Advanced Techniques in the Analysis and Prediction of Students' Behaviour in Technology-Enhanced Learning Contexts

Juan A. Gómez-Pulido [1],* , Young Park [2] and Ricardo Soto [3]

1   Department of Technologies of Computers and Communications, University of Extremadura,
    10003 Cáceres, Spain
2   Department of Computer Science and Information Systems, Bradley University, Peoria, IL 61625, USA;
    young@fsmail.bradley.edu
3   School of Computer Engineering, Pontificia Universidad Católica de Valparaíso, 4059 Valparaiso, Chile;
    ricardo.soto@pucv.cl
*   Correspondence: jangomez@unex.es

Abstract: The development and promotion of teaching-enhanced learning tools in the academic field is leading to the collection of a large amount of data generated from the usual activity of students and teachers. The analysis of these data is an opportunity to improve many aspects of the learning process: recommendations of activities, dropout prediction, performance and knowledge analysis, resources optimization, etc. However, these improvements would not be possible without the application of computer science techniques that have demonstrated a high effectiveness for this purpose: data mining, big data, machine learning, deep learning, collaborative filtering, and recommender systems, among other fields related to intelligent systems. This Special Issue provides 17 papers that show advances in the analysis, prediction, and recommendation of applications propelled by artificial intelligence, big data, and machine learning in the teaching-enhanced learning context.

1. Introduction

Analysing and predicting individuals' behaviour are important topics in academic environments, especially after the increasing development and deployment of software tools for supporting learning stages. The automation of many processes involved in the usual students' activity allows for processing massive volumes of data collected from teaching-enhanced learning (TEL) platforms, leading to useful applications for academic personnel. In this way, monitoring and analysing students' behaviour are key activities required for the improvement of students' learning.

Recommendations of activities, dropout prediction, performance and knowledge analysis, and resources optimization, among other students-centred interests, are complex tasks that involve many elements that need to be considered. Therefore, it becomes necessary that these efforts search for support from other fields in the computational science that have demonstrated a high effectiveness when handling data and processes that are strongly interconnected. Data mining, big data, machine learning (ML), deep learning, collaborative filtering, and recommender systems, among other fields related to intelligent systems, allow for the development of advanced techniques

that provide a significant potential for the above purposes, leading to new applications and more effective approaches in the analysis and prediction of the students' behaviour in academic contexts.

This Special Issue provides a collection of papers of original advances in the analysis, prediction, and recommendation of applications propelled by artificial intelligence, big data, and machine learning, especially in the TEL context.

2. Summary of the Contributions in This Special Issue

Although each paper published in this Special Issue covers different topics, we can identify three groups where the papers can be classified according to their main focus: performance and behaviour prediction, dropout and risk prediction, and intelligent analysis of different learning aspects. However, some of these papers could be classified into more than one of these groups. Finally, a review article is also provided to get a wide perspective of this field.

With regard to performance and behaviour prediction, we find four contributions.

The first article in this group is entitled "Implementing AutoML in Educational Data Mining for Prediction Tasks" [1], by Maria Tsiakmaki, Georgios Kostopoulos, Sotiris Kotsiantis and Omiros Ragos. This research focuses on examining the potential use of advanced ML strategies on educational settings from the perspective of hyperparameter optimization, because of the complexity of ML when it is applied for a given problem formulation and it must be optimally configured. To this end, the authors analyze the effectiveness of automated ML (autoML) for the task of predicting students' learning outcomes based on their participation in online learning platforms, limiting the search space to tree-based and rule-based models. After carrying out many experiments, the performance of AutoML is verified. This proposal allows educators and instructors in educational data mining (EDM) to perform experiments with good parameter configurations, thus achieving highly accurate results.

The second article in this group is entitled "Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques" [2], by Raza Hasan, Sellappan Palaniappan, Salman Mahmood, Ali Abbas, Kamal Uddin Sarker, and Mian Usman Sattar. The authors have developed a system for predicting student's overall performance at the end of the semester using video learning analytics and data mining techniques. They consider video-based learning with flipped teaching to improve student's academic performance. Particularly, the authors applied eight classification algorithms (where random forest obtained the best results) to data collected from the student information system, learning management system and mobile applications. Additionally, they used genetic search, principle component analysis, rule inducer and multivariate projection to improve different aspects of the study.

The third article in this group is entitled "Towards Portability of Models for Predicting Students' Final Performance in University Courses Starting from Moodle Logs" [3], by Javier López-Zambrano, Juan A. Lara, and Cristóbal Romero. This work focuses on the data sources rather than the prediction techniques. Particularly, the work studies the portability of prediction models obtained directly from Moodle logs, according to grouping similar courses by degree or level of usage of activities or using numerical or categorical attributes. To this end, the authors apply a classification algorithm to the datasets in order to obtain decision tree models and test their portability to other courses by comparing the obtained accuracies. The authors conclude that the prediction models can be transferred to different courses under some circumstances.

The fourth article in this group is entitled "Prediction of High Capabilities in the Development of Kindergarten Children" [4], by Yenny Villuendas-Rey, Carmen F. Rey-Benguría, Oscar Camacho-Nieto, and Cornelio Yáñez-Márquez. This paper focuses on a type of the student's behaviour: the early detection of high capabilities, particularly at kindergartens, when the students are children. The prediction of such students is difficult, due to its low number and the focus of the teachers in the learning process. The authors propose a prediction algorithm based on Nearest Neighbor able to tackle this problem with satisfactory results.

With regard to dropout and risk prediction prediction, we find five contributions.

The first article in this group is entitled "Predicting Students Success in Blended Learning-Evaluating Different Interactions Inside Learning Management Systems" [5], by Luiz Antonio Buschetto Macarini, Cristian Cechinel, Matheus Francisco Batista Machado, Vinicius Faria Culmant Ramos, and Roberto Munoz. The authors apply ML techniques for detecting at-risk students earlier in courses involving algorithms and programming topics in undergraduate programs, where dropout and failure rates are usually high. This research finds the best combination of datasets collected from Moodle (considering cognitive, social and teaching presence) and classification algorithms. The best ML model was able to detect students at-risk in the first week of the course.

The second article in this group is entitled "The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction" [6], by Sunbok Lee and Jae Young Chung. This researc tries to identify students who are at risk of dropping out of school. To this end, the authors developed a dropout early warning system characterized by two features. On the one hand, the system addresses the class imbalance issue using the synthetic minority oversampling techniques (SMOTE) and the ensemble methods in ML. On the other hand, the system evaluates the trained classifiers with both receiver operating characteristic (ROC) and precision–recall (PR) curves. The authors trained random forest, boosted decision tree, random forest with SMOTE, and boosted decision tree with SMOTE by using large datasets. Among these ML techniques, boosted decision tree obtained the best results.

The third article in this group is entitled "A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course" [7], by Emanuel Marques Queiroga, João Ladislau Lopes, Kristofer Kappel, Marilton Aguiar, Ricardo Matsumura Araújo, Roberto Munoz, Rodolfo Villarroel, and Cristian Cechinel. This article continues in the line of the early prediction of at-risk students. Here, the authors consider students' interactions with the virtual learning environment as data source. With the goal of maximizing the prediction results, the authors apply an elitist genetic algorithm for tuning the hyperparameters of some classifiers: classic decision tree (DT), random forest (RF), multilayer perceptron (MLP), logistic regression (LG), and the meta-algorithm AdaBoost (ADA).

The fourth article in this group is entitled "An Early Warning System to Detect At-Risk Students in Online Higher Education" [8], by David Bañeres, M. Elena Rodríguez, Ana Elena Guerrero-Roldán, and Abdulkadir Karadeniz. The authors centered the research effort on finding accurate predictive models to identify at-risk students. The authors considered several classifiers, whose prediction quality was evaluated by a proposed method. Furthermore, they developed an early warning system tested in a real educational setting, which demonstrated accuracy and usefulness for detecting at-risk students in online higher education. This system shows different dashboards where students and teachers can analyze the information and perform some interventions to reduce at-risk situations.

The fifth article in this group is entitled "Automated Assessment and Microlearning Units as Predictors of At-Risk Students and Students' Outcomes in the Introductory Programming Courses" [9], by Jan Skalka and Martin Drlik. This work predicts at-risk students particularly in the context of introductory programming courses, where some students with limited programming skills can be discouraged by key aspects as the ability to think abstractly, solve problems, and design solutions. This work analyzed the automated source code assessment of assignments and the implementation of a set of microlearning units as predictors of at-risk students and students' outcomes. The authors found a significant contribution of automated code assessment in students' learning outcomes and proved a certain dependence between the students' activity and achievement in the activities and final students' outcomes.

With regard to intelligent analysis of different learning aspects, we find seven contributions.

The first article in this group is entitled "Predicting Students' Behavioral Intention to Use Open Source Software: A Combined View of the Technology Acceptance Model and Self-Determination Theory" [10], by F. José Racero, Salvador Bueno and M. Dolores Gallego. This work focuses on students' behavioral intention to continue using open source software (OSS) after be trained in it. This intention

is predicted by applying Self-Determination Theory and the technological acceptance model (TAM). The dataset was built collecting data from a survey. The results obtained by the model confirmed the influence of the intrinsic motivations, autonomy and relatedness, to improve perceptions with regard to the usefulness of OSS and, therefore, on the intention to continue considering OSS.

The second article in this group is entitled "A Multi-Analytical Approach to Predict the Determinants of Cloud Computing Adoption in Higher Education Institutions" [11], by Yousef A. M. Qasem, Shahla Asadi, Rusli Abdullah, Yusmadi Yah, Rodziah Atan, Mohammed A. Al-Sharafi, and Amr Abdullatif Yassin. This work predicts the key aspects that influence on the managers of higher education institutions for adopting cloud computing as services provider. To this end, a variance-based structural equation modeling (PLS-SEM) and an artificial neural network (ANN) were applied to data collected from 134 managers involved in the decision making of the institutions. The PLS-SEM approch was used for extracting the significant relationships among the identified factors, whereas ANN ranked the normalized importance among those factors. It is interesting to know that technology readiness is the most important predictor for cloud computing adoption, followed by security and competitive pressure. Furthermore, the authors present an innovative approach useful for decision-makers to develop stategies for adopting cloud computing services.

The third article in this group is entitled "Predicting Student Grades Based on Their Usage of LMS Moodle Using Petri Nets" [12], by Zoltán Balogh and Michal Kuchárik. In this paper, the data source is the popular learning management system (LMS) Moodle. This platform provides the information needed for analyzing the correlations between access to materials and the final grade in order to predict student's grades. According to the highest correlation, a model with Petri nets predicts what grade the student would get based on their usage of Moodle.

The fourth article in this group is entitled "The Relationship between the Facial Expression of People in University Campus and Host-City Variables" [13], by Hongxu Wei, Richard J. Hauer and Xuquan Zhai. The authors evaluate the public attitude towards university campuses and detect the relationship with host-city variables by using data about facial expression scores on social networks. It is interesting to know this attitude since it matters for the resource investment to sustainable science and technology. To this end, ML techniques area applied on datasets composed of 4327 selfies collected from social networks. The photos provide scores of happy and sad facial expressions and a positive response index was calculated. After analyzing some interesting results, the main conclusion is that people tend to show positive expression at campuses in cities with more education infrastructures but fewer residences and internet users.

The fifth article in this group is entitled "How to Extract Meaningful Insights from UGC: A Knowledge-Based Method Applied to Education" [14], by Jose Ramon Saura, Ana Reyes-Menendez, and Dag R. Benn. Students and teachers are a rich source of user generated content (UGC) on social networks and digital platforms. The vast amount of this type of data can supply useful knowledge by extracting and visualizing samples of readily available content, particularly the tweets published in Twitter. The authors apply latent dirichlet allocation (LDA) to identify topics, which are then subjected to sentiment analysis by using ML and a data visualization algorithm for complex networks. This research allows practitioners to improve short-term education strategies and interventions.

The sixth article in this group is entitled "Short CFD Simulation Activities in the Context of Fluid-Mechanical Learning in a Multidisciplinary Student Body" [15], by Manuel Rodríguez-Martín, Pablo Rodríguez-Gonzálvez, Alberto Sánchez-Patrocinio, and Javier Ramón Sánchez. The learning goal in this research was the instruction of students in fluid simulation tools in industrial engineering bachelors. These tools usually require long training times. Therefore, the authors propose a methodology based on short lessons, whose statistical results show a good acceptance in many terms. Furthermore, a ML technique was applied to find group peculiarities and differences among them in order to identify the need for further personalization of the learning activity.

The seventh article in this group is entitled "Technology-Enhanced Learning for Graduate Students: Exploring the Correlation of Media Richness and Creativity of Computer-Mediated

Communication and Face-to-Face Communication" [16], by Shan-Hui Chao, Jinzhang Jiang, Chia-Hsuan Hsu, Yi-Te Chiang, Eric Ng, and Wei-Ta Fang. This article explores and compares the differences in potential creative thinking that media richness had on learners in creativity training by considering computer-mediated communication and face-to-face communication. The authors found that the computer-mediated communication format shows better fluency, flexibility, and originality dimensions of creative thinking than the face-to-face format. Moreover, the computer-mediated format provides a greater level of media richness perception.

Finally, this special issue includes the review article entitled "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review" [17], by Juan L. Rastrollo-Guerrero, Juan A. Gómez-Pulido, and Arturo Durán-Domínguez. This article provides a wide perspective in the field of predicting students' performance. Many promising algorithms and methods focused on predicting students' performance have been investigated, hence the need to provide a detailed review. In this article, almost 70 papers were analyzed to show different techniques and objectives, mainly in the context of the artificial intelligence (AI).

References

1. Tsiakmaki, M.; Kostopoulos, G.; Kotsiantis, S.; Ragos, O. Implementing AutoML in Educational Data Mining for Prediction Tasks. *Appl. Sci.* **2020**, *10*, 90. [CrossRef]

2. Hasan, R.; Palaniappan, S.; Mahmood, S.; Abbas, A.; Sarker, K.U.; Sattar, M.U. Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques. *Appl. Sci.* **2020**, *10*, 3894. [CrossRef]

3. López-Zambrano, J.; Lara, J.A.; Romero, C. Towards Portability of Models for Predicting Students' Final Performance in University Courses Starting from Moodle Logs. *Appl. Sci.* **2020**, *10*, 354. [CrossRef]

4. Villuendas-Rey, Y.; Rey-Benguría, C.F.; Camacho-Nieto, O.; Yáñez-Márquez, C. Prediction of High Capabilities in the Development of Kindergarten Children. *Appl. Sci.* **2020**, *10*, 2710. [CrossRef]

5. Buschetto Macarini, L.A.; Cechinel, C.; Batista Machado, M.F.; Faria Culmant Ramos, V.; Munoz, R. Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems. *Appl. Sci.* **2019**, *9*, 5523. [CrossRef]

6. Lee, S.; Chung, J.Y. The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction. *Appl. Sci.* **2019**, *9*, 93. [CrossRef]

7. Queiroga, E.M.; Lopes, J.L.; Kappel, K.; Aguiar, M.; Araújo, R.M.; Munoz, R.; Villarroel, R.; Cechinel, C. A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course. *Appl. Sci.* **2020**, *10*, 3998. [CrossRef]

8. Bañeres, D.; Rodríguez, M.E.; Guerrero-Roldán, A.E.; Karadeniz, A. An Early Warning System to Detect At-Risk Students in Online Higher Education. *Appl. Sci.* **2020**, *10*, 4427. [CrossRef]

9. Skalka, J.; Drlik, M. Automated Assessment and Microlearning Units as Predictors of At-Risk Students and Students' Outcomes in the Introductory Programming Courses. *Appl. Sci.* **2020**, *10*, 4566. [CrossRef]

10. Racero, F.J.; Bueno, S.; Gallego, M.D. Predicting Students' Behavioral Intention to Use Open Source Software: A Combined View of the Technology Acceptance Model and Self-Determination Theory. *Appl. Sci.* **2020**, *10*, 2711. [CrossRef]

11. Qasem, Y.A.M.; Asadi, S.; Abdullah, R.; Yah, Y.; Atan, R.; Al-Sharafi, M.A.; Yassin, A.A. A Multi-Analytical Approach to Predict the Determinants of Cloud Computing Adoption in Higher Education Institutions. *Appl. Sci.* **2020**, *10*, 4905. [CrossRef]

12. Balogh, Z.; Kuchárik, M. Predicting Student Grades Based on Their Usage of LMS Moodle Using Petri Nets. *Appl. Sci.* **2019**, *9*, 4211. [CrossRef]

13. Wei, H.; Hauer, R.J.; Zhai, X. The Relationship between the Facial Expression of People in University Campus and Host-City Variables. *Appl. Sci.* **2020**, *10*, 1474. [CrossRef]

14. Saura, J.R.; Reyes-Menendez, A.; Bennett, D.R. How to Extract Meaningful Insights from UGC: A Knowledge-Based Method Applied to Education. *Appl. Sci.* **2019**, *9*, 4603. [CrossRef]

15. Rodríguez-Martín, M.; Rodríguez-Gonzálvez, P.; Sánchez-Patrocinio, A.; Sánchez, J.R. Short CFD Simulation Activities in the Context of Fluid-Mechanical Learning in a Multidisciplinary Student Body. *Appl. Sci.* **2019**, *9*, 4809. [CrossRef]

16. Chao, S.H.; Jiang, J.; Hsu, C.H.; Chiang, Y.T.; Ng, E.; Fang, W.T. Technology-Enhanced Learning for Graduate Students: Exploring the Correlation of Media Richness and Creativity of Computer-Mediated Communication and Face-to-Face Communication. *Appl. Sci.* **2020**, *10*, 1602. [CrossRef]

17. Rastrollo-Guerrero, J.L.; Gómez-Pulido, J.A.; Durán-Domínguez, A. Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Appl. Sci.* **2020**, *10*, 42. [CrossRef]