set	Model	*P*	*F*	Total
2014	*Static*	$8,811$	$6,119$	$14,930$
Chunk #2	*Adaptive*	$4,474$	$2,611$	$7,085$
Chunk #3	*Adaptive*	$3,358$	$2,635$	$5,993$
Chunk #4	*Adaptive*	$5,453$	$3,484$	$8,937$

TABLE II
Classification Performance of Random Forest for the Binary Classification.

	Accuracy = 90.5%	
Class	**Precision**	**Recall**
P	0.92	0.92
F	0.88	0.88

Binary classification: Batch and adaptive predictive models were generated by training RF and ARF classifiers on the given data. Table I summarizes the class distribution of the data. We can observe that RF is highly accurate, as it is able to obtain 90.5% of accuracy. It is quite good in recognizing the *pass* class, but it had some difficulties in discriminating the *fail* class, despite the high values of the overall precision and recall (around 90%), as reported in Table II. Table III reports the classification results obtained with ARF on the three chunks. It is worth noting that the accuracy values increase as new data arrive. This suggests that the classifier is able to adapt the previous model, in accordance with the new distribution and patterns in the data. Moreover, we can observe that when all chunks have been seen, the last model reaches an overall accuracy of 89.63%, which is very close to that obtained with the batch classifier. Thus, with only a small loss of accuracy, the system was able to learn the model from small portions of data, which is significant since they represent semantically relevant temporal slots such as semesters. This allows for monitoring the evolution of the students' outcomes over time. A deeper analysis of the results in Table III suggests that, with the fist chunk, the model confuses the classes *fail* and *pass* (low precision), but all the elements retrieved for the target class have been correctly labeled (high recall). As shown in Table I, concerning the first chunk, the *fail* class is under represented (*pass* students are two times the *fail* ones); thus, the model does not have enough data to learn patterns for the *fail* class. When the second chunk arrives, the model improves the recall for the *pass* class, but decreases it for the *fail* class: some students, in fact, belonging to the *fail* class are mistakenly assigned to the *pass* class. However, the model reaches high precision values for the *fail* class: almost all the students that have been predicted as *fail*, actually failed. Similar results are observed with the third chunk.

Multi-class classification: When considering the four classes, instead of the two aggregated one, the classification

TABLE III
CLASSIFICATION PERFORMANCE OF ADAPTIVE RANDOM FOREST FOR THE BINARY CLASSIFICATION.

	Chunk #2		**Chunk #3**		**Chunk #4**	
	Acc. = 65.59%		**Acc.** = 89.92%		**Acc.** = 89.63%	
Class	**Prec.**	**Rec.**	**Prec.**	**Rec.**	**Prec.**	**Rec.**
P	0.91	0.51	0.85	0.99	0.87	0.97
F	0.52	0.91	0.98	0.78	0.94	0.78

TABLE IV
STATISTICS OF THE MULTI-CLASS CLASSIFICATION.

Test set	**Model**	***P***	***F***	***D***	***W***	**Total**
2014	*Static*	$6,863$	$3,137$	$1,948$	$2,982$	$14,930$
Chunk #2	*Adaptive*	$3,725$	$15,94$	749	$1,017$	$7,085$
Chunk #3	*Adaptive*	$2,574$	$1,444$	784	$1,191$	$5,993$
Chunk #4	*Adaptive*	$4,289$	$1,693$	$1,164$	$1,791$	$8,937$

TABLE V
CLASSIFICATION PERFORMANCE OF RANDOM FOREST FOR THE MULTI-CLASS CLASSIFICATION.

	Accuracy = 64.57%	
Class	**Precision**	**Recall**
P	0.72	0.91
F	0.46	0.53
D	0.70	0.22
W	0.64	0.43

TABLE VI
CLASSIFICATION PERFORMANCE OF ADAPTIVE RANDOM FOREST FOR THE MULTI-CLASS CLASSIFICATION.

	Chunk #2		**Chunk #3**		**Chunk #4**	
	Acc. = 44.37%		**Acc.** = 59.39%		**Acc.** = 63.16%	
Class	**Prec.**	**Rec.**	**Prec.**	**Rec.**	**Prec.**	**Rec.**
P	0.08	0.13	0.61	1	0.67	0.98
F	0.26	0.45	0.50	0.43	0.42	0.39
D	0	0	0	0	0.6	0.26
W	0.35	0.31	0.64	0.32	0.73	0.43

performance decrease. This is caused by the different class distribution, as shown in Table IV. Indeed, students who passed the course are much more than those belonging to the other classes. Particularly, among these three classes, *fail* is still more represented than *distinction* and *withdrawn*. This class imbalance inevitably affects the classification results. Concerning the predictions provided by RF (Table V), there is a high recall for the *pass* class, i.e. the model is able to retrieve students belonging to this class. The other classes have low recall values, because the model is not able to find all the samples belonging to these classes. However, it is worth noting that the classes *distinction* and *withdrawn* have good precision values, i.e. the model found some patterns to discriminate them. How it could be expected, when considering chunks instead of the whole dataset, performance decrease. Anyhow, as shown in Table VI, as new chunks arrive the model adapts its structure and better results are obtained. Particularly, after the third chunk processing, the model performs almost as in the batch mode. Differently from this, adaptive models consider the evolution of time, without forgetting the history of data, and save memory, since they do not need to save old data. Most importantly, it is not likely that there is "omniscience" about the students' interactions and behaviours within the VLE, which is a typical requirement of non-streaming learning algorithms, thus adaptive models can provide a better support as online courses evolve during time.

B. Feature Importance

A feature importance analysis was conducted in order to identify the most discriminant features for the prediction task. As previously described, three different feature selection methods were used and three rankings of features were obtained (Table VII). The top ten most discriminant features were considered for each algorithm and the intersection of these sets are shown in Fig. 1. No significant differences were observed between the binary and the multi-class case, so only the first one is discussed.

It is worth noting that RFE selects as relevant demographic data and students' performance, while ignores students' interaction with the platform. The most discriminant features are code module (the studied subject) and code presentation (the semester in which the exam has been given). On the contrary, both L1 and RFS select feature referring to students' interaction. Particularly, a subset of them is in common between the two algorithms. These results can be used as feedback to strengthen useful activities, while dismiss unnecessary ones.

It is interesting to note that while the L1 feature selection selects only behavioural features, just after the number of studied credits, that is the most discriminant feature for the prediction task, RFS takes into account also the subjects and the semesters, and the most discriminant feature refers to the number of executed tests.

IV. CONCLUSION

In this paper, we have presented a case study on educational data mining, involving the application of adaptive random forest. The Open University Learning Analytics Dataset has been processed as a data stream to develop a classification model capable of predicting the students' outcomes. The algorithm has been able to adapt and evolve its internal parameters to new incoming data. The obtained results have shown the effectiveness of the method in correctly classifying students' outcomes by processing educational data as a stream. Moreover, the method compares favorably with its non-streaming version which cannot reliably used in a realistic setting. Finally, a semantic analysis has shown the most important features for the prediction task, returned by three different algorithms. Despite some overlap in the feature rankings, some substantial differences are present. Future work will better investigate the effectiveness of adaptive learning algorithms when dealing with huge collections of complex and heterogeneous educational data. Moreover, further semantic

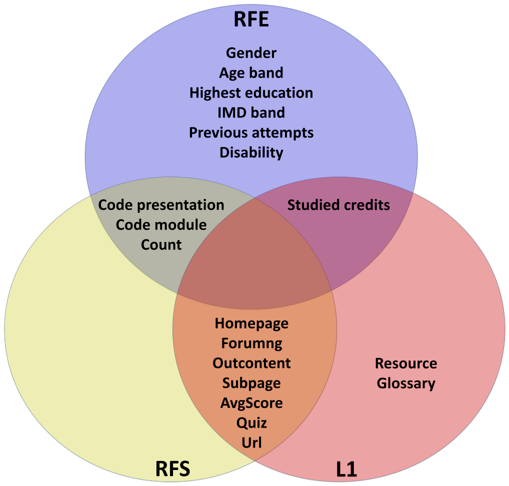

Fig. 1. Two class feature analysis.

TABLE VII
FEATURE RANKINGS FOR THE BINARY CLASSIFICATION.

RFE	L1	RFS
Code module	Studied credits	Count
Code presentation	Resource	AvgScore
Gender	Homepage	Outcontent
Highest education	Forumng	Code module
IMD band	Glossary	Homepage
Age band	Outcontent	Quiz
Previous attempts	Subpage	Forumng
Studied credits	Url	Subpage
Disability	Quiz	Url
Count	Avgscore	Code presentation

studies, regarding the influence of different feature categories on the prediction results, will be carried out.

ACKNOWLEDGMENT

Gabriella Casalino, Giovanna Castellano and Gennaro Vessio are members of the INdAM Research group GNCS.

REFERENCES

[1] M. C. Sáiz-Manzanares, C. I. G. Osorio, J. F. Díez-Pastor, and L. J. M. Antón, "Will personalized e-learning increase deep learning in higher education?" *Information Discovery and Delivery*, 2019.

[2] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, vol. 51, pp. 368–384, 2008.

[3] F. Sciarrone and M. Temperini, "Learning analytics models: A brief review," in *2019 23rd International Conference Information Visualisation (IV)*. IEEE, 2019, pp. 287–291.

[4] M. A. Chatti, V. Lukarov, H. Thüs, A. Muslim, A. M. F. Yousef, U. Wahid, C. Greven, A. Chakrabarti, and U. Schroeder, "Learning analytics: Challenges and future research directions," *eleed*, vol. 10, no. 1, 2014.

[5] P. S. Lopes, C. A. Silva, L. Valtensir, C. N. Targa, and D. B. Conrado, "A survey of recent developments in management cognitive neuroscience and machine learning in educational environment," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, 2018, pp. 1–4.

[6] A. Castro, L. Garcia, D. Prata, M. Lisboa, and M. Prata, "An exploratory study on data mining in education: Practiced algorithms and methods," *Int. J. of Information and Education Tech.*, vol. 7, no. 5, p. 319, 2017.

[7] G. Casalino, C. Castiello, N. Del Buono, F. Esposito, and C. Mencar, "Q-matrix extraction from real response data using nonnegative matrix factorizations," in *Int. Conference on Computational Science and Its Applications*. Springer, 2017, pp. 203–216.

[8] G. Castellano, A. Fanelli, and T. Roselli, "Mining categories of learners by a competitive neural network," in *IJCNN'01. Int. Joint Conference on Neural Networks. Proceedings*, vol. 2. IEEE, 2001, pp. 945–950.

[9] P. Ardimento, M. L. Bernardi, M. Cimitile, and F. M. Maggi, "Evaluating coding behavior in software development processes: A process mining approach," in *2019 IEEE/ACM International Conference on Software and System Processes (ICSSP)*. IEEE, 2019, pp. 84–93.

[10] B. De Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "Engaged Faces: Measuring and monitoring student engagement from face and gaze behavior," in *IEEE/WIC/ACM Int. Conf. on Web Intelligence-Companion*, 2019, pp. 80–85.

[11] M. Nakayama, S. Kikuchi, and H. Yamamoto, "Analysis of lexical features of online postings during a blended learning course," in *2019 18th Int. Conference on Information Technology Based Higher Education and Training (ITHET)*. IEEE, 2019, pp. 1–5.

[12] N. Di Blas, L. Lodi, P. Paolini, B. Pernici, N. Raspa, D. Rooein, and F. Renzi, "Sustainable chatbots supporting learning," in *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 2019, pp. 1358–1363.

[13] J. M. Alonso and G. Casalino, "Explainable artificial intelligence for human-centric data analysis in virtual learning environments," in *Int. Workshop on Higher Education Learning Methodologies and Technologies Online*. Springer, 2019, pp. 125–138.

[14] R. Paiva, I. I. Bittencourt, W. Lemos, A. Vinicius, and D. Dermeval, "Visualizing learning analytics and educational data mining outputs," in *Int. Conf. on Artificial Intelligence in Education*, 2018, pp. 251–256.

[15] A. F. D. Gonçalves, A. M. A. Maciel, and R. L. Rodrigues, "Development of a data mining education framework for visualization of data in distance learning environments," in *29th Int. Conf. on Software Eng. and Knowledge Eng., Pittsburgh, PA, USA*, 2017, pp. 547–550.

[16] N. Xu and B. Ruan, "An application of big data learning analysis based on MOOC platform," in *9th Int. Conf. on Information Technology in Medicine and Education (ITME)*. IEEE, 2018, pp. 698–702.

[17] C. Romero and S. Ventura, "Educational data science in massive open online courses," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 1, p. e1187, 2017.

[18] F. A. S. Herrera, R. G. Crespo, L. R. Baena, and D. Burgos, "A solution to manage the full life cycle of learning analytics in a learning management system: Analytic," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, vol. 14, no. 4, pp. 127–134, 2019.

[19] R. Pecori, V. Suraci, and P. Ducange, "Efficient computation of key performance indicators in a distance learning university," *Information Discovery and Delivery*, 2019.

[20] S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold, and L. Porter, "A robust machine learning technique to predict low-performing students," *ACM Transactions on Computing Education (TOCE)*, vol. 19, no. 3, pp. 1–19, 2019.

[21] C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes," *Computers & Education*, vol. 131, pp. 22–32, 2019.

[22] A. A. Saa, M. Al-Emran, and K. Shaalan, "Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques," *Technology, Knowledge and Learning*, vol. 24, no. 4, pp. 567–598, 2019.

[23] G. Casalino, G. Castellano, and C. Mencar, "Data stream classification by dynamic incremental semi-supervised fuzzy clustering," *International Journal on Artificial Intelligence Tools*, vol. 28, no. 08, p. 1960009, 2019.

[24] A. Abdullatif, F. Masulli, and S. Rovetta, "Clustering of nonstationary data streams: A survey of fuzzy partitional methods," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1258, 2018.

[25] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Scientific data*, vol. 4, p. 170171, 2017.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[27] M. Diaz, M. A. Ferrer, D. Impedovo, G. Pirlo, and G. Vessio, "Dynamically enhanced static handwriting representation for Parkinson's disease detection," *Pattern Recognition Letters*, vol. 128, pp. 204–210, 2019.

[28] D. Impedovo, G. Pirlo, G. Vessio, and M. T. Angelillo, "A handwriting-based protocol for assessing neurodegenerative dementia," *Cognitive Computation*, vol. 11, no. 4, pp. 576–586, 2019.
[29] M. T. Angelillo, F. Balducci, D. Impedovo, G. Pirlo, and G. Vessio, "Attentional pattern classification for automatic dementia detection," *IEEE Access*, vol. 7, pp. 57 706–57 716, 2019.
[30] T. Maggipinto, R. Bellotti, N. Amoroso, D. Diacono, G. Donvito, E. Lella, A. Monaco, M. A. Scelsi, and S. Tangaro, "DTI measurements for Alzheimer's classification," *Physics in Medicine & Biology*, vol. 62, no. 6, p. 2361, 2017.
[31] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, "Adaptive random forests for evolving data stream classification," *Machine Learning*, vol. 106, no. 9-10, pp. 1469–1495, 2017.
[32] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sensors and Actuators B: Chemical*, vol. 212, pp. 353–363, 2015.
[33] B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2001.