Predicting Student Performance in Massive Open Online Courses (MOOCs) Using Big Data Analysis and Convolutional Neural Network

Ibrahim Aboualgasem O Alrmah
School of Information Technology
Malaysia University of Science and Technology (Must), Malaysia

Dr. Asiah Lokman
School of Information Technology
Malaysia University of Science and Technology (Must), Malaysia

Abstract:- One of the understanding analytics in MOOC is to determine and forecast students' performance based on different antecedents that gathered as records from the system for the engagement activities of the students. Along with the visibility of big data, the usage of artificial intelligent methods can easily supply effective results in forecasting the students' standing and performance. This study aims to provide an artificial neural network design for forecasting students pass/fail status along with their band performance based on MOOC big data analysis. The data collection utilized is the one collected and discharged through Harvard and MIT "HarvardXMITx -Course Dataset AY2013" in May, 2014. MATLAB Convolutional Neural Networks (CNN) is used as a platform for simulating the proposed design. For the data set, the total cases were 641138 cases; the filtered cases with complete field were 58453 cases. The initial design has eight possible input variables, which is tested as a preliminary step to determine its importance to the model. The final design for predicting learners' performance and level have four inputs and two outputs. Predicting accuracy of success status (pass/fail) shows that 91.6% of learners' success status can be predicted by using testing data. Predicting accuracy of success Level (Band 1 to Band 5) shows that 82.6% of learners' success status can be predicted by using testing data. The proposed data mining model has four input variables and the precedence for its importance are day's activity, followed by played videos, then events number, and finally chapters opened.

Keywords:- *MOOC, Data Analytics, CNN, Student Performance, Educational Data Science.*

I. INTRODUCTION

The 2000s saw changes in online, or e-learning and distance education, with boosting on the internet existence, open learning possibilities, and the growth of MOOCs (Massive Open Online Courses) [1]. MOOC can be specified, as web-based online training courses that are different from the traditional e-learning systems by the adhering to characteristics [2], [3]. It is an on-line course targeted at large participation and open (free) accessibility through the web. They are similar to college courses, however do not tend to supply scholastic credit history. A number of web-based systems supported by leading colleges and universities offer MOOCs in a vast array of subjects [1]. MOOCs are commonly seen as a significant part of a bigger disruptive innovation taking location in greater education and learning. MOOCs endanger existing organisation models by possibly selling placement, mentor, and/or analysis separately from the present package of services [4]. Robert Zemsky (2014) says that MOOCs have actually passed their height: "They came; they dominated very little; as well as currently they encounter significantly diminished prospects" [5].

In enhancement to the learners' tasks steps, better investigation of just how the diversity of MOOC trainees based on their personal profile or set of individual qualities (e.g., learners with different histories, factors for signing up, degrees of self-regulation skills as well as previous knowledge) is associated to their state-level inspiration throughout the training course [6]. Future work ought to collect and evaluate information from various MOOCs across different disciplinary locations and also program systems; it is affordable to ask whether results may differ between MOOCs on topics, culture distinctions [7].

The value of evaluating extraordinary loudness of data on students in MOOCs is actually of wonderful rate of interest to data science (DS) scientists [8]. The specific use of data science (DS) in the education and learning field is actually recognized as educational data science (EDS), which functions along with data acquired coming from academic environments/settings to resolve academic issues [9].

This research is discovering to which level the students' tasks actions as well as learners' profile actions can predict learner's efficiency in MOOC. The major concern for this research study is "what are the essential components of students' efficiency in MOOC?".

II. LITEARTURE REVIEW

A. Course Completion

Some of the vital sensations that have actually been actually noticed in MOOCs, both in scholarly places as well as in the well-known push, is actually that several students stop working to finish MOOCs [10], [11]. It is actually confusing, as explained over, whether conclusion is actually also the target for numerous students. Students move toward MOOCs along with a range of inspirations as well as targets, past simply accomplishing the program or even getting a certification [12]. Several MOOC attendees sign up in training programs simply to delight their preliminary inquisitiveness along with no goal of finishing the training course [13]. For cMOOCs, several program individuals of the Connectivism as well as Connective Knowledge (CCK08) MOOC shared that program conclusion was actually certainly not their purpose [14].

One center research that took into consideration training course finalization as simply among numerous prospective methods to taking a MOOC recommended that MOOC students can be sorted right into four teams based upon their actions: Completing, Auditing, Disengaging, as well as Sampling. Their looking for proposed that MOOC students involve along with web content in at the very least 4 unique various methods; these 4 designs of actions might work with 4 various targets [15]. Another research watches the very same sensation as standing for a "channel of involvement", where students may be actually sorted through their level of involvement, along with each much deeper degree of engagement being actually embarked on through a much smaller amount of students [16].

A variety of research studies have actually been actually performed on student recognition in standard setups, e.g., failure evaluation of educational institution training. Some analysts have actually effectively made use of the amount and regularity of online forum messages, as well as succeeding interaction to assess student rate of interest as well as loyalty in MOOCs [17]. Whereas some of these strategies forecast student failure precisely one full week in advancement (eg. Kloft, Stiehler, Zheng, & Pinkwart, 2014), others have actually anticipated students' ultimate performance utilizing simply full week 1 data [19].

B. Learner Usage of MOOCs

In relations to learner utilization of online videos, scolars checked out video clip seeing data coming from 6.9 thousand video recording checking out treatments, along with the research study concern of exactly how layout selections throughout video recording creation influences student interaction [20]. They looked into student performance about video recording designs including whether the online video was actually taped in a real-time class, whether the grabbed video recording consists of genuine target market, whether teacher presents a pulling out freehand on an electronic tablet computer, and so on. They discovered briefer online videos, introduction of coach talking-head video clips, as well as existence of drawing-hand type guidelines triggered much better involvement [20].

In regards to learner tasks within the dialogue online forums, scholars considered just how social elements drawn out coming from conversation online forums affect training course fulfillment and also pinpointed forecasters of conclusion, locating that metrics like whether a student is actually a chat initiator and also a student's regularity of uploading are actually anticipating of finalization [21], [22]. Another study carried out an industry trying out 2 speculative online forums under an edX MOOC as well as analyzed whether the visibility of an online forum credibility component may affect student performance [23]. Their research study revealed that the existence of the discussion forum image is actually connected along with greater training program loyalty.

As an available knowing setting, MOOCs give students a higher level of flexibility in conditions of just how and also when the readily available discovering information may be actually utilized, producing navigating styles a likely beneficial device for knowing student interaction, targets, as well as finding out techniques [24]. authors accumulated student task data coming from 4 edX MOOCs to review whether students belonging to various demographic classifications show 'direct navigating' (accessing training course components and also video clips depending on to the offered pattern), or even 'non-linear navigating'.

C. Learner Motivation

There is actually a lengthy background of research study right into learner motivation in MOOCs and other similar approaches. Students along with finding out targets - likewise named knowledge targets - aim to enhance their proficiency and expert abilities; students along with performance objectives aim to be successful as well as secure positive analyses coming from others [25].

It has actually been actually said that various target positioning are in fact signs and symptoms of rooting student point of views. Students along with development mindsets keep the opinion that knowledge is actually manageable; whereas students along with a dealt with point of view take into consideration cleverness a permanent company [26]. Previous studies analyzed way of thinking throughout of different ages, via links in between mindset and also objective alignment, as well as additionally discovered that students along with a development point of view outrun their versions that approve a repaired mindset, over the long-term [27].

Scientists have actually likewise analyzed learner incentive in internet understanding atmospheres. Previous work have actually suggested that students of E-learning systems deal with additional inspirational difficulties due to the fact that they must operate separately far-off for the most part, lessening the forms of assistance readily available in a grounds atmosphere, consisting of each social communication and also specialized assistance [28].

Students taking MOOCs might encounter identical obstacles.

Inspirational parts unique to the MOOC as a knowing system have actually additionally been actually taken into consideration in research studies checking out why students enlist in MOOCs [29]. Scientists have actually inquired MOOC attendees whether they are actually geographically distant coming from the college where a program is actually located, considering that a MOOC can easily get to students in an extensive array of areas and also nations [30]. This kind of motivation might reveal why some students take a details MOOC, considering that it gives all of them information that is actually certainly not offered in your area.

As an available discovering atmosphere, MOOCs use students a higher level of liberty in conditions of just how and also when the offered knowing information can easily be actually utilized, producing navigating trends a possibly practical resource for knowing student interaction, objectives, and also discovering approaches. Guo and also Reinecke (2014) gathered student task data coming from 4 edX MOOCs to take a look at whether students belonging to various demographic types show 'straight navigating' (accessing training course products as well as online videos depending on to the offered pattern), or even 'non-linear navigating'. Many of the previous MOOC research studies pointed out over concentrated on student communications along with a MOOC, such as video clip enjoying behaviors and also navigating designs. In the current research study, our team seek to enhance this job through utilizing self-report musical instruments to analyze students' inspiration, accumulating a collection of survey things on students' inspirations, as well as connecting all of them to student results [31]. Analysts have actually inquired MOOC attendees whether they are actually geographically distant coming from the college where a program is actually located, because a MOOC can easily get to students in an extensive variety of areas as well as nations [32].

III. METHODS AND DESIGN

Before A theoretical framework is actually a logical resource along with numerous varieties and situations. Powerful visionary platforms record one thing genuine and perform this in a method that is actually very easy to administer and always remember.

- Input: Learners' activities measures
- Data Mining: CNN design – Multilayer Perception (MLP)
- Output: Pass/Fail status and Grade classification (band 0, band 1, band 2, band 3, band 4)

Based on the proposed aim, the following are the proposed conceptual framework of this study (Fig. 1).

- Learners' interaction activities (such as interaction, played videos, chapters, active days, and forum logs) in MOOC have a positive impact on the learners' performance.
- Learners' personal characteristics (such as age, gender, and qualification) in MOOC cause differences in the learners' performance.

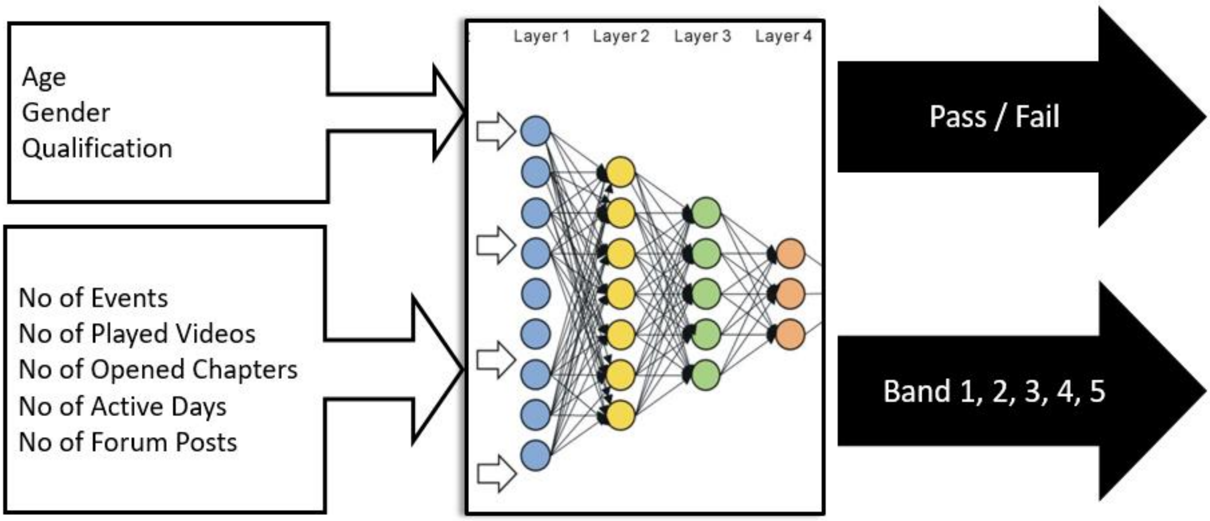

Fig 1:- The research Hypothesis

For drawing out the students' performance, the research study is making use of computational semantic network (CNN) due to the fact that it is a collection of non-linear details modelling devices including input as well as likewise outcome layers plus a couple of surprise layers. MATLAB Neural Networks provides non-linear details modeling treatments that permit you to locate added challenging collaborations in your data. The data collection made use of is the initial of its kind jointly launched by Harvard as well as MIT "HarvardXMITx Person-Course Dataset AY2013" in May, 2014 - the first (de-identified) details easily offered on MOOCs.

ISSN No:-2456-2165

IV. FINDINGS AND ANALYSIS

A. Data Screening

This procedure is vital in data study to acquire the precise participants. It is actually described as a method of evaluating data for mistakes and also fixing all of them just before carrying out data review. Data filtering might entail inspecting biting data, handling and determining outliers along with skipping data.

Relating to the real data established, the data is actually secondary collection, which picked up through MIT as well as harvared in 2013. The data collection consisting of all the instances whether it is actually finished or otherwise as well as in our cae our team require to filter merely the data of students that accomplish certification and also possess a full documents and also without a doubt possess a regular data behavior certainly not influencing the entire collection of data. The procedure filtering system the dataset is actually the following:

- Full Data set is the all student who register even if not doing any later activities.
- The data is cleaned to delete any cases which are not met with the following criteria
  - Grade must have a value higher than 0 and must not be empty
  - Number of event must have a number (starting 0) and must not be empty
  - Number of active days must have a number (starting 0) and must not be empty
  - Number of played videos must have a number (starting 0) and must not be empty
  - Number of opened chapters must have a number (starting 0) and must not be empty
  - Number of forum posts must have a number (starting 0) and must not be empty
  - For data of birth, country, gender; must have an entered value and must not be empty fields.

Table 1 demonstrates the related statistics. For the data set, the total cases were 641138 cases, the filtered cases with complete field were 58453 cases.

Description	**Total number of fields**
Data Full-Set	**641138**
Data with Complete-Fields	**58453**

Table 1:- Data Screening Analysis

B. Preliminary Design

Proposed design is looking for predicting the performance of learners based on their activities and characteristics; therefore, it is important to define how to assess performance and which activities or characteristics is useful for prediction. The possible predictor available in the secondary dataset include the following

1. Age – recoded into 6 ordinal categories
2. Gender (male/female)
3. Qualification - recoded into 6 ordinal categories
4. No of Events – continuous numbers
5. No of Played Videos – continuous numbers
6. No of Opened Chapters – continuous numbers
7. No of Active Days – continuous numbers
8. No of Forum Posts – continuous numbers

The variables is tested one by one to assess whether it have any successful prediction for success status and level by performing a simple Neural network design of 1 input and 2 outputs of 7 categories. As shown in Table II, the preliminary results show that learners personal profile attributes (age, gender, qualification) have no role in predicting neither success status nor success level. In addition the interaction attribute “forum posts”, has no role in both outcome measures. Therefore, the final design for predicting learners’ performance and level have four inputs and two outputs as the design in the following Fig. 2.

No	Variable	Predict at least 3 categories	Predict Pass/Fail	Comment
1	Age	No	No	Leave
2	Gender	No	No	Leave
3	Qualification	No	No	Leave
4	No of Events	Yes	Yes	Keep
5	No of Played Videos	Yes	Yes	Keep
6	No of Opened Chapters	Yes	Yes	Keep
7	No of Active Days	Yes	Yes	Keep
8	No of Forum Posts	No	No	Leave

Table 2:- Preliminary Results of the Variables Usability

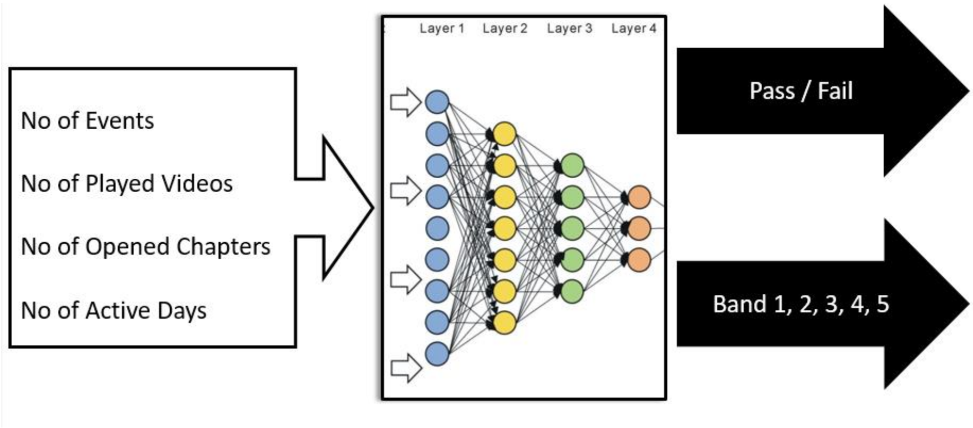

Fig 2:- Proposed Design of Prediction Design

C. *Neural Network Design*

Figure 3 shows a demonstration of the convolutional neural network proposed for this particular study. It is clear that the design have three layers, input, process, and output layers. Input layers consists of four neurons to key in standardized dataset of events, played videos, opened chapters, and active days. Processing layer is the data mining design, in which there are six neurons triggered by using hyperbolic tangent function. Finally, output layer have seven neuron associated with seven outputs, five outputs are the success levels categories and two outputs are the success/fail status. The output layer use softmax as a triggering function and cross entropy as an error function. The layers are as the following.

- ➢ Input layers with 4 variables of standardised dataset:
- CN_Events
- CN_ActDays
- CN_PlayVideo
- CN_Chapters
- ➢ 1 Hidden Layer with 6 Units (Neurons)
- Activation Function = Hyperbolic tangent
- ➢ Output Layer with 7 Unites
- Activation Function = Softmax
- Error Function = Cross-entropy

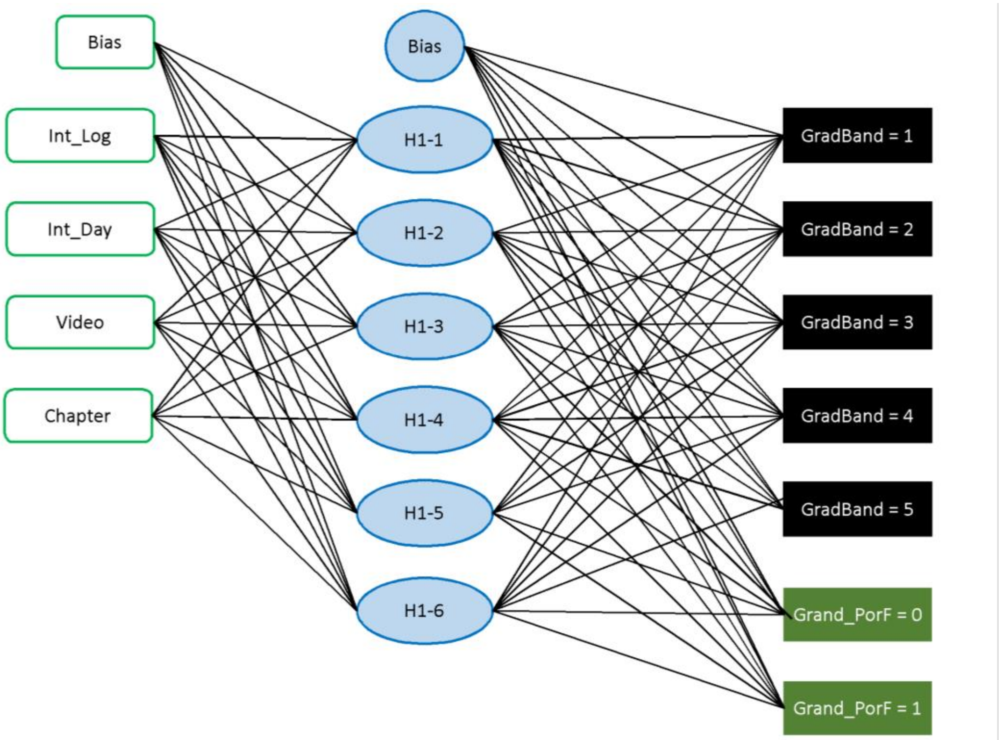

Fig 3:- Proposed Design of Neural Network

Training and Testing Data – the total dataset have 58453 cases, which is distributed into two samples; training and testing. Training dataset has 70% of the data with 40935 cases and testing has 30% of the data with 17518 cases. It is common in artificial intelligence models to assign the majority of the dataset for training the system then use a smaller fraction for testing purposed. The distribution ration differs as 60% to 80% are assigned for training and 20% to 40% are assigned for actual testing purposed. This study used the 70-30% distribution strategy as seen in Table III.

		Count	Percent
Sample	Training	40935	70.0%
	Testing	17518	30.0%
Total		**58453**	**100%**

Table 3:- Training Vs Testing Data

D. Findings for Classification of Success Status (Pass/Fail)

As seen in Table IV, predicting accuracy of success status (pass/fail) is quite high for both training and testing datasets. 92% of learners' success status can be predicted by using training data and 91.6% of learners' success status can be predicted by using testing data. For detailed discussion, accuracy is discussed for fail student (0 = 0.0 – 0.49), pass students (1 = 0.5 – 1.0), and for both together. Results show that accuracy for fail category is 94.5% for training and 94.4% for testing. However, the accuracy of pass category is much fewer, because accuracy for fail category is 83.6% for training and 85.1% for testing. The proposed design can classify 91.6% of the learning to be assigned to Pass or Fail Status.

Grade_PorF				
Sample		Predicted		
		0	1	Percent Correct
Training	0	29775	1720	94.5%
	1	1547	7893	83.6%
	Overall Percent	76.5%	23.5%	92.0%
Holdout	0	11597	688	94.4%
	1	781	4452	85.1%
	Overall Percent	**70.7%**	**29.3%**	**91.6%**

Table 4:- Prediction Accuracy for Success Status

E. Findings for Classification of Success Level (Band 1 to 5)

AS seen in Table V, predicting accuracy of success Level (Band 1 to Band 5) is high for both training and testing datasets. 87.6% of learners' success level can be predicted by using training data and 82.6% of learners' success status can be predicted by using testing data. For detailed discussion, accuracy is discussed based on the different five bands of success level (0.0 – 0.49 = Band 0; 0.5 – 059 = Band 1; 0.6 – 0.69 = Band 2; 0.7 – 0.79 = Band 3; 0.8 – 0.89 = Band 4; and 0.9 – 1.0 = Band 5). Results show that accuracy for band 0 is 96.6% for training and 95.2% for testing; accuracy for band 1 is 0% for training and 0% for testing; accuracy for band 2 is 60.6% for training and 64.8% for testing; accuracy for band 3 is 79.6% for training and 80.9% for pass category; accuracy for band 4 is 11.5% for training and 4.4% for testing; and accuracy for band 5 is 90.9% for training and 79.3% for testing. The proposed design can classify 82.6% of the learning to be assigned to bands from 1 to 5.

GradBands								
Sample		Predicted						
		0	1	2	3	4	5	Percent Correct
Training	0	30438	0	271	262	184	340	96.6%
	1	402	0	417	54	13	11	0.0%
	2	383	0	836	81	32	48	60.6%
	3	241	0	5	1207	56	7	79.6%
	4	1035	0	179	177	256	573	11.5%
	5	155	0	100	6	52	3114	90.9%

GradBands								
Sample		Predicted						
		0	1	2	3	4	5	Percent Correct
	Overall Percent	79.8%	0.0%	4.4%	4.4%	1.4%	10.0%	87.6%
Holdout	0	11691	0	129	267	31	167	95.2%
	1	145	0	171	69	1	1	0.0%
	2	122	0	459	112	7	8	64.8%
	3	132	0	4	690	25	2	80.9%
	4	365	0	142	146	57	594	4.4%
	5	231	0	70	23	86	1571	79.3%
	Overall Percent	72.4%	0.0%	5.6%	7.5%	1.2%	13.4%	**82.6%**

Table 5:- Prediction Accuracy for Success Level

F. Overall Prediction and Variables Importance

As seen in Table VI, when taking into account the overall accuracy of the model as one unit, the prediction accuracy are 87.1% for the testing data and 89.8% for the training data. Overall, the proposed design can predict up to 87.1% of learners' status and level in MOOC courses.

Variables importance in predicting outputs are discussed in Table VII based on the path coefficient score. Path coefficient is a score for how much contribution is added by the variable in the explination of the variance in the of students results. The higher value the higher impact; therefore the precedence of the four input variables re day's activity (0.322), followed by played videos (0.284), then events number (0.215), and finally chapters opened (0.180).

Sample	Grade_PorF	GradBands	Overall Percent Correct
Training	92.0%	87.6%	**89.8%**
Holdout	91.6%	82.6%	**87.1%**

Table 6:- Overall Predicting Accuracy of the Proposed Model

Rank	Variable	Importance (Regression)
3	CN_Events	0.215
1	CN_ActDays	0.322
2	CN_PlayVideo	0.284
4	CN_Chapters	0.180

Table 7:- Variables Importance in Predicting Outputs

V. CONCLUSIONS AND RECOMMENDATIONS

Predicting accuracy of success status (pass/fail) is quite high for both training and testing datasets. 92% of learners' success status can be predicted by using training data and 91.6% of learners' success status can be predicted by using testing data. Predicting accuracy of success Level (Band 1 to Band 5) is high for both training and testing datasets. 87.6% of learners' success level can be predicted by using training data and 82.6% of learners' success status can be predicted by using testing data. When taking into account the overall accuracy of the model as one unit, the prediction accuracy is 87.1% for the testing data and 89.8% for the training data. Overall, the proposed design can predict up to 87.1% of learners' status and level in MOOC courses. The proposed data mining model has four input variables and the precedence for its importance are day's activity, followed by played videos, then events number, and finally chapters opened.

This research has some academic and practical contribution. In theoretical and academic point of view, the study is unique in two main things, making empirical examination for predicting accuracy of learners' performance, and proposing a unique CNN system that fit for MOOC data analysis.

In addition, most of data science predictions models have empirical results based on traditional statistical analysis. However, this study is different in using actual big data with high diversities from the whole world to test the proposed CNN model. Any CNN network must be configured to fit with theses requirement, which is unique for this study. This research is important to scholars as it introduce a proposed solution for a specific type of Dataset. Assessing the applicability of deep learning techniques for prediction is important step towards diffusing the use of contemporary solutions. For decision makers and professional in education field such as universities, the results will increase the knowledge regarding the applicability of using AI in monitoring and predicting students' performance.

The proposed design of predicting learners' performance model has been tested by using MOOC dataset. However, the results could be limited to this type of data. Therefore, replicating the same design but with the use of different dataset such as historical university dataset can provide more insight. Actual analysis of this study shows that learners' personal profile such as gender, age, and qualification have no role in predicting learners' performance. Extra investigation is needed to figure out why it is not predictors on MOOC while many studies shows that it have a role but in different learning environment. Results shows that Forum activities have no role in predicting learners' performance. Extra investigation is needed to figure out why it is not predictors on MOOC while many studies shows that it has a role but in different learning environment.

REFERENCES

[1]. A. M. Kaplan and M. Haenlein, "Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster," Bus. Horiz., 2016.
[2]. Y. Pang, C. Liao, W. Tan, Y. Wu, and C. Zhou, "Recommendation for MOOC with Learner Neighbors and Learning Series," in International Conference on Web Information Systems Engineering, 2018, pp. 379–394.
[3]. D. Liberona, D. Fuenzalida, L. In, J. Uden, and Y. Sinclair, "Use of Moodle platforms in higher education: case'. Tao, and D.," Lib. eds Learn. Technol. Educ. Cloud MOOC Big Data Third Int. Work. Chile Sept. Proc. London Springer, vol. 2014 SRC, pp. 2–5, 2014.
[4]. N. Dwivedi, Exploring Institutional Logics for Technology-Mediated Higher Education. Routledge, 2019.
[5]. R. Zemsky, "With a MOOC MOOC here and a MOOC MOOC there, here a MOOC, there a MOOC, everywhere a MOOC MOOC," J. Gen. Educ., vol. 63, no. 4, pp. 237–243, 2014.
[6]. P. G. De Barba, G. E. Kennedy, and M. D. Ainley, "The role of students ' motivation and participation in predicting performance in a MOOC," no. March 2015, pp. 218–231, 2016.
[7]. Y. Wang and R. Baker, "Content or platform : Why do students complete MOOCs ?," vol. 11, no. 1, pp. 17–30, 2015.
[8]. C. Romero and S. Ventura, "Educational data science in massive open online courses," 2016.
[9]. C. Romero and S. Ventura, "Educational data science in massive open online courses," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 7, no. 1, p. e1187, 2017.
[10]. L. Rothkrantz, "Dropout rates of regular courses and MOOCs," in International Conference on Computer Supported Education, 2016, pp. 25–46.
[11]. E. B. Gregori, J. Zhang, C. Galván-Fernández, and F. de Asís Fernández-Navarro, "Learner support in MOOCs: Identifying variables linked to completion," Comput. Educ., vol. 122, pp. 153–168, 2018.
[12]. A. Sintjago, "International Participants' xMOOC Perspectives," 2018.
[13]. F.-R. Sheu, C. J. Bonk, and X. Kou, "A mixed methods look at self-directed online learning: MOOCs, open education, and beyond," in 25th Annual Ethnographic & Qualitative Research Conference. Cedarville, OH, 2013.
[14]. A. Fini, "The technological dimension of a massive open online course: The case of the CCK08 course tools," Int. Rev. Res. Open Distrib. Learn., vol. 10, no. 5, 2009.
[15]. R. F. Kizilcec, M. Pérez-Sanagustín, and J. J. Maldonado, "Recommending self-regulated learning strategies does not improve performance in a MOOC," in Proceedings of the third (2016) ACM conference on learning@ scale, 2016, pp. 101–104.
[16]. N. Hood and A. Littlejohn, "Quality in MOOCs: Surveying the terrain," 2016.
[17]. T. Mullaney, "Making sense of MOOCs: A reconceptualization of HarvardX courses and their students," Available SSRN 2463736, 2014.
[18]. M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, "Predicting MOOC dropout over weeks using machine learning methods," in Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, 2014, pp. 60–65.
[19]. K. Veeramachaneni, S. Halawa, F. Dernoncourt, U.-M. O'Reilly, C. Taylor, and C. Do, "Moocdb: Developing standards and systems to support mooc data science," arXiv Prepr. arXiv1406.2015, 2014.
[20]. P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: An empirical study of MOOC videos," in Proceedings of the first ACM conference on Learning@ scale conference, 2014, pp. 41–50.
[21]. J. J. George, "Transforming Viscous Data into Liquid Data: How Does Intermediating Through Digital Platforms Impact Data?" Baylor University, 2019.
[22]. N. Gov, "MOOCs," 2015.
[23]. K. Willcox, S. Sarma, and P. Lippel, "Online education: A Catalyst for higher education reforms," Final Report. Recuper. https//goo. gl/Vfrzne, 2016.
[24]. [24] S. White, "Learning designers and educators in the 'third space': the socio-technical construction of MOOCs in UK higher education." University of Southampton, 2018.
[25]. M. Jemni and M. K. Khribi, "Toward empowering open and online education in the Arab world through OER and MOOCs," in Open education: from OERs to MOOCs, Springer, 2017, pp. 73–100.

[26]. D. S. May, "Meeting the Challenges of Massive Open Online Courses in Higher Education." Pepperdine University, 2018.
[27]. P. G. De Barba, G. E. Kennedy, and M. D. Ainley, "The role of students' motivation and participation in predicting performance in a MOOC," J. Comput. Assist. Learn., vol. 32, no. 3, pp. 218–231, 2016.
[28]. S. Carliner and P. Shank, The e-learning handbook: past promises, present challenges. John Wiley & Sons, 2016.
[29]. J. O. Uchidiuno, A. Ogan, E. Yarzebinski, and J. Hammer, "Going global: Understanding English language learners' student motivation in English-language MOOCs," Int. J. Artif. Intell. Educ., vol. 28, no. 4, pp. 528–552, 2018.
[30]. C. J. Bonk and M. M. Lee, "Motivations, achievements, and challenges of self-directed informal learners in open educational environments and MOOCs," J. Learn. Dev., vol. 4, no. 1, 2017.
[31]. P. J. Guo and K. Reinecke, "Demographic differences in how students navigate through MOOCs," in Proceedings of the first ACM conference on Learning@ scale conference, 2014, pp. 21–30.
[32]. K. J. Branson, "Academic Self-Efficacy and Massive Open Online Course (MOOC) Completion Rates: Do Academic Self-Efficacy and Adult Learner Characteristics Influence MOOC Outcomes?" San Diego State University, 2017.