Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Predicting Student Dropout in MOOCs Using Genetic Algorithms and XGBoost

Houssam Eddine Aouarib[1, *], Salah Eddine Henouda[2], Fatima Zohra Laallam[3]

[1] *Ph. D candidate, Department of Computer Science and Information Technologies, Faculty of New Technologies of Information and Communication, Artificial Intelligence and Information Technology Laboratory (LINATI), Kasdi Merbah University, Ouargla, Algeria. Email: aouarib.houssam@univ-ouargla.dz*

[2] *Dr, Department of Computer Science and Information Technologies, Faculty of New Technologies of Information and Communication, Artificial Intelligence and Information Technology Laboratory (LINATI), Kasdi Merbah University, Ouargla, Algeria. Email: salaheddinehenoudafr@gmail.com*

[2]*Professor, Department of Computer Science and Information Technologies, Faculty of New Technologies of Information and Communication, Artificial Intelligence and Information Technology Laboratory (LINATI), Kasdi Merbah University, Ouargla, Algeria. Email: laallamfz@gmail.com*

ARTICLE INFO

Received: 12 Dec 2024

Revised: 14 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Massive open online courses (MOOCs) have become a transformative tendency in education due to their various benefits in the last few years. However, despite the various benefits offered to students by MOOCs, such as accessibility, flexibility, and affordability, in addition to the massive number of enrolments, MOOCs suffer from persistent high dropout rates. The emergence of machine learning (ML)and deep learning (DL) techniques, along with educational big data provided by MOOC platforms, allows researchers to address the student dropout problem through Big Data analytics and predictive models to reveal hidden patterns to improve academic outcomes. This study proposes a combination of genetic and XGBoost algorithms for forecasting student dropout from MOOCs as early as possible. It also employed and comprehensively compared several machine learning (ML) and deep learning (DL) predictive models. These models include DT, RF, LR, SVM, MLP, and LSTM. The models suggested in this study to investigate student dropout prediction (SDP) utilize the Open University Learning Analytics Dataset (OULAD). The results showed that the employed models could successfully predict student dropout. However, the proposed model outperforms the other models with 0.36-4.36%,1.34-3.78%, and 1.18-3.20% improvement in terms of accuracy, F1 Score, and AUC, respectively.

Keywords: Student Dropout Prediction (SDP), Massive open online courses (MOOCs), Machine learning (ML)& Deep learning (DL), Genetic Algorithm (GA), XGBoost.

INTRODUCTION

Education stands as a pivotal element in the advancement of socioeconomic development. The recent advancement of technology has opened up revolutionary educational opportunities, one of the most promising of which is online learning solutions. Massive Open Online Courses (MOOCs) are one of the notable inventions that have sparked the interest of online learners since their beginning in 2008 [1, 2] and in 2012, the New York Times declared it "The Year of the MOOC" [3]. MOOCs have become a popular style of education. However, despite the numerous advantages of online education, it has a higher student attrition rate than conventional education, as shown by statistical data [4]. The student dropout phenomenon poses a considerable challenge in MOOCs, marked by high attrition rates often exceeding those observed among traditional campus-based students. Dropout figures may reach up to 90% [5, 6].

Educational data mining (EDM) and learning analytics (LA) are interdisciplinary disciplines that involve multiple academic communities aiming to improve educational quality through the analysis of educational data and the

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

extraction of relevant information for researchers, employing various statistical, machine learning, and deep learning methodologies [7]. Researchers in these two disciplines have examined the phenomenon of student dropout.

Our study's primary aim was to identify students at risk of dropping out of MOOCs. We used machine and deep learning methodologies to analyze student interactions inside the Open University Learning Analytics dataset to achieve this objective. This study's contribution concerns developing and evaluating a predictive model using the genetic algorithms GA and XGBoost. We also employ and compare several ML/DL algorithms, including Support Vector Machine (SVM), Long Short-Term Memory (LSTM), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Decision Trees (DT), and Random Forest (RF) to identify students at risk of dropout as soon as possible. This predictive model identifies early withdrawals. As a result, stakeholders and instructors can intervene and implement timely and effective interventions to retain students.

The subsequent sections of this paper are structured in the following manner. Section 2 briefly overviews the academic disciplines relevant to predicting student attrition. A thorough examination of previous studies pertinent to anticipating student attrition in (MOOCs) is provided in Section 3. Section 4 outlines the methods used for predicting student dropout in this research. Section 5 provides the findings and discussion. Section 6 delineates the conclusion of this work and prospective research paths.

BACKGROUND

This section delineates essential ideas about student dropout in MOOCs referenced in this work and their interconnections, as represented in Figure 1.

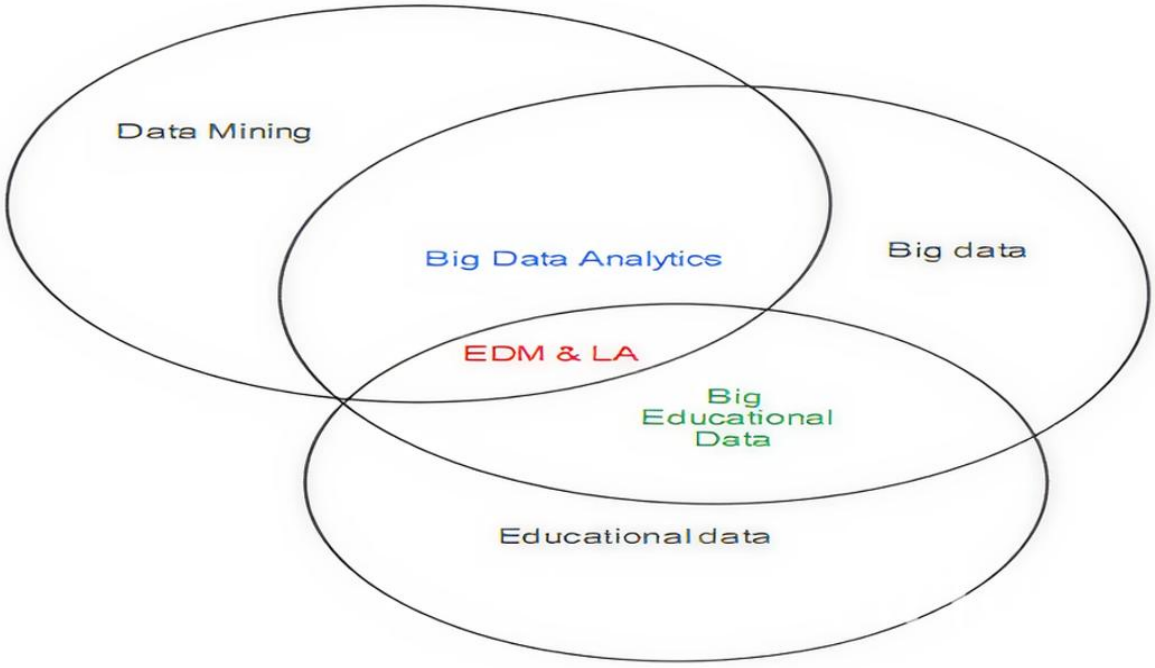

Figure 1. An illustration of EDM, LA, and Big data links and uses in higher education

Despite their increasing popularity since its inception in 2008 due to its potential to revolutionize education around the globe, MOOCs still face significant challenges, as represented by the increased student dropout rates. The two study domains that have tackled this issue LA and EDM, which may be delineated as follows: LA, as described by the First International Conference on Learning Analytics and Knowledge (LAK), can be articulated as: "The measurement, collection, analysis, and reporting of data about learners and their contexts for understanding and optimizing learning and the environments in which it occurs" [8].

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

EDM denotes the use of machine learning and data-mining techniques to analyze educational data primarily generated by students and educators in educational environments [9].

MOOC platforms generate extensive educational data from student interactions, assessments, forum contributions, and online tools, creating "big data" about the learning environments derived from the interaction of numerous students [10]. Big data analytics involves using analysis methods to examine extensive datasets, derive insights, and reveal hidden patterns [11]. Academics from different backgrounds and disciplines interested in addressing student dropout prediction and other educational challenges leverage big data analytics to analyze large amounts of educational data to solve this problem.

Overall, predicting student dropout in MOOCs is the initial measure in mitigating dropout through prompt intervention. The prediction is attainable through leveraging machine learning and deep learning techniques to develop a predictive model based on educational data about students.

RELATED WORK

In this section, we examine the latest advancements in previous research on student dropout predictions. Many research studies examine the topic of MOOC student dropout. Most studies use machine learning and deep learning techniques to tackle student attrition. These techniques represent one of the most effective methods for enhancing MOOC completion rates through early dropout prediction models.

[12] predict student attrition in a virtual learning environment (VLE)using data from the Open University Learning Analytics (OULA) dataset. The suggested approach was based on the Long Short-Term Memory (LSTM) algorithm and offers insights into the early identification of at-risk students and potential interventions to improve course completion rates in MOOCs.

[13] presented a deep learning technique called "MOOCVERSITY" that forecasts dropout rates in MOOCs. The proposed model offers a reliable method that outperforms decision trees and logistic regression when identifying students at risk of dropping out of MOOCs.

In [12] the authors proposed a learning analytics methodology to identify students at risk of dropping out, based on their behaviour in an online virtual learning environment. This study presents a model utilising a genetic algorithm for automated hyper-parameter tuning to predict early student dropout. While the authors in [15] trained deep artificial neural networks (ANNs) using manually curated features derived from online class log data to forecast student attrition. The proposed approach surpasses the baseline model in detecting at-risk students.

The authors in [16] provide a deep learning model that integrates convolutional neural networks with long short-term memory (LSTM) networks. This approach aims to assess student data and predict dropout tendencies. The proposed method demonstrates superior performance compared to traditional models, including SVM and logistic regression, facilitating the execution of preventive measures.

The Random Forest Model has been used in different studies as the best model for predicting at-risk students, such as in [17, 18]. The researchers in [17] employ various models such as decision trees, random forests, and support vectors to identify students who show dropout behavior based on performance and engagement data from the OULA dataset. In [18] the authors analyzed Open edX MOOC platform data to identify key predictors of dropout. RF achieved the best results and emphasized the potential of machine learning in predicting student dropout in both studies,

Furthermore, the work proposed in [19] assessed the efficacy of several deep-learning and Machine learning models in predicting student dropout in MOOCs. The experiments exhibited the capacity of proposed models to predict attrition in two different MOOCs with an accuracy ranging from 84% to 87%.

Multiple research studies have recently employed the XGBoost machine learning algorithm to predict student dropout. In the study of [20], the authors evaluated various predictive models to determine the best and most reliable

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

model for predicting student dropout. The experiment showed that XGBoost with the SMOTE algorithm for data balancing obtained the best results. Similarly, in [21] examine the application of an ensemble of classification methods to address the issue of student dropout. Their study illustrates the effectiveness of integrating the classification models AdaBoost and XGB to enhance prediction accuracy. Finally, in [22] the research analyze several methods to ascertain student attrition in MOOCs. The findings indicate that the XGBoost classifier surpasses the other classifiers, with an accuracy of 87%. This project aims to improve retention through early identification and tailored support strategies.

METHODOLOGY AND DATA DESCRIPTION

This section delineates the methodology used to address the study objective and emphasizes the dataset used in this study. It also conducted a comprehensive analysis of the criteria used to evaluate the effectiveness of our prediction model in identifying students at risk of dropping out.

A. Method and Proposed Model

This study aims to determine which students are most likely to drop out of massive open online courses (MOOCs). The forecast of student attrition depends on the examination of clickstream data generated during students' engagement with the virtual learning environments of MOOC platforms. Figure 2 [23] shows the Knowledge Discovery in Databases approach, which is the basis for SDP Tsak's methodology. Several classification techniques were employed and trained using data from the OULA dataset to build predictive models for early identifying at-risk students to prevent them from dropping out of the courses.

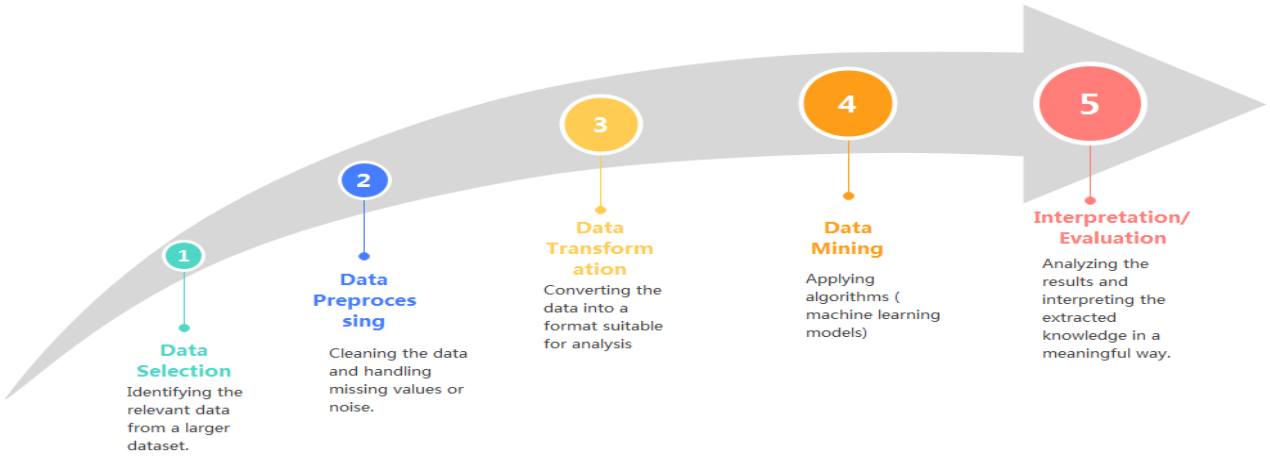

Figure 2: Student dropout prediction approach

Propose Model

A new architecture was proposed to overcome the challenge of student dropout prediction. The proposed architecture merges the genetic algorithm (GA) and XGBoost. The GA is used for this optimization task to navigate for the best hyperparameters, whereas XGBoost is used to predict student dropout from MOOCs. The experiments' flowchart is presented in Figure 3.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

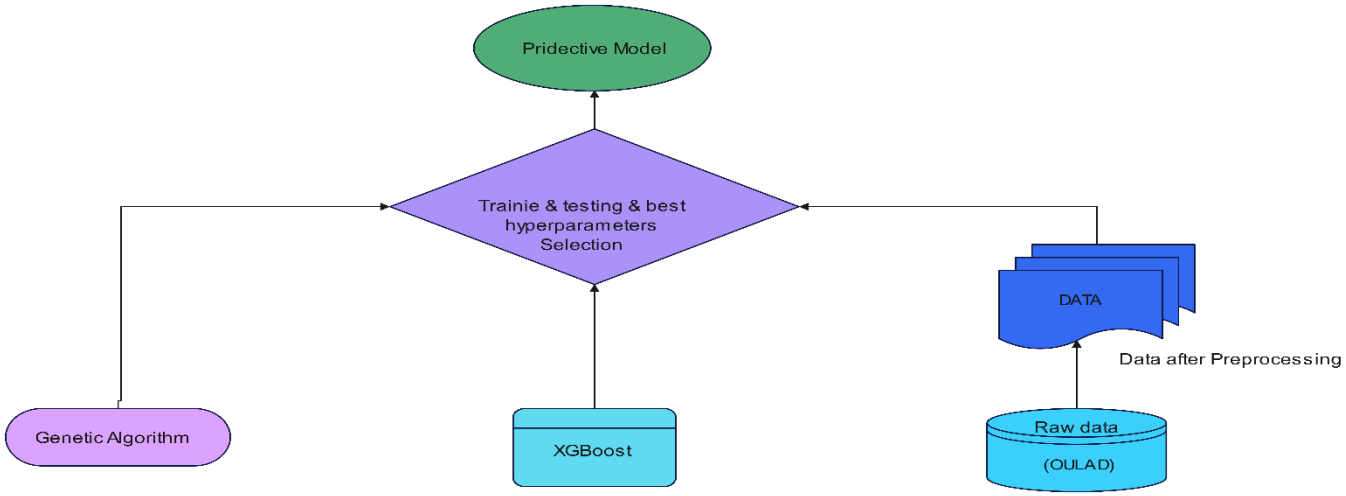

Figure 3: The experiments' flowchart.

B. Dataset

This investigation used the Open University Learning Analytics dataset (OULAD). This dataset includes demographic, performance, and clickstream data from 22 courses and 32,593 students. The clickstream data is produced by the daily summaries of student clicks (10,655,280 entries) in the virtual learning environment of Open University.

In this study, we selected 20 different types of student activities provided by the MOOC platform as features for training and testing our model. These features represent students' interactions with the VLE, as illustrated in Table 1.

Table 1: Description of the OULA dataset activities used as features

N°	Activity	Description
1	DataPlus	Interaction with supplementary information and views on videos, audio, and websites.
2	DualPane	Interaction with the site's information and associated activity
3	External Quiz	interaction with the external quiz
4	Folder	Interaction with course-related materials in public folders
5	Forumng	Participation the discussion forum
6	Glossary	Interaction with the course-related glossary of essential terms.
7	HomePage	Interaction with the course homepage
8	HtmlActivity	Interaction with the dynamic HTML page
9	Oucollaborate	Interaction with online video discussion forums
10	Oucontent	Interaction with the assignment's contents
11	Ouelluminate	Interaction with the online learning sessions
12	Ouwiki	Interaction with the Wikipedia content
13	Page	Interaction with the course-related information.
14	Questionnaire	Interaction with the course-related questionnaires.
15	Quiz	Interaction with the course quiz link.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

16	RepeatActivity	Interaction with course material from prior sessions.
17	Resource	Downloads pdf documents, such as books.
18	SharedSubPage	Broadcast the knowledge between courses and faculty
19	SubPage	Interaction with the additional websites that are accessible within the course.
20	Url	Interaction with hyperlinks leading to audio and video materials.

C. Preprocessing

This research examined student interaction with the VLE, which represents student behavior on the MOOCs platform. This latter dataset included 10,655,280 entries, each representing students' daily clickstream activity through 20 different activities. Since the Open University (OU) data is not directly applicable as inputs in the machine learning (ML) classification model.

The input features utilized in this study were initially obtained from various tables and required fusion into a unified table. Various preprocessing steps were conducted on the data, where the clickstream data log of each student was transformed into a weekly format by aggregating daily records into a weekly activity record. The final structure of the data, subsequent to the preparation step, is illustrated in Figure 4.

id_student	code_module	code_presentation	Week	Features(20 different type of VLE activites)				Label
				resource	quiz	.....	htmlactivity	
S1	C_M	C_P	W1	.....	.....	.....	.....	Pass
S2	C_M	C_P	W1	.....	.....		.....	Withdraw am
.....	.....	.....	.....	.....	.....		.....	.....
Sn	C_M	C_P	Wn	.....	.....		.....	Pass

Figure 4: Data structure obtained from preprocessing phase.

D. Used Techniques

We seek in this study to evaluate the effectiveness of early dropout model by comparing different conventional classification models frequently utilized in the literature for predicting student dropout at an early stage.

Logistic Regression (LR): is a linear model utilizing a sigmoid activation function to produce the probability of the positive class. It is utilized for binary classification problems (Hassan et al. 2019).
Decision Tree (DT): is a classification and regression technique designated for its arboreal configuration. A tree structure has root, internal, and leaf nodes. DT recursively divides data based on feature values to generate predictions [14].
XGBoost (XGB): XGBoost is a scalable tree-boosting system widely used by data scientists and provides state-of-the-art results on many problems. XGBoost algorithm handles sparse data, and its cache access patterns, data compression, and sharding features enable the algorithm to solve large-scale problems using a minimal amount of resources.
Random Forest (RF): combines multiple decision trees to make predictions by aggregating their outputs. Random subsets of data and features for each tree increase forecast accuracy and prevent overfitting [14].

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Support Vector Machine (SVM): is a classification and regression supervised machine learning technique. It works for linear and non-linear data by finding an ideal
hyperplane that maximum separates classes or matches the data with the largest margin [15].
Multilayer Perceptron (MLP): an artificial neural network commonly used for various
machine learning tasks. It consists of multiple layers of interconnected nodes, including input, hidden, and output layers [13].
Long-short-term memory (LSTM): is a recurrent neural network (RNN) architecture designed to alleviate the vanishing gradient problem in sequence data processing. Natural language processing and time series analysis employ LSTMs to capture long-term data relationships [12].

E. Employed Metrics

An assessment is crucial for evaluating the models' efficacy in accurately predicting student attrition. The chosen assessment metrics for this study may be described as follows:

Accuracy: a singular numerical measurement of the ratio of accurately predicted instances to the total number of instances [7].

$$\textbf{Accuracy} = \frac{\textbf{Number of correct predictions}}{\textbf{Total prediction number}}$$

F1 Score: It takes into account both the model's ability to correctly identify positive instances (precision) and its ability to capture all positive instances (recall), providing a balanced performance measure [16].

$$\textbf{F1 Score} = 2 \times \frac{\textbf{pr eci si on} \times \textbf{r ecal l}}{\textbf{pr eci si on} + \textbf{r ecal l}}$$

Area Under the Curve (AUC): It is a widely employed evaluation metric in binary classification tasks, quantifying the effectiveness of a model's prediction probabilities. [24]. It is a curve resulting from plotting the false positive rate of FPR (or 1-specificity) versus the true positive rate of TPR (or sensitivity) [25].

RESULTS AND DISCUSSIONS

This subsequent section presents and examines the empirical findings derived from our experimental investigations. The main goal of this project was to develop, test, and refine machine learning models that could predict when students would stop attending classes by analyzing their study habits recorded in their virtual learning environments (VLEs).
During the training process, the data from each week is concatenated with the data from the previous weeks to construct sequence data. The resulting data is fed to the classification model to predict at-risk students early. Thus, preventive action can be taken to prevent dropout.
This study proposes a predictive model based on genetic and XGBoost algorithms. In addition, several algorithms for a binary classification problem were employed to evaluate our model. The empirical results were evaluated using efficiency measures, such as accuracy, F1-Score, and AUC, as shown in Table 2.

Table 2: The experimental results of employed model.

Weeks	5			15			25			35		
Metrics	**Accuracy**	**F1 Score**	**AUC**	**Accuracy**	**F1 Score**	**AUC**	**Accuracy**	**F1 Score**	**AUC**	**Accuracy**	**F1 Score**	**AUC**
GA & XGboost	**81,09**	**74,2**	**79,05**	**90,31**	**87,08**	**89,3**	**96,9**	**96,16**	**96,75**	**98,39**	**98,03**	**98,14**
LR	74,61	71,4	75,57	85,05	82,38	85,63	92,74	91,17	93,1	96,83	96,06	96,96
MLP	80,25	73,05	78,08	86,05	82,69	85,75	95,18	93,75	94,55	97,17	96,42	97,02

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

DT	80,41	69,78	76,39	89,77	86,13	88,14	96,35	95,29	95,85	97,78	97,16	97,51
SVM	77,19	69,15	75,02	88,04	84,39	86,95	94,34	92,81	94	96,73	95,85	96,55
RF	80,98	71,36	77,34	90,28	86,6	88,4	95,94	94,67	95,1	98,02	97,46	97,67
LSTM	77,72	72,08	76,86	86,07	82,1	85,16	93,08	91,15	92,57	96,07	95,01	95,84

Table 2 illustrates the prediction summary obtained from various machine learning algorithms employed, namely DT, RF, MLP, SVM, LR, LSTM, and the combination of genetic algorithm and XGBoost on the OULAD dataset. The table presented in this study illustrates the specific results of the predictions made at intervals of 5, 15, 25, and 35 weeks regarding accuracy, F1 score, and AUC metrics.
The range from the fifth week to the 35th was selected because determining the behavior of the dropouts before the first five weeks is difficult due to the limited availability of information. At the same time, interaction and dropout rates are rare as the end of the course approaches after the 35th week.

The results presented in Table 2 demonstrate the effectiveness of our model in predicting student dropout with performance of 81.09%, 74.2%, and 90.31% in terms of accuracy, F1 Score, and AUC, respectively. These results were obtained in the first weeks when there was a lack of data related to student behaviors in the earlier weeks of the course. However, the predictive result got better with the advancement of the courses, to execute 98% regarding the three metrics.

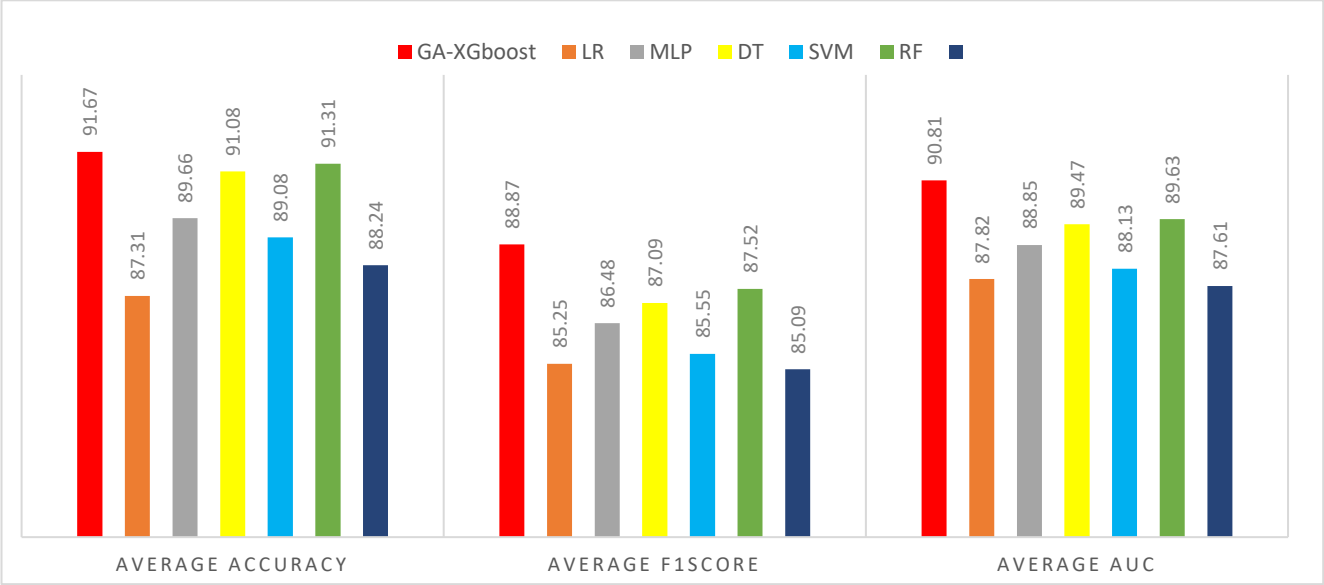

Figure 5: Average accuracy of the proposed model compared baseline models.

Figure 5 compares the average accuracy of our proposed model against Support Vector Machines (SVM), Long-Short-Term Memory (LSTM), Logistic Regression (LR), Multi-Layer Perceptron (MLP), Decision Trees (DT), and Random Forests (RF).
The results show the average accuracy of all presented methods. We observe that the RF and DT gave approximate accuracy results to our proposed model while outperforming the rest of the LR, LSM, and MLP modes. However, in terms of F1 score and AUC, our model exceeds the rest by at least 1%.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 3 delineates the variations in average accuracy, F1 score, and AUC. The findings demonstrate that our suggested model enhances accuracy by 0.36-4.36%, F1 score by 1.34-3.78%, and AUC by 1.18-3.20% relative to the baseline model.

Table 3: Performances differences

Genetic Algorithm & Xgboost	**LR**	**MLP**	**DT**	**SVM**	**RF**	**LSTM**
Accuracy Differences	4,365	2,01	0,595	2,597	0,368	3,438
F1Score Differences	3,615	2,39	1,778	3,318	1,345	3,783
AUC Differences	2,995	1,96	1,338	2,68	1,182	3,203

Figure 6 demonstrates that our proposed model outperformed the others, hence yielding superior average accuracy, F1 score, and AUC across all selected time intervals (5, 15, 25, and 35 weeks). Moreover, it shows a positive correlation between advancement in the course weeks and average performance for both our proposed model and the other models. The performance improves due to the accessibility of student-related data.

The findings indicate that our proposed model achieves a remarkable result and outperforms the baseline model. This strength is due to the combination of the Genetic Algorithm (GA) optimization capabilities for feature selection with the strengths of the XGBoost architecture.
Moreover, the DT and RF models also predict student dropout well. For the Random Forest (RF) method, accuracy rates vary from 80.89% in the first five weeks to each 98.02% in the 35th. Also, the Decision Tree (DT) algorithm has 80.41% to 97.78% accuracy. These findings revealed that these algorithms can predict at-risk students.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

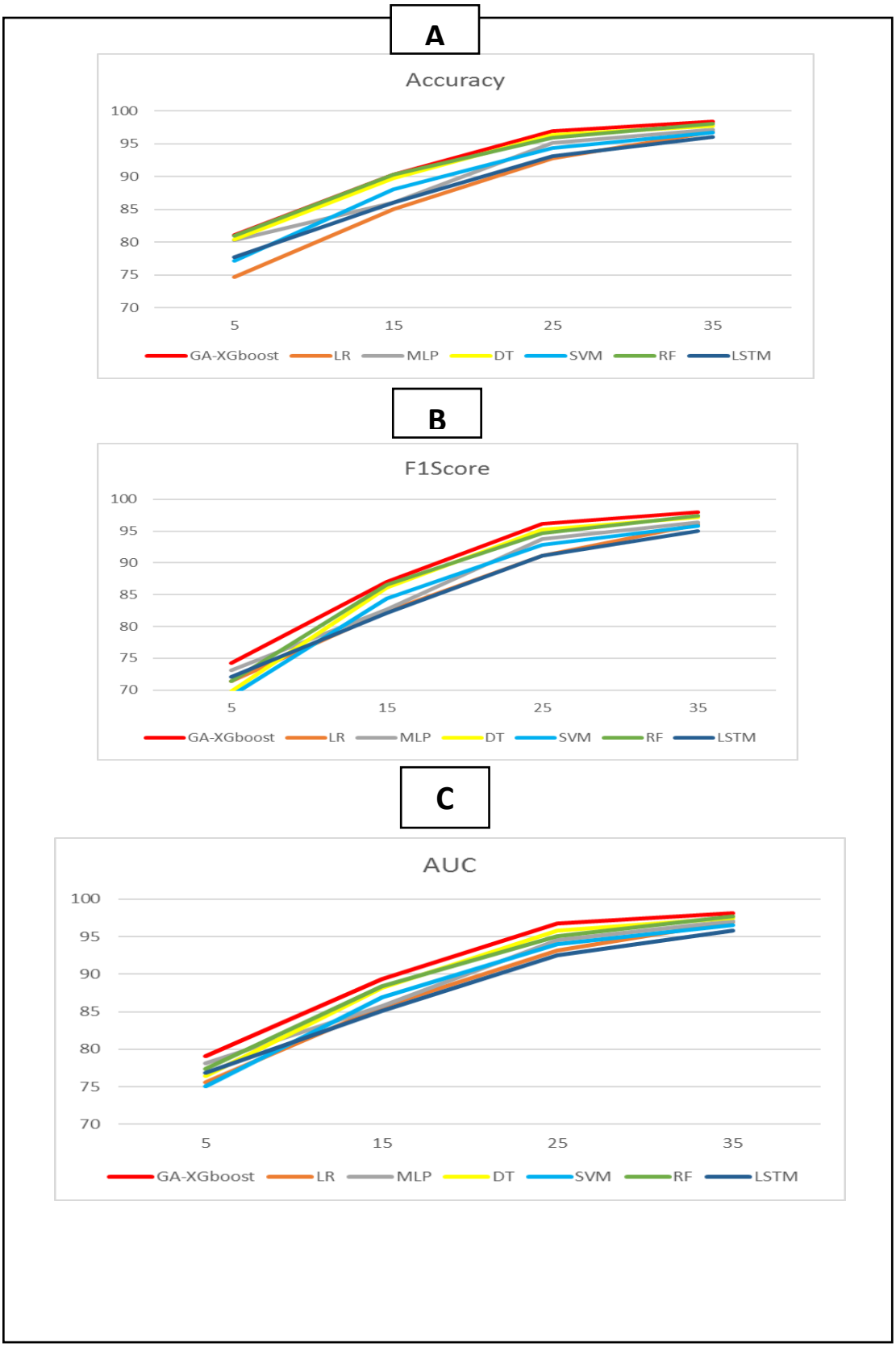

Figure 6: Performances of employed models during weeks of the course.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

CONCLUSION

This study aimed to develop an early predictive model of students at risk of withdrawal from courses in MOOC platforms. Implementing dropout prediction in (MOOCs) can be achieved by utilizing classification techniques. The proposed predictive model is composed of the Genetic and the XGBoost algorithms. This model has been trained and evaluated using the clickstream data from the OULA dataset. The assessment of the suggested model illustrates the effectiveness and resilience of Genetic and the XGBoost algorithm in forecasting at-risk students. The experimental results demonstrate that the model achieved average performances of 91.67% for Accuracy, 88.87% for F1 Score, and 90.81% for AUC, respectively. Furthermore, the model demonstrated enhanced performance relative to baseline models, including RF, DT, LR, SVM, MLP, and LSTM, in terms of performance measures. As a limitation, this research must evaluate our model against more advanced algorithms and assess its efficacy on more datasets. Finally, further tests and investigations are necessary for future studies to enhance accuracy, develop a daily prediction model, and include other variables.

REFRENCES

[1] Brinton, C. G., S. Buccapatnam, M. Chiang, and H. Poor. 2015. “Mining MOOC clickstreams: on the relationship between learner behavior and performance.” arXiv preprint arXiv:1503.06489.

[2] Cobos, R., A. Wilde, and E. Zaluska. 2017. “Predicting attrition from massive open online courses in FutureLearn and edX.” In Proceedings of the 7th International Learning Analytics and Knowledge Conference, Simon Fraser University, Vancouver, BC, Canada, 13–17.

[3] Pappano, L. 2012. “The Year of the MOOC.” The New York Times 2 (12): 2012.

[4] Nadar, N., & Kamatchi, R. (2018). A novel student risk identification model using machine learning approach. Int. J. Adv. Comput. Sci. Appl, 9 (11), 305–309

[5] Medina-Labrador, M., Martinez Quintero, S., Escobar Suarez, D., & Sarmiento Rincón, A. (2022).Dropout reduction in moocs through gamification and length of videos. In Technology-enabled innovations in education: Select proceedings of ciie 2020 (pp. 255–267). Springer.

[6] Chanaa, A., et al. (2022). Sentiment analysis on massive open online courses (moocs): Multi- factor analysis, and machine learning approach. International Journal of Information and Communication Technology Education (IJICTE), 18 (1), 1–22.

[7] Aljohani, N. R., Fayoumi, A., & Hassan, S.-U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability, 11 (24), 7238

[8] Calvet Liñán, L., & Juan Pérez, Á. A. (2015). Minería de datos educativos y análisis de datos sobre aprendizaje: Diferencias, parecidos y evolución en el tiempo. International Journal of Educational Technology in Higher Education, 12 , 98–112

[9] Bakhshinategh, B., Zaiane, O. R., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. Education and Information Technologies, 23 , 537–553

[10] Piety, P. J., Hickey, D. T., & Bishop, M. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 193–202).

[11] Russom, P., et al. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19 (4), 1–34.

[12] Hassan, S.-U., Waheed, H., Aljohani, N. R., Ali, M., Ventura, S., & Herrera, F. (2019). Virtual learning environment to predict withdrawal by leveraging deep learning. International Journal of Intelligent Systems, 34 (8), 1935–1952.

Journal of Information Systems Engineering and Management
2025, 10(57s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[13] Muthukumar, V., & Bhalaji, D. N. (2020). Moocversity-deep learning based dropout prediction inmoocs over weeks. Journal of Soft Computing Paradigm, 2 (3), 140–152.

[14] Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., .Cechinel, C. (2020). A learning analytics approach to identify students at risk of dropout: A case study with atechnical distance education course. Applied Sciences, 10 (11), 3998.

[15] Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from vle big data using deep learning models. Computers in Human behavior, 104 , 106189.

[16] Mubarak, A. A., Cao, H., & Hezam, I. M. (2021). Deep analytic model for student dropout prediction in massive open online courses. Computers & Electrical Engineering, 93 , 107271.

[17] Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., ... & Khan, S. U. (2021). Predicting at-risk students at different percentages of course length for early intervention using machine learning models. Ieee Access, 9, 7519-7539.

[18] Dass, S., Gary, K., & Cunningham, J. (2021). Predicting student dropout in self-paced MOOC course using random forest model. Information, 12(11), 476.

[19] Basnet, R. B., Johnson, C., & Doleck, T. (2022). Dropout prediction in Moocs using deep learning and machine learning. Education and Information Technologies, 27(8), 11499-11513.

[20] Porras, J. M., Porras, A., Fernández, J. A., Romero, C., & Ventura, S. (2023). Selecting the Best Approach for Predicting Student Dropout in Full Online Private Higher Education. In LASI Spain.

[21] Pecuchova, J., & Drlik, M. (2023). Predicting Students at Risk of Early Dropping Out from Course Using Ensemble Classification Methods. Procedia Computer Science, 225, 3223-3232.

[22] Patel, K. K., & Amin, K. (2024). Predictive modeling of dropout in MOOCs using machine learning techniques. The Scientific Temper, 15(02), 2199-2206.

[23] Fayyad, U. M. (1996). Piatetsky-Shapiro, G., Smyth. Advances in Knowledge Discovery and Data Mining.

[24] Bowers, A. J., & Zhou, X. (2019). Receiver operating characteristic (roc) area under the curve(auc): A diagnostic measure for evaluating the accuracy of predictors of education outcomes. Journal of Education for Students Placed at Risk (JESPAR), 24 (1), 20–46

[25] Fawcett, T. (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), 861-874.