Bachelor´s Thesis Proposal

Modeling and Interpretation of Student Performance in Virtual Learning Platforms with Neural Networks

Paz Barquín Madariaga
202009738@alu.comillas.edu

Tutor: Jenny Alexandra Cifuentes Quintero

Universidad Pontificia Comillas
E-6 + Analytics

1 Main and Specific Objectives

1.1 Main Objective

To develop a predictive model based on neural networks that estimates student performance in virtual learning environments, and to apply advanced interpretability techniques to identify the most influential variables in student performance.

1.2 Specfic Objectives

- To review recent scientific literature on interpretability techniques applied to neural networks, focusing on advancements made in recent years that enable understanding and explaining the results of complex models in the context of predicting educational performance.
- To develop a predictive model of student performance using a standard dataset in learning analytics, through the evaluation of open-access repositories and the application of neural networks for academic performance modeling.
- To apply a global interpretability technique from the state of the art to the results obtained from the neural network model, in order to identify and analyze the most influential variables in predicting student performance in virtual learning environments.

2 Motivation

In recent years, virtual learning environments have become increasingly significant, particularly following the transition to hybrid education driven by the COVID-19 pandemic. This shift has intensified researchers' interest in understanding and accurately predicting academic performance in these contexts, with the aim of identifying the key factors influencing student outcomes. To meet this demand, a broad range of predictive models has been proposed, encompassing traditional statistical methods as well as more advanced machine learning approaches, including neural networks [Agudo-Peregrina et al., 2012].

Despite the strong predictive performance of neural networks, their primary challenge is their perception as 'black boxes,' which hinders the interpretation of their results and restricts their applicability in educational contexts. In these settings, understanding the underlying factors driving predictions is essential. Consequently, enhancing the interpretability of these models has emerged as a key priority in the field. To address the challenge of neural network interpretability, several studies have proposed the application of interpretability techniques to facilitate understanding of these models [Christoph, 2020]. These techniques decompose complex predictions into more comprehensible components, enabling the identification of the most influential variables and offering clearer insights into the model's decision-making process [Lundberg, 2017]. For example, [Chen et al., 2022] applied LIME (Local Interpretable Model-agnostic Explanations) to a convolutional neural network (CNN) integrated with a long-short term memory model (LSTM) to provide interpretability in predicting academic performance. LIME identified the most relevant factors, such as weekly student engagement and interaction patterns, which significantly influenced the predictions. Other studies have employed techniques like SHAP (SHapley Additive exPlanations); for instance,

[Sahlaoui et al., 2021] which study demonstrated that SHAP highlighted key factors impacting student performance, including interaction data, submission timings, and participation rates.

This Bachelor's Thesis aligns with this research area by aiming to apply novel interpretability techniques to neural network models in the analysis of academic performance. This approach will enable educators and administrators to not only anticipate potential issues in student performance but also gain insight into the underlying factors contributing to these challenges, thereby facilitating more informed and effective decision-making.

3 Proposed Methodology

The methodology of this Bachelor's Thesis is structured into several phases to achieve the proposed objectives. First, the relevance of interpretability techniques applied to the analysis of academic performance will be presented, highlighting the role of neural networks and their potential to improve education in virtual environments. The need to interpret complex models for better decision-making in the educational field will be justified.

Next, a literature review focused on the use of neural networks in education will be conducted, identifying the most relevant approaches for predicting student performance and recent advances in interpretability techniques applied to these models. In the next phase, data will be collected and processed from open-access repositories, selecting the most suitable dataset. The necessary cleaning and transformation processes will be carried out to ensure the quality of the dataset by eliminating inconsistencies and noise.

With the processed data, a neural network model will be implemented to predict student performance. After training, advanced interpretability techniques will be applied to identify the most influential variables in the predictions. Finally, an analysis of the results obtained from the predictive model and the interpretations derived from the applied techniques will be conducted. The model's performance will be evaluated using specific metrics, and the utility of interpretability will be analyzed to understand the factors that most influence student performance, providing valuable information for improving educational strategies.

4 Proposed Table of Contents

1. Introduction: (a) Motivation (b) Objectives (c) Index
2. State of Art
3. Methodology: (a) Data Selection and Pre-processing (b) Descriptive Analysis (c) Predictive Model Development (d) Application of Interpretability Techniques
4. Results
5. Conclusions
6. Bibliography

References

[Agudo-Peregrina et al., 2012] Agudo-Peregrina, Á. F., Hernández-García, Á., and Iglesias-Pradas, S. (2012). Predicting academic performance with learning analytics in virtual learning environments: A comparative study of three interaction classifications. In *2012 International Symposium on Computers in Education (SIIE)*, pages 1–6. IEEE.

[Chen et al., 2022] Chen, H., Prasetyo, E., Tseng, S., and Putra, K. (2022). Prayitno; kusumawardani, ss; weng, ce week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence. *Appl. Sci*, 12:1885.

[Christoph, 2020] Christoph, M. (2020). *Interpretable machine learning: A guide for making black box models explainable*. Leanpub.

[Lundberg, 2017] Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

[Sahlaoui et al., 2021] Sahlaoui, H., Nayyar, A., Agoujil, S., Jaber, M. M., et al. (2021). Predicting and interpreting student performance using ensemble models and shapley additive explanations. *IEEE Access*, 9:152688–152703.