CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

MODELING AT-RISK LEARNER DETECTION IN ONLINE EDUCATION: DEEP LEARNING FRAMEWORK WITH LSTM AND ATTENTION MECHANISMS

Smruti Nanavaty , Ajay Khunteta
[1]Department of Computer Science, Poornima University, Jaipur, India
[2]Faculty of Computer Science and Engineering, Poornima University, Jaipur, India

ABSTRACT

Identifying and supporting at-risk students is a major challenge in a digital education environment. This study examines the use of deep learning methods, specifically Long Short-Term Memory (LSTM) neurons with cognitive mechanisms to identify at-risk learners based on their involvement in periodic assessment and engagement with learning components in online learning environments. It also takes into account the importance of the dependencies of temporal factors, thus augmenting accuracy of prediction. The findings highlight the potential of advanced data analytics techniques to improve support strategies for learners on virtual learning platforms, ultimately leading to greater learner success and retention in turn. By using the test results, the study highlights the robustness of the LSTM model in predicting learner's achievement and provides insight into the factors that have the greatest predictive impact. The model performance indicates that the approach of LSTM along with attention mechanism is effective in capturing the periodic behavior of the learner on virtual platforms and the early predictions would be useful to administrators for designing timely intervention and improve retention rates of learners.

Keywords: At-risk learners, Online learning platforms, Learning outcomes, Periodic activities, Predictive analysis

1. INTRODUCTION

In today's digital education context, online learning platforms essentially provide the ability for ability for learners to acquire knowledge anytime and anywhere around the globe. The one constant problem faced by the platforms offering online certification are with regards to at-risk learners is in terms of the identification and support of these learners who may be underperforming or are distracted due to various issues [1-2]. It is therefore necessary to monitor the learner’s activity logs to monitor behaviour as well as performances during periodic assessments continuously during the course. An earlier study has proved that attendance in live sessions, time spent on the learning content and frequency of visits to archived content are some useful predictors for identification of at-risk learners [3]. Periodic activities of the learners on these online platforms can be analysed using Deep Learning techniques like Artificial Neural Networks (ANN), Recurrent Neural Networks RNN including Long Short-term Memory and GRU, has provided promising results [1, 4-7]. When compared with other models deep learning techniques offer higher accuracy in the range of 84% to 99.7 in some cases [8]. Despite these advances, none of the behavioural aspects when taken alone are correlated to learner’s

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

final performance so a comprehensive impact of these non-linear predictors needs to be considered [9]. Group Convolution and Dilated Causal Convolution (CGDC) combined with LSTM helps in capturing the impact of various learning periods of the learner's performance [10]. Some weakness observed in earlier studies where CNN is used for automatic feature recognition is that local correlations structures need to be identified while input tables are created for numerical data and if the data is periodic use of LSTM should be considered [11]. Early warning systems thus developed may be used to build intervention strategies and administrative policies to improve educational planners for reducing attrition rates of learners [12].

Our article addresses integration of timed submission entries and a quiz in to online teaching environments in order to detect academically at-risk learners. Through the employment of machine learning technologies and advanced data analytics, the instructors may be able to discover behaviour of the learners, engagement patterns and evaluation statistics to point at weak areas and help to improve them [2]. Our specialization concerns the implementation of deep learning techniques, for example, Long Short-Term Memory networks with attention mechanisms, to be able to work with students' sequential data flows in order to join connections and make accurate predictions of outcome of a learner.

In addition to the practical implementation of LSTM networks and attention mechanisms, it is crucial to understand the mathematical underpinnings of these techniques. LSTM networks are characterized by their intricate architecture, involving recurrent connections and gating mechanisms that enable them to capture long-term dependencies in sequential data. Mathematically, the operations within an LSTM cell, including the computation of cell state updates and gate activations, are governed by a series of matrix multiplications, element-wise operations, and non-linear activation functions [6, 11]. Similarly, attention mechanisms employ mathematical formulations to dynamically weigh the importance of different input features, often through dot-product attention or other attention scoring mechanisms. By delving into the mathematical foundations of these techniques, educators and researchers can gain deeper insights into their functioning and optimize their application for improved prediction accuracy and interpretability.

The study outlines the process of the data preparation stage, the model architecture implementation and experimental results obtained, to prove the effectiveness of LSTM model augmented with attention mechanisms for predictive modelling of online learners' engagement based on their activity records from periodic assignments and the quiz components. Furthermore, we consider the educational research in reference to our conclusions and indicate the possible ways of helping high-risk learners to get earliest counselling concerned their difficulties.

Instructors and educational institutions can boost their support to learners online by referring to the recommendations in this paper and therefore learners achieve positive academic outcomes and the cases of dropouts be reduced.

2. RESEARCH METHOD

LSTM networks are a type of Recurrent Neural Network (RNN) architecture designed to capture

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

long-term dependencies in sequential data [10]. Unlike traditional RNNs, which often struggle to retain information over long sequences due to vanishing gradient problems, LSTM networks overcome this limitation by introducing specialized memory cells and gating mechanisms.

2.1 LSTM Architecture

The most critical components that make up an LSTM cell include:

- Cell State – Maintains the cell's memory, ensuring information preservation in lengthy sequences.
- Hidden State – As the cell's output, it combines relevant information from the current input and the most distant hidden state.
- Forget Gate – Determines which information should and should not be held in the cell state.
- Input Gate – Controls the cell state's adjustment by monitoring the inflow of new data.
- Output Gate – Determines, which part of the cell state, should be utilized as the output.

2.2 Exploring LSTM Network Training and Attention Mechanisms

LSTM networks update their parameters using BackPropagation Through Time (BPTT). It is done by propagating the errors gradients through the entire sequence and adjusting the weights to reduce the value of the loss function. Attention is another component of our model besides the LSTM layers. The attention mechanism enables the ability to focus on different parts of the input sequence dynamically. Instead of treating the feature vectors as uniform, the model may assign more value to some features within a sequence depending on their importance to the model's current decision. This mechanism enables to extract vital data from the input and hence provide a better performance in prediction.

2.3 Hyperparameters

Hyperparameters are the settings of parameters assigned prior to model training, affecting the model's behavior and the performance. They are frequently pre-defined and do not change with data and thus need to be tuned through investigation to achieve the desired model performance. Hyperparameters of an LSTM model network include:

- Number of Units – which is the dimension of the LSTM cell state and the hidden state;
- Learning Rate – which is the step size taken during parameter updates to avoid overshooting the optimal solution;
- Batch Size – the number of samples processed in each iteration;
- Number of Epochs – the number of times the entire training data is passed through the model; and the
- Validation Split – the percentage of training data held out for validation to prevent overfitting the model.

By utilizing, the capabilities of LSTM networks with attention mechanisms, our model can effectively capture time-based dependencies and feature importance in sequential data, thus making it well suited for predicting student performance based on their engagement and performance metrics.

2.4 Model Description and Architectural flow

The proposed LSTM model is created using Keras API. The model has an input layer, which is folloed by two LSTM layers, each defined to include 50 LSTM units. In addition, an attention

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

mechanism is placed in-between the LSTM layers making the model to focus on the parts of the input data for the prediction to be more accurate and logical. The output layer is proposed to have a single dense layer that uses sigmoid activation function. This layer is used to make binary prediction for each time step and is therefore a sequential layer formed using the Time Distributed Layer. The model architecture thus comprises of two LSTM layers having 50 LSTM units each, an attention mechanism and a Time Distributed Dense layer with sigmoid activation Function and the learning rate (LR) of 0.001. Table 1 presents the hyperparameters used for building the model.

Table 1: Hyperparameters for the proposed Model with their respective values

Hyper parameter	Value
LSTM Units	50
Learning Rate (LR)	0.001
Batch Size	32
Number of Epochs	10
Validation Split	0.2

Figure 1 explains the Architectural Flow of the model illustrating the steps from data selection, loading and preprocessing, model evaluation and outcome of the model.

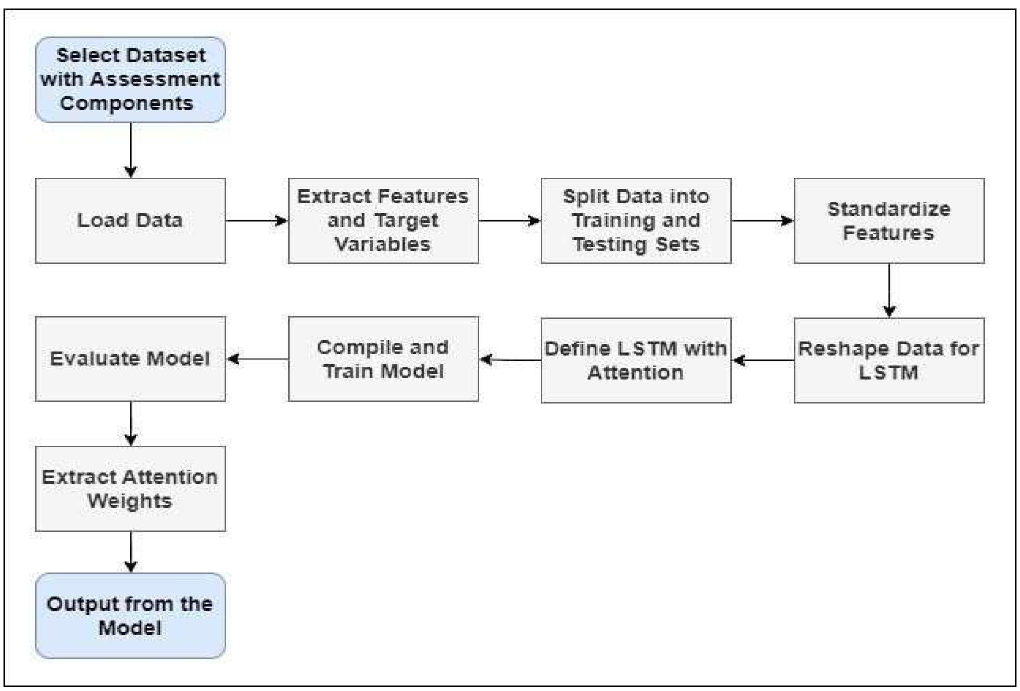

Figure 1: Architectural flow of the Model

In the table 2 is presented the list of technologies used for building the model which has been selected considering the nature of the dataset and suitability of the technology in optimising the results.

Table 2: Technology used for Model Building with its suitability

Technology	Suitability

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

Python	Offers a rich ecosystem with useful libraries for performing machine learning and Deep Learning tasks
Keras	Keras make available a comprehensiblebinterface building neural networks
TensorFlow backend	TensorFlow provides for efficient computational power while training Deep Learning Models.
Scikit-learn	Helps in effective Data Pre-processing and computing metrics for performance evaluation of the Model.
Pandas and NumPy	Effectively used for Manipulation and operations on numerical data

3. RESULTS AND DISCUSSION

The results obtained highlights the Model's performance through various metrics, supplying valuable insights for evaluating the effectiveness of LSTM in predicting the learning outcomes of online learners in our study.

3.1 Metrics and Evaluation

The evaluation metrics for the model are described as follows:

1. Loss is a metric used to find the dissimilarity between the predicted and the actual results. Lower values for this metric are indicative of better model performance.
2. Accuracy is a measure for how correctly the model is able to predict the learner's outcomes and how often the model produces correct predictions.
3. Precision reflects the level of accuracy of positive predictions demonstrating the model's capability to identify positive outcomes.
4. Recall is a metrics that focuses on the models ability to capture all the positive outcomes.
5. Area Under the Curve (AUC) signifies the model's ability to distinguish between positive and negative outcomes across various levels of thresholds.
6. Validation results are computed by working with separate dataset assisting the assessment of model's generalised performance.
7. The average attention weights are computed for finding the significant contribution of features like quiz component and assignments in model's prediction results. The performance of the models is remarkable as it showed an accuracy of 98.32%, Precision and recall were computed to be more than 98%, which denotes that the model is able to accurately predict for both the positive as well as negative outcomes. Moreover, the AUC values are 99.68% and impressively explain that the model is effectively able to discriminate between positive and negative outcomes as presented in Figure 2.

CAHIERS MAGELLANES-NS
Volume 06 Issue 1
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

```
Evaluation Metrics:
Loss: 0.05047372356057167
Accuracy: 0.9831944704055786
Precision: 0.9862446188926697
Recall: 0.9958183765411377
AUC: 0.9967505931854248
```

Figure 2: Screenshot showing model Model Performance

The average attention weights as seen in the Figure 3 provides perspectives of the features contributing to the model's predictions. It is observed that the features with higher absolute attention weights have a significant influence in determining the outcomes. The results inclusively suggest that the proposed LSTM model with attention mechanism has been effective in predicting the learner's performance based on periodic submissions and activities like quizzes conducted in phased out manner during the course duration on Virtual Learning Environments (VLEs).

```
Average Attention Weights:
Quiz1: 0.6278393268585205
Quiz2: -0.5719175338745117
Quiz3: -0.6198258399963379
Quiz4: 0.6252627372741699
Quiz5: -0.5584434866905212
Quiz6: -0.6002066731452942
Quiz7: -0.5940329432487488
Quiz8: 0.6019236445426941
Quiz9: 0.5762627124786377
Quiz10: 0.5688388347625732
Quiz11: 0.587897002696991
Quiz12: -0.43011271953582764
Quiz13: 0.6911751627922058
Quiz14: 0.5765851736068726
Quiz15: -0.6176942586898804
Assignment_1: 0.595747172832489
Assignment_2: -0.5661104321479797
```

Figure 3: Attention weights for Periodic Assessments

Our findings are consistent with previous research that utilized machine-learning algorithms for predicting at-risk learners. For instance, previous studies have demonstrated accuracies ranging from 80% to 95% using various machine-learning techniques like random forests and support vector machines. However, our LSTM model with attention mechanisms outperforms these studies by achieving an accuracy of 98.32%, a precision of over 98%, and an AUC of 99.68%. Compared to the neural network approaches used in self-paced education by other researchers, our model not only

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

captures the temporal dependencies more effectively but also leverages the attention mechanism to highlight the most significant features. This leads to more accurate and actionable predictions.

Our study also addresses the shortcomings of earlier methods that did not account for temporal patterns and dependencies adequately. By integrating LSTM networks with attention mechanisms, we can better capture and utilize the sequential nature of student interactions with online platforms, which previous models often overlooked. This is a significant innovation as it allows for more timely and personalized interventions.

3.2 Interpretation of Results

The results and findings from the proposed LSTM model with attention mechanism trained using data logs collected from VLEs upholds several key insights. The model evaluation shows that the model performs accurately and yield precise output in terms of AUC and recall to substantiate that the model is reliable for prediction of learner's performance based on periodic submission like assignments and quiz performances. Such a model could be useful to educator for providing early alerts by identifying underperforming learners. Moreover, the model may serve as a useful tool to develop personalized intervention as per the requirement of individual learners.

4. CONCLUSION

To sum up, the application of Long Short-Term Memory (LSTM) networks with attention mechanisms opens new horizons for educational predictive analytics [13]. Our study demonstrates the effectiveness of these advanced techniques in analyzing temporal patterns and key features, providing essential information about student performance. The LSTM model with attention mechanisms achieved an impressive accuracy of 98.32%, precision and recall of over 98%, and an AUC of 99.68%, outperforming previous studies that utilized traditional machine learning techniques. The novelty of this research lies in its ability to leverage temporal dependencies and periodic behaviors of learners, which previous models often overlooked. By integrating attention mechanisms, the model highlights significant features, such as quiz components and assignment submissions, allowing for more precise and actionable predictions. This innovation enables educators to identify at-risk learners earlier and design personalized interventions, enhancing learner support and retention on virtual learning platforms.

The strengths of this research include high predictive accuracy, the ability to identify key features influencing learner outcomes, and the provision of timely interventions. However, the model's performance is heavily reliant on the quality and comprehensiveness of the input data, and the implementation is computationally intensive. Future improvements could focus on enhancing scalability, interpretability, and integration with various virtual learning environments. Open questions remain about adapting the model to different educational contexts and understanding the long-term impacts on learners' outcomes. The prospects for applying LSTM networks with attention mechanisms in educational predictive analytics are vast, with future research needed to refine these models and explore the integration of other deep learning techniques. This study contributes to the theoretical understanding of using deep learning in educational settings, providing a foundation for future research and motivating other researchers to explore similar approaches. The quantitative data and insights

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

highlight the potential of these techniques to revolutionize educational predictive analytics, paving the way for more personalized and effective interventions to support student success.

REFERENCES

[1] H. Waheed, S. U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: A neural network approach," *Expert Syst Appl*, vol. 213, Mar. 2023, doi: 10.1016/j.eswa.2022.118868.

[2] D. Liu, Y. Zhang, J. Zhang, Q. Li, C. Zhang, and Y. Yin, "Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction," *IEEE Access*, vol. 8, pp. 194894–194903, 2020, doi: 10.1109/ACCESS.2020.3033200.

[3] Ş. Aydoğdu, "Predicting student final performance using artificial neural networks in online learning environments," *Educ Inf Technol (Dordr)*, vol. 25, no. 3, pp. 1913–1927, May 2020, doi: 10.1007/s10639-019-10053-x.

[4] Y. Ren, S. Huang, and Y. Zhou, "Deep learning and integrated learning for predicting student's withdrawal behavior in MOOC," in *Proceedings - 2021 2nd International Conference on Education, Knowledge and Information Management, ICEKIM 2021*, Institute of Electrical and Electronics Engineers Inc., Jan. 2021, pp. 81–84. doi: 10.1109/ICEKIM52309.2021.00026.

[5] A. Bouamrane and H. Zarzour, "A Deep Learning-based System for Detection At-Risk Students," in *4th International Conference on Pattern Analysis and Intelligent Systems, PAIS 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/PAIS56586.2022.9946878.

[6] A. S. Aljaloud *et al.*, "A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM," *IEEE Access*, vol. 10, pp. 85255–85265, 2022, doi: 10.1109/ACCESS.2022.3196784.

[7] A. Al-Sulami, M. Al-Masre, and N. Al-Malki, "Deep Learning to Predict At-Risk Students' Achievement in a Preparatory-year English Courses," in *1st International Conference in Advanced Innovation on Smart City, ICAISC 2023 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICAISC56366.2023.10085097.

[8] T. Motaung, K. Padayachee, and R. Ajoodha, "Predicting Students At-Risk Using Deep Learning Neural Network: A comparison of performance of different models," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICECCME55909.2022.9988117.

[9] T. Y. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang, "Behavior-Based Grade Prediction for MOOCs Via Time Series Neural Networks," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 5, pp. 716–728, Aug. 2017, doi: 10.1109/JSTSP.2017.2700227.

[10] Y. Zhou, K. Niu, H. Lv, G. Lu, and Y. Pan, "CGDC- LSTM: A novel hybrid neural network model for MOOC dropout prediction," in *Proceedings of the International Joint Conference on Neural Networks*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/IJCNN54540.2023.10191794.

ISSN:1624-1940
DOI 10.6084/m9.figshare.26096446
http://magellanes.com/

[11] K. Talebi, Z. Torabi, and N. Daneshpour, "Ensemble models based on CNN and LSTM for dropout prediction in MOOC," *Expert Syst Appl*, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121187.

[12] M. Skittou, M. Merrouchi, and T. Gadi, "Development of an Early Warning System to Support Educational Planning Process by Identifying At-Risk Students," *IEEE Access*, pp. 1–1, Dec. 2023, doi: 10.1109/access.2023.3348091.

[13] E. S. Karabacak and Y. Yaslan, "Comparison of Machine Learning Methods for Early Detection of Student Dropouts," in *UBMK 2023 - Proceedings: 8th International Conference on Computer Science and Engineering*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 376–381. doi: 10.1109/UBMK59864.2023.10286747.