A Novel Analysis of MOOC's Prediction

S. Nithya[1*] and Dr S.Umarani[2]

[1*] Research Scholar, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

[2] Professor, Department of Computer Science, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

Abstract:

Massive Open Online Courses (MOOCs) are exclusively designed for virtual teaching resources of high quality that benefit large audiences. This paper explores the current status of meta-analysis on MOOCs. The key issues identified during the study in predicting MOOCs include characterizing the prediction learning outcomes, identifying the prediction features, and determining the methodologies utilized to forecast the variables. The assessment is one of the research efforts to collate the leading technologies and concepts used in learning to forecast the achievement of student learning outcomes as well as the leading characteristics utilized, which are a factor in educational outcomes. A wide range of prediction features are available, but video data and behavioral analysis of the platform are the most prominent ones. Our findings suggest a strong desire to predict MOOC dropouts in various learning approaches and two of the most commonly used forecasting analytics are Logistic Regression and Support Vector Machines.

Keywords: *MOOC prediction, Learning Analytics, Performance Prediction, Learning Outcomes, Learning Techniques.*

I. INTRODUCTION

A large number of enrollees in MOOCs are one of their most distinguishing features, as it allows for the collection of a large amount of data on what is happening in the course for further analysis. R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight[1]state that the phenomenon of big data, consisting of volume, variety, and velocity (3Vs), is a result of the of continuously created data. The professional education industry where these 3Vs prevail forms the data of this analysis. The online educational environment, which includes platforms and systems like course management and LMS, MOOCs and others, captures and generates educational data quickly.

MOOCs provide educational matter to multiple participants worldwide using online platforms with better accessibility. J. L. M. Nunez, E. T. Caro, and J. R. H. Gonzalez[2] have stated that the online education method was designed in the mid-1990s.

The assessment by S. Fu, J. Zhao, W. Cui, and H. Qu [3] emphasizes the educational sector in the early 2000s. With MOOC progressing from a teaching tool to globally revolutionizing teaching methods, the New York Times magazine proclaimed 2012 as "Year of the MOOC." From 2013 onward, self-learning concepts have thrived in the education market. In 2017 and 2018, the number of MOOC platforms has risen. I. Jo[4] has stated that most researchers have based their LMS and EDM evaluations. MOOC conference papers of high impact have traditionally been delivered at ACM and IEEE conferences.

P. G. de Barba, G. E. Kennedy, and M. D. Ainley[5] state that MOOCs have evolved into three distinct forms over time: cMOOCs, xMOOCs, and hybrid MOOCs. Users of cMOOCs platform exchange knowledge and participate in a shared teaching/ learning experience and an xMOOCs environment, users are either learners or instructors. Later MOOCs versions were hybrid, incorporating the features of all earlier versions.

Table1. In an online academic journal, the findings of several phases of a search.

	Google Scholar	IEEE Explorer	ACM	Springer	Total
Initial	60	52	76	10	198
Scanning the Keywords	49	41	66	8	164
Remove Conference Papers	32	26	34	5	97
Remove Duplicates	28	23	26	3	80
Reading Full text	21	20	21	3	65

The data in the Table.1 came from a number of electronic sources, including IEEE Explorer, Google Scholar, ACM, and Springer. In the exploration, the papers were initially counted at 198, with the majority of the papers being peer-reviewed at the ACM Conference. Then take out the conference list, taking the number down to 97.The review eventually refined 65 online-learning assessment papers. The analyzed papers revealed that LMS and EDM research has extended into a variety of sub-areas and themes.

Student learning outcomes are used to measure academic performance; analysis of groundwork between 2000 and 2020 provided a basic grasp of various intelligent approaches used to forecast student results. The pictorial view of the electronic database portrayed in fig.1 signifies the phases of the research work.

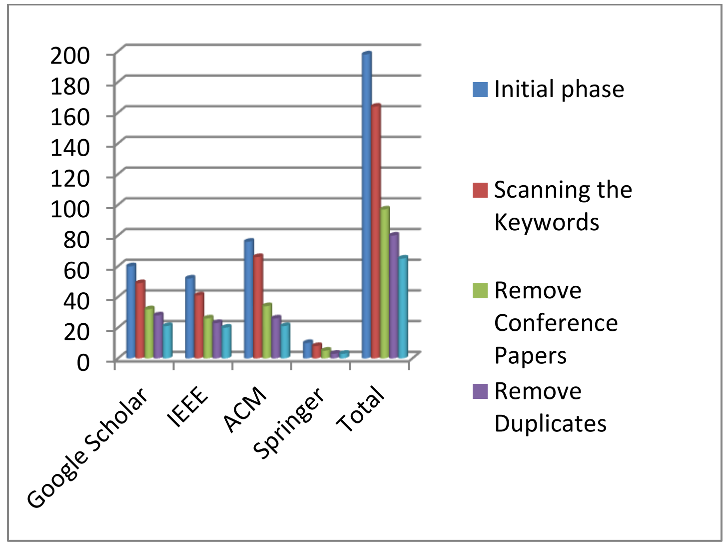

Fig1. The findings of several phases of a search

Only a few assessments focusing on the estimation of student academic achievement from the perspective of learning outcomes have been performed, according to our analysis. Table2 summarizes the best-known inquires on student progress prediction, with an emphasis on their findings and research directions. It also covers the various learning techniques and algorithms involved in this research decade.

Table2. Established Analysis on Predicting Student Achievement, their Outcome, and Challenges

S.No	Focus of Survey	Database explored	Year of study	No.of Papers Reviewed	Platform discussed	Questionnaire	Learning Techniques	Summary	Provocation / Challenges
1	Prediction of student engagement & learning outcome	EdiTlib, EBSCOhost, Scopus, Web of Science, Science Direct, Taylor & Francis, and Willey.	2012 - 2015	38 article	Coursera, edX, Udacity, and Code academy	1) To examine student activity through a systematic review of the literature. 2) To refine a well-established model of student engagement. 3) The metrics are calculated and outcome measured.	Machine learning approach, Correlation, Chi-square test, ANOVA or MANOVA, social network analysis, survival analysis, and mixed-effects regression.	1) Common Operationalization of Learning Outcomes 2)Academic Performance 3)Cognitive Change 4)Persistence and Dropout 5)Multidimensional Measures 6))Association Between Metrics Identified and Measured Outcome	1) Coded the platform where a MOOC was delivered, the educational level suggested for each of the offered courses, course domain, and course completion rates.
2	Early prediction & Learning performance	IEEE	*Not Reported*	55 paper=40 special issue + 11 jour	*Not Reported*	1) Generalization, 2) Multi-view, and 3) Frameworks.	Deep learning, Neural network	1) Early prediction and supporting of learning performance. 2) Developing and testing Early Warning Systems (EWS) and Response To Intervention (RTI) in real education environment.	1) Preventing & monitoring student at risk on real time. 2) Providing personalized intervention and feedback to each student at risk. 3) Evaluate, student who can attend in risk.
3	Predict student performance using Discussion forum	WOS, Sci.Dir, ERIC, ACM, and IEEE Xplore.	2008 - 2017	84 articles	Coursera, edX, FutureLearn, XuetangX and MiríadaX.	LARM consists of 4 components: 1) Data, 2) Stakeholders, 3) Objective, and 4) Method.	1) Observation, 2) Statistics, 3) Qualitative data, 4) Data mining, 5) Social network analysis and 6) Visualization.	1) Descriptive analysis: 5 dim - time of the publication, stake-holders involved in the analysis, and objectives, methods, and data sources used in the studies. 2) Content analysis: The main issues covered by research on discussion forums in the MOOC context. 3) The general areas of investigation: the factors affecting the participant's behavior and learning outcome.	1) To close the learning analytics loop and contribute to the intervention and visualization objectives. 2) The models to data from different courses and platforms to measure the reliability and external validity of their models. 3) To investigate how discussion forums are best used to optimize learning programming.
4	Prediction & Evaluation metrics in MOOC - Review	Scopus and ISI Web of Knowledge, IEEE, Elsevier, ACM, Springer, Taylor & Francis, and so.	2001 - 2017	88 articles=72 Conf+16 jour	edX, Coursera, Open edX, Canvas, Stanford platforms, Telescopio and MiríadaX, Iversity	**RQ1:** The most common characteristics of the MOOCs that have been used for prediction. **RQ2:** Outcomes of MOOCs? **RQ3:** Features used to build prediction models in MOOCs. **RQ4:** Techniques/models used for prediction in MOOCs. **RQ5:** Metrics have been used to evaluate prediction results in MOOCs.	Regression and Support Vector Machines. Decision trees (DTs),Random Forest (RF),Naive Bayes, Gradient Boosting Machine (GBM),Neural Networks, PSL (Probabilistic Soft Logic), K-means and Gaussian Processes.	Evaluation metrics used as, Confusion matrix, Cohen's kappa, RMSE.	1) To focus on developing and applying predictive models that can be used in more heterogeneous contexts. 2) To enhance the predictive power of current models by improving algorithms or adding novel higher-order features. 3) The prediction features, but clickstream data about platform use stands out.
5	Different Learning Styles & problems - Review	Science Direct, Springer, IEEE, Taylor and Francis, Emerald, Wiley and other Journal and Conference	2000 - 2015	129 articles	Not Reported	1)Individual E-learning problems, learning styles and methods - a)LPG (learning path generation), b)POC/CLP c)CLS or d)Doc e)IR	Bayesian networks, Neural networks, Swarm and genetic intelligence, Fuzzy logic, Association rules etc.	1) To focus on different learning styles using variable adaptive tech, E-learning problems and learning styles. 2) To exp and predictive tech like HMM, petri nets etc. are used in 17 research papers. 3) Ontology & other hybrid recommendation-based technique.	1) To predictive models, hybrid models and experimental work are discussed. 2) Many digitized areas like digital game-based learning, etc., are still at the early stages. 3)The combined usage of learning styles like Kolbe's-Five factor model, Mumford-Cattell etc.

S.No	Focus of Survey	Database explored	Year of study	No.of Papers Reviewed	Platform discussed	Questionnaire	Learning Techniques	Summary	Provocation / Challenges
6	MOOC literature-student mapping	WoS, Sci.Dir, IEEE Xplore, and ACM Digital Library	2009 - 2018.	311 articles	Coursera, Udacity, EdX, FutureLearn and XeutangX	**RQ1:** The research methods used in MOOC studies. **RQ2:** The researches focus in MOOC studies? **RQ3:** Type of contribution did MOOC studies proposed? **RQ4:** Research approaches MOOC studies have been focusing on?	1) Linear regression, logical regression, multivariate and correlation. 2)ML-SVM, DT, NN, Random forest	1) Articles in 'high impact' only considered. 2) It examines frequently used research methods, data collection methods, data analysis methods, research focuses, author's affiliations, publication avenues, publication journals and conferences in MOOCs. 3) Data collection methods are survey, interviews and dataset from MOOC platforms.	1)Use any kind of database 2)To written in other languages than English 3)Data's collected by voting exercise
7	Big Educational Data & Analytics: Survey, Architecture and Challenges	IEEE Xplore, Springer, Sci.Dir,ACM conference proceedings.	2000 - 2020	585 journals and 1452 conference papers.	'iCourse' app	1) Big educational data 2) Technological aspects for Big data for education 3) Data analytics for Big education data and 4) Future challenges for Big education data.	*Bigdata - Hadoop, Spark and Samza.*	1) The comprehensive literature review on data analytics from both technology and education aspects. 2) Big education data including the data sources, data collection, technological aspects, data analytics and challenges. 3) The different sources for input include, learning management systems (LMS), open educational resources (OER), MOOC, social media and linked data. 4) Various approaches for analytics have been discussed.	1)The social (privacy and ethical issues) and technological challenges for Big education data to be addressed
8	Student dropout predictions	ACM	2010 - Nov 2020.	*Not Reported*	Coursera, edX	1)Field of study, 2)Gathered Data 3)Student Modelling 4)Methods 5)Evaluation measures	1) Analytic Examination. 2)Machine Learning Prediction Strategies 3)Deep Learning Prediction Strategies	1) An in-depth analysis in the field of SDP, but not exclusive, of ML predictive algorithms. 2) Comprehensive hierarchical classification of existing literature. 3) Comparative analysis - alternative dropout models investigated. 4) Given less attention, such as evaluation metrics, gathered data, and privacy concern.	1) It doesn't focus only on the dropout phenomenon but also on the time that students take to complete course stages.
9	Student performance prediction using Datamining & ML techniques	ACM, IEEE Xplore, Google Scholar, Science Direct, Scopus, Springer, and Web of Science.	2010 - Nov 2020.	62 relevant papers	The real-life data set of 10,554 students of a global subscription-based online learning provider.	1) Learning outcomes are predicted, 2) Analytics models developed to forecast student learning 3) The dominant factors impacting student outcomes.	Regression and supervised machine learning models.	1) To predict the attainment of student learning outcomes, which represent a proxy for student performance? 2) Lack of unsupervised learning techniques.	1) The lack of explanatory analytics of the learning outcomes 2) The predictive models using multiple datasets from different majors and disciplines. 3) To develop explanatory predictions rather than models that merely forecast student performance. 4) The predicting learning outcomes should extend to other majors, such as humanities.
10	Strategies for retention in MOOC students	EBSCOhost, Google Scholar, and Research Gate	Jan 2015 – Mar 2019	18 articles = 17 jour +1 Conf	*Not Reported*	1) Student affective factors. 2) Institutional, faculty, and course factors 3) Environmental and demographic factors 4) Student academic factors and 5) Technological factors.	*Not Reported*	1) Retention Solution: (a) Course development, (b) Student success support, (c) Faculty involvement, (d) Social engagement, and (e) Emotional engagement. 2) The retention rates are high.	1) To extend beyond the 5-year timeframe. 2) To utilize quantitative research methods to examine online student retention. 3) To thematic research conducted in more specific areas. 4) The students from specific populations.

According to the results of this study, there are a few issues with the MOOC platform. This paper will help to explain the situation and propose new approaches in this area. Track down some problems in the MOOC content, and then list out the questions, according to this SLR.

Q1: In terms of prediction, what is the overall learning outcome?

Q2: What characteristics are utilized to create models for predictions?

Q3: What are the approaches used in these predictions?

A brief description of the various sections is as follows: The first section is the introduction. The relevant work is described in Section II. Section III presents the methodologies used in this analysis. Section IV covers the results and discussions of this SLR and provides the details needed to address the research questions. Gathering some real-time datasets for realistic evaluation as part of this analysis and the performance measurements are addressed in section V. Section VI indicates directions for future research. Finally, the entire analysis is summarized in Section VII.

II. RELATED WORK

A large number of learner log data collected on MOOC reveals the different learning habits of each enrollee including assignment submission and performance information, student demographic information, curriculum forum posts, and interactive information, clickstream data, and so on.

On educational datasets, various data mining approaches are used to forecast student progress, evaluate slow learners, and predict dropouts. Experiments using data from 100 plus students enrolled in multiple online learning modules are presented in this paper, demonstrating the predictive potential of our proposal. Using the SVM algorithm, C. Burgos, et al.[16]might cut the dropout rate by 14% compared to earlier academic period by using this dropout prevention tool. Machine learning methods are used in many experiments to assess academic performance and predict students who are at risk of failing. X. Xu, J. Wang, H. Peng, and R. Wu[17]used these features for predicting academic performance using three basic algorithms of machine learning: decision trees, neural networks, and support vector machines. Their dataset was generated by analyzing undergraduate students' final grades (enrolled in compulsory courses at Beihang University in Beijing, China).

F. Dalipi, A. S. Imran, and Z. Kastrati [18], have provided a detailed overview of the ML approaches to solving MOOCs dropout problems. In the MOOC world, Logistic Regression (LR) is the *de facto* technique for predicting student dropout. Other prediction models for dropouts employing support vector machines and Bayesian networks are also being proposed by other researchers.

Deep learning approaches combined with learning analytics for forecast progress of borderline students is an upcoming field being researched. Deep learning, a technique for learning representations from raw data, involves building a model with multiple layers. H. Waheed, *et al.*[19]used the most widely used technique of Educational Data Mining (EDM), Artificial Neural Networks (ANNs) and the accuracy rate of their findings is between 84 and 93 percent.

To visualize clickstream data, a number of analytics models have been developed. PeakVizor is a tool for exploring and understanding the learning behaviors that underpin MOOC video clickstream peaks. It assists educators with analyzing peaks in video click streams in MOOCs from various perspectives. According to Q. Chen et al. [20], instructional assessment and various trends revealed by the existing system can be examined after additional information from more programmes can be acquired and processed. As a result, a visualization framework that facilitates course-level analysis is beneficial, as it allows users to investigate data from various perspectives.

Q. Chen,, et al.[20] employed two methods to evaluate student activity and performance. Video-watching clickstreams may be interpreted as sequences of events generated, or sequences of visited positions. This event-based paradigm extracts repeated sub-sequences of student activities like reflecting (i.e. pause-play) and revision (i.e. backward and role-play).They discover that some of these actions[21] are linked to whether a user would be Correct on First Attempt (CFA) or never when attempting to answer multiple choice questions.

The fora participation patterns among participants, their communication and learning styles in collaborative learning environments, their roles played in the fora, and their connection in forming and maintaining the well-being of their online communities were investigated in papers aimed at exploration and understanding. User behavior, tweets, data collected from web logs, course information, threads, and performance data were six categories of data which S. Fu, J. Zhao, W. Cui, and H. Qu[3]collected for review.

The papers collected and collated are often related to learners' and instructors' views and requirements in relation to the fora, as well as how these requirements could be better met. For predicting engagement levels and learning outcomes, researchers used techniques like[22]Naive Bayes, Regression models, Support vector machine, Random forest, and Neural networks.

III. METHODOLOGY

People keep updating their expertise by taking new courses in various fields. Companies and their IT departments encourage employees to use the e-learning process.

As per this SLR's inclusion guidelines:

i. Data is collected from databases such as Google Scholar, IEEE, Springer, ACM, and Taylor & Francis.
ii. The papers selected for this paper were published between 2000 and 2020.
iii. In this investigation, only journal publications were used.
iv. MOOC Prediction OR MOOC Forecast OR Self-learning OR E-Learning OR Online Learning OR Learning Analytics are the main words used in this database scan.
v. This quest has also taken word variations into account.
vi. The field research was greatly biased in favor of Computer Science articles by practitioners of applied technology.

As per this SLR's exclusion rules, the following categories are not included:

i. Conference papers not subjected to peer-review.
ii. Articles written on non-English papers, and duplicate papers with the same name and topic.

The selected papers were taken from various publications as given in the list of references. IEEE Transactions on Learning Technologies, ACM Transactions on Computer- Human Interaction are among the journals that have published it.

IV. RESULTS AND DISCUSSION

This section contains general details about the selected papers, the process in which student outcomes were forecast, the intelligent models built for success prediction, and the predictors of student learning outcomes achievement.

Q1: What is the entire learning outcome in prediction?

In a MOOC, the prediction can focus on a variety of goals, such as learning outcomes and learner behaviors. The student dropout predictions were spotted by most of the studies in this study. Also, posts that are often used to categorize articles have been grouped together. Following this procedure, four classifications were made:

a) Grade prediction

It is primarily concerned with the assignment scores. S. Jiang, *et al.*[23] Simultaneously it estimated the attempts needed to complete the task. Correct on First Attempt (CFA) is the highest priority for clearing the task.

b) Certificate earning

If learners are interested in learning a new concept, they should first go through the course material. After enrolling in the course, a small percentage of them did not take the exams. As a result, they did not receive completion certificates at the end of the course in the study by Y. Lee[24].

c) Dropout

Most of the MOOC researchers focus on the student dropout predictions, which consider the performance of the students, and how they are involved in the learning process. Initially, the numbers of enrolled students are high, but number of students decrease during the study period. M. Şahin[25]employed Multiview ensemble learning to predict the dropouts. The aim of his evaluation was to predict whether or not students would abandon the course before its completion.

Q2: What characteristics are utilized to create models for predictions?

To predict the specified outcomes, the models with prediction features have considered a wide range of MOOC components; the final list is long since each prediction model uses a different set of features. As a consequence, the most significant prediction characteristics have been divided into four possible variable subsections:

a. *User logs*

In the work of K. Coussement, *et al.* [26], users who entered their details in the registration process contained demographic variables like age, gender, basic degree, experience, family status, and other relevant subject areas. It also entails learners providing feedback on the course's content.

b. *Academic performance*

The primary variable in the model by C. G. Brinton, *et al.*[21]is the score of a quiz or assignment. They can retrieve the resource materials like pdf files, etc., and have already obtained certificates or GPA scores for the participants. Attendance percentages are also considered for this prediction.

c. *Learning style*

In the study by Z. Xie[27], the video-viewing behavior (is the learner continuously watching the full video or skipping the video, how long takes to complete the video, etc.) is examined. Ease of access and inconveniences, if any, were taken into account.

d. *Student behavior*

The variables used in the behavioral analysis of A. Ramesh, et al. [28] are the interaction between tutors and forum respondents, the on-time task submission, the number of tries required to complete the questionnaire, the content delivery of the tutors, and the presentation of the content.

Q3: What techniques have been used in the prediction?

The models were built to forecast the predictions and appraise the learners' outcome; the results were dependent on the dataset and the features decided by the researchers. The new approaches in the prediction pertain to the neural network from 2017 onwards. For every algorithm included in this review,

A. *Statistical analysis*

An evaluation metric, the researcher uses method such as chi-square test, t-test, ANOVA and MANOVA. In the work by C. Schumacher and D. Ifenthaler[29], the software packages used for statistical analysis include SPSS, EPInfo, and Minitab, etc.

B. *Data mining algorithms and Machine learning Techniques*

Most of the models were developed by using data mining algorithms from 2000 onwards. The analyzed data of EDM and LMS are mostly examined by A. Namoun and A. Alshanqiti [14] using Supervised Machine learning algorithms. Linear regression has the easiest and utmost used algorithm for prediction. Many of the research studies in the overall evaluation used SVM algorithms.

C. *Bigdata analytics*

Big data analysis is currently dominated by Hadoop, Samza, and Spark.

MapReduce, a parallel processing paradigm that uses several machines to process large datasets, was employed by S. Fu, J. Zhao, W. Cui, and H. Qu[3]. HDFS is a distributed file system that stores large files across multiple servers. High fault tolerance and scalability are two of their advantages.

Apache Spark, a MapReduce-like distributed computing platform[30], was used by Y. Mourdi, *et al*.

Spark, a query analyzer, works on interactive mode; the high-volume graph processer/ analyzer Bagel was used by H. H. S. Ip *et al.*[31].Real-time analyzer Spark Streaming and the machine learning library Mllib are all part of the Spark ecosystem.

Samza, a real-world data platform based on distributed stream processing, was employed by K. L. M. Ang, F. L. Ge, and K. P. Seng [12]. Samza keeps storage and analysis on the same computer and consumes minimal memory space, all while maintaining computational performance and laying the groundwork for a versatile, tunable API.

V. PRACTICAL IMPLICATIONS

The studies listed in this paper have gathered data from a variety of sources before analyzing the relationship between the extracted function and the dropout mark. They have used datasets like KDD CUP 2015[32], iCourse app[28], and XuetangX [33]in this paper to evaluate different performance parameters.

Many popular measurement metrics are used in machine learning, statistics, and deep learning. There are several different metrics, but in the set of papers studied, seven of them predominate. Accuracy[34], AUCROC, AUCPR, Confusion matrix[23], F-Score, and others are commonly employed in machine- and deep-learning.

In statistical analysis, metrics such as Cohen's kappa, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Underestimated/Overestimated Prediction Error Rate (OPER/UPER) are evaluated. R. Al-Shabandar, *et al.*[35]have stated that all the other datasets, with the exception of KDDCup15 and the Open University (OULAD), are not freely accessible; hence data providers need to be contacted for access.

VI. GAPS AND FUTURE DIRECTION

Our analysis of various probes on this subject revealed that none of them has performed any focusing on the prediction of student academic success from the perspective of learning outcomes.

There were a number of flaws with some of the posts examined, including the fact that they were not peer-reviewed as,

a) Concentrate on various forms of research, such as humanities, natural science, and so on.
b) This paper makes no mention of the article's qualities.
c) The knowledge is still obtained from the same source. (For example, Coursera and Edx).

Our works strongly encourage researchers to continue their research on the following factors.

a. Requisite attributes used to predict learning outcome.
b. Consider the spatial distributions for further study.
c. Build the models using fuzzy, deep learning methods and advanced techniques used as Tensorflow, Pycharm, etc.
d. Compare the datasets from the different LMS models.
e. Include short duration courses – 4 weeks.

VII. CONCLUSION

In this analysis, the most significant papers in prediction-related MOOCs were examined. The research recommendations were applied by this assessment to examine the prediction of student performance, to absorb the leading technologies and the features implemented to construct the models. The predictive models were released from 2000 until 2020 in peer-reviewed journals.

A synthesis of 65 primary papers resulted from an exhaustive analysis of ten bibliographic databases. It only required a small search of the database and excluded non-English papers and conference papers. In this work, we conferred a few datasets in realistic terms and tested the output metrics with distinct learning techniques. Machine learning and statistical analysis are the most commonly employed metrics in this assessment.

In the future, the various databases will be extended to inspect multiple papers, duration of program and focus on the various datasets on multiple platforms in order to better predict the evaluation metrics.

REFERENCES

[1] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach," *IEEE Access*, vol. 6, no. c, pp. 73669–73685, 2018, doi: 10.1109/ACCESS.2018.2876755.

[2] J. L. M. Nunez, E. T. Caro, and J. R. H. Gonzalez, "From Higher Education to Open Education: Challenges in the Transformation of an Online Traditional Course," *IEEE Trans. Educ.*, vol. 60, no. 2, pp. 134–142, 2017,

doi: 10.1109/TE.2016.2607693.

[3] S. Fu, J. Zhao, W. Cui, and H. Qu, "Visual Analysis of MOOC Forums with iForum," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 201–210, 2017, doi: 10.1109/TVCG.2016.2598444.

[4] I. Jo, "Analysis of Online Behavior and Prediction of Learning Performance in Blended Learning Environments," *Educ. Technol. Int.*, vol. 15, no. 2, pp. 71–88, 2014.

[5] 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, doi: 10.1111/jcal.12130.

[6] S. Joksimović *et al.*, *How do we model learning at scale? A systematic review of research on moocs*, vol. 88, no. 1. 2018.

[7] C. Romero and S. Ventura, "Guest Editorial: Special Issue on Early Prediction and Supporting of Learning Performance," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 145–147, 2019, doi: 10.1109/TLT.2019.2908106.

[8] O. Almatrafi and A. Johri, "Systematic Review of Discussion Forums in Massive Open Online Courses (MOOCs)," *IEEE Trans. Learn. Technol.*, vol. 12, no. 3, pp. 413–428, 2019, doi: 10.1109/TLT.2018.2859304.

[9] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Munoz-Merino, and C. D. Kloos, "Prediction in MOOCs: A Review and Future Research Directions," *IEEE Trans. Learn. Technol.*, vol. 12, no. 3, pp. 384–401, 2019, doi: 10.1109/TLT.2018.2856808.

[10] A. Khamparia and B. Pandey, "Association of learning styles with different e-learning.pdf," 2019.

[11] R. A. Rasheed, A. Kamsin, N. A. Abdullah, A. Zakari, and K. Haruna, "A systematic mapping study of the empirical MOOC literature," *IEEE Access*, vol. 7, pp. 124809–124827, 2019, doi: 10.1109/ACCESS.2019.2938561.

[12] K. L. M. Ang, F. L. Ge, and K. P. Seng, "Big Educational Data Analytics: Survey, Architecture and Challenges," *IEEE Access*, vol. 8, pp. 116392–116414, 2020, doi: 10.1109/ACCESS.2020.2994561.

[13] B. Prenkaj, P. Velardi, G. Stilo, D. Distante, and S. Faralli, "A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses," *ACM Comput. Surv.*, vol. 53, no. 3, 2020, doi: 10.1145/3388792.

[14] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, pp. 1–28, 2021, doi: 10.3390/app11010237.

[15] K. Seery, A. Barreda, S. Hein, and J. Hiller, "Retention strategies for online students: A systematic literature review," *J. Glob. Educ. Res.*, vol. 5, no. 1, pp. 72–84, Jun. 2021, doi: 10.5038/2577-509X.5.1.1105.

[16] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018, doi: 10.1016/j.compeleceng.2017.03.005.

[17] X. Xu, J. Wang, H. Peng, and R. Wu, "Prediction of academic performance associated with internet usage behaviors using machine learning algorithms," *Comput. Human Behav.*, vol. 98, no. April, pp. 166–173, 2019, doi: 10.1016/j.chb.2019.04.015.

[18] F. Dalipi, A. S. Imran, and Z. Kastrati, "MOOC dropout prediction using machine learning techniques: Review and research challenges," *IEEE Glob. Eng. Educ. Conf. EDUCON*, vol. 2018-April, pp. 1007–1014, 2018, doi: 10.1109/EDUCON.2018.8363340.

[19] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, 2020, doi: 10.1016/j.chb.2019.106189.

[20] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, "PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses," *IEEE Trans. Vis. Comput. Graph.*, vol. 22, no. 10, pp. 2315–2330, 2016, doi: 10.1109/TVCG.2015.2505305.

[21] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC Clickstreams: On the Relationship Between Learner Behavior and Performance," no. ii, 2015, [Online]. Available: http://arxiv.org/abs/1503.06489.

[22] Jesús Peral, Alejandro Maté and Manuel Marco, Application of Data Mining techniques to identify relevant Key Performance Indicators, Computer Standards & Interfaces, http://dx.doi.org/10.1016/j.csi.2016.11.006

[23] S. Jiang, A. E. Williams, K. Schenke, M. Warschauer, and D. O. Dowd, "Predicting MOOC Performance with Week 1 Behavior," *Proc. 7th Int. Conf. Educ. Data Min.*, pp. 273–275, 2014.

[24] Y. Lee, "Effect of uninterrupted time-on-task on students' success in Massive Open Online Courses (MOOCs)," *Comput. Human Behav.*, vol. 86, pp. 174–180, 2018, doi: 10.1016/j.chb.2018.04.043.

[25] M. Şahin, "A Comparative Analysis of Dropout Prediction in Massive Open Online Courses," *Arab. J. Sci. Eng.*, 2020, doi: 10.1007/s13369-020-05127-9.

[26] K. Coussement, M. Phan, A. De Caigny, D. F. Benoit, and A. Raes, "Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model," *Decis. Support Syst.*, vol. 135, no. December 2019, p. 113325, 2020, doi: 10.1016/j.dss.2020.113325.

[27] Z. Xie, "Bridging MOOC Education and Information Sciences: Empirical Studies," *IEEE Access*, vol. 7, pp. 74206–74216, 2019, doi: 10.1109/ACCESS.2019.2921009.

[28] A. Ramesh, D. Goldwasser, B. Huang, H. Daume, and L. Getoor, "Interpretable Engagement Models for MOOCs Using Hinge-Loss Markov Random Fields," *IEEE Trans. Learn. Technol.*, vol. 13, no. 1, pp. 107–122, 2020, doi:

10.1109/TLT.2018.2889953.

[29] C. Schumacher and D. Ifenthaler, "Investigating prompts for supporting students' self-regulation – A remaining challenge for learning analytics approaches?," *Internet High. Educ.*, vol. 49, p. 100791, 2021, doi: 10.1016/j.iheduc.2020.100791.

[30] Y. Mourdi, M. Sadgal, H. El Kabtane, and H. E. A. El Abdallaoui, "A Multi-Layers Perceptron for predicting weekly learner commitment in MOOCs," *J. Phys. Conf. Ser.*, vol. 1743, p. 012027, 2021, doi: 10.1088/1742-6596/1743/1/012027.

[31] H. H. S. Ip *et al.*, "Design and Evaluate Immersive Learning Experience for Massive Open Online Courses (MOOCs)," *IEEE Trans. Learn. Technol.*, vol. 12, no. 4, pp. 503–515, 2019, doi: 10.1109/TLT.2018.2878700.

[32] Y. Zheng, Z. Gao, Y. Wang, and Q. Fu, "MOOC Dropout Prediction Using FWTS-CNN Model Based on Fused Feature Weighting and Time Series," *IEEE Access*, vol. 8, pp. 225324–225335, 2020, doi: 10.1109/ACCESS.2020.3045157.

[33] K. Liang, Y. Zhang, Y. He, Y. Zhou, W. Tan, and X. Li, "Online Behavior Analysis-Based Student Profile for Intelligent E-Learning," *J. Electr. Comput. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/9720396.

[34] Y. Meier, J. Xu, O. Atan, and M. Van Der Schaar, "Predicting grades," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 959–972, 2016, doi: 10.1109/TSP.2015.2496278.

[35] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Detecting at-risk students with early interventions using machine learning techniques," *IEEE Access*, vol. 7, pp. 149464–149478, 2019, doi: 10.1109/ACCESS.2019.2943351.