Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance

I. Introduction

PREDICTING student's learning performance in traditional face-to-face learning, online learning (LMS, MOOCs, etc.), and blended learning is a challenging but essential task in education [1]. On the one hand, it has become a difficult challenge due to the high number of factors that can influence a student's final status. On the other hand, it is a critical issue in education because it concerns many students of all levels (primary education, secondary education, and tertiary or higher education) and institutions over the entire world. Moreover, also, an increase in the number of low performing students can cause a lower graduation rate, an inferior institution reputation in the eyes of all involved, and it usually results in overall financial loss.

The task of predicting students' performance is one of the oldest and most studied tasks in Educational Data Mining (EDM) and Learning Analytics (LA), and a wide range of classification and regression approaches have been successfully applied [1]. Nowadays, there is a great interest in adapting these previous approaches and developing new ones for predicting as soon as possible what is the final status of the students in the course. In particular, the early identification of vulnerable students who are prone to fail or drop their courses is crucial for the success of any learning method [2]. To reduce the problem mentioned above, it is necessary to detect students who are at risk as early as possible and thus provide some care to intervene early to facilitate student success and to prevent these students from quitting or failing their studies. Decreasing the time lag between identification and real at-risk state, it may significantly reduce the risk of failure. Prevention is better than cure; likewise, it is better to prevent the student from failure, drop out, withdraw, disengage, wheel spinning or stop out than providing them with remedies later on. The sooner we know if a student is at risk, the better we can provide the right kind of help to the students.

It is important to realize that identifying students is only the first step in truly addressing this issue. The next step is to implement programs to provide effective and appropriate prevention strategies, to give advice or recommendations, and to carry out remedial or intervention actions [3]. For example, to identify the specific needs and problems of each student who is in danger of failure or dropping out, or to providing generic (more than specific) support to students at risk, to inform timely at-risk students by e-mail or using dashboards, etc. Therefore, stakeholders should be able to attend to students' needs to help them in time to avoid possible problems during all their learning process.

Digital Object Identifier 10.1109/TLT.2019.2908106

It is important to notice, that there are two important related terms with this special issue that we want to define:

Early Warning Systems (EWS) is any system that is designed to alert decision makers of potential dangers. Its purpose is to allow for the prevention of the problem before it becomes an actual danger [4]. In the educational domain, an academic EWS consists of a set of procedures and instruments for early detection of indicators of students at risk of dropping out and also involves the implementation of appropriate interventions to make them stay in school [5].

Response To Intervention (RTI) integrates assessment and intervention within a multi-level prevention system to maximize student achievement and to reduce behavioral problems. It uses data to identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness, and identify students with learning disabilities or other disabilities [6].

The objective of this special issue is to show recent trends and reports that discuss key successes, lessons learned, and challenges for predicting student's performance as early as possible and methods for providing appropriate intervention based on predictions.

II. Papers in the Special Issue

The response to our call for papers was 55 submissions. Of these, 40 were in scope of the special issue and went through several rounds of peer review. Finally, only 11 papers were accepted for publication in the journal. After reading these papers, we have detected three main topics or issues in which we can group their research: generalization, multi-view, and frameworks.

The first group of five papers deals with generalization problem, in which their objective is to obtain models or predictors that can be transferable to other different courses to the originals.

In the first paper "Improving Predictive Modeling for At-Risk Student Identification: A Multi-Stage Approach," J.L. Hung, B.E. Shelton, J. Yang, and X. Du propose a two-step approach by construction three predictive models (successful model, at-risk model, and coordination model) for improving prediction accuracy. They also identify two generalizable predictors from instruction-intensive and discussion-intensive courses.

In the second paper "Developing Early Detectors of Student Attrition and Wheel Spinning Using Deep Learning," A.F. Botelho, A. Varatharaj, T. Patikorn, D. Doherty, S.A. Adjei,

and J.E. Beck apply a transfer learning methodology using deep learning to study high and low representations of unproductive persistence. They focus on developing early detectors of student dropout and wheel spinning behaviors. The models developed to demonstrate that it is possible to automatically construct features that generalize and apply to new prediction tasks.

In the third paper "A Quest for a One-Size-Fits-All Neural Network: Early Prediction of Students at Risk in Online Courses," D. Monllao Olive, D.Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese propose to use deep learning neural networks to contextualize predictor variables and to improve the model's portability. The objective is to obtain predictive models that can generalize well to many different types of courses using data available in Moodle.

In the fourth paper "How Widely Can Prediction Models Be Generalized? An Analysis of Performance Prediction in Blended Courses," N. Gitinabard, Y. Xu, S. Heckman, T. Barnes, and C.F. Lynch propose a prediction model on student performance based upon persistent characteristics. They evaluate the performance of the models on different segments of the course to determine how early reliable predictions can be made and across courses to test how well they generalize.

The second group of four papers deals with multi-view problem, in which the data used for making predictions come from multi-sources or different data sources.

In the fifth paper "Interpretable Multiview Early Warning System Adapted to Underrepresented Student Populations," A. Cano and J.D. Leonard propose to detect students at risk of failure from multiple data repositories and to combine the information to make more accurate predictions. They use multi-view genetic programming for extracting interpretable classification rules and make it comprehensible to instructors which facilitates early intervention.

In the sixth paper "Multiview Learning for Early Prognosis of Academic Performance: A Case Study," G. Kostopoulos, S. Karlos, and S. Kotsiantis propose a co-training method for the early prediction of undergraduate student's success or failure in the final examinations of a distance learning course. They use two different sources: the tutor that provides attributes regarding students' characteristics and academic achievements and the online learning management system that provides attributes regarding student's online activities.

In the seventh paper "Predicting the Risk of Academic Dropout With Temporal Multi-Objective Optimization," F. Jiménez, A. Paoletti, G. Sánchez, and G. Sciavicco describe an innovative temporal multi-objective optimization model for identifying the earliest moment in a student's career in which a reliable prediction can be made. They predict the risk of dropout from the course of studies based on academic behavior alone ignoring classical used attributes.

In the eighth paper "Feature Extraction for Next-Term Prediction of Poor Student Performance," A. Polyzou and G. Karypis try to accurate identify students that are at risk before they even take a class by formulating the problem as binary classification. They try to gain insight to which are the factors that lead to poor performance by engineering some human-interpretable features that quantify these factors.

Finally, the last group of three papers deals about developing frameworks, early warning systems, and applying intervention strategies in education environments.

In the ninth paper "An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course," D. Baneres, M. E. Rodríguez, and M. Serra present an Early Warning Systems for early prediction in first-year higher education courses. The system can semi-automate the exploration of the predictive models, the identification of at-risk students, and the application of an intervention mechanism.

In the 10th paper "From Lab to Production: Lessons Learnt and Real-Life Challenges of an Early Student-Dropout Prevention System," A. Ortigosa, R.M. Carro, J. Bravo-Agapito, D. Lizcano, J.J. Alcolea, and O. Blanco describe a system to support dropout risk prediction and retention-oriented interventions in a fully online university. They discussed the main challenges faced on the way from the laboratory to the system exploration and beyond, as well as the lessons learned during the whole process.

The 11th and last paper "Pedagogical Intervention Practices: Improving Learning Engagement Based on Early Prediction," H. Wan, K. Liu, Q. Yu and X. Gao propose a transfer learning model of predicting the learner's performance based on their study behavior features. They transfer knowledge from one course to another by boosting a basic learner. They also show a general process of building an academical social media platform for applying teacher intervention strategies.

III. Summary

The papers in this special issue provide one view on the current state of the art in the early prediction and supporting of learning performance. They show us that the current research in this area is focused on:

- Generalizing early prediction models in order can be applied or transferred to other different courses.
- Selecting and evaluating what are the most important very early factors or indicators that affect to student's final status from different sources of data.
- Developing and testing Early Warning Systems (EWS) and Response To Intervention (RTI) in real education environment.
- Additionally, we have also detected some other important issues and future challenges in this area, such as:
- Preventing and monitoring students at risk on real-time.
- Providing personalized intervention and feedback to each students at risk.
- Evaluating what are the best remediation/intervention actions that could be done to address students at risk.

We hope these papers will offer an informative snapshot of the current state in early prediction and supporting of learning performance and can stimulate more productive research in the near future.

The guest editors of this special issue gratefully acknowledge the financial subsidy provided by the Spanish Ministry of Science and Technology TIN2017-83445-P.

CRISTÓBAL ROMERO, *Guest Editor*
Computer Science Department and Numerical Analysis
University of Cordoba, Cordoba 14071, Spain
(e-mail: cromero@uco.es)

SEBASTIAN VENTURA, *Guest Editor*
Computer Science Department and Numerical Analysis
University of Cordoba, Cordoba 14071, Spain
(e-mail: cromero@uco.es)

REFERENCES

[1] C. Romero and S. Ventura, "Data mining in education." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.

[2] C. Márquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. M. Fardoun, and S.Ventura, "Early dropout prediction using data mining: a case study with high school students," *Expert Syst.*, vol. 33, no. 1, pp. 107–124, 2016.

[3] A. L. Sønderlund, E. Hughes and J. Smith, "The efficacy of learning analytics interventions in higher education: A systematic review " *British J. Edu.Technol.*, pp. 1–25, 2018.

[4] V. F. Grasso, "Early warning systems: state-of-art analysis and future directions," *Draft Report United Nations Environment Programme* (UNEP), vol. 1, pp. 1–66, 2009.

[5] J. B. Heppen and S. B. Therriault, "Developing early warning systems to identify potential high school dropouts. Issue brief " *National High School Center*, American Institutes for Research, 1–13. 2008.

[6] D. Fuchs and L. S. Fuchs, "Introduction to response to intervention: What, why, and how valid is it?," *Reading Res. Quarterly*, vol. 41, no. 1, pp. 93–99, 2006.

Cristóbal Romero is a Full Professor with the University of Córdoba, Córdoba, Spain, and member of the Knowledge Discovery and Intelligent Systems research group. He has authored/coauthored more than 100 papers about Educational Data Mining (EDM) in international conferences and journals, from which 40 have been published in Thomson-Reuters Impact Factor journals, and some of them are important EDM surveys. He is also the co-editor of two books entitled *EDM: Data Mining in e-learning* (Wit Press, 2006) and *Handbook of Educational Data Mining* (Chapman & Hall/CRC, 2010). He is currently an Associate Editor of the IEEE-TLT journal, a member of the International EDM Society and on the program committee of a great number of conferences about education, personalization, and data mining.

Sebastian Ventura received the B.Sc. and Ph.D. degrees in sciences from the University of Cordoba, Cordoba, Spain, in 1989 and 1996, respectively. He is currently a Full Professor with the Department of Computer Science and Numerical Analysis, University of Cordoba, where he heads the Knowledge Discovery and Intelligent Systems Research Laboratory. He has authored/coauthored more than 300 papers in journals and scientific conferences, and he has edited three books and several special issues in international journals. He has also been engaged in 12 research projects (being the coordinator of five of them) supported by the Spanish and Andalusian governments and the European Union. His research interests include machine learning, data mining, and their applications. He is a senior member of the IEEE Computer, the IEEE Computational Intelligence and the IEEE Systems, Man and Cybernetics Societies, as well as the Association of Computing Machinery.