

Article

Predicting At-Risk Students Using Clickstream Data in the Virtual Learning Environment

Naif Radi Aljohani ¹, Ayman Fayoumi ¹ and Saeed-Ul Hassan ^{2,*} 

¹ Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; nraljohani@kau.edu.sa (N.R.A.); afayoumi@kau.edu.sa (A.F.)

² Department of Computer Science, Information Technology University, Lahore 54600, Pakistan

* Correspondence: saeed-ul-hassan@itu.edu.pk; Tel.: +92-322-228-9756

Received: 10 November 2019; Accepted: 26 November 2019; Published: 17 December 2019



Abstract: In higher education, predicting the academic performance of students is associated with formulating optimal educational policies that vehemently impact economic and financial development. In online educational platforms, the captured clickstream information of students can be exploited in ascertaining their performance. In the current study, the time-series sequential classification problem of students' performance prediction is explored by deploying a deep long short-term memory (LSTM) model using the freely accessible Open University Learning Analytics dataset. In the pass/fail classification job, the deployed LSTM model outperformed the state-of-the-art approaches with 93.46% precision and 75.79% recall. Encouragingly, our model superseded the baseline logistic regression and artificial neural networks by 18.48% and 12.31%, respectively, with 95.23% learning accuracy. We demonstrated that the clickstream data generated due to the students' interaction with the online learning platforms can be evaluated at a week-wise granularity to improve the early prediction of at-risk students. Interestingly, our model can predict pass/fail class with around 90% accuracy within the first 10 weeks of student interaction in a virtual learning environment (VLE). A contribution of our research is an informed approach to advanced higher education decision-making towards sustainable education. It is a bold effort for student-centric policies, promoting the trust and the loyalty of students in courses and programs.

Keywords: students-at-risk; virtual learning environment (VLE); classification; deep learning; long short-term memory (LSTM)

1. Introduction

The abundance of the vast available educational data provides opportunities to utilize it for various purposes, such as tapping the learning behaviors of the stakeholders involved, improving these behaviors by addressing the issues, and optimizing the learning environment [1]. With the readily accessible educational data, several research communities have exhibited noticeable interest in predicting students' patterns and extracting meaningful insights from these patterns. Such information extraction is not only bound to the data mining community. However, new communities have also emerged that focus not only on the agenda of improving students' performance but, as a whole, optimizing the learning environment, referred to as learning analytics [2]. Educational data, accumulated due to the interactional activity between learners and instructors, has been substantiated as a multidisciplinary field of study, involving researchers from various research communities, which has yielded to the inclusion of numerous terms associated with the exploration of the educational data, such as academic analytics, predictive analytics, and learning analytics [3].

With the emergence of the learning analytics research community, much emphasis has been laid on the investigation of students' behavior, assembling methods to improve understanding to

yield an optimal environment with enhanced learners' performance by early predicting potential grades [4,5]. Such practices contribute to maintaining and achieving a positive educational atmosphere that subsequently supports an institute to maintain its conduct [6,7]. In higher education, the learning analytics paradigm is cumulatively defined with multiple dimensions, including academic analytics, assisting institutes in maintaining their finance sectors by providing proper resource allocation practices to reduce student attrition, understanding learner's behavior to propose counteractive policies for enhancing the learning mechanism [8]. Furthermore, it assists an institute in maintaining student retention, ultimately contributing to a higher number of graduation rates [9].

An array of various data analytic techniques and machine learning practices are administered for the prediction of several measures and events, with deep artificial neural networks (ANNs) being a prominent practice among these due to their learning abilities [10]. The paradigm of deep learning is defined as hierarchical representational learning, encompassing various layers of computation and enabling the system to learn from prevailing examples, intervening in the traditional feature engineering methods [11,12]. In the literature, only a limited number of studies can be observed that measure the effectiveness of deep learning approaches in the learning analytics paradigm, especially for early intervening with students to improve their performance. Recently, a systematic literature review conducted by LeCun and Silveira [10] analyzed the learning analytics studies that deployed deep learning practices, demarcating some prominent problems where such practices superseded the conventional statistical practices; such as learners' performance [6,9,13], their knowledge assessments [14], and their writing recognition patterns [15]. To intervene early with the students for optimal performance, various forms of ANNs, such as recurrent neural network (RNN) and long short-term memory (LSTM), have been employed. These practices correspond to the course duration as sequential data by analyzing each learner's daily, weekly, or monthly performance. Overall, in this discipline, the research community is embarking towards the adoption of these sequential practices to predict early on the students at risk of poor performance and intervene with them on time for optimal results.

In the recent past, online educational systems have emerged as a rising phenomenon, contributing in the generation of educational data repositories encompassing learners' interactions, activities, and engagement patterns, which can be further analyzed to capture the behavior of students and extracting critical differences in the engagement patterns of successful students and those at risk. Such analysis assists the academic and administrative community to formulate optimal policies and corrective strategies for the improvement of risky students and to yield a supportive pedagogical system [14,16]. This study intends to investigate the effectiveness of deep learning practices to early predict the students at risk of a failure in a virtual learning environment, from the easily accessible Open University Learning Analytics (OULA) dataset. The objectives of this study are as follows:

- Firstly, we intended to leverage deep learning models by transforming the dataset into a sequential format by assembling students' engagement with the virtual learning environment (VLE) weekly.
- Secondly, we delivered an understanding of the behavior of students at risk of failure, contributing to the decision making policies to devise early intervention strategies to improve student performance, enforcing student retention.
- Lastly, ascertaining the effectiveness of the deployed deep LSTM model in the early prediction of at-risk students compared to conventional approaches.

The organization of this paper is as follows. Section 2 briefly discusses the existing studies co-relating deep learning with various online learning platforms. Section 3 presents the data and the deployed methods for its analysis. Section 4 formulates the experimental setup and results, their discussion, and evaluation. The concluding remarks with the limitations and future directions are proposed in Section 4.

2. Literature Review

In several studies, the problem of predicting the academic performance of students is either classified as a regression problem, where the learners' perspective scores are predicted, or a classification problem where a learner's final result is predicted in terms of pass, fail, or dropout. The behavior of students in online learning platforms varies from that of the traditional classroom settings, where learners' inherent motivation to succeed and excel plays a significant contribution in their performance, making them solely responsible for their good/bad performance [17]. In the literature, substantial studies are found that emphasize the prediction of student performance by applying various data analytic techniques and highlighting various factors impacting a learners' performance, categorizing different features contributing to low student retention [18]. In online educational platforms, the attribute of time is considered as the vital feature of impacting learners and their performance, followed by the incoherent support provided by the instructors. Moreover, effective curriculum content is paramount to drive an intrinsic motivation in learners, encouraging them in positive and active participation, ultimately influencing a learner's performance and intention to complete a particular program [19,20].

Another dimension involves predicting students' grades for the next term potential courses [5,21]. Morsy et al. [22] characterized the course space as a latent vector and suggested regression models for next term grade prediction. Similarly, another study predicted course-specific next term grades through Markovian models [4]. Marbouti et al. [23] deployed logistic regression to assess students at risk of a failure by incorporating attributes of their attendances, quizzes, and examination behavior. They identified the students at risk of low performance in several weeks of the courses, and their predictions improved for the final weeks. Furthermore, logistic regression was also applied as the baseline evaluation technique to predict at-risk students [24]. The previous history of students, encompassing past grades in previous courses, assessments, and entry tests, is also an essential element in classifying and predicting an individual's performance [25].

The correlation of the learning analytics community with deep learning techniques is still in its preliminary stages, with few pieces of evidence observed in the existing literature. Deep learning, constituting several non-linear layers, enables self-adaptive models through the hierarchal representation phenomena, where each layer transcends the learned information, abstractly, to the successive layers [26]. Corrigan & Smeaton [13] deployed a variation of recurrent neural networks (RNN) to predict the success of students by incorporating interactional activities of students with the VLE. Their deployed deep learning approach outperformed the traditional baseline approach. Similarly, another study predicted success rate by including student attendance and tapped their behavior through log data information [27]. They predicted the grades of students, including their engagement patterns and interactions with the VLE, through the application of RNN and LSTM models. Through the deployed sequential model, they intended to predict the grades of students and identify at early stages those at risk. The deployed technique was compared with the conventional regression analysis and was found to be more effective in the early prediction on grades compared to traditional regression approaches. Fei and Yeung [28] employed a feature set, consisting of the lectures watched and downloaded, assessments scores and attempts, forum activities, and the number of times commented on forums, to predict the student performance and assess at-risk students. They implemented an array of techniques, such as support vector machines, logistic regression, input-output hidden Markov model, RNN, and LSTM, and found LSTM to surpass the other techniques.

An exciting dimension in the educational community has been the use of virtual reality in the learning context of distance education, facilitating teacher-to-student interaction more conveniently and elaborately [29]. This implicitly assists instructors in customized teaching practices for different categories of students. Furthermore, Kohler provided a theoretical framework for effective teaching to produce optimized learning outcomes in traditional and blended classroom settings [30]. E-portfolios are another tool established to facilitate the instructors in personalizing their teaching methods and supporting student interventions [31]. Overall, in a VLE, we only found a limited number of studies

embarking on the adoption of deep learning techniques to evaluate and comprehend student behavior in a rigorous manner. Our study leverages the power of deep learning for the early prediction of at-risk students in a VLE.

3. Data and Experimentation

The procured Open University Learning Analytics dataset (OULAD) constituted student demographics, clickstream history, and assessment submission information of 32,593 students over a course duration of 9 months, from 2014 to 2015 [32]. The data were composed of several courses, with each course being taught at different intervals in a year. Four distinct performance classes were defined: distinction, pass, fail, and withdrawal.

The OULAD comprised students' information regarding their interaction with the VLE—their assessments, quizzes, and course performances. The interaction with the VLE was further categorized into 20 different activity types with each activity referring to a specific action, such as downloading or viewing lectures, course content, or quizzes. The names of each of these activity types are as follows: dataplus, dualpane, externalquiz, folder, forumng, glossary, homepage, htmlactivity, oucollaborate, oucontent, ouilluminate, ouwiki, page, questionnaire, quiz, repeatactivity, resource, sharedsubpage, subpage, and url.

The aggregated average clicks per student were processed weekly to visualize the students' weekly interactions. Figure 1 depicts the number of clicks for the two classes—pass and fail—where the aggregated activities for each class were normalized according to per student. It can be observed that the two classes demarcated in terms of the interaction level in the VLE, with 'pass' instances being more active than 'fail' instances.

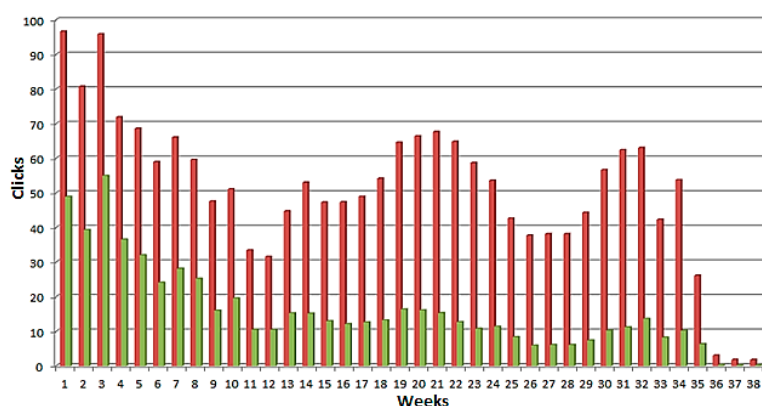


Figure 1. Week-wise average clicks per student.

3.1. Data Preprocessing

The OULAD was procured in a raw structured format with several data files. The log-file data were computed to obtain features catering to the various actions signifying students' interactions with the VLE. These features were formulated by processing the provided data tables in the database. The data were computed in a week-wise manner, with each week constituting the same activity features, and each week comprising a homogenous set of students, that is, students in the week i were also present in the week $(i-1)$ and so on. Each student was identified with a unique ID in the data. Only a ratio of 5.8% in the data took more than one course and hence were repeated; however, we did not intend to find academic performances of a student in multiple courses. Therefore, each student was identified by a unique ID for each course. Similarly, we did not intend to analyze student performance on a course-granular level; therefore, students repeating the same course were ignored. The unique student IDs were computed by a combination of their older IDs, the course that was taken, and the interval in which the course was presented. This study intended to analyze 'pass' and 'fail' instances,

where the 'pass' instances were merged with the distinction instances to formulate one class. Hence, the formulated data consisted of 22,437 instances.

3.2. Approach

The structure of deep learning methods is composed of many non-linear levels, where each level plays a significant role in transcending the learned representation more abstractly to its higher layers, consequently assisting in learning complex functionalities [33]. As opposed to the conventional statistical approaches, such self-adaptive techniques effectively determine the underlying associations between the data by generalizing the input sequence and learning from it [34]. RNNs are designed to learn the long-term dependencies in the sequential data through a recursive loop at each cell that supports it to keep a check on the previous input data along with the current input. RNN weights are updated by backpropagation through time, which transmits the error and the gradient over the whole vector [35]. Though RNNs were designed for long-term dependencies, pieces of evidence in the empirical research demonstrate the opposite. Due to the vanishing and exploding gradient problems, the error in RNN can only be back-propagated to a short distance [36]. To resolve these issues, LSTM was introduced, where a memory cell is augmented in the network, enabling it to retain longer sequences.

LSTM constitutes three gates; at some particular time instance say t , the input gate i_t administers the writing of the data in the LSTM unit, forgetting gate f_t is responsible for controlling the data that is the amount of data to be written, memory cell C_t is responsible for retaining the past information, and output gate o_t manages the representation of the delivered output. The manipulation of forgetting gate f_t at some time t , held responsible for administering the amount of data to be retained, is described mathematically in Equation (1):

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f). \quad (1)$$

Further on, the input gate computes the information by multiplying the input x_t with the activation of the input gate and determining the relevant information to be retained, as shown in Equations (2) and (3):

$$\xi_t = \text{Tanh}(W_\xi[h_{t-1}, x_t] + b_\xi), \quad (2)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i). \quad (3)$$

A layer of LSTM constitutes multiple blocks, each with the required gates and memory component [37]. The memory component C_t is updated at each interval,

$$C_t = f_t C_{t-1} + i_t \xi_t. \quad (4)$$

Three parameters—input, memory component, and previous hidden states h_{t-1} —cumulatively update the currently hidden state h_t via the output gate o_t , as shown in Equation (5):

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (5)$$

$$h_t = o_t \cdot \text{Tanh}(C_t), \quad (6)$$

where b_f , b_ξ , b_i , b_o , W_f , W_ξ , W_i , W_o are the corresponding biases and weights for the gates, and σ is the component-wise logistic sigmoid function. Tanh in Equations (2) and (5) is the activation function that computes the candidate values and adds it to the memory component C_t .

The augmented memory cell in LSTM blocks facilitates the model with a lookback window that is flexible enough to retain longer sequences, which ultimately assists in making decisions on the basis of both past and current input sequences [38]. The proposed study deployed a deep sequential model for the early prediction of at-risk students on the basis of their week-wise clickstream interactions with the VLE. To attain this purpose, the OULA dataset was processed to retrieve weekly clickstream data for each activity. The clickstream information for each student was recorded in a weekly manner,

where each i^{th} week consisted of the interactions and activities of students for that specific i^{th} week, as illustrated in Figure 2, where S1, S2 ... Sn represents the unique students that are the same for all weeks.

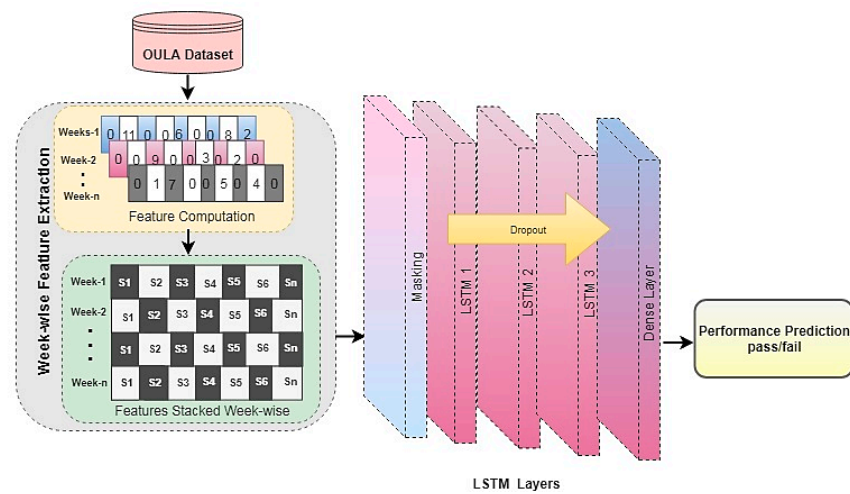


Figure 2. Proposed architecture of long short-term memory (LSTM)-based deep learning.

This week-wise stack of students forms a vector consisting of the appended weeks' sequence, which was passed to the sequential model. Week-wise data were manipulated in such a way that for each i^{th} week, the sequence vector was appended until i^{th} weeks. These i^{th} weeks ranged from the starting first week to the last 38th week in the course. Because the week-wise vector length depended on the i^{th} week, therefore, the padding was implemented to enable an equal length of the vector. These padded values were masked before passing to the model. The masked values were ignored by the model, and the model did not learn its impact. Thus, LSTM layers were implemented with a flexible lookback window, enabling early prediction for each i^{th} week. Multiple layers of the LSTM model were implemented in the architecture, depicted in Figure 2, assisting the model to learn complicated and intricate details of student engagement vector where each higher layer presented its output to its lower layers [39]. A layered architecture assisted the model in learning the inherent data representations more accurately.

3.3. Experimentation and Evaluation

This section presents the experimental setup for the evaluation of the deployed deep LSTM model. To early predict the students' academic behavior, the problem was first converted to a binary classification through the inclusion of 'fail', 'distinction', and 'pass' categories. A total number of 22,437 instances were included, where the classes 'pass' and 'distinction' were formulated as 'pass' class. Three layers were implemented in the deep LSTM architecture, and each layer constituted a range of neurons, from 100 to 300 units. The dropout was implemented between the layers to reduce the overfitting of the data. This also inhibited the inter-dependency between neurons and enabled the model to learn more effectively and rigorously [40]. At each instant of time t_i , data constituting of a vector at particular instance t_i was passed to the deep LSTM model. ADAM optimizer - an adaptive learning rate optimization algorithm that's been designed specifically for training deep neural networks - was used as the optimizer with some hyper-parameter tuning, where the learning rate was set from 0.0 to 0.0001. To calculate the efficiency of the deployed model, in terms of the difference between the actual and predicted values, binary cross-entropy was applied as the loss function. The problem of students' early prediction was constituted as a binary problem, with a student either passing or failing a course. Therefore, this loss function produced optimal results. The equation for binary cross-entropy

is provided in Equation (7) with p_i representing the likelihood of the actual data and q_i representing the likelihood of the predicted data to pass or fail.

$$\sum_0^n p_i \log \frac{1}{q_i} \quad (7)$$

The predictions of the deployed deep LSTM model tended to improve over the weeks, as illustrated in Figures 3 and 4, where with additional weeks, the model gradually depicted the improvement in the prediction of academic performance. The accuracy ranged from week 5 to week 38, showing an increase in the predicted academic performance, as shown in Figure 3a. A particular behavior of students could not be determined in the initial weeks. Therefore, the model was deployed on the dataset of the 5th week and onwards. Results were displayed of selective weeks to depict the refinement and improvement of the model. As previously mentioned, with additional history of students' interaction and their clickstream behavior, the model predicted the academic performance for each student (pass/fail) with a confidence of 69.69% achieved in the 1st week, whereas 80.82% was acquired in the 5th week, and 95.23% was acquired in the last week. Figure 3a also depicts an increasing trend in the accuracy of the predictions after the initial 5th week, depicting the robustness of the deployed deep LSTM performing better with accumulated clickstream data. From the 10th week, the model can predict a student as passing or failing with reasonable accuracy of over 85%. Therefore, this pattern of leaning is a vital determinant in the early prediction of students' academic performance. As accuracy increases with additional weeks, the loss values tend to decrease with additional week-wise information, insinuating the robustness of the model, which increases as the student behavior is determined with additional week-wise engagement patterns. Note that Figure 3b shown the learning loss of the model.

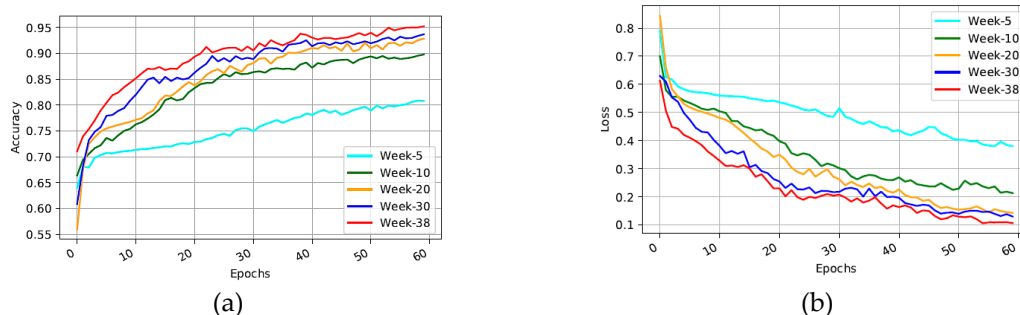


Figure 3. Week-wise learning metrics: accuracy and loss curves. (a): Learning accuracy of employed model across 60 epochs; (b): Loss value of employed model across 60 epochs.

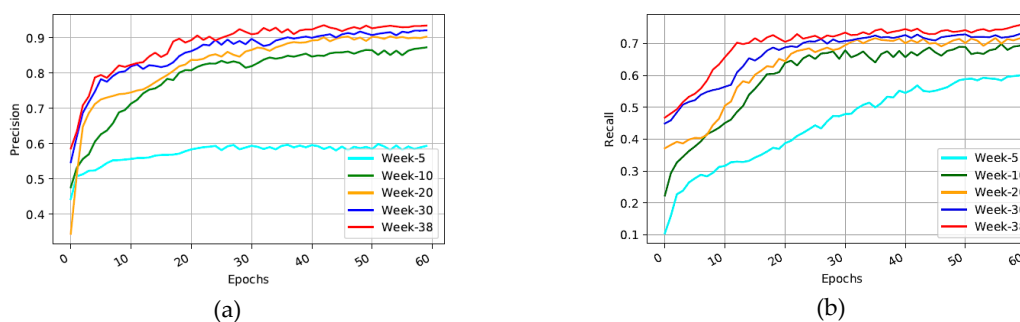


Figure 4. Week-wise validation metric: precision and recall curves. (a): Precision score of employed model across 60 epochs; (b): Recall score of employed model across 60 epochs.

The precision and recall curves of the validation data are depicted in Figure 4 that tend to improve with additional week-wise information. Precision was defined as the ratio of the at-risk students

identified correctly from the total number of the students identified as at-risk by the model. The recall was defined as the ratio of the students captured as at-risk by the model from the total number of at-risk students in the actual data. As illustrated in Figure 4, the precision and recall curves tended to significantly elevate after the 5th week, insinuating the significance of additional information in the model. As the model was fed more information about students' interaction and their engagement activities, the model tended to learn their behavior and produce better results. Precision values tended to improve from 59.36% achieved in the 1st week to 93.46% achieved in the last week. Similarly, recall values improved from 60.99% achieved in the 1st week to 75.79% achieved in the last week. Moreover, the learning accuracy and loss for all the weeks, from 1st to 38th, are depicted in Figure 5, illustrating the improvement in the accuracy values with additional weeks and the week-wise degradation in the loss values, implying the robustness of the predicted model. These values were obtained at the 60th epoch of the trained model.

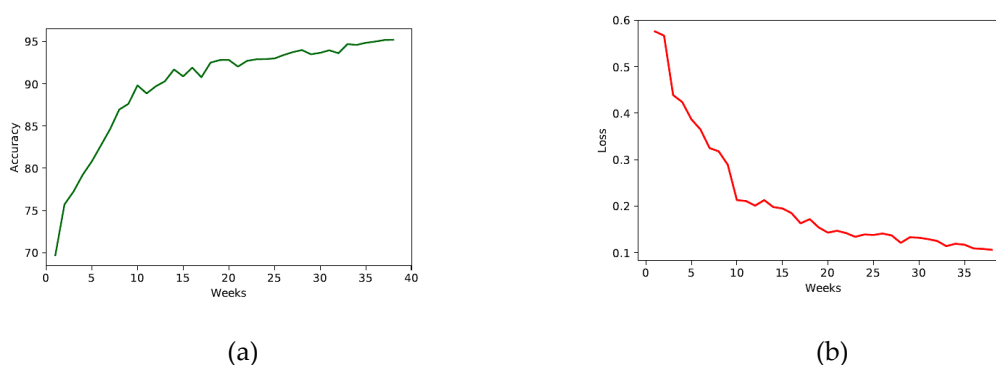


Figure 5. Learning accuracy and loss for all weeks at the 60th epoch. (a): Learning accuracy of employed model across the weeks; (b): Loss value of the model across the week.

3.4. Evaluation with Baseline

To evaluate the deployed deep LSTM model for the early prediction of at-risk students, several machine learning algorithms were deployed as baselines. ANN, Support Vector Machine (SVM), and Logistic Regression (LR) have been frequently adopted in the educational research community for evaluating other proposed models [23]. Aggregated data were processed for the machine learning algorithms, where each specific i^{th} week data were aggregated until that i^{th} week. A week-wise flat vector was hence computed for each student and passed to the model. The results in comparison to LSTM are illustrated in Figure 6, where W5, W10, W20, W30, W38, represent the 5th, 10th, 20th, 30th, and 38th weeks, respectively. It can be observed that LSTM performed significantly better than other baseline models in predicting at-risk students.

LSTM works sufficiently well for sequential data; hence, it efficiently analyzed the behavior of students in a weekly manner and produced optimal results compared to baseline techniques. Because the conventional non-sequential models worked with the aggregated data, which constituted the entirety of information until that particular time, hence such models were unable to analyze the behavior on a time granularity level. Due to the complete week-wise information aggregated into values, the model was unable to learn this weekly behavior or predict the at-risk students from their collective interactions aggregated into a vector. Such aggregation hindered the learning predictability of the model; therefore, the sequential model, with a good fit, captured the learning behavior of students and efficiently predicted the at-risk students on the basis of their interactions and engagement patterns.

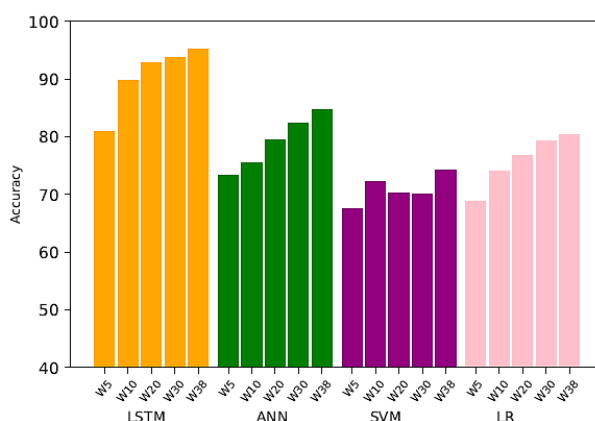


Figure 6. Evaluation with the baseline techniques. LR: logistic regression, ANN: artificial neural network, SVM: support vector machine. W5, W10, W20, W30, W38, represent the 5th, 10th, 20th, 30th, and 38th weeks, respectively.

3.5. Implication of Results

This research presented the critical concern of the identification of students at risk of failure. Different computation methods were deployed to predict the students at risk of a failure on the basis of their behavior and engagement patterns with the VLE. Deep LSTM improved the predictability of the students' decisions and assisted the educational community to develop guidelines for helping the at-risk students. Such behavioral models craft a path for the administrative authorities to contribute to formulating policies and strategies to implement timely interventions, regulate the decision-making process, and ultimately assist students through the provision of support systems. Moreover, such settings will also help establish guidance committees and regular student counseling sessions for maintaining a motivational infrastructure and tapping the behavior of students for data-driven decision-making processes.

4. Concluding Remarks, Limitations, and Future Extensions

This study examined at-risk students by converting the problem into a sequential weekly format and measuring the effectiveness of the deep sequential model versus the conventional machine learning baseline models. We intended to deliver an understanding of the behavior of students at risk of failure, contributing to the decision making policies to devise early intervention strategies to improve student performance, enforcing student retention. The deep LSTM tended to monitor the sequential week-wise pattern of students and their activities, performing better compared to conventional classifiers that deal with students' interactions in a collective and aggregated manner. Such early predictions will facilitate institutions to timely intervene in the students at risk by providing them a support system through counseling and alert emails. Such data-driven analysis will also assist the decision-makers to formulate optimal policies for students' success on the basis of their behavior and interaction patterns. The deployed criteria to determine early on the risky students will assist the educational community in capturing their activities and behavior for student retention by intervening on time, providing pedagogical support in terms of the formulation of guidance committees and corrective strategies.

This study does not cater to the variation in the performances of the students who repeat their courses. Therefore, analyzing their behavior is another dimension that requires sufficient data of such students. Similarly, a course-level analysis is also required to capture the behavioral differences of the same students in different courses and identify the influential elements to tap their academic performance. Because the deployed dataset did not have sufficient course-level records, this is hence a limitation of our study. Moreover, analyzing the behavior of repeating students in one course and examining the differences in their first and second attempts is another crucial area of research. In the

future, we plan to enrich our model prediction by including students' assessment scores, analyzing the association between their assessment submission pattern and performance.

Furthermore, we also seek to investigate activities having an influential impact on the performance by mining textual data [41–43] of students' feedback by employing deep advance learning [44] and natural language processing techniques [45]. A framework catering to the prominent attributes—extrinsic and intrinsic—associated with students' performance may enable the learning analytics community to adhere towards more effective decision-making systems. Moreover, analyzing students' behavior on a day-to-day basis is another dimension of interest that will assist the educational community in identifying the most influential phase where the students tend to demonstrate positive performance

Author Contributions: N.R.A.: supervision, investigation, design, validation, writing—original draft preparation, and funding. A.F.: supervision, methodology, design, validation, writing—original draft preparation, and funding. S.-U.H.: conceptualization, design, methodology, investigation, writing—original draft preparation, and writing—review and editing.

Funding: This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant no. D-106-611-1440. The authors, therefore, greatly acknowledge the DSR technical and financial support.

Conflicts of Interest: The authors declare that they have no known competing for financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Yi, C.; Kang-Yi, C. Predictive analytics approach to improve and sustain college students' non-cognitive skills and their educational outcome. *Sustainability* **2018**, *10*, 4012. [CrossRef]
2. Schumacher, C.; Ifenthaler, D. Features students really expect from learning analytics. *Comput. Hum. Behav.* **2018**, *78*, 397–407. [CrossRef]
3. Viberg, O.; Hatakka, M.; Bälter, O.; Mavroudi, A. The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* **2018**, *89*, 98–110. [CrossRef]
4. Hu, Q.; Rangwala, H. Course-Specific Markovian Models for Grade Prediction. In *Advances in Knowledge Discovery and Data Mining. PAKDD 2018. Lecture Notes in Computer Science*; Phung, D., Tseng, V., Webb, G., Ho, B., Ganji, M., Rashidi, L., Eds.; Springer: Cham, Germany, 2018; Volume 10938, pp. 29–41.
5. Polyzou, A.; Karypis, G. Grade prediction with course and student specific models. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Cham, Germany, 2016; pp. 89–101.
6. Baker, R.S.; Inventado, P.S. Educational data mining and learning analytics. In *Learning Analytics*; Springer: New York, NY, USA, 2014; pp. 61–75.
7. Daniel, B.K. Big Data in Higher Education: The Big Picture. In *Big Data and Learning Analytics in Higher Education*; Daniel, B.K., Ed.; Springer: Cham, Germany, 2017; pp. 19–28.
8. Rienties, B.; Boroowa, A.; Cross, S.; Kubiak, C.; Mayles, K.; Murphy, S. Analytics4Action Evaluation Framework: A Review of Evidence-Based Learning Analytics Interventions at the Open University UK. Available online: <https://eric.ed.gov/?id=EJ1089327> (accessed on 16 December 2019).
9. Palmer, S. Modelling engineering student academic performance using academic analytics. *Int. J. Eng. Educ.* **2013**, *29*, 132–138.
10. Coelho, O.B.; Silveira, I. Deep Learning Applied to Learning Analytics and Educational Data Mining: A Systematic Literature Review. Proceedings of the Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), Nova Scotia, Canada; 2017. Available online: <https://br-ie.org/pub/index.php/sbie/article/view/7543> (accessed on 16 December 2019).
11. Poplin, R.; Varadarajan, A.V.; Blumer, K. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2018**, *2*, 158. [CrossRef] [PubMed]
12. Guo, B.; Zhang, R.; Xu, G.; Shi, C.; Yang, L. Predicting Students Performance in Educational Data Mining. In Proceedings of the 2015 International Symposium on Educational Technology (ISET), Wuhan, China, 27–29 July 2015.

13. Corrigan, O.; Smeaton, A.F. A Course Agnostic Approach to Predicting Student Success from VLE Log Data Using Recurrent Neural Networks. In Proceedings of the European Conference on Technology Enhanced Learning, Tallinn, Estonia, 12–15 September 2017; pp. 545–548.
14. Li, J.; Wong, Y.; Kankanhalli, M.S. Multi-stream Deep Learning Framework for Automated Presentation Assessment. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016.
15. Gross, E.; Wshah, S.; Simmons, I.; Skinner, G. A Handwriting Recognition System for the Classroom. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Poughkeepsie, NY, USA, 16–20 March 2015.
16. Wang, L.; Sy, A.; Liu, L.; Piech, C. Deep Knowledge Tracing on Programming Exercises. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, Cambridge, MA, USA, 20–21 April 2017.
17. Davis, H.C.; Dickens, K.; Leon Urrutia, M.; Vera, S.; del Mar, M.; White, S. MOOCs for Universities and Learners an Analysis of Motivating Factors. In Proceedings of the 6th International Conference on Computer Supported Education, Barcelona, Spain, 1–3 April 2014.
18. Hone, K.S.; El Said, G.R. Exploring the factors affecting MOOC retention: A survey study. *Comput. Educ.* **2016**, *98*, 157–168. [CrossRef]
19. Fidalgo-Blanco, Á.; Sein-Echaluce, M.L.; García-Peñalvo, F.J.; Conde, M.Á. Using learning analytics to improve teamwork assessment. *Comput. Hum. Behav.* **2015**, *47*, 149–156. [CrossRef]
20. Khan, I.U.; Hameed, Z.; Yu, Y.; Islam, T.; Sheikh, Z.; Khan, S.U. Predicting the acceptance of MOOCs in a developing country: Application of task-technology fit model, social motivation, and self-determination theory. *Telemat. Inform.* **2018**, *35*, 964–978. [CrossRef]
21. Bydžovská, H.A. Comparative Analysis of Techniques for Predicting Student Performance. In Proceedings of the 9th International Conference on Educational Data Mining 2016, Raleigh, NC, USA, 29 June–2 July 2016.
22. Morsy, S.; Karypis, G. Cumulative Knowledge-based Regression Models for Next-Term Grade Prediction. In Proceedings of the 2017 SIAM International Conference on Data Mining, Houston, TX, USA, 27–29 April 2017.
23. Marbouti, F.; Diefes-Dux, H.A.; Madhavan, K. Models for early prediction of at-risk students in a course using standards-based grading. *Comput. Educ.* **2016**, *103*, 1–15. [CrossRef]
24. Marbouti, M.F.; Diefes-Dux, H.A. Building course-specific regression-based models to identify at-risk students. *Age* **2015**, *26*, 1.
25. Leitner, P.; Khalil, M.; Ebner, M. Learning analytics in higher education—A literature review. In *Learning Analytics: Fundamentals, Applications, and Trends*; Springer: Houston, TX, USA, 27–29 April 2017; pp. 1–23.
26. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
27. Okubo, F.; Yamashita, T.; Shimada, A.; Ogata, H. A Neural Network Approach for Students' Performance Prediction. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, BC, Canada, 13–17 March 2017.
28. Fei, M.; Yeung, D.Y. Temporal Models for Predicting Student Dropout in Massive Open Online Courses. In Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW), Atlantic, NJ, USA, 14–17 November 2015.
29. Klampfer, A. Virtual/Augmented Reality in Education Analysis of the Potential Applications in the Teaching/Learning Process. Available online: https://www.researchgate.net/publication/318680101_VirtualAugmented_Reality_in_Education_Analysis_of_the_Potential_Applications_in_the_TeachingLearning_Process (accessed on 28 November 2019).
30. Gettinger, M.; Kohler, K.M. Process-outcome approaches to classroom management and effective teaching. In *Handbook of Classroom Management*; Routledge: Abingdon, UK, 2013; pp. 83–106.
31. Klampfer, A.; Köhler, T. Learners' and teachers' motivation toward using e-portfolios. An empirical investigation. *Int. J. Cont. Eng. Educ. Life-Long Learn.* **2015**, *25*, 189. [CrossRef]
32. Kuzilek, J.; Hlosta, M.; Zdrahal, Z. Open university learning analytics dataset. *Sci. Data* **2017**, *4*, 170–171. [CrossRef] [PubMed]
33. Nielsen, M.A. *Neural Networks and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015.
34. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62. [CrossRef]

35. Karpathy, A. The Unreasonable Effectiveness of Recurrent Neural Networks. Available online: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/> (accessed on 27 November 2019).
36. Sak, H.; Senior, A.; Beaufays, F. Long Short-term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. In Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore, 14–18 September 2014.
37. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [[CrossRef](#)] [[PubMed](#)]
38. Okubo, F.; Yamashita, T.; Shimada, A.; Konomi, S. Students' Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network. In Proceedings of the 25th International Conference on Computers in Education, Christchurch, New Zealand, 4–8 December 2017.
39. Ballesteros, M.; Dyer, C.; Smith, N.A. Improved Transition-Based Parsing by Modeling Characters Instead of Words with LSTMs. Available online: <https://arxiv.org/abs/1508.00657> (accessed on 27 November 2019).
40. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
41. Ananiadou, S.; Thompson, P.; Nawaz, R. Enhancing Search: Events and Their Discourse Context. In *International Conference on Intelligent Text Processing and Computational Linguistics*; Springer: Berlin, Germany, 2013; pp. 318–334.
42. Shardlow, M.; Batista-Navarro, R.; Thompson, P.; Nawaz, R.; McNaught, J.; Ananiadou, S. Identification of Research Hypotheses and New Knowledge from Scientific Literature. Available online: <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-018-0639-1> (accessed on 27 November 2019).
43. Hassan, S.U.; Visvizi, A.; Waheed, H. The 'who' and the 'what' in international migration research: data-driven analysis of Scopus-indexed scientific literature. *Behav. Inf. Technol.* **2019**, *38*, 924–939.
44. Jahangir, M.; Afzal, H.; Ahmed, M.; Khurshid, K.; Nawaz, R. An Expert System for Diabetes Prediction Using Auto Tuned Multi-layer Perceptron. In Proceedings of the 2017 Intelligent Systems Conference (IntelliSys), London, UK, 7–8 September 2017.
45. Batista-Navarro, R.; Theresa, G.K.; Mihăilă, C.; Thompson, P.; Rak, R.; Nawaz, R.; Korkontzelos, I.; Ananiadou, S. Facilitating the analysis of discourse phenomena in an interoperable NLP platform. In *International Conference on Intelligent Text Processing and Computational Linguistics*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 559–571.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).