International Journal of Engineering & Technology, 7 (2.29) (2018) 1113-1120

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Malaysia MOOC: Improving Low Student Retention with Predictive Analytics

Nadirah Mohamad[1]*, Nor Bahiah Ahmad[2], Dayang Norhayati Abang Jawawi[3]

[1,2,3]Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Skudai Johor, Malaysia Author a, Author b, Author c
**Corresponding author E-mail:* nadirah23@live.utm.my

Abstract

Massive Open Online Courses MOOCs have become more acceptable as a learning program globally, including Malaysia. One main issue that has been discussed since the implementation of MOOCs is the issue of low student retention or high dropout rates from the course. Various factors have been found to play a role in this issue including the interaction factor. Previous studies have experimented with various strategies to monitor student retention and apply intervention programs to improve the situation. The strategies include the usage of machine learning and data mining techniques in analysing students' online interactions to predict student retention rates. The implementation of these strategies produced promising result. However, in Malaysia, these strategies are not really implemented yet. Therefore, this paper discusses the issue of student retention in MOOCs, explores possible intervention plans using data mining and its suitability with the current platforms used for MOOCs. The proposed method includes predictive analytics that involves classification analysis. This paper suggests that the method can be applied to the current platform and complement intervention programs for the issue of low retention or high dropouts with several improvements.

Keywords: *Massive Open Online Course MOOC; Student Retention; Student's Online Interaction; Predictive Analytics*

1. Introduction

Since the World Wide Web came into existence, people turn to the Internet to find answers to their questions. The powerful capability of websites in providing immediate and convincing answers to almost all questions attracts people to revisit the Internet. With more people sharing information, more open and free resources are available online. Many providers and institutions have launched their open courses which start from small class initiatives to larger scale open online courses offered to the public. Nowadays, Massive Open Online Courses MOOCs have become more acceptable as a learning program globally, including in Malaysia. Previous studies on MOOCs have done research in various areas. These areas can be categorized into case studies, learning theories, business models, target groups survey and opinions, assessments and design (1).

However, despite the increase in research interest and MOOC participants, the main issue that is discussed is the low student retention or high dropout rate (2). This issue has been mentioned in almost all previous studies on MOOC. Henek (3) in his study stated that on average, only 10% of students completed a course. This report is similar to Gomez-Zermeno and De La Garza (4) who found that 86% of their study sample were not retained in MOOCs. The low retention is caused by many factors including interaction, which plays a great role. Several suggestions have been proposed previously which include improving interaction, whether social interaction interaction with peers and instructor or the student's interaction with the content MOOC (5); (6). In terms of social interaction, most of the instructors have other commitments that are of greater priority (7). Moreover, in several institutions, MOOC is a medium that acts more as aided learning. This is because face-to-face or conventional learning still plays the major role and MOOC is used for blended learning. Therefore, maintaining interaction with students at all times is challenging.

However, this gap can be filled with a strategy that helps the instructor to monitor and detect low retention students and apply intervention programs in order to improve the situation. Previous studies have experimented with various strategies to predict student retention. These strategies include predictive analytics that use data mining techniques to analyse a student's online interaction. The implementation of this strategy produced promising results. Also, the analytics results is important especially in the early stages as the instructor can motivate the students to improve their status by increasing their interaction Tseng, Tsao (8). The need for such results is also mentioned by Fadzil, Latif (9) who stated that knowing the student's completion progress and dropout status will help to provide a clear view of MOOCs as well as contribute towards the development of national policies and plans. However, the implementation of the analytics program, particularly predictive analytics, has not been really emphasised in Malaysia (10).

Therefore, the questions that this study investigated from the literature are whether or not predictive analytics have been implemented in other countries, what are the current plans on the implementation of such programs and if it is possible for it to be implemented in Malaysia MOOCs. Malaysia MOOC is an initiative by Malaysia government to offer open courses to students which the content developed by public universities. The questions on predictive analytics implementation will be answered throughout this paper by explanation, discussion and recommendations. Beginning with the next section, this paper discusses the importance of predictive analytics and its role in intervention programs in Malaysia. This is followed by the discussion on examples of implementation in Malaysia MOOC with several suggestions. The study then ends with the conclusion.

2. Predictive Analytics

Predictive analytics are commonly found under the context of learning analytics and educational data mining studies (11). Over the past few years, the term predictive learning analytics or PLA has started to emerge. The task is defined as "technology that learns from experience data to predict the future behaviour of individuals in order to drive better decisions" (12). This section will explain the benefits of predictive analytics and the role of predictive analytics in intervention programs.

2.1. Benefits of Predictive Analytics

Predictive analytics are implemented for various reasons such as to predict student dropout, student retention rate when using MOOCs, predict performance whether or not the student achieves the correct answer on the first attempt, and to predict the student's grade (13); (14); (15); (16). From previous research, it is clear that many areas in MOOC can be improved with predictive analytics. Next, the study will explore the potential that predictive analytics has in MOOCs.

2.1.1. Improve the Teaching and Learning Process

The results from the analytics will be beneficial to the entire teaching and learning process by its impact on participant satisfaction and course quality. When the instructor helps students based on the analytics results, the students would show improvement and become more active, hence interacting more. Then, the learning process would become livelier. This can encourage a supportive environment and increase positive feedback about student satisfaction during the teaching and learning process.

2.1.2. Identify Students Who Are Left Behind

During the teaching and learning process, identifying a student's progress and participation status, or the whole class performance just by looking at the log file is challenging. This is because the files are unprocessed and analysis needs to be done first to extract the meaningful information. This includes information about whether the students are actively interacting or have been left behind (17). This information is available because it is needed to predict the student's next status or progress. After identifying the students who are left behind, the instructor can decide on the best possible solution, including encouraging the student to become more active by increasing their interaction levels. This would enable assistance and feedback to be given to them almost immediately, thus preventing them from feeling isolated.

2.1.3. Identify Weakness of Components or Materials

Apart from improving the component of students, the analytics can help in monitoring course content and materials (18); (19). For example, to predict material of poor quality, the analytic produces result of user's access to each topic and subtopic of materials. From the results, the instructor will be able to identify topics that need to be investigated, based on the low hits of access to that topic. If there are problems such as redundant or unreadable videos, the instructor can improve that section. Or, if a student views a certain video multiple times, it could be that the video is interesting or that the student could not understand the content of the video. Other than that, the instructor will know if certain questions need to be reviewed when a high percentage of students could not answer the question.

2.1.4. Provide Better Services than Other Facilities

It is an advantage for a course provider to provide a facility like predictive analytics, not only for students and instructors to exchange knowledge but also for course rankings when compared to other courses (20). With the tools provided by analytics, students are able to know their progress as well as predicted status and the instructors are able to know when to intervene compared to other courses that can't provide this information. The tools help to increase student satisfaction levels with regards to course effectiveness.

The benefits explained previously are to increase the number of students who complete the course, achieving certification or according to the targeted milestone for the course. From a different perspective, a course with many completers and achievers present a positive image and indirectly promote the course to other students. Simply put, the image reflects the course quality.

2.2. Predictive Analytics in Intervention Programs

Analytic tools which are used in intervention programs have been applied in many other countries for their e-learning courses (21). The same also applies to MOOCs which have been conducted by many research studies such as (16); (14); and (22). However, there is a lack of research in predicting retention or dropouts among students for intervention programs in Malaysia. Within our literature review, there is one study that proposed an intelligent automatic analytics to predict and recommend action (21). However, this study is designed for the e-learning of an institution, which focuses on the performance of the institution.

Several other studies which have investigated intervention are for the traditional mode of learning face-to-face or for brick-and-mortar schools (23); (24); (25). The prevention programs can be said to involve several stages including identifying the factors and grouping the students based on their behaviour to identify focus groups as well as the program to be chosen according to the student's situation. However, the programs that were implemented for traditional modes of learning addressed the issue of dropouts because of the serious consequences present such as theft, robbery and substance abuse (25). This is in contrast with MOOC as MOOC itself is not compulsory and is just a medium to aid learning. A heavy penalty would not be suitable for the identified focus group. However, the penalty component can be considered by implementing point reductions for hint requests during problem solving tasks. The need of such program in MOOC research is still important and it can provide input in developing a model. Therefore, to set up a virtual or online course, several points can be considered.

Generally, intervention programs for MOOC can be divided into instructional and technological programs (26). A large proportion of previous MOOC research emphasizes more on instructional programs compared to technological programs. However, nowadays, as technology evolves, there are more studies about intervention programs utilizing technology such as using predictive analytics, which is automated and gives more efficient results. An analytics task needs to apply appropriate data mining and machine learning techniques. From a global point of view on MOOC predictive analytics, the studies can be used as guideline in developing a model for implementation as shown in Table 1.

From Table 1, the studies according to group project analyse data from various platforms with a goal to solve the issue of low retention, high dropout or student at risk. The works include a study from Sunar, White (27) who investigated the dataset of the FutureLearn platform to predict interaction, mainly social interaction, with plans on continuing the project with the recommended system. The other study worth mentioning is (22) which was continued by Taylor, Veeramachaneni (28). They conducted an extensive study from data collection involving feature engineering to the implementation of predictive analytics with plans towards developing open source software. The latest work also provides extensive research and can be a good reference like Whitehill,

Mohan (16) for predicting dropout and Yang, Brinton (15) for predicting a student's grade.
The study concluded that predictive analytics play an important role in supporting and executing intervention programs for MOOCs. The tasks included in preparing for intervention programs are investigating the factors that lead to low student retention, identifying target groups and performing analytics which include predictive analytics with the appropriate tools and data mining techniques applied. Then, the necessary actions can be recommended to prevent students from dropping out or to improve a student's low grades during the learning process in MOOC. Even though most studies are still under development and implemented as a stand-alone tool, the studies can be improved and subsequently used to outline and develop a model or framework for more efficient embedded or plug-in analytic tools. The process model for predictive analytics is concluded and explained in the next section along with an example of implementation in a Malaysia MOOC course.

Table 1: Previous Studies on Predictive Analytic or Model

Authors	University	Prediction goal	Process identified
(16); (29)	Open edX HarvardX, U.S	Dropout	Data collection, data pre-processing, feature selection, weekly classification, prediction, intervention
(15)	Coursera, U.S	Performance Grade	Data collection, data pre-processing, partitioning students, weekly classification, prediction
(30)	Open edX xuetangX, China	Dropout Will do activities or not	Data collection, data pre-processing, weekly classification, prediction
(14)	iMooX, Austria	Retention	Data collection, data pre-processing, partitioning students, weekly classification, prediction.
(27)	FutureLearn, U.K	Recurrent interaction	Data collection, data pre-processing, feature selection, partitioning students, weekly classification, prediction
(31)	not mentioned from UCLA	Performance Pass or fail grade	Data collection, data pre-processing, weekly classification, prediction.
(13)	Coursera, U.S	Performance Correct on first attempt	Data collection, data pre-processing, partitioning students, weekly classification, prediction.
(22, 28)	edX, U.S	Dropout Stopout	Data collection, data pre-processing, feature selection, partitioning students, weekly classification, prediction.

3. Implementation and Discussion

This section discusses the possibility of implementing predictive analytics in the current Malaysia MOOC platforms with several areas of improvement proposed. Before the discussion, Table 2 shows the courses created by several public universities in Malaysia on the provider's platform partnership with OpenLearning, Sydney, Australia since 2014 which have reached 204 courses. The statistics show the government's initiative in implementing MOOC at the national level. Also, the high number of students registered an average of 1584 students per course, shows that the students are interested to participate in the courses offered.

Table 2: Malaysia MOOC with Number of Courses Offered www.openlearning.com

No	Platform	Total courses created	Average students
1	Universiti Kebangsaan Malaysia UKM	65	1826
2	Universiti Teknologi Mara UiTM	15	2392
3	Universiti Teknologi Malaysia UTM	14	439
4	Universiti Malaysia Sarawak UNIMAS	13	1887
5	Universiti Malaysia Pahang UMP	12	338
6	Universiti Putra Malaysia UPM	9	11104
7	Universiti Sultan Zainal Abidin UniSZA	9	152152
8	Universiti Malaysia Terengganu UMT	8	762
9	Universiti Teknikal Malaysia Melaka UTeM	8	522
10	Universiti Utara Malaysia UUM	8	480
11	Universiti Pertahanan Nasional Malaysia UPNM	7	121
12	Universiti Malaysia Sabah UMS	7	81
13	Universiti Sains Islam Malaysia USIM	6	694
14	Universiti Malaysia Perlis UniMAP	6	148
15	Universiti Sians Malaysia USM	5	77
16	International Islamic University of Malaysia IIUM	4	330
17	Uniersiti Tun Hussein Onn Malaysia UTHM	3	287
18	Universiti Malaysia Kelantan UMK	3	68
19	University of Malaya UM	2	288

From the next subsection, this study has provided a brief explanation on the process involved in implementing predictive analytics as discussed in Section 2 and shown in Figure 1. The process is focused more on predicting dropout or retention. Figure 1 shows the process involved which includes data collection, data pre-processing for feature construction, feature selection, partitioning students, weekly student status and applying student prediction with predictive analytics. Finally, the appropriate intervention program can be executed.
The data used in this example of implementation is from the Data Structures and Algorithms course with 652 students for the first cohort. For data collection in preparation for constructing features, the process depends on the prediction goal which has been discussed in Section 2. Therefore, instead of data collection, the type of data will be explained in more details. For more insight on the process of collecting the features, studies like Veeramachaneni, Halawa (22), Hone and El Said (32) and Kloft, Stiehler (33) are great references. For this study, the features that are available pertain to content, forum and task.

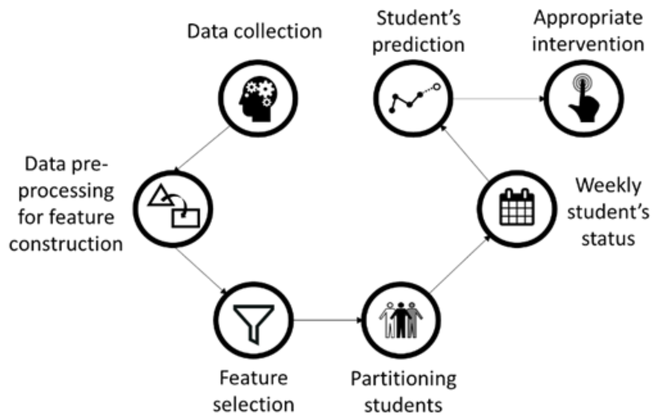

Fig. 1: Process Involved in Implementing Predictive Analytics

3.1. Type of Data

Generally, online learning consists of simple to complex data. The data contains attributes for information like demographics, events and scores which range from a few to hundreds of attributes. For MOOCs alone, there are various platforms for people to choose from like edX, Coursera and Udacity (34). These platforms generate data in form of tracking log data which are extracted into a JSON format and transformed into CSV files. There are many attributes that can be extracted but the most commonly extracted from the files are student id, timestamp, events and the resource accessed like in the studies of Whitehill, Mohan (16) and Khalil and Ebner (14). Figure 2 shows an example of log file from a Malaysia MOOC course which can be accessed by administrators or the instructor.

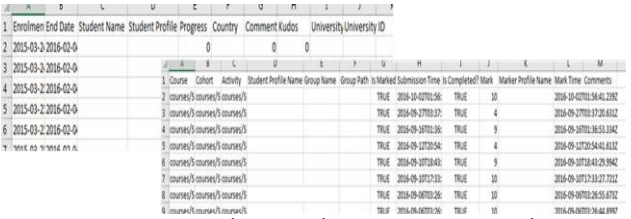

Fig 2: Example of Log File from OpenLearning Platform

From Figure 1, the attributes available are enrolment date, end date, student information like ID, country and university, progress, number of kudos likes, comments from student posts and active time updated. For student information, university information is used as Malaysia MOOCs focus more on blended learning (35). This information is valuable in easily identifying and monitoring the on-campus students. Meanwhile, for an assignment or task, the log file contains attributes like submission data and mark time, comments given, activity assignment, task, quiz or the topic involved, mark status and the mark given. The files provide meaningful insight into a student's background, their progress and marks. However, compared to the common attributes extracted in predictive analytics, several attributes can be added such as events with their resources, information and timestamps. Those attributes are needed because they represent the interaction which the students have made with the system the MOOC. All interaction hits by students are recorded as 'event' attributes, with a 'timestamp' that shows when the hits clicked and 'resource' which represents the activity that was involved during the hits. Therefore, providing event, resource and time attributes for any prediction of dropout and retention is crucial. Limited data is also mentioned in Gunawardena (36) as one of the challenges faced by instructors. Additional attributes like 'success' is needed to predict the student's next move or correctness in an exam or quiz. These data then are prepared for pre-processing to construct features which are needed for the predictive analytics.

3.2. Pre-processing and Statistical Analysis

Maintaining and processing the raw data from MOOCs into files is challenging and needs a lot of effort because of the high space, time and cost required (37); (28). However, from previous studies, data pre-processing is rarely discussed in depth. The MOOC db project has been deployed by Veeramachaneni, Dernoncourt (38) to minimize this effort by developing a standard scheme for all MOOCs which contribute to data sharing where at the end, the file can be accessed in form of a csv file. The challenge becomes apparent when the data is transformed from the CSV files into the ready dataset. To make them into a ready dataset, the common attributes discussed in previous section are extracted to proceed with feature construction that combines the total of each event and is based on timestamp information for every student. The time slice is determined, using which the data is divided into day or weeks, according to the study. This becomes more challenging for massive data from many students. Several issues during this phase have been discussed more in our previous study (39) which also involves the feature construction that will be used for predictive analytics.

3.3. Analytics

Typically, the analytics that are available in MOOC is general analysis and student statistics as shown in Figure 3. Figure 3 shows the 'analytics tool' through which the instructor can view the course current progress such as how many pages are created, how many comments posted and how many students have registered. The tools also enable the instructor to view the percentage and number of students who have or have not viewed each content, task and quiz uploaded. The pages also show total number of view and time spent for each page created. Other than that, a heat map is available to show the varying number of registered students represent by colours for each countries. Apart from that, there is the 'students tool' shown in Figure 4 provided for more detailed analytics on each student, on whether he or she has completed certain tasks or not, and has completed viewing the content or not.

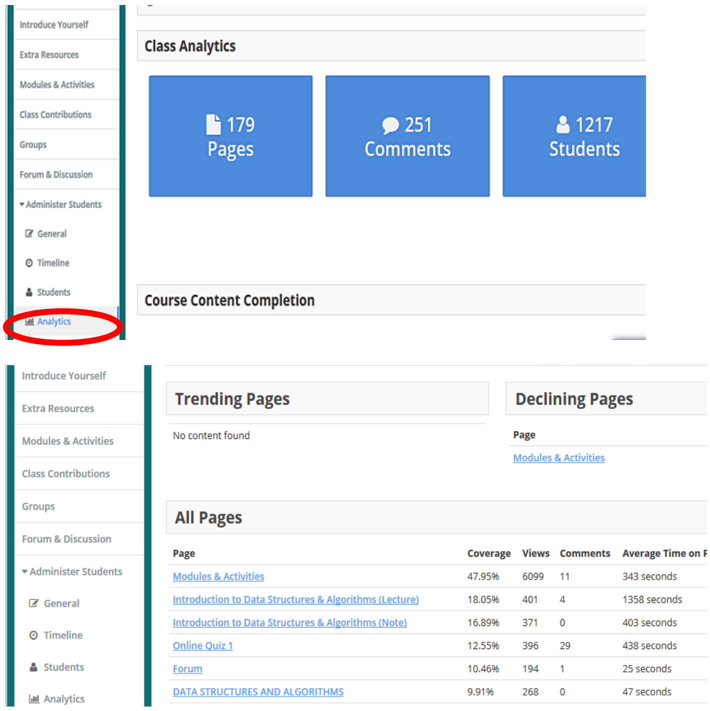

Fig 3: Analytics Tool Available to View Whole Class Statistics and Progress

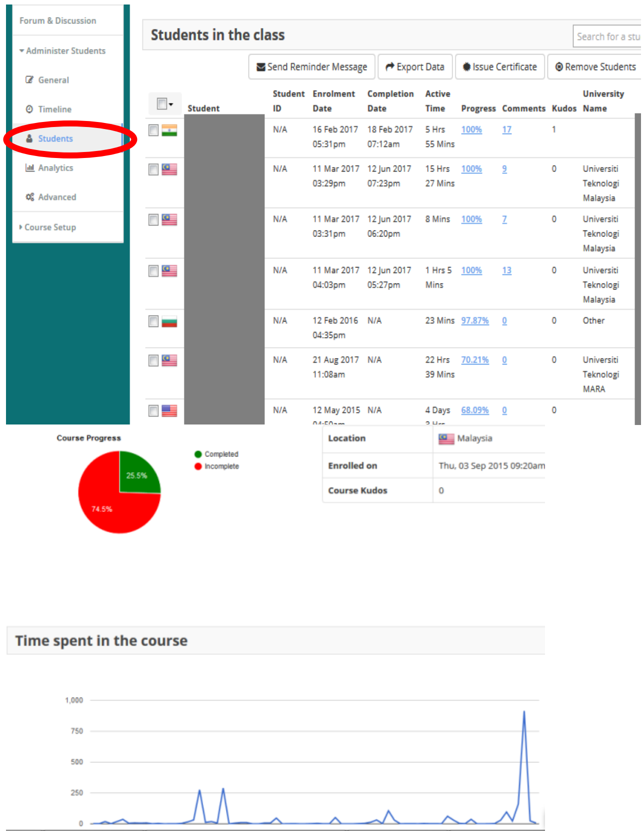

Fig 4: Analytics Tool Available to View Each Student's Progress

One of the processes that are generally done after having a dataset with selected features is partitioning the students. Partitioning students/learners mentioned is discussed in Taylor, Veeramachaneni (28) and appears to be used by others like Khalil and Ebner (14) and Brinton and Chiang (13). Most studies did not mentioned the techniques used in partitioning the students but the studies about student retention like Khalil and Ebner (14) developed a model to divide the student according to their own groups. The objective of the process is to ease the monitoring process for students in different groups with different learning patterns which may require different intervention programs. Figure 5 shows an example of students in the whole group, grouped by their level of interaction that can show their level of retention. This figure is the result from analysing available data from the 'students tool' using the data generated in form of log files and combined with the quiz and forum data shown in Figure 2. The groups are clustered using the self-organizing map technique, an unsupervised classification technique with many advantages including the ability to handle and group various sizes of data (40). This technology is also able to handle high dimensional data which suits the MOOC data which has many features. From the results, the students can be divided into groups of very low, low, medium and high retention rates. The results also show that many students have high access to content especially students with very low retention levels. This is probably because they prefer activities like reading and watching videos. Meanwhile, students with high retention have low interaction levels with the forum compared to content and task while students with medium retention rates have high interaction levels with the content compare to forum and task. The result shows that the students have various learning preferences that need different strategies for intervention.

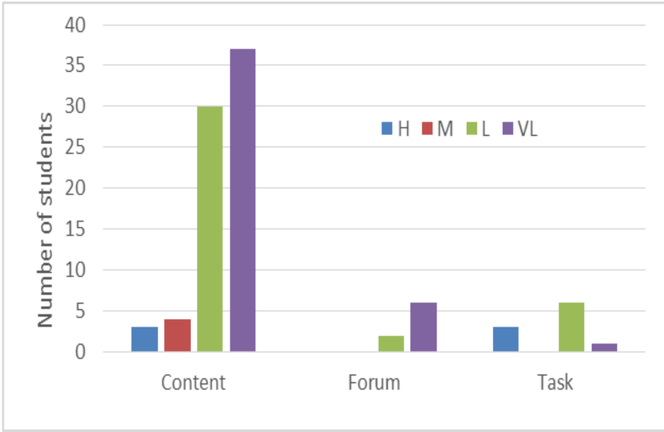

Fig 5: Retention Rates According to Activities for the First Cohort

The next process is weekly classification. Time delineation by week is the most common practice used in previous studies like Taylor, Veeramachaneni (28). Instead of results that analyse the total duration shown in Figure 5, the analysis that generates each week's result is needed for the weekly classification process as shown in Figure 6. The analysis also focuses more on the target group such as the group of students who have low interactivity levels. Figure 7 shows an example of feature construction for dropout prediction using data from edX platform. 1 means the student has done at least one interaction for each feature X1, X2, X3… and for each week W1, W2, W3…. The feature values depend on the analysis of time, events and resources information.

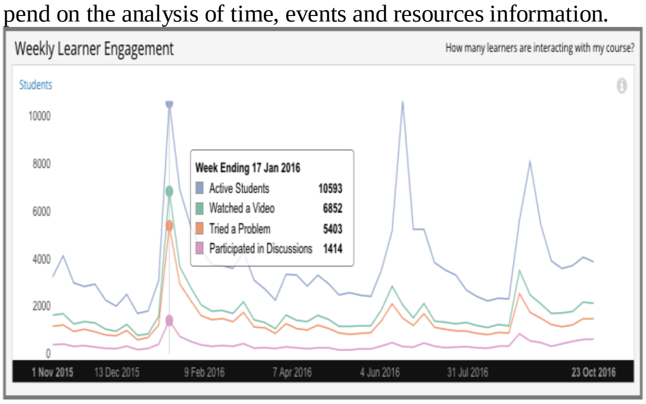

Fig 6: Example of weekly student engagement visualization open.edx.org

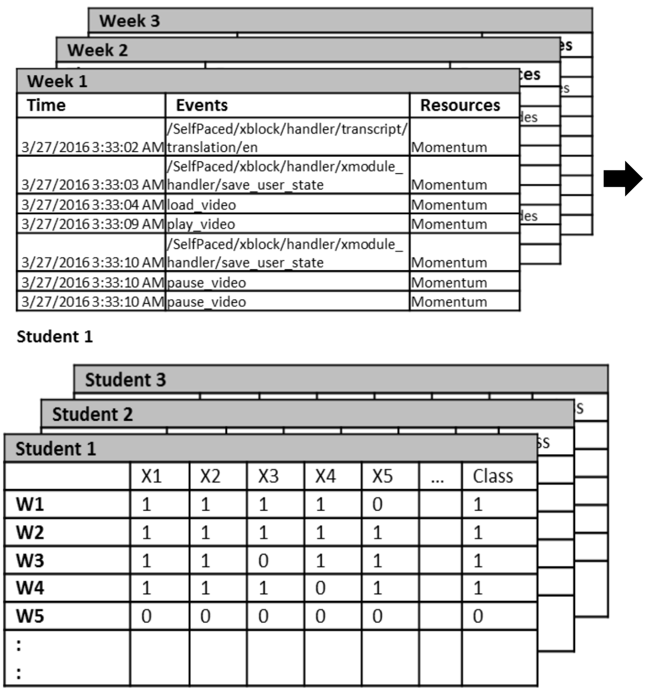

Fig 7: Example of weekly status for each student

A good prediction depends on reliable factors with a sufficient dataset. This means that apart from having good features, the analyser needs to determine the right predictors according to the case study. For example, in predicting student probability to drop out of a course, several studies use assignment submission status or the last activities occurred that define the dropout status. Meanwhile, for studies that predict student success in a quiz or exam, the student's previous score is used. For studies that focus on predicting student retention, an example of the input value is shown in Table 3. Table 3 shows ten students and their learning process over the course of ten weeks. In this example, the last interaction is used as a predictor. Based on average week of dropout, W5 determines whether the students will retain or not. If the student has interaction 1 until W5 and no more interaction 0 after the coming week, the student will be labelled with 'N'. On the other hand, if the student retains after the fifth week or dropout, the student is labelled with 'Y'. This is an example for 'lead and lag' which refers to the Veeramachaneni, Halawa (22) study.

Table 3: Example of Retention Status for Ten Students

ID	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	Class
1	1	1	1	1	0	0	0	0	0	0	N
2	1	1	1	0	0	0	0	0	0	0	N
3	1	0	1	1	1	0	1	0	1	1	Y
4	1	1	0	1	0	0	0	0	0	0	N
5	1	1	0	0	1	1	0	1	0	0	Y
6	1	1	1	0	1	1	1	1	0	1	Y
7	1	1	0	0	0	0	0	0	0	0	N
8	1	1	0	1	1	0	0	0	0	0	N
9	1	1	1	0	0	0	0	0	0	0	N
10	1	1	0	0	0	0	0	0	0	0	Y

Typically, the dataset is divided into training and test dataset for task prediction. There are various techniques applied previously that can be summarized into regression, classification or latent knowledge estimation (41). Earlier studies experimented with techniques like Hidden Markov Model (42), Support Vector Machine (33), and n-gram techniques (43). Later on, there were more studies that applied logistic regression in their case studies. However, recent studies show there are better result with techniques like decision tree or by improving the logistic regression technique itself (44); (16). The reason behind logistic regression is because its advantage in the binary predictive model that is mostly helpful for data with classes like dropout or not dropout, pass or fail.

After the technique is applied, the results can be visualized through graphics, for example by using line graphs or pie charts as shown in Figure 8. Simple but informative visualization is very helpful for instructors to immediately grasp the significance of the student's current and future progress. There are other online learning systems that already embed or plug-in analytic tools into their systems which can become a model or guideline for future study and in actualizing the predictive analytics. The system includes Knewton a K12 online learning that claims to have a 47% reduction in withdrawal rates which predicts score for particular exams and can even provide suggestions for the next action (36). Other examples include Noel-Levitz, Desire2Learn Insights and Blackboard analytics (45).

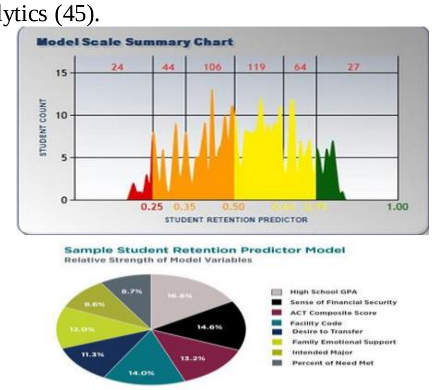

Fig 8: Example of Report from Noel-Levitz Prediction Analytics

After the result is generated, the instructor may provide the appropriate action for the intervention program. Examples of intervention programs include; sending private messages about the student's progress, email intervention or status reminders on how many activities are not complete. In conclusion, there are possibilities for the model to be applied to existing platforms. The initial process until partitioning of students is achievable. As the Malaysia MOOCs are still in their infancy, the teams are in the right phase for predictive analytics processes and implementing intervention programs, not only to mitigate the low retention rates and drop outs, but to also help on course development and improve the MOOC learning environment as a whole unit.

4. Conclusion

In conclusion, this study has emphasized the importance of predictive analytics and its application in Malaysia as well as outside of Malaysia. The processes involved in predictive analytics including the relevant techniques have been presented in this paper as a guideline to be applied in the local context. There are several policy implications that can be outlined from this study. These can be divided into short-term, medium term and long term policy implications. In order to implement the usage of automated predictive analytics tools in Malaysia MOOCs, short term policy implications include extensive research and development at the levels of institution and provider which includes;

i) The improvements in technique especially for partitioning students. This requires more exploration as this part of analytics is still unpopular yet has benefits for prediction and complements the next process to complete the prediction tasks. Suitable techniques like clustering and classification can be investigated further.

ii) The provision of standardized schemes in generating the data to complement the process of predictive analytics is required because different institutions might produce different strategies in constructing data. Such schemes can improve the process when developing the automated analytics tool. In addition, protecting the user's privacy is always priority. Therefore, developing policies about releasing such protected data de-identification based on request is also crucial.

iii) There are various goals of prediction task such as predicting questions or predicting a student's favourite materials. For this reason, the prediction task requires different strategies. Therefore, a guideline for prediction tasks that addresses different goals need to be developed.

The results from the studies are important in developing a better predictive analytics tool. In order to execute the development of the tool, the medium term policy implication is proposed;

i) The predictive analytics tool needs to be developed based on the needs of the system and users. Therefore, surveys on the usage of MOOCs and on the needs of analytic tools are required to be collected and analysed from users of MOOCs.

ii) The consent and cooperation from providers is also important to be acquired in order to execute the project in developing the tool, especially in the collaboration between the learning institutions with the MOOC providers as both entities may have different plans and expectations.

Lastly, for long term policy implications, several recommendations need to be considered;

i) The cooperation and mutual support from institutions and the government need to be established in order to widely use the MOOC itself whether the institution's role in practicing the MOOC as a learning medium in class or the government's role in encouraging the implementation at national level. Especially in this digital era where artificial intelligence is a constant topic, the application of intelligence tools for online learning is timely executed.

ii) Active promotion and campaigns need to be launched to encourage greater usage of MOOCs and maintain contact with students to strengthen relationships whether online such as through Twitter and Facebook or offline such as through symposiums and conferences.

iii) Institutions must observe and maintain their good image by providing quality content, accredited courses creators and provide timely feedback when deemed possible. Also, after the predictive tool is embedded into the MOOC system, the tool is not only for display but needs to be fully utilized by users to monitor and improve their learning performance.

Those are the areas which can be covered and implemented in the future which this study believes would encourage the usage of Malaysia MOOCs and mitigate the issues of drop outs. The tools developed benefit the users of MOOCs where the tools can minimize downtime by detecting defective components of modules like error in contents or questions and improve the component immediately. With automatic prediction, the students can perform better by keeping track of their performance status, whether on par with peers or belonging to the group who are left behind. More importantly, Malaysia MOOCs can become better at meeting the government's aspirations whether in providing quality education or for life-long learning regardless of the student's background. The implications of this also enable the MOOC team to build trust to retain students and have a better image to attract new students to register for the courses offered in the future.

Acknowledgements

The authors would like to express our deepest gratitude to Research Management Centre RMC, Universiti Teknologi Malaysia UTM for the support in R&D FRGS, 4F496; Tier 1, 17H75 and UTM Big Data Centre, Soft Computing Research Group SCRG for the inspiration in making this study a success.

References

[1] Yousef AMF, Chatti MA, Schroeder U, andHarald Jakobs MW, editors. A Review of the State-of-the-Art. Proceedings of CSEDU2014, 6th International Conference on Computer Supported Education; 2014.
[2] Ejreaw AM, Drus SM. The challenges of massive open online courses (mooc)–a preliminary review. Proceedings of the 6th International Conference on Computing and Informatics. 2017:7.
[3] Henek O. MOOC Retention Rate and Motivation. SDU: University of Southern Denmark, Kolding; 2017.
[4] Gomez-Zermeno MG, De La Garza LA. Research analysis on mooc course dropout and retention rates. Turkish Online Journal of Distance Education. 2016.
[5] Guo Z, Xiao L, Van Toorn C, Lai Y, Seo C. Promoting online learners' continuance intention: an integrated flow framework. Information & Management. 2016;53(2):279-95.
[6] Hew KF. Promoting engagement in online courses: What strategies can we learn from three highly rated MOOCS. British Journal of Educational Technology. 2016;47(2):320-41.
[7] Ghazali NB, Nordin MS. The Perception of University Lecturers of Teaching and Learning In Massive Open Online Courses (MOOCs). Journal of Personalized Learning. 2017;2(1):52-7.
[8] Tseng S-F, Tsao Y-W, Yu L-C, Chan C-L, Lai KR. Who will pass? Analyzing learner behaviors in MOOCs. Research and Practice in Technology Enhanced Learning. 2016;11(1):8.
[9] Fadzil M, Latif LA, Azzman TAMTM. MOOCs in Malaysia: a preliminary case study. E-Asem Forum: Renewing The Lifelong Learning Agenda For The Future. 2015.
[10] Shahiri AM, Husain W. A review on predicting student's performance using data mining techniques. Procedia Computer Science. 2015;72:414-22.
[11] Gašević D, Dawson S, Rogers T, Gasevic D. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. The Internet and Higher Education. 2016;28:68-84.
[12] Siegel E. Predictive analytics. Hoboken: Wiley. 2013.
[13] Brinton CG, Chiang M, editors. Mooc performance prediction via clickstream data and social learning networks. Computer Communications (INFOCOM), 2015 IEEE Conference on; 2015: IEEE.
[14] Khalil M, Ebner M, editors. Learning Analytics in MOOCs: Can Data Improve Students Retention and Learning? EdMedia: World Conference on Educational Media and Technology; 2016: Association for the Advancement of Computing in Education (AACE).
[15] Yang T-Y, Brinton CG, Joe-Wong C, Chiang M. Behavior-Based Grade Prediction for MOOCs via Time Series Neural Networks. IEEE Journal of Selected Topics in Signal Processing. 2017.
[16] Whitehill J, Mohan K, Seaton D, Rosen Y, Tingley D. Delving Deeper into MOOC Student Dropout Prediction. arXiv preprint arXiv:170206404. 2017.
[17] Freitas S, Gibson D, Du Plessis C, Halloran P, Williams E, Ambrose M, et al. Foundations of dynamic learning analytics: Using university student data to increase retention. British Journal of Educational Technology. 2015;46(6):1175-88.
[18] Wong BTM, Wong BTM. Learning analytics in higher education: an analysis of case studies. Asian Association of Open Universities Journal. 2017;12(1):21-40.
[19] Cao G, Duan Y. Understanding Learning Analytics from an IT Affordance Perspective. 2017.
[20] Sclater N. Developing a Code of Practice for learning analytics. Journal of Learning Analytics. 2016;3(1):16-42.
[21] Mat U, Buniyamin N, Arsad PM, Kassim R, editors. An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. Engineering Education (ICEED), 2013 IEEE 5th Conference on; 2013: IEEE.
[22] Veeramachaneni K, Halawa S, Dernoncourt F, O'Reilly U-M, Taylor C, Do C. Moocdb: Developing standards and systems to support MOOC data science. arXiv preprint arXiv:14062015. 2014.
[23] Kadir ZA, Baboo SB, Rosni NS, Rahman ZHA, Bakar NA. Design & development of digital learning resource (BMT): blended learning approach. Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication; Beppu, Japan. 3022234: ACM; 2017. p. 1-7.
[24] Bakar Z, Ali R, Zainal Azrie Shuib M. The effect of learning potential enhancement intervention (lpei) on student's regulation of learning at school2017.
[25] Tai SS, Ah SHAB. The untold truancy stories: a case study in Malaysia. The Malaysian Journal of Social Administration. 2017;12(1):76-99.
[26] Gasevic D, Kovanovic V, Joksimovic S, Siemens G. Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. The International Review of Research in Open and Distributed Learning. 2014;15(5).
[27] Sunar A, White S, Abdullah N, Davis H. How learners' interactions sustain engagement: a MOOC case study. IEEE Transactions on Learning Technologies. 2016.
[28] Taylor C, Veeramachaneni K, O'Reilly U-M. Likely to stop? Predicting stopout in massive open online courses. arXiv preprint arXiv:14083382. 2014.
[29] Whitehill J, Williams JJ, Lopez G, Coleman CA, Reich J. Beyond prediction: First steps toward automatic intervention in MOOC student stopout. 2015.
[30] Wang F, Chen L, editors. A Nonlinear State Space Model for Identifying At-Risk Students in Open Online Courses. EDM; 2016.
[31] Meier Y, Xu J, Atan O, Schaar Mvd. Predicting Grades. IEEE Transactions on Signal Processing. 2016;64(4):959-72.
[32] Hone KS, El Said GR. Exploring the factors affecting MOOC retention: A survey study. Computers & Education. 2016;98:157-68.
[33] Kloft M, Stiehler F, Zheng Z, Pinkwart N, editors. Predicting MOOC dropout over weeks using machine learning methods. Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs; 2014.
[34] Tsironis A, Katsanos C, Xenos M, editors. Comparative usability evaluation of three popular MOOC platforms. Global Engineering Education Conference (EDUCON), 2016 IEEE; 2016: IEEE.
[35] Sa'don N, Alias R, Ohshima N. Stakeholders' view on MOOCs sustainability in Malaysian higher education: A preliminary study2015. 18104-12 p.
[36] Gunawardena A. BRIEF SURVEY OF ANALYTICS IN K12 AND HIGHER EDUCATION2017. 24 p.
[37] Miguel J, Caballé S, Prieto J, editors. Providing Information Security to MOOC: Towards effective student authentication. Intelligent Networking and Collaborative Systems (INCoS), 2013 5th International Conference on; 2013: IEEE.
[38] Veeramachaneni K, Dernoncourt F, Taylor C, Pardos Z, O'Reilly U-M, editors. Moocdb: Developing data standards for mooc data science. AIED 2013 Workshops Proceedings Volume; 2013.

[39] Mohamad N, Ahmad N, Sulaiman S. Data pre-processing: A case study in predicting student's retention in MOOC. Journal of Fundamental and Applied Sciences. 2017;9(4S):598-613.
[40] Alias UF, Ahmad NB, Hasan S, editors. Mining of E-learning behavior using SOM clustering. Student Project Conference (ICT-ISPC), 2017 6th ICT International; 2017: IEEE.
[41] Baker RS, Inventado PS. Educational data mining and learning analytics. Learning analytics: Springer; 2014. p. 61-75.
[42] Balakrishnan G, Coetzee D. Predicting student retention in massive open online courses using hidden markov models. Electrical Engineering and Computer Sciences University of California at Berkeley. 2013.
[43] Brooks CA, Thompson C, Teasley SD, editors. Towards A General Method for Building Predictive Models of Learner Success using Educational Time Series Data. LAK Workshops; 2014.
[44] Nagrecha S, Dillon JZ, Chawla NV, editors. MOOC Dropout Prediction: Lessons Learned from Making Pipelines Interpretable. Proceedings of the 26th International Conference on World Wide Web Companion; 2017: International World Wide Web Conferences Steering Committee.
[45] Noh K-S. Plan for vitalisation of application of big data for e-learning in South Korea. Indian Journal of Science and Technology. 2015;8(S5):149-55.