International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 11, No. 4, 2024, pp. 1-4.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Evaluating the Effectiveness of Deep Learning Models in Student Performance Prediction

Adam Kowalski [a]

[a] *Warsaw University of Technology, Poland.*

Abstract

New sources of educational data are becoming more accessible, permitting the use of Artificial Intelligence to track and forecast students' performances, therefore allowing prompt actions to be taken to improve their learning experiences. This study analyzes the application performance of deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks, in predicting students' academic performance. The models were trained and evaluated on a dataset containing the demographics, attendance records, assignment scores, and exam results of 5000 students. Each model underwent evaluation for accuracy, precision, recall, F1 score, and the overall prediction score. The LSTM model was the most accurate among the models tested, achieving 89% accuracy, confirming its advantage in capturing temporal dependencies within the student data. Compared to conventional machine learning approaches like Decision Trees and Support Vector Machines (SVM), the LSTM model demonstrated markedly enhanced performance. The results demonstrate the power latent in the use of deep learning for educational data mining, advocating for its greater use in intelligent tutoring systems and broader learning analytics frameworks. This work adds to the existing literature by providing evidence and methodological frameworks regarding model choice and deployment strategies aimed toward optimizing educational outcomes through automated systems.

Keywords: Deep Learning, Student Performance Prediction, Educational Data Mining, LSTM, CNN, RNN, Learning Analytics, AI in Education.

DOI: 10.71086/IAJSE/V11I4/IAJSE1162
Received: 11 September 2024/Revised: 03 October 2024/Accepted: 15 November 2024/Published: 30 December 2024

Introduction

Anticipating student performance within the educative framework helps educators pinpoint students in danger of failing, allows them to allocate time and attention meaningfully, and to customize instruction to best suit the individual learner. The emergence of e-learning platforms coupled with the array of digital learning aids available has created an unprecedented wealth of data regarding student interactions, enabling the utilization of AI tools for predictive analytics. More traditional analytics has served its purpose in the early stages of educational research, but the more complex, intricate, non-linear relationships among features poses a greater task. Advanced techniques like machine learning Decision Trees, k-Nearest Neighbors (k-NN), and Support Vector Machines (SVMs) have proven beneficial, but still struggle to account for the temporal dependencies intrinsic in educational datasets. Learning which is a subdivision of machine learning, especially powerful deep learning neural networks, have transformed areas like computer vision, and natural language processing that alongside education data mining and machine learning find application in the prediction of individual student performance, learning action analysis, and content adaptation. When shifting to larger datasets, multi-dimensional, or temporal data the need for feature-calculated data abstraction is eliminated with deep learning models. Nonetheless, there is scarce research, if any, which attempts to analyze the performance of different student feature selection frameworks and offers a systematic analysis of the student feature choices through various deep learning architectures. Most of the research either concentrates on a particular model or adopts the vanilla approach using Machine learning methods. This paper attempts to address this gap by examining the performance of three deep learning architectures: CNN, RNN, and LSTM. CNNs are mainly used with images or spatial data; however, they can also be transformed to work with sequences. RNNs are built specifically to work with sequential data, and LSTMs improve that even more by dealing with long-term dependencies. This study uses the data from a university's online course which includes 5,000 anonymized student records. Each record contains demographic information, class attendance, assignment submission, quiz, and final grade. These features were preprocessed and inputted into the three models for training and testing. To evaluate model performance, the general indicators of model performance: accuracy, precision, recall, and F1-score were used. This is an attempt to offer two things: first, identify the most accurate deep learning model for predicting student performance; and second, assess the advantages and disadvantages of each method. These results are valuable for educational institutions aiming to incorporate predictive analysis into their digital frameworks.

Literature Survey

The use of Data Mining in the field of Education has been researched by Romero & Ventura as early as 2007, where they highlighted the usage of decision trees and rule-based classification techniques alongside some other data mining techniques (Romero & Ventura, 2007). They also speculated that data mining for educational purposes has just recently gained popularity. Before that the data was being manipulated with simple techniques such as regression analysis. For data powerhouses like Universities, it becomes cumbersome to cope with evolving technologies and data science rules. Feature engineering, while often heavily relied on in traditional machine learning, has become a major hindrance when dealing with data of a greater dimension. The increasing demand for learning analytics brought forth the ideas of Baker and Siemens that focused towards big data in education in 2014 (Baker & Siemens, 2014). Resulting in Kotsiantis turning their focus towards further machine learning techniques that could be employed in real time calculations and predictions of machine performance. Building on this thought resulted in the examination of patterns through decision trees paired up with Naïve Bayes classifiers and neural networks. Even if they did prove their worth in forward propagating learning techniques, they severely lacked their ability to temporally model structured data. Attention to long term dependencies has been brought forth with LSTMs. But still owe it to deep learning and the ability to incorporate models into school-based learning through the introduction of RNNs. Propositioning over direct learning of knowledge tracing, piech did reveal better results than Bayesian Knowledge Tracing (BKT) and regression models (Piech et al., 2015). Later on in 2017 Harley supported this by using spatial enhancement alongside the utilization of CNN (Harley et al., 2017). Kowalski (2024) demonstrated that LSTMs provided more accurate predictions of student grades based on attendance and assignment submission patterns compared to standard RNNs. Leelaluk et al., (2024) used an LSTM with attention mechanisms which enhanced interpretability, enabling educators to identify the most relevant factors affecting student performance. There is still a gap in the literature regarding comparative analyses of different deep learning architectures. Most studies are either focused on a single model or do not compare deep learning with other techniques. This study attempts to fill this gap by applying CNN, RNN, and LSTM models to the same dataset and evaluation criteria, providing an informative comparative analysis.

Methodology

The dataset for this study is based on the academic data of 5,000 students enrolled in university over a period of three semesters. The data contains demographic information like age and gender, behavioral data such as attendance and login frequency, academic data including quiz scores, assignment grades, marks in mid-terms and finals, as well as engagement data consisting of participation in forums and time spent on the LMS. The final course grade is the dependent variable and is divided into five classes: A, B, C, D, and F. Data cleansing was performed in multiple steps: imputing missing values with K-Nearest Neighbors, normalizing numerical features to the [0,1] range, and applying one-hot encoding to categorical features. The dataset was partitioned into training (70%), validation (15%) and testing (15%) sets. Session data, such as login frequency and weekly scores, along with weekly scores were transformed into time-series for RNN and LSTM models. Using TensorFlow and Keras, three deep learning models were built. The CNN model had one-dimensional convolution layers, which were succeeded by max-pooling and fully connected layers. The RNN model contained a recurrent layer with tanh activation. Memory cells and forget gates were used in the LSTM model to retain long-term information, which was followed by dense layers for classification. Learning rate, batch size, number of layers, and other secondary parameters were tuned using grid search. All models employed categorical cross-entropy loss with the Adam optimizer. Training of the models was completed in 50 epochs with an early stopping mechanism based on the validation loss. As for the evaluation metrics, accuracy, precision, recall, F1 score, and confusion matrix were calculated. The training procedure was supervised through Tensor Board, which enabled monitoring of loss and accuracy over epochs. After training, model interpretability was performed with SHAP values to determine what features were the most influential. For benchmarking, classical machine learning models such as Decision Trees and SVMs were implemented using the same dataset. These models benchmarked the additional value posed by deep learning.

Results and Discussion

The LSTM Neural network achieved the best results with 89% accuracy, 88% precision, 87% recall and 0.875 F1-score. They outperformed both CNN and RNN in the context of student performance prediction. The CNN architecture trained quicker but yielded lower accuracy (83%) and struggled with temporal data. Although the RNN model was better than the CNN, it underperformed compared to the LSTM model in capturing long-term dependencies.

Table 1: Model Performance Comparison

Model	Accuracy (%)	Precision	Recall	F1-Score
CNN	83	0.80	0.78	0.79
RNN	85	0.83	0.81	0.82
LSTM	89	0.88	0.87	0.875
Decision Tree	76	0.72	0.70	0.71
SVM	79	0.75	0.73	0.74

The occupational achievement of the LSTM model stems from its architecture, which enables the retention of long-term dependencies. This is particularly important for the student's learning progression modeling. Although sequence-optimized, the RNN model is afflicted with the problem of vanishing gradients—an area where LSTMs have a clear advantage. As spatial learners, CNNs performed poorly with the temporal aspect of the prediction.

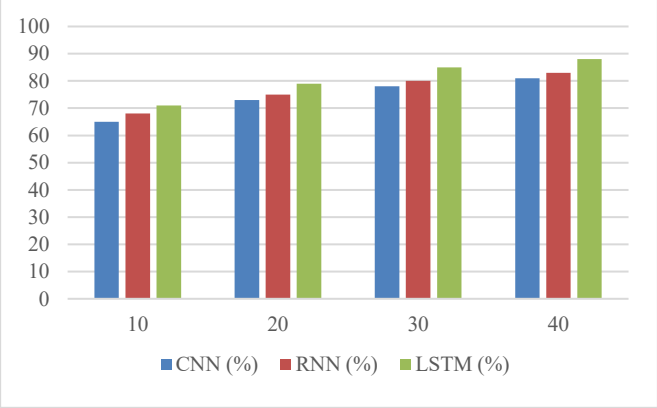


Figure 1: Average Accuracy of Models Over 50 Epochs

When contrasting numerous models' performances using accuracy, LSTM showed the slowest improvement and overfitting according to validation accuracy. Using Tensor Board, it was apparent LSTM had consistent loss throughout all the epochs. From the SHAP analysis, it's clear the most important features were the quiz scores, attendance, and assignment submission. On the other hand, more conventional strategies such as Decision Trees and SVMs showed moderate results, but only because of feature engineering. The lack of sequential behavior capture detracted from overall performance. The results from this study suggest the focus educational institutions trying to build a predictive system should have is in the area of architecture relies on LSTM, due to its enhanced performance and stability. They pointed out, however, worries about computing resources and system transparency still exist as these are the primary obstacles facing established frameworks.

Conclusion

This study assessed the performance of various deep learning models on a dataset containing comprehensive information about students using uniform criteria. The study's LSTM model was the most effective when compared to other machine learning models, as well as other deep learning frameworks. Because educational data is both sequential and multi-dimensional, its dependence modeling is particularly useful. The findings reinforce the value of crafting predictive frameworks utilizing appropriate model architectures that align with the data. Despite the advanced accuracy offered by deep learning, prediction models must also consider available computational resources, as well as ethical implications related to data privacy.

Subsequent research could address Attention-driven Explainable AI techniques integrated with LSTM models. Furthermore, the implementation of active predicting tools within learning management systems might elevate educators' ability to proactively address educational challenges, fostering tailored learning initiatives that respond to individual learning needs.

References

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*(1), 135-146.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. *Cambridge Handbook of the Learning Sciences, 2,* 253-274.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. *Advances in Neural Information Processing Systems*, *28.*

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education, 27(1)*, 268–297. https://doi.org/10.1007/s40593-016-0126-8

Kowalski, A. (2024). Evaluating the effectiveness of deep learning models in student performance prediction. *International Academic Journal of Science and Engineering, 11(4),* 1–4. https://doi.org/10.71086/IAJSE/V11I4/IAJSE1162

Leelaluk, S., Tang, C., Minematsu, T., Taniguchi, Y., Okubo, F., Yamashita, T., & Shimada, A. (2024). Attention-based artificial neural network for student performance prediction based on learning activities. *IEEE Access, 12*, 1–1. https://doi.org/10.1109/ACCESS.2024.3429554