Predictors and Early Warning Systems in Higher Education — A Systematic Literature Review

Martín Liz-Domínguez, Manuel Caeiro-Rodríguez, Martín Llamas-Nistal, and Fernando Mikic-Fonte

University of Vigo, Spain
{mliz,mcaeiro,martin,mikic}@gist.uvigo.es

Abstract. The topic of predictive algorithms is often regarded among the most relevant fields of study within the data analytics discipline. Nowadays, these algorithms are widely used by entrepreneurs and researchers alike, having practical applications in a broad variety of contexts, such as in finance, marketing or healthcare. One of such contexts is the educational field, where the development and implementation of learning technologies led to the birth and popularization of computer-based and blended learning. Consequently, student-related data has become easier to collect. This Research Full Paper presents a literature review on predictive algorithms applied to higher education contexts, with special attention to early warning systems (EWS): tools that are typically used to analyze future risks such as a student failing or dropping a course, and that are able to send alerts to instructors or students themselves before these events can happen. Results of using predictors and EWS in real academic scenarios are also highlighted.

Keywords: Predictive analytics · Early warning systems · Learning analytics · Learning technologies.

CEUR-WS.org/Vol-2415/paper08.pdf

1 Introduction

1.1 Context

Over the last couple decades, the meteoric rise of information technologies (IT) has caused deep social and economic transformations worldwide, leading to the growth of new disciplines and activities which are of utmost importance today. Among these disciplines is data analytics, which is currently a huge source of income for many companies — especially, but not exclusively, those in the IT field —, as well as a very relevant topic for researchers.

Data analytics encompasses the collection of techniques that are used to examine data of a variety of types to reveal hidden patterns, unknown correlations and, in general, obtain new knowledge [18]. The discipline is often coupled with the term "big data", since analysis tasks are often performed over huge data sets. Other fields of study which are very popular nowadays, such as data mining or machine learning, are close to data analytics and share many relevant techniques.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

Depending on the nature of the data that is being analyzed and the objective that the analysis task should fulfill, several sub-disciplines can be defined under data analytics. Examples of these are text analytics, audio analytics, video analytics and social media analytics. The main focus in this paper, however, will be *predictive analytics*, which includes the variety of techniques that make predictions of future outcomes relying on historical and present data [13].

The capability of predicting future events is essential for the proper functioning of some applications. Notable examples among these are early warning systems (EWS), which are capable of anticipating potential risks in the future thanks to present information, accordingly sending alerts to the person or group of people who may be affected by these risks and/or that are capable of countering them. Their degree of reliability on information technologies greatly varies depending on the context they are applied on.

Early warning systems are mostly known for their use to reduce the impact of natural disasters, such as earthquakes, floods and hurricanes. Upon detection of signs that a catastrophe might happen in the near future, members of the potentially affected population are alerted and given instructions to prevent or minimize damage [24]. However, other kinds of EWS have been implemented in a variety of different contexts. For instance, they are used in financial environments to predict economic downturns at an early stage and provide better opportunities to mitigate their negative effects [10]. In healthcare, early warning systems are used by hospital care teams to recognize the early signs of clinical deterioration, enabling the initiation of early intervention and management [27].

1.2 Research objectives

This document will explore the reported uses of predictive algorithms and early warning systems in the educational context, focusing on higher education environments, most notably university courses. This scenario falls under the umbrella of learning analytics (LA), a particularization of data analytics which is usually defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" [26].

The study is presented as a systematic literature review, following the general guidelines established by Kitchenham and Charters [19], attempting to provide an answer to the following research questions:

- *RQ1: What are the most important purposes of using predictive algorithms in higher education, and how are they implemented?*
- *RQ2: Which are the most notable examples of early warning systems applied in higher education scenarios?*

Following this introductory section, this report explains the literature search process and the criteria that was followed to assess the relevance of analyzed documents. Next, the contents of the most relevant papers are summarized, addressing the research questions proposed above. Finally, some insights and discussion are presented at the end of this document.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

2 Document retrieval

2.1 Search

The search process consisted in the retrieval of relevant documents available in online libraries and repositories. The selected sources were:

- IEEE Xplore Digital Library.
- ACM Digital Library.
- Elsevier (ScienceDirect).
- Wiley Online Library.
- Springer (SpringerLink).
- Google Scholar.

The following query string was run in each one of these platforms:

```
("early warning system" OR "predictive analysis"
        OR "predictive analytics"
        OR "predictive algorithm")
        "education" "university"
        -"disaster" -"medical" -"health"
```

The purpose of this query was to obtain documents related to the use of EWS and predictive algorithms in university contexts, while disregarding unrelated applications in the fields of natural disaster prediction and healthcare technologies — uses so common that they have entire journals dedicated to them. Additionally, publication dates were restricted to 2012 or newer, and only journal articles, conference proceedings and book extracts were considered.

Table 1 summarizes the results of the search procedure. Notice that due to Google Scholar's nature as an indexer of many different sources, some overlapping results with the rest of the libraries are expected. This search engine was included in order to obtain potentially relevant papers which are not available in any of the other digital libraries.

Table 1. Summary of the document search process.

Library	Search results	Selected documents
IEEE Xplore Digital Library	36	5
ACM Digital Library	45	6
Elsevier (ScienceDirect)	412	6
Wiley Online Library	255	0
Springer (SpringerLink)	91[1]	6
Google Scholar	$\sim$ 13800[2]	2

[1] Search limited to the "Education" discipline.
[2] Only the first 200 returned documents, according to Google Scholar's order of relevance, were considered.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

2.2 Rating

In order to rate the retrieved documents according to their relevance, the criteria listed in Table 2 are defined. For the sake of this study, it is considered important that the paper presents a predictor or EWS which is useful for higher education scenarios, that the inner workings of their algorithms are clearly explained, and that the system has been tested and results exist.

Table 2. Rating criteria for the considered documents.

Criterion	Very relevant	Relevant	Not relevant
Experiment relevance	The presented predictive algorithm or EWS is very relevant for higher education scenarios.	The usefulness of the predictive algorithm or EWS in a higher education scenario is limited.	The paper does not present a predictive algorithm or EWS applied to higher education.
Analysis description	The data analysis process in its entirety is thoroughly documented.	A fair explanation of the data analysis process is provided, although with missing information or technicalities.	The document does not give any details about the data analysis process or the algorithms that were used.
Testing and results	The predictive algorithm or EWS is tested in real scenarios with solid results.	The predictive algorithm or EWS is tested in limited scenarios. Results may not be fully convincing.	No tests are performed.

The document rating and selection process was carried out in three steps. First, papers were filtered by reading titles and abstracts, discarding those unrelated to the educational field. Next, introductions and conclusions were analyzed in order to confirm that the documents address the points that were established as rating criteria. The resulting document list was more thoroughly analyzed, disregarding papers that fail to achieve at least a "relevant" rating in any of the three aspects considered for evaluation.

After completing the rating process, a narrower list including the most relevant papers is obtained. Table 1 indicates the amount of documents per source that satisfactorily meet the established criteria.

As previously stated, predictive analytics and EWS are extremely popular disciplines with applications in many different knowledge fields, which makes efficiently filtering search results an arduous task. This explains the fact that the amount of selected documents is relatively low compared to the quantity of yielded results. This is particularly true for the Elsevier and Wiley repositories — in the latter case, not even one document was found to be relevant for the educational context. It is also worth mentioning that most of the relevant papers

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

returned by the Google Scholar searcher had already been selected from one of the other online libraries, and only unique articles are reflected as selected in the table.

3 Content review

3.1 Overview

Tables 3 and 4 briefly showcase the most important characteristics of the covered predictive models and EWS, respectively, for the sake of comparison. More detailed descriptions of each one of the documents' contents are provided in following subsections.

Table 3. Defining characteristics of the selected predictive models.

Document	Year	Input data	Prediction goal	Key aspects
Ornelas [22]	2017	Engagement and performance indicators.	Course success/failure.	Applied in 11 different courses.
Thompson [28]	2018	Reasoning and math tests prior to the course.	Course success/failure.	Very early risk estimation.
Benablo [6]	2018	Engagement indicators.	Course success/failure.	Analyzes procrastination.
Umer [30]	2018	Test results and LMS log data.	Letter grades (five-point system).	Applied in continuous assessment.
Kostopoulos [20]	2019	Demographics, academic achievements, LMS data.	Course success/failure	Uses co-training to improve accuracy.
Hirose [15]	2018	Results of weekly tests.	Course success/failure.	Uses item response theory.
Schuck [25]	2017	Demographics, performance, violence indicators.	Graduation rate.	Measures the influence of crime and violence.
Tsiakmaki [29]	2018	Final course scores.	Grades from upcoming subjects.	Predicts outcomes of future courses.
Adekitan [1]	2019	GPA from first three years.	Final cumulative GPA.	Predicts total GPA of a five-year degree program.
Jovanovic [17]	2019	Interactions with pre-class material.	Final course grades.	Applied in a flipped classroom setting.
Chen [11]	2013	Quality of students' notes	Final course grades.	Using note-taking as input data.
Amirkhan [4]	2018	Stress level surveys.	Final course GPA.	Impact of stress overload.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

Table 4. Defining characteristics of the selected EWS.

Document	Year	Input data	Prediction goal	Key aspects
Arnold [5]	2012	Demographics, performance, academic history.	Level of risk (three-point scale).	Course Signals EWS. Well-established and widely tested.
Krumm [21]	2014	LMS effort and performance data.	Level of risk (three-point scale).	Student Explorer EWS. Used in several studies.
Waddington [31]	2014	LMS resource use.	Final course grade.	Improvement upon Student Explorer.
Brown [7]	2016	Student Explorer data.	Risk level changes.	Study using the Student Explorer EWS.
Brown [8]	2017	Student Explorer data.	Best measures to help struggling students.	Study using the Student Explorer EWS.
Brown [9]	2018	Student Explorer data.	Influence of co-enrollment.	Study using the Student Explorer EWS.
Gutiérrez [14]	2018	Grades, data from enrolled courses.	Risk of failing the course.	LADA EWS. Supports decision-making of advisors.
Akhtar [3]	2017	Attendance, neighbors, location within the laboratory.	Risk of failing the course.	SurreyConnect EWS. Targets laboratory sessions.
Howard [16]	2018	Demographics, intermediate task results.	Final course grades.	Tries to find the optimal time to apply an EWS in continuous assessment.
Wang [32]	2018	Grades, attendance, engagement, library and dorm records.	Level of risk for several different events.	Incorporates data on students' life habits.
Cohen [12]	2017	LMS log data.	Dropout risk.	Detecting student inactivity in order to predict dropout.
Akçapınar [2]	2019	E-book management system data.	Risk of failing the course.	Input data related to e-book interaction.
Plak [23]	2019	Student demographics and performance.	Risk of failing the course.	Using the EWS does not lead to dropout reduction.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

3.2 Predictive analysis in education

This section addresses RQ1 by summarizing the contents of papers related to the topic of predictive analytics in higher education. As will be shown, student success, performance and grades stand out as the most popular prediction objectives. Within this category, two different approaches can be identified: success predictors, which try to estimate whether a student will pass or fail a course; and grade predictors, which attempt to anticipate the final grade of a student. Unique traits in each study include the nature of input data, the data processing algorithms that are used and the scenarios in which they are tested.

Success prediction. These applications are mostly based on classifier algorithms.

Ornelas and Ordonez [22] proposed a Naive Bayesian classifier which was applied in a dozen courses taught at Rio Salado Community College (Arizona, USA). They used data from the institution's LMS as input, divided into two categories: engagement indicators (LMS logins and participation in online activities) and performance (points earned in course tasks). The classifier was able to predict success — that is, the student getting a C grade or better — with an accuracy of over 90% for eleven different courses, although not early enough so that it could properly work as an early warning system. This experiment was applied to a fairly big population, with a training sample of 5936 students and a validation sample of 2722.

Thompson et al. [28] used logistic regression to estimate the chances of student success in an introductory biology course taught during the first semester of a university major program, with a total of 413 enrolled students. As opposed to the previous case, they exclusively used results from tests with no direct relationship with the course, which were taken right at the beginning of the semester. These were Lawson's Classroom Test of Scientific Reasoning and the ACT Mathematics Test. Although this was not a perfect model to predict success by any means, it provided a first estimation of students at risk before the course had even started.

Benablo et al. [6] introduced procrastination into the picture by surveying students on the time that they spend using social networks and playing online games. A SVM classifier was able to successfully identify underperforming students: 100% precision and 96.7% recall on a 100-instance data set.

Umer et al. [30] tried to estimate the earliest possible time within a course at which a reliable identification of students at risk could be made. The targeted course, an Australian introductory mathematics module with 99 enrolled students, used the continuous assessment system, in which multiple assignments are performed throughout the duration of the course, instead of just a final exam. The input data were a combination of assignment results and LMS log data. Students were classified regarding their final performance estimation — grades A, B, C, D, as well as failing or dropping the course. After one week, a Random Forest classifier was able to identify students at risk with 70% accuracy.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

This percentage increased to 87% after five weeks, a point at which students had already completed 2 out of 7 total assignments.

Kostopoulos et al. [20] tried to improve the performance of traditional student success classifiers by implementing a co-training method. This technique consists on splitting the available data features into two independent and sufficient views. It is particularly useful when the amount of unlabeled data is large compared to the number of labeled examples, since it allows to expand the labeled set by adding initially unlabeled entries that both views can classify with high certainty. This study targeted an introductory informatics module from a Greek open university, involving 1073 students. The feature split that was performed created one view containing students' demographic characteristics and academic achievements provided by their tutors, and a second view including LMS activity data. The co-training method was observed to outperform traditional classifiers such as Naive Bayes, k-NN and Random Forest, providing very accurate identifications of poor performers towards the middle of the course.

Hirose [15] attempted to make an estimation of students' abilities using item response theory (IRT). The study was performed in the context of introductory mathematics courses under the continuous assessment system, in which students needed to answer a set of multiple-choice questions each week. Data from around 1100 students was available for this test. Thanks to IRT, question difficulty was assessed together with students' abilities, resulting in a more fair judgment. At specific times during the course, students were classified into "successful" or "not successful" using the Nearest Neighbor method. After seven weeks, roughly half the course, a misclassification rate as low as 18% was achieved; however, the number of false positives was noticeably high, meaning that many well-performing students were identified as being at risk of failing.

Schuck [25] presented a unique study which tried to establish a correlation between the level of crime and violence around campus and student success. The experiment was possible thanks to data provided by university representatives of the US Department of Education, as well as the United States' National Center for Education Statistics. Overall, complete data from 1281 higher education institutions was available. The study used multivariate regression models in order to predict graduation rate, that is, the fraction of students who finish their degree within the intended number of years. Input data for this model included the amount of violent incidents and disciplinary measures per number of students, the percent of disciplinary actions that ended up with arrests, as well as student demographic information and school characteristics. As a result of analysis, rates of violence were observed to negatively affect graduation years, as opposed to the rate of disciplinary measures, which is a positive indicator. Additionally, use of the student conduct system was observed to be better than criminal justice system for minor offenses.

Grade prediction. These applications are mostly built upon regression-based estimators.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

There are several examples of grade predictors which take students' previous results as their main source of input data. Tsiakmaki et al. [29] used final scores from courses imparted during the first semester of a Business Administration degree (592 students) in order to predict grades from second semester subjects, obtaining fair results using Random Forest and SVM algorithms. On the other hand, Adekitan and Salau [1] ran an experiment in a Nigerian engineering school trying to determine how well the grade point average (GPA) over the first three years of a degree could predict the final, cumulative GPA over the entire five-year program. Out of the tested analysis algorithms, logistic regression yielded the best result, with a 89.15% accuracy over a 1841 student sample.

Jovanovic et al. [17] proposed a predictive model to be applied in courses following the flipped classroom teaching method, focusing on student interaction with pre-class learning activities. These activities included videos and documents with multiple choice questions, as well as problem sequences. The model was tested in a first-year engineering course at an Australian university for three consecutive years, with a number of students ranging between 290 and 486. The study concluded that indicators of regularity and engagement related to pre-class activities had significantly superior predictive power than generic indicators such as the frequency of LMS logins.

Chen [11] assessed the quality and quantity of students' note-taking, both in and after class, to explore the effects that this could have on academic performance. A population of 38 freshmen students from a Taiwanese university participated in the experiment. Students' notes were retrieved and copied after each lecture by the professor, who rated their quality based on accuracy and completeness regarding the contents of the lecture. The word count, in this case number of Chinese characters, was also recorded. The studio concluded that only the quality of the notes taken during the class was a significant predictor of the students' final grade.

Amirkhan and Kofman [4] studied the effects of stress overload — the destructive form of stress — over the GPA obtained by students. The experiment was conducted over two consecutive semesters with a population of 600 freshmen students, who were surveyed mid-semester in order to assess their levels of stress. As a result of predictive analysis, stress was found to be among the strongest performance predictors, having significant and negative relationship with final GPA. However, it did not seem to have a direct relationship with dropout rate.

3.3 Early warning systems in education

This section addresses RQ2 by showing some of the most important EWS that were found in the literature. The applications listed below have the objective of identifying certain risks that students may be exposed to, and do it as soon as possible in order to take proper corrective measures in time. The most common risks to identify are high chances of a student failing or dropping out of a course or degree.

Arnold and Pistilli [5] present Course Signals, an EWS first implemented at Purdue University which has become one of the most popular and referenced by

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

the research community. This tool works in conjunction with the LMS Blackboard Vista, using data related to student demographics, performance, effort and prior academic history. Thanks to an on-demand student success algorithm, instructors can obtain an estimation of the risk level of a student, color coded as green, yellow and red for increasing degrees of risk. The application then allows the instructor to take measures if required, such as sending a message to the student or scheduling a face-to-face meeting. Course Signals has been employed in many courses at Purdue since 2007, registering a significant improvement in student grades, as well as a decrease in dropout rate. As opposed to most other EWS, which are not past their experimental stage, Course Signals is a mature and well-established application with proven positive results throughout the last decade.

Another well-known EWS is Student Explorer. As described by Krumm et al. [21], and similarly to Course Signals, Student Explorer mines effort and performance data from the institutional LMS in order to assess the likelihood of a student's academic failure. Students are classified with the labels "encourage", "explore" and "engage", in increasing order of risk, and student advisors can use this information to take corrective action.

Multiple other papers presented further experiments and improvements over the base Student Explorer application. Waddington and Nam [31] incorporated LMS resource use as input data, including information such as access to lecture notes or completion of assignments. Analysis using logistic regression determined a direct correlation between resource use and final grade, with activities related to exam preparation having the strongest positive relationship with performance. Brown et al. performed several studies revolving around Student Explorer. They observed that students had a greater chance of entering the "explore" category if they were in large classes, sophomore level courses and courses belonging to pure scientific degrees; while underperformance was still the most significant reason students entered the "engage" category [7]. They also used the tool to investigate how to best help struggling students recover. They concluded that students with moderate difficulties benefited the most from assistance planning their study behaviors, while those with severe difficulties benefited from better exam preparation [8]. Finally, these authors studied the effect of co-enrollment in multiple courses over performance, establishing a correlation between being enrolled in at least one "difficult course" and a higher chance to experience academic struggles [9].

Gutiérrez et al. [14] were the developers of LADA, a Learning Analytics Dashboard for Advisors. As its name implies, the main goal of this tool is to support the decision-making process of academic advisors. LADA incorporates a predictive module that estimates the students' odds of success by means of multilevel clustering. Input data includes student grades, courses booked by a student and the number of credits per course. The risk level of a student is calculated by comparing to other students with similar profiles from previous cohorts. LADA was deployed in two different universities, and student advisors

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

claimed that the biggest advantage that it provides is being able to analyze a greater amount of scenarios within a given time frame.

Akhtar et al. [3] created SurreyConnect, a teaching assistant with the objective of supporting computer-aided design (CAD) courses at University of Surrey (England). Most of the utilities of this tool were useful for laboratory sessions, allowing the instructor to share her or his computer screen with students, broadcast the screen of a specific student to the rest of the class or remotely connect to a student's computer in order to provide help. SurreyConnect also implements an analytics module with the purpose of identifying students at risk of failing the course. In order to do this, the application passively collects data during lab sessions regarding student attendance, location and neighbors within the lab, and time spent in class and doing exercises. A sample of 331 undergraduate students was selected to assess the usefulness of this feature, running an ANOVA test to identify the statistical significance of the input data, as well as applying Pearson correlation to identify the independent variables that influence final outcomes. Class attendance and time spent on tasks were shown to have a direct connection with learning outcomes, while student positioning in the classroom and sitting with a particular group of students also impacted performance.

Howard et al. [16] tried to find out the optimal time to apply an EWS in a course using the continuous assessment system. The study targeted a Practical Statistics course at University College Dublin (UCD) with 136 participant students, in which 40% of the final grade was awarded for completing certain tasks that were assigned each week throughout the course. The results of these tasks, as well as student demographic information and the number of times they accessed course resources, were collected as input for grade prediction. The data source was the institution's LMS, Blackboard. After testing multiple predictive models, Bayesian Additive Regressive Trees (BART) yielded the best results, being able to predict students' final grade with a mean absolute error of 6.5% as early as at week 6, exactly halfway through the course. This provides a decently precise prediction for the teacher, early enough so that corrective measures can be taken.

Wang et al. [32] are the designers of an EWS applied in Hangzhou Normal University (China), with the goal of reducing student dropout and minimizing delays in graduation. This application stands out because it includes types of input data that are not seen anywhere else: besides information regarding students grades, attendance and engagement, it also includes records from the university library and dorm. These extra data enable a closer monitoring of study habits. The EWS assigns students different labels depending on the kind of risks they are exposed to, such as the risk of obtaining low grades, graduation delay or dropout. After three semesters, and using a sample of 1712 students, a Naive Bayes algorithm was able to perform risk classification with an accuracy of 86%, with grades and library borrowing data being among the key indicators.

Cohen [12] focused on quantitative analysis of student activity data in order to provide an early identification of learner dropout. The study hypothesized that students who drop out of the course will first become inactive in course

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

websites. The proposed EWS collects student activity data from the Moodle LMS, including number and types of actions performed, as well as their timing and frequency. For the reported test, data from 362 students was collected. The input data were analyzed in a monthly basis in order to detect significant activity drops by a student, who would be subsequently flagged as at risk. The study concluded that two thirds of flagged students would indeed end up dropping their courses or degrees.

Akçapınar et al. [2] built an EWS intended to be used in courses that use e-books as learning material, using reading data in order to identify students at risk of academic failure. The data were collected from an e-book management system named BookRoll, used in several Asian universities and which students utilize in order to access course materials. This particularly study obtained information from 90 students registered in an Elementary Informatics course, registering their interactions with BookRoll, for instance, e-book navigation, page highlighting or note taking. Each week, analysis was performed to label students as low or high performing, trying multiple prediction algorithms. It was observed that an accuracy of 79% was achieved just with data from the first three weeks. Random Forest performed the best with raw data, however, Naive Bayes became the best performing when transforming the input into categorical data.

Finally, Plak et al. [23] document a case in which the use of an EWS does not provide the expected benefits. This experiment, conducted at Vrije Unversiteit in Amsterdam, provided student counselors with an analytics monitor that allowed them to identify low-performing students. The EWS used data related to student progress and demographics. However, the introduction of the tool did not lead to a reduction in dropout or an increase in obtained credits. While early identification of at-risk students is useful, the underlying problem that causes poor performance is ultimately undetermined.

4 Conclusion

Within the enormous world that is data analytics, the learning analytics field could be seen as just a small niche, mostly covered by academic research. However, a closer look into the discipline reveals that learning analytics is an extensive subject in its own right. The literature review presented in this document revealed that there exists a considerable variety of studies and applications revolving around predictive algorithms in the educational field, which is itself an important subject of study within learning analytics.

As an answer to RQ1, it was observed that most of the tools and predictive algorithms that were presented shared similar goals: most commonly, predict student grades, assess their chance of failure or their risk of dropping off a course or degree. Nevertheless, the great variety of educational contexts meant that each study had a unique approach in order to achieve said goals, meaning that specific implementation aspects greatly differ from case to case. Naturally, the availability of certain types of data, such as those related to student engagement and performance, is the factor that influences the analysis method the most.

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

However, many other aspects need to be taken into account in order to design a good predictor. Examples of elements that can significantly affect the analytics process are teaching strategy — such as the flipped classroom [17] — , assessment method — such as continuous assessment [16, 30] — , geographical context, student demographics or the year within a degree program. Thus, there is not a single predictive algorithm that can be considered better than the rest in all possible scenarios.

As for RQ2, this document covered some of the most important instances of EWS in education. Course Signals [5] and Student Explorer [21] are worthy of a special mention. The former has been applied in practical scenarios for over a decade and is one of the most referenced in the literature, while the latter has a pivotal role in several learning analytics experiments.

In general, the introduction of predictors and EWS in educational environments has been helpful in order to optimize the learning process and improve student performance. However, as of the year 2019, they have mostly been used in experimental environments only, with the notable exception being the Course Signals EWS. Additionally, analysis results mean nothing if there is not a person or group of people that are able to interpret them and react accordingly. As of today, these tools are not able to fully take on the figure of a student advisor.

It is worth noticing that data analytics in general has been an extremely active research area for many years, and it still is today. This also applies to learning analytics. As a matter of fact, most of the papers included in this literature review were published in 2017 or later. This means that the subject is most definitely not fully explored, and that many innovative pieces of work will keep arising in the foreseeable future, influenced by changes in teaching trends and progress in data analysis techniques.

Overall, this study highlights the many possibilities that predictive analytics provides in order to boost the learning process. At the same time, it is evident that building a single solution that will work well for many different types of learning environments is a very difficult task. This remains one of the greatest challenges within the learning analytics discipline.

Acknowledgment. This work is partially financed by public funds granted by the Galician regional government, with the purpose of supporting research activities carried out by PhD students. ("Programa de axudas á etapa predoutoral da Xunta de Galicia — Consellería de Educación, Universidade e Formación Profesional")

References

1. Adekitan, A.I., Salau, O.: The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon **5**(2), e01250 (Feb 2019). https://doi.org/10.1016/j.heliyon.2019.e01250, http://www.sciencedirect.com/science/article/pii/S240584401836924X

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

2. Akçapınar, G., Hasnine, M.N., Majumdar, R., Flanagan, B., Ogata, H.: Developing an early-warning system for spotting at-risk students by using eBook interaction logs. Smart Learning Environments **6**(1), 4 (May 2019). https://doi.org/10.1186/s40561-019-0083-4, https://doi.org/10.1186/s40561-019-0083-4
3. Akhtar, S., Warburton, S., Xu, W.: The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. International Journal of Technology and Design Education **27**(2), 251–270 (Jun 2017). https://doi.org/10.1007/s10798-015-9346-8, https://doi.org/10.1007/s10798-015-9346-8
4. Amirkhan, J.H., Kofman, Y.B.: Stress overload as a red flag for freshman failure and attrition. Contemporary Educational Psychology **54**, 297–308 (Jul 2018). https://doi.org/10.1016/j.cedpsych.2018.07.004, http://www.sciencedirect.com/science/article/pii/S0361476X17301108
5. Arnold, K.E., Pistilli, M.D.: Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In: Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge. pp. 267–270. LAK '12, ACM, New York, NY, USA (2012). https://doi.org/10.1145/2330601.2330666, http://doi.acm.org/10.1145/2330601.2330666
6. Benablo, C.I.P., Sarte, E.T., Dormido, J.M.D., Palaoag, T.: Higher Education Student's Academic Performance Analysis Through Predictive Analytics. In: Proceedings of the 2018 7th International Conference on Software and Computer Applications. pp. 238–242. ICSCA 2018, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3185089.3185102, http://doi.acm.org/10.1145/3185089.3185102, event-place: Kuantan, Malaysia
7. Brown, M.G., DeMonbrun, R.M., Lonn, S., Aguilar, S.J., Teasley, S.D.: What and when: The Role of Course Type and Timing in Students' Academic Performance. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. pp. 459–468. LAK '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2883851.2883907, http://doi.acm.org/10.1145/2883851.2883907, event-place: Edinburgh, United Kingdom
8. Brown, M.G., DeMonbrun, R.M., Teasley, S.D.: Don't Call It a Comeback: Academic Recovery and the Timing of Educational Technology Adoption. In: Proceedings of the Seventh International Learning Analytics & Knowledge Conference. pp. 489–493. LAK '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3027385.3027393, http://doi.acm.org/10.1145/3027385.3027393, event-place: Vancouver, British Columbia, Canada
9. Brown, M.G., DeMonbrun, R.M., Teasley, S.D.: Conceptualizing Coenrollment: Accounting for Student Experiences Across the Curriculum. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge. pp. 305–309. LAK '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3170358.3170366, http://doi.acm.org/10.1145/3170358.3170366, event-place: Sydney, New South Wales, Australia
10. Bussiere, M., Fratzscher, M.: Towards a new early warning system of financial crises. Journal of International Money and Finance **25**(6), 953–973 (Oct 2006). https://doi.org/10.1016/j.jimonfin.2006.07.007, http://www.sciencedirect.com/science/article/pii/S0261560606000532

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

11. Chen, P.H.: The Effects of College Students' In-Class and After-Class Lecture Note-Taking on Academic Performance. The Asia-Pacific Education Researcher **22**(2), 173–180 (May 2013). https://doi.org/10.1007/s40299-012-0010-8, https://doi.org/10.1007/s40299-012-0010-8
12. Cohen, A.: Analysis of student activity in web-supported courses as a tool for predicting dropout. Educational Technology Research and Development **65**(5), 1285–1304 (Oct 2017). https://doi.org/10.1007/s11423-017-9524-3, https://doi.org/10.1007/s11423-017-9524-3
13. Gandomi, A., Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management **35**(2), 137–144 (Apr 2015). https://doi.org/10.1016/j.ijinfomgt.2014.10.007, http://www.sciencedirect.com/science/article/pii/S0268401214001066
14. Gutiérrez, F., Seipp, K., Ochoa, X., Chiluiza, K., De Laet, T., Verbert, K.: LADA: A learning analytics dashboard for academic advising. Computers in Human Behavior **PP** (Dec 2018). https://doi.org/10.1016/j.chb.2018.12.004, http://www.sciencedirect.com/science/article/pii/S0747563218305909
15. Hirose, H.: Success/Failure Prediction for Final Examination Using the Trend of Weekly Online Testing. In: 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI). pp. 139–145 (Jul 2018). https://doi.org/10.1109/IIAI-AAI.2018.00036
16. Howard, E., Meehan, M., Parnell, A.: Contrasting prediction methods for early warning systems at undergraduate level. The Internet and Higher Education **37**, 66–75 (Apr 2018). https://doi.org/10.1016/j.iheduc.2018.02.001, http://www.sciencedirect.com/science/article/pii/S1096751617303974
17. Jovanovic, J., Mirriahi, N., Gašević, D., Dawson, S., Pardo, A.: Predictive power of regularity of pre-class activities in a flipped classroom. Computers & Education **134**, 156–168 (Jun 2019). https://doi.org/10.1016/j.compedu.2019.02.011, http://www.sciencedirect.com/science/article/pii/S0360131519300405
18. Kempler, S., Mathews, T.: Earth Science Data Analytics: Definitions, Techniques and Skills. Data Science Journal **16**(0), 6 (Feb 2017). https://doi.org/10.5334/dsj-2017-006, http://datascience.codata.org/articles/10.5334/dsj-2017-006/
19. Kitchenham, B., Charters, S.: Guidelines for performing Systematic Literature Reviews in Software Engineering (2007)
20. Kostopoulos, G., Karlos, S., Kotsiantis, S.B.: Multi-view Learning for Early Prognosis of Academic Performance: A Case Study. IEEE Transactions on Learning Technologies **PP** (2019). https://doi.org/10.1109/TLT.2019.2911581
21. Krumm, A.E., Waddington, R.J., Teasley, S.D., Lonn, S.: A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering. In: Larusson, J.A., White, B. (eds.) Learning Analytics: From Research to Practice, pp. 103–119. Springer New York, New York, NY (2014). https://doi.org/10.1007/978-1-4614-3305-7_6, https://doi.org/10.1007/978-1-4614-3305-7_6
22. Ornelas, F., Ordonez, C.: Predicting Student Success: A Naïve Bayesian Application to Community College Data. Technology, Knowledge and Learning **22**(3), 299–315 (Oct 2017). https://doi.org/10.1007/s10758-017-9334-z, https://doi.org/10.1007/s10758-017-9334-z
23. Plak, S., Cornelisz, I., Meeter, M., van Klaveren, C.: Early Warning Systems for More Effective Student Counseling in Higher Education – Evidence from a Dutch Field Experiment. In: SREE Spring 2019 Conference. p. 4. Washington, DC, USA (Mar 2019)

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

24. Reid, B.: Global early warning systems for natural hazards: systematic and people-centred. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **364**(1845), 2167–2182 (Aug 2006). https://doi.org/10.1098/rsta.2006.1819, https://royalsocietypublishing.org/doi/full/10.1098/rsta.2006.1819
25. Schuck, A.M.: Evaluating the Impact of Crime and Discipline on Student Success in Postsecondary Education. Research in Higher Education **58**(1), 77–97 (Feb 2017). https://doi.org/10.1007/s11162-016-9419-x, https://doi.org/10.1007/s11162-016-9419-x
26. Siemens, G.: 1st International Conference on Learning Analytics and Knowledge 2011 | Connecting the technical, pedagogical, and social dimensions of learning analytics (Jul 2010), https://tekri.athabascau.ca/analytics/
27. Smith, M.E.B., Chiovaro, J.C., O'Neil, M., Kansagara, D., Quinones, A., Freeman, M., Motu'apuaka, M., Slatore, C.G.: Early Warning System Scores: A Systematic Review. VA Evidence-based Synthesis Program Reports, Department of Veterans Affairs, Washington (DC) (2014), http://www.ncbi.nlm.nih.gov/books/NBK259026/
28. Thompson, E.D., Bowling, B.V., Markle, R.E.: Predicting Student Success in a Major's Introductory Biology Course via Logistic Regression Analysis of Scientific Reasoning Ability and Mathematics Scores. Research in Science Education **48**(1), 151–163 (Feb 2018). https://doi.org/10.1007/s11165-016-9563-5, https://doi.org/10.1007/s11165-016-9563-5
29. Tsiakmaki, M., Kostopoulos, G., Koutsonikos, G., Pierrakeas, C., Kotsiantis, S., Ragos, O.: Predicting University Students' Grades Based on Previous Academic Achievements. In: 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). pp. 1–6 (Jul 2018). https://doi.org/10.1109/IISA.2018.8633618
30. Umer, R., Susnjak, T., Mathrani, A., Suriadi, S.: A learning analytics approach: Using online weekly student engagement data to make predictions on student performance. In: 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube). pp. 1–5 (Nov 2018). https://doi.org/10.1109/ICECUBE.2018.8610959
31. Waddington, R.J., Nam, S.: Practice Exams Make Perfect: Incorporating Course Resource Use into an Early Warning System. In: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. pp. 188–192. LAK '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2567574.2567623, http://doi.acm.org/10.1145/2567574.2567623, event-place: Indianapolis, Indiana, USA
32. Wang, Z., Zhu, C., Ying, Z., Zhang, Y., Wang, B., Jin, X., Yang, H.: Design and Implementation of Early Warning System Based on Educational Big Data. In: 2018 5th International Conference on Systems and Informatics (ICSAI). pp. 549–553 (Nov 2018). https://doi.org/10.1109/ICSAI.2018.8599357

Copyright © 2019 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.