EasyChair Preprint

№ 6965

Predicting At-Risk Students' Performance Based on Their Interactions and Assessments in Foundation Year English Courses at King Abdul-Aziz University

Amnah Alsulami, Maiada Almasre and Norah Almalki

October 31, 2021

Predicting At-Risk Students' Performance Based on their Interactions and Assessments in Foundation Year English Courses at King Abdul-Aziz University

Abstract*—Predicting online learners' final course achievement is most of the time dependant on their course interactions with LMS-based graded assessments, which can be used as an indicator to identify at-risk students in various academic contexts; ultimately, support students' success and academic advancement. This is where Learning Analytics (LA) representing learners' behavior inside the Learning Management Systems (LMS), and Deep Learning (DL) techniques come into play as academic data, which can be used as the gateway to informed predictions of learners' future performance. Not surprisingly, the need for developing at-risk profiling models becomes apparent in cases when a large student cohort is taking a foundation course online, for example, and where instructors cannot easily or effectively monitor their students' progress in real-time. The proposed study aims to utilize Artificial Neural Network (CNN, Encoder/Decoder LSTM, and GRU); to develop models that predict students' final grade (pass or fail) based on their interactions with the assessments' types (assignment or test). The LMS data used in this study was for 3,929 Students enrolled in a four-module English foundation course in King AbdulAziz University (KAU) which was collected during the second semester of 2020. The results show that the CNN model with the lowest RMSE value (RMSE = 3.15) performed better than the other models.

Keywords*—Predict at-risk student, Artificial Neural Network, Learning Management System, and Educational Data Mining.

I. INTRODUCTION

Educational Data Mining (EDM) is a new subfield of Data Mining (DM) that focuses on analyzing educational data to develop models that can be useful in improving learning experiences and enhancing institutions' efficiency. Researching EDM involves implementing techniques from statistical analyses, data mining, machine learning (ML), and deep learning (DL) to detect patterns in extensive educational data [1]. Many researchers established the effectiveness of using EDM techniques in the educational domain. The most critical issue that demands attention, as they believe, is related to the prediction of at-risk students based on their LMS behaviour.

There is no definite agreement on the scope and implications of the concept of at-risk students, which is attributed more or less to the ever-changing landscape of probable risks that learners might encounter, but the ultimate impact of all the identified risk factors is always on students' final grades [2][3].

Users' behaviour refers to the data representing their activities in all areas inside a system, an LMS, for example. This data can be in the form of counts, averages, or percentages, indicative of course accesses, submissions, clicks, time details, video logs, lectures, online evaluations, discussion forums, and even live video discussions, etc. [4] [5]. In an ideal situation, this data is analysed to predict students' performance, develop recommendation systems, analyse learners' behaviour, provide reports, and improve courses [6]

The inability to identify at-risk students in the four-module English course delivered during the Foundation year in King Abdul-Aziz University (KAU) is a critical issue that continually demands attention. The ultimate aim is to improve students' performance by giving them the opportunity to enhance their performance and avoid dropout or being academically dismissed from the programs. Based on the predictive output, an educational institution can enhance the responses and engagement of low-performing students. Although at-risk students' prediction is widely studied, it remains a challenge, as students' performance is influenced by numerous factors, including demographic, academic, social factors, etc. [7].

There are limited number of studies that address practical methods for detecting at-risk students based on Deep Learning (DL) algorithms, which can deal with a big dataset and provide high-quality. Hence, some of those studies didn't based on real data.

Therefore, the purpose of this study is to develop an Artificial Neural Network (CNN, Encoder/Decoder LSTM, GRU) models that capable to predicts at-risk students enrolled in the 4-level English course taught at KAU, based on their LMS assessments.

This research is organized as follows: Section II is a literature review about Machine and Deep learning methods used in students' performance prediction. Section III, describes the research methodology. Section IV, presents the findings of our research. Section V, offers an interpretation of the cited results. Section VI, is the conclusion with reference to future work.

II. LITERATURE REVIEW

Predicting students' final performance helps students change their study patterns and get better grades. For that purpose, various studies consider the prediction of students' final grades based on their term assessments; as in [8] who proposed a model to be applied to 136 students LMS activity. The researchers implemented RF; BART; XGBoost; Principal Components Regression (PCR); SVM; NN; Multivariate Adaptive Regression Splines; and K-NN. They used the final percentage grade as the main variable (predictor). To compare the predicted grade and actual grade, mean absolute error (MAE) was used; and they reported that PCR achieved the lowest MAE value. They found the optimal time to expect at-risk students that is week 5/6. With the same objective, [9]

proposed a model based on an RNN. They experimented with 937 students from six courses over eight weeks. Therefore, the predicted model is evaluated for each week separately. The result showed that the model achieved good accuracy of up to 91% in week eight.

Many studies focus on predicting students' approaching the risk zone by relying on their LMS data. For instance, [10] identified at-risk students to find the best combination of algorithms and datasets. Thirteen datasets which consisted of 89 students from the Moodle LMS were used. The classification algorithm used is K-NN, Multilayer Perceptron (MLP), Naive Bayes, AdaBoost, and RF. As a result, they found that the combination between the AdaBoost with DB2 and DB5 have better performance but nearly close to other combinations. Also, warning models that combine ML and DL algorithms were developed in [11] to identify at-risk students early. The dataset used for the model was 12,869 students collected from a K-12 virtual school in the USA. The algorithms used for the model were RF, SVM with the sigmoid kernel, SVM with Polynomial kernel, SVM with Gaussian radial basis function, and neural network. The model performed well by capturing 51% of at-risk students with 86% accuracy.

Some researchers assess the effect of different features on students' performance, such as in [12], where the researcher evaluated the influence of various characteristics on students' success using disposition analysis, which encompass the dispositions communicated by the student to the learning environment. The data contains 500 students collected from Kalboard 360 by a learner behaviour tracker tool named experience API (xAPI) that tracks learning progress and learner action. Then the data classified into Demographic features, Academic features, Parental Involvement features, and Interactional features. And using K-means, they found that Students' Dispositions Analytics are responsible for the performance of students and these Dispositions are dependent on Parental Involvement and Demographic features. In the same line, [13] authors assess the behavioural features' effect to enhance students' performance. They used 500 student's dataset gathered from LMS using xAPI. Their model use ID3, Naive Bayes, K-NN, SVM. The algorithms implemented on Waikato Environment for Knowledge Analysis (WEKA); ID3 works better than other methods. The proposed model achieved 90% accuracy.

Other studies looked into studying the effect of various features besides their behaviours in LMS. As in [7] the authors investigated the essential features that impact students' performance. The dataset used in this research was 241 undergraduate student's records from different courses on Blackboard. The study utilized five algorithms; decision tree (J48), RF, Sequential Minimal Optimization (SMO), MLP, and logistic regression (Logistic) in the WEKA. The researchers compared the algorithms' performance and concluded that the assessment grades are the useful features; RF achieves 99% accuracy.

With respect to DL methods, The researchers in [14] developed an ANN for student final performance prediction. The dataset used in this research was 3518 students. The study achieves an accuracy of 80%. Besides, [15] produced a model to predict students' performance in a computer science subject at Al-Muthanna University (MU), which was tested on data representing the user behaviour of 161 students. Authors were applying ANN, Naïve Bayes, Decision Tree, and Logistic Regression. The proposed model achieved 77% accuracy. In the same line, the researchers in [16] produced a method to predict students' results, which was tested on data representing the user behaviour of 10140 students and implemented the RNN algorithm. The result showed the effectiveness of RNN by achieving 95% accuracy.

The early evaluation of students' performance enables them to improve their learning strategies as well as evaluate the features that could impact their performance. Many studies focus on evaluation, which targets learners. For example, [17] predicted a mechanism to warn students who have poor performance based on cognitive and non-cognitive features to minimize students' dropout. The dataset used in this research was 128 students collected from different universities and 650 students from an online repository. Authors applied logistic regression, decision tree, Naïve Bayes, and neural networks. The result proved that cognitive features improve the accuracy of algorithms. Besides, [18] evaluated students' performance by proposing a Hybrid Educational Data Mining (HEDM) model. The model combines the effectiveness of Naive Bayes and the J48 Classifier classification technique. The model was evaluated on an online dataset and achieve 98% accuracy.

The previously cited researches have studied the effect of various features on students' academic performance and final performance prediction, but few researches have focused on predicting at-risk students with various methods for data extraction, size, and processing in the context of predicting or forecasting learners' achievement. However, few of these cited papers use advanced DL techniques, whereas the rest focus on using traditional ML techniques. The present study utilizes Deep Learning technique to develop CNN, encoder/decoder LSTM, and GRU models capable to predict at-risk students. Finally, the results of the developed prediction model are compared and evaluated.

III. RESEARCH METHODOLOGY

This study's methodology consists of four key phases: data collection, data pre-processing, development of prediction models, and evaluation (see Fig. 1).

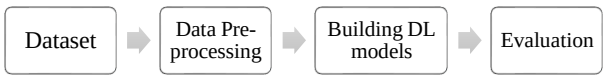

Fig. 1 Methodology Pipeline

A. Dataset

The dataset used in this study is retrieved via a collaboration with the Deanship of E-Learning and Distance Education at KAU using the official Learning Analytics system, A4L which supports the Blackboard LMS. This service facilitates the extraction of educational big data; specifically, in our case, detailed reports on students' interaction with two types of assessments and their final grade in the course. The dataset is comprised of the activity data from four-module English Foundation Course (ELIA

101, ELIA 102, ELIA 103, ELIA 104), which were delivered during the second semester of 2020. Ideally, this course is delivered at least twice a year. The dataset is basically time series made up of students' assessments (Assignment / Test) matched up with a value indicating the level of their activity inside the course assessments, and a final course grade that is indicative of their passing or failing the Final Exam. Out of the 3929 unique student IDs, only (11) failed the course. However, and according to the institutional Warning System data, 543 students were either warned once or twice of being academically dismissed from the program.

The structure of the extracted data consists of eight columns (Fig. 1): Date, (student)ID, Course, Item Type (Assessment type: assignment or test), Academic Standing (1st Warning, 2nd Warning, 3rd Warning, No Warning, and Dismissed), Average Grade (corresponding Final Course Grade), and Item Interaction (a count of students' interactions with assessments by type).

Date	ID	Course	Item_Type	Academic_Standing	Avg_Grade	Item_Interactions
1/19/2021	2126112	1	1	1	67	13
1/19/2021	2137032	1	1	1	71	1
1/19/2021	2143033	1	1	2	74	6
1/19/2021	2109035	1	1	4	60	7
1/19/2021	2115094	1	1	4	60	3

Fig 1 Dataset Structure

The following inclusion and exclusion criterion for dataset creation have been observed:

1) Inclusion Criterion:
   - The extracted reports included features (measures) representative of potential risk factors.
   - Comparability is ensured through extracting data about a definable student cohort (Foundation-year students enrolled in the 4-level English Course at King Abdul-Aziz University. The courses are: ELIA 101, 102, 103, 104).
   - The data was extracted from both course and students' perspectives for 3929 students.
   - The researchers verified that the courses included actual activities so that the extracted reports reflect actual user behaviour.
   - The time series data was extracted for the second semester of 2020.
2) Exclusion Criterion:
   - Measures which relate to logins were excluded, because they display data representative of all the courses a student is enrolled in, not just the English courses.
   - Measures which perform complex statistical operations on the data (change rate, moving averages) were as well not considered.

B. Data Pre-processing

Data cleaning and data transformation are two forms of data pre-processing techniques that are used to improve the quality of the data. This experiment ensured that there were no nosiness and inconsistency of the data. Three of the dataset features (Course, Item Type, and Academic Standing) were categorial in nature, so we defined an encoding function for these. While retaining both Date and Item Interactions Attributes, XGBoost was used to decide feature importance, eventually it retained: Item Type_1 (Assignment), Course_1, Course_2, Course_3, Course_4, ID, Academic Standing_2 (2nd Warning) and Average Grade. Finally, this multivariate timeseries dataset was split into temporal folds for training, and testing. Adopting the Horizon Style, our training set consists of observations that precede the observation that forms the test set.

C. Building the deep learning models

The main objective of this research was to develop deep learning models to predict the presence of at-risk students (Pass/Fail) depending on their LMS interaction with the assessments that made up the final grade. As we are considering a dataset with multiple predicators, we opted for developing and comparing the performance of models that are capable of prediction based on multiple indicators. To achieve these objectives, we experimented with three deep learning models (CNN, Encoder Decoder LSTM, and GRU), illustrated in Fig 2.

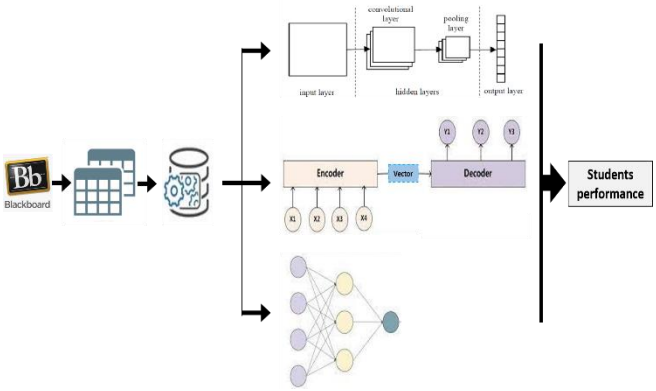

Fig. 2 General Proposed Models' Architecture

All deep learning models were trained and tested using Python 3 and Tensorflow 2.6.0. The models' hyperparameters were set according to the requirements of each model.

1) CNN model

Traditionally, CNNs models are developed for two-dimensional image data problems, however, they can be used as well to model multivariate time series prediction ones. The CNN model was constructed of Convolution 1D layer network with 64 filters and 3 kernel size, pooling 1D layer with 2 size, dropout layer with 0.2, flatten layer, dense hidden layer with 30 neurons and relu activation, dropout layer with 0.2, and the output layer with 10 neurons. The model training stopped at 13 epochs, as early stopping was used to avoid overfitting.

.

2) Encoder/Decoder LSTM model

The LSTM model was constructed of the input layer with 40 neurons, hidden layer with 20 neurons. Then, the Repeat Vector layer was applied. After that, two hidden layers with (40,25) neurons, respectively. Finally, fully connected time distributed layer with one neurons were applied. As early stopping was used, model training stopped at 11 epochs.

3) GRU model

The GRU was constructed with 3 layers; the input layer has 100 neurons and applied a 0.2 dropout layer, one hidden layer with 50 neurons and applied a 0.2 dropout layer, and the output layer have 10 neurons. The model training stopped at 76 epochs.

D. Evaluation

To assess the prediction models' performance, we used the following metrics:

- Mean Absolute Error (MAE): is a metric that measure the average of errors between the predictions; it is computed based on the following equation.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad (1)$$

- Mean Squared Error (MAE): is a metric that measure the average of errors between the true and predictions values; it is computed based on the following equation.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \quad (2)$$

- Root Mean Squared Error (RMSE): is a square root for MSE; it is computed based on the following equation.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} \quad (3)$$

IV. Results

In order to assess the proposed models (CNN, Encoder/Decoder LSTM, GRU), the main regression metrics were used for evaluation purposes. The results achieved by the metrics are as follows. The MAE achieved 2.375, 2.370, 2.602; MSE's values were 9.975, 10.244, 13.543; and the RMSE's were 3.158, 3.200, 3.680, respectively. Overall, these results indicate that the CNN model performed better than the other models in predicting at-risk students as its obtained the least loss in MSE and RMSE. Table 1 summarizes the comparison between the proposed models. Besides, Fig 3 (a, b, and c) show a comparison between the training and validation loss for each model. Finally, Fig 4 (a, b, and c) provide a comparison between the actual and predicted data for each model.

Table 1 Results comparison

Model	MAE	MSE	RMSE
CNN	2.3755	**9.97536**	**3.158381**
Encoder Decoder LSTM	**2.37011**	10.24498	3.200780
GRU	2.602066	13.54373	3.680181

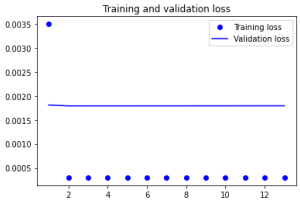

Fig 3 a) CNN Training and Validation Loss

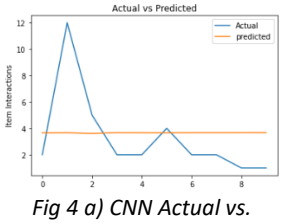

Fig 4 a) CNN Actual vs. Predicted

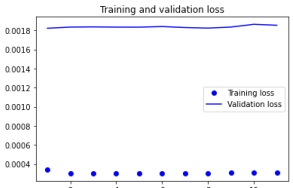

Fig 3 b) Encoder-Decoder LSTM Training and Validation Loss

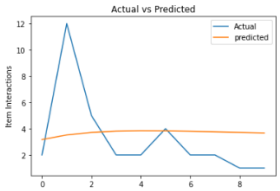

Fig 4 b) Encoder-Decoder LSTM Actual vs. Predicted

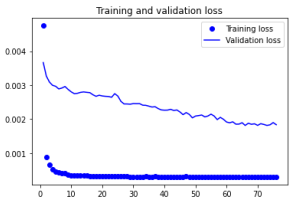

Fig 3 c) GRU Training and Validation Loss

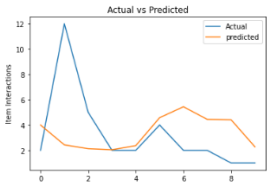

Fig 4 c) GRU Actual vs. Predicted

The results indicate that the CNN model performed better than the encoder/decoder LSTM, and GRU. Based on the findings, future studies can implement CNN model with more changes in the layers, and hyper-parameters. Moreover, attention to other evaluation metrics like R2 should be given its due as it's the closest metric to accuracy in the context of a regression problem.

However, certain limitations were faced during this project such as, the dataset was collected during Covid-19 lockdown where a heavy use of the Blackboard LMS was reported. Also, the used dataset for a limited student cohort which represents students enrolled in the English language Foundation course.

V. Conclusion and Future Work

In this study, we propose three neural networks models (CNN, encoder/decoder LSTM, and GRU) for predicting at-risk students based on a dataset of 3,929 students extracted from the A4L; KAU Blackboard system. The results show that the CNN performs better than other models because it achieved a smallest RMSE by 3.15. In the future, we plan to apply other deep learning models to achieve better performance; most importantly, adding other predicators of students' user behaviour inside the LMS and explore their relation to students' final achievement.

Acknowledgment

The researchers would like to thank Deanship of E-Learning and Distance Education at King Abdul-Aziz University to provide the dataset

REFERENCES

[1] A. Peña-Ayala, Ed., *Educational Data Mining*, vol. 524. Cham: Springer International Publishing, 2014.

[2] Lewis, R., & McCann, T. (2009). Teaching "at risk" students: Meeting their needs. In *International handbook of research on teachers and teaching* (pp. 895-905). Springer, Boston, MA.

[3] Allen, C. B., Higgs, Z. R., & Holloway, J. R. (1988). Identifying students at risk for academic difficulty. *Journal of Professional Nursing*, *4*(2), 113-118.

[4] P. Ibañez, C. Villalonga, and L. Nuere, 'Exploring Student Activity with Learning Analytics in the Digital Environments of the Nebrija University', *Technol. Knowl. Learn.*, vol. 25, no. 4, pp. 769–787, Dec. 2020, doi: 10.1007/s10758-019-09419-4.

[5] Y. He *et al*., 'Online At-Risk Student Identification using RNN-GRU Joint Neural Networks', *Information*, vol. 11, no. 10, p. 474, Oct. 2020, doi: 10.3390/info11100474.

[6] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, 'A Systematic Review of Deep Learning Approaches to Educational Data Mining', *Complexity*, May 12, 2019. https://www.hindawi.com/journals/complexity/2019/1306039/ (accessed Oct. 01, 2020).

[7] A. Alhassan, B. Zafar, and A. Mueen, 'Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data', *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 4, pp. 185–194, Apr. 2020.

[8] E. Howard, M. Meehan, and A. Parnell, 'Contrasting prediction methods for early warning systems at undergraduate level', *Internet High. Educ.*, vol. 37, pp. 66–75, Apr. 2018, doi: 10.1016/j.iheduc.2018.02.001.

[9] F. Okubo, T. Yamashita, A. Shimada, and S. Konomi, 'Students' Performance Prediction Using Data of Multiple Courses by Recurrent Neural Network', p. 6.

[10] L. A. Buschetto Macarini, C. Cechinel, M. F. Batista Machado, V. Faria Culmant Ramos, and R. Munoz, 'Predicting Students Success in Blended Learning—Evaluating Different Interactions Inside Learning Management Systems', Appl. Sci., vol. 9, no. 24, Art. no. 24, Jan. 2019, doi: 10.3390/app9245523.

[11] J.-L. Hung, K. Rice, J. Kepka, and J. Yang, 'Improving predictive power through deep learning analysis of K-12 online student behaviors and discussion board content', Inf. Discov. Deliv., May 2020, doi: 10.1108/IDD-02-2020-0019.

[12] S. Bharara, S. Sabitha, and A. Bansal, 'Application of learning analytics using clustering data Mining for Students' disposition analysis', Educ. Inf. Technol., vol. 23, no. 2, pp. 957–984, Mar. 2018, doi: 10.1007/s10639-017-9645-7.

[13] P. Kumari, P. K. Jain, and R. Pamula, 'An efficient use of ensemble methods to predict students academic performance', in 2018 4th International Conference on Recent Advances in Information Technology (RAIT), Mar. 2018, pp. 1–6, doi: 10.1109/RAIT.2018.8389056.

[14] Ş. Aydoğdu, 'Predicting student final performance using artificial neural networks in online learning environments', Educ. Inf. Technol., vol. 25, no. 3, pp. 1913–1927, May 2020, doi: 10.1007/s10639-019-10053-x.

[15] H. Altabrawee, O. Ali, and S. Qaisar, 'Predicting Students' Performance Using Machine Learning Techniques', J. Univ. BABYLON Pure Appl. Sci., vol. 27, pp. 194–205, Apr. 2019, doi: 10.29196/jubpas.v27i1.2108.

[16] S. Hussain, Z. F. Muhsion, Y. K. Salal, P. Theodoru, F. Kurtoğlu, and G. C. Hazarika, 'Prediction Model on Student Performance based on Internal Assessment using Deep Learning', Int. J. Emerg. Technol. Learn. IJET, vol. 14, no. 08, p. 4, Apr. 2019, doi: 10.3991/ijet.v14i08.10001.

[17] S. Sultana, S. Khan, and M. A. Abbas, 'Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts', Int. J. Electr. Eng. Educ., vol. 54, no. 2, pp. 105–118, Apr. 2017, doi: 10.1177/0020720916688484.

[18] V. G. Karthikeyan, P. Thangaraj, and S. Karthik, 'Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation', Soft Comput., Jun. 2020, doi: 10.1007/s00500-020-05075-4.