ISSN No:-2456-2165

Assessment of Offline and Online Learning Mode in the Performance of Student during Corona Virus Pandemic Era using Machine Learning in Python

Dr. Wilson MUSONI[1] PhD, Florence MUKAMABANO[2] Master's
[1,2] Information communication and Technology (ICT), University of Kigali,Kigali, Rwanda.
Subject area: Machine Learning

Abstract:- The shift from offline to online learning mode that many countries adopted during the covid-19 pandemic necessitates a greater knowledge on how students cope with these changes and its effect on their performance. The offline/traditional learning system has many advantages especially in facilitating interactions and idea exchanges between students and teachers. However little is known about the effect of this transition to online learning considering the skills that students and teachers have in using innovative technologies and programs. We assessed the students' performance in both offline and online learning mode during the coronavirus pandemic period. We also examined the effective learning mode between the online and offline mode just comparing both learning modes through python libraries. We found that there is a significant difference in the students' performance between online and offline learning modes. Our results show that during this covid-19 pandemic students performed well in online learning mode compared to students who used and continued using offline learning. The prediction tools used suggest that the use of online and offline learning modes also improved the quality of education and empowered higher learning institutions to enhance strategies of more flexibility in offering more web-based classes. Consequently, the online learning mode appears to be particularly important in improving the education quality and the academic performance.

I. INTRODUCTION

Many years ago, Rwanda has been using an education system based offline in all levels of education(Moshman, 2015). This traditional learning provides real student-teacher interactions and classroom management where teachers are believed to have the ability to manage the classroom interactions properly by "setting appropriate seating arrangements, managing learning activities, choosing the most interesting topic to discuss and making students disciplined by setting the rules(Muluk et al., 2021).

With the current technology development, learners want quality programs that can be accessed from different sources at any time(Paul & Jefferson, 2019). Based on this demand combined with COVID-19 pandemic requirements in education sector, online education became a pedagogical shift from traditional method to the modern approach of teaching-learning from classroom to Zoom, from personal to virtual and from seminars to webinars (Mishra, Gupta, & Shree, 2020). Using technology in classrooms has the potential to create increased student motivation, increased social interactions, positive outcomes, enhanced student learning, and enhanced student engagement(Costley, 2014).

(Howard, McGee, Hong, & Shia, 1998) realized that students' knowledge from online classes might be more highly positive than those of traditional classes as lecturers are requested to teach and assess achievement in ways that enable students to analyze, create and apply their knowledge.

According to(Wargadinata, Maimunah, Dewi, & Rofiq, 2020), the online learning models used during the COVID-19 pandemic equipped students with knowledge in a similar way to face to face learning. However, online learning provides more experience and the importance of process learning that balances the development of time and technology based on self-regulatory capabilities, which is definitely owned by each student. According to(Wargadinata et al., 2020), the regulations enforced by the head of an educational institution are most important when delivering online learning.

Consequently, it is important to examine how the student performance was during the COVID-19 pandemic to examine the effective learning mode between the online and offline mode for implementing an improved education systems and related strategies (Selvaraj, Radhin, Ka, Benson, & Mathew, 2021).

Many comparative studies have been carried out to prove the point to explore whether face-to-face or traditional teaching methods are more productive or whether online learning mode is better (Pei & Wu, 2019; Lockman & Schirmer, 2020;(González-Gómez, Jeong, Airado Rodríguez, & Cañada-Cañada, 2016) Results of all these studies show that the students perform much better in online learning than in traditional learning.

II. METHODOLOGY

This study assessed the student performance based on their final grade, just to model predicts the result of student's performance based on other properties, created by using the student's dataset. The dataset used for input data does include various types of information related to: gender, department, intake, education level, assignment score, Continuous Assessment Test score (CAT) and final score.

We estimated the effect of various machine learning models and algorithms. Algorithms were applied in generating predictive rules just taking into consideration four of them:Random Forest Regressor, Lasso Regression , Gradient Boosting Regressor and Decision Tree Regressor.

A. Data analysis

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains(datos, 2014).

Students' performance analysis is considered Supervised Machine Learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output. Y = f(X), The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance.

B. Cleaning the data

Data cleaning is the process of removing incorrect, duplicate, or otherwise erroneous data from a dataset.These errors can include incorrectly formatted data, redundant entries, mislabeled data, and other issues; they often arise when two or more datasets are combined together. Data cleaning improves the quality of your data as well as any business decisions that you draw based on the data(Sherrer, 2022).

If data is incorrect, outcomes and algorithms are unreliable, even though they may look correct. There is no one absolute way to suggest the exact steps in the data cleaning process because the processes will vary from dataset to dataset. But it is important to establish a template for your data cleaning process so you know you are doing it the right way every time. The dataset used needed the following changes to be considered clean:

- Dropping of Rows with Null Values
- Adding final_grade as a column which results from final_score converting.

C. Sample of dataset

	gender	department	level_of_study	intake	assign_score	cat_score	final_Score
0	female	General nursing	level I	September	73	69	71
1	female	Midwifery	level I	September	50	46	48
2	male	General nursing	level I	September	77	77	77
3	female	General nursing	level II	September	77	74	76
4	female	General nursing	level II	September	72	63	68

Fig. 1: Sample of online mode dataset

	gender	department	level_of_study	intake	assign_score	cat_score	final_Score
0	male	General nursing	level I	September	73	70	71
1	female	General nursing	level I	September	72	76	74
2	male	General nursing	level I	September	61	59	60
3	female	General nursing	level I	September	71	71	71
4	female	General nursing	level I	September	68	61	64

Fig. 2: Sample of offline mode dataset

D. Null Values

Most data science algorithms do not tolerate nulls (missing values). So, one must do something to eliminate them, before or while analyzing a data set. There are many techniques for handling nulls. Which techniques are appropriate for a given variable can depend strongly on the algorithms you intend to use, as well as statistical patterns in the raw data, in particular, the missing values, and the randomness of the locations of the missing values. Moreover, different techniques may be appropriate for different variables, in a given data set. Sometimes it is useful to apply several techniques to a single variable. Finally, note that corrupt values are generally treated as nulls.

The figure below shows the rows in the dataset used in this article and the number of missing values in each row.

```
gender           0
department       0
level_of_study   0
intake           0
assign_score     0
cat_score        0
final_Score      0
dtype: int64

gender           0.0
department       0.0
level_of_study   0.0
intake           0.0
assign_score     0.0
cat_score        0.0
final_Score      0.0
dtype: float64
```

Fig. 3: Null Rows dataset of online mode

```
gender           0
department       0
level_of_study   0
intake           0
assign_score     0
cat_score        0
final_Score      0
dtype: int64

gender           0.0
department       0.0
level_of_study   0.0
intake           0.0
assign_score     0.0
cat_score        0.0
final_Score      0.0
dtype: float64
```

Fig. 4: Null Rows dataset of offline mode

It is clear that both datasets either online or offline learning mode dataset have no null values. It means that our research datasets are accuratewith no missing values.

E. Adding final_grade column

It is important to make sure all columns are in the correct format to be processed by the chosen algorithm. So, the first step was to transform final_Score column by final_ grade.

Final_grade contains the excellent, very good, good, pass and fail categories. To properly see how each students performance has classified into the defined categories. Below is a picture of the final result after the mentioned operations.

F. Sample of dataset after changes

	gender	department	level_of_study	intake	assign_score	cat_score	final_Score	final_grade
0	female	General nursing	level I	September	73	69	71	Very Good
1	female	Midwifery	level I	September	50	46	48	Fail
2	male	General nursing	level I	September	77	77	77	Very Good
3	female	General nursing	level II	September	77	74	76	Very Good
4	female	General nursing	level II	September	72	63	68	Good

Fig. 5: Online mode dataset after changes

	gender	department	level_of_study	intake	assign_score	cat_score	final_Score	final_grade
0	male	General nursing	level I	September	73	70	71	Very Good
1	female	General nursing	level I	September	72	76	74	Very Good
2	male	General nursing	level I	September	61	59	60	Good
3	female	General nursing	level I	September	71	71	71	Very Good
4	female	General nursing	level I	September	68	61	64	Good

Fig. 6: Offline mode dataset after changes

G. Data Visualization

Data visualization is a graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. This blog on data visualization techniques will help you understand detailed techniques and benefits(Team, 2020).

III. RESULT

A. Machine learning models

We divided our datasets into two variables X as the features we defined earlier and Y as the **student final score** the target value we want to predict.

This is a regression problem so we used Regression methods.

Train test split will be ratio respectively.

B. Machine Learning Models used:

- Random Forest Regressor
- Lasso Regressor
- Gradient Boosting Regressor
- Decision Tree Regressor

C. The Process of Modeling the Data:

- Importing the model
- Fitting the model
- Predicting final score

D. Regression metrics

Score Metrics for Regression: Mean Absolute Error (MAE) - Mean of the absolute value of errors (absolute distance from true value):

Mean Squared Error (MSE) - Mean of the squared value of errors (squared distance from true value):

R^2 (coefficient of determination) - Regression score function.

E. Random Forest Regressor

Random forest is a Supervised Learning algorithm which uses ensemble learning method for classification and regression. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees

Mean Absolute Error: 1.75
Mean Squared Error: 5.14
R^2 Score: 0.8774

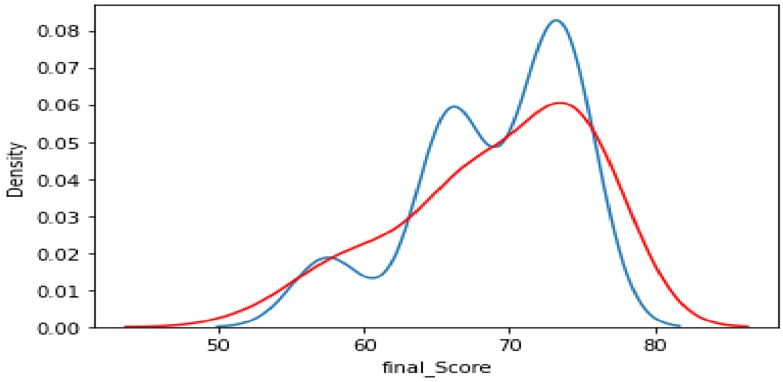

Fig. 7: Random forest plot of online mode

Mean Absolute Error: 2.44
Mean Squared Error: 13.42
R^2 Score: 0.7481

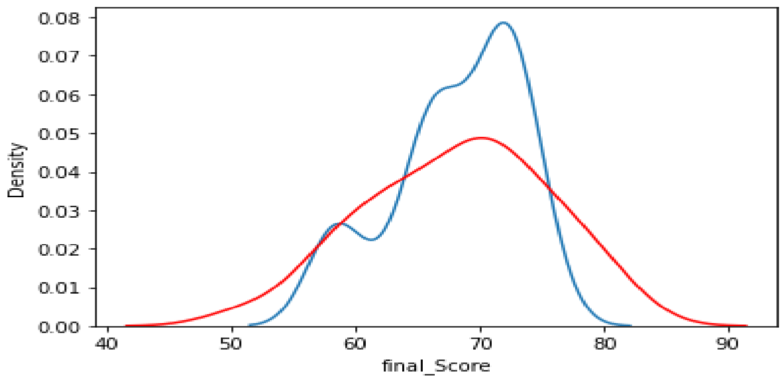

Fig. 8: Random forest plot of offline mode

F. Lasso Regressor

In statistics, to increase the prediction accuracy and interpret-ability of the model, Least Absolute Shrinkage and Selection Operator (LASSO) is extremely popular. It is a regression procedure that involves selection and regularization(Chetty, 2018).

Mean Absolute Error: 0.8
Mean Squared Error: 1.22
R^2 Score: 0.9709

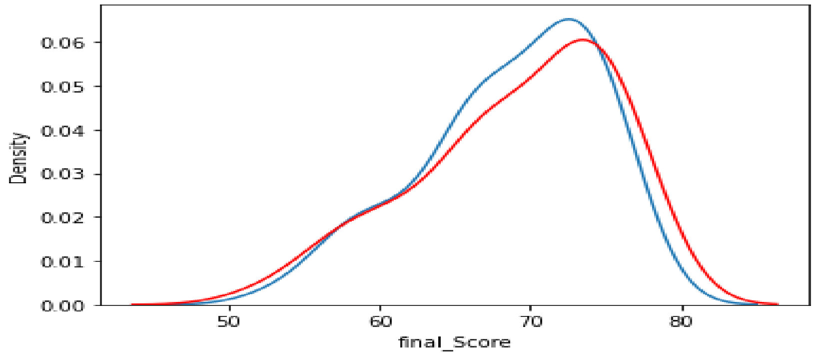

Fig. 9: Lasso plot of online mode

Mean Absolute Error: 1.22
Mean Squared Error: 8.82
R^2 Score: 0.8344

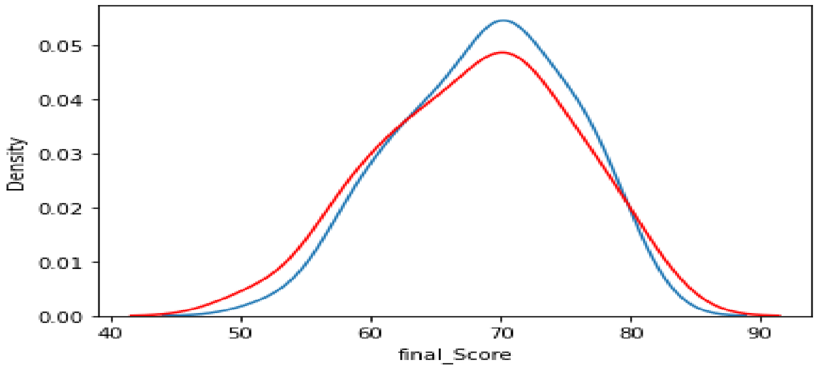

Fig. 10: Lasso plot of offline mode

G. Gradient Boosting Regressor

Gradient boosting algorithm is one of the most powerful algorithms in the field of machine learning.As we know that the errors in machine learning algorithms are broadly classified into two categories i.e. Bias Error and Variance Error. As gradient boosting is one of the boosting algorithms it is used to minimize bias error of the model (Tarbani, 2021).

Mean Absolute Error: 0.95
Mean Squared Error: 2.02
R^2 Score: 0.9518

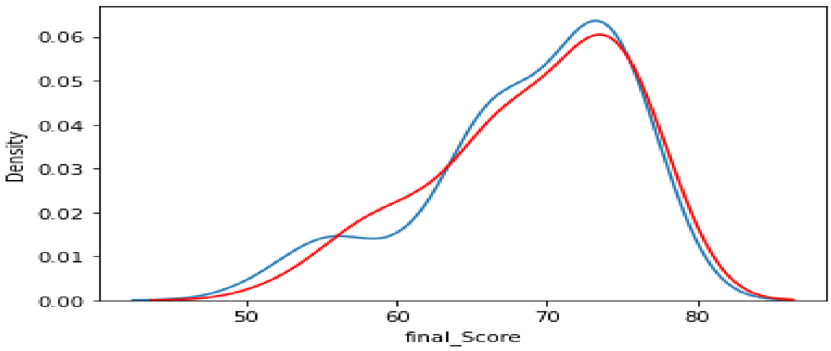

Fig. 11: Gradient boosting plot of online mode

Mean Absolute Error: 1.17
Mean Squared Error: 4.95
R^2 Score: 0.9072

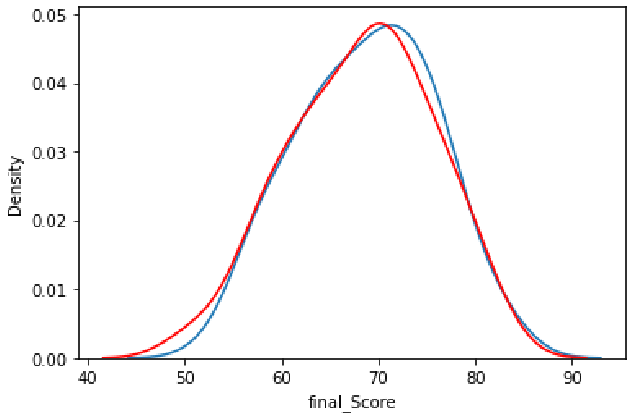

Fig. 12: Gradient boosting plot of offline mode

H. Decision tree

Decision Trees are a non-parametric supervised learning method used for both classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features(Abdulhafedh, 2022).

Mean Absolute Error: 1.54
Mean Squared Error: 4.07
R^2 Score: 0.9029

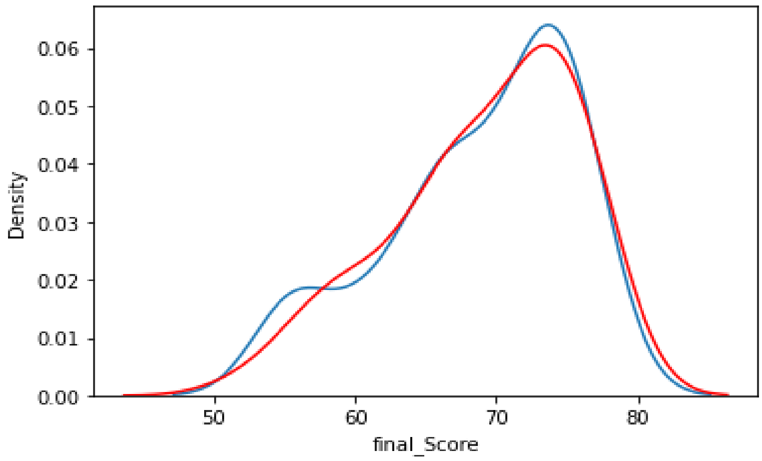

Fig. 13: Decision tree plot of online mode

Mean Absolute Error: 2.08
Mean Squared Error: 10.28
R^2 Score: 0.807

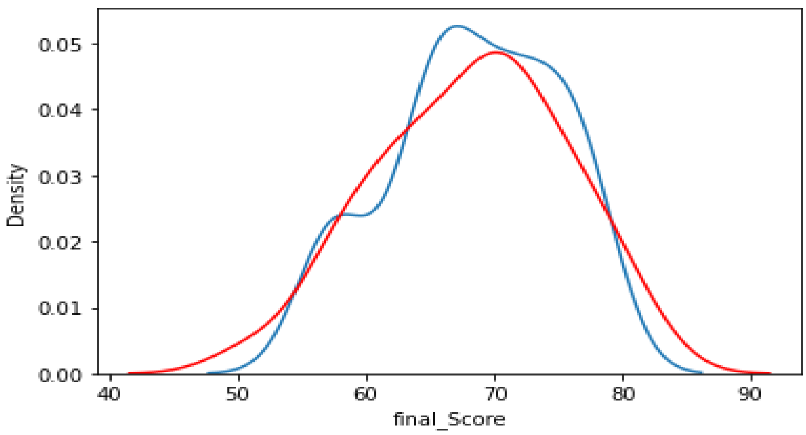

Fig. 14: Decision tree plot of offline mode

IV. FINAL RESULT OF ALL MODELS THAT HAS BEEN TESTED

No	Algorithms	Accuracy score	Mean Absolute Error	Mean Squared Error
1.	Random Forest Regressor	0.8774	1.75	5.14
2.	Lasso Regressor	0.9709	0.8	1.22
3.	Gradient Boosting Regressor	0.9518	0.95	2.02
4.	Decision Tree Regressor	0.9029	1.54	4.07

Table 1: Comparison of evaluation results of online learning mode

The above table shows the comparison of evaluation results of online learning mode where Lasso Regressor performed well with all metrics accuracy score, Mean absolute Error and Max Error. Lasso Regressor had the minimum error in predicting the students' performance through their final score when compared to the Random Forest Regression, Gradient Boosting Regression and Decision Tree Regression. Random Forest Regression demonstrated the worst performance with the highest error in all metrics. A simplified tabular form based on the results is created above.

The Lasso Regressor model is well suited for online learning mode because of the following reasons:

- Lasso Regressor method is used to predict the final score of the upcoming period. According to results there are high similarities between predictors and actual data.
- Lasso regressor is the suitable model in online learning mode because it shows the minimum error and high accuracy scorein predicting the students' performance compared with other models.

No	Algorithms	Accuracy score	Mean Absolute Error	Mean Squared Error
1.	Random Forest Regressor	0.7481	2.44	13.43
2.	Lasso Regressor	0.8344	1.22	8.82
3.	Gradient Boosting Regressor	0.9072	1.17	4.94
4.	Decision Tree Regressor	0.807	2.08	10.28

Table 2: Comparison of evaluation results of offline learning mode

The table below shows the comparison of evaluation results of offline learning mode where GBR stand for Gradient Boosting Regressor performed well with all metrics Accuracy Score, Mean Absolute Error and Max Error. GBR had the minimum error in predicting the students' performance through their final score when compared to the Random Forest Regression, Lasso Regression and Decision Tree Regression. Random Forest Regression demonstrated the worst performance with the highest error in all metrics. A simplified tabular form based on the results is created above.

The GBR model is well suited for offline learning mode because of the following reasons:

- GBR method is used to predict the final score of upcoming period. According to results there are high similarities between predictors and actual data.
- GBR is suitable for regression prediction problems where it shows the minimum error and high accuracy scorein predicting the students' performance compared with other models.

V. CONCLUSION

The main objective of this study was to assess the effect of offline and online learning mode in the performance of students which would help universities and higher learning institutions to be aware of which effective learning mode must be used for improving quality of education through the students' performance. The COVID-19 pandemic obliged learners to shift the normal learning process, offline learning mode, to the online learning process. In every country around the world, many students and teachers were forced to use online learning tools to teach and learn their courses. Based on the students' academic performance results before and during the outbreak of covid-19, our results show that students performed well in online learning mode compared to offline learning. Hence, our results suggest that teachers and studentsare likely tochange their ways in the learning methods to impact the education quality and performance. We realized that during the Covid-19 the online mode can help students do the task more efficiently and can always use the recorded online learning tools to enhance their knowledge and understating on different subject. The extent to which different universities and higher learning institutions have implemented online learning strategies, will continue to be influential in the future success of the education sector.

REFERENCES

[1.] Abdulhafedh, A. (2022). Comparison between Common Statistical Modeling Techniques Used in Research, Including: Discriminant Analysis vs Logistic Regression, Ridge Regression vs LASSO, and Decision Tree vs Random Forest. *Open Access Library Journal*, *9*(2), 1–19. https://doi.org/10.4236/oalib.1108414

[2.] Chetty, P. S., Priya. (2018, April 3). How to perform LASSO regression test? Retrieved November 16, 2022, from Knowledge Tank website: https://www.projectguru.in/perform-lasso-regression-test/

[3.] Costley, K. C. (n.d.). *The Positive Effects of Technology on Teaching and Student Learning*. 11.

[4.] datos, M. (2014, January 14). What is analysis of data. Retrieved November 14, 2022, from Manejando datos website: https://www.manejandodatos.es/2014/01/analysis-data/

[5.] González-Gómez, D., Jeong, J. S., Airado Rodríguez, D., & Cañada-Cañada, F. (2016). Performance and Perception in the Flipped Learning Model: An Initial Approach to Evaluate the Effectiveness of a New Teaching Methodology in a General Science Classroom. *Journal of Science Education and Technology*, *25*(3), 450–459. https://doi.org/10.1007/s10956-016-9605-9

[6.] Mishra, L., Gupta, T., & Shree, A. (2020). Online teaching-learning in higher education during lockdown period of COVID-19 pandemic. *International Journal of Educational Research Open*, *1*, 100012. https://doi.org/10.1016/j.ijedro.2020.100012

[7.] Moshman, D. (2015). Identity, history, and education in Rwanda: Reflections on the 2014 Nobel Peace Prize. *Child Abuse & Neglect*, *44*, 1–4. https://doi.org/10.1016/j.chiabu.2015.01.004

[8.] Muluk, S., Habiburrahim, H., Zulfikar, T., Akmal, S., Nasriyanti, N., & Safrul, M. S. (2021). Classroom Management in EFL Speaking Class: Strategies and Challenges. *IJEE (Indonesian Journal of English Education)*, *1*(1), 127–146.

[9.] Paul, J., & Jefferson, F. (2019). A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course From 2009 to 2016. *Frontiers in Computer Science*, *1*, 7. https://doi.org/10.3389/fcomp.2019.00007

[10.] Selvaraj, A., Radhin, V., Ka, N., Benson, N., & Mathew, A. J. (2021). Effect of pandemic based online education on teaching and learning system. *International Journal of Educational Development*, *85*, 102444. https://doi.org/10.1016/j.ijedudev.2021.102444

[11.] Sherrer, K. (2022, June 30). Data Cleaning: Definition, Methods & Steps. Retrieved November 14, 2022, from TechnologyAdvice website: https://technologyadvice.com/blog/information-technology/data-cleaning/

[12.] Tarbani, N. (2021, April 19). Gradient Boosting Algorithm | How Gradient Boosting Algorithm Works. Retrieved October 31, 2022, from Analytics Vidhya website: https://www.analyticsvidhya.com/blog/2021/04/how-the-gradient-boosting-algorithm-works/

[13.] Team, G. L. (2020, October 1). Understanding Data Visualization Techniques. Retrieved November 14, 2022, from Great Learning Blog: Free Resources what Matters to shape your Career! website: https://www.mygreatlearning.com/blog/understanding-data-visualization-techniques/

[14.] Wargadinata, W., Maimunah, I., Dewi, E., & Rofiq, Z. (2020). Student's Responses on Learning in the Early COVID-19 Pandemic. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *5*(1), 141–153. https://doi.org/10.24042/tadris.v5i1.6153