Educational data mining: using knowledge tracing as a tool for student success

Joel Mandlazi
School of Computer Science
and Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa
thabo.mandlazi1@students.wits.ac.za

Ashwini Jadhav
Science Teaching and Learning Unit
Faculty of Science
University of the Witwatersrand
Johannesburg, South Africa
Ashwini.Jadhav@wits.ac.za

Ritesh Ajoodha
School of Computer Science
and Applied Mathematics
University of the Witwatersrand
Johannesburg, South Africa
Ritesh.Ajoodha@wits.ac.za

***Abstract*—Student success is important to institutions of learning. Learning institutions offer academic support to their students to ensure success. Knowledge Tracing (KT), the task of modelling student knowledge based on their learning history is a significant topic in the field of Educational Data Mining and has numerous applications in the development of interactive and adaptive learning technology. KT can be used to understand each student's unique learning behaviour, individual needs, and skill-levels.**

Machine learning classification models were implemented to trace the knowledge state of students. The Support Vector Machine outperformed all other models in this article, with Random Forest performing the worst, with accuracy of 76.8% and 70.76%, respectively. One important goal of KT is to personalise study materials in order to assist student's understand learning concepts more efficiently.

***Index Terms*—Machine Learning, Educational Data Mining, Knowledge Tracing**

I. Introduction

For many students, the thrill of being accepted into a South African university is short-lived, as the hurdles they experience are sometimes overwhelming, leading to many dropping out. Letseka and Maile [1] reported that, according to the Department of Higher Education and Training (DHET), roughly 30% of students drop out in their first year of study. Another 20% drop out during their second and third years of school, with only 22% graduating in record time. The National Treasury invested R4.5 billion in grants and appropriations to institutions of higher learning, with no return on their investments, this was due to students dropping out. The Department of Higher Education and Training (DHET) recognises that dropouts and student retention are a problem [1], [2]. In order to improve the problem, student retention methods have been implemented. However, the student retention issue continues to exist year after year.

With the dark cloud of covid-19 looming over our heads, institutions of higher learning are being compelled to shift from face-to-face instruction to online teaching and learning, indicating that institutions of higher learning are to produce more data than they been producing. Educational Data Mining (EDM) is a subset of data mining that focuses on the creation of tools for analysing specific types of educational data. Its overarching purpose is to help better understand how students learn and to uncover factors that can help them learn more effectively [3].

Researchers have been drawn to developing models for predicting students' knowledge states in order to provide appropriate feedback as a result of the availability of large-scale student performance datasets [4]. Knowledge tracing (KT) is the process of tracing a student's knowledge state, which demonstrates their level of mastery of topics based on previous learning activities [5]. KT is a supervised sequence learning task - given student's previous interactions with an exercise $X = (x_1, x_2, ..., x_t)$, predict the probability that the student will answer the next question of the exercise correctly, i.e., $P(r_t = 1 = 1|e_{t+1}, X)$. The interactions are shown as $x_t = (e_t, r_t)$, where e_t is the exercise attempted by the student at time t and r_t is the correctness of the student's answer. The goal of KT is to forecast whether the student will be able to correctly solve the next exercise or not [6], [7].

This paper's main contribution was to implement KT models that trace a student's level of knowledge (whether or not a student understands a concept being taught in class), with the goal of assisting institutions of higher learning in developing individually tailored guidances to assist students to master concepts being taught in class.This will also improve student achievement because a student will receive study material that is completely focused on their learning patterns. As a result, student dropouts and retention rates will be decreased.

II. Literature Review

Many traditional strategies have been developed to evaluate student performance, identify student behaviour patterns in academic settings, anticipate their performance, and find patterns relevant for course content delivery. Table I summarises the relevant literature aiming at predicting student performance.

A. Data

1) Institutional Datasets: Higher education institutions generate a significant amount of data, which includes a number of attributes such as student demographic details, accomplishments, and behavioural characteristics [14]. Despite the fact that the majority of researchers obtain their data from

Authors	Background	Academic	Socio-Economic	Model Used	Accuracy
Ajoodha, Jadhav, and Dukhan [8]	✓	✓	✓		
Daud, Aljohani, Abbasi, *et al.* [9]	✓		✓	SVM	86%
Osmanbegovic and Suljic [10]	✓	✓	✓	NB	76%
Kabakchieva [11]	✓	✓	✓	NN	72%
Minaei-Bidgoli, Kashy, Kortemeyer, *et al.* [12]		✓		KNN	82%
Lonia Masangu, Ashwini Jadhav [13]	✓	✓		RF	69%
Daud, Aljohani, Abbasi, *et al.* [9]	✓		✓	J48	76%

Table I: Review of key papers that use various approaches to predict student performance

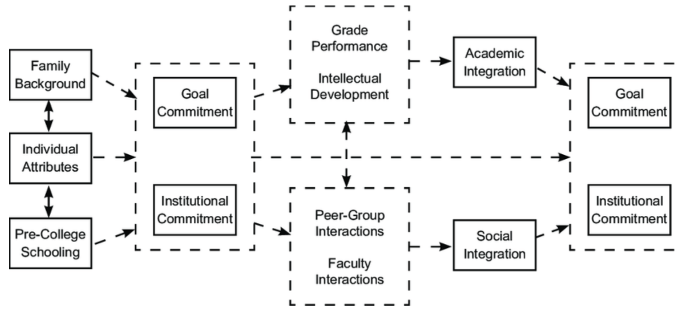

Figure 1: Conceptual framework by Tinto [19], adopted from [8]

institutions of higher learning, various attributes are used.
Researchers use data from educational institutions to predict student success. Three separate datasets from colleges in India, each with 24 attributes were used [15]. Student records from various universities in Pakistan were used in [9]. Student data from two separate databases containing 10330 student records from Bulgarian universities were used [11]. Datasets from two public schools in Portugal and nine separate high schools in Kancheepuram district were used in [16], [17], respectively.
For the purposes of this article, we will use EdNEt Dataset [5], which is made up of online courses, similar to how [18] used a dataset of 15 online courses from MIT and Harvard. All of the researchers discussed in this section have the same aim in mind: to predict whether or not a student will answer the next question correctly.

2) Questionnaire: One method for gathering data is through a survey in the form of questionnaires. Questionnaires are used by researchers to investigate students' comfort levels, educational backgrounds, and teaching methods.
Questionnaires that included personal and school-related variables that have effect on student performance were used in [17]. [14] also followed the method in [16], [17] of collecting data using a questionnaire, however with a different purpose of evaluating teaching methods that also affect student performance.

B. Tinto's Conceptual Framework and Features

1) Tinto's Conceptual Framework: Tinto [19] presented a widely used conceptual framework (see figure 1) for understanding the student attrition process. In this model, three types of attributes, namely Demographic, Academic, and Socio-Economic, are interconnected and are thought to influence student's academic performance [20]. Despite the fact that most authors employ the Tinto [19] conceptual framework, the model fails to account for numerous social-psychological causes for student retention, which are especially important for minority students. Transitional interactions for underrepresented minority students to college can mirror catastrophic situations [21]. The use of these attributes are discussed in the section that follows.

2) Demographic Information: The demographic information, which may also be refer to as background of the student is one the most used feature by researchers when predicting student success. Demographic information includes, age, gender, disability and birth place, to mention few, this information is used to predict student performance [9], [12], [16], [18], [22]. Some Authors, instead of using all the student background features, they rank features to discover the most influential ones. For an example, relief attribute evaluation method was used to rank highly influential features, to which gender was amongst the eleven highly influential attributes selected to predict student performance [15].
Furthermore, [23], took a slightly different approach by applying data reprocessing techniques to transforming data into First Normal Form, to acquire fields required for data mining such as gender and language of instruction.

3) Academic Background: Academic background, referred to as pre-college or schooling characteristics by Tinto [19], includes summative assessments [12], university or high school results [9], [16] and participation of student in class activities, are among the attributes used by researchers. Some researchers used online academic information, like [12] used introductory to physics course for scientist and engineers homework sets with 184 problems to predict students performance.
Study related data that contain common attributes such as admission point scores and number of credits gained from passed courses were used in [14], [24]. Collection of detailed marks of students [17] and assignment grades [18], were acquired from different educational institutions, with the sole purpose of predicting student success from both.

4) Socio-Economic: Research shows that socio-economic issues such as family income, parents' education, and parents' occupations, do affect student performance. Machine learning classifiers may be used to analyse the effect of each feature for predicting students' performance [9]. After using machine learning to analyse feature influence, being *"self employed"* was considered the best predictor of student performance [9]. Furthermore, family income and whether or not the student lived with their parents also played a role in predicting student performance [16], [17].

C. Models

1) Neural Networks (NN): Different problems including prediction, approximation, function, classification, and pattern

recognition have been solved successfully using neural networks [10]. The first known work was published in 1993, where NN was used to predict student performance [25]. Thereafter many authors followed on their foot steps.
In recent years, poor quality graduates from several Nigerian institutions have been observed, pointing to the inadequacy of the national university admission examination system in achieving this. An ANN was created to forecast a candidate's likelihood of being accepted for admission [26]. ANN was used to forecast a sophomore's success in engineering majors at the Faculty of Engineering and Information Technology [27]. According to their test data, the ANN is capable of accurately predicting performance with the accuracy of 70% and 80% in [26] and [27], respectively.

Multiple authors implemented NN to predict student success, as seen in table I. Despite the fact that their individual efforts utilised different datasets and properties. It is worth noting that NN did exceptionally well in all of their papers, with [13] having the lowest accuracy of 64% and [15] having the highest accuracy of 99%.

2) Decision Trees: developed by Professor Ross Quinlan in 1993, is a greedy algorithm that selects the next characteristics based on knowledge gathered from the previous attributes. The attribute with the greatest information gain or entropy reduction is chosen as the test attribute for the current node [28].
$C4.5$ use the Gain ratio as an attribute selection criterion and is the most commonly and potentially widely used decision tree algorithm [10], [29]. This may be attributable to a variety of factors, including its additional features such as managing missing data, categorising continuous attributes, and others.
$C4.5$ was used to predict student success by nearly every author in literature shown in table I. It is clear that the $C4.5$ algorithm performs exceptionally well; however, the [14] shows that it performed poorly despite the use of common features such as academic information and student behaviour. This may be due to the form of data used.

3) Naïve Bayes: (NB) is a basic classification method based on probability theory. It is referred to as naïve because it simplifies difficulties by relying on two primary assumptions: that the prognostic qualities are conditionally independent of familiar classification and that there are no hidden attributes that could impact the prediction process [10].
NB is also a common algorithm for predicting student success. It is also one of the most commonly used models as evident in literature shown in table I. It is one of the best performing algorithms. However, in the paper [22] and [11], the accuracy is less than 60%.

4) K-Nearest Neighbour: A slightly different educational prediction challenge was taken in [30]. Where KNN is used in this study to estimate which level of study material students would see based on their talents and performance. According to the research paper, numerous different versions of each course should be created and categories for different levels of skill. Based on their current performance, the KNN algorithm is then utilised to identify the best suited lesson level for each student.
Despite its simplicity and lack of assumptions regarding the prior probabilities of the training data, the KNN technique has been proved to perform effectively in real-world classification challenges [31]. Given its simplicity, the method's performance when compared to other classifiers is quite good [8], [12], [17].

III. Methodology

We use Machine Learning algorithms in this research to predict whether a student will answer the next question correctly or incorrectly. To achieve this, we trained the 7 machine learning classification models using 10-fold cross validation.

A. Data Collection and Pre-processing

EdNet dataset [5], includes information about student behaviours such as the learning resources he/she has consumed and their response. Santa compiled the student interactions of over 780K users over a two-year period, making it the largest of the ITS datasets made available to the public thus far. The dataset is freely available for research purposes. EdNet offers four layers of datasets, each labeled KT1, KT2, KT3, and KT4. The number and diversity of processes involved grows in direct proportions to the level of the dataset. Due to memory limits, we randomly selected data from 100 users who attempted to solve exercises from the KT1 dataset, further went to limit the number of attempts to 50 per user.

B. Features

KT1 is made up of question-answering logs from students, and it is the most fundamental and crucial knowledge that may be used by various deep-learning KT models. Table II shows features of KT1 dataset.

Feature	Type	Description
timestamp	unix time	time student accessed the question
solving_id	numeric	bundle id in which question was taken from
question_id	numeric	question number
user_answer	Categorical (a, b, c, d)	possible student answer(s)
elapsed_time	unix time	time it took student to answer

Table II: Description of Features

C. Classification and Evaluation

Machine learning classification methods were used in this paper. 10fold cross validation was used in each of the classifications instances. The machine learning algorithms were implemented using a total of 5000 student attempts. The following are the seven classification models that were used:

a) Multilayer Perceptron (MLP): A feed forward neural network that uses back propagation to learn the parameters. This model is a multi-layer neural network using sigmoid functions as nodes at each layer. In cases where the class is numeric, the output nodes become unthresholded linear units [25]. The implementation of MLP in this paper is based on [25].

b) $C4.5$: is the most commonly and potentially widely used decision tree algorithm, which was developed by Professor Ross Quinlan in 1993 [10], [28]. This may be attributable to a variety of factors, including its additional features such as managing missing data, categorising continuous attributes, and others. The $C4.5$ classification method used in this paper is based on [28].

c) Naïve Bayes (NB): is a probabilistic model that uses Bayesian inference to calculate the probability of a degree of causation. Using prior probability of in- stances feature A given class B. The Bayes rule can be interpreted as: $P(Y|X) = \frac{P(X|Y)}{P(X)}$, where $P(Y|X)$ represents the desired posterior probability. $P(X|Y)$ represents the prior probability. $P(X)$ and $P(Y)$ are the features observed and target class respectively [32]. The implementation of the NB follow that of [32].

d) Support Vector Machines (SVM): The classification model combines the training data with a non-probabilistic binary linear classifier that splits the training data classes using a multidimensional hyperplane. Then, in the same space, test instances are mapped and forecast based on which side of the hyper-plane they fall on. Using a kernel trick and one-versus-all class splitting, SVM may be scaled for nonlinear and high-dimensional classification. The SVM implementation used in this study is based on Hussain, Dahan, Ba-Alwib, *et al.* [15]'s implementation.

e) Random Forests (RF): is made up of many decision trees that work together as an ensemble to combine the results and implementation of various models and instances. These uncorrelated decision trees have their properties merged and function together in what is known as a committee, which outperforms any individual tree model. This paper's implementation of RF is focused on [16].

f) K:* is based on how close instances are to each other and then analysis is made on this. This works by calculating the distance between test data instances and training data formulations using an entropy-based function.

g) K-Nearest Neighbor (KNN): is a method for classifying objects in the feature space using the most recent training instances. KNN is a method of instance-based learning or lazy learning in which the function is only approximated locally and all computation is postponed until classification [24]. The KNN implementation used in this work is based on Ajoodha, Jadhav, and Dukhan [8]'s implementation.

To evaluate all of the above-mentioned classification models, the following metrics were used:

Confusion Matrix: is used to evaluate model performance and consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). The number of negative examples correctly classified is denoted by TN, the number of negative examples incorrectly classified as positive is denoted by FP, the number of positive examples incorrectly classified as negative is denoted by FN, and the number of positive examples correctly classified is denoted by TP. A confusion matrix is shown in table III.

		actual	
		Positive	Negative
Predicted	Positive	TP	FP
		FN	TN

Table III: Confusion Matrix

Accuracy: is a widely used statistic for assessing classification methods. It is derived by dividing the total number of forecasts by the number of correct guesses 1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Recall: is defined as the number of correct positive predictions divided by the total number of positives. This is also known as the true positive rate. Equation 2 shows how to calculate the true positive rate.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision: defined as the number of correct positive predictions as a proportion of the overall number of positive predictions. Equation 3 shows how to calculate precision.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

F1 score: represents the average of recall and precision (shown in equation 4). It acts as a warning sign when a performance has been erroneously classified.

$$\mathrm{F}_1 = (\frac{recall^{-1} + precision^{-1}}{2})^{-1} = 2 * (\frac{recall * precision}{recall + precision}) \tag{4}$$

IV. Results and Discussion

We implemented and tested the classification models to predict whether a student will get the next question correctly or not, based on their previous interaction with the exercises. A confusion matrix was used to test the efficiency of the classification algorithm after applying 10-fold cross validation. The classification model used are RF, SVM, $C4.5$, $K*$, KNN, MLP and NB. Figure 2 represents the outcome of each of these classifiers in predicting the class variable.

Figure 2a shows the confusion matrix for the $C4.5$ classification model which the fourth best performing model with an accuracy of 72.14%. $C4.5$, performed better than KNN, K* and RF. We also notice that 84% of instances of class a were correctly classified, where as 6% instance of class a were misclassified as a instance d. Figure 2b shows the confusion matrix for the MLP classification model is the third best performing model in this paper with an accuracy of 73.64%. MLP took the most time to build with over 60 seconds. This may be as a results of its multi-layers. Furthermore, we observe that 75% of instances of a were correctly classified

Actual

Predicted	a	b	c	d
a	1067	163	154	77
b	58	849	135	102
c	43	102	818	198
d	82	136	142	873

(a) A confusion Matrix describing the performance of the **C4.5** classification model. The **C4.5** classification model has an accuracy of **72.14**%. There were 3607 correctly classified instances and 1393 incorrectly classified.

Actual

Predicted	a	b	c	d
a	902	177	76	88
b	114	914	184	127
c	91	89	879	48
d	143	70	111	987

(b) A confusion Matrix describing the performance of **MLP** classification model. The neural network classification model has an accuracy of **73.64**%. There were 3682 correctly classified instances and 1318 incorrectly classified.

Actual

Predicted Answer	a	b	c	d
a	910	48	141	49
b	120	1083	159	95
c	84	62	818	160
d	136	57	132	946

(c) A confusion Matrix describing the performance of the **Naïve Bayes** classification model. The NB classification model has an accuracy of **75.14**%. There were 9757 correctly classified instances and 1243 incorrectly classified.

Actual

Predicted	a	b	c	d
a	1106	86	96	64
b	43	951	184	163
c	61	104	871	111
d	40	109	99	912

(d) A confusion Matrix describing the performance of the **SVM** classification model. The SVM classification model has an accuracy of **76.8**% .There were 9840 correctly classified instances and 1160 incorrectly classified.

Actual

Predicted	a	b	c	d
a	758	114	135	128
b	102	941	165	64
c	236	136	847	66
d	154	59	103	992

(e) A confusion Matrix describing the performance of the **RF** classification model. The RF classification model has an accuracy of **7.76**%. There were 3538 correctly classified instances and 1462 incorrectly classified.

Actual

Predicted	a	b	c	d
a	880	83	129	114
b	75	914	71	107
c	135	147	987	176
d	160	106	63	823

(f) A confusion Matrix describing the performance of the **KNN** classification model. The random forests classification model has an accuracy of **72.08**% .There were 3604 correctly classified instances and 1396 incorrectly classified.

Actual

Predicted	a	b	c	d
a	822	172	122	79
b	181	962	131	80
c	174	66	848	120
d	73	50	149	971

(g) A confusion Matrix describing the performance of the **K*** classification model. The random forests classification model has an accuracy of **72.06**% .There were 3603 correctly classified instances and 1397 incorrectly classified.

Figure 2: A set of confusion matrices describing the performance of classification models. Each classification model's accuracy is indicated along with the correctly and incorrectly classified instances.

and over 14% instances of class b were misclassified as class c instances.

Figure 2c shows the confusion matrix for the Naïve Bayes classification model is the second best performed model with accuracy of 75.14% following SVM. We note that with approximately 86% instances of actual class b were correctly classified, with over 12% instances of class c misclassified as instances of d. Figure 2d shows the confusion matrix for the SVM classification model is the best performed model with the accuracy of 76.8%. Followed by NB. We also note that 88% instances of class b were correctly classified, with the lowest 3% instance class d being misclassidied as instances of class a. Figure 2e shows the confusion matrix for the Random Forest classification model which marks it as the least performed model with the accuracy of 70.76%. It took the second longest time to build with 0.42 seconds.

Figure 2f shows the confusion matrix for the KNN classification model marks the third least performed model, following RF and K* with the accuracy of 72.08%. We also not that 12% of class a instances were misclassified as class d instances.

Figure 2g shows the confusion matrix for the $K*$ classification model achieved the second least performing classification

Model	Accuracy	Precision	Recall	F1-Score
SVM	76.8	0.768	0.768	0.768
NB	75.14	0.752	0.7514	0.7517
MLP	73.64	0.7387	0.7364	0.7375
C4.5	72.14	0.7211	0.7214	0.7212
KNN	70.08	0.7231	0.7208	0.7219
K*	72.06	0.7203	0.7206	0.7204
RF	70.76	0.7062	0.7076	0.7069

Table IV: Performances of the classification models

model with the accuracy of 72.06%. We also note that 13% of class a instances are misclassified as class b instances.
Figure 3 depicts the results from table IV. This was done to assess the classification models' results in greater depth. In addition to accuracy, we also evaluated precision and recall. How these were calculated from the confusion matrices is given in detail in III-C0g. We discovered that precision and recall were quite close to accuracy and hence provided presumably very comparable interpretations.

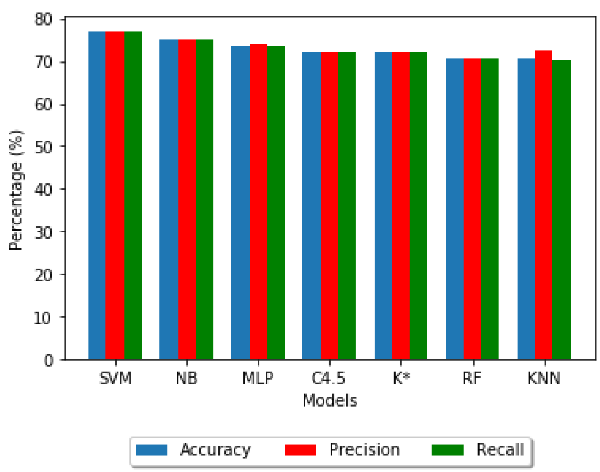

Figure 3: Bar graph illustrating classification model performances.

V. Conclusion

In conclusion, we reiterate that tertiary education is fraught with uncertainty when it comes to the prospects of new students, with many falling along the way struggling to keep up with the demands and pressures of the tertiary learning environment, and many failing to obtain their degree qualifications. We demonstrated in this research study that employing Machine Learning classification models to trace a student's knowledge is a viable. Thus, it may be used to track students' knowledge levels, and study recommendations can be given to students in order to improve their performance. Machine Learning algorithms were implemented using the EdNet KT1 dataset to predict student success, our data shows that SVM outperforms alternative classification algorithms, with random forest performing the worst. We intend to improve our results by implementing Knowledge tracing models on other EdNet datasets, which will necessitate more powerful computing capacity.

ACKNOWLEDGEMENT

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant numbers: 121835).

References

[1] M. Letseka and S. Maile, "High university drop-out rates: A threat to south africa's future, hsrc policy brief", in. Jan. 2008, pp. 1–9, ISBN: 9780-7969-2228-1.

[2] Department of Higher Education and Training, *2000 to 2016 First Time Entering Undergraduate Cohort Studies For Public Higher Education Institutions*, https://www.dhet.gov.za/HEMIS/2000 TO 2016 FIRST TIME ENTERING UNDERGRADUATE COHORT STUDIES FOR PUBLIC HEIs.pdf.

[3] A. V. Manjarres, L. G. M. Sandoval, and M. S. Suárez, "Data mining techniques applied in educational environments: Literature review", *Digital Education Review*, no. 33, pp. 235–266, 2018.

[4] J. Self, "Theoretical foundations for intelligent tutoring systems", *Journal of Artificial Intelligence in Education*, vol. 1, no. 4, pp. 3–14, 1990.

[5] Y. Choi, Y. Lee, D. Shin, J. Cho, S. Park, S. Lee, J. Baek, C. Bae, B. Kim, and J. Heo, *Ednet: A large-scale hierarchical dataset in education*, 2020. arXiv: 1912.03072 [cs.CY].

[6] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, "Dynamic key-value memory networks for knowledge tracing", in *Proceedings of the 26th International Conference on World Wide Web*, International World Wide Web Conferences Steering Committee, 2017, 765–774. DOI: 10.1145/3038912.3052580.

[7] S. Pandey, G. Karypis, and J. Srivastava, *An empirical comparison of deep learning models for knowledge tracing on large-scale dataset*, 2021. arXiv: 2101.06373 [cs.AI].

[8] R. Ajoodha, A. Jadhav, and S. Dukhan, "Forecasting learner attrition for student success at a south african university", in *Conference of the South African Institute of Computer Scientists and Information Technologists 2020*, 2020, pp. 19–28.

[9] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting student performance using advanced learning analytics", in *Proceedings of the 26th international conference on world wide web companion*, 2017, pp. 415–421.

[10] E. Osmanbegovic and M. Suljic, "Data mining approach for predicting student performance", *Economic Review: Journal of Economics and Business*, vol. 10, no. 1, pp. 3–12, 2012.

[11] D. Kabakchieva, “Predicting student performance by using data mining methods for classification”, *Cybernetics and information technologies*, vol. 13, no. 1, pp. 61–72, 2013.

[12] B. Minaei-Bidgoli, D. Kashy, G. Kortemeyer, and W. Punch, “Predicting student performance: An application of data mining methods with an educational web-based system”, in *33rd Annual Frontiers in Education, 2003. FIE 2003.*, vol. 1, 2003, T2A–13. DOI: 10.1109/FIE.2003.1263284.

[13] R. A. Lonia Masangu, Ashwini Jadhav, “Predicting Student Academic Performance Using Data Mining Techniques”, *Advances in Science, Technology and Engineering Systems Journal*, vol. 6, no. 1, pp. 153–163, 2021. DOI: 10.25046/aj060117.

[14] H. Bydžovská, “A comparative analysis of techniques for predicting student performance.”, *International Educational Data Mining Society*, 2016.

[15] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, “Educational data mining and analysis of students’ academic performance using weka”, *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447–459, 2018.

[16] P. Cortez and A. M. G. Silva, “Using data mining to predict secondary school student performance”, 2008.

[17] V. Ramesh, P Parkavi, and K Ramar, “Predicting student performance: A statistical and data mining approach”, *International journal of computer applications*, vol. 63, no. 8, 2013.

[18] R. Al-Shabandar, A. Hussain, A. Laws, R. Keight, J. Lunn, and N. Radi, “Machine learning approaches to predict learning outcomes in massive open online courses”, in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 713–720. DOI: 10.1109/IJCNN.2017.7965922.

[19] V. Tinto, “Dropout from higher education: A theoretical synthesis of recent research”, *Review of educational research*, vol. 45, no. 1, pp. 89–125, 1975.

[20] J. P. Bean, “Conceptual models of student attrition: How theory can help the institutional researcher”, *New directions for institutional research*, vol. 1982, no. 36, pp. 17–33, 1982.

[21] C. Bennett and A. M. Okinaka, “Factors related to persistence among asian, black, hispanic, and white undergraduates at a predominantly white university: Comparison between first and fourth year cohorts”, *The Urban Review*, vol. 22, no. 1, pp. 33–60, 1990.

[22] B. Sekeroglu, K. Dimililer, and K. Tuncal, “Student performance prediction and classification using machine learning algorithms”, in *Proceedings of the 2019 8th International Conference on Educational and Information Technology*, 2019, pp. 7–11.

[23] U. Pandey and S. Pal, “Data mining : A prediction of performer or underperformer usingclassification”, *Journal of Computer Science and Technology*, vol. 4, Apr. 2011.

[24] D. Kabakchieva, “Student performance prediction by using data mining classification algorithms”, *International journal of computer science and management research*, vol. 1, no. 4, pp. 686–690, 2012.

[25] T. Gedeon and H. Turner, “Explaining student grades predicted by a neural network”, in *Proceedings of 1993 International Conference on Neural Networks (IJCNN-93-Nagoya, Japan)*, IEEE, vol. 1, 1993, pp. 609–612.

[26] V. Oladokun, A. Adebanjo, and O. Charles-Owaba, “Predicting students academic performance using artificial neural network: A case study of an engineering course”, 2008.

[27] S. S. Abu-Naser, I. S. Zaqout, M. Abu Ghosh, R. R. Atallah, and E. Alajrami, “Predicting student performance using artificial neural network: In the faculty of engineering and information technology”, 2015.

[28] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, 1993.

[29] R. Kabra and R. Bichkar, “Performance prediction of engineering students using decision trees”, *International Journal of computer applications*, vol. 36, no. 11, pp. 8–12, 2011.

[30] B.-Y. Shih and W.-I. Lee, “The application of nearest neighbor algorithm on creating an adaptive on-line learning system”, in *31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education. Conference Proceedings (Cat. No.01CH37193)*, vol. 1, 2001, T3F–10. DOI: 10.1109/FIE.2001.963925.

[31] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*. Cambridge University Press Cambridge, 2008, vol. 39.

[32] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers”, in *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo: Morgan Kaufmann, 1995, pp. 338–345.