Student's Performance Prediction based on Personality Traits and Intelligence Quotient using Machine Learning

Samar El-Keiey[1], Dina ElMenshawy[2], Ehab Hassanein[3]
Information Systems Department
Faculty of Computers and Artificial Intelligence
Cairo University
Egypt

***Abstract*—Apparently, most life activities that people perform depend on their unique characteristics. Personal characteristics vary across people, so they perform tasks in different ways based on their skills. People have different mental, psychological, and behavioral features that affect most life activities. This is the same case with students at various educational levels. Students have different features that affect their academic performance. The academic score is the main indicator of the student's performance. However, other factors such as personality features, intelligence level, and basic personal data can have a great influence on the student's performance. This means that the academic score is not the only indicator that can be used in predicting students' performance. Consequently, an approach based on personal data, personality features, and intelligence quotient is proposed to predict the performance of university undergraduates. Five machine learning techniques were used in the proposed approach. In order to evaluate the performance of the proposed approach, a real student's dataset was used, and various performance measures were computed. Several experiments were performed to determine the impact of various features on the student's performance. The proposed approach gave promising results when tested on the dataset.**

***Keywords*—Prediction; student performance; machine learning; personality; intelligence quotient**

I. Introduction

People have a wide range of cognitive abilities, including intelligence, memory, attention, and so on. People can carry out brain-based operations or activities in a variety of ways and for varying lengths of time. Two people can perform the same operation in two different ways. This is because people's personal characteristics differ, and most life activities are dependent on these personal characteristics.

Each person has certain abilities which he/she uses to deal with various real-life activities. Also, people have different ways to process and memorize information. Actually, all operations or activities a person performs are based on his/her features either the behavioral features or the psychological ones.

For example, the tasks that need some sort of intelligence, people who have higher intelligence quotients usually finish their assigned tasks in fewer steps and in less time than others who have lower intelligence quotients. Moreover, people who have higher memory skills usually remember things better and more quickly than other people who have lower memory skills.

Apparently, most life operations or actions which people perform are done based on their various skills. The same idea goes for students at all educational levels. Students have different characteristics either psychologically or mentally which affect their academic performance.

Some students perform better in theoretical questions that need remembering skills while other students perform better in practical questions which need brain-based skills. Mainly the academic score of a student depends on his/her unique characteristics.

The main purpose of this research is to measure the impact of the personality traits and the intelligence of students on their academic performance. In addition to these features, some personal data and the academic score have been used to predict the academic performance of undergraduate students.

To the best of our knowledge, no work exists answering the question of the impact of personal data, personality traits, intelligence quotient, and academic score on the student's performance.

The research questions of this research are:

1) What effect does the Intelligence Quotient (IQ) have on the undergraduate student's performance?

2) What effect do personality traits have on the undergraduate student's performance?

3) What effect does the combination of IQ, personality traits, personal data, and academic score have on the undergraduate student's performance?

4) What are the most significant features in the student's performance prediction?

II. Student Performance Prediction

Student performance prediction has played a significant role in educational systems. Predicting student performance helps students to select appropriate courses which match their skills as the student's performance can vary across different courses. Prediction helps students choose courses that match their abilities. Also, student performance prediction can assist in designing appropriate future study plans for students. Beyond predicting student performance, it helps teachers and managers to monitor and support the students, and to offer training programs to achieve the best results. Other benefits of student performance prediction are reducing official warning signs and discarding students for inefficiency.

Machine learning techniques have played a significant role in the creation of effective educational systems over the past two decades. These techniques helped in offering better learning techniques and in enhancing the academic performance of students [1].

Applying machine learning techniques in educational systems has played a crucial role in discovering concealed and unexpected methods to impart knowledge across all educational levels. As a result, some prediction models have been proposed by numerous researchers to enhance the student's performance and learning quality as in [2], [3].

III. Motivation

Student performance mainly can be predicted through historical records of quizzes and exams, and Grade Point Average (GPA). This is usually common in various educational levels and ages. However, sometimes there are other factors that may influence a student's performance such as mentality skills, behavioral characteristics, cognitive skills, personality traits, and psychological factors.

Also, some of these factors may affect the performance of students at certain ages but may have no effect at other ages or at other educational levels. For example, personality traits may have a great influence on older students like university students but may not have a noticeable effect on younger students at primary or high school levels.

Moreover, the factors that are considered important for student performance prediction for preschool students may not have the same importance as for older students like undergraduates in universities. Also, gender can play a role in student performance prediction. For example, sometimes male students have better scores in different courses or specialties than female students and vice versa.

The importance of a certain factor or a feature in student performance prediction usually depends on the age or the educational level in which the student is enrolled. Consequently, in this research, we focus on a certain educational level which is undergraduates in universities.

As a result, in this paper, an approach based on personal data, personality traits, and intelligence quotient is proposed to predict the student's performance of university undergraduate students.

The main contributions of this paper are as follows:

1) Proposed an approach that utilized academic score, gender, region/city, number of brothers/sisters, Intelligence Quotient (IQ), and personality features to predict student's performance.

2) Applied five machine learning techniques in the proposed approach to predict student's performance.

3) Applied the proposed approach on real students' datasets.

4) Compared the performance of different features across the five machine learning techniques.

5) Predicted the best indicators that assist in student performance prediction.

The rest of the paper is organized as follows: Section 4 presents the literature work on students' performance prediction. Section 5 presents the proposed approach. Section 6 presents the results. Section 7 presents the analysis and discussion of the results. Section 8 presents the conclusion and future work.

IV. Related Work

In the following paragraphs, related work on students' performance prediction using different techniques is presented.

In recent decades, many attempts have been made to predict students' academic performance before students start the learning process to make their outcomes predictable. This is also necessary for instructors to know the areas where students have defects so that students' skills in these areas can be improved. By predicting future results in a timely manner, instructors can know the areas that need improvement while teaching students.

Educational institutions and governments also want to know the performance of the current educational system in order to perform improvements in the long term. In [4], the authors showed that students' performance depends on various factors such as demographics, behavior, previous outcomes, and habits. Unexpected factors, such as the address of students, had a great impact on the student's performance.

Many studies have been conducted to discover the effect of various variables on student academic performance. These factors are not the same all over the world and may vary from university to university, from university to school, and also from individual to individual.

In today's world, data has become very powerful, and machine learning can be very helpful in harnessing the power of this data. Machine learning techniques, along with deep learning techniques, have played a very important role in predicting student academic performance.

Various machine learning and deep learning techniques, such as Support Vector Machine (SVM), Neural Networks (NN), and clustering have been studied on different datasets in different institutions to find hidden and unexpected patterns. Several machine learning techniques were applied in [5].

In [6], various models for academic performance prediction have been developed using Decision Tree (DT), Naïve Bayes (NB), and Rule-Based (RB) for the Bachelor of Computer Science students at the University Sultan Zainal Abidin. The results showed that DT and RB provided better accuracies than NB.

In [7], the authors applied SVM, NN, DT, and NB models on two independent datasets. The results showed that SVM performed better than the other methods in terms of statistical significance.

Predicting student performance is a key challenge in the educational process that uses technology to help students towards success [8]. Previous research has detected several factors that influence the student's performance. Some of these factors are student demographics such as gender [9], previous academic grades [10], and interaction with the learning environment [11], [12].

In [13], DT, NB, logistic regression, SVM, K-nearest neighbors, iterative minimum optimization, and NN were utilized to study the student's performance in final undergraduate exams. Logistic regression performed the best compared to other models used in this study, with an accuracy of 66%.

In [14], the proposed model applied NN, logistic regression, and SVM algorithms in a virtual school learning environment.

In [15], recurrent neural networks (RNNs) for detailed knowledge and engagement studies were used. In this study, the authors achieved an accuracy value of 88.3%. SVM, linear discriminant analysis, random forests, K-nearest neighbors, and classification regression trees (CART) were used in this study. Random forests performed the best, with an accuracy of 90%.

In [16], the research applied a logistic regression prediction model on some variables which are level of involvement, level of prestige, level of visibility, number of student visits, and management system by subject, experience, age, and gender. The predictive models used in the analysis were NN, DT, and NB using bagging, boosting, and ensemble techniques. The DT classifier gave the highest accuracy value of 82%.

In [17], the authors provided the most comprehensive assessment to ensure the strength of the relationship between the big five personality traits and academic performance through the synthesis of 267 independent samples (N = 413,074) in 228 original studies. Also, the progressive validity of personality traits beyond cognitive ability in predicting academic performance was examined.

In [18], the research introduced a Contextual Cognitive Skill Scores (CCSS) approach to predict the student's performance. To determine the CCSS score, the cognitive skills required to solve the exam questions were combined with the exam scores. In doing so, the authors focused on generalizing the student's performance to understand its utility in predicting student risk profiles.

Exam questions and their scores vary by course and by exam. Such data heterogeneity is very important when generalizing the model. This is because feature variation affects the results. Therefore, the authors built the CCSS so that the feature space is the same for all courses so that students' performance can be visualized in the same dimension.

In [19], the proposed model predicted the success of students based on their behavioral patterns/activities in the learning process. Six machine learning prediction models were presented. Then, accuracy measures were computed to evaluate the proposed model.

In [20], data was collected from different students with different parameters, then these parameters were analyzed using techniques similar to those used in [21]. After that, decision trees for all considered attributes were built. Next, "if-then" rules for the relationship between various attributes and the student's performance were generated.

In [22], some machine learning algorithms were utilized to predict and classify different types of educational data. Three machine learning algorithms, namely, backpropagation (BP), support vector regression (SVR), and long short-term memory (LSTM) were used to predict the student's performance. In addition to these algorithms, a Gradient Boosting Classifier (GBC) is implemented in the classification phase.

V. Proposed Approach

In this section, the proposed approach is presented along with the features used, the techniques applied, and the dataset used.

The main idea of our proposed approach is to discover the impact of the big five personality traits and the intelligence quotient on the performance of students when these features are integrated with the academic score feature and some personal data namely, gender, number of brothers/sisters and the region/city where each student lives in. The proposed approach was implemented in Python.

A. Features Used and Data Collection

For the proposed approach, the data was collected from students in the Faculty of Computers and Artificial Intelligence, Cairo University. The data was collected from students from various levels, namely, 1st, 2^{nd,} and 3rd year undergraduate students. Approximately 300 students shared their data and their grades in different courses which are an introduction to database systems, fundamentals of computer science, and Web-based information systems.

The features collected are:

- Gender
- Number of brothers/sisters
- Region/city
- GPA
- IQ
- Five personality traits, namely,
    1. Openness (O)
    2. Conscientiousness (C)
    3. Extraversion (E)
    4. Agreeableness (A)
    5. Neuroticism (N)

The 300 students who shared their data were divided into 117 females and 183 males. The average GPA is 2.9 and the average IQ is 3.7.

A snapshot of the dataset is shown in Fig. 1.

GPA	Gender	brothers/sisters	O	C	E	A	N	IQ	city/region
2.32	Male	3	4	2	3	3	3	5	Helwan / Cairo
2.9	Male	1	2	5	1	4	1	4	Cairo
3.6	Male	2	3	5	5	1	1	4	Giza
2.1	Male	2	3	4	4	3	4	2	Helwan/Egypt
3.65	Female	2	2	5	2	5	2	5	Ciaro
2.99	Male	2	2	5	4	5	2	4	Giza
2.9	Male	2	2	3	4	4	4	3	Maadi , Cairo
2.67	Male	3	1	4	3	3	1	4	Nasr city
3.4	Male	3	2	5	3	3	1	5	fisal-guza
3	Female	2	4	3	3	3	5	4	giza
3.45	Female	2	4	4	4	4	3	5	New cairo
3.67	Female	2	4	5	4	4	4	5	Maadi-Cairo
3.95	Male	2	3	5	1	4	1	5	Giza/Haram
2.93	Male	3	2	3	3	4	3	4	giza
3.7	Female	2	3	5	4	3	1	4	Cairo
2.96	Male	2	3	4	3	4	1	4	Giza
3.01	Female	2	4	3	4	3	4	4	Giza
1.91	Male	2	2	1	4	1	5	2	Cairo
2.7	Female	3	3	3	3	1	4	4	Giza
2.87	Male	2	2	3	2	3	3	4	Cairo
3.43	Female	2	3	4	3	3	2	5	cairo

Fig. 1. Snapshot of the Dataset.

B. Data Preprocessing

The data was collected from a questionnaire via Google form so some preprocessing steps were performed so that the data can be used in the proposed prediction approach. The preprocessing steps focused mainly on removing noise and minimizing redundancy. Also, some features whose values were text were converted to numeric so that they can be used in the prediction model.

Moreover, the numeric scores were converted to categorical classes which are low, medium, and high so that the class label to which each student belongs can be predicted. Each categorical class included a range of values, as follows:

- Low (0 - <40)
- Medium (40 - <80)
- High (>= 80)

C. Techniques Used

Mainly, five machine learning techniques were used in our proposed approach in predicting the student's performance. The used techniques are:

1) **k-nearest neighbor (KNN):** KNN is a non-parametric supervised machine learning classifier algorithm, which is used for regression and classification [23]. After performing various experiments to determine the best value of k, a value of 6 gave the best accuracy, as presented in Fig. 2.

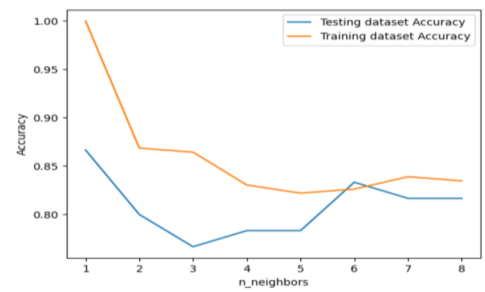

Fig. 2. The Best K Value.

2) **Support Vector Machine (SVM)** is a supervised learning model used for regression as well as for classification [24].

3) **Decision Tree (DT)** is a non-parametric supervised machine learning classifier algorithm, which is used for regression and classification. It has a hierarchical tree structure [25].

4) **Random Forest (RF)** is a classification algorithm used for classification and regression. It consists of several decision trees [26].

5) **Naive Bayes (NB)** is a probabilistic classifier that applies Bayes' theorem [27].

D. Features Importance

To know the most influential features in the prediction, the Scikit-learn (Sklearn) python library and the feature importance python function [28] were used to determine the most important features as presented in Fig. 3.

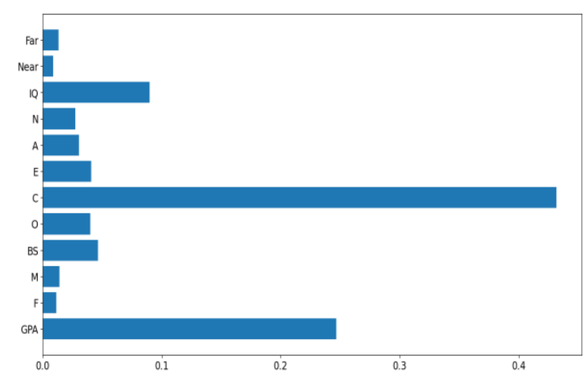

Fig. 3. The Most Important Features.

As shown in Fig. 3, the features are abbreviated as follows:

- GPA: Grade Point Average
- F: Female
- M: Male
- BS: Number of Brothers/Sisters
- O: Openness
- C: Conscientiousness
- E: Extraversion
- A: Agreeableness

- N: Neuroticism
- IQ: Intelligence Quotient

After performing several experiments, it was deduced that the most important features that have the highest significance on the student's performance are Conscientiousness, GPA, and IQ.

Also, it was noticed that gender and location have the least significant on the student's performance. The location feature is not of great importance, so our proposed approach can be applied easily in online learning.

E. Correlation Matrix

To know the correlation and the dependencies between the features, the correlation matrix was constructed, and it is shown in Fig. 4.

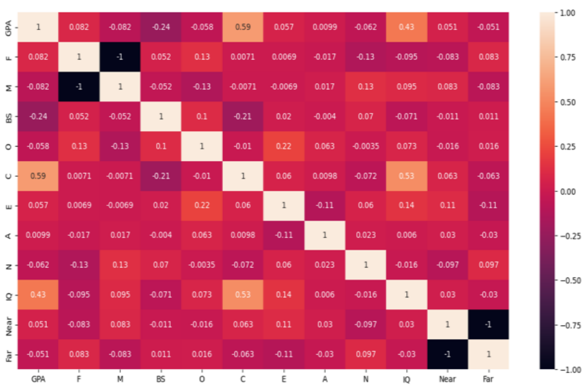

Fig. 4. Correlation Matrix.

The correlation between the features is the indicator of the extent of the effect of one feature on another one. The higher the value (either positively or negatively) between a pair of features, the higher is the correlation between them.

On the other hand, the less correlated features gave values closer to zero. When a value becomes closer to zero, this means that the features are less correlated.

When a value of a certain feature increases and the value of the other correlated feature increases, this denotes a positive correlation. On the other hand, when a value of a certain feature increases and the value of the other correlated feature decreases, and vice versa, this denotes a negative correlation.

After performing several experiments on the various features, the most correlated features were GPA, Conscientiousness, and IQ.

VI. RESULTS

The collected dataset was used to evaluate the performance of the proposed approach. The prediction accuracy, precision, recall, F1 measure, and the confusion matrix (true positive, true negative, false positive, and false negative) were computed for all techniques. Also, the results are compared across the five techniques. The results are presented in Table I.

TABLE I. PREDICTION RESULTS OF THE FIVE TECHNIQUES

Technique	Accuracy	Precision	Recall	F1 measure
KNN	0.85	0.84	0.85	0.84
NB	0.86	0.86	0.86	0.86
RF	0.83	0.82	0.83	0.83
DT	0.90	0.89	0.90	0.89
SVM	0.88	0.87	0.88	0.88

As shown in Table I, the decision tree technique gave the best performance among the other techniques with accuracy = 0.90%, precision = 0.89%, recall = 0.90% and F1 measure = 0.89%.

A. Confusion Matrix for All Techniques

The confusion matrix is used to analyze the performance of the classification techniques by computing the True positive (TP), the True Negative (TN), the False Positive (FP), and the False Negative (FN) for the testing data which are described as follows.

- TP: The true positive value is the case in which the actual value and expected value are identical [29].
- TN: A class's True Negative value is the sum of all columns and rows, except those for the class for which we are computing the values [29].
- FP: A class's False-positive value is the sum of all the values in the relevant column, except for the TP value [29].
- FN: A class's False-negative value is the sum of the values in the relevant rows, except for the TP value [29].

Using the confusion matrix, we can assess the model's performance in terms of recall, precision, and accuracy.

The confusion matrices for the prediction classification techniques are shown below in Tables II to VI.

1) K-Nearest Neighbor

TABLE II. KNN CONFUSION MATRIX

Actual values

	High	Medium	Low
High	31	0	0
Medium	1	11	4
Low	2	2	9

Predicted values

2) Naïve Bayes

TABLE III. NAIVE BAYES CONFUSION MATRIX

Actual values	High	Medium	Low
High	31	0	0
Medium	1	13	2
Low	3	2	8

Predicted values

3) Random Forest

TABLE IV. RANDOM FORESTS CONFUSION MATRIX

Actual values	High	Medium	Low
High	31	0	0
Medium	1	12	3
Low	3	2	8

Predicted values

4) Decision Tree

TABLE V. DECISION TREE CONFUSION MATRIX

Actual values	High	Medium	Low
High	31	0	0
Medium	0	14	2
Low	3	2	9

Predicted values

5) Support Vector Machine

TABLE VI. SUPPORT VECTOR MACHINE CONFUSION MATRIX

Actual values	High	Medium	Low
High	31	0	0
Medium	0	13	3
Low	2	2	9

Predicted values

VII. RESULT ANALYSIS AND DISCUSSION

The proposed approach proved that the big five personality traits, especially the "conscientiousness" feature, are the most significant features in predicting student's performance for undergraduate students in the Faculty of Computers and Artificial Intelligence, Cairo University in Egypt.

The intelligence quotient score also has a significant role in our prediction approach.

In order to test the efficiency of our proposed approach in integrating several features, personality traits and IQ features were removed from the dataset and the results were compared before and after removing these features as will be described in the following paragraphs.

A. The Impact of Removing the Big Five Personality Traits

Another experiment has been conducted to evaluate the performance of all techniques after removing all big five personality traits features from the dataset. This experiment was performed to monitor the impact of the big five personality traits on the academic performance of Egyptian students. The results are presented in Table VII.

TABLE VII. PREDICTION RESULTS WITHOUT BIG FIVE PERSONALITY FEATURES

Technique	Accuracy	Precision	Recall	F1 measure
KNN	0.66	0.68	0.66	0.67
NB	0.75	0.77	0.75	0.75
RF	0.76	0.77	0.76	0.76
DT	0.71	0.71	0.71	0.71
SVM	0.76	0.77	0.76	0.77

Table VII shows that accuracy, precision, recall, and F1 measure had decreased after removing the big five personality traits from the dataset, compared to the results in Table I.

Also, Table VII shows that the SVM provides the best performance in terms of accuracy, precision, recall, and F1 measure.

On the other hand, Table I shows that the decision tree technique provides the best performance in terms of accuracy, precision, recall, and F1 measure.

This emphasizes that the existence of the big five personality traits has a very significant role in predicting a student's academic performance.

B. The Impact of Removing the IQ Feature

Another experiment has been conducted to evaluate the performance of all techniques after removing the IQ feature from the dataset. This experiment was performed to discover the extent of the impact of the IQ feature on the academic performance of students. The results are presented in Table VIII.

Table VIII shows that the decision tree technique provides the best performance in terms of accuracy, precision, recall, and F1 measure; this corresponds to the results in Table I.

Table VIII shows that accuracy, precision, recall, and F1 measure had a little decrease after removing the IQ feature from the dataset, compared to the results in Table I.

This means that the existence of the IQ feature has an important role in the prediction of student academic performance in the proposed prediction approach.

TABLE VIII. PREDICTION RESULTS WITHOUT IQ FEATURE

Technique	Accuracy	Precision	Recall	F1 measure
KNN	0.76	0.76	0.76	0.76
NB	0.84	0.83	0.84	0.83
RF	0.82	0.81	0.82	0.82
DT	0.89	0.88	0.89	0.88
SVM	0.86	0.86	0.87	0.86

After conducting several experiments, the results have proved that "conscientiousness" is the most significant predictor when it is integrated with IQ.

To conclude, it was found that merging the big five personality traits with IQ and the academic feature besides the other moderator features namely, gender, region/city, and the number of brothers/sisters provided better results in all performance measures (accuracy, precision, recall, and F1 measure) compared to each one of them separately.

A comparison of accuracy with adding and removing the important features for all applied machine learning techniques is shown in Fig. 5.

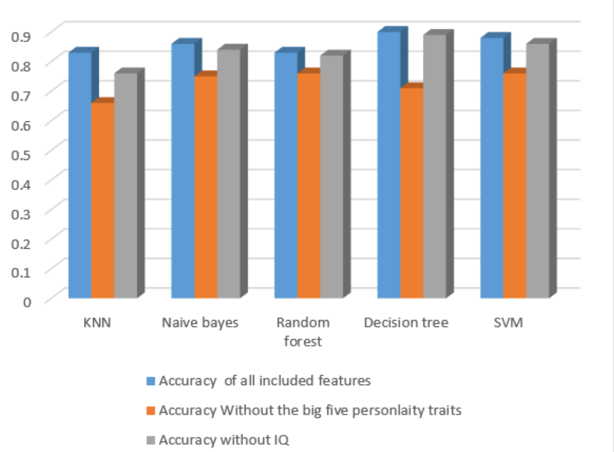

Fig. 5. Accuracy Comparison of All Techniques with and without Important Features.

VIII. CONCLUSION AND FUTURE WORK

Predicting student academic performance is very helpful for educators and learners to improve the teaching and learning processes. In this paper, student academic performance was predicted by applying various machine learning techniques with different features.

The main idea of this research is to discover the impact of the student's personality and the IQ along with other moderator features namely, gender, region/city, and the number of brothers/sisters integrated with the student's GPA on the student's academic performance.

Classification algorithms are widely used in education. K-nearest neighbor, decision trees, support vector machine, random forests, and Naive Bayes techniques were used to predict the student's academic performance. The decision tree technique performed the best in predicting the student's academic performance.

To enhance the effectiveness of the proposed approach, the personality traits and the IQ features were removed. After that, the proposed approach was re-implemented after removing these features then the performance measures were re-computed. The results showed a decrease in accuracy, precision, recall, and F1 measure which emphasizes the significant role of personality traits and IQ in predicting student academic performance.

To conclude, a lot of work can be done in predicting student academic performance so further research can be conducted as future work. This helps the educational systems to track the student's academic performance in a structured way. Furthermore, further research can use deep learning techniques and neural networks besides machine learning techniques to enhance the results.

ACKNOWLEDGMENT

We would like to express our appreciation to the students in the Faculty of Computers and Artificial Intelligence, Cairo University for their valuable role in helping us to collect the data needed for this research work.

REFERENCES

[1] K. Phanniphong, P. Nuankaew, D. Teeraputon, W. Nuankaew, M. Boontonglek and S. Bussaman "Clustering of learners performance based on learning outcomes for finding significant courses", *In Proc.* of ECTI DAMT-NCON - 4th International Conference on Digital Arts, Media and Technology and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, Nan, Thailand, doi: 10.1109/ECTINCON. 2019. 8692263, 2019, pp. 192-196.

[2] S. Bussaman, W. Nuankaew, P. Nuankaew, N. Rachata, K. Phanniphong and P. Jedeejit. "Prediction models of learning strategies and learning achievement for lifelong learning", *In Proc.* of IEEE International Conference on Teaching, Assessment and Learning for Engineering, Wollongong, Australia, doi: 10.1109/TALE.2017.8252332, January 2018, pp. 192-197.

[3] Nuankaew, Wongpanya, and J. Thongkam, "Improving student academic performance prediction models using feature selection", *In Proc.* of 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications, and Information Technology (ECTI-CON), Phuket, Thailand, 2020, pp. 392-395.

[4] S. Hirokawa, "Key attributes for predicting student academic performance", *In Proc.* of the 10th International Conference on Education Technology and Computers, New York, USA, DOI: https://doi.org/10.1145/3290511.3290576, 2018, pp. 308-313.

[5] Katarya, Rahul, J. Gaba, A. Garg, and V. Verma. "A review on machine learning based student's academic performance prediction systems", *In Proc.* of International Conference on Artificial Intelligence and Smart Systems (ICAIS), India, 2021, pp. 254-259.

[6] A. A Aziz and N. H. I. F. Ahmad, "First semester computer science students academic performances analysis by using data mining classification algorithms", *In Proc.* of the International Conference on Artificial Intelligence and Computer Science (AICS), Bandung, Astanaanyar, 2014, pp. 15-16.

[7] 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", Computers in Human Behavior, vol. 73, doi: 10.1016/j.chb.2017, pp. 247-256, 2017.

[8] A. Dutt and M. A. Ismail, "Can we predict student learning performance from lms data? a classification approach", *In Proc.* of 3rd International Conference on Current Issues in Education, Yogyakarta, Indonesia, 2019, pp. 24-29.

[9] Z. Cai, X. Fan, and J. Du, "Gender and attitudes toward technology use: A meta-analysis", Computers and Education, vol. 105, pp. 1-13, 2017.

[10] C. J. Asarta and J. R. Schmidt, "Comparing student performance in blended and traditional courses: Does prior academic achievement matter?", The Internet and Higher Education, vol. 32, pp. 29-38, 2017.

[11] 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", Computational Intelligence and Neuroscience, vol. 2018, pp. 1-21, 2018.

[12] E. A. Amrieh, T. Hamtini, and I. Aljarah, "Mining educational data to predict student's academic performance using ensemble methods", International Journal of Database Theory and Application, vol. 9, no. 8, pp. 119-136, 2016.

[13] A. S. Hashim, W. A. Awadh, and A. K. Hamoud, "Student performance prediction model based on supervised machine learning algorithms", *In Proc.* of IOP Conference Series: Materials Science and Engineering, Thi-Qar, Iraq, 2020.

[14] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from big data using deep learning models", Computers in Human Behavior, vol. 104, p. 106-189, 2020.

[15] K. Mongkhonvanit, K. Kanopka, and D. Lang, "Deep knowledge tracing and engagement with moocs". *In Proc.* of the 9th International Conference on Learning Analytics and Knowledge, Tempe, Arizona, 2019, pp. 340–342.

[16] J. C. S. Silva, J. L. Ramos, R. L. Rodrigues, A. S. Gomes, F. d. F. de Souza, and A. M. A. Maciel, "An edm approach to the analysis of students' engagement in online courses from constructs of the transactional distance", *In Proc.* of IEEE 16th International Conference on Advanced Learning Technologies (ICALT), Austin, TX, USA, 2016, pp. 230-231.

[17] Mammadov and Sakhavat, "Big Five personality traits and academic performance: A meta-analysis", Journal of Personality, vol. 90, no. 2, pp.222-255, 2022.

[18] Krishnamoorthy and Shivsubramani, "Student performance prediction, risk analysis, and feedback based on context-bound cognitive skill scores", Education and Information Technologies vol. 27, no. 3, pp. 3981-4005, 2022.

[19] Buraimoh, Eluwumi, R. Ajoodha, and K. Padayachee. "Application of machine learning techniques to the prediction of student success", *In Proc.* of IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, Canada, 2021, pp. 1-6.

[20] G. Shidaganti, I. Yadav, and H. Dagdi. "Identification of Student Group Activities in Educational Institute Using Cognitive Analytics", *In Proc.* of International Conference on Innovative Computing and Communications Springer, Singapore, 2022, pp. 275-284.

[21] Prateek, Student Group Activity Recognition, GitHub repository. https://github.com/prateekiiest/Student-Group-Activity-Recognition, 2017.

[22] S. Boran, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms", *In Proc.* of the 8th International Conference on Educational and Information Technology, Cambridge, UK, 2019, pp. 7-11.

[23] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression", The American Statistician, vol. 46, no. 3, pp.175-185, 1992.

[24] Wang, Lipo, ed, "Support vector machines: theory and applications", Springer Science and Business Media, vol. 177, 2005.

[25] Myles, J. Anthony, N.Robert, Feudale, Y. L., Nathaniel A. Woody, and Steven D. Brown. "An introduction to decision tree modeling", Journal of Chemometrics: A Journal of the Chemometrics Society, vol. 18, no. 6, pp.275-285, 2004.

[26] Biau, Gérard, and E. Scornet, "A random forest guided tour", vol. 25, no. 2, pp.197-227, 2016.

[27] Rish and Irina, "An empirical study of the naive Bayes classifier", In IJCAI 2001 workshop on empirical methods in artificial intelligence, vol. 3, no. 22. pp. 41-46, 2001.

[28] "Feature importances with a forest of trees". [Online]. Available: https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html [Accessed: 11-June-2022].

[29] "Confusion matrix for multi-class classification". [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/confusion-matrix-for-multi-class-classification/ [Accessed: 14-July-2022].