Journal of Autonomous Intelligence (2024) Volume 7 Issue 4
doi: 10.32629/jai.v7i4.1443

ORIGINAL RESEARCH ARTICLE

Predictive assessment of learners through initial interactions with encoding techniques in deep learning

Mariame Ouahi[1,*], Samira Khoulji[1], Mohammed Laarbi Kerkeb[2]

[1] *Innovative Systems Engineering Lab. (ISI), Artificial intelligence and digitalization department, National School of Applied Sciences, Abdelmalek Essaâdi University, Tetouan 93000, Morocco*

[2] *Ibn Tofail University, Kenitra BP 242, Morocco*

*** Corresponding author:** Mariame Ouahi, mouahi@uae.ac.ma

ABSTRACT

Recent academic research has prominently focused on predicting student achievements in online learning. However, ongoing challenges persist for teachers and researchers, primarily revolving around the selection of relevant features. To address this issue, the study aims to predict the correlation between the use of a Virtual Learning Environment (VLE) and students' performance based solely on their initial interactions with the platform during the initial segment of the course module within the annual study sequence of the academic year 2014–2015. The assessment employed a specialized data model that integrates the chronological order of a learner's interactions with the educational materials provided on the online learning platform throughout the sequence employing data from the Open University Learning Analytics Dataset (OULAD), which included 32,593 undergraduates in all courses. Innovative methods, such as the VLE, and Long Short-Term Memory (LSTM) neural network rooted in deep learning, were applied to ensure the prediction of student achievement. Preliminary findings demonstrated a 60% accuracy of the model using only learner interaction data within the first third of the course duration and employing various variable encoding techniques. The successful integration of the esteemed OULAD, combined with the implementation of an LSTM neural network architecture in our data model, proved to be a highly effective strategy. This approach yielded valuable perceptions facilitating the identification of crucial data points essential for predicting student success and contributing to informed decision-making in formulating strategies to assess and support student performance effectively.

Keywords: virtual learning environment; chronological sequence; deep learning; predicting a student's success; variable encoding; e-learning; LSTM; OULAD

ARTICLE INFO

Received: 20 November 2023
Accepted: 5 December 2023
Available online: 26 January 2024

1. Introduction

E-learning is a broad term that encompasses various learning approaches involving technological tools like computers, web applications, and smartphones. It is not a pedagogical approach in itself but serves as a set of tools for implementing different pedagogical methods[1]. The success of e-learning is on the rise offering numerous advantages. It enables learners to overcome obstacles such as time and location constraints. In other words, individuals can engage in training entirely online at their own pace and convenience[2].

An additional significant advantage of e-learning is the accessibility of learning trace data, which includes a variety of information about learners' interactions with educational activities[3,4]. Today, online learning platforms collect vast and diverse data on each learner's learning journey[5–7]. These data are like a treasure trove providing important ideas for training managers and learners alike. For educational managers, these data serve as relevant indicators regarding training quality and avenues for optimizing the learning process. Learners, on the other hand, can gain clarity on the most suitable learning styles for their needs. However, extracting meaningful information from raw data necessitates the application of data analysis methods[8]. The concept of a virtual learning environment (VLE) could be considered as a dynamic concept due to the constant evolution of digital technologies, its features and potentialities, and the importance that such environments have within the learning processes[9].

An increasing number of universities, schools, and companies are incorporating web-based educational systems, not only to integrate web technology into their courses but also to complement traditional face-to-face instruction. These systems accumulate a substantial amount of data valuable for analyzing course content and students' utilization. Learning environments relying on technology and digital resources play a mediating role in the learning process by enabling various activities. They facilitate interaction and interrelation within a continuous communication process, thereby enhancing the construction and reconstruction of knowledge, meanings, as well as the formation of habits and attitudes within a common framework for all involved in the educational process[10].

From a pedagogical perspective, Virtual Learning Environments (VLEs) employed in educational institutions play a crucial role in advancing education and fostering innovative experiences[11], even though their primary emphasis tends to be on content production and distribution. These environments often replicate traditional teaching methods by digitally disseminating content, messages, and notices, and facilitating communication through discussion forums and chats[12]. Previous studies have leveraged VLEs to facilitate learning, considering the elements present in the learning environment, ranging from specified elements to those emerging from use. For example, the study conducted by Dahlstrom et al[13] revealed that 74% of teachers consider VLEs highly useful tools for improving teaching, with 71% believing they significantly contribute to enhancing students' learning. The research also indicated extensive VLE usage, with 99% of institutions employing a VLE, 85% of teachers utilizing it, and 56% using it daily. Furthermore, 83% of students engage with the VLE, with 56% indicating its use in all or most course units. Miranda et al[14] identified that VLE tools most valued by over 90% of teachers include resources supporting the course unit, notices, messages, students' registers, and summaries. The same authors emphasized teachers' high regard for digital resource features such as accessibility, user-friendliness, integration with the virtual learning environment, and PDF download.

In a separate study at the University of Maryland in the United States, it was concluded that students who received low grades used the VLE 40% less than those earning C grades or higher. Another study at California State University, Chico, found that VLE use can serve as a proxy for student effort, explaining 25% of the variation in final grades[15]. Additionally, Wolff et al.[16] developed and tested models aiming to predict students' failure by utilizing VLE data in conjunction with assessment data, drawing from the historical record of activities in the VLE and other sources of data. On the other hand, Alves et al.[17] asserted that the primary potential of VLEs lies in providing a set of tools to support the production and distribution of content, communication, and the assessment of the teaching and learning process. Despite the primary focus on content production and distribution, VLEs used in educational institutions contribute to pedagogical advancement and the creation of innovative learning experiences. These environments, in their emulation of traditional teaching methods, digitally disseminate content, messages, notices, and facilitate online communication through discussion forums and chats.

To address this issue, there is an imperative need for a precise selection of characteristics, variables, or pertinent data to predict learner performance. We propose a strategic approach to this challenge. Initially, we will define a comprehensive data model to summarize the learning efforts of each learner. Subsequently, we will formulate a prediction problem derived from this data model, specifically aimed at forecasting learner performance using only the data collected during the initial third of the course duration. To validate this approach, we will conduct an experiment using OULAD and implement a classification model based on the LSTM architecture. The LSTM architecture, an enhanced form of Recurrent Neural Network (RNN)[18], is particularly well-suited for processing time series data[19]. This architectural choice enables us to effectively consider the chronological aspect inherent in a learner's diverse interactions with activities available on an e-learning platform.

Overall, the primary goal of this study is to depict learner data in the form of a sequence of activities. The experiment is centered on predicting a learner's outcome, which is classified as either pass with distinction, just pass, fail, or dropout; thus, presenting a multi-class classification challenge utilizing the OULAD dataset. Our approach unfolds in the following manner:

1) Starting with a concise literature review highlighting models to expect learners' performance in e-learning, with a specific emphasis on deep learning techniques.
2) Demonstrating our proposal, providing details on the learner data model and the formulated learning problem
3) Presenting the key components of our experiment, including the characteristics of the OULAD dataset and the implemented LSTM model with using variable encoding which is the one-hot encoding and word embedding.
4) Analyzing the achieved results, drawing conclusions, and discussing future considerations.

2. Materials and methods

2.1. Literature review: productive learning in the learning environment

Predicting a learner's performance in an online course presents various challenges, as discussed in recent literature[20], emphasizing its primary application in classifying and regressing problems. The learning contexts range from specific courses at individual universities[21,22] to MOOC courses[23,24] and modules in secondary schools[25,26]. Many studies focus on employing deep learning techniques, particularly highlighting the use of LSTM architecture, and consider temporal aspects of the data for early predictions with the possibility of implementation. One noteworthy study conducted by Al-azazi and Ghurab[27] proposes a daily prediction model for a multi-class classification problem using the OULAD dataset. The approach combines LSTM architecture[28] with a simple neural network, termed ANN-LSTM[25]. The results demonstrate a significant improvement, surpassing state-of-the-art models on the dataset by 6 to 14%. However, a redundancy issue arises in learner information representation, as learners are characterized by both demographic data and the number of clicks on learning activities. In this daily prediction model, demographic information is repeatedly associated with the number of clicks over several days, posing a potential optimization challenge due to increased memory capacity demands.

In addition, Okubo et al.[21] proposes a model for predicting the final result of learners from learning traces collected through an LMS. Here, the particularity is the nature of the data collected; they are well organized and codified by the LMS. This data only contains learner learning efforts; it does not contain demographic information. They did not mention any data concerning pre-processing tasks; however, it is important to specify the encoding techniques used. Their approach takes into account the temporality of the data thanks to a simple architecture of RNN. The results obtained demonstrate the relevance of this type of architecture predicting the result of a learner.

The study by Aljaloud et al.[29] focusses on expecting learners' performance within a university training program comprising seven courses. Unlike typical scenarios, this study takes into account the learning efforts of the learners across multiple courses simultaneously. The data is sourced from the Blackboard system, with a unique feature of excluding demographic information and relying solely on seven entirely quantitative variables. Notably, the study achieves significant results (94.2%) by employing all seven variables in combination, highlighting the effectiveness of considering these variables collectively.

In Qu's et al. study[30], an integrated model is proposed, comprising an LSTM neural network for capturing the learner's learning process. This is followed by a DSP to categorize the learner based on four specific behaviors. Lastly, an attention mechanism is employed to predict whether the learner is likely to fail the exam within the context of MOOCs, taking into account the information extracted in the prior stages. In the conducted experiment, the selected characteristics mainly revolve around the time of submission of evaluations. The achieved accuracy is notably high at 91%, and an intriguing conclusion is drawn. The study observes that learners who fail exhibit a certain similarity in their learning behavior, distinguishing them from those who succeed.

According to these authors, this observation suggests that success in learning is not confined to a singular path, rather, there exist numerous learning behaviors that can lead to success. Conversely, failure tends to be associated with consistently poor learning behaviors. In recent research on predicting learner outcomes in e-learning, various models emphasize specific performance metrics such as accuracy, precision, recall, and F1-score. However, it is crucial to acknowledge the substantial differences, including the classification problems addressed, the data employed, the training context, and the content itself. This variability is evident not only in the previously mentioned works but also in studies by Liu et al. and Arsad et al.[31,32].

2.2. Predicting a learner's outcome based on types of learning activities, their dates, and the spent duration

Our methodology adhered to a workflow illustrated in **Figure 1**. It starts with data collection, proceeds to variable selection and transformation, and then involves the training of various models. Ultimately, the performance of these models is assessed in the final phase of our approach.

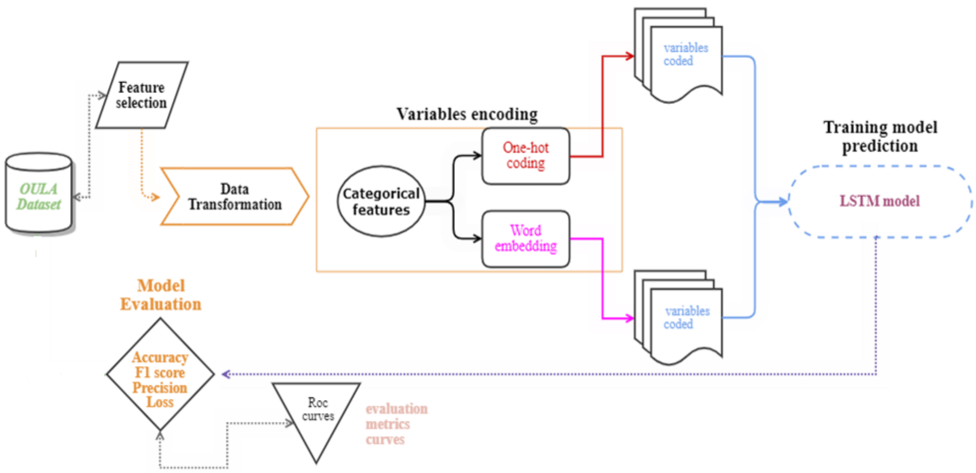


Figure 1. The followed workflow.

2.2.1. Feature selection

To simplify the practical application of corrective measures for possible failures, effective feature engineering is essential. It is an integral aspect of our approach, encompassing the selection or generation of variables within a dataset to enhance the performance of machine learning models. Our model involves three essential components of a learner's learning process. This includes feature selection, a process involving the removal of unnecessary or redundant variables. The elimination of irrelevant variables necessitates a thorough evaluation of their significance, often achieved by constructing a model to assess their correlation with the dependent variable. Following the application of feature selection techniques, the relevant variables used to meet the objectives are outlined in the "studentVle" table. This table provides information about each student's interactions with the VLE. Specifically, the used columns from this file include:

- "date": representing the date of the student's interaction with the material, measured as the number of days since the start of the module presentation.
- "sum click": denoting the number of times a student engages with the material on a given day.

The second table, "VLE," encompasses details regarding the materials accessible in the VLE such as HTML pages and PDF files. Students can access these materials online, and their interactions are systematically recorded. For our analysis, we exclusively utilize the "activity type" column, which specifies the role associated with the module material.

The "studentInfo.csv" table comprises demographic details about the students and their module-presentation results. Our analysis focuses exclusively on the "final result" column, which indicates the students' outcome in the module-presentation. Furthermore, another contributing factor to the success of deep learning is its ability to circumvent the necessity for the feature engineering process. In traditional machine learning, feature engineering involves selecting the most relevant features essential for algorithm functionality and discarding non-informative attributes. This process is challenging and time-consuming, as the accurate choice of features is crucial to the system's performance. In contrast, deep learning employs feature learning to autonomously discover the representations needed for the particular task at hand.

2.2.2. Data transformation

To facilitate the development of effective prediction models for learner performance, we suggest transforming each learner's data into a two-dimensional matrix illustrated in **Figure 2**. This matrix comprises N rows and M columns, where N (greater than or equal to 1) denotes the number of interactions based on the date of consultation and the type of resources. M, with a minimum value of 3 variables, includes at least the type of resource, date of consultation, and the number of clicks. Additional information determining the learner's effort, such as consultation duration and progress level, can be incorporated into these columns.

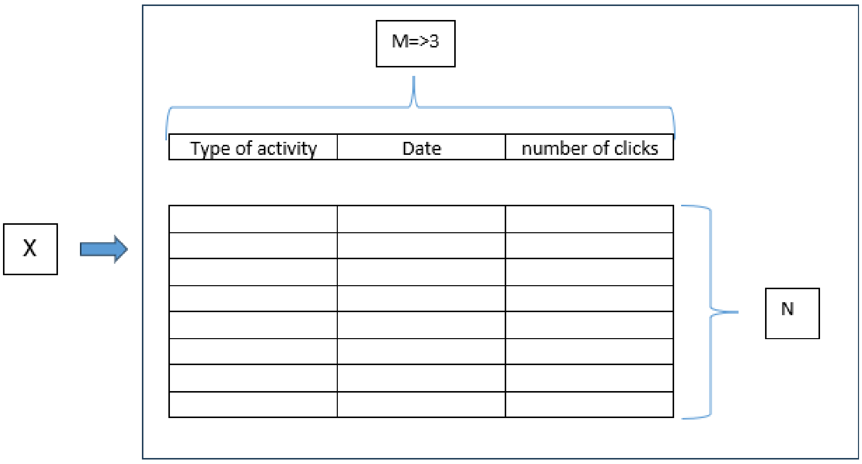


Figure 2. The proposed data model.

2.2.3. Variables encoding

The data comprises both qualitative and quantitative variables. The qualitative variables, such as activity types (forum, web page, video), will be indexed, assigning each modality an integer value. This indexing process is crucial for machine learning to process the data, converting it into a binary vector representation suitable for machine learning algorithms. To encode these indices, two distinct techniques, namely one-hot encoding and word embedding, will be employed. On the other hand, quantitative variables (number of clicks, duration of consultation, date) will undergo normalization using the mean and standard deviation. This normalization step helps to improve the performance of the model (Equation (1)).

$$\frac{x - mean}{standard\ deviation} \tag{1}$$

One-hot encoding

One-hot encoding is widely used in machine learning and deep learning applications due to its efficacy and straightforward implementation. The process involves creating a binary vector for each category within the original categorical variable. For instance, with a categorical variable featuring three categories ("A," "B," and "C"), one-hot encoding would generate three binary vectors: [1, 0, 0], [0, 1, 0], and [0, 0, 1][33].

An essential advantage of one-hot encoding lies in its prevention of introducing ordinal relationships between categories. Each category is treated as an independent entity, eliminating any numerical assumptions or biases in the model's performance[34].

Moreover, one-hot encoding is capable of handling categorical variables with any number of categories, making it suitable for diverse datasets. Furthermore, it is well-supported by most machine learning libraries and frameworks, ensuring ease of implementation across various applications.

However, it is crucial to acknowledge that one-hot encoding can result in a high-dimensional feature space, especially when dealing with categorical variables featuring numerous categories. This might increase the complexity of the model and potentially lead to the "curse of dimensionality," where computational resources and the risk of overfitting increase significantly.

Word embedding

Word embedding is also a widely employed technique in natural language processing (NLP) that represents words as dense, continuous vectors in a lower-dimensional space. The primary objective of word embedding is to capture the semantic relationships and contextual information of words, enhancing the ability of machine learning models to comprehend and process language more effectively[35]. Unlike conventional encoding methods like one-hot encoding, where each word is represented as a sparse binary vector, word embedding employs dense vectors that encode the meaning and context of words based on their usage in a given corpus of text.

The process of generating word embedding entails training a neural network on a substantial corpus of text data. The network learns to map words to their corresponding vector representations by discerning the context in which they appear. Word embedding offers several advantages over traditional encoding methods:

- Semantic Relationships: Word embedding captures semantic relationships between words. Words sharing similar meanings or contexts tend to have comparable vector representations, enabling the model to understand analogies and associations between words[36].
- Dimensionality Reduction: Word embedding transforms high-dimensional, one-hot encoded word representations into lower-dimensional continuous vectors. Reducing dimensionality improves computational efficiency and decreases memory demands for NLP tasks[37].
- Contextual Information: Embeddings consider the context in which words appear, resulting in better representations of polysemous words (words with multiple meanings) and resolving ambiguities based on surrounding words[38].
- Generalization: Word embedding generalize well to unseen or rare words that were not present in the training data. They capture the underlying semantic properties shared between words, allowing for

effective handling of new vocabulary[39].

Several popular word embedding techniques, including Word2Vec[34], GloVe (Global Vectors for Word Representation), and Fast Text, have been pre-trained on extensive text corpora. These methods are easily applicable to various NLP tasks such as sentiment analysis, machine translation, and text classification[40]. While word embedding was initially developed for NLP, its applicability has expanded beyond this field.

2.3. Deep learning model suitable for temporal data: case of an LSTM architecture

To ensure the temporal dimension of the data, we opted for an LSTM neural network architecture. These networks are specifically engineered to address the challenge of long-term dependencies encountered by (RNNs); a problem often associated with the vanishing gradient issue. LSTMs incorporate feedback connections, setting them apart from traditional feed-forward neural networks. This unique characteristic enables LSTMs to handle entire sequences of data, such as time series, by retaining valuable information from preceding data points. This approach facilitates the processing of new data points in the context of the entire sequence. Consequently, LSTMs prove particularly effective in handling sequential data types like text, speech, and general time series[41].

2.4. Experimentation

OULA dataset

The Open University Learning Analytics Dataset (OULAD) is an extensive educational dataset generously provided by the Open University (OU) in the United Kingdom. Tailored for research endeavors, particularly within the field of learning analytics, this dataset is a trove of detailed information pertaining to students' engagements with online learning platforms. It involves a wide array of data, ranging from demographics and academic performance to learning activities and engagement patterns. It is noteworthy that the dataset has been anonymized to safeguard the privacy of both students and researchers working with this valuable activity[42].

Our study relies on the OULAD dataset, which provides a comprehensive collection of data related to courses, students, and their interactions with the VLE. The dataset is specifically chosen for research purposes and focuses on seven selected courses, referred to as modules within the dataset. This dataset serves as the foundation for our investigation, allowing us to analyze various aspects of students' interactions and learning experiences within the specified VLE (**Table 1**). Here are some of the key components of the OULAD:

- Student information: This includes demographic information about students, such as age, gender and location. It allows researchers to analyze the influence of these factors on students' academic performance and engagement.
- Course information: Details of the courses and modules taken by students are provided. This includes the course code, the length of study and the final outcome of each module.
- Assessment data: Information on student performance in assessments, quizzes, exams and assignments is included. This data allows researchers to explore the relationship between learning activities and academic performance.
- VLE interaction data: The dataset records students' interactions with the VLE, the online platform used by the Open University. This interaction data includes logins, clicks, downloads, forum participation and other VLE activities.
- Clickstream data: Clickstream data refers to the sequence of actions taken by students as they navigate through the online learning platform. It provides insight into how students engage with course materials and resources.
- Dropout and retention information: The dataset also contains information about students who have

dropped out of courses and those who have remained enrolled, allowing researchers to investigate the factors that influence student retention.

The OULAD has been widely used by researchers in the field of learning analytics and educational data mining to develop predictive models, analyzing student behavior and gaining perceptions into effective learning strategies. Its availability has been instrumental in advancing the understanding of online learning and student success. Again, it is important to note that the dataset used in this study is the OULAD dataset. This dataset involves information about courses, students, and their interactions with the VLE across seven selected courses referred to as modules within the dataset.

Table 1. The various aspect of the dataset (OULA dataset description)[42].

Description	Value
Years	2015
Courses	7 courses
Number of courses introduced	22
Min-max course period	235–270 days
Number of unique students	28,785
Number of students in all courses	32,593
Number of assessments	206
Number of activities	6364

3. Results

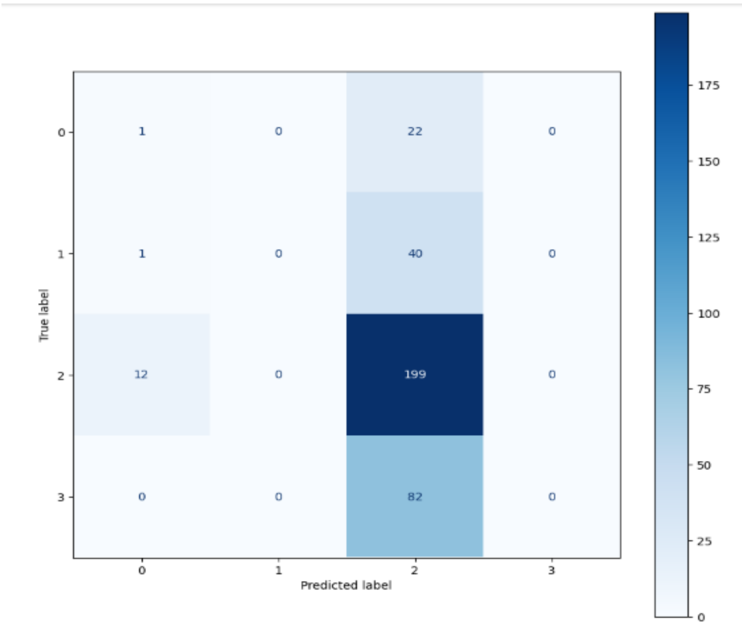
In our study, we conducted a comparative analysis of the LSTM model using two approaches: one-hot encoding and word embedding. This comparison aimed to evaluate their respective effectiveness in the classification task at hand.

The results, as presented in **Table 2**, play a crucial role in offering clear insights into the model's performance with each encoding method. These findings enable us to make informed decisions regarding the most suitable approach for our specific classification task. Following a thorough analysis of these results, we have concluded that one-hot encoding proves to be the most suitable method for optimizing the performance of our LSTM model.

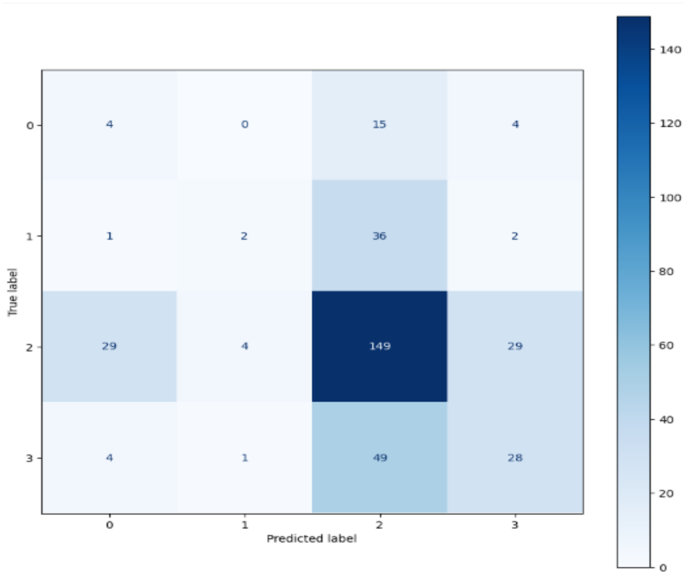
Table 2. Model performance.

Evaluation metrics	LSTM	
	One-hot	Word embedding
Accuracy	0.55	0.60
Loss	1.135	1.186

Certainly, the application of word embedding has produced noteworthy outcomes, showcasing an accuracy rate of 60% and a cost function value of 1.186. This marks a 5% improvement in accuracy compared to the use of the one-hot encoding technique. These metrics indicate a relatively high level of performance and an enhanced ability of the model to generalize when incorporating word embedding. Consequently, the adoption of word embedding contributes to an overall improvement in the LSTM model's precision and reliability in classification tasks. This confirms the efficacy of the word embedding approach in representing data within the specific context of our experiment. Additionally, the examination of confusion matrices for both techniques reinforces our confidence in the superiority of word embedding. Upon closer examination of these matrices, a clear pattern emerges, depicting a more favorable distribution of predictions for word embedding in comparison to one-hot encoding (**Figure 3**).



(a)



(b)

Figure 3. Confusion matrix for LSTM **(a)** with word embedding **(b)** and with one-hot encodin.

The enhanced accuracy achieved with word embedding has been translated into a confusion matrix with fewer classification errors. The values for true positives, true negatives, false positives, and false negatives collectively indicate an overall superior performance for this approach, which is particularly encouraging.

Tables 3 and **4** provide the metrics of precision, recall, and F1 score[43]. Precision assesses the proportion of true positives among all positive predictions, while recall, also known as sensitivity or true positive rate, evaluates the proportion of true positives among all actual positives. The F1 score measures the harmony between precision and recall, calculated as their harmonic mean. It proves beneficial when striving for a balance between precision and recall, especially in cases where classes are unbalanced. Support indicates the total number of real instances for each class in the dataset, representing the number of actual occurrences of each class.

Table 3. Accuracy prediction models with word embedding.

Labels	Precision	Recall	F1-score	support
class 0	0.07	0.04	0.05	23
class 1	0.00	0.00	0.00	41
class 2	0.59	0.96	0.73	211
class 3	0.38	0.06	0.11	82
Accuracy	0.00	0.00	0.60	357
Macro avg	0.24	0.25	0.21	357
Weighted avg	0.44	0.58	0.46	357

It is evident from both approaches that predicting Class 2 is notably more straightforward. The performance metrics and accuracy rates consistently demonstrate a higher level of predictability for Class 2 in comparison to other classes. This observation suggests that the model exhibits greater efficacy and reliability when distinguishing instances belonging to Class 2.

Table 4. Accuracy prediction models with one-hot encoding.

Labels	Precision	Recall	F1-score	support
class 0	0.07	0.04	0.05	23
class 1	0.00	0.00	0.00	41
class 2	0.58	0.94	0.72	211
class 3	0.00	0.06	0.11	82
Accuracy	0.00	0.00	0.55	357
Macro avg	0.16	0.25	0.19	357
Weighted avg	0.35	0.56	0.43	357

In the final results of the anticipated student performance experiment using initial interactions, our approach yields an accuracy of 60%, as highlighted in **Table 3**. This table illustrates the experimental outcomes, indicating that the LSTM model achieved satisfactory performance across various indicators, including precision, recall, F1-score, and accuracy using word embedding.

The relatively lower accuracy value of 60% compared to other studies can be attributed to the unique focus on early predictions of cases at risk of failure or learner outcomes based on the first interactions with educational resources in an online course. This approach involves using data from the initial third of the total learning period, setting it apart from other studies that use data throughout the entire training period. The accuracy of such studies may be higher because they incorporate data from the entire training duration. However, our model aims to showcase the ability to predict outcomes quite early in the training process, offering insights before the training concludes. This emphasis on early prediction adds a valuable dimension to the understanding of learner performance dynamics.

The findings demonstrate that the LSTM model consistently achieved the highest accuracy in predicting students' performance across all datasets without dividing the results into multiple parts. This success can be attributed to the LSTM's distinctive ability to effectively recall and consider previous student behaviors, an inherent feature of the LSTM network architecture.

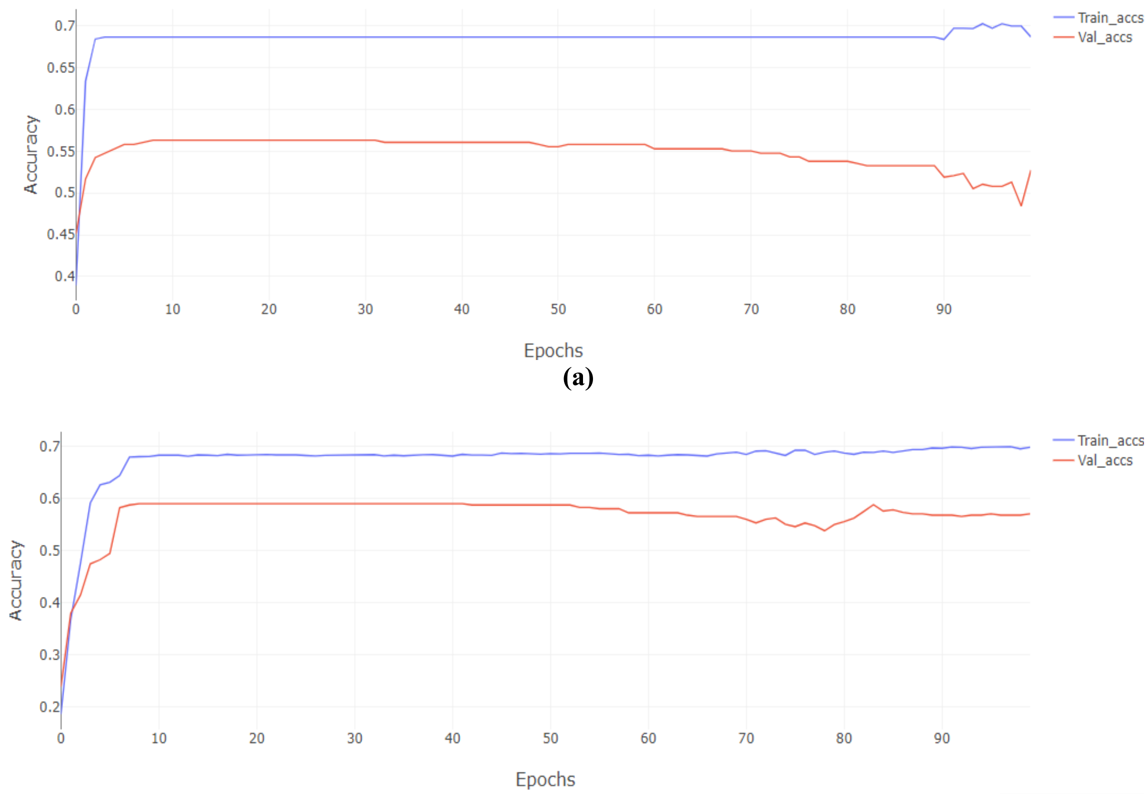
In the evaluation of categorical methods, four fundamental criteria are commonly employed: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). Each criterion is defined as follows:

- Precision = TP/(TP + FP)
- Recall = TP/(TP + FN)
- Accuracy = (TP + TN)/(TP + TN + FP + FN)
- F1 Score = 2 × Precision × Recall/(Precision + Recall)

These criteria offer a comprehensive assessment of the model's predictive performance, taking into account both positive and negative classifications. The quality of the LSTM model across these criteria confirms its effectiveness in capturing and impacting the sequential nature of student behaviors for accurate performance predictions.

In our experiment, we trained the LSTM model to classify the student performances. First, we used the one-hot encoding method to encode a qualitative variable "type of activity". Then, for a different approach, we also used a word embedding encoder to represent this categorical variable.

Figure 4 illustrates the evolution of accuracy concerning the number of iterations for two distinct data representation approaches. The first, depicted in **Figure 4a**, displays the combination of one-hot encoding and word embedding. On the other hand, the second figure (**Figure 4b**) represents the one-hot approach. Notably, the one-hot approach demonstrates greater stability in performance across the various iterations as mentioned in the **Tables 3** and **4**.



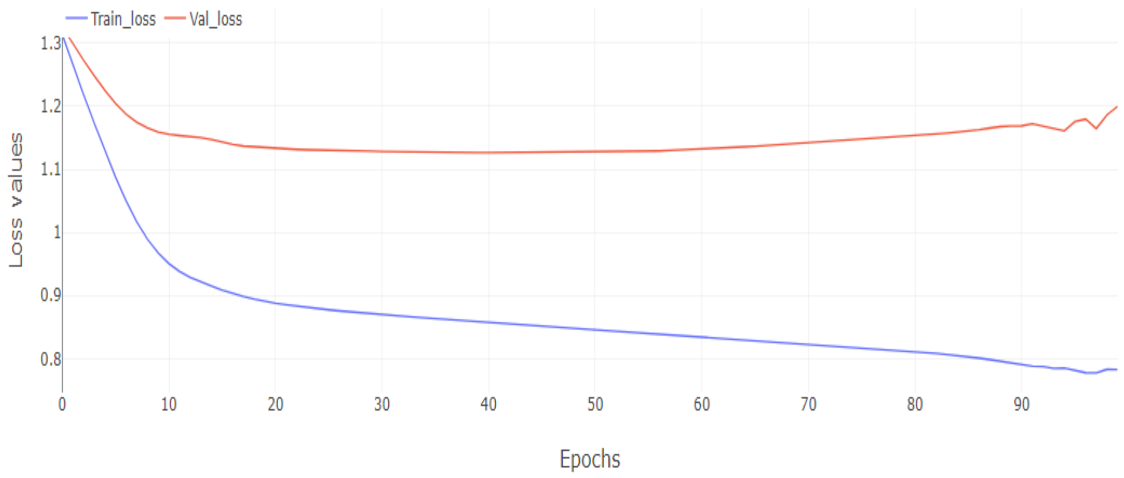
(b)

Figure 4. Line plots the training an validation **(a)** with word embedding **(b)** and with one-hot.

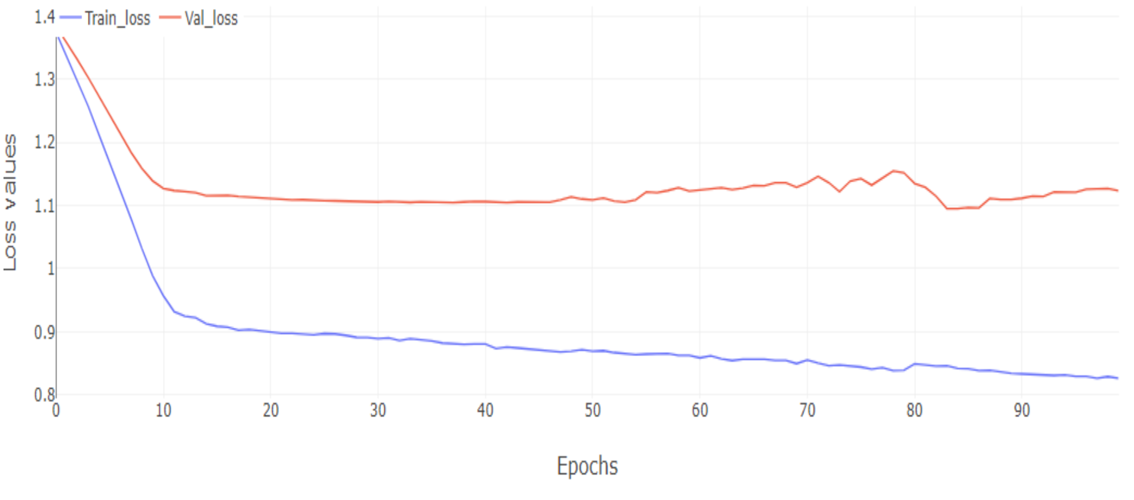
On the contrary, when employing the word embedding approach, we observed a rapid increase in the accuracy of the training set with an augmented number of iterations. However, a notable decline in the accuracy of the test set was observed simultaneously. This phenomenon suggests that the model is entering a state of overfitting to the training data, signifying that it is becoming too tailored to the specifics of the training data and is consequently struggling to generalize effectively to new data in the test set.

Careful observation in monitoring the performance evolution over the test set is crucial to forestall overfitting. Mitigation strategies such as adjusting hyper parameters, incorporating regularization techniques, or simplifying the model's complexity can be instrumental in enhancing its ability to generalize to the test data[44].

As illustrated in **Figure 5**, in both methodologies namely, one-hot encoding and word embedding (**Figure 5a** and **5b** respectfully) we witnessed a progressive convergence of the cost function toward 0.5 on the training data. Nevertheless, distinct patterns emerged when considering the validation data. In the case of word embedding, the cost function began to rise after approximately 90 iterations, signaling the onset of overfitting. Conversely, when employing one-hot encoding, the cost function exhibited instability in the validation data.



(a)



(b)

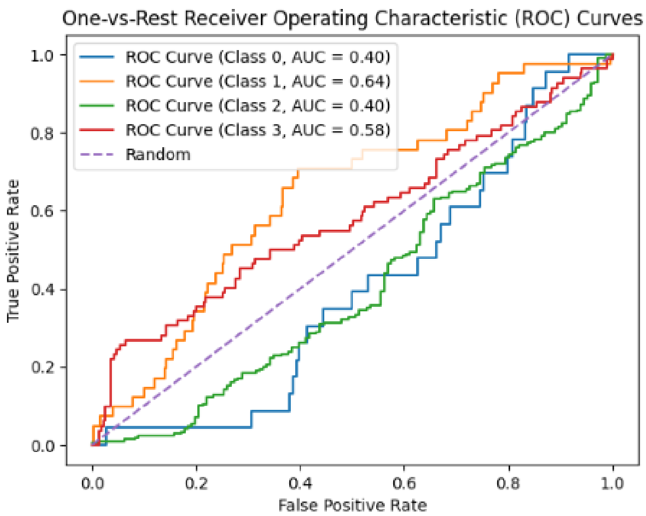
Figure 5. Loss values over several training epochs **(a)** with word embedding **(b)** and with one-hot.

The inconsistent behavior of the cost function on the validation data with one-hot encoding implies a potential challenge in the model's ability to generalize accurately to unknown data. This instability may be attributed to the high dimensionality inherent in one-hot encoded features, leading to increased model complexity.

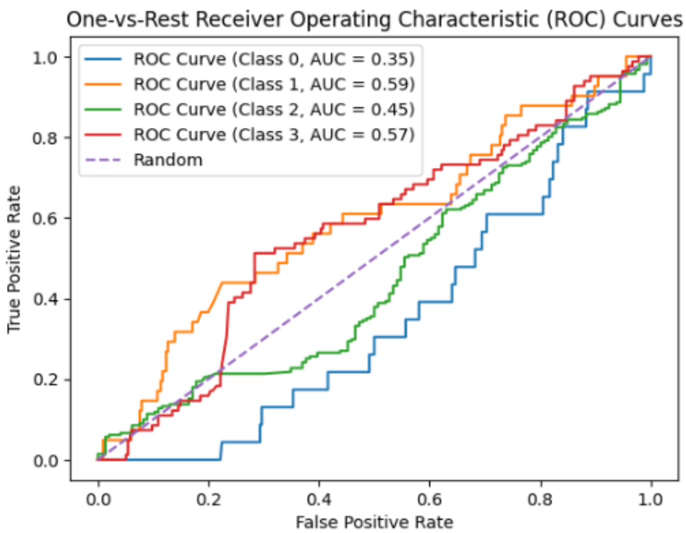
On the other hand, in the case of word embedding, the uptick in the cost function on the validation data around the 90th iteration points to an issue of overfitting, as previously discussed. This suggests that the model, while effectively memorizing the intricacies of the training data, struggles to extend its understanding to new examples, indicating a limitation in its capacity to generalize.

The ROC curves presented in **Figure 6** reveal that the model excels in detecting classes 1 and 3, as evidenced by the curves positioned above the random reference line (AUC > 0.5). This implies that the model can effectively differentiate between these two classes with a certain degree of accuracy, and the true positive rates excel the false positive rates for these classes. However, for the remaining classes, the ROC curves may be closer to the random reference line, indicating that the model encounters greater difficulty distinguishing between them. This is to say that several factors could contribute to this challenge, including potential class imbalance in the data.

Our approach facilitates detection and measurement at multiple points throughout the course. For instance, within a 6-month course, the model can undergo training based on the outcomes of the initial 2 months to identify students who may encounter challenges or face difficulties.



(a)



(b)

Figure 6. Courbe ROC curves **(a)** with word embedding **(b)** and with one-hot.

4. Discussion

In this research, we introduced a model distinguished by its rapidity in forecasting the academic performance of students based on initial interactions with educational resources early in an online course sequence. To construct our predictive model, we incorporated feature selection techniques, prioritizing the creation of simpler and more comprehensible models, enhancing data mining efficiency, and ensuring the preparation of clean data. We implemented variable coding, including methods like one-hot encoding and word embedding, to represent text words through real-number vectors. Our methodology, employing Artificial Neural Networks on the OULAD, yielded a prediction accuracy and precision of 60%. A comparative analysis with other studies revealed varying outcomes.

The study conducted by Kusumawardani and Alfarozi[45], the Transformer Encoder Model for Sequential Prediction of Student Performance based on Log Activities, diverged in results because of the application of a deep learning-based learning analytics method, specifically Transformer Encoding. This method sequentially predicted the final performance of at-risk students based on OULAD log activities. In contrast to our approach, Kusumawardani and Alfarozi proposed model, compared with the LSTM model, demonstrated early-stage predictive accuracy of 83.17% for withdrawal classes versus success-distinction classes. In other tasks like withdrawal-failure classes versus pass-distinction and fail versus pass-distinction, early-stage accuracy remained at least 76%. The study concluded that the Transformer Encoder outperformed LSTM, with mean differences of 1% to 3% in accuracy and 3% to 7% in F1-score.

Contrastingly, Al-Zawqari et al.[46] adopted a flexible feature selection approach for predicting students' academic performance in online courses, employing raw data directly for prediction model construction without a feature engineering step. Feature selection was based on model interpretability, applying two different classifiers-random forest and artificial neural networks to OULAD. Without feature engineering, both achieved high prediction accuracy for at-risk students at 86% and 88% compared to all pass and distinction students.

Similarly, Sehaba[47] focused on analyzing learners' performance to define indicators predicting their results based on interactions with a learning tool. This approach was applied to a dataset from a real training course with 32,593 learners and 10,655,280 events, resulting in a prediction accuracy of around 80%. Rule extraction methods were also employed to elucidate the rules governing the prediction indicator.

Based on these findings, it is evident that the achieved accuracy of 60% in our study represents a noteworthy outcome. This result gains particular significance due to the exclusive emphasis on early predictions of potential failure cases or learner outcomes, especially considering the data from the initial third of the overall learning period. The obtained perceptions from our model contribute to the evolving landscape of predicting student performance and offer valuable considerations for enhancing early-stage predictive analytics in educational settings.

5. Conclusions

The VLE is a dynamic concept and is considered a primary method for predicting student achievements in online learning. However, it faces persistent challenges for universities and teachers. Our study focuses on predicting student performance using the Open University Learning Analytics Dataset (OULAD) of the Virtual Learning Environment (VLE) and an LSTM model. This model analyzes students' performance based solely on their initial interactions with the platform during the initial segment of the course module within the annual study sequence of the academic year 2014–2015. The study encompasses approximately 32,593 undergraduates in all courses. The results derived from the word embedding analysis have yielded notable outcomes, revealing an accuracy rate of 60% and a cost-function value of 1.186. This marks a 5%

improvement in accuracy compared to the use of the one-hot encoding technique. These metrics indicate a relatively high level of performance and an enhanced ability of the model to generalize when incorporating word embedding. Consequently, the adoption of word embedding contributes to an overall improvement in the LSTM model's precision and reliability in classification tasks. This confirms the efficacy of the word embedding approach in representing data within the specific context of our experiment.

In other words, these findings ensure the efficacy of the word embedding approach in capturing meaningful patterns and relationships within the data, contributing valuable perceptions to the overall assessment of the model's performance. The achieved accuracy rate signifies a commendable level of predictive success, demonstrating the model's capability to perceive and predict outcomes effectively. Additionally, the associated cost-function value provides a quantitative measure of the model's performance, indicating its ability to minimize errors and enhance overall efficiency. These results affirm the applicability and effectiveness of the word embedding methodology in the context of the study, paving the way for further exploration and refinement in predictive analytics for educational settings.

Furthermore, the examination of confusion matrices for both techniques strengthens our confidence in the superiority of word embedding. Upon closer exploration of these matrices, an apparent pattern emerges, depicting a more favorable distribution of predictions for word embedding compared to one-hot encoding.

In our forthcoming endeavors, we are committed to expanding and deepening our research scope beyond the current study. Our focus will shift toward a meticulous examination of learner profiles, showing the power of unsupervised learning techniques. This strategic exploration is designed to uncover crucial perceptions into the diverse characteristics and patterns within the learner population, particularly those at risk. Our main objective is to advance the field of educational analytics by refining our understanding and providing a more comprehensive perception into personalized remediation strategies. By tailoring interventions based on individual learner profiles, we aspire to address the main challenges that learners may encounter on their educational journey.

This avenue of research holds immense potential to make substantial contributions to the field. By employing unsupervised learning techniques, we anticipate unveiling hidden patterns, connections, and potential areas of improvement that may not be immediately apparent through traditional analytical methods. The knowledge gained from this exploration will empower educators, institutions, and educational policymakers with accurate and targeted approaches for addressing specific challenges. Ultimately, our goal is to facilitate improved educational outcomes by fostering a more adaptive and responsive educational environment that serves the unique needs of each learner. Through this research initiative, we aim to make lasting contributions to the ongoing evolution of educational practices and methodologies.

Author contributions

Conceptualization, OM, KS and MLK; methodology, OM, KS, and MLK.; software, KS and MLK; validation OM, MLK and KS; formal analysis, OM, MLK and KS.; writing—original draft preparation, OM, KS; writing-review and editing, OM, KS; visualization OM, MLK and KS; supervision, KS and MLK; funding acquisition, OM, KS and MLK; All authors have read and agreed to the published version of the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Abbreviation

LTSM, Long short-term memory; E-learning, Electronic learning; ROC, Receiver Operating

Characteristic; OULAD, Open University Learning Analytics Dataset; RNN, Recurrent Neural Network; DSP, Digital Signal Processing; MOOC, Massive open online course; LMS, Learning Management System.

References

1. Müller FM, Souza MV. The role of Knowledge Media in Network Education. International Journal for Innovation Education and Research. 2020, 8(7): 76-93. doi: 10.31686/ijier.vol8.iss7.2457
2. Goyal S. E-Learning: Future of Education. Journal of Education and Learning (EduLearn). 2012, 6(4): 239. doi: 10.11591/edulearn.v6i4.168
3. Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery. 2020, 10(3). doi: 10.1002/widm.1355
4. Slimani K, Bourekkadi S, Messoussi R, et al. Sharing emotions in the distance education experience: attitudes and motivation of university students. International Conference on Intelligent Systems and Computer Vision (ISCV); IEEE. 2020. pp. 1-10.
5. Bekele R, Menzel W. A bayesian approach to predict performance of a student (bapps): A case with ethiopian students. Algorithms. 2005; 22(23): 24.
6. Mubarak AA, Cao H, Zhang W, et al. Visual analytics of video-clickstream data and prediction of learners' performance using deep learning models in MOOCs' courses. Computer Applications in Engineering Education. 2020, 29(4): 710-732. doi: 10.1002/cae.22328
7. Khadija S, Ruichek Y, Messoussi R. Compound facial emotional expression recognition using cnn deep features. Engineering Letters. 2022; 30(4): 1402-1416.
8. AhmetBerk U. Effects of Mobile Learning in Blended Learning Environments. Journal of Information and Communication Technologies. 2019; 1(1): 1–14.
9. Torres Martín C, Acal C, El Homrani M, et al. Impact on the Virtual Learning Environment Due to COVID-19. Sustainability. 2021, 13(2): 582. doi: 10.3390/su13020582
10. Schneckenberg D, Ehlers U, Adelsberger H. Web 2.0 and competence-oriented design of learning—Potentials and implications for higher education. British Journal of Educational Technology. 2011, 42(5): 747-762. doi: 10.1111/j.1467-8535.2010.01092.x
11. Annansingh F. Mind the gap: Cognitive active learning in virtual learning environment perception of instructors and students. Education and Information Technologies. 2019, 24(6): 3669-3688. doi: 10.1007/s10639-019-09949-5
12. Galán JG, Pérez CL, López JÁM, et al. VLE Environments and MOOC Courses. Innovation and ICT in Education. Published online September 1, 2022: 77-91. doi: 10.1201/9781003338567-9
13. Jokiaho A, May B, Specht M, et al. Barriers to using E Learning in an Advanced Way. International Journal of Advanced Corporate Learning (iJAC). 2018, 11(1): 17. doi: 10.3991/ijac.v11i1.9235
14. Miranda L, Alves P, Morais C. Assessment of virtual learning environments by higher education teachers and students. ECEL 2013-Proceedings for the 12th European Conference on eLearning: ECEL 2013. Academic Conferences Limited. 2013. p. 311.
15. Sclater N, Peasgood A, Mullan J. Learning analytics in higher education. A review of UK and international practice. 2016; 8(2017): 176.
16. Kuzilek J, Zdrahal Z, Fuglik V. Student success prediction using student exam behaviour. Future Generation Computer Systems. 2021, 125: 661-671. doi: 10.1016/j.future.2021.07.009
17. Alves P, Miranda L, Morais C. The Influence of Virtual Learning Environments in Students' Performance. Universal Journal of Educational Research. 2017, 5(3): 517-527. doi: 10.13189/ujer.2017.050325
18. Das CR, Das S, Panda S. Groundwater quality monitoring by correlation, regression and hierarchical clustering analyses using WQI and PAST tools. Groundwater for Sustainable Development. 2022, 16: 100708. doi: 10.1016/j.gsd.2021.100708
19. Mondal A, Mukherjee J. An Approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN). International Journal of Computer Applications. 2018, 181(6): 1-5. doi: 10.5120/ijca2018917352
20. Alamri R, Alharbi B. Explainable Student Performance Prediction Models: A Systematic Review. IEEE Access. 2021, 9: 33132-33143. doi: 10.1109/access.2021.3061368
21. Okubo F, Yamashita T, Shimada A, et al. A Neural Network Approach for Students' Performance Prediction. In the Proceedings of the Seventh International Learning Analytics & Knowledge Conference. 2017; 598-599.
22. Hood N, Littlejohn A, Milligan C. Context counts: How learners' contexts influence learning in a MOOC. Computers & Education. 2015, 91: 83-91. doi: 10.1016/j.compedu.2015.10.019
23. Al-azazi FA, Ghurab M. ANN-LSTM: A deep learning model for early student performance prediction in MOOC. Heliyon. 2023, 9(4): e15382. doi: 10.1016/j.heliyon.2023.e15382
24. Kim BH, Vizitei E, Ganapathi V. GritNet: Student Performance Prediction with Deep Learning. arXiv preprint arXiv.2018; 1804.07405:1-5.
25. Imran M, Latif S, Mehmood D, et al. Student Academic Performance Prediction using Supervised Learning Techniques. International Journal of Emerging Technologies in Learning (iJET). 2019, 14(14): 92. doi: 10.3991/ijet.v14i14.10310

26. Sekeroglu B, Dimililer K, Tuncal K. Student Performance Prediction and Classification Using Machine Learning Algorithms. In: Proceedings of the 8th International Conference on Educational and Information Technology; New York, NY, USA; 2 March 2019. pp. 7–11.
27. Al-azazi FA, Ghurab M. ANN-LSTM: A deep learning model for early student performance prediction in MOOC. Heliyon, 2023, 9(4). doi: 10.1016/j.heliyon.2023.e15382
28. Rather AM. Deep Learning and Autoregressive Approach for Prediction of Time Series Data. Journal of Autonomous Intelligence. 2021, 3(2): 1. doi: 10.32629/jai.v3i2.207
29. Aljaloud AS, Uliyan DM, Alkhalil A, et al. A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM. IEEE Access. 2022, 10: 85255-85265. doi: 10.1109/access.2022.3196784
30. Qu S, Li K, Wu B, et al. Predicting Student Achievement Based on Temporal Learning Behavior in MOOCs. Applied Sciences. 2019, 9(24): 5539. doi: 10.3390/app9245539
31. Liu Q, Huang Z, Yin Y, et al. EKT: Exercise-Aware Knowledge Tracing for Student Performance Prediction. IEEE Transactions on Knowledge and Data Engineering. 2021, 33(1): 100-115. doi: 10.1109/tkde.2019.2924374
32. Arsad PM, Buniyamin B, Manan JLA. A Neural Network Students' Performance Prediction Model (NNSPPM). IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA); IEEE. 2013; 1-5.
33. Cerda P, Varoquaux G, Kégl B. Similarity encoding for learning with dirty categorical variables. Machine Learning. 2018, 107(8-10): 1477-1494. doi: 10.1007/s10994-018-5724-2
34. Hancock JT, Khoshgoftaar TM. Survey on categorical data for neural networks. Journal of Big Data. 2020, 7(1). doi: 10.1186/s40537-020-00305-w
35. Yang H, Kong J, Hu H, et al. A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges. Remote Sensing. 2022, 14(8): 1770. doi: 10.3390/rs14081770
36. Jiao Q, Zhang S. A Brief Survey of Word Embedding and Its Recent Development. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Published online March 12, 2021. doi: 10.1109/iaeac50856.2021.9390956
37. Raunak V, Gupta V, Metze F. Effective Dimensionality Reduction for Word Embeddings. Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019). Published online 2019. doi: 10.18653/v1/w19-4328
38. Peters M, Neumann M, Zettlemoyer L, et al. Dissecting Contextual Word Embeddings: Architecture and Representation. arXiv 2018; arXiv:1808.08949:1-3.
39. Zhao J, Mudgal S, Liang Y. Generalizing Word Embeddings Using Bag of Subwords. arXiv 2018; arXiv:1809.04259:3-5.
40. Sabbeh SF, Fasihuddin HA. A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification. Electronics. 2023, 12(6): 1425. doi: 10.3390/electronics12061425
41. Yu Y, Si X, Hu C, et al. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Computation. 2019, 31(7): 1235-1270. doi: 10.1162/neco_a_01199
42. Kuzilek J, Hlosta M, Zdrahal Z. Open University Learning Analytics dataset. Scientific Data. 2017, 4(1). doi: 10.1038/sdata.2017.171
43. Wang X, Yu X, Guo L, et al. Student Performance Prediction with Short-Term Sequential Campus Behaviors. Information. 2020, 11(4): 201. doi: 10.3390/info11040201
44. IOFFE, S, SZEGEDY. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In International conference on machine learning. 2015; pp. 448-456.
45. Kusumawardani SS, Alfarozi SAI. Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities. IEEE Access. 2023, 11: 18960-18971. doi: 10.1109/access.2023.3246122
46. Al-Zawqari A, Peumans D, Vandersteen G. A flexible feature selection approach for predicting students' academic performance in online courses. Computers and Education: Artificial Intelligence. 2022, 3: 100103. doi: 10.1016/j.caeai.2022.100103
47. Sehaba K. Learner Performance Prediction Indicators Based on Machine Learning. In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020). 2020, 1: 47-57.