



Journal of Advanced Zoology

ISSN: 0253-7214

Volume 44 Issue 5 Year 2023 Page 844:849

Early Prediction of 'At-Risk' Learners on Virtual Platforms using ODFs

Smruti Nanavaty^{1*}, Dr. Ajay Khunteta²

^{1*}Research Scholar, School of Basic and Applied Sciences Poornima University Jaipur, India,

Email: phdodd14.smruti@poornima.edu.in

²Professor and Dean, School of Computer Science and Engineering Poornima University Jaipur, India,

Email: dean.fce@poornima.edu.in

***Corresponding author:** Smruti Nanavaty

^{1*}Research Scholar, School of Basic and Applied Sciences Poornima University Jaipur, India,

Email: phdodd14.smruti@poornima.edu.in

<i>Article History</i>	<i>Abstract</i>
Received: 10/07/2023 Revised: 12/08/2023 Accepted: 10/09/2023 CC License CC-BY-NC-SA 4.0	— This Learning analytics are one of the most important assistance tools used by educators for early identification of at-risk learners. Researchers have used many AI based tools for monitoring learning and improving learner's performances by using any early intervention strategies to reduce dropout rates on online platforms that lacks face-to-face acknowledgement and feedback. Online platforms have Online Discussion Forums (ODFs) where a learner can post his queries and interact with other learners or the instructor. It becomes one of the useful indicators of tracking participation of a learner in the teaching learning process. Learners who actively participate in interaction on these online discussion platforms and contribute to the learning content required by other users are believed to give better performance as compared to those who do not participate in forum discussion. This paper focuses on the aspects of forum discussion like frequency of posts, sentimental analysis of forum post, number of threads initiated or replied to, and also how recent the post to predict the learners who could be at-risk of dropping out. The prediction model uses a data set from secondary resource. Various metrics like Confusion Matrix and Loss curve are employed to measure the accuracy of the model. Results indicate that data captured using forum posts can help in early identification of At-risk Learners. Keywords- <i>Forum Discussion, sentimental analysis, online learners, Natural Language processing, at-risk learners</i>

I. INTRODUCTION

Online learning platforms have transformed the education industry and more and more learners are now using online learning for acquiring degrees and enhancing their skills. MOOCs like Coursera, Udacity, edX and others are augmenting the traditional face-to-face instruction and have gained huge popularity in the past decade[1]. Although the online learning mode has various advantages, which include accessibility and self-

paced learning, it lacks the personal face-to-face feedback mechanism. Therefore, the MOOC creators often face a challenge in effectively engaging the learners on these virtual platforms. It is therefore necessary to understand the key indicators that may hinder or promote learning on the virtual platforms. Learner performance is one such measure for used by many researchers to measure the success rate of any teaching learning process. The researchers focused on various parameters and factors affecting the performance of a learner on virtual platform. Some of these include motivation levels, interactions and behaviour on the platform, participation in activities conducted on virtual platforms, type of resources and content on these platforms including the frequency of participation in various activities available on the online platforms. Many researchers also conclude that the learners who actively participate in various activities on the online platform have better performance than the learners who are inactive for a long period. Discussion forum is one such tool, which can capture various aspects such as the sentiments of the learner as well as the motivation level of the learner. The data captured on these forums can be further analysed for early detection and identification of at-risk learners. Since the MOOC platforms are evolving to become more Learnercentric from instructor centric, discussion forums could be considered for leveraging effective peer to peer learning thereby improving engagement levels of the learner on online learning platforms. Authors have proved such elements of learner centric interaction have improved the learner participation on discussion forums in MOOCs [1].

II. LITERATURE REVIEW

Past research has shown that parameters like social habits, engagement levels, social behaviours, interactions on VLEs and demographics have a correlation with student's performance in not only face-to-face learning but also blended learning and online learning. Authors have studied the persistent student's characteristics including their study habits and social interactions [2] for providing early and timely intervention. Some authors used social learning networks and predicted the student's performance for short-term courses [3]. Authors have used sentimental analysis to detect emotions of the learners and track stress and engagement levels by analysing data captured on MOOC platforms. [4]. Authors studied the correlation of gender, grades and stream with participation on discussion forums [5]. Authors investigated various factors for enhancing the design of forum discussion leading to prioritising and clustering queries for effective learning [6]. [7] used t-statistics to conclude that the learners who participated in forum discussion had better grades than those who did not actively participate in forum discussions. Authors found that ODFs (Open Discussion Forums) positively affect learner's behavior, emotional and cognitive engagement and enhances critical thinking ability. The Challenges in ODFs like Time constraints, poor linguistic proficiency of learners may affect the learner's participation. [8] Authors [9] found that there is Positive correlation between ODF participation and final course grades. Social interaction and collaborative nature in ODFs increase achievements while Contextual dimension need to be given attention. Authors used Spearman Rank Correlation Coefficient and Mann Whitney U test for significance Social Network Analysis of the posts Centrality analysis and found that social network is polycentric and less intermediary [10]. [11] proposed a temporal Convolution enhanced deep learning Framework called SocialGrid to model and forecast temporal dynamics at different granularities concluding that it excels at various levels compared to others. [12] studied quality of affirmation of the forum data and classified the post using Random forest, XGB and Naïve Bayes classifiers for drawing necessary conclusions about the quality of the post. The researchers found that Random Forest had better accuracy in comparison to the other classifiers. [13] in their study determined level of similarity between student and lecturer post in online discussion forums. They found that more the similarity better is the learning outcome of the student.

III. DATA AND METHODOLOGY

A. About the Dataset

The research focuses on understanding student interactions in online forums. We gathered data on various aspects like how often students post and comment, how quickly they respond, the emotions conveyed in their posts, their active participation in discussions, the number of new topics they introduce, and the recency of their interactions. Each of these aspects is tied to a "Result" column, which indicates the overall outcome or response variable. Think of it like observing how talkative, responsive, emotional, and engaged students are in forums and linking these observations to an ultimate result. This allows us to not only grasp the dynamics of student discussions but also train a computer model to learn from these patterns and predict or understand the outcomes in these online interactions.

B. Methodology Used

The SentimentIntensityAnalyzer is a part of the Natural Language Toolkit (NLTK) library in Python and is specifically available in the `nltk.sentiment` module. This analyzer is used to assess the sentiment of a piece of text, providing a compound score that represents the overall sentiment polarity. It is commonly used for sentiment analysis in various applications, such as social media monitoring, customer feedback analysis, and text-based sentiment classification tasks. We used this on data to classify the post as positive or negative.

C. Model Architecture

Model Description: The model uses neural network, structured as a sequential model, which means that layers are added sequentially, one after the other.

- **Input Layer:** The first layer is a dense layer with 64 units and a rectified linear unit (ReLU) activation function. This layer serves as the input layer.
- **Hidden Layer:** The second layer is another dense layer with 32 units and a ReLU activation function. This layer is positioned as a hidden layer, capturing complex patterns and representations within the data.
- **Output Layer:** The final layer is a dense layer with 1 unit and a sigmoid activation function. This layer acts as the output layer and is particularly suitable for binary classification problems. The sigmoid activation function squashes the output values between 0 and 1, representing the probability of the binary outcome.
- **Compilation of the Model:** The model is compiled using the Adam optimizer, a popular optimization algorithm for training neural networks. The binary crossentropy loss function is chosen, indicating that the model is designed for binary classification tasks. The accuracy metric is used to evaluate the performance of the model during training and testing.

In summary, the neural network comprises an input layer, a hidden layer, and an output layer as shown in Figure. 1. The choice of the ReLU activation function in the hidden layers helps introduce non-linearity and allows the model to learn complex relationships within the data. The binary crossentropy loss, Adam optimizer, and accuracy metric are selected to optimize and assess the model's performance in binary classification tasks.

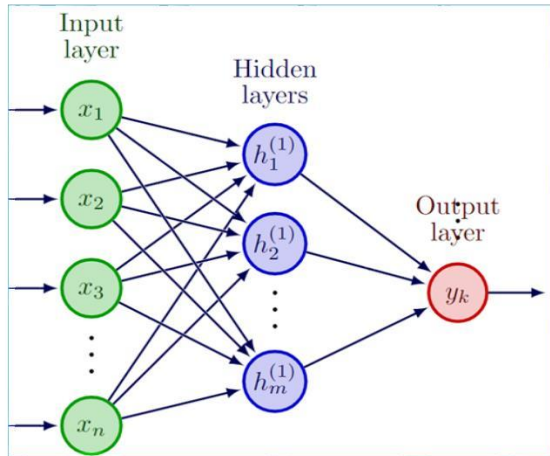


Figure. 1 Neural Network Architecture

Additionally, the SHAP (SHapley Additive exPlanations) was used to explain the output of a machine learning model by computing SHAP values for a test dataset and generating a summary plot to highlight the impact of each feature in influencing model predictions. The magnitude of SHAP values is used to determine how strong the impact is. The summary plot provides a clear visualization of the features' contributions to the model's output.

D. Experiments and Results

Suitable Machine learning algorithms were used to train the model. A decision tree is a machine-learning algorithm that makes decisions based on splitting the data into subsets according to the values of input features. In this context, a decision tree model was trained on features related to forum interactions (Frequency of Posting and Commenting, Timeliness of Submissions, Sentiment Analysis of Posts, Thread Involvement and Engagement, Number of Threads Initiated, Recency of Forum Interactions) to predict a target feature labelled as 'Final Result.'

The decision tree model, trained on forum interaction features, achieved a moderate accuracy of 66.2%, indicating its ability to make correct predictions, while further analysis of feature importance can provide insights into the factors influencing the model's decisions.

The Random Forest model, trained on forum interaction features, achieved a higher accuracy of 78.6% compared to the Decision Tree.

Results obtained are as follows:

Decision Tree accuracy: 0.662

Random Forest accuracy: 0.786

Feature Importance for the features of the online forum discussion being studied using Random forest are as follows:

[0.14046857 0.26016092 0.11202461 0.10207462
0.12571107 0.25956021]

The provided feature importance scores from the Random Forest model offer insights into the contribution of each feature to the model's predictive accuracy. Here's an interpretation based on the given feature importance values: Feature 0: Frequency of Posting and Commenting (0.1405): This feature contributes moderately to the model's predictions. A higher value indicates that a higher frequency of posting and commenting has a positive impact on the model's accuracy.

Feature 1: Timeliness of Submissions (0.2602):

The highest feature importance is assigned to the timeliness of submissions. This suggests that the timeliness of forum submissions has a significant impact on the model's accuracy, and variations in this feature strongly influence predictions.

Feature 2: Sentiment Analysis of Posts (0.1120):

Sentiment analysis contributes to the model but has a lower importance compared to the timeliness of submissions. The sentiment expressed in forum posts influences the model's predictions, but to a lesser extent.

Feature 3: Thread Involvement and Engagement (0.1021):

Thread involvement and engagement have a moderate impact on the model's predictions. While important, this feature contributes less than the timeliness of submissions and sentiment analysis.

Feature 4: Number of Threads Initiated (0.1257):

The number of threads initiated also plays a substantial role in the model's predictions, with a contribution that falls between moderate and high.

Feature 5: Recency of Forum Interactions (0.2596): The 'recency' of the post means how latest is the post in the timeline of the posts being analyses. Recency of a post in forum interactions is the second most important feature according to these scores. Recent interactions have a significant influence on the model's accuracy.

Accuracy of Neural Network model is observed to be 0.7940000295639038 and the loss = 0.5107 as shown in the loss curve in the Figure. 2.

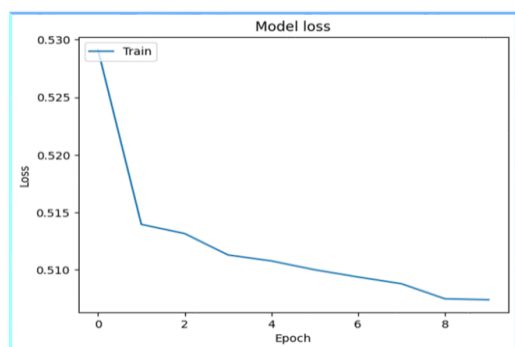


Figure. 2 Loss Curve

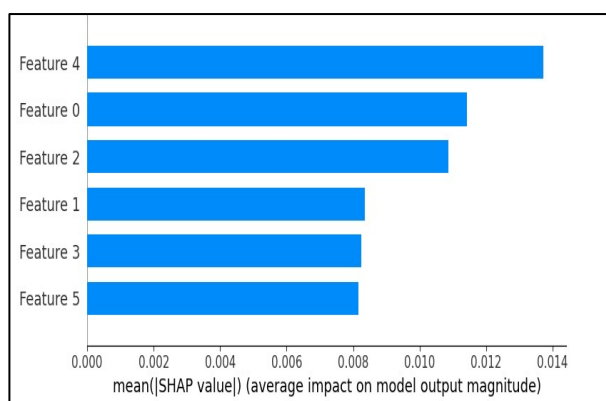


Figure. 3 Average impact on model output magnitude

The above Figure 3 shows the average impact of the features of our study on model output magnitude.

Where,

Feature 0: Frequency of Posting and Commenting

Feature 1: Timeliness of Submissions

Feature 2: Sentiment Analysis of Posts

Feature 3: Thread Involvement and Engagement

Feature 4: Number of Threads Initiated

Feature 5: Recency of Forum Interactions

E. Performance Evaluation

The analysis of the forum interaction is carried out on the data. The post on the forum were first classified using machine learning algorithms as positive or negative to capture the sentiments of the posts. The other parameters used to train the model were Frequency of Posting and Commenting, Timeliness of Submissions, Sentiment Analysis of Posts, Thread Involvement and Engagement, Number of Threads Initiated, Recency of Forum Interactions. Accuracy of different algorithms were compared and Random Forest was found to perform better than the other algorithms. Further feature importance was studied to find that number of thread initiation during forum interaction had a most significant impact on the final scores of a learner. A neural network was trained with the parameters to understand the how various parameters contributed towards the final prediction. The neural network model performed well with 79.40% accuracy proving that forum interaction have a significant contribution in prediction of the final results of a learner on virtual platforms.

IV. 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.

V. CONCLUSION

There are various features contributing to prediction of final performance of a learner on virtual platforms. Learner demographics and behaviour on the platform are among the predictors for learner performance. Sentiments of the learners and level of engagement are also factors contributing to the final scores of a learner. Forum post can enable a researcher to study these factors. The study shows that there are various features of the forum interactions which can be studied to analyse the extent to which they contribute to accurately predict the target variable. Timeliness of submission of post plays the most important role in determining the learner's final performance and the latest time when the post was submitted is the next important feature in determining the performance of the learners on online platforms.

VI. IMPLICATION 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 offering a better and more comprehensive understanding of student success factors.

IV. REFERENCES

1. N. Ayer, H. Sukhathankar, U. Deshmukh, and S. Sahasrabudhe, "Impact of learner-centric discussion forums on learner engagement in skill development MOOC," in Proceedings - IEEE 9th International Conference on Technology for Education, T4E 2018, Institute of Electrical and Electronics Engineers Inc., Dec. 2018, pp. 69–72. doi: 10.1109/T4E.2018.00021.
2. N. Gitinabard, Y. Xu, S. Heckman, T. Barnes, and C. F. Lynch, "How Widely Can Prediction Models Be Generalized? Performance Prediction in Blended Courses," IEEE Transactions on Learning Technologies, vol. 12, no. 2, pp. 184–197, Apr. 2019, doi: 10.1109/TLT.2019.2911832.
3. IEEE Communications Society and Institute of Electrical and Electronics Engineers, IEEE INFOCOM 2017 - IEEE Conference on Computer Communications : 1-4 May 2017.
4. X. Tao et al., "Data Analytics on Online Student Engagement Data for Academic Performance Modeling," IEEE Access, vol. 10, pp. 103176–103186, 2022, doi: 10.1109/ACCESS.2022.3208953.
5. M. T. Shaikh, U. Koppikar, and M. Vijayalakshmi, "Identifying Factors that Influence Student's Participation in an Online Discussion Forum," in 2nd Edition of IEEE Delhi Section Owned Conference, DELCON 2023 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/DELCON57910.2023.10127471.
6. Neha and E. Kim, "Investigating Responsible Factors for Interaction between Learners and Instructors in the Discussion Forum of MOOC," in 2021 9th International Conference on Information and Education Technology, ICIET 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 204–207. doi: 10.1109/ICIET51873.2021.9419599.
7. B. H. Olivier, "The Impact of Contact Sessions and Discussion Forums on the Academic Performance of Open Distance Learning Students," 2016.
8. A. A. Gasmi, "Through the Lens of Students: How Online Discussion Forums Affect Students' Learning," International Journal of Technology in Education, vol. 5, no. 4, pp. 669–684, Oct. 2022, doi: 10.46328/ijte.291.
9. M. Alzahrani, "The effect of using online discussion forums on students' learning." [Online]. Available: <https://www.researchgate.net/publication/312136267>
10. H. Wan, L. Tang, K. Liu, and X. Gao, "Exploring the Relationship between SPOC Forum Behaviors and Learning Outcomes Based on Social Network Analysis."
11. C. Ling, D. Cui, G. Tong, and J. Zhu, "ON FORECASTING DYNAMICS IN ONLINE DISCUSSION FORUMS," in Proceedings - IEEE International Conference on Multimedia and Expo, IEEE Computer Society, 2021. doi: 10.1109/ICME51207.2021.9428382.
12. A. Lalingkar, P. Mishra, S. Mandyam, J. Pattanaik, and S. Srinivasa, "Building a model for finding quality of affirmation in a discussion forum," in Proceedings - IEEE 20th International Conference on Advanced Learning Technologies, ICALT 2020, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 122–126. doi: 10.1109/ICALT49669.2020.00043.
13. D. Fitria Murad, Y. Heryadi, S. Muhamad Isa, W. Budiharto, and B. Dwi Wijanarko, "Text Mining Analysis in the Log Discussion Forum for Online Learning Recommendation Systems."