The Efficiency of Artificial Recurrent Neural Network (RNN) in Predicting Academic Performance for Students

Abdulellah Abdullah Alsulaimani

Department of Educational Technologies
Faculty of Education, King Abdulaziz University, Saudi Arabia.
Email: aalsulaimani@kau.edu.sa

Abstract - This study examines the assessment of two different models, Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP), specifically in terms of their effectiveness in predicting the academic advancement of students at King Abdulaziz University in Saudi Arabia. The study focuses mostly on courses that include infographics and animated infographics. It employs three main statistics metrics: Symmetric Mean Absolute Percentage Error (SMAPE), Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). The findings of our analysis demonstrate the higher performance of the LSTM model compared to the MLP model in all three-evaluation metrics. More specifically, the LSTM model regularly performs better than the MLP model, with lower values for SMAPE, MSE, and APE. The decreased error metrics in the LSTM column indicate improved overall prediction accuracy in comparison to the MLP model. The study's extensive findings and powerful prediction capabilities signify a substantial advancement in comprehending and utilizing these technologies for educational objectives. This research has ramifications that go beyond academia and can provide real benefits to educators, policymakers, and organizations. This study enhances the continuous endeavors to enhance educational outcomes and student performance by introducing more effective methods for detecting and assisting children who are in danger of expulsion.

Keywords - statistics metrics, prediction, analysis, Long Short-Term Memory, Multilayer Perceptron.

I. INTRODUCTION

In the current educational context, the use of data to make informed decisions requires the implementation of advanced predictive models that can effectively analyze and anticipate student performance [1]. This study focuses on analyzing a dataset that specifically looks at student grades, to develop a strong predictive framework. The study used two separate models, Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP), to evaluate their efficacy in capturing and predicting complex patterns in the dataset. The goal is to improve our comprehension of the elements that impact student academic achievements and to identify a model that provides better effectiveness and precision in this context.

The paper is structured as follows: the introduction is presented in the first section, followed by the second and third sections which cover fundamental ideas related to the models utilized in the study. The fourth section focuses on accuracy measurements, while the fifth section provides a comprehensive discussion of the results. The final portion concludes the paper.

II. ROLE OF MACHINE LEARNING IN THE PREDICTION OF STUDENT ACADEMIC ACHIEVEMENT

The automated forecasting of student performance is essential, considering the substantial volume of data available in educational systems. Educational data mining (EDM) utilizes approaches to extract meaningful insights from data generated in educational settings. Artificial intelligence (AI) can analyze educational data to predict student performance and provide interventions to prevent academic failure and improve learning results. Educational platforms augment conventional learning settings by assessing student performance, hence reducing the likelihood of student failure [2]. Undoubtedly, the utilization of machine learning in learning analytics is crucial for comprehending and enhancing the educational process. Below is an in-depth examination of the role of machine learning in learning analytics:

Data Collection: Machine learning algorithms collect data on students' interactions with various forms of information, including text, videos, interactive simulations, and evaluations [3].

The data may encompass the duration allocated to each category of material, the degree of involvement, and the results achieved in assessments [4].

Algorithm Training: Data can be used to train supervised learning algorithms, specifically classification models. The program acquires the ability to identify patterns and correlations between students' interactions and their favored learning techniques [5].

Analysis of behavior: Machine learning algorithms can assess how students interact with different educational materials, such as textbooks, online resources, and interactive content. Patterns, such as the duration allocated to particular subjects, the frequency of reviewing material, and the sequence of interactions, can be discerned [6].

DOI 10.5013/IJSSST.a.25.01.02 ISSN: 1473-804x online, 1473-8031 print

Adaptive Learning Platforms: Machine learning algorithms in adaptive learning systems continuously assess a student's performance and adapt the content based on their strengths and weaknesses. These platforms provide personalized learning experiences by adjusting difficulty levels, content presentation, and assessment methods [7].

Identifying Learning Styles: Machine learning can identify individual learning styles by analyzing how students respond to different types of content and assessments. Understanding learning styles allows educators to tailor their teaching methods to better suit the preferences and needs of each student [8].

Automated Feedback Systems: Automating the processing of written assignments and providing feedback to students can be achieved through the application of Natural Language Processing (NLP) techniques. This can facilitate the provision of prompt and tailored feedback, so fostering ongoing enhancement [9].

Increased Student Engagement: Increased student engagement can be achieved by customizing information to match individual learning styles. Increased student engagement correlates with improved comprehension and enhanced information retention [10].

Resource Allocation: Institutions can employ machine learning to enhance resource allocation by identifying areas that require greater support, resources, or staff [11].

III. LONG SHORT-TERM MEMORY

Long Short-Term Memory (LSTM) is a specific architecture of recurrent neural network (RNN) that is specifically developed to overcome the issue of vanishing gradients commonly encountered by standard RNNs [12]. LSTMs are highly suitable for jobs that use sequential data, such as natural language processing, audio recognition, and time series prediction [13]. Below is a fundamental explanation and overview of Long Short-Term Memory networks:

Background: Vanishing Gradient Problem: Conventional Recurrent Neural Networks (RNNs) encounter difficulties in effectively capturing long-term relationships in sequential data, mostly because of the vanishing gradient problem [14]. During the process of error backpropagation through time, the gradients frequently diminish significantly, resulting in challenges when it comes to updating the weights of preceding layers [15].

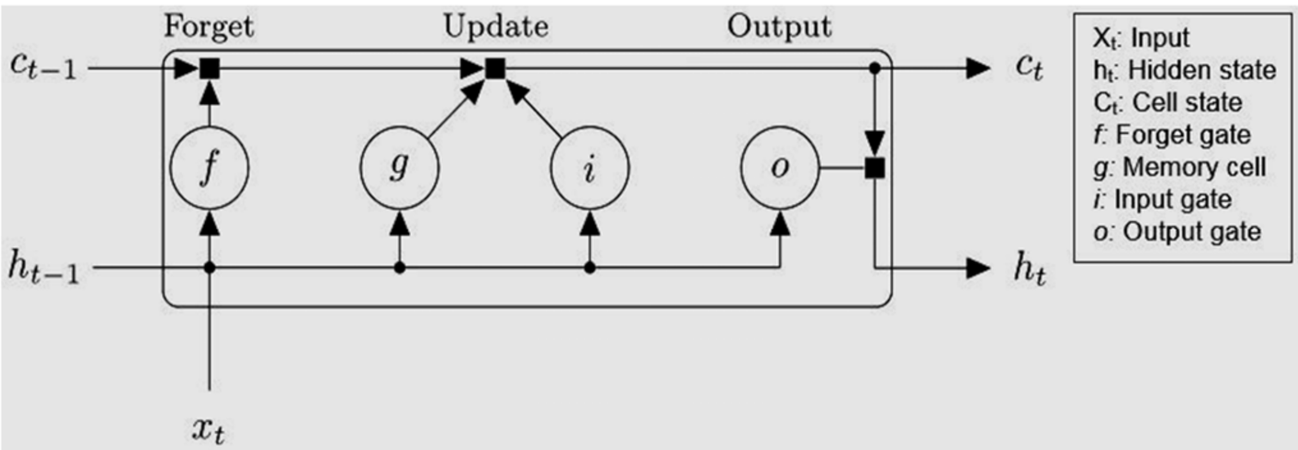
Architecture: Memory Cells: LSTMs incorporate the notion of memory cells, which have the ability to retain information for extended durations. The memory cells possess the ability to selectively incorporate or exclude information using gating processes, enabling the network to retain or discard information as required [16].

Gating processes: Long Short-Term Memory (LSTM) models include three gating processes to regulate the transmission of information [17]:

Forget Gate: Determines which information from the prior cell state should be excluded.

Input Gate: Determines the selection of fresh information to be stored in the cell state.

Output Gate: Determines the selection of information from the cell state to generate the output.



IV. MULTILAYER PERCEPTRON

A Multilayer Perceptron (MLP) is an artificial neural network (ANN) specifically developed for supervised learning tasks [18]. It belongs to the feedforward neural network category, where information moves through the network in a unidirectional manner, from the input layer to the output layer. MLPs are extensively utilized for a range of activities, encompassing classification and regression challenges [19].

Neurons: also known as nodes, are the fundamental components of a Multi-Layer Perceptron (MLP). Every node receives one or several inputs, performs operations on them, and generates an output [20].

Edges: The links between nodes are assigned weights. The weights govern the magnitude of the connection, affecting the influence of the input on the output of the node [21].

Layers: Nodes are arranged systematically into distinct layers. A Multilayer Perceptron (MLP) generally comprises three sorts of layers: an input layer, one or more hidden layers, and an output layer.

Architectural Design and Structure:

The input layer: is responsible for receiving the characteristics of the incoming data. Every node in the input layer represents a specific characteristic or attribute in the dataset [22].

Hidden levels: One or more levels can exist between the input and output layers, known as hidden layers. Every node within a concealed layer obtains input from every node within the preceding layer and transmits its output to every node within the subsequent layer.

The output layer is responsible for generating the ultimate outcomes of the network's calculations. The quantity of nodes in the output layer is contingent upon the type of task being performed, such as binary classification, multi-class classification, or regression.

Activation Function:
Every node inside the hidden layers and output layer applies an activation function to the weighted sum of its inputs. Popular activation functions comprise the sigmoid, hyperbolic tangent (tanh), and rectified linear unit (ReLU).

Backpropagation as a Training Method

MLPs are trained by a method called supervised learning. The backpropagation algorithm is frequently employed for training purposes. The process entails propagating input data forward through the network, computing the error, and subsequently modifying the weights in the opposite direction to minimize said error[23].

Loss Function: The loss function quantifies the disparity between the expected output and the true target values. The objective during training is to limit the amount of loss.

Optimization algorithms: Optimization algorithms, such as gradient descent, iteratively modify the weights in the network. The purpose of these changes is to identify the most favorable combination of weights that will reduce the loss function [24].

V. PRECISION

Three accuracy measurements are used to determine the performance of the proposed method, the symmetric Mean Absolute Percentage Error (sMAPE), the Mean Absolute Scaled Error (MASE), and the mean absolute percentage error (MAPE) to measure the accuracy of the segmentation are used. sMAPE, MASE, and MAPE are given by the following formulae [25]:

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n} |e_i| / (|y_i| + |\hat{y}_i|) \quad (1)$$

$$MASE = (\frac{1}{n}\sum_{i=1}^{n} |e_i|) / (\frac{1}{n-1}\sum_{i=2}^{n} |y_i - y_{i-1}|) \quad (2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |e_i / y_i| \quad (3)$$

VI. DISCUSSION

The table below offers a comprehensive study of a dataset pertaining to student grades. The dataset has data from a total of 45 students. The mean student grade within the sample is around 3.7811. The standard error of the mean (SE Mean) is 0.0931. The standard error is a measure of the precision of the sample mean estimate. The standard deviation is 0.6244. The standard deviation quantifies the degree of variability or spread within a dataset. The minimum student grade in the dataset is 2.22. The first quartile (Q1) is 3.445. The first quartile is the threshold below which 25% of the data is located. The median value is 3.94. The median represents the central value of a dataset after it has been arranged in ascending or descending order. The third quartile, denoted as Q3, is the threshold below which 75% of the data is located. The value in this instance is 4.19. The highest attainable grade contains a student grade with the highest reported value of 4.8.

TABLE I. DATASET DESCRIPTION

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
s'studentgrade	45	3.7811	0.0931	0.6244	2.22	3.445	3.94	4.19	4.8

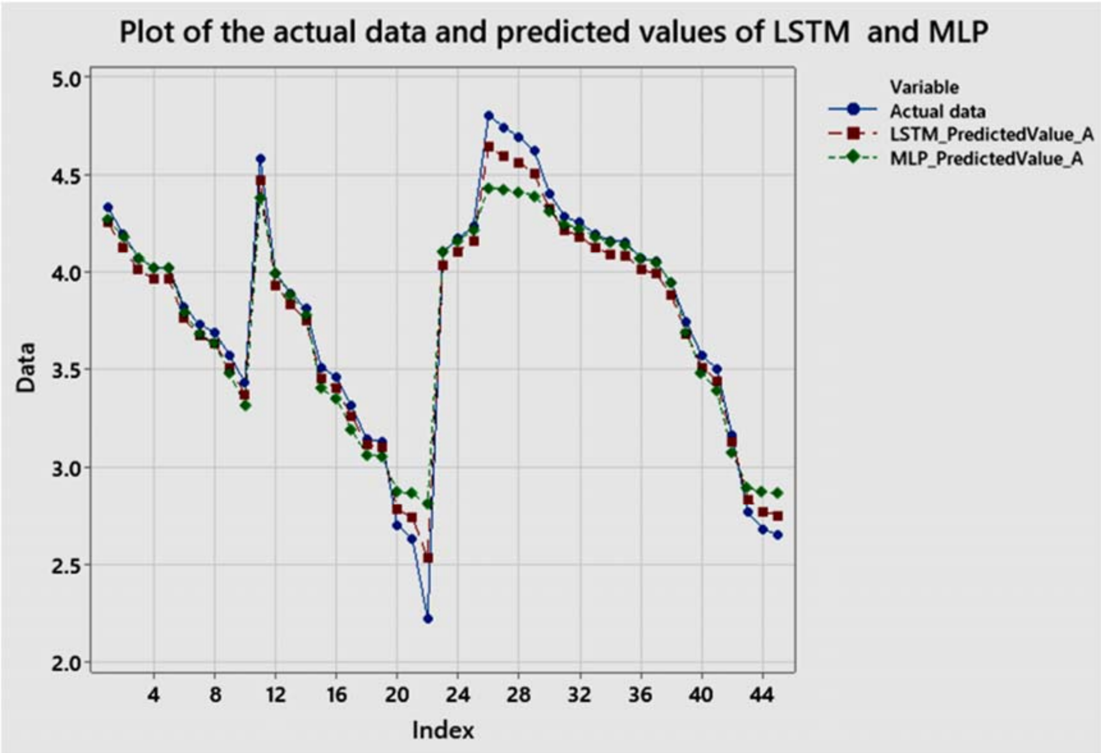


Figure 1. Plot of the actual data and predicted values of LSTM and MLP

Describing the plot of observed data and forecasted values for both LSTM and MLP models entails analyzing the extent to which each model accurately captures the patterns and trends present in the data. Here is a comprehensive descriptive examination of such plots:

- The x-axis reflects the sequential data items here are the students.
- The y-axis represents the observed and forecasted values.
- Data points are represented graphically as a line with dots marking the true values of students' grades and show analyzing the dispersion of real values over some among different data points.
- The Predictions made using LSTM are depicted by a red line distinguishable marker, Conducted a comparison between the predictions made by the LSTM model and the actual data.
- Observed that the LSTM models had the ability to accurately represent the general patterns, high points, and low points in the data.
- The evaluation of the LSTM model responds adequately to changes in the real values.
- Furthermore, the graphic displays the projected values generated by the MLP, which is depicted by a green line. The figure shows a comparison between the predictions made by the MLP model and the actual data

	SMAPE	MSE	MAPE
LSTM	0.010826	0.37335278	0.021774807
MLP	0.014484	0.47391199	0.029597513

The table above displays performance metrics for two distinct models, LSTM and MLP, utilizing three assessment measures: SMAPE (Symmetric Mean Absolute Percentage Error), MSE (Mean Squared Error), and MAPE (Mean Absolute Percentage Error). The findings indicate that the LSTM model surpasses the MLP model in terms of all three measures (sMAPE, MSE, and MAPE). The reduced values for sMAPE, MSE, and MAPE in the LSTM column indicate superior overall prediction accuracy in comparison to the MLP model.

VII. CONCLUSION

The thorough examination of the student grades dataset, along with the visual examination of the plot illustrating the actual data and model predictions, provides useful insights into the predictive capabilities of LSTM and MLP models.

The description of the dataset, focused on student grades, offered a comprehensive insight into academic performance indicators, such as average, standard deviation, quartiles, and the general grade distribution. The visual inspection of the plot displaying real student grades and prediction using both LSTM and MLP models unveiled a conspicuous disparity in performance. The predictions made by LSTM were highly accurate and closely matched the actual data trends, demonstrating its capability to capture complex patterns and fluctuations in student performance.

The utilization of precise accuracy metrics such as SMAPE, MSE, and MAPE were employed as quantitative indicators of the model's performance. The LSTM model consistently outperformed the MLP model in all metrics, exhibiting lower values for SMAPE, MSE, and MAPE. This indicates a higher level of accuracy, a small margin of error, and improved predictive precision. The LSTM model, known for its intricate structure, had exceptional efficacy in forecasting student grades, outperforming the MLP model in terms of accuracy measures. Lower values in SMAPE, MSE, and MAPE not only imply reduced prediction errors but also reflect a more sophisticated comprehension of the underlying patterns in the dataset.

The superior effectiveness of the LSTM model indicates its practical suitability for accurate student grade predictions.

Decision-makers, educators, and stakeholders can utilize the knowledge obtained from the LSTM model to make well-informed decisions on academic interventions and resource allocation. Continued investigation into the elements that contribute to the exceptional performance of LSTM can provide valuable insights for improving future models. Continual improvement of the LSTM structure and ongoing assessment of the model will enhance the long-term effectiveness in forecasting student grades. In summary, the results confirm that the LSTM model is highly efficient and accurate, making it the ideal option for predictive modeling in the context of student grades. These findings provide both immediate advantages for academic decision-making and establish the foundation for continuous progress in predictive analytics within the educational field.

REFERENCES

[1] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1), 237.

[2] Abuzinadah, N., Umer, M., Ishaq, A., Al Hejaili, A., Alsubai, S., Eshmawi, A. A., ... & Ashraf, I. (2023). Role of convolutional features and machine learning for predicting student academic performance from MOODLE data. Plos one, 18(11), e0293061.

[3] Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences, 10(11), 3894.

[4] Abuzinadah, N., Umer, M., Ishaq, A., Al Hejaili, A., Alsubai, S., Eshmawi, A. A., ... & Ashraf, I. (2023). Role of convolutional features and machine learning for predicting student academic performance from MOODLE data. Plos one, 18(11), e0293061.

[5] Chen, J.F., Hsieh, H. N., & Do, Q. H. (2014). Predicting student academic performance: A comparison of two meta-heuristic algorithms inspired by cuckoo birds for training neural networks. Algorithms, 7(4), 538-553.

[6] Kovacic, Z. (2010). Early prediction of student success: Mining students' enrolment data.

[7] Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews), 40(6), 601-618.

[8] Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. Sustainability, 15(16), 12451.

[9] Bauer, E., Greisel, M., Kuznetsov, I., Berndt, M., Kollar, I., Dresel, M., ... & Fischer, F. (2023). Using natural language processing to support peer‐feedback in the age of artificial intelligence: A cross‐disciplinary framework and a research agenda. British Journal of Educational Technology.

[10] Hu, J., Peng, Y., Chen, X., & Yu, H. (2021). Differentiating the learning styles of college students in different disciplines in a college English blended learning setting. PLoS One, 16(5), e0251545.

[11] Van der Schaar, M., Alaa, A. M., Floto, A., Gimson, A., Scholtes, S., Wood, A., ... & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. Machine Learning, 110, 1-14.

[12] Staudemeyer, R. C., & Morris, E. R. (2019). Understanding LSTM--a tutorial into long short-term memory recurrent neural networks. arXiv preprint arXiv:1909.09586.

[13] Alshamari, M. A. (2023). Evaluating User Satisfaction Using Deep-Learning-Based Sentiment Analysis for Social Media Data in Saudi Arabia’s Telecommunication Sector. Computers, 12(9), 170.

[14] Jehangir, B., Radhakrishnan, S., & Agarwal, R. (2023). A survey on Named Entity Recognition—datasets, tools, and methodologies. Natural Language Processing Journal, 3, 100017.

[15] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

[16] Syed, M. A. B., & Ahmed, I. (2023). A CNN-LSTM Architecture for Marine Vessel Track Association Using Automatic Identification System (AIS) Data. arXiv preprint arXiv:2303.14068.

[17] Wang, G., Wei, W., Jiang, J., Ning, C., Chen, H., Huang, J., ... & Ye, L. (2019). Application of a long short-term memory neural network: a burgeoning method of deep learning in

forecasting HIV incidence in Guangxi, China. Epidemiology & Infection, 147.

[18] Alkadri, S., Ledwos, N., Mirchi, N., Reich, A., Yilmaz, R., Driscoll, M., & Del Maestro, R. F. (2021). Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. Computers in Biology and Medicine, 136, 104770.

[19] Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D. J., & Maharaj, K. T. (2023). The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. Sustainability, 15(13), 10543.

[20] Li, H., Gao, W., Xie, J., & Yen, G. G. (2023). Multiobjective bilevel programming model for multilayer perceptron neural networks. Information Sciences, 642, 119031.

[21] Kang, T., Ding, W., Zhang, L., Ziemek, D., & Zarringhalam, K. (2017). A biological network-based regularized artificial neural network model for robust phenotype prediction from gene expression data. BMC bioinformatics, 18, 1-11.

[22] Chen, T. (2015). Analyzing and forecasting the global CO2 concentration-a collaborative fuzzy-neural agent network approach. Journal of applied research and technology, 13(3), 364-373.

[23] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., ... & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. Information Fusion, 99, 101805.

[24] Mehmood, F., Ahmad, S., & Whangbo, T. K. (2023). An Efficient Optimization Technique for Training Deep Neural Networks. Mathematics, 11(6), 1360

[25] H., Mandelkow, J. A.,De Zwart, & J. H., Duyn, . Linear discriminant analysis achieves high classification accuracy for the BOLD fMRI response to naturalistic movie stimuli. Frontiers in human neuroscience, 10, 128.(2016).

.