PREDICTION OF STUDENTS' PERFORMANCE IN e-LEARNING ENVIRONMENT OF UTMSPACE PROGRAM

ZAHARADDEEN HUSSAINI

A dissertation submitted in partial fullfilment of the
requirements for the award of the degree of
Master of Computer Science

Faculy of Computing
Universiti Teknologi Malaysia

DECEMBER 2017

I dedicate my research work to my late parents Alhaji Hussaini Kaita and Hajia Habiba Hussaini Kaita for their moral support, to my lovely brothers, sister and also to my lovely wife. May Allah (SWA) forgive their shortcomings and grant them Jannatul Firdaus.

ACKNOWLEDGEMENT

All praise be to Allah (SWA) the Beneficent the Merciful, the creator of the heavens and the earth, May the peace and blessings of Allah be upon our Noble Prophet Muhammad (SAW) who instruct his followers to seek knowledge where ever it may take them.

I would like to appreciate the management of Universiti of Teknologi Malaysia in general and in particular to the Faculty of Computing for the provision of qualified human resources and material resources on ground to enable the students to have a good studying environment.

I humbly appreciate the effort of my able supervisor in person of Dr. Nor Bahiah HJ Ahmad who worked with me tirelessly in making this research work a success and worthy of presentation, May Almighty Allah bless you and your family.

I would also like to acknowledge my brothers Alh. Muntari Hussaini Kaita, Dikko Hussaini Kaita and Haj Magajia Hussaini Kaita for the fatherly advises and support during my studies, not forgetting my lovely wife for her patience in bearing with absence for the period of studies. May Almighty Allah bless us well.

My friends and classmates in the Universiti of Teknologi Malaysia who are numerous to mention are also worthy of acknowledging for their good friendships and educative collaboration in the success of my studies.

ABSTRACT

Part-time educational programmes enable workers in both private and public sectors a means of acquiring knowledge and advancing themselves in their career. However, part-time students face some emanating challenges in their studies such as time constraint, inability to see lecturers and utilizing the educational resources due to their work commitments. With the advancement in e-learning technologies, the part-time students are able to empower themselves by interacting with eLearning environment so that the instructor may not be the gatekeeper of education. This dissertation is aimed at predicting the performance of part time students registered in UTMSPACE program based on their interactivity with the eLearning activities in MOODLE and MOOCs, this was achieved with the use of the student log files and some additional data about the particular student. The performance prediction was investigated using Decision Tree (C4.5 algorithm) and Neural Network algorithm techniques, in order to find the best technique for the student's prediction. Neural Networks out-performed Decision Tree C4.5 algorithms by giving 92% accuracy which was validated using precision and recall analysis of the classifier, while Decision Tree obtained 89.2% accuracy. In addition, the analysis of log files indicates that the rate of interactivity with e-learning environment has a significant impact on their performance as the students with highest interactivity on the MOODLE tend to have higher performance than those with low interactivity rate. From the analysis of the log files we can observe that the students spend more time on e-learning MOODLE than MOOCs, and because of that they are missing advantages of the available resources on MOOCs such as watching lecture videos, participating in quizzes, which may assist them in their study.

ABSTRAK

Program pendidikan separuh masa membolehkan pekerja di sektor swasta dan sektor awam menyambung pengajian di peringkat tinggi dan seterusnya memajukan diri dalam kerjaya mereka. Walau bagaimanapun, pelajar separuh masa menghadapi cabaran dalam pengajian mereka seperti kekangan masa, ketidakupayaan untuk berjumpa pensyarah dan kekangan menggunakan sumber pendidikan disebabkan komitmen kerja mereka. Dengan kemajuan dalam teknologi e-pembelajaran, pelajar separuh masa dapat memantapkan pengetahuan diri mereka dengan mencapai dan berinteraksi dengan e-pembelajaran supaya pengajar tidak menjadi sumber penyampai ilmu semata-mata. Disertasi ini bertujuan untuk meramalkan prestasi pelajar separuh masa yang didaftarkan dalam program UTMSPACE berdasarkan interaktiviti mereka dengan aktiviti eLearning di MOODLE dan MOOCs. Interaksi pelajar diperolehi daripada fail log pelajar dan data-data tambahan mengenai profil pelajar. Ramalan prestasi dilaksanakan menggunakan teknik Pepohon Keputusan (algoritma C4.5) dan teknik algoritma Neural Network. Hasil kajian mendapati ramalan yang dihasilkan oleh Rangkaian Neural mengatasi Pohon Keputusan, algoritma C4.5 dengan memberikan ketepatan 92%, sementara Pohon Keputusan mendapat 89.2% ketepatan. Kajian ini telah disahkan menggunakan analisis ketepatan dan penarikan semula pengelas. Di samping itu, analisis logfiles menunjukkan bahawa kadar interaktiviti dengan persekitaran e-pembelajaran mempunyai kesan yang signifikan terhadap prestasi mereka kerana para pelajar yang mempunyai interaktiviti tertinggi di MOODLE mempunyai prestasi yang lebih tinggi daripada mereka yang mempunyai kadar interaktiviti yang rendah. Dari analisis logfiles kita dapat melihat bahawa para pelajar menghabiskan lebih banyak masa dalam e-learning MOODLE daripada MOOCs. Pelajar juga didapati tidak menggunakan sumber yang ada pada MOOCs sepenuhnya seperti menonton video kuliah dan membuat kuiz, yang boleh membantu mereka dalam pengajian mereka.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	-	Artificial Neural Network
CS	-	Computer Science
CSV	-	Comma Separated Version
CMS	-	Course Management System
EDM	-	Educational Data Mining
LA	-	Learning Analytics
LMS	-	Learning Management System
MOOCS	-	Massive Open Online Courses
MOODLE	-	Modular Object Oriented Dynamic Learning Environment
VLE	-	Virtual Learning Environment
SVM	-	Support Vector Machine
UTM	-	University of Technology Malaysia

LIST OFAPPENDICES

CHAPTER 1

INTRODUCTION

1.0 Overview

Moro-Egido and Panades (2010), indicated that there is tendency of having distinct effects on the academic performance of full-time and part-time students. The research finds little conclusive evidence that working affects the average outcomes of part-time students overall and for any part-time students. With the expanding rates of studying while working, the interest and approach activities has extended into higher educational institutions, it is basic to comprehend the expenses and advantages of working while in school, particularly for diverse sections of students. The study further suggests that working while in school can have future work showcase adjustments and enhance delicate aptitudes, for example, time effectiveness, correspondence, critical thinking capacity, and moral duty. On the other hand, part-time students are less contented than full-time students; which confirms research hypothesis regarding the negative effect of part-time work on students' general level of satisfaction of their studies experiences. It indicated that part-time students cannot take full advantage of the facilities available to full-time students. Lack of access or participation may lead part-time students to assess their academic performance more negatively.

Universiti Teknologi Malaysia, School of Professional and Continuing Education (UTMSPACE) offered professional development programmes such as short courses, seminars, workshops and in-house training. UTMSPACE offers Academic Programmes courses at Diploma and Undergraduate level, in line with the courses offered by the various Faculties at UTM. For customers' convenience, the classes are conducted at 17 learning centres all over the country. UTMSPACE offers courses due to an overwhelming demand for part-time and full-time educational programmes in both private and public sectors, UTMSPACE utilises the technological advancement of eLearning MOODLE and MOOCs.

Researchers predicting student performances in in-class education have typically obtained data for analysis using validated questionnaires, interviews, and observational techniques, with relevant theoretical concepts in mind so that the analysis can be prepared towards the notions that the researcher thinks need to be measured. The utilization of eLearning management system takes into account analyzing students' online behaviour without the need of tedious information accumulation or data-collection. Nonetheless, eLearning management system gives basic log information that are not solid estimations of already laid out hypothetical ideas. It is along these lines imperative to comprehend whether and how this information can be utilized for learning investigation (Conijn *et al.,* 2017).

With substantial class sizes at colleges, significantly bigger class sizes in Massive Open Online Courses (MOOCs), that have experienced a fast advancement in the previous couple of ages, it has turned out to be incomprehensible for the educator and instructing aides to monitor the execution of every student separately. This can lead to students drooping out in a class of students who could have perform better if fitting beneficial moves were made sufficiently early or brilliant students not getting vital advancement to profit maximally from the course (Meier *et al.,* 2016).

Romero *et al,* (2013) indicated that these days there is expanding enthusiasm for the utilization of exchange gatherings as an indicator of student performance, on-line discussion forums constitute group of people gaining from each other, which not just notify the students about their peers' inquiries and issues yet can moreover prompt educators about their students' questions and issues, and can likewise advise teachers about their students' information of the course contents. It could be noted that student's performance might be impacted by a few variables, for example, age, gender, parent's financial circumstance, occupation, nature of school being gone to, class medium of educating, number of study hours spent every day, and nature of settlement which might be school possess inn or something else.

Meier *et al*., (2016) recommends that; it is of overwhelming significance to create customized frameworks that anticipate the performance of a student in a course before the course is over and as quickly as time permits, and outlining courses to have early in-class assessments that identifies students who have the potential to do poorly without intervention.

Therefore, this research intends to explore, exploit, and analyze the learning processes of students using data mining techniques and predict their performances. The result is expected to show the correlation between the students' performance with the rate of interactivity in eLearning environment/MOOCs in Introduction Data Structure and Algorithm in UTMSPACE program, with the aim of finding out whether interactivity could have an effect in the student performance of students in eLearning/MOOCs process positively or negatively.

The study therefore focuses on developing prediction model of students' academic performance based on their interaction with e-learning management system and MOOCs using data mining techniques (Neural Networks and Decision tree) in the prediction of student performance with the aim of achieving high prediction accuracy and better performance.

1.2 Background of the study

The advancement in information technology has enhanced the effectiveness of web based education (e-learning) system. The e-learning system allows students from anywhere and at any time to have access and carry out different learning activities such as reading, downloading and uploading documents, presentations and assignments. The activities normally take place in a platform called Learning Management System (LMS), which utilizes various technology mostly on the internet which facilitate the student's communication between students and tutors or among the students. The LMS provides a huge volume of varied data for usage by the students and the instructors as well, it also contains data about the student's personal data, learning styles, practices, behaviors and usability preferences (Dráždilová *et al,* 2010).

In addition, Massive open online courses (MOOCs) empower students around the world to gain from top notch instructive substance with ease. A standout amongst the most conspicuous attributes of MOOCs is that, mostly because of the ease of enlistment, numerous students may calmly select in a course, peruse a couple recordings or discourse gatherings, and after that stop participation (Whitehill *et al*, 2015).

Student's academic performance is a significant factor in any educational institution, for that reason it could ensure strategic programms be planned in continuing enlightening or guiding the students for a better performance that may leads them to a better future (Quadri and Kalyankar, 2010).
According to Tasir *et al*, (2011), e-learning is an essential apparatus to help and encourage instructing and learning process. It gives the instruments to learners to be in contact with companions and educators outside the classroom. It additionally engages students to deal with their particular learning way and in the most suitable route for every student. Students learn in various ways that may involve perusing, watching, investigating, inquiring about, cooperating, conveying, working together, examining, and sharing information and encounters.

Various educational data mining (EDM) techniques have been used in the prediction of students' performance such as classification, regression and density estimation for predicting student's activities, and correlating the interactivity of students in the LMS environment. It is worth noting that the recent researches on EDM for students' performance prediction were primarily applied to cases of University of high school students.

With e-Learning, students can have access to an extensive variety of learning assets and learning can happen anyplace, at whatever time, and there are no longer any geological or geographical limitations to learning (Tasir *et al*, 2011). E-Learning has empowered the student so that the educator is no longer the gatekeeper of information. E-learning has strengthened the significance of casual learning and brought merging between learning, working, correspondence and entertainment (Andersson, 2010). With the advancement in e-learning, and being an e-service product makes interactivity to become more and more part and parcel of e-learning, it is possible to explore the learning processes of students from their interaction data within the e-learning system.

According to Wang *et al,* (2011), the concept “interactivity” means reciprocal activity between the learning man and the e-learning system. Action/reaction of the learning man depends on the action and reaction of the system itself. However online instruction turns out to be increasingly a vital instructive environment with the advance of Internet and PC innovation, and intuitiveness has been proclaimed by numerous as one of highlight of this innovation. Intelligence, as an apparatus of creating capacities and abilities of the under research, is absolutely a reasonable supplement inside an e-learning backing of training and it is an imperative advantage of mixed media electronic learning guideline.

For online training, intelligence can be given liveliness, recreations, sounds, recordings, and movies. It can likewise be utilized to show substance in a way that conventional showing materials can't, for instance, movements of components and procedures can help understudies envision how frameworks cooperate. An intuitive sight and sound module can outwardly fortify an under research and change learning into a dynamic drawing in process.

Interactivity of students with e-learning environment could be used to analyze the students’ behavior; that is to know the behavioral level and knowing that could enable us to understand whether the interactivity can affect the outcome or to predict the student performance.

1.3 Problem Statement

According to Shahiri and Hussain, (2015) predicting students’ performance in Malaysia becomes very challenging due to the huge volume of data and the lack of existing system to analyze and study the progress and performance of students. Their studies also suggest that using appropriate data mining techniques (such as Decision Tree, Neural Networks, Naïve Bayesian, K-Nearest Neighbor and SVM) the achievements and success of students could actually be improved more effectively.

The main question of this research is whether interactivity affects performance outcomes and how students implicitly interact with the learning materials in the e-Learning environment. The effect of interactivity in e-learning on the expansion in quality and capability with instructive process has been demonstrated by numerous authors (Moreno and Meyer, 2005). Regardless, in the meantime a few researchers indicate the potential confinements with a high level of interactivity of the framework does not really guarantee a high level of comprehension cognitive load.

Sanchez-Santillan *et al,* (2016) analyzes some datasheets obtained for two years from 2012 (111 students) to 2013 (84 students), using three different algorithms; Classification rules, Decision tree and Bayesian Networks respectively. The results were tested using cross validation which shows the classifiers (11 classifiers) accuracy for each model. The study revealed that in general there is not one single algorithm that obtains the best classification accuracy in all the three models.

Due to the computerized advancement, tertiary institutions and organizations are confronted with new opportunities and difficulties. MOOCs constitute an essential advancement in open training: MOOCs are another device in a computerized setting, and they speak to above and beyond from customary free open instruction. Free open education, from one perspective, experiences physical and topographical confinements, has low individual communication, and has low media assortment in educating assets. MOOCs, then again, are accessible to anybody with a web connection, regardless of the number of students taking the course or any physical or geological state of the college conveying the course (Ospina-Delgado and Zorio-Girima, 2016).

1.4 Research Question and Hypothesis

Rothe *et al,* (2014) while evaluating information systems that may support instructors and course designers discovered that the indicators used in measuring the interactivity is normally manual data collection. Rothe research work introduces new measures in evaluating online discussions; the interstactivity rate and contrast them with producers perspective and based on indicators used in evaluating online discussions comes up with question: How can data mining techniques be used to evaluate and analyze online discussions against the background of a complicated view of interactivity? Thus, the research intends to investigate whether interactivity could have an effect in the student performance of students in eLearning process positively or negatively. Therefore, this study intends to answer these questions:

i). Is the rate of active interactivity in e-learning and MOOC's affect the performance of part-time students?

ii). Which prediction technique is significant in predicting part-time student performance in e-learning and MOOCs?

iii). How does interactivity affect part time students' performance in MOOCs and e-learning?

The entire research work revolves around the hypothesis H. Which is a tangible, testable and measurable statement on the proposed research work. Thus,

H = "The rate of interactivity of part-time students with e-learning/MOOCs environments affect their performance which may lead to student dropping out or having low performance in the course".

1.5 Aim and Objectives of the Research

The aim of this dissertation is to predict student academic performance based on their interaction with e-learning management system and MOOCs using Neural Networks and Decision tree.

Shahiri and Hussain, (2015) indicated that the above techniques are proven to be highly used by various researchers for predicting students' performance. As such the research intend to apply the algorithms in the obtained data to know which technique outperform the other, the techniques will be applied to the same dataset to find out the technique that gives a better performance prediction.

The research is intended to achieve the following objectives:

i). To study and analyze the rate of interactivity of part-time students with e- learning and MOOC's environment, if it affects the students' performance.
ii). To predict the performance of part time students interacting with eLearning using Decision Tree (C4.5) and Neural Network techniques.

iii) To validate the impact of interactivity on the part-time students with e-learning environment on their academic performances.

The above objectives could be achieved by measuring the understandability of their learning and interactivity processes and couple with some factors (demography and working experience) through the use of data mining methods, tools and techniques.

1.6 Scope of the Research

The scopes of this dissertation are described below:

1. The boundary of this research work focuses on the significance interactivity with eLearning environment in MOODLE and MOOCs on students' learning behavior.

2. The data are obtained from log files upon interaction with the e-learning environment and MOOCs of UTMSPACE Program part-time students taking Intro. to Data Structure and Algorithm course SCSJ 2013, 2016/2017 Session.

3. MATLAB 2015 tool is used to implement the prediction model.

1.7 Significance of the Research

Predicting students that are vulnerable to dropping out from the part time program would be very important to the institution administration and the students themselves. For example, it could give an appropriate warning to students who may be at risk of failing or dropping by forecasting their grades, and offer them to avoid problems and overcome all difficulties in their study with the e-learning environment.

However, this research is aimed at predicting the part time student performance based on their interaction with e-learning/MOOC systems, the end result of the study is a prediction model that classify students based on some described attributes. The research result will help in understanding the behavior of part time students while interacting with e-learning/MOOCs environment and how it affects their performance. It can also be used in finding problems faced by part time students, and improves the students learning outcome. Furthermore, it will assist in knowing the significance of e-learning/MOOC in the learning process of part time students, and above all research findings could be used by the UTMSPACE program, to enable them have a view of the part time student's perception of the e-learning/MOOC system and know the comprehensibility of the part time students of e-learning and MOOC system in helping them acquiring knowledge.

REFERENCE

Agaoglu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. *IEEE Access, 4,* 2379-2387.

Agudo-Peregrina, Á. F., Hernández-García, Á., & Iglesias-Pradas, S. (2012). *Predicting academic performance with learning analytics in virtual learning environments: A comparative study of three interaction classifications.* Paper presented at the Computers in Education (SIIE), 2012 International Symposium on.

Aher, S. B., & Lobo, L. (2011). *Data mining in educational system using Weka.* Paper presented at the IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT).

Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data Mining: A prediction for Student's Performance Using Classification Method. *World Journal of Computer Application and Technology, 2*(2), 43-47.

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015). *Preprocessing and analyzing educational data set using X-API for improving student's performance.* Paper presented at the Applied Electrical Engineering and Computing Technologies (AEECT), 2015 IEEE Jordan Conference on.

Andersson, A. (2010). Learning e-learning: The restructuring of students' beliefs and assumptions about learning. *International Journal on E-learning, 9*(4), 435.

Arbaugh, J. B. (2014). "System, scholar or students? Which most influences online MBA course effectiveness?" Journal of computer assisted learning 30(4): 349-362.

Ashenafi, M. M., Riccardi, G., & Ronchetti, M. (2015). *Predicting students' final exam scores from their course activities.* Paper presented at the Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE.

Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*.

Bralić, A., & Divjak, B. (2016). USE OF MOOCS IN TRADITIONAL CLASSROOM: BLENDED LEARNING APPROACH.

Brinton, C. G., & Chiang, M. (2015). *Mooc performance prediction via clickstream data and social learning networks.* Paper presented at the Computer Communications (INFOCOM), 2015 IEEE Conference on.

Capovilla, D., Hubwieser, P., & Shah, P. (2016). *DiCS-Index: Predicting student performance in computer science by analyzing learning behaviors.* Paper presented at the Learning and Teaching in Computing and Engineering (LaTICE), 2016 International Conference on.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning, 4*(5-6), 318-331.

Cheewaprakobkit, P. (2013). *Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program.* Paper presented at the Proceedings of the International MultiConference of Engineers and Computer Scientists.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies, 10*(1), 17-29.

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review, 38*, 38-50.

Dawson, S., Bakharia, A., Lockyer, L., & Heathcote, E., (2011). "Seeing" networks: Visualising and evaluating student learning networks." A report prepared for the Australian Learning and Teaching Council funded project.

Dollinger, S. J., Matyja, A. M., & Huber, J. L., (2008). "Which factors best account for academic success: Those which college students can control or those they cannot?" Journal of research in Personality 42(4): 872-885.

Dráždilová, P., Obadi, G., Slaninová, K., Al-Dubaee, S., Martinovič, J., & Snášel, V. (2010). Computational intelligence methods for data analysis and mining of elearning activities. In *Computational intelligence for technology enhanced learning* (pp. 195-224): Springer.

El Gamal, A. (2013). An Educational Data Mining Model for Predicting Student Performance in Programming Course. *International Journal of Computer Applications, 70*(17).

Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting Student Performance Using Personalized Analytics. *Computer, 49*(4), 61-69.

Fernández-Delgado, M., Mucientes, M., Vázquez-Barreiros, B., & Lama, M. (2014). *Learning analytics for the prediction of the educational objectives achievement.* Paper presented at the Frontiers in Education Conference (FIE), 2014 IEEE.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68-84.

Guoqiang Zhang, B. Eddy Patuwo, Michael Y. Hu, Forecasting with artificial neural networks:: The state of the art, In International Journal of Forecasting, Volume 14, Issue 1, 1998, Pages 35-62, ISSN 0169-2070, https://doi.org/10.1016/S0169-2070(97)00044-7.

Hämäläinen, W., & Vinni, M. (2010). Classifiers for educational data mining. *Handbook of educational data mining*, 57-74.

Hu, Y.-H., Lo, C. -L, & Shih, S.-P, (2014). "Developing early warning systems to predict students' online learning performance." Computers in Human Behavior **36**: 469-478.

Jayaprakash, S. M., Moody, E. W., Lauría, E. JM, Regan J.R., Baron J. D. (2014). "Early alert of academically at-risk students: An open source analytics initiative." Journal of Learning Analytics **1**(1): 6-47.

Kabra, R., & Bichkar, R. (2011). Performance prediction of engineering students using decision trees. *International Journal of Computer Applications, 36*(11), 8-12.

Kamavisdar, P., Saluja, S., & Agrawal, S. (2013). A survey on image classification approaches and techniques. *International Journal of Advanced Research in Computer and Communication Engineering, 2*(1), 1005-1009.

Kember, D., Ying, C. K., Wan, C. S., Yung, C. S., Wai, C. T., Mui, C. Y., . . . Chi, L. W. (2005). How students cope with part-time study An analysis of coping mechanisms through an on-line forum. *Active Learning in Higher Education, 6*(3), 230-242.

Klimasauskas, C.C., 1991. Applying neural networks. Part 3: Training a neural network, PC-AI, May/ June, 20–24.

Kovanović, V. Gašević, D., Dawson S., Joksimović S., Baker R. S., & Hatala M. (2015). Penetrating the black box of time-on-task estimation. Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, ACM.

Kovacic, Z. J. (2012). Predicting student success by mining enrolment data. *Research in Higher Education Journal, 15*, 1.

Lauría, E. J., Baron, J. D., Devireddy, M., Sundararaju, V., & Jayaprakash, S. M (2012). Mining academic data to improve college student retention: An open source perspective. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM.

MacCann, C., Fogarty, G. J., & Roberts, R. D. (2012). Strategies for success in education: Time management is more important for part-time than full-time community college students. *Learning and Individual Differences, 22*(5), 618-623.

Márquez-Vera, C., Cano, A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied intelligence, 38*(3), 315-330.

Meier, Y., Xu, J., Atan, O., & van der Schaar, M. (2016). Predicting grades. *IEEE Transactions on Signal Processing, 64*(4), 959-972.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of educational psychology, 97*(1), 117.

Moro-Egido, A. I., & Panades, J. (2010). An analysis of student satisfaction: Full-time vs. part-time students. *Social Indicators Research, 96*(2), 363-378.

Morris, L. V., Finnegan, C., & Wu, S. -S., (2005). "Tracking student behavior, persistence, and achievement in online courses." The Internet and Higher Education **8**(3): 221-231.

Nandi, D., Hamilton, M., Harland, J., & Warburton, G., (2011). How active are students in online discussion forums? Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114, Australian Computer Society, Inc.

Ospina-Delgado, J., & Zorio-Grima, A. (2016). Innovation at universities: A fuzzy-set approach for MOOC-intensiveness. *Journal of Business Research, 69*(4), 1325-1328.

Pistilli, M. D., Arnold, K., & Bethune, M. (2012). Signals: Using academic analytics to promote student success. *EDUCAUSE Review Online*.

Quadri, M. M., & Kalyankar, N. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology, 10*(2).

Rafaeli, S. & G. Ravid (1997). Online, web based learning environment for an information systems course: Access logs, linearity and performance. ISECON.

Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence, 32*(3), 569-575.

Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, *68*, 458-472.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications, 33*(1), 135-146.

Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601-618.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education, 51*(1), 368-384.

Rothe, H., Sundermeier, J., & Gersch, M. (2014). *Analyzing interactivity in asynchronous video discussions.* Paper presented at the International Conference on Learning and Collaboration Technologies.

Ruffaldi, E., Dabisias, G., Landolfi, L., & Spikol, D. (2016). *Data collection and processing for a multimodal learning analytic system.* Paper presented at the SAI Computing Conference (SAI), 2016.

Sanchez-Santillan, M., Paule-Ruiz, M., Cerezo, R., & Nuñez, J. (2016). *Predicting Students' Performance: Incremental Interaction Classifiers.* Paper presented at the Proceedings of the Third (2016) ACM Conference on Learning@ Scale.

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72*, 414-422.

Shukor, N. A., Tasir, Z., & Van der Meijden, H. (2015). An examination of online learning effectiveness using data mining. *Procedia-Social and Behavioral Sciences, 172*, 555-562.

Siemens, G. (2012). *Learning analytics: envisioning a research discipline and a domain of practice.* Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.

Sorour, S. E., El Rahman, S. A., & Mine, T. (2016). *Teacher interventions to enhance the quality of student comments and their effect on prediction performance.* Paper presented at the Frontiers in Education Conference (FIE), 2016 IEEE.

Sun, P.-C., Tsai, R. J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education, 50*(4), 1183-1202.

Sunar, A. S., Abdullah, N. A., White, S., & Davis, H. C. (2015). Analysing and predicting recurrent interactions among learners during online discussions in a mooc.

Tair, M. M. A., & El-Halees, A. M. (2012). Mining educational data to improve students' performance: a case study. *International Journal of Information, 2*(2).

Tasir, Z., Al-Dheleai, Y. M. H., Harun, J., & Shukor, N. A. (2011). *Student's Perception towards the Use of Social Networking as an e-Learning Platform.* Paper presented at the 10th WSEAS International Conference on Education and Educational Technology, Penang: Malaysia.

Vahdat, M., Ghio, A., Oneto, L., Anguita, D., Funk, M., & Rauterberg, M. (2015). Advances in learning analytics and educational data mining. *Proc. of ESANN2015*, 297-306.

Van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE learning initiative, 1*, 1-11.

Violante, M. G., & Vezzetti, E. (2015a). Design and implementation of 3D Web-based interactive medical devices for educational purposes. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1-14.

Violante, M. G., & Vezzetti, E. (2015b). Virtual interactive e-learning application: An evaluation of the student satisfaction. *Computer Applications in Engineering Education*, *23*(1), 72-91.

Wang, P.-Y., Vaughn, B. K., & Liu, M. (2011). The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education, 56*(1), 300-311.

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann.

Wu, J.-Y. (2011). MIMO CMAC neural network classifier for solving classification problems. *Applied Soft Computing, 11*(2), 2326-2333.

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior, 58*, 119-129.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming:

Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168-181.

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Data mining applications: A comparative study for predicting student's performance. *arXiv preprint arXiv:1202.4815*.

Zacharis, N. Z. (2015). "A multivariate approach to predicting student outcomes in web-enabled blended learning courses." The Internet and Higher Education **27**: 44-53.

Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*: Cambridge University Press.

Zhou, M. (2016). Chinese university students' acceptance of MOOCs: A self-determination perspective. *Computers & Education, 92*, 194-203.

Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J. (2005). *Web usage mining project for improving web-based learning sites.* Paper presented at the International Conference on Computer Aided Systems Theory.