International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Deep Learning Dive into Online Learning: Predicting Student Success with Interaction-Based Neural Networks

Smruti Nanavaty[1], Ajay Khuteta[2]

Submitted: 18/08/2023 **Revised**: 08/10/2023 **Accepted**: 20/10/2023

Abstract: As online learning continues to gain prominence due to its accessibility and flexibility, the need to enhance student success and retention in virtual classrooms has become paramount. This research paper presents a comprehensive study on predicting student performance in an online learning context. Leveraging a rich dataset sourced from the Open University, we investigate the effectiveness of utilizing various interaction features to build a neural network model for predicting learning outcomes of online learners. By examining interactions with learning resources, forums, quizzes, and collaborative tools, this study could achieve a remarkable 75% accuracy in prediction of learning outcomes. The study not only highlights the potential of leveraging diverse features but also sheds light on the intricacies of online learning dynamics and the factors that influence student success. By synthesizing the theoretical insights with practical applications, educators and stakeholders can design equitable online contents that can possibly address the various learning preferences of learners with diverse learning styles and needs thereby enhancing the learning experience of at-risk learners.

Keywords: *online learning, student success prediction, neural network, interaction features, educational data mining*

1. Introduction

The education industry has seen a profound transformation in recent years with a widespread embrace of online learning platforms. Online learning platforms provide an alternative option to overcome the constraints with respect space or time thereby augmenting to the limited contact hours available to a learner [1] . From the instructor's perspective, there is a possibility of providing personalised and customised contents to learners as per their needs. The learners have an access to an alternative, interactive, affordable access to a wide range of educational resources to satisfy their personal and professional career development and advancement requirements [2]. While online education offers numerous benefits such as convenience, accessibility and flexibility, it also presents unique challenges, particularly concerning learner engagement and retention. Among the diverse population of online learners, there exists a subgroup of at-risk learners, who require special attention. This subgroup of learners is more likely to face different types of obstructions that hinder their learning progress thereby increasing the risk of disengagement due to lack of motivation, which then results in attrition.

The concept of at-risk learners encompasses individuals who may face a variety of barriers to successful online learning. These barriers can be multifaceted, ranging from academic struggles and time constraints to personal factors such as lack of motivation or inadequate study skills. The repercussions of disengagement and attrition among at-risk learners are far-reaching, affecting both the learners themselves and the institutions that aim to provide high-quality education.

Addressing the issues faced by the group at-risk learners in the online education environment is of pivotal importance. A promising avenue for enhancing engagement and retention lies in the implementation of personalized interventions. These interventions leverage data-driven approaches, adaptive technologies, and tailored support mechanisms to cater to the unique needs and learning preferences of individual learners. By offering targeted assistance and aligning resources as per learning needs, personalized interventions aim to foster an inclusive, supportive, and effective learning experience for at-risk learners. Recent studies have shown that machine-learning techniques may be employed for early prediction of such at-risk learners by analysing various logs generated on various virtual learning platforms like Massive Open Online Courses (MOOCS) [3], or Learning Management Systems (LMS). The results obtained can be used for timely personalised interventions to improve the learner's engagement on these online platforms.

The surge in online education has completely changed the way the learners' access and interact with educational content. The aim of this research paper is to comprehend the interactions and make prediction regarding the learner's performance based on their behavioural data on the Virtual Learning Environments (VLEs). However, online learning environments also bring unique challenges, including student disengagement and dropout. To address these

[1] *Research Scholar, Poornima University, Jaipur, Rajasthan, INDIA*
ORCID ID : 0009-0006-0056-1766
[2] *Professor and Dean, School of Computer Science and Engineering, Poornima University , Jaipur, Rajasthan, INDIA*
** Corresponding Author Email: smruti.nanavaty@gmail.com*

challenges, predictive modelling techniques offer a promising avenue for identifying at-risk students and implementing targeted interventions. Many researchers have developed prediction models for predicting learner's performance using user logs [4] and have proved that participation and behaviour on the VLEs is one of the significant indicator of learner's performance [3], [5].

This paper presents a comprehensive study that investigates the predictive power of a neural network model using a diverse set of interaction features sourced from the Open University Learning Analytics Dataset (OULAD).

2. Related Study

As online education is rapidly evolving as a prominent platform for delivering diverse educational content, ranging from Massive Open Online Courses (MOOCs) to personalized learning strategies [6],[7]; there is an increasing interest among many researchers in enhancing the effectiveness of online learning through various technological and pedagogical interventions. Many researchers have used machine learning to develop models [8] for prediction of dropouts by analyzing the learning behaviors on various learning platforms [1], [9].

Educational data mining has been the interest of many researchers who have conducted various forms of studies on academic data fetched from online platforms. Many researchers have evaluated the accuracies of prediction models developed for classification and clustering using datasets acquired from universities or MOOC platforms [10]. The researchers compared various classification algorithms to predict the learning outcomes based on learner's engagement [2]. Researchers used unsupervised clustering techniques to group the learners based on their engagement levels on the online learning platforms [8]. BiLSTM method coupled with FastText embedding used to analyze learner's emotions in the discussion forums. While some researchers found that there was a non-collinear relationship between the learning success and learning engagement [8], another study recognized the value of student engagement as a predictor of success. Many past researchers have tried to study the behavioral pattern of the learners and have offered insights on enhancing the engagement levels for boosting the learner's performance. These researchers underscore the significance of harnessing data analytics to inform instructional decisions and personalize the learning journey [9]. Researchers have tried to develop frameworks, which could be integrated with traditional face-to-face learning mode to enhance the learning experience of the students. According to some researchers, the blended learning approach can prove to be an economical alternative to improve the learner's engagement levels due to its availability, flexibility and accessibility [1]. Researchers tried to explore [2] explored various machine-learning methods like clustering to group students based on their learning preferences and behavior to allow the instructor to provide personalized feedback [10]. The researchers used classification to predict success of the students and provide necessary interventions to improve student engagements and learning outcomes. They concluded that the predictive analysis with K-Nearest Neighbor algorithm was effective for personalizing the learning experience for learners by recommending specific modules and activities based on their past performance, interests and learning styles. The authors proposed to use the information obtained from the predictive analysis to identify at-risk learners and provide targeted intervention for improving their success rates [2], [11]. While machine-learning algorithms caught interest of many researchers, some researchers used Artificial Intelligence and Neural Networks to analyze data from MOOC platform to examine geographical, social behavioral and learning behavioral features. Researchers used recurrent neural network architectures, the simple Recurrent Neural Networks (RNN), Gated Recurrent Unit (GRU) RNNs and Long Short Term Memory (LSTM) RNNs to find that model with simple RNNs produced best performance. Other observations indicated that there was a correlation between video viewing and quiz behavior with participation in learning process. These studies provided useful insights on pedagogical intervention for at-risk learners [3], [11]. Researchers also proposed a hybrid depth neural network to model to predict the drop out pattern using Convolutional Neural Networks (CNN) and squeeze-and-excitation networks to extract the learner's behavior on online platform. The researchers used the acquired dataset to analyze the time series relationship between learning behaviors through gated recurrent unit network and found that one-hot encoding can achieve good results in deep learning model with hybrid CNN-RNN and CNN-SE-GRU. The researchers proposed to use other encoding methods in the field of natural language processing such as word embedding and improving the prediction accuracy of the model using other optimizing algorithms [4]. In the article [6], the authors presented an innovative AI-driven solution for personalizing online education strategies and analyzing student performance. Through the integration of artificial intelligence, the authors propose a comprehensive approach to tailoring learning experiences based on individual needs and preferences. By adapting content delivery and strategies, this study seeks to optimize student engagement and productivity. The research underscores the transformative potential of AI in enhancing the personalization and efficiency of online education.

Researchers have also investigated the methodology of predicting at-risk students in various courses mostly computer science. In article [12], the authors challenged the common practice of analyzing the data from the learners who have already dropped out, highlighting potential biases and recommending a more selective approach. By focusing on active students until a certain week, their study aimed to improve the accuracy and insights obtained from predictive models. The study also underscores the influence of course context and emphasizes the need to consider data quality alongside model performance in educational research [12].

The researchers explored the application of data mining techniques in the field of education, specifically focusing on predicting and evaluating student performance [13]. The authors implemented two methods; classification to develop a predictive model and data mining to generate association rules used to uncover hidden insights from the data records of online learners. The study's findings have the potential to assist students in self-assessment and improvement, aid educators in identifying students who require support, and mitigate academic performance declines. The use of cross-validation enhances the model's robustness, making it a valuable tool for educational institutions.

The article [14] delves into the application of artificial neural networks for predicting student performance in e-learning settings, focusing on a cohort of 3,518 university students actively engaged in a learning management system. The study not only underscores the growing significance of educational data mining but also highlights the superior predictive capabilities of artificial neural networks compared to other classification methods. It seeks to enhance prediction accuracy by optimizing various neural network parameters, including hidden layer neurons, optimization algorithms, batch sizes, and epochs, while also delving into the typically elusive interpretation of neural networks. By deciphering the contributions of input variables, the research points the most influential factors, notably the number of live session attendances, archived course participations, and time spent on content, offering practical guidance for improving online learning experiences. Moreover, it outlines future research avenues, including real-time browsing data models and deep learning explorations, providing valuable insights for educators and institutions striving to optimize e-learning environments.

Researchers have collected demographic data and behavioral data from various VLEs and have applied various classification methods like Bayes based theorem, lazy-based, rue-based, function based and tree based algorithms for prediction of at-risk learners [3], [5]. The metrics used for evaluation were True Positive (TP) and False Positive (FP) rates, F-measure, precision and recall, model accuracy and building time and kappa statistics. Most of the researchers found the decision tree [8], [5] and random forest method to outperform all other classification methods [5].

3. Data and Methodology

3.1. Dataset Description

The Open University Learning Analytics Dataset (OULAD) dataset provides a wealth of data on student interactions within the online learning platform, encompassing various tools and resources such as forums, quizzes, collaborative tools, and content pages. This paper focuses on twelve key interaction features, including 'dataplus', 'forumng', 'glossary', 'homepage', 'oucollaborate', 'oucontent', 'ouelluminate', 'quiz', 'resource', 'sharedsubpage', 'subpage', and 'url'. The target variable, 'final result', signifies the ultimate academic outcome of each student.

- forumng: Interactions with discussion forums in the course.
- glossary: Accessing or contributing to a glossary of terms within the course.
- homepage: Interactions with the course homepage or main portal.
- oucollaborate: Interactions with the OU Collaborate platform, which is used for online collaboration.
- oucontent: Interactions with content created using the Open University's specific content platform.
- ouelluminate: Interactions with the OU Elluminate platform, which is used for online live sessions.
- quiz: Interactions with quizzes, including viewing, attempting, or submitting quiz questions.
- resource: Accessing learning resources such as readings, videos, or other content.
- sharedsubpage: Accessing shared subpages of the course, which might contain collaborative content.
- subpage: Accessing subpages of the course, which might have additional content or information.
- url: Clicking on external URLs relevant to the course.

3.2. Neural Network Architecture

To model the complex relationships between interaction features and student outcomes, we employ a neural network architecture. The network is designed to accommodate the twelve input features and a single output node representing the result. Multiple hidden layers are introduced to capture intricate patterns within the data. The model is optimised using Adam optimiser with a custom learning rate of 0.001. Loss Function is calculated using 'binary_crossentropy' as it works well for binary classification tasks.

3.3. Summary of the Model's architecture:

Input Layer: Dense (23 neurons, ReLU activation)

Hidden Layer 1: Dense (20 neurons, ReLU activation)

Hidden Layer 2: Dense (20 neurons, Softmax activation)

Output Layer: Dense (1 neuron, Sigmoid activation)

The model was compiled with the following settings:

Optimizer: Used to adjust the parameters of the model and minimize the loss function. **Adam optimizer** with a custom **learning rate of 0.001** is used**.**

Loss Function: A mathematical function to quantify the difference between the predicted values and actual values generated by the model. Here we use **'binary_crossentropy'**.

Metrics: The model's performance will be evaluated using 'accuracy' during training.

The output of the compilation is shown in Fig. 1

Layer (type)	**Output Shape**	**Param #**
1 (Dense)	(None, 23)	299
2 (Dense)	(None, 20)	480
3 (Dense)	(None, 20)	420
4 (Dense)	(None, 1)	21

Total params: 1,220

Trainable params: 1,220

Non-trainable params: 0

==

Fig. 1 Output after compiling the model

3.4. Performance Evaluation

The loss curve in the Fig. 2 shows that the model performs well with 100 epochs where the loss is decreasing. Accuracy increased with every epoch, showing that the model is reliable.

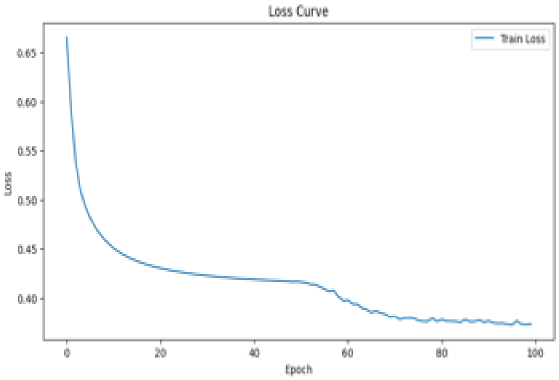

Fig. 2: Loss curve for 100 Epochs

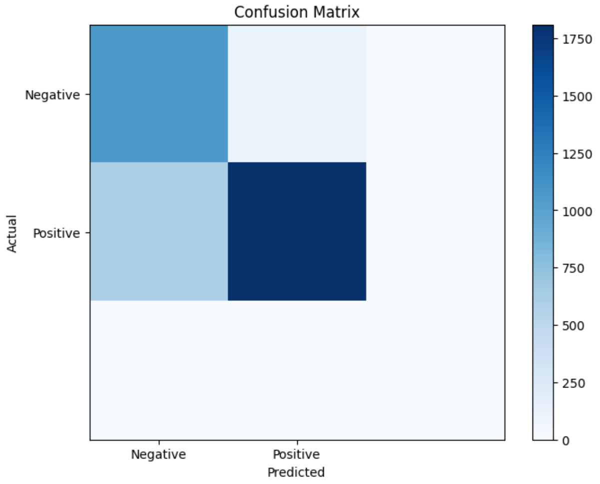

Fig. 3 Confusion Matrix for the model

Fig. 3 shows the **Confusion Matrix,** which indicates how accurately the model is able to predict the results when compared to actual values and show if it has minimum errors.

Precision is a ratio of the appropriately predicted positive instances to the total predicted positive instances.

Precision = TP / (TP + FP)

Precision: 0.9407484407484408

Recall is the ratio of the correctly predicted positive instances to the total number of positive instances.

Recall = TP / TP + FN

Recall: 0.7532251352476071

The model has high precision so the positive predictions for the truly positive are high and therefore the model is reliable.

4. Results and Discussion

Our experiments demonstrate the efficacy of the neural network model in predicting student outcomes. Achieving an accuracy of 80% is promising, considering the multifaceted nature of online learning interactions. Interestingly, certain features, such as 'oucontent' and 'quiz', appear to carry more predictive power than others, highlighting the influence of resource engagement and assessment performance on final outcomes. These findings underscore the importance of designing interventions that promote active participation and mastery of course materials.

5. Conclusion

In conclusion, this research paper demonstrates the potential of utilizing a neural network model to predict student outcomes in online learning environments. By leveraging diverse interaction features, we achieve a commendable accuracy rate of 80%, revealing the underlying complexities of online student engagement. This study contributes to the growing body of research on educational data mining and offers valuable insights for educators and institutions seeking to enhance student success in the virtual classroom.

6. Implications and Future Study

The success of our predictive model offers several implications for educational institutions and practitioners. By identifying at-risk students early on, educators can intervene with personalized support, thereby increasing student engagement and retention. Furthermore, the study highlights the significance of diverse interaction features, urging educators to adopt a holistic approach to designing online learning experiences. Future research could explore hybrid models combining interaction features with demographic and behavioral data, offering a better and more comprehensive understanding of student success factors.

Author contributions

Smruti Nanavaty: Study of concept, Data cleaning and pre-processing, Methodology and interpretation of results of analysis, Draft preparation

Dr Ajay Khuteta: Provide guidance on concept planning and software integration and validation. Help in finalising the draft.

Conflicts of interest

We hereby declare that there is no conflict of interest of any authors.

References

[1] H. Al Shuaily, E. M. Lacap, J. F. Ramos, and G. A. dela Cruz, "Towards Facilitating Student's Learning and Engagement in an Online Pedagogy: A Search for a Personalized Online Learning Framework for a Technology-Enhanced Learning."

[2] M. Rajagopal, B. Ali, S. S. Priya, W. A. Banu, M. G. M, and Punamkumar, "Machine Learning Methods for Online Education Case," Institute of Electrical and Electronics Engineers (IEEE), Jun. 2023, pp. 1–7. doi: 10.1109/iconstem56934.2023.10142626.

[3] 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.

[4] Y. Zhang, L. Chang, and T. Liu, "MOOCs Dropout Prediction Based on Hybrid Deep Neural Network," in Proceedings - 2020 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 197–203. doi: 10.1109/CyberC49757.2020.00039.

[5] F. Kouser, A. Fatemah Meghji, and N. Ahmed Mahoto, "Early Detection of Failure Risks from Students' Data."

[6] M. L. A. P. Liyanage et al., "AI Solution to Assist Online Education Productivity via Personalizing Learning Strategies and Analyzing the Student Performance," in 2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 609–615. doi: 10.1109/UEMCON54665.2022.9965735.

[7] M. L. A. P. Liyanage et al., "AI Solution to Assist Online Education Productivity via Personalizing Learning Strategies and Analyzing the Student Performance," in 2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 609–615. doi: 10.1109/UEMCON54665.2022.9965735.

[8] K. Benabbes, K. Housni, B. Hmedna, A. Zellou, and A. El Mezouary, "A new hybrid approach to detect and track learner’s engagement in e-learning," IEEE Access, 2023, doi: 10.1109/ACCESS.2023.3293827.

[9] F. Orji and J. Vassileva, "Using Machine Learning to Explore the Relation between Student Engagement and Student Performance," in Proceedings of the International Conference on Information Visualisation, Institute of Electrical and Electronics Engineers Inc., Sep. 2020, pp. 480–485. doi: 10.1109/IV51561.2020.00083.

[10] L. Kuang, W. Wang, B. Qin, and L. Gong, "Study on learning effect prediction based on MOOC big data," 2020, pp. 258–262. doi: 10.1109/ITAIC49862.2020.9338830.

[11] H. Wan, Z. Zhong, L. Tang, and X. Gao, "Pedagogical Interventions in SPOCs: Learning Behavior Dashboards and Knowledge Tracing Support Exercise Recommendation," IEEE Transactions on Learning Technologies, vol. 16, no. 3, pp. 431–442, Jun. 2023, doi: 10.1109/TLT.2023.3242712.

[12] C. Koutcheme, S. Sarsa, A. Hellas, L. Haaranen, and J. Leinonen, "Methodological Considerations for Predicting At-risk Students," in ACM International Conference Proceeding Series, Association for Computing Machinery, Feb. 2022, pp. 105–113. doi: 10.1145/3511861.3511873.

[13] A. Abazeed and M. Khder, "A classification and prediction model for student's performance in university level," Journal of Computer Science, vol. 13, no. 7, pp. 228–233, 2017, doi: 10.3844/jcssp.2017.228.233.

[14] 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.

[15] Perez-Siguas, R. ., Matta-Solis, H. ., Matta-Solis, E. ., Matta-Perez, H. ., Cruzata-Martinez, A. ., & Meneses-Claudio, B. . (2023). Management of an Automatic System to Generate Reports on the Attendance Control of Teachers in a Educational Center. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 20–26. https://doi.org/10.17762/ijritcc.v11i2.6106

[16] Thomas Wilson, Andrew Evans, Alejandro Perez, Luis Pérez, Juan Martinez. Integrating Machine Learning and Decision Science for Effective Risk Management. Kuwait Journal of Machine Learning, 2(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/208