2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)

How Well a Student Performed? A Machine Learning Approach to Classify Students' Performance on Virtual Learning Environment

Fatema Alnassar
Department of Computing
Goldsmiths, University of London
London, UK
falna001@gold.ac.uk

Tim Blackwell
Department of Computing
Goldsmiths, University of London
London, UK
t.blackwell@gold.ac.uk

Elaheh Homayounvala
Department of Computing
Goldsmiths, University of London
London, UK
e.homayounvala@gold.ac.uk

Matthew Yee-king
Department of Computing
Goldsmiths, University of London
London, UK
n.yee-king@gold.ac.uk

2021 2nd International Conference on Intelligent Engineering and Management (ICIEM) | 978-1-6654-1450-0/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICIEM51511.2021.9445286

***Abstract*—Prediction of student's performance using different relevant information has emerged as an efficient tool in educational institutes towards improving the curriculum and teaching methodologies. Automated analysis of educational data using state of the art Machine Learning (ML) and Artificial Intelligence (AI) algorithms is an active area of research. The research addresses the problem of student performance prediction by using three ML algorithms (i.e., Support Vector Classifier (SVC), *k*-Nearest Neighbour (*k*-NN), Artificial Neural Network (ANN)) on Open University (OU) dataset. Educational data is analyzed for three main indicators including demographic, engagement and performance. From the experimental analysis, the *k*-NN approach emerged as best for OU experiments when compared among applied and with existing literature. Improvement of results is attributed to change in dealing with missing values and data standardization approaches.**

***Index Terms*—Virtual Learning Environment (VLE), Machine Learning (ML), Neural Networks (NN), Student Performance Classification**

I. INTRODUCTION

Significant technological advancements in computers and Internet has revolutionized the education system and introduced Virtual Learning Environment (VLE) as a new research dimension for researchers [1]. Jani et al. [2] argued that blended learning of face-to-face and using the VLE platform increased the student's understanding and performance. The recent COVID-19 pandemic has impacted in-campus attendance and almost all education institutions are shifted online thus increasing the importance of VLE [3]. One of the biggest challenges the educational sector encountered during online education was the effective and reliable evaluation of student's performance on VLE. The complexity of the process arises from cheating during the e-assessment by accessing information from the Internet, written materials, and other helpful stuff [4]. The prediction of student performance will help instructors at the initial stages of the course to care for students who need help by encouraging them to do the classwork [5].

Among others, Educational Data Mining (EDM) which is a process used to extract useful information and patterns from a huge educational database [6] has emerged as most effective approach. It involves the application of data mining (DM) techniques to data obtained from student use of VLEs, actions, and behaviors recorded in it [7]. Machine Learning (ML) classification and regression approaches have been found effective in predicting students performance based on the data collected through VLE. Effectiveness and accuracy of prediction is highly dependent on the data type of features being used, size of a dataset, and diversity in the dataset.

ML approaches including *k*-NN, SVC, ANN, Random Forest (RF), AdaBoost and Decision Tree (DT) have been used as key techniques predicting the performance of students based on regression and classification analysis on the VLE datasets. This paper explores the idea of investigating the effect of different features (i.e., Demographic, Engagement, Performance) on prediction performance using *k*-NN, SVC and ANN. Individual and combination of above features from the Open University (OU) dataset have been explored in this paper. Improved performance classification has been achieved in comparison to literature mainly based on the different data pre-processing (i.e., NaN value replacement and Standard Scalar Transformation).

II. RELATED WORK

This section presents the review of literature where AI and ML are used for the student's performance prediction based on learning material, teaching style and access patterns datasets. Literature is presented in chronological order to highlight the shift in trends over the years in this domain.
In 2015, Elbadrawy et al. [8] implemented a class of linear multi-regression models to predict the preforming of students using the educational data. Models made use of data features

including past performance, interaction with Learning Management System (LMS) and course related activities. Proposed models were validated on a custom collected dataset of 11,556 student entries, and 832 courses. From the results, Root Mean Square Error (RMSE) for the multi-regression models were reported as 0.147, improved from the single regression model. Next year, Yee-King et al. [9] proposed a *k*-NN based module to predict the student's grades from collaborative social learning. A multivariate classification approach was used to avoid the weak classification. The proposed approach was validated on the custom generated dataset from a virtual programming course at Coursera in 2014. The gathered dataset consisted of the total number of User Interface (UI) clicks and mouse-overs generated during the course. Authors were able to achieve the classification accuracy of 88%, 77% and 31% for 2, 3 and 10 grade bands, respectively.

In 2017, Al-Shehri et al. [10] predicted the performance of students in the final exam using *k*-NN and Support Vector Machine (SVM) ML approaches. A custom collected dataset from the University of Minho, Portugal with 395 data samples were used to validate the performance of ML models. Dataset consisted of student background and personal information attributes. From the analysis, relatively comparative results were reported for both approached with SVM slightly better (96% accuracy) than *k*-NN (95% accuracy). In the same year, Iqbal et al. [11] performed a comparative study of three different ML approaches including CF, Matrix Factorization (MF) and Restricted Boltzman Machines (RBM) towards predicting the grade of students. A custom collected dataset from International Technical University (ITU), Pakistan with 225 student entries were used for the validation of ML algorithms. Dataset consisted of performance based features including previous academic performance and interview score. From the results, the RBM approach was reported best among implemented three with an RMSE of 0.3.

In 2018, Hussain et al. [12] performed a comparative study to predict the student engagement and its impact on performance using several learning-based algorithms. Authors implemented DT, Classification, and Regression Tree (CART), JRIP Decision Rules, Gradient Boosting Trees (GBT) and Naive Bayes Classifier (NBC) on OU acquired dataset to predict the student engagement. Dataset of only July 2013 session (384 records) was used with demographic, performance, and learning behaviour features. The authors reported that J48 decision tree algorithm outclassed others with the highest ac- curacy of 88.52% and recall of 93.4%. Later in the same year, Heuer and Breiter [13] implemented several ML approaches to identify the at-risk students from their first assessments. The authors used standard OULAD Dataset with 32,593 student entries. Activity-based and performance features were used to predict the performance. Authors implemented SVM, NB, RF, XGBoost and Logistic Regression (LR) ML approaches. From the results, SVM outperformed other implemented algorithms with 87.98% accuracy.

In 2019, Sekeroglu et al. [14] investigated the student performance prediction and classification using a variety of ML algorithms. Authors used Long-Short Term Memory (LSTM), Back propagation (BP) and Support Vector Regression (SVR) for prediction while BP, SVM and Gradient Boosting Classifier (GBC) for classification. Student Performance Dataset (SPD) was used for prediction analysis while Students Academic Performance Dataset (SAPD) was used for classification analysis. Datasets mainly included student's demographic information, academic background history and behavioural pattern features. Authors reported SVR is the best algorithm for prediction while BP for the classification. Later, El Fouki et al. [15] proposed an advanced classification model established on deep learning and Principal Component Analysis (PCA) for the prediction of student's performance. The proposed multi-dimensional approach aimed at reducing the dimensions of data and extracting relevant information from the data to improve the model classification accuracy. A custom collected dataset with 496 records consisting of features including student's performance, section information and activity participation. Dataset was pre-processed using PCA for dimensionality reduction and then analysed using deep learning model, Multi-Layer Perceptron (MLP) and BayesNet. The authors reported the highest classification accuracy of 92.54% for the deep learning model. In the same year, Hussain et al. [16] proposed a module based on internal assessment using deep learning with Adam optimizer to predict student performance. In addition to the deep learning model, two other approaches including Artificial Immune Recognition System (AIRS) v2.0 and AdaBoost were also implemented for comparative investigation. The authors used custom collected dataset of 10,140 records from 3 different colleges in India. Performance of students in multiple tests was the main feature of the used dataset towards predicting the final grades. From the results, a deep learning model with binary cross-entropy loss and sigmoid activation was reported as best with a classification accuracy of 95.34%. Later in 2019, Ajibade et al. [17] implemented various classification algorithms on behavioural learning data of students to predict the performance. In addition, authors used Differential Evolution (DE) for behavioural feature selection. Proposed approaches were validated against the custom collected dataset with a record of 500 students. Dataset consisted of the demographic, academic, learning process, and behavioural learning features. DT, *k*-NN and SVM approaches were applied and DT was reported best among three but not with a huge margin. The authors also implemented ensemble of multiple models using bagging, boosting, and RF approaches towards improving the classification results. Classification accuracy was improved to 91.5% by using the ensemble approach.
Recently in 2020, Tomasevic et al. [18] performed a comparative study to investigate the effect of different features on student's assessment prediction using a variety

of ML and statistical approaches. Authors used SVM, *k*-NN, ANN, DT, Bayesian Linear Regression (BLR), and Regularized Linear Regression (RLR). The authors used a part of the OULAD dataset with demographic, engagement and performance features. F1 score and RMSE were used as performance measures for classification and regression models. Authors reported 96.62% F1 score for ANN using engagement and performance features, while 96.04% SVM (RBF kernel) using demographic, engagement and performance features. In the same year, Hooshyar et al. [19] proposed a novel approach PPP based on the procrastination behaviour of students to predict their performance. The proposed algorithm focused on student's assignment submission behaviour as the main indicator in predicting their performance. The authors validated the proposed approach for a custom collected dataset of 242 students from the University of Tartu, Estonia. Common ML approaches including Linear SVM (L-SVM), Radial SVM (R-SVM), DT, Gaussian Process (GP), RF, NN, AdaBoost, and NB were implemented on the dataset. From the results, NN was reported as best for categorical features with 96% accuracy while L- SVM was reported as overall best with 95% classification accuracy. In one of the most recent publications, Waheed et al. [20] proposed the use of Deep Neural Network (DNN) for predicting the students performance from the VLE big data. The authors used the OULAD open-source dataset consisted of 32,593 student records. Dataset features included demographics, clickstream behaviour and assessment performance. From the results, the authors reported that the proposed deep learning based approach outclassed conventional regression and SVM approaches with an accuracy of up to 93%.

Educational datasets with features of student's access patterns, availability of learning materials, different teaching styles, and students activities have been used to predict the students performance. Some standard datasets being used by researchers for performance prediction include OU, OULAD, SPD and SAPD. ML and AI have emerged as a key role in exploiting the educational datasets in comparison to conventional statistical approaches. A shift has been observed from literature from conventional ML approaches (i.e., SVR, DT, GBT) towards deep learning approaches (i.e., DNN, MLP, ANN, LSTM). However, availability of training datasets for deep learning approaches has been one of major shortcomings till date. F1-Score, accuracy, recall score and RMSE are reported to be commonly used evaluation measures for trained ML models.

III. Background

This section presents brief theoretical background about ML approaches used in this research. Three ML models including *k*-NN, SVC and ANN were implemented for students performance classification. A brief theoretical background of each is presented as follows.

1. k Nearest Neighbour (k-NN)

k-NN is a non-parametric ML approach was first introduced in 1951 by Fix et al. [21]. Algorithm for classification works on the principle of classifying an input into one of the target classes based on popularity among its neighbours (i.e., classes of nearest neighbours). In the ML domain, *k*-NN classification is most commonly used approach for the case when there is no knowledge about the data distribution. The original algorithm has been extended over the years in terms of the definition of formal properties [22], introduction of new rejection approaches [23], Bayes error rate refinements [24], distance weighted technique [25], soft computing approaches [26], and fuzzy approaches [27].

The algorithms basically works on computing the Euclidean distance between test and training samples. Let yi = (yi1; yi2; : : : yim) be the input sample with m features where (i = 1; 2; : : : n). The Euclidean distance between yi (training sample) and yt (test sample) can be determined using the expression in Equation 1.

$$d(y_i, y_t) = \sqrt{\sum_{m}^{m=1} (y_{im} - y_{tm})^2} \quad (1)$$

A major shortcoming of the majority-based voting occurs for the unbalanced class dataset which is a common scenario in real-world. Since in this case, each new test example will be biased to be classified to the class with a greater number of samples. One approach to address this problem is to assign the weights to neighbours (i.e., weighted *k*-NN). The most important part of this algorithm is to select the appropriate value of K which is highly dependent on the dataset. In general, a greater value of K reduces the noise effect in the dataset, however, makes classes boundary less distinct [28]. Since *k*-NN is primarily based on the Euclidean distance, normalizing the training data can significantly improve the classification performance.

A. Support Vector Classifier (SVC)

SVC is one of the most robust supervised ML algorithms introduced originally by Vapnik in 1963 and extended by Boser et al. [29]. Algorithm is based on the construction of multiple hyperplanes in the high-dimensional space with the aim to achieve good separation between the hyperplanes. The high margin between hyperplanes indicates the reduced generalization loss. Given the test samples to be classified into one of two classes, each test sample is represented as an m-dimensional vector and separated by an (*m - 1*) dimensional hyperplane. Test samples may be separated by multiple hyperplanes, however, the bet one is selected based on the maximum separation in linear classification case [30], [31]. For a binary classification case, let us consider (*xi; yi*) represents the labeled input samples, where xi are a feature vector and *yi* are their respective class labels i.e., -1 or 1. SVM aims to construct a model to assign each input feature vector to one of the target classes. Equation 2 mathematically expresses the functionality of the binary classifier [31].

$$f(x, w, b) = < w, x > + b \tag{2}$$

Where w denotes the decision hyperplane and b is the intercept.

A. Artificial Neural Network (ANN)

ANNs are the ML algorithms inspired by the biological functionality of the animal brain and have proven effective. The network consists of nodes, connections and layers. Nodes are the representations of artificial neurons and capable of processing the input signal and transmitting it to other neurons. Each neuron transforms the input by some non-linear function and transmits the output. These neurons are connected with each other over number of layers where each layer responsible certain transformations. Each layer is assigned the weights which represent the strength of a signal at a given neuron and improves over the training iterations. Typically, ANN contains of input, output layer and number of hidden layers with artificial neurons at each layer connected with each other [32], [33].

Listed are the main components of ANN:

- **Artificial Neurons** are conceptually inspired by the biological neurons and process the given input using a non-linear activation function to generate output.
- **Connections** Neurons in the network are connected with each other by connections which are responsible for transmitting the output of one neuron as an input to other neuron. One neuron is usually connected to multiple neurons.
- **Weights:** Each connection in the network has a weight which determines its significance. These weights are updated during the training process to optimize the performance of network.

IV. Experimental Design

This section provides information about the dataset and protocols used to perform the analysis.

A. Dataset

Open University (OU) dataset [34] from the Kaggle competition was used for the experiment. Dataset consisted of total 32,593 student entries from 15 different countries. Furthermore, dataset contained the information about courses taken by students, students demographics and students interactions with the VLE. The dataset was pre-processed to clean and extract the relevant features to be used for analysis. Dataset cleaning mainly included dealing with missing values and assigning numerical values to phrases for classification analysis. Demographic (D), Engagement (E) and Performance (P) are the input features in the dataset, while student performance (pass/fail) is the target variable.

B. Data Pre-Processing

For this experiment, extracted data in CSV format was prepared using the guidelines of [18]. Dataset was split based on three main features including Demographic (D), Engagement (E) and Past Performance (P). Details of dataset features under each category are presented in Table I.

The dataset contained entries with missing (Nan) value for those students who withdrawn from the course or did not attend the final exam. In this study, Nan values were replaced with zero in contrary to [18] approach who removed the Nan entries. Feature Selection during the training process has impacted the performance of the models. In this experiment, filter based feature selection approach was applied to the metric, 10 features were resulted as the best number of features to be selected for efficient training.

C. Experimental Protocol

For all three algorithms; *k*-NN, SVC and ANN, dataset division of 80:20 was used where 80 percent data was used for training and 20 percent data was used for testing. For *k*-NN algorithm, 5 number of neighbours were used with uniform weights. For SVC, rbf kernel was used with shrinking heuristic set as true. Finally, for ANN classifier, two hidden layers were used with 10 and 15 number of neurons, respectively. Relu activation function with Adam optimizer was used and constant learning rate of 0.001 was set for training. ANN was trained for 200 iterations in this experiment. Number of iterations were selected from trial and error to see if the loss is converged. To assess the performance of ML algorithms for student performance prediction applied for binary classification for student pass or fail, number of measures were used including classification accuracy, F1 score and J-Index [35].

V. Results And Analysis

This section provides the results of implemented *k*-NN, SVM and ANN. ML algorithms for the student performance prediction based on the combination of different input features. ML algorithms were applied to classify the students performance for permutations of three features (demographic, engagement, past performance) from the OU education dataset. Basic protocol of the experiments was based on the [18] guidelines. Table II presents the results of experiment.

From Table II, it observed that *k*-NN performed best for E, P and D+E cases among the implemented three with accuracy of 0.9622, 0.9993 and 0.9627, respectively. However, for D, D+P, E+P and D+E+P cases, ANN performed best with the accuracy of 0.7596, 0.9981, 0.9983 and 0.9980, respectively. To highlight the scope of performed analysis, results were compared with the study performed by the [18] using same models, same features from the same dataset as used in this experiment. Table III presents the comparative analysis. From the results, it can clearly observed that analysis performed in this study achieved better results for all the permutation cases of features. Improved results from what already reported may be attributed to replace missing values by 0 instead of removing them. Furthermore, for the proposed investigation standard scaling (i.e, standardize features by taking off the mean and scaling to unit variance) is used.

Students learning behavior and student historical data were reported as the source for instructors to determine the student performance. However, some student has different assessment type counts on the same module and course period, which make the data relay on the student behavior rather than module design. Moreover, the number of assessment student has been taken during the course based on different factors such as the extent level of student engaged in the course. Also, how the course materials and VLE dashboard attract student to spend a reasonable time to get better outcome. Teaching materials and style on the VLE effect how the student engaged during the course and their final results.

Table I. Description of Dataset Features

Indicator Feature Term	Value	Description
Demographic		
Gender	[0,1]	0: male 1: female
Highest education	[0,0.25,0.5,0.75,1]	0: Below high school 0.25: High school 0.5: Diploma 0.75: Bachelor 1: Post graduate
Age	[0,0.5,1]	0: <35 0.5: 35-55 1: >55
Engagement		
Sum of clicks	[0-N]	0 - N
Performance		
Score per assessment	[0-1]	0 - 100
No. of Attempts	[0-1]	0 - N
Final exam score	[0-1]	0 - N

Table II. Performance comparison for three classification ML Algorithms (k-NN, SVM and ANN) to predict final exam result (D-Demographics, E-Engagement, P-Performance Data)

		D	E	P	D+E	D+P	E+P	D+E+P
Accuracy	*k*-NN	0.6905	**0.9622**	**0.9993**	**0.9627**	0.9966	0.9966	0.9934
	SVC	0.7594	0.9516	0.9990	0.9340	0.9963	0.9983	0.9958
	ANN	**0.7596**	0.9219	0.9984	0.9298	**0.9981**	**0.9983**	**0.9980**
F1 Score	*k*-NN	0.8108	**0.9756**	**0.9995**	**0.9759**	0.9977	0.9977	0.9956
	SVC	0.8632	0.9685	0.9993	0.9564	0.9975	0.9988	0.9972
	ANN	**0.8633**	0.9480	0.9989	0.9535	**0.9987**	**0.9988**	**0.9986**
J-Index	*k*-NN	0.6818	**0.9524**	**0.9991**	**0.9530**	0.9955	0.9955	0.9913
	SVC	0.7593	0.9390	0.9987	0.9165	0.9951	0.9977	0.9945
	ANN	**0.7596**	0.9012	0.9979	0.9112	**0.9975**	**0.9977**	**0.9973**

F1 Score based Performance Comparison of this paper with Tomasevic Et. AI. [18] for three classification ML Algorithms (k-NN, SVM, ANN)

	D		E		P		D+E		D+P		E+P		D+E+P	
	*	**	*	**	*	**	*	**	*	**	*	**	*	**
k-NN	**0.81**	0.61	**0.97**	0.91	**0.99**	0.94	**0.97**	0.88	**0.99**	0.94	**0.99**	0.93	**0.99**	0.94
SVC	**0.86**	0.72	**0.96**	0.94	**0.99**	0.93	**0.95**	0.93	**0.99**	0.94	**0.99**	0.95	**0.99**	0.96
ANN	**0.86**	0.71	**0.94**	0.95	**0.99**	0.94	**0.95**	0.94	**0.99**	0.94	**0.99**	0.96	**0.99**	0.96

* This paper
** Tomasevic et al. [18]

VI. CONCLUSION

This paper investigated the OU dataset towards predicting performance of students using variety of ML algorithms. Experiment consisted of three main stages (a) data preparation (b) implementation of ML algorithms (c) presentation of results and critical analysis. From the experiments, *k*-NN algorithm performed better in comparison to SVM and ANN for different feature permutations. Student historical data indicated that assessment type and a number of the previous attempt have a high influence on the student's final result.

REFERENCES

[1] J. M. Crotty, "Distance Learning Has Been around since 1892, You Big MOOC," section: Tech. [Online]. Available: https://www.forbes.com/sites/jamesmarshallcrotty/2012/11/14/distance- learning-has-been-around-since-1892-you-big-mooc/

[2] J. Jani, R. Muszali, S. Nathan, and M. S. Abdullah, "Blended learning approach using frog vle platform towards students' achievement in teaching games for understanding," *Journal of Applied and Fundamental Sciences*, vol. 10, pp. 1131–1151, Jan. 2018.

[3] I. Chatziralli, C. V. Ventura, S. Touhami, R. Reynolds, M. Nassisi, T. Weinberg, K. Pakzad-Vaezi, D. Anaya, M. Mustapha, A. Plant, M. Yuan, and A. Loewenstein, "Transforming ophthalmic education into virtual learning during COVID-19 pandemic: a global perspective," *Eye*, pp. 1–8, Jul. 2020, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41433-020-1080-0

[4] H. Mellar, R. Peytcheva-Forsyth, S. Kocdar, A. Karadeniz, and B. Yovkova, "Addressing cheating in e-assessment using student authentication and authorship checking systems: Teachers' perspectives," *International Journal for Educational Integrity*, vol. 14, p. 2, Feb. 2018.

[5] C. Vegega, P. Pytel, and M. F. Pollo-Cattaneo, "Application of the Requirements Elicitation Process for the Construction of Intelligent System-Based Predictive Models in the Education Area," in *Applied Informatics*, ser. Communications in Computer and Information Science, H. Florez, M. Leon, J. M. Diaz-Nafria, and S. Belli, Eds. Cham: Springer International Publishing, 2019, pp. 43–58.

[6] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Computer Science*, vol. 72, pp. 414–422, Jan. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915036182

[7] A. S. J. Abu Hammad, "Mining Educational Data to Analyze Students' Performance (A Case with University College of Science and Technology Students)," *Central European Researchers Journal*, vol. 4, no. 2, 2018.

[8] A. Elbadrawy, R. Studham, and G. Karypis, "Personalized Multi-Regression Models for Predicting Students' Performance in Course Activities," Mar. 2015.

[9] M. Yee-King, A. Grimalt-Reynes, and M. d'Inverno, "Predicting student grades from online, collaborative social learning metrics using K-NN." in *EDM*, 2016, pp. 654–655.

[10] H. Al-Shehri, A. Al-Qarni, L. Al-Saati, A. Batoaq, H. Badukhen, S. Alrashed, J. Alhiyafi, and S. O. Olatunji, "Student performance prediction using support vector machine and k-nearest neighbor," in *2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE)*, 2017, pp. 1–4, tex.organization: IEEE.

[11] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, "Machine learning based student grade prediction: A case study," *arXiv preprint arXiv:1708.08744*, 2017.

[12] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores," Oct. 2018, iSSN: 1687-5265 Pages: e6347186 Publisher: Hindawi Volume: 2018. [Online]. Available: https://www.hindawi.com/journals/cin/2018/6347186/

[13] H. Heuer and A. Breiter, "Student success prediction and the trade-off between big data and data minimization," *DeLFI 2018-Die 16. E-Learning Fachtagung Informatik*, 2018, publisher: Gesellschaft für Informatik eV.

[14] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student Performance Prediction and Classification Using Machine Learning Algorithms," Mar. 2019, pp. 7–11.

[15] M. El Fouki, N. Aknin, and K. E. El Kadiri, "Multidimensional approach based on deep learning to improve the prediction performance of DNN models," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, no. 02, pp. 30–41, 2019.

[16] S. Hussain, Z. Muhsen, Y. Salal, P. Theodorou, F. Kurto ğlu, and Hazarika, "Prediction Model on Student Performance based on Internal Assessment using Deep Learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 14, p. 4, Apr. 2019.

[17] S. Ajibade, N. Ahmad, and S. M. Shamsuddin, "Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods," *IOP Conference Series: Materials Science and Engineering*, vol. 551, p. 012061, Aug. 2019.

[18] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers & Education*, vol. 143, p. 103676, Jan. 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360131519302295

[19] D. Hooshyar, M. Pedaste, Y. Yang, L. Malva, G.-J. Hwang, M. Wang, H. Lim, and D. Delev, "From Gaming to Computational Thinking: An Adaptive Educational Computer Game-Based Learning Approach," *Journal of Educational Computing Research*, p. 0735633120965919, Oct. 2020, publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/0735633120965919

[20] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior*, vol. 104, p. 106189, 2020, publisher: Elsevier.

[21] E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.

[22] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[23] M. E. Hellman, "The nearest neighbor classification rule with a reject option," *IEEE Transactions on Systems Science and Cybernetics*, vol. 6, no. 3, pp. 179–185, 1970.

[24] K. Fukunaga and L. Hostetler, "K-nearest-neighbor bayes-risk estimation," *IEEE Transactions on Information Theory*, vol. 21, no. 3, pp. 285–293, 1975.

[25] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 4, pp. 325–327, 1976.

[26] S. Bermejo and J. Cabestany, "Adaptive soft k-nearest-neighbour classifiers," *Pattern Recognition*, vol. 33, no. 12, pp. 1999–2005, 2000.

[27] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE transactions on systems, man, and cybernetics*, no. 4, pp. 580–585, 1985.

[28] D. Coomans and D. L. Massart, "Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules," *Analytica Chimica Acta*, vol. 136, pp. 15–27, 1982.

[29] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[30] Y. Pu, D. B. Apel, and H. Xu, "Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier," *Tunnelling and Underground Space Technology*, vol. 90, pp. 12–18, 2019.

[31] D. Fradkin and I. Muchnik, "Support vector machines for classification," *DIMACS series in discrete mathematics and theoretical computer science*, vol. 70, pp. 13–20, 2006.

[32] A. Abraham, "Artificial neural networks," *Handbook of measuring system design*, 2005.

[33] K. Mehrotra, C. K. Mohan, and S. Ranka, *Elements of artificial neural networks*. MIT press, 1997.

[34] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Scientific data*, vol. 4, p. 170171, 2017.

[35] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.