A LEARNING ANALYTICS APPROACH FOR STUDENT PERFORMANCE ASSESSMENT

Mohamed H Haggag[1], Mahmood Abdel Latif[2], Deena Mostafa Helal[3]

[1,2,3]Department of Computer Science, Helwan University, Cairo, Egypt.

ABSTRACT

Due to the increasing interest in big data especially in the educational field and online education has led to a conflict in terms of performance indicators of the student. In this paper we discuss the methodology of assessing the student performance in terms of the success indicators revealing a number of indicators that is recommended to indicate success of the final academic achievement.

KEYWORDS

Big data, student success, performance indicators

1. INTRODUCTION

Learning analytics, educational data mining, and academic analytics are closely related concepts that we need to differentiate between them according to [1] and [2]. We need to differentiate between LA (Learning Analytic), AA (Academic Analytics) And EDM (Educational Data Mining). Each of which has a different pattern in terms of process and purpose pointing out to the LA learning analytics it is the process of collecting measuring analysing and reporting data patterns with the purpose of optimizing the content according to the learner's experience learning analytics has its origins roots from business intelligence.[25] Moreover, the idea behind the optimization of the content in the learning experience for the student is defined as adaptive learning system that to gather and analyse the information in an intelligent way,   Educational Data Mining- Two areas that are specific to the use of big data in education are educational data mining and learning analytics. Although there is no hard and fast distinction between these two fields, they have had somewhat different research histories and are developing as distinct research areas. Generally, educational data mining looks for new patterns in data and develops new algorithms and/or new models, while learning analytics applies known predictive models in instructional systems. [1] Educational data mining focuses on developing and implementing methods with a goal of promoting discoveries from data in educational settings. It examines patterns in a large data set related to students' actions. As [3] provided a definition of educational data mining that uses data mining algorithms with the objective of solving certain educational issues. The methods may be utilized to form a better understanding of the educational settings and learners where [4] defined educational data mining as data analysis. Techniques which when applied extract hidden knowledge consisting of tasks consisting of pattern discovery as well as predictive modelling. Educational data mining is emerging as a research area with a number of computational and psychological methods and research approaches for understanding how students learn. New computer-supported interactive learning methods and tools—intelligent tutoring systems, simulations, games— have opened up opportunities to collect and analyse student data, to discover patterns and trends in those data, and to make new discoveries and test hypotheses about how students learn. Data collected from online learning systems can be aggregated over large numbers of students and can contain many variables that data mining algorithms can explore for model building. EDM develops methods and applies techniques from

statistics, machine learning, and data mining to analyse data collected during teaching and learning. EDM tests learning theories and informs educational practice [1].

 Academic analytics refers to an application of the principles and tools of business intelligence to academia with the goal of improving educational institutions' decision-making and performance Academic analytics combines "large data sets, statistical techniques, and predictive modelling" [5]. Learning Analytics and Educational data mining have risen as two independent research communities because of the increasing interest in the application of analytics in education, and the large amount of available data and software platforms and tools as well.

Learning analytics uses predictive models that provide actionable information. As [4] emphasized that a multidisciplinary approach based on data processing, technology-learning enhancement, educational data mining, and visualization. Learning analytics applies techniques from information science, sociology, psychology, statistics, machine learning, and data mining to analyse data collected during education administration and services, teaching, and learning. Learning analytics creates applications that directly influence educational practice [1].

The purpose of LA is to tailor educational opportunities to the individual learner's need and ability through actions such as intervening with students at risk or providing feedback and instructional content. Conversely, educational data mining tries to generate systematic and automated responses to learners. While LA focuses on the application of known methods and models to address issues affecting student learning and the organizational learning system, educational data mining focuses on the development of new computational data analysis methods. [1]. One of the main applications of learning analytics is tracking and predicting learners' performance as well as identifying potential problematic issues and students at risk [6][7]. Some universities have already used LA in various courses to improve learning.

For example, Purdue University used predictive modelling based on data collected from the course management system to identify students at risk and provide intervention. [21]

The University of Alabama improved student retention by forming a predictive model for students at risk based on the large data set of learners' demographics. In another case, Northern Arizona University connected resource use, risk level, and students' achievement by forming a predicting model to identify which students would benefit from which resource [5]. These are some examples of pioneer higher education institutions that applied LA. As in [8] When discussing learning analytics methods in education, it is important to provide a background regarding the flow of analytical information. The flow of analytical information can be traced from the students to the stakeholders within the framework of a hierarchy. When provided the opportunity to offer input and make recommendations, stakeholders can help enrich the learning experiences of students.   According to [9] the current literature related to the use of analytics in education had introduced to us many terms such as academic analytics, learning analytics, predictive analytics, social learning analytics, and educational data mining most of them are dominant than others according to the use of the researchers which the terms had been used by. According to [5], analytics marries large data sets, statistical techniques, and predictive modelling that would identify the implementation of analysing institutional data to produce actionable intelligence" in education and the appearance of "big data" term fiercely in the educational field associated with the technological development. Large data sets are present from students' interactions with educational software. online learning - among other sources - with public data repositories helping the researchers in the field because capturing the data would be difficult to handle in the normal way to show the student final performance [10]. In [11] who have followed the learning analytics approach by working on the importance of studying the student features that determine the

academic performance through the analysis of LMS data to allows instructors to discover meaningful patterns by identifying at-risk students and providing proactive feedback, and to adjust instructional strategies. [12].This method would enable the data driven decision making and improving the institutional overall performance at the same time.

 As [13] and [8] suggest that educational data mining and learning analytics will be extensively employed to improve the decision making process in institutions working on both improving the academic performance and finding optimizing solutions for student performance in higher education within the next few years. Moreover there are many researchers as [14] emphasized that using student features that would be primarily time based series that gives indication the performance from students' log all over time is much better than the one time discrete event spent on a one certain problem in terms of describing students' self-regulated learning. In this context, the present study aims to identify significant success indicators, including to predict course achievement. Additionally, this study examines whether the data collected in the middle of the course can successfully predict final course achievement, which would contribute to the possibility of early prediction based on the learning analytics approach. This paper is organized as follows into Six Section Two that discuss the related Literature Review that is related to this subject ,Section Three Methods Section Four Implementation Section Five the Results and Section Six the Conclusion.

2. THE ORETICAL BACKGROUND

2.1. Data use in learning analytics

Data-driven approach had been introduced by analytics in a way in terms that individuals and organizations conduct business and make decisions. Education, and in particular higher education became proof against the effect of analytics. Many universities around the world have started to apply analytics to gain new insights on a variety of business and educational issues aiming for improvement.

Online learning systems—learning management systems, learning platforms, and learning software—have the ability to capture streams of behaviours of learners are analysed in details can operate on the data providing a number of stakeholders with feedback to improve teaching, learning, and educational decision making. To clarify how such adaptive systems, operate, using the predictive models created by educational data mining and from the perspective of the system in terms of learning analytics. According to [11] who suggested that predicting the student performance along with providing proactive feedback have been two of the most frequently adopted tasks associated with learning analytics. Moreover, [15] reviewed the learning analytics tasks as categorizing students' behaviours after extracting them in terms of learning experience.

Many papers suggested the [16] even relates that the results shows considerate, and practically meaningful relationship between student features , parent supervision and academic achievement revealing that there is a strong relationship between the student characteristics and surrounding environment and the academic performance which is the aim of this study.

2.2. Determining student-at Risk through Predictive Modelling

 Tracking the student participation online and offline, student progression as well as highlighting any possible issues with teachers or the administration become one of the major tasks being adopted by learning analytics system in the universities nowadays.

For many institutions predictive modelling is the most usable method that enable them get have an access for the use of their historical data of their students and identifying the student at risk giving them instant support that they need for the course progression.

Access to LMS use logs for their courses and the data use became a driver for many scholar to develop coping software applications and by the use of those applications that the entire faculty would easily identify at risk students and give the needed support in some cases early intervention [28].

Not to mention one of the most important uses for the LMS which is reporting to improve student achievement, which depends on the entire faculty and learner responses. As [17] mentioned in their research that a student at-risk model can be tailored or built inside the institution itself in case data is collected and stored correctly using the appropriate method in collecting reporting and storing the data as well.

Back in [18] who categorized students by classifying their features extracted from logged data in a web-based system as a first step and their second step was prediction for their final grades using a genetic algorithms for the purpose of accuracy in their prediction through combining more than one classifier together and weighted feature vector as well. Then in 2005 [19] evaluates forums of discussions in online courses with the purpose of improving the ability of the tutor for assessing the progression of the discussions online . Also [20]used discriminant analysis on high school data (GPA and standardized test scores) to predict the successful completion of online courses.

 A tutorial case study is being exempted by [3] with Moodle giving a detailed description explaining that various data mining techniques and its packages are mainly used for improving the student's learning process and the course itself. As the study being developed by [21] in the Course Signals project which is one of most highlighted projects in the learning analytics field that builds model from student data to predict which students may be struggling academically and subsequently provides proactive intervention. Reports on pilots between fall 2007 and fall 2009 showed significant improvement on course completion, and mastery of content learning outcomes, making Course Signals one of the most successful proofs of concept of the use of data mining and statistical techniques to develop early alert systems). Further, [22] collected data from LMS and practiced the use of it as a significant indicator for the student success revealing that the use of data from course level perspective shows a great improvement and better predicting than other indicators for the student performance. However, the results for at-risk student were questionable and needs more clarification. Also [22] had developed a model for the code of practice that would help institutions. In dealing and getting the best use of learning analytics.
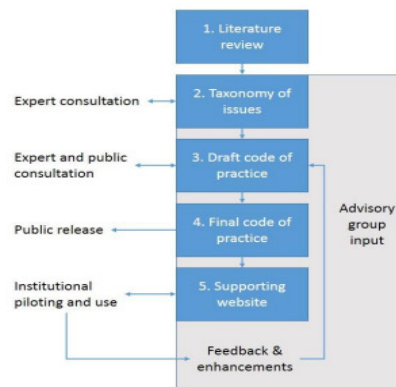
Fig 1 : Example of developing the Code of practice

There are obstacles for many institutions in proceeding in the field because of the lack of code practice affecting the technologies and the process resulting in affecting the individual learners as well. There are obstacles for many institutions in proceeding in the field because of the lack of code practice affecting the technologies and the process resulting in affecting the individual learners as well.

Back in [18] who categorized students by classifying their features extracted from logged data in a web-based system as a first step and their second step was prediction for their final grades using a genetic algorithms for the purpose of accuracy in their prediction through combining more than one classifier together and weighted feature vector as well. Then in 2005 [19] evaluates forums of discussions in online courses with the purpose of improving the ability of the tutor for assessing the progression of the discussions online . Also [20]used discriminant analysis on high school data (GPA and standardized test scores) to predict the successful completion of online courses.

 A tutorial case study is being exempted by [3] with Moodle giving a detailed description explaining that various data mining techniques and its packages are mainly used for improving the student's learning process and the course itself. As the study being developed by [21] in the Course Signals project which is one of most highlighted projects in the learning analytics field that builds model from student data to predict which students may be struggling academically and subsequently provides proactive intervention. Reports on pilots between fall 2007 and fall 2009 showed significant improvement on course completion, and mastery of content learning outcomes, making Course Signals one of the most successful proofs of concept of the use of data mining and statistical techniques to develop early alert systems). Further, [22] collected data from LMS and practiced the use of it as a significant indicator for the student success revealing that the use of data from course level perspective shows a great improvement and better predicting than other indicators for the student performance. However, the results for at-risk student were questionable and needs more clarification. Also [22] had developed a model for the code of practice that would help institutions. In dealing and getting the best use of learning analytics.

2.3. Student Intervention and providing a proactive feedback

Early intervention with students most at risk of attrition can be effective in improving college student retention.

Predictive models do not influence course completion and retention rates without being matched with effective intervention strategies aiming at helping at-risk students succeed. In 2004, Northern Arizona University (NAU) launched an initiative that used multiple data sources to identify at-risk first- year students that assess which proactive interventions have the best influence on their academic success and retention. The model measured utilization of services and resources (e.g., academic services recreational resources, social resources such as student organization membership, academic referrals, and advising sessions), levels of risk (e.g., standardized high school test scores, high school GPA), and outcomes (measured by first-year student GPAs and enrolment retention status). The initiative has evolved into an early alert system called GPS. Also [5] used factor analysis and logistic regression on a set of student features. derived from data extracted from Blackboard revealing 11 variables which included learning management system usage and student demographics. This research gave way to the development of Course Signals [21] dissertation at Purdue University a prominent early intervention system originally developed at Purdue.

Staff may have a greater level of consciousness about potential privacy violations than students

but their concerns may also stem from a feeling that learning analytics could be used to monitor their own activities which might act as an opposition. This may be justified: highlighting lack of timely feedback on assignments is an underlying motivation in some of the institutions interviewed. The role of staff in the analytics process is also mentioned as vital by some participants. Steps have been taken at Bridgwater and Derby, for instance, to allow staff to contextualise automated email interventions before they are sent on to students. Not to mention there is no one size that fits all according to [23] that while LMS activity and the student engagement are to be considered the most important ways to be specific about Students' performance because of the technique the data had been captured, gathered reported and intervened if an support is needed as well. According to JISC report 2016 in some situations the interventions are operated in an automated way such that LMS send the emails to students or staff. But still managed by the teachers and adapted based on the tutor's knowledge and management according to the actual of the learner's individual circumstances as in the case At Derby, letters are sent to students who are absent and the tone of the letters become more severe as the absence continues.

2.4. Early Prediction and self-Regulating

Providing learners with better information about their progress is considered one of the main drivers to the success throughout the learning process. According to [22], the findings of Campbell's study recommends to combine student demographic features and LMS behaviour which is eventually considered to be the future trend increasing ability to the predict the student achievement based on the input variables.

A considerable gap in this field is to be mentioned causing a gap in current education technology research. Highlighting this gap is of a great importance as that in Campbell's research he had a much higher R2 (Meaning that that there is stronger relationship at Campbell's between the student features and the academic performance other than the other studies) value than other studies, and student characteristics provided most of the predictive value. The purpose of those studies is still questioned such that the past research has been conducted largely to demonstrate that variables in LMS activity logs do have a statistically significant relationship to student achievement. The research does not seek to advance knowledge about how the LMS activity was related to the pedagogical function behind the online activities, the course content, or student characteristics. [24],[25] and [26]. Learning and academic analytics in higher education are used to predict student success by examining how and what students learn and how success is supported by academic programs and institutions. [27]concludes that, students who are more engaged with the content and discussions in an online course will achieve and complete successfully. However, the studies provide a guideline for those who would be interested in the interactions between students, faculty, course materials and course structures.

Additionally, If only we focused on individual learners, these applications do not help us to evaluate if a course activity or overall design can be improved for all learners in a course, which has a much larger potential impact to improve student achievement. [29]they ensured according to their findings that correlations do not confirm and extend earlier propositions ([5], [26] and [30]) mentioned that pedagogically meaningful information can be extracted from LMSs and made available to educators via a dashboard- like interface where predictive models could be built and network visualization tools. Student tracking data are readily accessible, scalable and non-intrusive and provide lead indicators of final student performance or failure. Campbell (2007) has reported that LMS data added only minimally to the power of his model for predicting student academic success which we could refer to the generalisability of the model Campbell sought to predict student overall success across an entire academic program, and consequently was unable to consider the differential intentions and differential tool

implementation of individual instructors within individual courses. The purposes of monitoring student activity and achievement, predictive models must be course oriented in the first level. Furthermore, future developments of any evaluative and data visualization resources must be highly tailored according to the instructor differences for adopting LMS tools and the pedagogical purposes as well. Providing students themselves with better information on their progress is also mentioned as being an important driver. That is driven primarily by the need to improve student success, retention and the learning experience, learning analytics is a rapidly growing area of interest in educational institutions worldwide.

Here [31]emphasized that because self-regulated learning is critical to success in online Learning, therefore measures reflecting self-regulated learning should be included to examine the relationship between LMS data measures and course achievement revealing the importance of self-regulated learning and verify the advantages of using measures related to meaningful learning behaviours rather than simple frequency measures.

Furthermore, the measures collected in the middle of the course significantly predicted course achievement, and the findings support the potential for early prediction using learning performance data.

Resulting in extraction and aggregation of meaningful indicators from the students' behavioural data and the development of reliable prediction models are valuable.

Instructors who seek to understand their students' learning statuses and to provide actionable feedback during the course. In this paper, we will examine the role of learning and learning analytics and their relevance to academic performance in undergraduate and graduate programs.

2.5. The Research Questions

a-What are the important features/indicators for student to measure and asses their performance?
b-How to measure / combine those features together to indicate the student performance?

3. METHODOLOGY

Analytics collectively use the applied techniques from computer science mathematics and statistics for generating information that has a meaning from a very large dataset [1]. ive Stages suggested by Campbell and Oblinger (2007) .
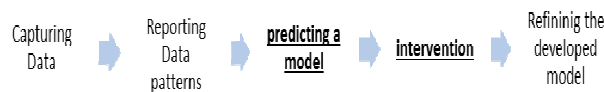
Capturing Data → Reporting Data patterns → **predicting a model** → **intervention** → Refininig the developed model

Fig 2: 5 steps for designing the LA Models Suggested by Campbell and Oblinger (2007)

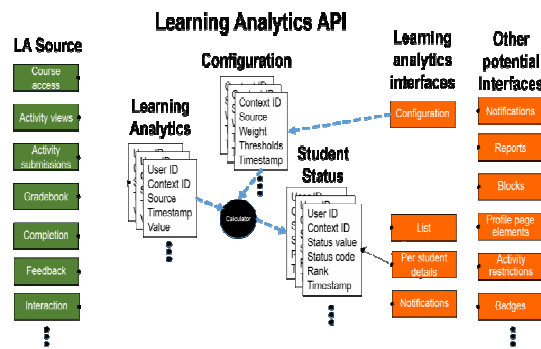### 3.1 Capturing data from the VLE (Virtual Learning Environment

Fig 3: Learning analytics process

Each of the designated learning analytics sources should reveal the information it encounters through day to day and easy queries so we can use them in real time without buffering. Instructors should be able to determine the meaning and effect of the measures of the indicators on the students as well. Here are some of the learning analytics sources that are made available for the data.

• Content/announcement views relative within students of a course
• Activity/resource views, reads (from event observer, using R in CRUD), relative to number of possible views.
• Interacting with the tutors.
• Interacting with the students and the actual content through online discussion.
• Mid-year and final academic performance.
• Assessment Feedback views.
• Content/announcement.

Capturing data from VLE is a first step that happens when the user/learner interacts with the announcement or the digital content providing that LMS provides restricted data about the whole process or requires even more complicated procedure for capturing purposes. Gathering involves the selection and capture of data; processing involves the aggregation of information; displaying involves the sharing of the information; and dissemination involves decision-making based on that information. The process of capturing itself includes (a) recording the activity by capturing the interaction of the learner with the actual learning content (b) and detailed activity which could be used for further purposes such as measuring the student engagement degree with the learning content[32]. Combining the data with the statistical techniques after the process of capturing comes a second step identifying the technologies and describing the methods been used, then analysing and assessing the effect of the activities and plan which will lead eventually to predict the student engagement features that are important in the business process. Summarizing the collection phase :

(1) Collect the information about learners from their learning profiles on the virtual learning environment.
(2) Filter the information and data gathered from student.
(3) Select the information needed according to the information to be built and the   purpose for this model.

3.2 Reporting data pattern and trends

The Society for Learning Analytics Research (SoLAR, 2011) defines LA 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".

[33] Learning analytics reveal data and information about usage, trends, and patterns of learning. Traditionally, student course evaluations provide data and information at the end of the semester and have been based on openness and reflections of students who have completed the course. Learning analytics are collected before the end of the course and can help to identify students who may be struggling during the course and may include feedback about how participants are using the course content, and their level of participation.

3.3 Predicting a model based on the data available using statistical regression

3.3.1 Selecting the Indicator (outcome measure)

Considering the student performance as the only dependent variable represented in Y indicating if the learning process is in total success or fail such that we the student performance would be the academic final performance

$$Y = a + b^1 x^1 + b^2 x^2 + b^3 x^3 \qquad (1) \quad \text{(Multiple Regression)}$$

3.3.2 Selecting the Predictors (KPI)

Before any determination can be made as to whether any data-based analytics are effective. Indicative metrics should be primarily identified in which they should be meaningful, measurable, and monitored.

Any decisions about objectives and outcomes should be made in the context of an overall strategic plan. In such context learning analytics can be a powerful tool for discovering which modules, sections, or pages of a site are most popular and effective for learners. (e.g., demographics, academic ability/performance/history, financial, and other information) to make strategic decisions not only about a particular course section, but also about individual students and academic programs, and institutional planning.[32] Multiple correlation coefficient that is noted by R that describes of how much of the variability in the student performance is related to many of the student features so in order and to get this we will be squaring the "R".

So R could be computed as follows:

$$R = \sqrt{\frac{\left[\left(r_{y,x1}\right)^2 + \left(r_{y,x2}\right)^2\right] - \left(2 r_{y,x1} r_{y,x2} r_{x1,x2}\right)}{1 - \left(r_{x1,x2}\right)^2}}$$

$$(2)$$

3.3.3 Assessing the accuracy of the Prediction

Assessing accuracy of the model is best accomplished by analysing the standard error of estimate (SEE) and the percentage that the SEE represents of the predicted mean. The SEE represents the degree to which the predicted scores vary from the observed scores on the criterion measure, similar to the standard deviation used in other statistical procedures. The relevant formulas are:

$$s_{est} = \sqrt{\frac{\sum (Y-Y')^2}{N-2}}$$

$$(3)$$

AND

$$s_{est} = S_y \sqrt{1-r^2} \sqrt{\frac{N}{N-2}}$$

(4)

where r is the sample correlation

S (estimated) = the standard deviation of Y
r is the sample correlation

3.3.4 Acting by using an intervention based on the model to improve learning

System-generated interventions can range from a simple alert about a student's likelihood of success to requiring at-risk students to take specific actions to address concerns. In [35]Interventions should be made as early as possible to ensure that students have the opportunity to develop effective learning habits before they are negatively affected. However, this drive must be balanced against the need for informative data on a student's patterns of engagement. Many data models will begin to provide meaningful patterns in the earliest weeks.

Moreover, the timing of follow-up interventions should be based on an institution's understanding of its own students' needs and also keeping track of the a) type b) timing   c) outcome of interventions will enable institutions to identify patterns of success. Analytics may associate the recommended intervention through the LMS according to the impact and the use of interventions and for enabling such an association we either tailor the existing system or build a novel system for the purpose of sharing data between systems for privacy purposes. As being emphasized in [36] tracking the evidence is of great importance that interventions and activities should also be based upon evidence such as time event logs. It is highly recommended for institutions to enable access to the tracking of the whole process including the intervention within the entire institution and to emphasize that the research is used initially to serve the analytics strategy.

3.3.5 Refining the developed Model

Acting upon the proposed model in addition to the actively produced feedback such that it could be timely updated and acted upon the model learning analytics analysis. Learning analytics focuses on data related to learners' interactions with course content, other students, and instructors.

*Data Use in Adaptive Learning-*Online learning systems—learning management systems, learning platforms, and learning software—have the ability to capture streams of fine-grained learner behaviours, and the tools and techniques described above can operate on the data to provide a variety of stakeholders with feedback to improve teaching, learning, and educational decision making.

To demonstrate how such adaptive systems, operate, using the predictive models created by educational data mining and the system-level view of learning analytics, this section describes a prototypical learning system with six components.

A content management, maintenance, and delivery component interacts with students to customized subject content and assessments to support student learning.

• A student learning database (or other big data repository) stores real time student input and behaviours captured as students interacts within the system.

- A predictive model combines demographic data (from an external student information system) and learning/behaviour data from the student learning database to track a student's progress and make predictions about his or her future behaviours or performance
- A reporting server uses the output of the predictive model to produce indicative dashboards that provide visible feedback for many users (tutors/students).
- An adaption engine customizes actual content based on the output of the predictive model to deliver material according to a student's performance level in a being improved continuously.
- An intervention engine allows teachers, administrators, or system developers to intervene and override the automated system to better serve a student's learning.

Additionally, the idea behind the optimization of the content in the learning experience for the student is featured as adaptive learning system for the purpose of gathering and analysing the information in an intelligent way to better shape learning pedagogy.

The adaptive learning is to be considered as an artificial intelligent tutor as the new technology learns the way people learn and adapt its material according to how do each one gets the information in which accordingly they adjust the learning content presented to learner.

In this context adaptive learning will help the higher education and the students who fail or might be at risk.
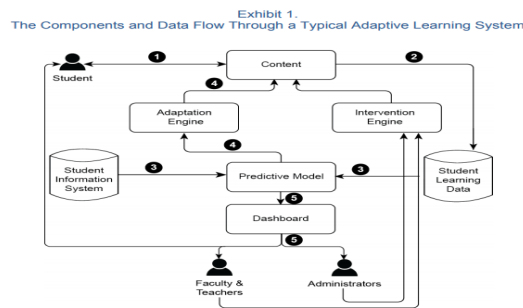

Fig.4: Adaptive Learning system

Analytics, broadly defined, comprises applied techniques from computer science, mathematics, and statistics for extracting usable information from very large datasets.

4. Implementation

The data collection process is accomplished using a learner activity tracker tool, which called experience API (XAPI). The collected features are classified into three categories.

Demographic features, academic background features and behavioural features. The behavioural features are a new feature category that is linked to the leaner experience during educational process[37].

The data used set consists of 150 student's record with eleven different features. The features are classified into three main categories: (1) Demographic features such as gender and nationality. (2) Academic background features such as stage, grade and section. (3) Behavioural factors such as raised hand on class, opening resources, participating in discussions groups, viewing messages and announcements.

Methodology applied

In the dataset that we are using we chose to use to investigate the research questions using descriptive statistics and correlation analyses that are performed, and throughout the statistical methods the hierarchical regression had been chosen as a method to identify 4the important features that would assess the student performance.

The descriptive statistics for the study variables are presented in Table 2 the high mean of the "Visited resources" variable indicated that the participants visited on a regular basis, but their announcement view varied widely.
.
The class was very large, and it was designed more for individual learning than for collaborative learning. The results of the correlation analyses of the study variables are presented in Table 2
we chose to investigate the research questions using descriptive statistics and correlation analyses were performed in order to decide among all the student features which are more indicative one than others, and throughout the statistical methods the hierarchical regression had been chosen as a method to identify the most important features that could possibly assess the student performance.

4.1 Descriptive and correlation analyses

Table1: Descriptive Data

	Mean	SD	Minimum	Maximum
Discussion	43.3319415	27.6460829	9	82
Announcements View	37.993737	26.5882513	5	98
Visited Resources	54.8789144	33.0669087	0	99
Raised Hands	46.8413361	30.777034	0	100

In this Paper we chose to use to investigate the research questions using descriptive statistics and correlation analyses were performed, and throughout the statistical methods the and hierarchical regression had been chosen as a method to identify the important features that would assess the student performance.

The descriptive statistics for the study variables are presented in Table 1 the high mean of the visiting the resources variable indicated that they were engaged with content on a regular basis, but the view of the announcement was not high which reflects that they were not engaged on a regular basis. Speaking of discussion, it was moderate in comparison to other indicators.

The class was very large, and it was designed more for individual learning than for collaborative learning. The results of the correlation analyses of the student behavioural features are presented in Table 2.

Step 1, Discussion, Parent Answering Survey, Announcements View and Visited Resources was entered in the correlation analysis as. A stepwise method was used, Discussion, Parent Answering Survey, Announcements View and Visited Resources that are considered to be under the behavioural features category were entered in this step three variables were revealed to be significant: Visited Resources, Discussion, Announcements View.

Table 2 : Correlation analysis

Result	4	3	2	1	
				1	1.Raised Hands
			1	0.691571705	2.Visited Resources
		1	0.594500027	0.643917768	3.Announcements View
	1	0.417289999	0.243291769	0.339385991	4.Discussion
1	0.307824597	0.540758185	0.699285843	0.650906541	Result

As presented in table 3, raised hands (B=0.10, p<0.5) Visited resources (B=0.14, p<0.5) Announcement View (B=0.09, p<0.5) Discussion (B= 4.18, p<0.5) indicating that those variables are more important than others in predicting the student final achievement.

The regression model predicted 57.7% of the variance in the student final achievement ($R2 = 0.579$, p b .001).

Table 3: hierarchical regression

Model	Predictors	B	SE	R2
M0	Raised Hands	1.7	23.3	0.42
M1	Raised Hands	0.12	7.95	0.54
	Visited Resources	0.16		
M2	Raised Hands	0.10	7.9	0.54
	Visited Resources	0.16		
	Announcements View	0.04		
M3	Raised Hands	0.10	7.88	0.54
	Visited Resources	0.16		
	Announcements View	0.02		
	Discussion	0.03		
M4	Raised Hands	0.10	7.66	0.57
	Visited Resources	0.14		
	Announcements View	0.09		
	Discussion	0.02		
	PAS	4.18		

The current research is being held to answer some questions in regards to the indicators for the student performance to help instructors in their final academic assessment.

Our Current study was sought to show the most indicative variable in assessing the student performance. The results of the actual study showed that visiting resources, view of announcements and the discussion were the selected indicators in the performance assessment of the student among all the characteristics.

Shifting to the online environment is no longer a choice therefore critical measures were determined to explain it. our results showed that visiting the resources was the most critical behaviour in assessing the student performance followed by raised hands, announcement view and discussion.

Four Indicators from the xAPI indicated the student final achievement and the predictivity of the regression model with four variables explained of variance of 58% of the final score where the predictability of the model showed to be high compared those from previous research (e.g., 33% in McFadden & Dawson, 2010, and 31% in Morris et al., 2005)
To sum the results up, the outcome indicated with clearance the importance of the engagement of students online and the importance of online learning , submitting tasks /emails in actionably timely manner , and regular visiting of the course content online and reading the course guidelines which are emphasized and highlighted by other researchers[38].

The current research results support many indications that need to be explained.

Firstly despite that the xAPI gave many variables online and many offline variables been collected , the fact that [38]highlighted the idea of not using many actual predictors as it would result in better results therefore researchers still need to work in finding more predictors that would help in measuring the student engagement with online content. The current research implies the importance of online learning and the engagement of the student with the course material online is extremely important. Although studying offline has fixed schedules with regular attendance but there is no critical measure for checking the material and to interact with course guidelines but studying online need to be timely managed to cope up with the course so the present research highlight the importance of frequent visits to the resources online. Tracking the student participation online and offline, progression, highlighting any possible issues with instructors in regards to the engagement with the course content become one of the major tasks being adopted by learning analytics system academically nowadays. the extraction and gathering of meaningful indicators from the students' behavioural data and the building prediction models that can be relied on are important for tutors who need to understand the level of their students and intervene in the actionably timely manner.

REFERENCES

[1]   Bienkowski M, Feng M and Means B 2012 Enhancing teaching and learning through educational data mining and learning analytics: An issue brief Washington, DC SRI Int. 1–57

[2]   Elias T 2011 Learning Analytics : Definitions , Processes and Potential Learning 23 134–48

[3]   Romero C and Ventura S 2013 Data mining in education Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 3 12–27

[4]   Hung J, Hsu Y-C and Rice K 2012 Integrating Data Mining in Program Evaluation of K-12 Online Education Educ. Technol. Soc. 15 27–41

[5]   O B L I N G E R By G, Campbell J P, Deblois P B and Oblinger D G 2007 "Academic Analytics: A New Tool for a New Era"

[6]   EDUCAUSE 2010 Next generation learning challenges: Learner analytics premises Educause

[7]   Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood K 2011 The horizon report: 2011 edition

[8]   Reyes J A 2015 The Skinny on Big Data in Education: Learning Analytics Simplified. TechTrends: Linking Research and Practice to Improve Learning 59(2), 75-80

[9] Jayaprakash S M, Moody E W, Lauría E J M, Regan J R and Baron J D 2014 Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative J. Learn. Anal. 1 6–47

[10] Koedinger K R, Baker R S J, Cunningham K and Skogsholm A 2010 A Data Repository for the EDM community : The PSLC DataShop Handb. Educ. Data Min. 43–55

[11] Gaševic D, Dawson S and Siemens G 2015 Let ' s not forget : Learning Analytics are about Learning Course Signals : Lessons Learned TechTrends 71–64 59

[12] Dietz-Uhler B and Hurn J 2013 Using learning analytics to predict (and improve) student success: a faculty perspective J. Interact. Online Learn. 12 17–26

[13] Johnson L, Adams Becker S, Estrada V, Freeman A, Kampylis P, Vuorikari R and Punie Y 2014 Horizon Report Europe: 2014 Schools Edition_Sintese e Introdução PT

[14] Hadwin A F, Nesbit J C, Jamieson-Noel D, Code J and Winne P H 2007 Examining trace data to explore self-regulated learning Metacognition Learn. 2 107–24

[15] Blikstein P 2011 Using learning analytics to assess students' behavior in open-ended programming tasks Proc. 1st Int. Conf. Learn. Anal. Knowl. - LAK '11 110

[16] Fan X and Chen M 2001 Parental involvement and students' academic achievement: A meta-analysis Educ. Psychol. Rev. 13 1–22

[17] Agnihotri L and Ott A 2014 Building a Student At-Risk Model : An End-to-End Perspective Proc. 7th Int. Conf. Educ. Data Min. 209–12

[18] Minaei-Bidgoli B, Kashy D a., Kortemeyer G and Punch W F 2003 Predicting student performance: an application of data mining methods with an educational web-based system 33rd Annu. Front. Educ. 2003. FIE 2003. 1T2A_13-T2A_18

[19] Dringus L P and Ellis T 2005 Using data mining as a strategy for assessing asynchronous discussion forums Comput. Educ. 45 141–60

[20] Morris L V, Wu S S and Finnegan C L 2005 Predicting Retention In Online General Education Courses Am. J. Distance Educ. 19 23–36

[21] Arnold K E, Pistilli M D and Arnold K E 2012 Course Signals at Purdue: Using Learning Analytics to Increase Student Success 2nd Int. Conf. Learn. Anal. Knowl. 2–5

[22] John Whitmer K F and W A 2012 Analytics in Progress: Technology Use, Student Characteristics, and Student Achievement

[23] Gašević D, Dawson S, Rogers T and Gasevic D 2016 Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success Internet High. Educ. 28 68–84

[24] McAfee A, Brynjolfsson E, Boyd D, Crawford K and Lohr S 2012 Critical Questions for Big Data Information, Commun. Soc. 15 1–5

[25] Shum S B and Ferguson R 2012 Social Learning Analytics Educ. Technol. Soc. 15 3–26

[26] Siemens G and Long P 2011 Penetrating the Fog: Analytics in Learning and Education Educ. Rev. 46 30–2

[27] Morris L V, Finnegan C and Wu S 2005 Tracking student behavior , persistence , and achievement in online courses 8 221–31

[28] Mazza R and Dimitrova V 2007 CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses Int. J. Hum. Comput. Stud. 65 125–39

[29] Macfadyen L P and Dawson S 2010 Mining LMS data to develop an "early warning system" for educators: A proof of concept Comput. Educ. 54 588–99

[30] Richard Lynch M D 2002 The Relationship Between Self-Regulation and Online Learning in a Blended Learning Context Int. Rev. Res. an open Distrib. Learn.

[31] You J W 2016 Identifying significant indicators using LMS data to predict course achievement in online learning Internet High. Educ. 29 23–30

[32] Pardo A Data Capturing Mechanisms for Learning Analytics

[33] Baker R S J D, Yacef K, et al 2012 Learning analytics for online discussions: a pedagogical model for intervention with embedded and extracted analytics Am. Behav. Sci. 15 2–5

[34] Lane D Logic of Hypothesis Testing - David Lane

[35] Linda Baer J C Chapter 4: From Metrics to Analytics, Reporting to Action: Analytics'TM Role in Changing the Learning Environment Educ. Publ.

[36] Anon Student Intervention Guide Learn. Anal. Educ. data Min. Learn. impact

[37] Amrieh E A, Hamtini T and Aljarah I 2015 Preprocessing and analyzing educational data set using X-API for improving student's performance 2015 IEEE Jordan Conf. Appl. Electr. Eng. Comput.

Technol. 1–5

[38]   Chen J F, Hsieh H N and Do Q H 2014 Predicting student academic performance: A comparison of two meta-heuristic algorithms inspired by cuckoo birds for training neural networks Algorithms 7 538–53.

Authors

Professor Mohamed H.Haggag computer science department in faculty of computer science Helwan university he received his B.Sc. 1992 Faculty of Engineering/Mansoura University, his masters from in 1995 Faculty of Engineering/Mansoura University and then in 2000 he received his phd 2000 Faculty of Engineering/Mansoura University he was a previous director to the computing center for helwan university and previous vice dean for the faculty for student affairs and he is currently a head for the post graduate studies department for the faculty.

Prof. Faculty of Computers & Info., Helwan University , Egypt - Assoc. prof at Uni. of Jeddah, SA he was the head of dept. of post graduate studies and he was vice dean for the faculty in environmental issues and his industry knowledge is in datamining ,project management, software development, data analysis and process improvement and has around 20 published paper in data mining , semantics and ontology.

Third author

Deena Mostafa Helal she is currently a demonstrator in information system department. She received her Bsc. IN 2012 and she is a current a master student working on her thesis and her intersest in network , big data and learning analytics.