5Published by The College of Science & Information Technology (COSIT), TASUED, Vol. 18, No. 2, pp. 115-125.

COSIT, TASUED
Journal of Science and Information Technology (JOSIT)

# Student Performulator: Classification and Prediction of Academic Performance of Students Using Machine Learning.

**Adedeji, O.B.; Olusanya, O.O.; Adebare, A.O.; Omilabu, A.A.**

*Department of Computer and Information science, Tai Solarin University of Education, Ijagun, Ogun state.*
*Corresponding Author: adedejiob@tasued.edu.ng*
*Other Authors: olusanyaoo@tasued.edu.ng; adebaredj@gmail.com; omilabuoo@tasued.edu.ng*

**Abstract**

Predicting student academic performance is critical for enhancing personalized learning and improving educational outcomes. Traditional assessment methods, while useful, often fail to capture the complex factors influencing performance, such as socio-economic background and engagement metrics. This study explores the development of a predictive model using a machine learning algorithms to classify students' academic performance in higher institutions. By leveraging data collected from Department of Computer Science, Tai Solarin University of Education records, relevant features were selected using the mutual information method. The model was formulated and simulated using machine learning algorithm, Support Vector Machines (SVM) in the Google CoLaboratory environment. The model's predictive accuracy was evaluated based on key performance metrics, including accuracy, precision, and F-measure. Results indicate that the ensemble approach outperforms single-model methods by enhancing prediction robustness and reducing variance. This study demonstrates the effectiveness of machine learning techniques in identifying at-risk students early with SVM having 100% accuracy allowing for timely interventions and improved resource allocation. Moreover, it contributes to evidence-based decision-making in educational institutions, helping to optimize learning experiences and boost student retention rates.

**Keywords**: Academic Performance, Classification, Machine Learning, Model, Support Vector Machine

## INTRODUCTION

In educational institutions, understanding and predicting student performance play a crucial role in facilitating personalized learning, early intervention, and academic success (Deda *et al*, 2021). Traditionally, educators have relied on various assessment methods, such as exams, quizzes, and assignments, to evaluate student performance. While these methods offer valuable insights into students' understanding and progress, they often provide only a snapshot of their academic abilities and may not capture the complex interplay of factors influencing performance (kumar and Salal, 2019).

Machine learning (ML) techniques offer promising avenues for analyzing large volumes of educational data and uncovering patterns that may be difficult to discern through manual analysis alone (Romero and Ventura, 2020). By leveraging ML algorithms, researchers and educators can develop predictive models capable of classifying student performance based on various input variables, such as demographic information, previous academic records, and engagement metrics (Albeike *et al*, 2021). These models have the potential to enhance educational outcomes by identifying at-risk students (karalar *et al*, 2021), tailoring

**Cite as:**

Adedeji, O.B.; Olusanya, O.O.; Adebare, A.O.; Omilabu, A.A. (2024). Student Performulator: Classification and Prediction of Academic Performance of Students Using Machine Learning. *Journal of Science and Information Technology (JOSIT)*, Vol. 18 No. 2, pp. 115-125.

instructional interventions, and optimizing resource allocation (Ramaswami *et al*, 2019).

Machine learning approaches offer several advantages over predictive models based on a single machine learning algorithm (karalla *et al*, 2021). By combining multiple models, machine learning techniques can leverage the strengths of individual models while mitigating their weaknesses, thereby improving predictive accuracy and robustness (Hussain and khan, 2023). Moreover, machine learning methods can handle diverse types of data and modelling techniques, enabling a more comprehensive analysis of factors influencing prediction. Ensemble modelling is widely used in various machine learning tasks, including classification, regression, and clustering. It has been shown to improve predictive performance, reduce variance, and increase model robustness compared to single-model approaches. However, ensemble modelling requires careful tuning of hyperparameters, selection of diverse base models, and consideration of computational resources to achieve optimal results.

Institutions of higher learning are currently facing the challenging task of attracting new students who can effectively meet their diverse academic demands. With these demands come the need for those institutions to develop strategies that can enhance students' learning experiences at various educational levels. Predicting the academic success at an early stage would allow academic institutions to develop specific enrolment guidelines while avoiding poor performance. Traditional methods of assessing student performance, such as standardized tests and course grades, have several limitations. These methods often rely on summative assessments that provide retrospective insights into students' abilities but offer limited predictive power regarding future performance. Moreover, traditional assessments may not capture the full spectrum of students' skills, knowledge, and competencies, leading to incomplete or biased evaluations. Furthermore, traditional methods may overlook non-academic factors that influence student performance, such as socio-economic background, motivation, learning style, and mental health. Failing to account for these factors can result in inaccurate predictions and missed opportunities for intervention. Additionally, traditional assessments are often labour-intensive, time-consuming, and subject to human biases, making them less scalable and efficient for large-scale predictive modelling tasks.

This research addresses this gap by developing a predictive model using an ensemble of machine learning algorithms which can be used to classify the academic performance of students in higher institutions based on information about features associated with influencing academic performance.

Predicting student performance holds significant implications for both students and educational institutions. For students, early identification of academic challenges can lead to timely support interventions, personalized learning experiences, and improved outcomes. By identifying struggling students early on, educators can provide targeted interventions, such as tutoring, counselling, or additional resources, to address academic difficulties and prevent dropout.

## RELATED WORKS

Hussain and Khan (Baashar *et al*, 2022) , worked on the development of a student performance estimator using machine learning algorithms. The dataset considered in the study consisted of 90,000 secondary school student records consisting of information about features however all confidential information were removed from the dataset. The study adopted the use of generatic algorithm for the selection of relevant features which are most important in the determination of students' performance. The study adopted the use of kNN and decision trees algorithm for the development of the predictive model requred for estimating students' academic performance. The results showed that the use of feature selection of relevant features improved the performance of machine learning algorithms. Decision trees algorithm showed better performance by achieving an accuracy of 96.64%. The study was limited to the use of dataset collected from secondary school students.

Baashar, *et al*., (Owosu-Boadu *et al*, 2021) worked on the assesssment of the application of AI models for the assessment of the academic performance of postgraduate students in Malaysia. The study identified the various features that are associated with the prediction of students acadmeic performance such as: demographic infromation, program name, program structure, sponsorship, attendance and final CGPA. The model simulation involved the use of the holdout method based on the use of 90% for training and 10% for testing which was fed to artificial neural network (ANN), support vector machines (SVM), decision trees (DT), and Guassian process regression. The results revealed that the best performance was achieved using ANN with an $R^2$ value of 0.89 and mean squared error (MSE) of 0.080. The study was limited to a regression task as it was focused on estimating the value of the students' CGPA and data colected from postgraduate students.

(Yağcı, 2021) worked on the application of machine learning algorithms to the prediction of the acadmeic performance of Turkish students. The study collected data about students taking a course in a Turkish university consisting of mid-term exam grades, department data and faculty data were used for predicting the final grade of the course. The study fed the dataset to a number of machine learning algorithms, namely: kNN, SVM, logistic regression (LR), random forest (RF) and naïve Bayes (NB). The results of the study revealed that the best performance was achieved using random forest with an accuracy of 74.6%. The study concluded that the ensemble model performed better than other machine learning algorithms. The study was limited to the prediction of the performance of student taking a particular course based on the comparative analysis of machine learning algorithms.

Owosu-Boadu *et al*, 2021), worked on the assessment of the academic performance of students in Ghana using machine learning algorithms. The study collected data fron third year students of three secondary schools located in Ghana based on a number of identified features. The features included demographic features such as gender, nationalit, place of birth, level, class group, topic, term, relation to guardian, class participation, library visits, involvement in group discussions, parent survey responses, parent satisfaction and student absence days. The model was formulated using kNN, DT, ANN, RF, SVM, LR and AdaBoost. The results revealed that random forest (RF) showed the best performance among all the selected algorithms with an accuracy of 82%. The results concluded that ensemble models performed better than machine learning algorithms. The study was limited to data collected from secondary schools.

## METHODOLOGY

Relevant data containing information about the features that are associated with the assessment of the academic performance of students was collected from the departmental records. Table 1 provides a description of the features that were considered for the classification of academic performance. The features in the dataset collected were subjected to feature selection using the mutual information method. The ensemble model for the classification of academic performance was formulated using a number of machine learning algorithms based on information about the features. Predictive models were simulated by using the holdout method based on 5 simulation runs for each machine learning algorithm such that the training dataset was used to build the model using the Google CoLaboratory; a Python Jupyter notebook for Gmail users. The models were evaluated using on a number of performance metrics, namely: accuracy, true positive (TP) rate, false positive (FP) rate, precision and f-measure based on the test dataset.

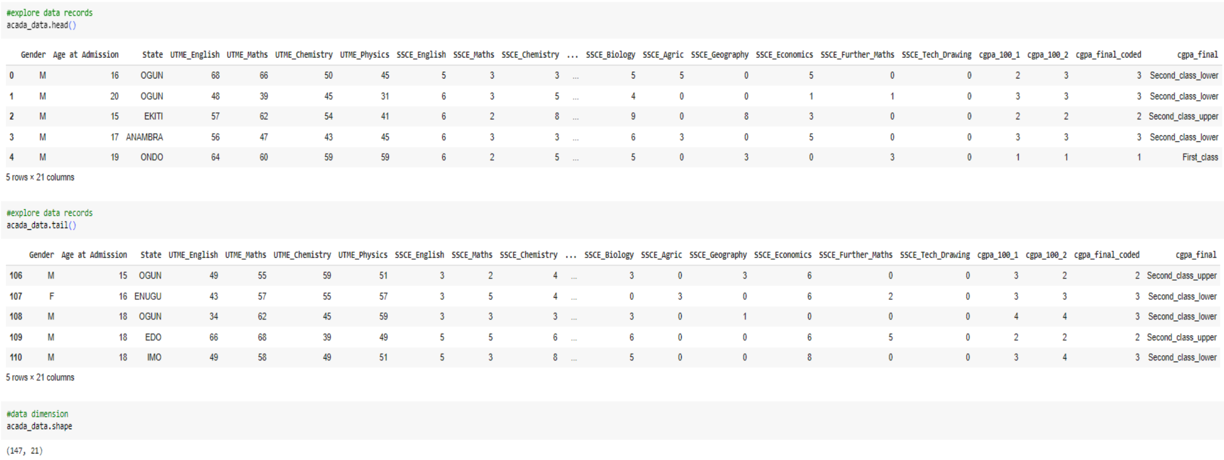
**Table 1.** Identification of features associated with credit worthiness.

| Class of Variable | Name | Label values |
|---|---|---|
| Socio-Demographic Information | Gender | Categorical (Male, Female) |
| | Age at Admission | Numeric – Integer type |
| | State of Origin | Categorical |
| UTME Results | English | Numeric – Integer Type |
| | Mathematics | Numeric – Integer Type |
| | Chemistry | Numeric – Integer Type |
| | Physics | Numeric – Integer Type |
| O'Level Results (SSCE) | English | Numeric – Integer type |
| | Mathematics | Numeric – Integer type |
| | Chemistry | Numeric – Integer type |
| | Physics | Numeric – Integer type |
| | Biology | Numeric – Integer type |
| | Agricultural Science | Numeric – Integer type |
| | Geography | Numeric – Integer type |
| | Economics | Numeric – Integer type |
| | Further Mathematics | Numeric – Integer type |
| | Technical Drawing | Numeric – Integer type |
| 100 Level Results | First Semester CGPA | Numeric – Float type |
| | Second Semester CGPA | Numeric – Float type |
| Target Class | Graduating Class of Degree | Categorical (First Class, Second Class Upper, Second Class Lower, Third Class) |

## RESULTS/DISCUSSIONS

### Results

The results of the exploration of the numerical and categorical features within the dataset was presented using appropriate tools such as tables, bar charts and box plots. Afterwards, the results of the transformation of the categorical string-type valued features into numeric types was presented alongside the assessment of the feature importance of the features. Finally, the results of the simulation and evaluation of the comparative analysis of the adoption of the machine learning and machine learning models was determined based on a number of performance evaluation metrics.



```
#explore data records
acada_data.head()
```

| | Gender | Age at Admission | State | UTME_English | UTME_Maths | UTME_Chemistry | UTME_Physics | SSCE_English | SSCE_Maths | SSCE_Chemistry | ... | SSCE_Biology | SSCE_Agric | SSCE_Geography | SSCE_Economics | SSCE_Further_Maths | SSCE_Tech_Drawing | cgpa_100_1 | cgpa_100_2 | cgpa_final_coded | cgpa_final |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | M | 16 | OGUN | 68 | 66 | 50 | 45 | 5 | 3 | 3 | ... | 5 | 5 | 0 | 5 | 0 | 0 | 2 | 3 | 3 | Second_class_lower |
| 1 | M | 20 | OGUN | 48 | 39 | 45 | 31 | 6 | 3 | 5 | ... | 4 | 0 | 0 | 1 | 1 | 0 | 3 | 3 | 3 | Second_class_lower |
| 2 | M | 15 | EKITI | 57 | 62 | 54 | 41 | 6 | 2 | 8 | ... | 9 | 0 | 8 | 3 | 0 | 0 | 2 | 2 | 2 | Second_class_upper |
| 3 | M | 17 | ANAMBRA | 56 | 47 | 43 | 45 | 6 | 3 | 3 | ... | 6 | 3 | 0 | 5 | 0 | 0 | 3 | 3 | 3 | Second_class_lower |
| 4 | M | 19 | ONDO | 64 | 60 | 59 | 59 | 6 | 2 | 5 | ... | 5 | 0 | 3 | 0 | 3 | 0 | 1 | 1 | 1 | First_class |

5 rows × 21 columns

```
#explore data records
acada_data.tail()
```

| | Gender | Age at Admission | State | UTME_English | UTME_Maths | UTME_Chemistry | UTME_Physics | SSCE_English | SSCE_Maths | SSCE_Chemistry | ... | SSCE_Biology | SSCE_Agric | SSCE_Geography | SSCE_Economics | SSCE_Further_Maths | SSCE_Tech_Drawing | cgpa_100_1 | cgpa_100_2 | cgpa_final_coded | cgpa_final |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 106 | M | 15 | OGUN | 49 | 55 | 59 | 51 | 3 | 2 | 4 | ... | 3 | 0 | 3 | 6 | 0 | 0 | 3 | 2 | 2 | Second_class_upper |
| 107 | F | 16 | ENUGU | 43 | 57 | 55 | 57 | 3 | 5 | 4 | ... | 0 | 3 | 0 | 6 | 2 | 0 | 3 | 3 | 3 | Second_class_lower |
| 108 | M | 18 | OGUN | 34 | 62 | 45 | 59 | 3 | 3 | 3 | ... | 3 | 0 | 1 | 0 | 0 | 0 | 4 | 4 | 3 | Second_class_lower |
| 109 | M | 18 | EDO | 66 | 68 | 39 | 49 | 5 | 5 | 6 | ... | 6 | 0 | 0 | 6 | 5 | 0 | 2 | 2 | 2 | Second_class_upper |
| 110 | M | 18 | IMO | 49 | 58 | 49 | 51 | 5 | 3 | 8 | ... | 5 | 0 | 0 | 8 | 0 | 0 | 3 | 4 | 3 | Second_class_lower |

5 rows × 21 columns

```
#data dimension
acada_data.shape
```

(147, 21)

**Figure 1.** Screenshot of visualization of contents of the collected dataset.

Figure 1 shows a screenshot of the description of the datasets showing the values of the features that were identified in the dataset collected for the purpose of this study. According to the figure, it was shown that majority of the features were stored using numeric values while the features: gender and state were stored as categorical string type values.

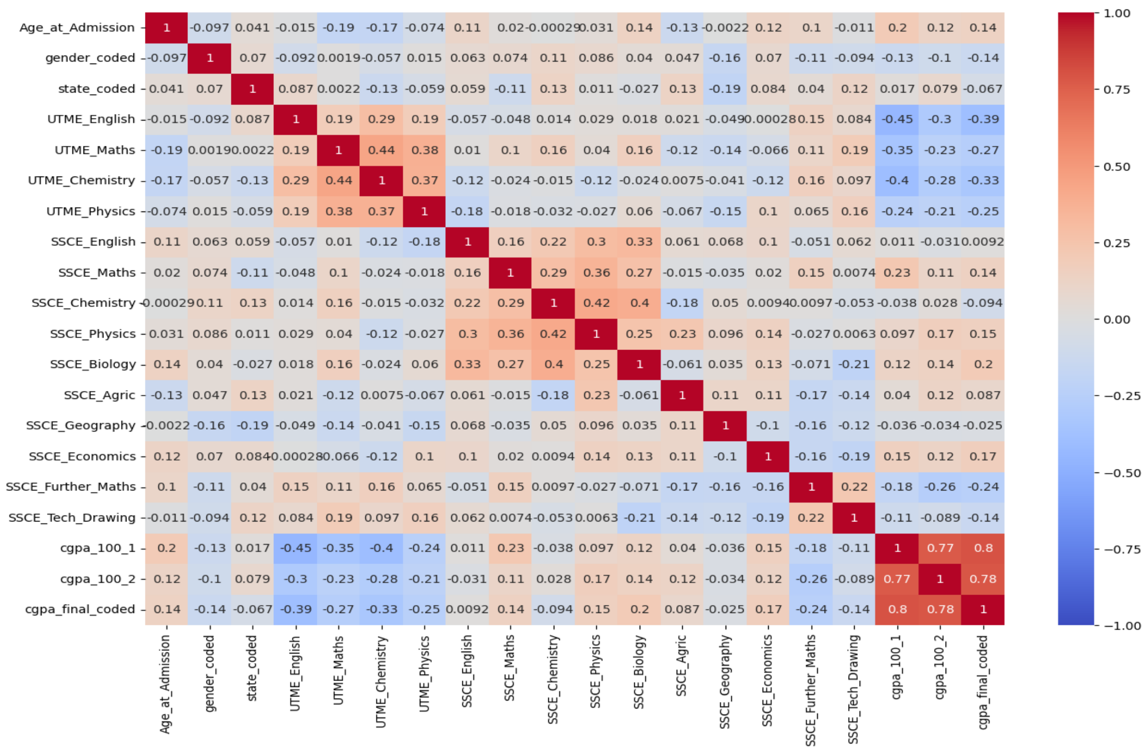



Figure.2: Visualization of heatmap of feature-feature intercorrelation.

Figure 2 displays the value of the correlation of the features with respect to one another such that darker colours reflect higher correlation while light colours reflect lower correlations. Also, red colours signified negative correlation and blue colour signified positive correlation.

However, since the focus of the study is on the association between the features and the classification of academic performance among students, the values displayed in the last row which was called *cgpa_final*, was considered. The values in the cell of the last row shows the correlation of the features with respect to the classification of the academic performance of students. On the other hand, figure 6 shows the graphical plot of feature importance in decreasing order based on the mutual information metric. As shown in figure 6, the mutual information revealed the amount of information about the classification of academic performance among students that can be explained by each feature identified in the dataset based on the information collected about them in the dataset. Table 2 gives a summary of the ranking of features based on the value of their Pearson's correlation and mutual information.

**Table 2.** Identification of Feature Importance.

| S/N | Pearson's Correlation | | Mutual Information | |
|---|---|---|---|---|
| | **Feature Name** | **Value** | **Feature Name** | **Value** |
| 1 | CGPA 100 Level First | 0.8000 | CGPA 100 Level First | 0.459060 |
| 2 | CGPA 100 Level Second | 0.7800 | CGPA 100 Level Second | 0.404737 |
| 3 | UTME-English | -0.3900 | UTME-English | 0.193618 |
| 4 | UTME-Chemistry | -0.3300 | SSCE-Physics | 0.147297 |
| 5 | UTME-Mathematics | -0.2700 | UTME-Chemistry | 0.134541 |
| 6 | UTME-Physics | -0.2500 | UTME-Physics | 0.068669 |
| 7 | SSCE-Further Maths | -0.2400 | SSCE-Further Maths | 0.041710 |
| 8 | SSCE-Biology | 0.2000 | SSCE-Mathematics | 0.021190 |
| 9 | SSCE-Economics | 0.1700 | SSCE-Agric | 0.00000 |
| 10 | SSCE-Physics | 0.1500 | SSCE-Tech Drawing | 0.00000 |
| 11 | Age at Admission | 0.1400 | SSCE-Economics | 0.00000 |
| 12 | SSCE-Mathematics | 0.1400 | SSCE-Geography | 0.00000 |
| 13 | Gender | -0.1400 | Age at Admission | 0.00000 |
| 14 | SSCE-Tech Drawing | -0.1400 | SSCE-Biology | 0.00000 |
| 15 | SSCE-Chemistry | -0.0940 | Gender | 0.00000 |
| 16 | SSCE-Agric | 0.0870 | SSCE-English | 0.00000 |
| 17 | State | -0.0670 | UTME-Mathematics | 0.00000 |
| 18 | SSCE-Geography | -0.0250 | State | 0.00000 |
| 19 | SSCE-English | 0.0092 | SSCE-Chemistry | 0.00000 |

According to the Pearson's correlation coefficient, it was revealed that the two most important features were *CGPA 100 Level First* and *CGPA 100 Level Second* both with positive correlation, followed by *UTME-English*, *UTME-Chemistry*, *UTME-Mathematics*, *UTME-Physics*, and *SSCE-Further Maths* all with negative correlation, followed by *SSCE-Biology*, *SSCE-Economics*, *SSCE-Physics*, *Age at Admission*, and *SSCE-Mathematics* with positive correlation. Features with the least correlation include: *Gender*, *SSCE-Tech Drawing*, and *SSCE-Chemistry* with negative correlation followed by *SSCE-Agric* with positive correlation followed by *State, and SSCE-Geography* with negative correlation and the least correlation was found in *SSCE-English* with positive correlation.

## Discussion

This section presents the results of the evaluation of the predictive models that were generated across the five simulations based on the machine learning and ensemble modelling techniques that were adopted in this study. The results are presented for each simulation following which the results of the performance of the algorithms were presented.

## Results of the Simulation of Predictive Model

This section presents the results of the application of the machine learning algorithm, namely: support vector machines (SVM) classifier. The model simulation was conducted by splitting the dataset into two parts, train and test dataset using five simulations such that 50/50, 60/40, 70/30, 80/20 and 90/10 percent of the dataset was adopted for training/testing the predictive model. Table 4.2 shows the number of records that were adopted for each simulation that were considered in this study. As stated earlier, the train datasets were used to build the predictive model while the test data were used to evaluate the performance of the predictive models based on a number of performance evaluation metrics.

**Table 3.** Description of the number of records adopted for training and testing predictive models across five simulations.

| Simulation# | Train Data | | | | | Test Data | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **2.1** | **2.2** | **First** | **Third** | **Total** | **2.1** | **2.2** | **First** | **Third** | **Total** |
| Simulation 1 (50/50) | 28 | 19 | 3 | 11 | **61** | 15 | 12 | 9 | 14 | **50** |
| Simulation 2 (60/40) | 32 | 21 | 4 | 14 | **71** | 11 | 7 | 11 | 11 | **40** |
| Simulation 3 (70/30) | 37 | 26 | 2 | 16 | **81** | 6 | 8 | 9 | 7 | **30** |
| Simulation 4 (80/20) | 42 | 31 | 3 | 15 | **91** | 5 | 6 | 4 | 5 | **20** |
| Simulation 5 (90/10) | 49 | 38 | 4 | 20 | **101** | 1 | 4 | 2 | 3 | **10** |

Figure 3 shows the confusion matrices that were used to interpret the results of the evaluation of the machine learning model adopted in simulation 1 based on the test dataset. Using SVM classifier, it was observed that all 15 actual second-class lower records were correctly classified, all 12 actual second-class lower records were correctly classified, all 9 actual first-class records were correctly classified and all 14 actual third-class records were correctly classified owing to an accuracy of 100.0%.

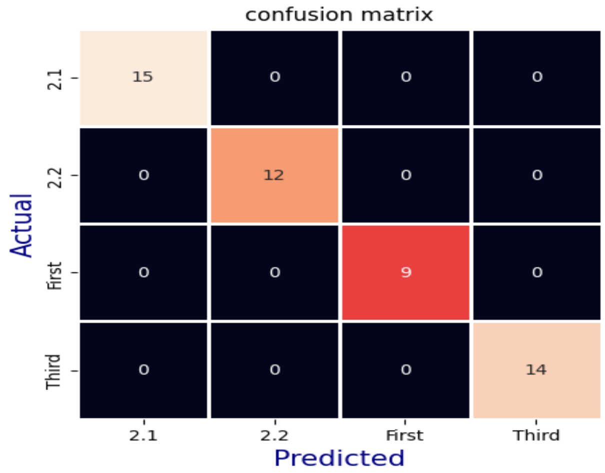



**Figure 3.** Confusion matrices for the evaluation of Support Vector Machines (center) for simulation 1.

Figure 4 shows the confusion matrices that were used to interpret the results of the evaluation of machine learning model adopted in simulation 2 based on the test dataset. Using SVM classifier, it was observed that all 11 actual second-class lower records were correctly classified, all 7 actual second-class lower records were correctly classified, all 11 actual first-class records were correctly classified and all 11 actual third-class records were correctly classified owing to an accuracy of 100.0%.

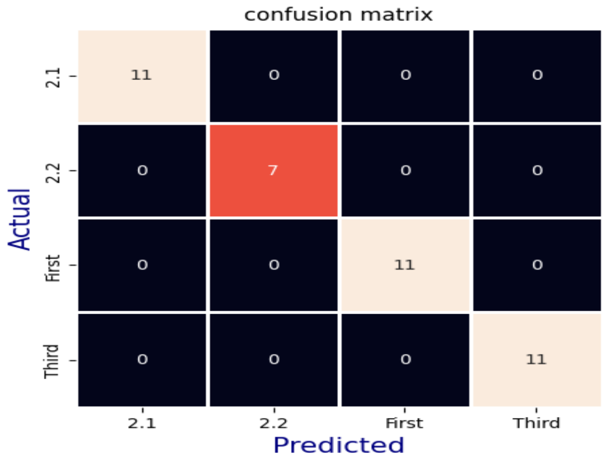



**Figure 4.** Confusion matrices for the evaluation of Support Vector Machines for simulation 2.

Figure 5 shows the confusion matrices that were used to interpret the results of the evaluation of both machine learning models adopted in simulation 3 based on the test dataset. Using SVM classifier, it was observed that all 6 actual second-class lower records were correctly classified, all 8 actual second-class lower records were correctly classified, all 9 actual first-class records were correctly classified and all 7 actual third-class records were correctly classified owing to an accuracy of 100.0%.

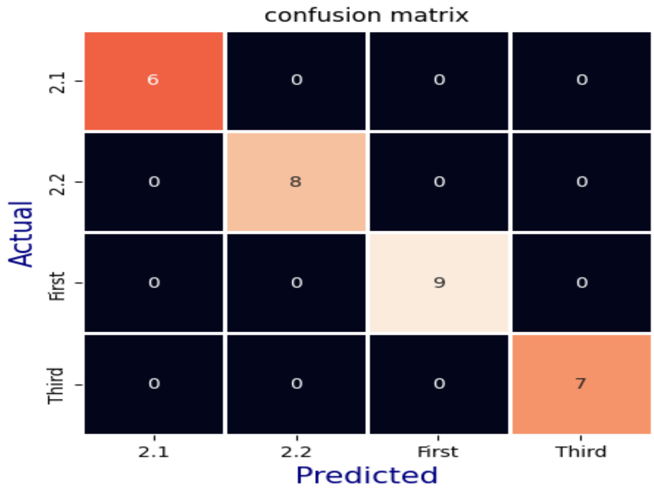



**Figure 5.** Confusion matrices for the evaluation of Support Vector Machines for simulation 3.

Figure 6 shows the confusion matrices that were used to interpret the results of the evaluation of both machine learning models adopted in simulation 4 based on the test dataset. Using SVM classifier, it was observed that all 5 actual second-class lower records were correctly classified, all 6 actual second-class lower records were correctly classified, all 4 actual first-class records were correctly classified and all 5 actual third-class records were correctly classified owing to an accuracy of 100.0%.

5Published by The College of Science & Information Technology (COSIT), TASUED, Vol. 18, No. 2, pp. 115-125.

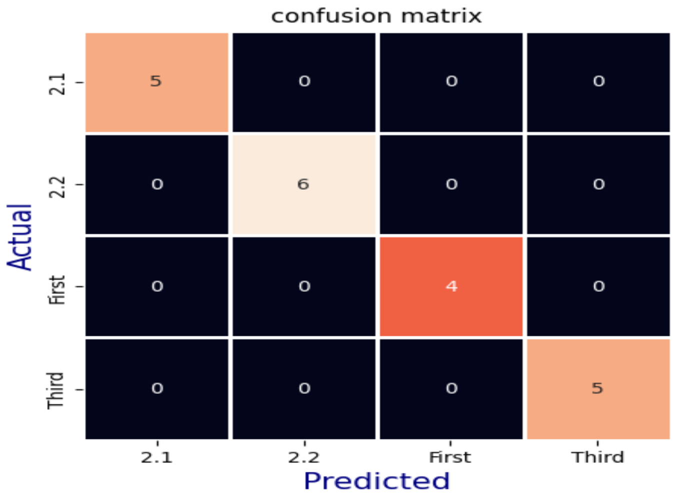



**Figure 6.** Confusion matrices for the evaluation of Support Vector Machines for simulation 4.

Figure 7 shows the confusion matrices that were used to interpret the results of the evaluation of both machine learning models adopted in simulation 2 based on the test dataset. Using SVM classifier, it was observed that all 1 actual second-class lower records were correctly classified, all 4 actual second-class lower records were correctly classified, all 2 actual first-class records were correctly classified and all 3 actual third-class records were correctly classified owing to an accuracy of 100.0%.

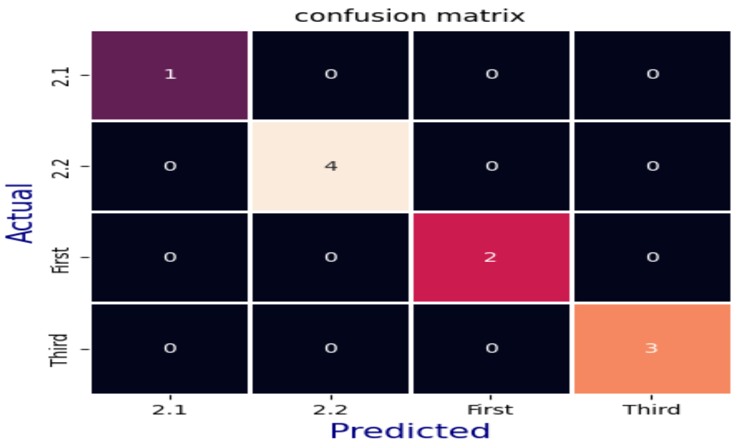



**Figure 7.** Confusion matrices for the evaluation of Support Vector Machine for simulation 5.

**Table 4.** Results of the evaluation of the predictive models across five simulations based on performance metrics.

| Simulation# | Algorithm | Correct Records | Accuracy (%) | Precision | | | | Recall | | | | F1-Score | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 2.1 | 2.2 | First | Third | 2.1 | 2.2 | First | Third | 2.1 | 2.2 | First | Third |
| Simulation 1 | SVM | 50 | 100.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Simulation 2 | SVM | 40 | 100.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Simulation 3 | SVM | 30 | 100.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Simulation 4 | SVM | 20 | 100.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Simulation 5 | SVM | 10 | 100.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

## CONCLUSION

The study examined the performance of a machine learning algorithms: Support Vector Machine in classification and predicting the performance of students. The datasets were obtained from the Department of Computer Science Tai Solarin University of Education. The model simulation was conducted by splitting the dataset into two parts, train and test dataset using five simulations such that 50/50, 60/40, 70/30, 80/20 and 90/10 percent of the dataset was adopted for training/testing the predictive model. The study concluded that machine learning model is very effective in the classification of the academic performance of students, especially the Support Vector classifiers that out performed with 100% accuracy.

## REFERENCES


Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Review of Student's Performance Prediction using Machine Learning Techniques. *Education Sciences*, *11*(9), 1-12.

Baashar, Y., Hamed, Y., Alkawsi, G., Capretz, L., Alhussian, H., Alwadain, A., & Al-Amri, R. (2022). Evaluation of Postgraduate Academic Performance using Artificial Intelligence Models. *Alexandria University Journal, 61*, 9867-9878.

Deda, E., Pacukaj, E., & Vardari, L. (2021). Education and Its role in the Economic Development of the COuntry and Government Policies to be undertaken to Increase the Quality of Education: The Case of Albania. *Journal of*

*Educational and Social Research, 11*(1), 188-199.

Hussain, S., & Khan, M. (2023). Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level using Machine Learning. *Annals of Data Science, 10*(3), 637-655.

Karalar, H., Kapucu, C., & Gürüler, H. (2021). Predicting Students at Risk of Academic Failure using Ensemble Model during Pandemic in a Distance Learning System. *International Journal of Educational Technology in Higher Education, 18*(1), 1-13.

Kumar, M., & Salal, Y. (2019). Systematic Review of predicting Student's Perfomance in Academics. *International Journal of Engineering and Advanced Technology, 8*(3), 54-61.

Owosu-Boadu, B., Nti, I., Nyarko-Boateng, O., Aning, J., & Boafo, V. (2021). Academic Performance MOdelling with Machine Learning based on Cognitive and Non-Cognitive Features. *Applied Computer Systems, 26*(2), 122-131.

Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J., & Garcia, P. (2019). Using Educational Data Mining Techniques to increase the Prediction Accuracy of Student Academic Performance. *Information and Learning Science, 120*, 451-467.

Romero, C., & Ventura, S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. *Wiley Interdisciplinary Reviws: Data Mining and Knowledge Discovery, 10*(3), 1-21.

Yağcı, M. (2022). Educational Data Mining: Prediction of Students' Academic Performance using Machine Learning Algorithms. *Smart Learning Environments, 9*(11), 1-19.