

An Improved Early Student's Performance Prediction Using Deep Learning

<https://doi.org/10.3991/ijet.v16i12.20699>

Nida Aslam^(✉), Irfan Ullah Khan, Leena H. Alamri, Ranim S. Almuslim
Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
naslam@iau.edu.sa

Abstract—Presently owed to technological revolution, a massive amount of data is generated in every field. Extracting the useful insights from consequential data is a very critical task. Moreover, advancement in the Deep Learning (DL) models ensued in the effective prediction of data. In the proposed study DL model is used for predicting the student's academic performance. Early prediction of the student's performance will reduce the risk of failure. The study used two courses data, i.e., mathematics and Portuguese language, containing demographic, socio-economical, educational and student's course grades data. The data set suffers from the imbalance; SMOTE (synthetic minority over-sampling technique) is used to overcome imbalance issue. The performance model is evaluated using several feature sets (all features, excluding G2 and G3) and evaluation measures such as precision, recall, F-score, and accuracy. The results showed the significance of the proposed DL model in early prediction of the students' academic performance. The model achieved an accuracy of 0.964 for Portuguese course data set and 0.932 for mathematics course data set. Similarly, the precision of 0.99 for Portuguese and 0.94 for mathematics is achieved. Furthermore, all the features are significant in predicting the student's academic performance.

Keywords—Deep Learning (DL), Educational Data mining (EDM), Early prediction, Academic Performance

1 Introduction

Education is one of the fundamental components in the development of the country. In the last decade, there is a progression in the educational field such as development of electronic educational applications, learning management system, automated assessment systems, and gamification in learning [1], [2], [3]. Several robots have been developed for teaching young children. These educational robots have greatly enhanced the learning and promote the activity-based learning. Recently, a survey has been made to explore the robots for kindergarten children [4] Similarly; numerous mobile educational applications have been developed for pre-school and kindergarten children, high school, and university students. Papadakis et al. [1] proposed a tool for the assessment of educational mobile applications for the kindergarten children under

the age of 3-6 using thirteen items. These electronic applications generate the paramount of data day by day. Due to the large size and complex nature, previously the data was unused. However, this data could aid in analyzing the quality of the education, early prediction of student's failure and dropout. Furthermore, it can also help policy makers and educational administrators to find the best strategies to enhance the quality of education. One of the challenges is to gain an insight of this huge amount of data for early prediction of student's performance and to explore the factors that led to the failure.

Student performance is an important part of the learning process. Predicting student performance is a key to identifying students who are more likely to have low academic performance in the future. This data could be helpful and will be utilized in prediction if that data has been processed into knowledge. Therefore, that knowledge could enhance teaching and learning quality and help students achieve their academic goals. Educational Data Mining (EDM) is a research field that applies data mining techniques to information generated from educational backgrounds [5]. Furthermore, applying EDM helps policymakers to plan how to improve student performance. Consequently, it will enhance the learning process and the students' experience in the educational institution. Renz et al. investigate the impact and integration of data driven and Artificial Intelligence models in education [9].

The emergence of the Machine Learning (ML), specifically the DL models have greatly enhanced the prediction performance in various domains such as health, business, agriculture, and many more [6][7][8]. This motivates the implication of DL in educational data as well. In the proposed study, we used DL model to predict the final grades of the students. This paper intends the early prediction of the students' grades and to find the most significant factors that affect the students' results. These features will help to find ways to improve the educational system in serving the students get better academic results and enhance the overall results. In the proposed study, two courses data, i.e., Portuguese and mathematics course was used to predict the student's performance.

The main objectives of the study are:

- To propose an improved deep learning-based model for the student's early performance prediction.
- Review and compare the performance of the previous studies and the proposed model using Portuguese and mathematics course dataset.
- Find the correlation among student's academic performance and various demographic, educational, and socio-economical features.

In this paper, section 2 covers the review of some related studies. Section 3 explains the proposed methodology, containing the description of the dataset, DL model, and the evaluation measures used to evaluate the performance of the model. Section 4 explains the experimental setup and results. Finally, conclusion is presented in section 5.

2 Related Studies

Several studies address student's academic performance prediction using ML models in various settings. The below section attempts to explore some of the previous work related to predicting student's academic performance. The studies were arranged in chronological order.

Cortez and A. Silva [10] presented a study that aimed to predict students' performance and identify the key variables that impact educational success/failure. Using two Portuguese secondary school's data: mathematics and Portuguese language course. The data modeled under binary / five-level classification and regression algorithms such as Decision Trees (DT), Random Forest (RF), Artificial Neural Networks (ANN) and Support Vector Machine (SVM). The results showed that providing the first and/or second school period grades can achieve good predictive accuracy for the final grade. Similarly, DT classifier was used to predict the student's performance using the dataset from agriculture college [11]. The model generated twenty-seven rules using 49 features. The model achieved an accuracy of 0.84, however the size of the dataset is very small.

Furthermore, another study was made on student's performance prediction using two data sources on introductory programming courses [12]. The first data set covers remote education student, and the second on-campus students' data. Furthermore, the study explores the impact of preprocessing and optimization techniques on classifiers performance such as SVM, J48, ANN, and Naïve Bayes (NB), respectively. The results showed that it is possible to early predict the students that likely to fail and SVM outperformed the other classifiers.

Yang and Li [13] proposed student attribute descriptors, performance estimator, attribute causal association predictor, and progress indicators. Fuzzy C-means and soft-margin was used for clustering the data set from the students' achievement management system of Huaqiao University. The results showed that students with good performance have a higher potential of achieving higher density on high grades and vice versa.

Lui study used NN-backpropagation (BP) NN-feedforward recurrent models. The study aimed to explore the impacting features for the English course score using the college students' data [14]. The data set contains 101 freshmen's NCEE (National College Entrance Examination) score. The results indicate that the most significant for the student academic performance in the first semester final grade is the NCEE marks, the learning attitude, and the gender. In contrast, age has a small effect on the scores. After training the model 20 times, the relative error is smaller, and the accuracy is higher.

Similarly, another study was performed to measure the association between the cognitive admission entry requirements and academic performance in their first year [15]. The data set was 1445 undergraduates between 2005-2009. The study used several classifiers such as RF, Tree Ensemble, DT, NB, Logistic Regression (LR), and Resilient backpropagation (Rprop) Multi-Layer Perceptron (MLP), respectively using student CGPA (Cumulative Grade Point Average) and class of degree. Experiments were performed using KNIME and Orange data mining tools, achieved maximum accuracies of 0.502 and 0.519. Additionally, using regression models the study achieved R^2 of 0.207 and 0.232. The study found that the first-year performance is not

defined by cognitive entry requirements. Furthermore, [16] proposed a study using models such as J48, Non-Nested Generalization exemplar classifiers (NNge), and MLP to predict the students' performance. The data set is academic session 2005-2006 of two schools of Alentejo region, Portugal and indicate that J48 outperformed the other model with an accuracy of 0.957.

Though, another study was performed to predict the engineering student's performance in the final graduation CGPA using the GPA for the first three years [17]. The data set contains 1841 students' academic performance within 2002-2014 across seven engineering programs from admission until graduation. Probabilistic Neural Network (PNN), RF, DT, NB, tree Ensemble, and LR classifiers were used. The results showed that LR achieved maximum accuracy of 0.891. The result was further validated using regression models and achieved R^2 of 0.955 and 0.957.

Furthermore, a study used classifiers such as J48, SVM, RF, NB, and MLP to predict students' performance levels using the data set of Kalboard 360 [18]. The result showed that the MLP outperformed other classifiers with 0.760 accuracy. Similarly, another study [19] was performed using the same data set, applying three classifiers such as DT, NB, and ANN. The study achieved 10 to 15% increase in the performance prediction when compared with results after removing some features. They focused on students' preparation and the impact of parents' participation in the learning process.

Additionally, a study used two data sets i.e., the academic year (2016-2017) over a 12-week semester for 240 students, and the other one, consisting of students' assessment scores for predicting the performance [20]. Several algorithms were used such as NB, RF, K-nearest neighbor (KNN), and LR, respectively. The study found that the RF was more accurate in a statistically significant way when compared with other algorithms.

Zohair's study [21] explore the possibility of training and modeling a small size data set and achieve a trusted accuracy rate. Additionally, finding the possibility of identifying the small data set's key indicators to create the prediction model. They used visualization and clustering techniques such as MLP-ANN, NB, SVM, KNN, and Linear Discriminant Analysis (LDA), respectively using 50 Records of graduated students. The results showed that the SVM model (with the radial kernel) achieved the highest accuracy rate. Recently, [22] RF and SVM was used to predict the student's final grades using mathematics course and Portuguese language course grades. They achieved the highest accuracy of 0.943 using RF for Portuguese data set and 0.924 for mathematics data set using SVM with all the features.

Doleck et al. [23] has recently made a comparative analysis of DL model and several conventional ML classifiers such as SVM, KNN, NB and LR, respectively. The study found that performance of the model depends on the data set used. There were two data sets used in the research: the MOOC data set consists of ten video-viewing features and outcome variables. The other data set was the CEGEP academic performance consisting of information about students' age, gender, prior academic performance, and the outcome variable. Additionally, the study made a comparison among the DL libraries such as Keras, TensorFlow, Pytorch, Theano and fast.ai, respectively. The results showed that on the first data set Keras (using TensorFlow as backend), and on the second data set TensorFlow (BoostedTrees) achieved the best performance. DL showed comparable performance to other ML algorithms.

Recently, a study has been made to investigate the performance of students in online courses using machine learning [24]. The study aimed to find the relationship among the student’s performance and activity logs. Additionally, they proposed a strategy for enhancing the student’s performance. The data is collected using Moodle platform of 107 students that were enrolled in 9-week online course and features related to course contents, assessment, grades, and forum were used. DT and LR models were used for training. Experiments were performed on original and SMOTE data. The study found that the student’s interaction has a high correlation with the student’s final grades. Similarly, the inclusion of the log files also enhanced the early prediction of the student’s grades.

Despite of several studies already been made in predicting the students’ academic performance but there is still a room of improvement. The proposed study is an extension of our previous study [22]. In the previous study we explore the significance of ML models in predicting the students’ performance using two open-source data sets. The data sets contain the student’s grades in two courses i.e., mathematics and Portuguese language course. The study achieved the result of 0.943 accuracy of Portuguese data set and 0.924 for mathematics. The results need to be further enhanced using the deep learning model. To achieve this the current study used DL model for predicting the students’ performance using both the data sets.

3 Materials and Methods

This section will cover the methodology used in the study to implement the proposed system. Figure 1 contains the steps of the proposed methodology. This section initially explains the description of the two data sets used in the study. It also covers the correlation and ranking of the features in the data sets using boruta feature selection. Boruta feature selection algorithm uses RF classifier for finding the association among the features and the target. Following the description of the deep learning model and finally the discussion of the evaluation matrix to compute the performance of the proposed model.

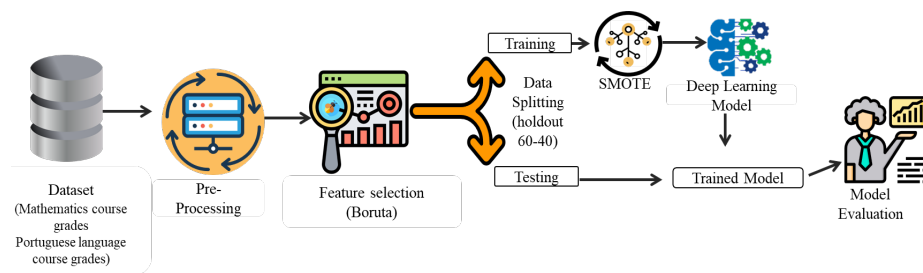


Fig. 1. Pipeline of the proposed methodology.

3.1 Data set description

The study used two data sets mathematics and Portuguese language course. The mathematics course data set contains the 395 records while Portuguese language

course contains 649 records [10]. The total number of attributes in the both the data sets are 32. Furthermore, the data set also contains three grades i.e., Grade 1, Grade 2 and Grade 3. Grade 1 represents the first period grade, grade 2 represents the second period grade, while grade 3 represents the final grade of the student. The student is declared as successful if the total grade of Grade 3 is equal or greater than 10. Table 1 contains the statistical description of data set. For the numeric features mean (μ) and standard deviation (σ), while count is used for categorical features. In the table below C represents categorical data and N represents numeric datatype. The data set contains demographic, family structure and status, lifestyle, educational information, and student academic performance, respectively. The class label contains binary class i.e., pass or fail. Figure 2 contains the number of records per class label for both the data sets. Both the data sets suffer from data imbalance. SMOTE was applied to overcome the imbalance.

Table 1. Description of the data set using G3 class label.

Feature Type	Features Name	Data Type	Mathematics(n=395)		Portuguese(n=649)	
			Pass(n=265)	Fail(n=130)	Pass(n=549)	Fail(n=100)
Demographic	School	C	GP (236) MS (29)	GP (113) MS (17)	GP (391) MS (158)	GP (32) MS (68)
	Gender	C	F (133) M (132)	F (75) M (55)	F (333) M (216)	F (50) M(50)
	Age	N	16.53 ± 1.19	17.02 ± 1.39	16.69 ± 1.20	17.06 ± 1.28
	Address	C	U (210) R (55)	U (97) R (33)	U (396) R (153)	U (56) R(44)
	Guardian	C	M (184) F (65) O (16)	M (89) F (25) O (16)	M (380) F (137) O (32)	M (75) F (16) O (9)
Family status	Famsize	C	GT3(185) LE3(80)	GT3(96) LE3(34)	GT3(381) LE3(168)	GT3(76) LE3(24)
	Pstatus	C	T (235) A (30)	T (119) A (11)	T (481) A (68)	T (88) A (12)
	Medu	C	0(2), 1(34) 2(67), 3(64) 4(98)	0(1), 1(25) 2(36), 3(35) 4(33)	0(6), 1(106) 2(159), 3(117) 4(161)	0(0), 1(37) 2(27), 3(22) 4(14)
	Fedu	C	0(2), 1(47) 2(77), 3(67) 4(72)	0(0), 1(35) 2(38), 3(33) 4(24)	0(7), 1(129) 2(178), 3(121) 4(114)	0(0), 1(45) 2(31), 3(10) 4(14)
	Mjob	C	Athome (37) Health (27) Other (88) Services (74) Teacher (39)	Athome (22) Health (7) Other (53) Services (29) Teacher (19)	Athome (106) Health (42) Other (216) Services (117) Teacher (68)	Athome (29) health (6) other (42) services (19) teacher (4)
	Fjob	C	Athome (12) Health (12) Other (148) Services (71) Teacher (22)	Athome (8) Health (6) Other (69) Services (40) Teacher (7)	Athome (35) Health (20) Other (313) Services (148) Teacher (33)	Athome (7) Health (3) Other (54) Services (33) Teacher (3)
	Famrel	C	1(6)-2(11)- 3(42)-4(131)- 5(75)	1(2)-2(7)- 3(26)-4(64)- 5(31)	1(13)-2(24)- 3(85)-4(383)- 5(144)	1(9)-2(5)- 3(16)-4(34)- 5(36)

	Famsup	C	Y (157) N (108)	Y (85) N (45)	Y (208) N (341)	Y (57) N (43)
Educational Information	Reason	C	Course (89) Home (73) Reputation (76) Other (27)	Course (56) Home (36) Reputation (29) Other (9)	Course (230) Home (132) Reputation (133) Other (54)	Course (55) Home (17) reputation (10) other (18)
	Failure	C	0(234)-1(24)-2(3)-3(4)	0(78)-1(26)-2(14)-3(12)	0(498)-1(38)-2(8)-3(5)	0(51)-1(32)-2(8)-3(9)
	Traveltime	C	1(177)-2(68)-3(15)-4(5)	1(80)-2(39)-3(8)-4(3)	1(366)-2(213)-3(54)-4(16)	1(51)-2(34)-3(12)-4(3)
	Studytime	C	1(68)-2(128)-3(49)-4(20)	1(37)-2(70)-3(16)-4(7)	1(315)-2(179)-3(42)-4(13)	1(50)-2(41)-3(7)-4(2)
	Schoolsup	C	Y (28) N (237)	Y (23) N (107)	Y (60) N (489)	Y (8) N (92)
	Paid	C	Y (130) N (135)	Y (51) N (79)	Y (30) N (519)	Y (9) N (91)
	Activities	C	Y (136) N (129)	Y (65) N (65)	Y (272) N (277)	Y (43) N (57)
	Nursery	C	Y (210) N (55)	Y (104) N (26)	Y (365) N (87)	Y (81) N (19)
	Higher	C	Y (258) N (7)	Y (117) N (13)	Y (440) N (109)	Y (67) N (33)
	Absence	N	5.19 ± 6.41	6.77 ± 10.58	3.49 ± 4.45	4.61 ± 5.50
Life Style	Internet	C	Y (225) N (40)	Y (104) N (26)	Y (430) N (119)	Y (68) N (32)
	Romantic	C	Y (80) N (185)	Y (52) N (78)	Y (193) N (356)	Y (46) N (54)
	Freetime	C	1(12)-2(49) 3(101)-4(74) 5(29)	1(7)-2(15) 3(56)-4(41) 5(11)	1(38)-2(95) 3(218)-4(149) 5(49)	1(7)-2(12) 3(33)-4(29) 5(19)
	Goout	C	1(17)-2(79) 3(93)-4(48) 5(28)	1(6)-2(24) 3(37)-4(38) 5(25)	1(36)-2(129) 3(182)-4(120) 5(82)	1(12)-2(16) 3(23)-4(21) 5(28)
	Dalc	C	1(193)-2(44) 3(17)-4(4) 5(7)	1(83)-2(31) 3(9)-4(5) 5(2)	1(393)-2(97) 3(36)-4(12) 5(11)	1(58)-2(24) 3(7)-4(5) 5(6)
	Walc	C	1(101)-2(60) 3(55)-4(31) 5(18)	1(50)-2(25) 3(25)-4(20) 5(10)	1(219)-2(219) 3(97)-4(71) 5(33)	1(28)-2(21) 3(23)-4(16) 5(12)
	Health	C	1(37)-2(30) 3(60)-4(42) 5(96)	1(10)-2(15) 3(31)-4(24) 5(50)	1(73)-2(69) 3(106)-4(96) 5(205)	1(17)-2(9) 3(18)-4(12) 5(44)
Academic Performance	Grade1	N	12.45 ± 2.78	7.76 ± 1.70	12.06 ± 2.39	7.78 ± 1.45
	Grade2	N	12.62 ± 2.53	6.82 ± 2.71	12.31 ± 2.34	7.53 ± 2.42
	Grade3	N	12.88 ± 2.50	5.39 ± 3.65	12.82 ± 2.28	6.89 ± 3.10

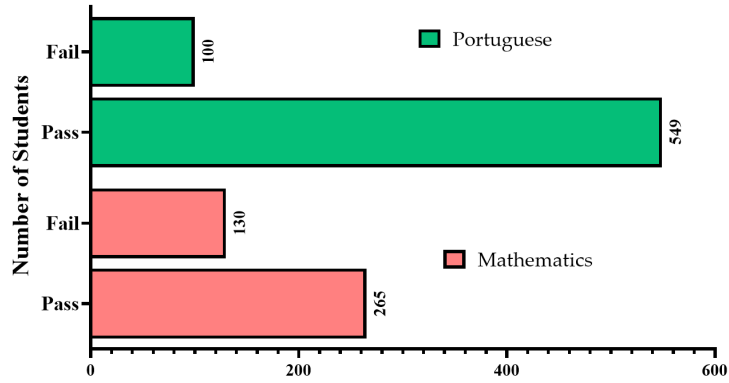


Fig. 2. Number of records per class in data sets using G3 as class label.

Figure 3 contains the correlation of the features with target (G3) for mathematics course data set. While Figure 4 represents the feature significance for Portuguese language course data set. The figures below present top 20 features in the data sets. The academic grades i.e., Grade 1 and Grade 2 has high correlation when compared with other features for mathematics course data set. However, for the Portuguese language course data set the academic grades are not amongst the top 20 features. Failure is the most significant feature for Portuguese language course data set.

The correlation among the attributes in the data set is shown in the Figure 5 and Figure 6. Figure 5 represents the heatmap for mathematics course grades, while Figure 6 contains the heatmap for Portuguese language course grades.

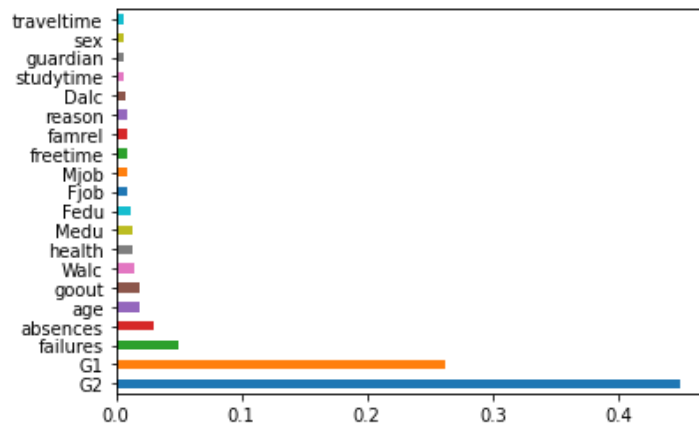


Fig. 3. Features Correlation in mathematics course grades data sets

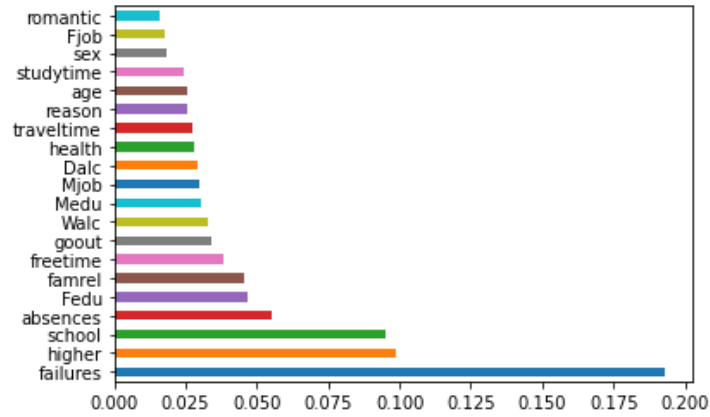


Fig. 4. Features Correlation in Portuguese language course grades data sets

Please Update the Title

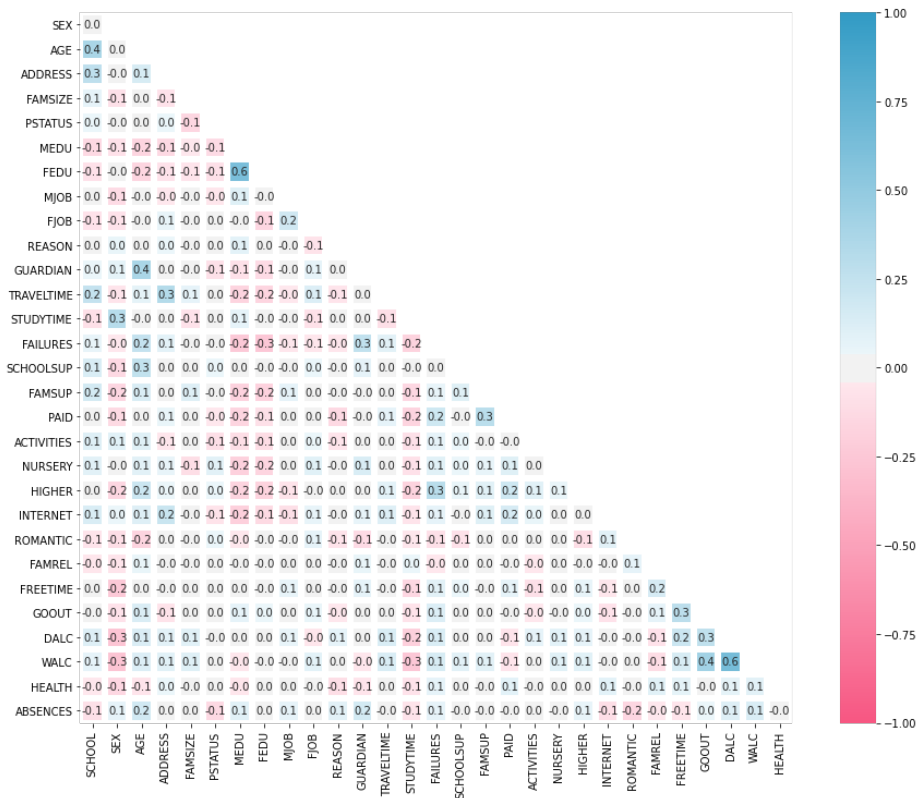


Fig. 5. Attributes correlation heatmap for mathematics course grades

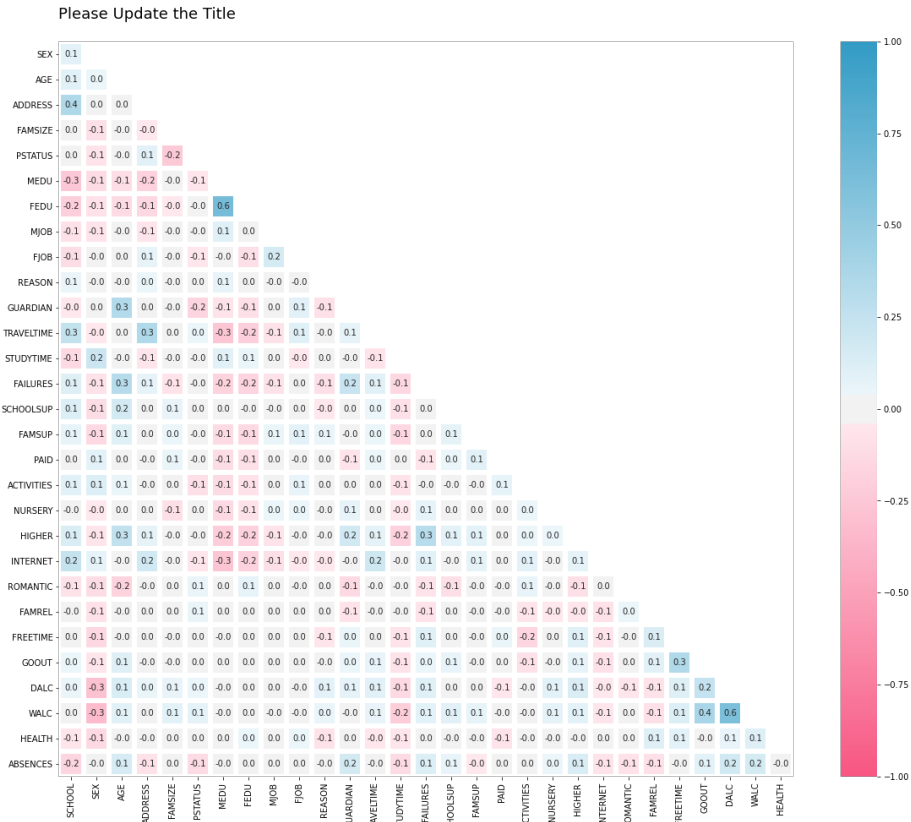


Fig. 6. Attributes correlation heatmap for Portuguese language course grades

3.2 Deep learning model

Neural networks (NN) are human brain inspired classification, regression, and clustering algorithm. NN can be utilized to solve complex and ambiguous problems using parallel processing. The composition of nodes in NN vary amongst algorithms such as feedforward and backward (sequential or convolution). These nodes interconnected with each other to exchange information, simulating the manner our neurons correlate via dendrites and axons for information exchange. Due to the promising results of the DL in many studies, it has been also applied in a broad field of EDM by researchers [25]. In our study we developed a DL model, section below discusses the details of the deep learning model used.

Our proposed DL model used a dense model and after rigorous experiments, DL model with eight hidden layers is implemented, where each layer consisting of 512, 256, 256, 256(dropout_layer), 64, 64, 64(dropout_layer) and 1(output_layer) neurons, respectively. The two dropout layers were added to force the DL network to learn more robust features. Moreover, the input layers and all the hidden layers used the

Rectified Linear Unit “ReLU” activation function while the output layer used ‘sigmoid’ as an activation function. Adam algorithm is used for model optimization, while “binary_crossentropy” is employed to calculate the loss of the model. The accuracy metrics is used to evaluate the accuracy of the model. For training model, the epochs set to 300, batch size to 128 along with additional setting of callbacks where the monitor is set to “Validation accuracy”, save best only to “True” and mode to “max”. After training the model, we create another model and reload the best saved weights to the new model and then perform the prediction on the new model. The structure of the deep learning model is mentioned in the Table 2. The result of the proposed DL model will be discussed in section 4.

Table 2. Summary of the DL model.

Layers (type)	Output Shape	Param#
Input Layer	(None, 512)	15872
Hidden Layer # 1 (Dense)	(None, 256)	131328
Hidden Layer # 2 (Dense)	(None, 256)	65792
Hidden Layer # 3 (Dropout)	(None, 256)	0
Hidden Layer # 4 (Dense)	(None, 64)	16448
Hidden Layer # 5 (Dense)	(None, 64)	4160
Hidden Layer # 6 (Dropout)	(None, 64)	0
Output Layer	(None, 1)	65

3.3 Evaluation matrix

The performance of the proposed model is evaluated using several standard evaluation measures such as precisions, recall, accuracy, and F-score. Precision represents the ratio of correctly predicted students’ grade in total students’ grades samples. Recall represents ratio of correctly predicted records in the data set. Accuracy represents the ratio of correctly identified grades from the total number of student’s grades in the data set. However, F-score indicates the harmonic mean of the precision and recall.

4 Experiments and Results

The model was implemented using python 3.8.5 version using Jupyter notebook 6.0.3. Several libraries are used such as sklearn, matplotlib and Keras. The data set was partition into training and testing using holdout method with the ratio of 60-40. The data sets suffer from data imbalance and which increases the risk of model overfitting. To overcome this problem, SMOTE was applied. Experiments were performed using original and SMOTE data, and using four feature sets such as all features, excluding G1, excluding G2, excluding G1 and G2, respectively. The results showed that the model produced the best outcome using all the features for both the data sets. For the Portuguese language course data set the highest accuracy was 0.964, precision 0.99, recall 0.933 and F-score of 0.962. However, for the mathematics course data set the highest accuracy was 0.93, precision 0.949, recall 0.903 and F-score 0.926, respec-

tively. The outcome of the model for the Portuguese language data set is shown in the Table 3. However, Table 4 represents the performance of the model for the mathematics course data set. To compare the performance of the proposed model and the other studies using both the data sets, benchmarking was performed. Table 5 contains the results achieved by the proposed model and the previous studies.

Table 3. Performance evaluation of the DL model using original and SMOTED Portuguese language course grades data set

Data set	SMOTE	Features	Accuracy	Precision	Recall	F-Score
Portuguese	With	Excluding G1	0.96	0.976	0.941	0.958
		Excluding G2	0.964	0.984	0.941	0.962
		Excluding G1&G2	0.927	0.946	0.903	0.924
		All-Features	0.964	0.992	0.933	0.962
	Without	Excluding G1	0.914	0.937	0.964	0.951
		Excluding G2	0.902	0.956	0.929	0.942
		Excluding G1&G2	0.878	0.876	1.0	0.934
		All-Features	0.914	0.963	0.936	0.949

Table 4. Performance evaluation of the DL model using original and SMOTED Mathematics course grades data set

Data set	SMOTE	Features	Accuracy	Precision	Recall	F-Score
Mathematics	With	Excluding G1	0.909	0.946	0.855	0.898
		Excluding G2	0.895	0.929	0.839	0.881
		Excluding G1&G2	0.805	0.773	0.826	0.797
		All-Features	0.932	0.949	0.903	0.926
	Without	Excluding G1	0.909	0.948	0.902	0.924
		Excluding G2	0.848	0.883	0.869	0.876
		Excluding G1&G2	0.718	0.747	0.82	0.781
		All-Features	0.909	0.919	0.934	0.927

Table 5. Comparison of the proposed model accuracy with the previous studies using Mathematics and Portuguese data set

Ref	Target	Mathematics data set		Portuguese data set	
		Technique	Accuracy	Technique	Accuracy
[10]	G3	SVM	0.706	RF	85
[22]	G3	RF	0.739	SVM, RF	89.74
Current Study	G3	DL	0.932	DL	96.4

As shown in Table 5 the proposed study outperformed the previous studies in the literature. Alamri et al. [22] used optimized RF without using SMOTE. They achieved better accuracy as compared to Cortez and Silva. [10] for both the data sets. In the previous two studies experiments were conducted using original data sets only with all the features. Data imbalance usually leads to model overfitting. Hence the results of both the studies might be due to model overfitting. However, for the proposed study the experiments were performed using original and SMOTED data to reduce the chance of model overfitting.

The study found that to predict the student's final academic performance all the features need to be considered. Some features in the Figure 3 and 4 shows low correlation with the target however, during the prediction it contribute significantly. Conclusively as per the study's objectives the proposed model produced better results when compared with the benchmark studies using similar data sets.

Despite of the outperformed performance of the proposed model there is still a room for the improvement. The proposed study suffers from some of limitations such as, the model needs to be further investigated on other data sets. Similarly, the data set suffers from the imbalance. Therefore, the model needs to be further validated using a large and balanced data set.

5 Conclusion

In the current study, we developed a dense DL model to predict the student's grades using the mathematics and Portuguese language course grades data set. The model serves as a tool for the early prediction of the student's grades. The study proved the significance of the DL model in EDM. The performance of the model was evaluated in terms of precision, recall, F-score, and accuracy. SMOTE was applied to reduce the risk of model overfitting. The proposed model outperformed when compared with the benchmark studies using the same data set. The study found that all the features contribute significantly to predicting the study academic performance. Furthermore, the deep learning models has greatly enhanced the prediction performance. One of the challenges faced is the data imbalance that sometimes leads to model overfitting. Despite of the several advantages, still there is a room for further improvement. The limitation of the current study is that the model is only validated using two open-source data sets and need to be further validated on other large size balanced data sets.

6 References

- [1] S. Papadakis, "Tools for evaluating educational apps for young children: a systematic review of the literature," 2020, doi: [10.1108/ITSE-08-2020-0127](https://doi.org/10.1108/ITSE-08-2020-0127).
- [2] M. S. A. El-seoud, N. Seddiek, M. M. El-khouly, and A. Nosseir, "E-Learning and Students' Motivation: A Research Study on the Effect of E-Learning on Higher Education," pp. 20–26, 2009. <https://doi.org/10.1109/icl.2013.6644685>
- [3] C. Hursen, "Use of Gamification Applications in Science Education," pp. 4–23.
- [4] S. Papadakis, "Robots and Robotics Kits for Early Childhood and First School Age," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 18, pp. 34–56, 2020, <https://doi.org/10.3991/ijim.v14i18.16631>
- [5] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007, doi: [10.1016/j.eswa.2006.04.005](https://doi.org/10.1016/j.eswa.2006.04.005).
- [6] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers," *Plants*, vol. 9, no. 10, pp. 1–17, 2020, <https://doi.org/10.3390/plants9101319>

- [7] F. Alshakhs, H. Alharthi, N. Aslam, I. U. Khan, and M. Elasheri, "Predicting postoperative length of stay for isolated coronary artery bypass graft patients using machine learning," *Int. J. Gen. Med.*, vol. 13, 2020, <https://doi.org/10.2147/ijgm.s250334>
- [8] I. U. Khan and N. Aslam, "A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images," *Inf.*, vol. 11, no. 9, 2020, <https://doi.org/10.3390/info11090419>
- [9] A. Renz, S. Krishnaraja, and E. Gronau, "Demystification of Artificial Intelligence in Education – How much AI is really in the Educational Technology?" *Int. J. Learn. Anal. Artif. Intell. Educ.*, vol. 2, no. 1, p. 14, 2020, <https://doi.org/10.3991/ijai.v2i1.12675>
- [10] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," 15th Eur. Concurr. Eng. Conf. 2008, ECEC 2008 - 5th Futur. Bus. Technol. Conf. FUBUTEC 2008, vol. 2003, no. 2000, pp. 5–12, 2008.
- [11] F. Berhanu and A. Abera, "Students' Performance Prediction based on their Academic Record," *Int. J. Comput. Appl.*, vol. 131, no. 5, pp. 27–35, 2015, <https://doi.org/10.5120/ijca2015907348>
- [12] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, "Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses," *Comput. Human Behav.*, vol. 73, pp. 247–256, 2017, <https://doi.org/10.1016/j.chb.2017.01.047>
- [13] F. Yang and F. W. B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Comput. Educ.*, vol. 123, no. October 2017, pp. 97–108, 2018, <https://doi.org/10.1016/j.compedu.2018.04.006>
- [14] W. Liu, "An improved back-propagation neural network for the prediction of college students' English performance," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 16, pp. 130–142, 2019, <https://doi.org/10.3991/ijet.v14i16.11187>
- [15] A. I. Adekitan and E. Noma-Osaghae, "Data mining approach to predicting the performance of first year student in a university using the admission requirements," *Educ. Inf. Technol.*, vol. 24, no. 2, pp. 1527–1543, 2019, <https://doi.org/10.1007/s10639-018-9839-7>
- [16] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 14, pp. 92–104, 2019, <https://doi.org/10.3991/ijet.v14i14.10310>
- [17] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, p. e01250, 2019, <https://doi.org/10.1016/j.heliyon.2019.e01250>
- [18] C. Jalota and R. Agrawal, "Analysis of Educational Data Mining using Classification," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Perspectives Prospect. Com.* 2019, pp. 243–247, 2019, <https://doi.org/10.1109/comitcon.2019.8862214>
- [19] S. Tuaha, I. F. Siddiqui, and Q. Ali Arain, "Analyzing Students' Academic Performance through Educational Data Mining," *3C Tecnol. innovación Apl. a la pyme*, no. May, pp. 402–421, 2019, <https://doi.org/10.17993/3ctecno.2019.specialissue2.402-421>
- [20] G. Ramaswami, T. Susnjak, A. Mathrani, J. Lim, and P. Garcia, "Using educational data mining techniques to increase the prediction accuracy of student academic performance," *Inf. Learn. Sci.*, vol. 120, no. 7–8, pp. 451–467, 2019, <https://doi.org/10.1108/ils-03-2019-0017>
- [21] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019, <https://doi.org/10.1186/s41239-019-0160-3>
- [22] L. H. Alamri, "Predicting Student Academic Performance using Support Vector Machine and Random Forest."

- [23] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazalais, “Predictive analytics in education: a comparison of deep learning frameworks,” *Educ. Inf. Technol.*, vol. 25, no. 3, pp. 1951–1963, 2020, <https://doi.org/10.1007/s10639-019-10068-4>
- [24] S. Azimi, C. G. Popa, and T. Cucić, “Improving students performance in small-scale online courses: A machine learning-based intervention,” *arXiv*, pp. 80–95, 2020, <https://doi.org/10.3991/ijai.v2i2.19371>
- [25] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, “A Systematic Review of Deep Learning Approaches to Educational Data Mining,” *Complexity*, vol. 2019, 2019, <https://doi.org/10.1155/2019/1306039>

7 Authors

Dr. Nida Aslam currently is an Assistant Professor in College of Computer Science and Information Technology at Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia. She received her Ph.D. degree from Middlesex University, London. Her research interests include Machine Learning, Data Mining, Image Processing, Computer Vision and specifically the application of AI in health and Education. Email: naslam@iau.edu.sa

Irfan Ullah Khan currently is an Assistant Professor in College of Computer Science and Information Technology at Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia. He received his Ph.D. degree from Middlesex University, London. His research interests include Machine Learning, Data Mining, Big Data processing, Image Processing, Computer Vision and specifically the application of AI in health and Education. Email: iurab@iau.edu.sa

Leena Alamri is a computer scientist. She received her bachelor’s degree in Computer Science from Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia in 2020. Her research interest focuses on data mining, educational data mining, and Artificial Intelligence. Email: 2160003034@iau.edu.sa

Ranim Almuslim is a computer scientist. She received her bachelor’s degree in Computer Science from Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia in 2020. Her research interest focuses on data mining, educational data mining, and information technology. Email: 2160006436@iau.edu.sa

Article submitted 2020-12-22. Resubmitted 2021-03-08. Final acceptance 2021-03-09. Final version published as submitted by the authors.