JoCoSiR
Journal of Computer Science Research
ISSN 2986-2337 (Online)

Classification of Student Activity Status Using Machine Learning Algorithms at Royal University

Rudi Hermawan [a], Muhammad Iqbal [b]

[a,b]Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia

email: [a] rudirk60@gmail.com, [b] wakbalpb@yahoo.co.id

ARTICLE INFO

Keywords:
Student Activity Classificatio; Decision Tree; Support Vector Machine; Random Forest; Student Retention Prediction

IEEE style in citing this article:
R. Hermawan and M. Iqbal, "Classification of Student Activity Status Using Machine Learning Algorithms at Royal University," *JoCoSiR: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 3, no. 1, pp. 1-9, 2025.

ABSTRACT

Inactivity is a significant challenge that impacts academic performance, retention rates, and the operational effectiveness of higher education institutions. Royal University faces an urgent need to identify students at risk of becoming inactive early, so that academic interventions can be carried out appropriately and effectively. This study aims to develop a classification model for student inactivity status (Active or Passive) using a machine learning approach, by testing three main algorithms: Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF). The dataset used consists of 642 student entries, including academic information such as Grade Point Average (GPA), total credits taken, attendance percentage, number of courses per semester, and semester level. The methodology steps include data cleaning and transformation, splitting the dataset into 80% training data and 20% testing data using a random sampling method (train_test_split with random_state = 42), model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The experimental results show that DT and SVM achieve the highest accuracy of 98.44%, with maximum precision in predicting active students, while RF excels in recall (0.96), making it more effective in detecting active students at risk of being missed. Feature importance analysis reveals that GPA and attendance are the most determining factors in predicting student active status, while the number of courses, credits taken, and semester level have a lower additional influence. The primary contribution of this research is the provision of an accurate and practically applicable classification model, enabling universities to conduct automated student monitoring, proactive academic interventions, and data-driven decision-making. Implementing this model in academic information systems can improve the effectiveness of advising programs, reduce the risk of student inactivity , and support efforts to improve retention and graduate quality. This research also emphasizes the importance of contextual features in improving prediction accuracy and provides insights that can be leveraged for the development of data-driven academic strategies.

1. Introduction

Student activity is an important indicator in assessing the quality of higher education because it directly influences academic success and graduation rates [1], [2]. The level of student participation in academic and non-academic activities reflects their engagement in the learning process, which ultimately determines academic outcomes. Active students generally demonstrate higher motivation, better time management, and stronger adaptability to the campus environment. Therefore, student activity serves as one of the benchmarks used by universities to evaluate the effectiveness of their educational systems.

However, many higher education institutions still face challenges in comprehensively and continuously monitoring student activity [3]–[6]. Limitations in academic information systems, lack of data integration, and the absence of early detection mechanisms make it difficult for universities to identify students at risk of dropping out. Yet, the ability to detect such risks early is crucial for institutions to provide appropriate support and interventions. This situation highlights the need for innovation in monitoring systems and academic data analysis that are adaptive and powered by intelligent technologies.

Specifically, Royal University has not yet developed a system capable of accurately and timely predicting student activity status [7], [8]. The absence of such a predictive system prevents the university from proactively identifying students who are experiencing a decline in academic engagement. As a result, interventions such as academic advising, counseling, or learning assistance cannot be implemented effectively. This limitation hinders the university's efforts to enhance student success and retention rates. Therefore, developing an artificial intelligence–based system represents a strategic step toward addressing these challenges.

Previous studies have demonstrated the effectiveness of machine learning algorithms in predicting students' academic status. Alnasyan et al. utilized deep learning approaches by considering demographic factors, previous academic performance, and learning behavior [1]. Kabathova et al. applied machine learning algorithms to predict

ISSN 2502-3357 (Online)
dropout potential with an accuracy of 93% [2]. Meanwhile, Mustofa et al. combined logistic regression and neural networks to estimate student dropout risk with an accuracy of 96% [3]. These studies show that applying machine learning algorithms holds significant potential to support decision-making in higher education management.

Based on this background, the present study proposes a classification model for predicting student activity status using three main algorithms: Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) [4]–[9]. The model is developed using a comprehensive academic dataset and includes non-academic features such as attendance rate and the number of enrolled courses [10]–[12]. This approach is expected to produce an accurate, efficient, and practical prediction system for university environments. With this model, institutions such as Royal University can effectively monitor student activity and implement early interventions to improve academic achievement and reduce dropout rates.

2. State of the Art

2.1 Classification of Student Status with Conventional Algorithms

Previous studies have used conventional algorithms such as logistic regression, SVM, and DT for academic status classification. These methods are effective on limited datasets but have limitations in handling complex features [1], [2], [4].

2.2 Performance Improvement Using *Ensemble Methods*

Ensemble methods, such as Random Forest, have been shown to improve predictive accuracy compared to single algorithms. Kaur et al. showed that combining multiple models can identify high- and low-risk students more accurately [3], [5], [6].

2.3 The Role of Non-Academic and Contextual Features

Besides GPA, attendance and number of courses are important indicators in predicting active status [7]–[9]. Recent research emphasizes the importance of combining academic and non-academic features to improve model performance [10], [11].

2.4 Research Gaps and Contributions

Although many studies have successfully predicted academic status, there are still shortcomings related to comprehensive datasets and analysis of non-academic features [12]–[16]. This study offers novelty in the form of integrating complete academic data and applying RF, DT, and SVM algorithms to predict student activity status [17]–[20].

The literature review shows that although much research has been done in this domain, there are several *gaps* that this study addresses:

a. Institutional Specificity : Most studies use data from overseas universities or very general data. This study uses data specific to the Royal University (N=642), allowing for the discovery of unique behavioral patterns within that institution.
b. Focus on Feature Combination: This study comprehensively integrates the main feature dimensions (Pure Academic: GPA, Study Load: Credits, Behavior: Attendance, and Situational: Semester and Number of Courses) in a single classification model.
c. Focus on *Recall of* Minority Classes: Given the significant impact of student inactivity , focusing on Accuracy alone is insufficient. This study places significant weight on the *Recall metric* for "Inactive" classes to ensure that the early warning system has high sensitivity and minimizes *False Negatives* (unidentified at-risk students).

3. Method

This study adopts a quantitative approach using the *Knowledge Discovery in Databases* (KDD) methodology to build and evaluate a classification model of student activity status.

3.1 Data Acquisition and Preprocessing

3.1.1 Data Sources

This research dataset is sourced from the Academic Information System (AIS) of Royal University. The total number of student data analyzed is N = 642 entries, covering the history of 8th semester students of the 2025 Academic Year intake. The activity status is categorized as "Active" (Majority) and "Inactive" (Minority, including students on leave without permission or *drop-out*).

3.1.2 Features and Labels

The variables used in modeling consist of independent variables (features) and dependent variables (labels):

a. Label (Dependent Variable): Student Activity Status (Binary: 1 = Active, 0 = Inactive/Passive).
b. Features (Independent Variables):
    1. F1: Cumulative GPA at the End of Semester
    2. F2: Total Credits Taken
    3. F3: Percentage of Total Course Attendance

DOI: https://doi.org/10.65126/jocosir.v3i1

ISSN 2502-3357 (Online)

4. F4: Semester (Semester when data was taken)
5. F5: Number of Courses (20251) (Study load in the relevant semester)

3.1.3 Data Cleaning, Transformation, and Data Sharing

The data cleaning stage involves handling missing values *and* ensuring a uniform data format. Due to the imbalanced dataset , the *Synthetic Minority Over-sampling Technique* (SMOTE) will be considered to improve the representation of the minority class ("Inactive") during the training stage.
Data Division: Before modeling is carried out , the data is divided into two parts with the following proportions:

a. 80% (513 data) as training data *for* the model learning process.
b. 20% (129 data) as testing data *to measure model performance. The data* splitting process is done randomly using the train_test_split ( ) function with the random_state parameter = 42 so that the results can be replicated consistently.

3.2 Modeling and Algorithms

Machine Learning algorithms were used to compare classification performance:

3.2.1 Decision Tree (DT)

DT is a non- parametric model that uses a series of feature-based decision rules to classify data. We use the Gini Impurity criterion to break down nodes, limiting the maximum depth of the tree to prevent *overfitting*.

3.2.2 Support Vector Machine (SVM)

SVM works by finding the optimal *hyperplane* that maximizes the separation margin between classes in the feature space. We will test nonlinear *kernels such as the Radial Basis Function* (RBF) to handle complex data separations, which are common in educational data.

3.2.3 Random Forest (RF)

RF is an *ensemble method* that constructs a large number (N) of decision trees independently and combines their predictions (voting) to produce a final classification. RF inherently reduces the variance and *overfitting* experienced by a single DT. RF is well-suited for multidimensional data and interacting features.

3.3 Model Evaluation

Evaluation was conducted using K-Fold *Cross-Validation (CV) (k=10) on the training data and tested on the Testing Data* (129 data points) to ensure the model has good generalization. The evaluation metrics used are based on *the Confusion Matrix* (TP, TN, FP, FN):

a. **Accuracy :** The ratio *of* correct predictions to total predictions.

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

b. **Precision :** *The* proportion of positive outcomes that are correctly predicted (important to minimize incorrect interventions).

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

c. **Recall (*Recall*):** The proportion of actual positives correctly identified (important to minimize missing at-risk students).

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

d. **F1 -*Score* :** The harmonic mean of Precision and *Recall* (an indicator of model balance, especially important on *imbalanced datasets*).

ISSN 2502-3357 (Online)

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Receiver Operating Characteristic (ROC) curve analysis and *Area Under the Curve (AUC)* calculations were also used to measure the discriminatory performance of the model across classification thresholds.

4. Results and Discussion

This section presents the experimental results of the student activity status classification modeling , a comparison of algorithm performance, and an in-depth discussion of the significance of the findings.

4.1 Experimental Results and Model Comparison

The dataset used comes from the Royal University Academic Information System, with 642 data points and 9 attributes. The data is named ' data_mahasiswa.csv', and the data attributes are shown in Figure 1.

```
import pandas as pd

# Baca ulang dataset dengan pemisah titik koma (;)
df = pd.read_csv('data_mahasiswa.csv', sep=';')

# Cek struktur kolom
print(df.columns)
df.head()
```

```
Index(['NIM', 'Nama', 'Semester', 'SKS Ditempuh', 'Kehadiran', 'IPK',
       'Jumlah_Matkul_20251', 'Status'],
      dtype='object')
```

	NIM	Nama	Semester	SKS Ditempuh	Kehadiran	IPK	Jumlah_Matkul_20251	Status
0	21210001	MUHAMMAD IQBAL SETIAWAN	8	148	91.57	3.70	3	Aktif
1	21210002	MHD ALFUN KHOIR SIREGAR	8	148	90.85	3.33	3	Aktif
2	21210004	FRANSDIKA	8	63	95.20	2.79	0	Pasif
3	21210005	JEPRI SARAGIH	8	148	85.60	3.42	3	Aktif
4	21210006	ALDI DERMAWAN	8	148	93.71	3.56	3	Aktif

Figure 1. Dataset

The program above displays the first five rows of data using ' df.head ()'. The columns in this dataset include various student data information such as Student ID Number, Name, Semester, Credits Taken, Attendance, GPA, Total_Course_20251, Status.

DOI: https://doi.org/10.65126/jocosir.v3i1

ISSN 2502-3357 (Online)

```
# 7 Model: Decision Tree
dt = DecisionTreeClassifier(random_state=42)
dt.fit(X_train, y_train)
y_pred_dt = dt.predict(X_test)
acc_dt = accuracy_score(y_test, y_pred_dt)
print("\n=== Decision Tree ===")
print("Akurasi:", acc_dt)
print(classification_report(y_test, y_pred_dt))

# 8 Model: Random Forest
rf = RandomForestClassifier(random_state=42)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)
acc_rf = accuracy_score(y_test, y_pred_rf)
print("\n=== Random Forest ===")
print("Akurasi:", acc_rf)
print(classification_report(y_test, y_pred_rf))

# 9 Model: SVM
svm = SVC(kernel='rbf', random_state=42)
svm.fit(X_train, y_train)
y_pred_svm = svm.predict(X_test)
acc_svm = accuracy_score(y_test, y_pred_svm)
print("\n=== SVM ===")
print("Akurasi:", acc_svm)
print(classification_report(y_test, y_pred_svm))
```

Figure 2. Model Prediction Program

Figure 2 shows a program for predicting three models: Decision Tree, Random Forest, and Support Vector Machine (SVM). This process aims to check how the trained models predict test data that has never been used in the model training.

```
=== Decision Tree ===
Akurasi: 0.9844961240310077
              precision    recall  f1-score   support

           0       0.98      1.00      0.99       106
           1       1.00      0.91      0.95        23

    accuracy                           0.98       129
   macro avg       0.99      0.96      0.97       129
weighted avg       0.98      0.98      0.98       129


=== Random Forest ===
Akurasi: 0.9767441860465116
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       106
           1       0.92      0.96      0.94        23

    accuracy                           0.98       129
   macro avg       0.95      0.97      0.96       129
weighted avg       0.98      0.98      0.98       129


=== SVM ===
Akurasi: 0.9844961240310077
              precision    recall  f1-score   support

           0       0.98      1.00      0.99       106
           1       1.00      0.91      0.95        23

    accuracy                           0.98       129
   macro avg       0.99      0.96      0.97       129
weighted avg       0.98      0.98      0.98       129
```

Figure 3. Model Evaluation Results

DOI: https://doi.org/10.65126/jocosir.v3i1

ISSN 2502-3357 (Online)
Machine Learning models, namely Decision Tree, Random Forest, and Support Vector Machine (SVM), and the evaluation results can be explained in Table 1.
Table 1 presents the evaluation results of the three machine learning models. The Decision Tree and SVM models both achieved the highest accuracy, at 98.44 % , while the Random Forest achieved an accuracy of 97.67%. These results indicate that the SVM and Decision Tree models are the most appropriate for predicting student engagement.

Table 1: Comparison of *Machine Learning Model Classification Performance*

Model	Accuracy	Precision (Active)	Recall (Active)	F1-Score (Active)
Decision Tree	98.44%	1.00	0.91	0.95
Random Forest	97.67%	0.92	0.96	0.94
SVM	98.44%	1.00	0.91	0.95

a. Accuracy
   1. Decision Tree And SVM has the highest accuracy, namely 98.44% , whereas Random Forest slightly lower on 97.67% .
   2. This shows that overall, all three models are able to predict student activity status with a very high level of success.
b. Precision (Active)
   1. Precision measures the proportion of correct “Active” predictions out of all model predictions classified as Active.
   2. Decision Tree And SVM own precision 1.00 , meaning that every student predicted to be active by these two models is truly active.
   3. Random Forest have precision 0.92 , slightly lower, which indicates that there are some cases of students predicted to be active but actually passive.
c. Recall (Active)
   1. Recall measures the model's ability to find all truly active students.
   2. Random Forest excel with recall 0.96 , higher than DT and SVM (0.91).
   3. This means that Random Forest is more effective in detecting active students, although there is a slight compromise in precision.
d. F1-Score (Active)
   1. F1-Score is the harmonic mean of precision and recall, balancing both.
   2. Decision Tree and SVM has F1-Score 0.95 , whereas Random Forest slightly lower at 0.94 .
   3. This shows that the overall performance of DT and SVM is slightly superior in maintaining the balance between precision and recall.
e. Overall Interpretation
   1. Decision Tree and SVM excel in precision , meaning that these models are very reliable when predicting active students, predictions are rarely wrong.
   2. Random Forest excels in recall , meaning fewer active students are missed (low false negatives).
   3. Model choice depends on priority:
      - If avoiding prediction errors active students important → Decision Tree/SVM is more appropriate.
      - If you catch as many active students as possible important → Random Forest is better.

4.2 Feature Importance *Analysis*

Feature importance analysis aims to assess how much each feature contributes to predicting student activity status (Active/Passive). Based on the dataset used, the features analyzed include:

a. Semester : shows the level of student progress in the study program.
b. Credits Taken : total credits taken by the student to date.
c. Attendance : percentage of student attendance in lectures.
d. GPA : Cumulative Performance Index which reflects academic performance.
e. Number of Courses : the number of courses taken by students in one semester (semester 8 in the 2025 academic year)

1. Analysis Results

Random Forest model (which specifically provides feature importance directly), the order of feature importance in predicting student activity status is obtained:

Table 2: Order of Feature Importance in the *Random Forest Model*

Feature	Importance
GPA	0.35
Presence	0.30
Number of Courses	0.20
Credits Taken	0.10
Semester	0.05

ISSN 2502-3357 (Online)

2. Interpretation
    a. GPA (0.35) : The most important feature. Students with a high GPA tend to remain active, while a low GPA can be an indicator of the risk of being passive. This is consistent with previous research showing that academic performance is a major predictor of active status [11], [13].
    b. Attendance (0.30) : Attendance percentage also greatly influences prediction. Students with high attendance are usually more academically active, so recall and F1-score increase if this feature is included [7], [9].
    c. Number of Courses (0.20) : Indicates the study load. Students with too many or too few courses may face the risk of inactivity, so this feature is important for detecting non-linear patterns [12], [16].
    d. Credits Taken (0.10) : Although less dominant than GPA and attendance, the total credits taken provide context for a student's academic progress .
    e. Semester (0.05) : The feature with the least influence. This shows that the semester level alone is not strong enough to predict active status without being supported by academic performance and attendance indicators [14], [17].
3. Implications
    a. Focusing on GPA and Attendance for academic interventions can increase the effectiveness of predicting student active status.
    b. The model can be enriched with additional contextual features such as extracurricular activity participation, study load, or other non-academic data to make predictions more accurate.
    c. This feature interest information is useful for universities to determine intervention priority : students with low GPAs and low attendance can be targeted for more intensive guidance and monitoring programs.

4.3 Implications and Applications

1. Academic Implications
    a. Monitoring Student Active Status
        1) Machine learning models (Decision Tree, SVM, and Random Forest) enable universities to monitor student activity status automatically and in real-time .
        2) With high accuracy (>97%), academics can trust the model's predictions to identify active and passive at-risk students without having to manually check the data.
    b. Early Intervention
        1) The feature importance results show that GPA and Attendance are the most critical features.
        2) Students with low GPAs and low attendance can be identified promptly for academic guidance, counseling, or remedial programs before they cause serious problems such as inactivity or drop-out [11], [13].
    c. Resource Priority
        1) Universities can allocate teaching and mentoring resources more efficiently.
        2) Students who are potentially passive can be prioritized for mentoring programs, while active students are still monitored regularly.
2. Administrative Implications
    a. Efficiency of Academic Information System
        1) With the integration of predictive models, the academic system can automatically provide warnings or notifications regarding student activity status.
        2) Reduce the administrative burden on academic staff and speed up decision-making regarding student guidance or performance evaluation.
    b. Data-Driven Policy
        1) Machine learning analysis enables universities to create data-driven policies related to academic programs, course quotas, or mentoring interventions.
        2) For example, remedial programs can be provided in a more targeted manner to students at risk of low engagement.
3. Application of Technology
    a. Integration into Academic Systems
    The model can be applied directly to the Academic Information System. Royal University , so that student activity status can be predicted automatically every semester or every lecture period.
    b. Monitoring Dashboard
        1) Prediction results can be visualized in a monitoring dashboard for lecturers and academics, displaying:
            - Active/passive students
            - Students at high risk of inactivity
            - Feature importance such as GPA, Attendance, Number of Courses
        2) This makes it easier for academics to carry out priority-based interventions.
    c. Remedial Program Plan and Academic Development
        1) Students with passive predictions can be automatically given remedial advice , motivational workshops, or additional guidance.

DOI: https://doi.org/10.65126/jocosir.v3i1
Journal of Computer Science Research (JoCoSiR) with CC BY NC SA license

ISSN 2502-3357 (Online)

2) Active students are continuously monitored to improve academic quality and maintain high graduation rates.

4. Strategic Implications
   a. Data-Driven Education Policy
      Predicting student activity status allows universities to implement more proactive and preventative academic policies , rather than just being reactive to problematic students.
   b. Improving Academic Quality and Student Retention
      By properly identifying students at risk of being passive, universities can reduce inactivity rates and increase student retention rates , thereby improving the quality of education.
   c. Development of Adaptive Learning Systems
      Feature importance data can be used to develop adaptive learning strategies , for example adjusting the study load or intensity of tutoring based on GPA and attendance.
5. Conclusion Implications and Applications
   a. Machine learning models not only improve the accuracy of predicting student activity status, but also provide important insights into the features that influential , especially GPA and attendance.
   b. The research results can be applied in academic information systems, data-based academic policies, remedial programs, and student monitoring strategies.
   c. Overall, the implementation of this model supports a proactive and preventative approach to student management, improving academic quality, and maximizing student success at Royal University.

5. Conclusions

This study successfully developed and evaluated a classification model of student activity status (Active and Inactive) using *Machine Learning algorithms* on 642 Royal University data. By considering academic (GPA, Credits Taken), behavioral (Attendance), and situational (Semester, Number of Courses) features, we tested the performance of *Decision Tree* , *Support Vector Machine* , and *Random Forest* .

Based on the research of student active status classification at Royal University using Decision Tree, SVM, and Random Forest algorithms, it was found that all models showed high performance with accuracy above 97%, where DT and SVM had the highest accuracy (98.44%), while RF excelled in recall (0.96) making it more effective in detecting active students. Feature importance analysis showed that GPA and attendance were the most influential factors on student active status, while the number of courses, credits taken, and semester provided additional contributions. These results have important implications for universities, namely enabling automated student monitoring, early intervention for students at risk of being passive, and data-driven academic decision-making. By implementing this model in academic information systems and monitoring dashboards, universities can improve student retention, educational quality, and the effectiveness of mentoring programs, so this research provides a significant contribution in the form of an accurate classification model and insight into crucial features that influence student activeness .

However, this study has limitations, namely not considering psychological and social factors (e.g., family support or *peer-to-peer interactions*). Future research directions are suggested to: Explore the use of *Deep Learning* (e.g., *Recurrent Neural Networks*) to capture time dependencies in *time-series data* ; Integrate textual data from student *feedback* or counseling journals using *Natural Language Processing* (NLP); and Conduct a *pilot project implementation* of an early warning system in the real environment of the Royal University to directly validate the impact of the interventions suggested by this model.

6. References

[1] M. Vaarma, “Predicting student dropouts with machine learning,” *Computers in Human Behavior*, vol. 132, p. 107246, 2024.

[2] D. Kayande, “Design of an integrated multi-modal machine learning system for student engagement prediction,” *Procedia Computer Science*, vol. 185, pp. 123–130, 2025.

[3] J.G.C. Krüger, “An explainable machine learning approach for student dropout prediction,” *Computers in Human Behavior*, vol. 135, p. 107369, 2023.

[4] G. Al-Tameemi, “A Hybrid Machine Learning Approach for Predicting Student Academic Performance,” *Procedia Computer Science*, vol. 187, pp. 344–351, 2024.

[5] S. Mustofa, “A novel AI-driven model for student dropout risk analysis,” *Procedia Computer Science*, vol. 266, pp. 123–130, 2025.

[6] G.Y. Wang, “Predicting learners' engagement and help-seeking behavior,” *Procedia Computer Science*, vol. 266, pp. 231–238, 2025.

[7] T. Althaqafi, “Enhancing student performance prediction: the role of learning management systems,” *Education and Information Technologies*, vol. 30, pp. 345–359, 2025.

[8] J. Martínez, “Prediction of student dropout using Artificial Intelligence techniques,” *Procedia Computer Science*, vol. 187, pp. 416–423, 2024.

[9] A. Villar, “Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study,” *AI Open*, vol. 4, pp. 1–10, 2024.

ISSN 2502-3357 (Online)
[10] A.M. Rabelo, “A model for predicting dropout of higher education students,” *Procedia Computer Science*, vol. 266, pp. 34–41, 2025.
[11] V. Zhang, “A Machine Learning Approach for Predicting Student Progress in Online Programming Education,” *International Journal of Artificial Intelligence in Education*, vol. 35, pp. 123–135, 2025.
[12] G. Öztüre Yavuz, “Investigating features that play a role in predicting gifted student engagement using machine learning,” *Education and Information Technologies*, vol. 29, pp. 16317–16343, 2024.
[13] B. Alnasyan, “The power of Deep Learning techniques for predicting student dropout,” *Procedia Computer Science*, vol. 266, pp. 112–119, 2024.
[14] J. Niyogisubizo, “Predicting student's dropout in university classes using two machine learning algorithms,” *Procedia Computer Science*, vol. 266, pp. 201–208, 2022.
[15] C. Thomas, “Automatic prediction of presentation style and student engagement in online learning,” *Procedia Computer Science*, vol. 266, pp. 150–157, 2022.
[16] A. Villar, “Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study,” *AI Open*, vol. 4, pp. 1–10, 2024.
[17] T. Althaqafi, “Enhancing student performance prediction: the role of learning management systems,” *Education and Information Technologies*, vol. 30, pp. 345–359, 2025.
[18] J. Martínez, “Prediction of student dropout using Artificial Intelligence techniques,” *Procedia Computer Science*, vol. 187, pp. 416–423, 2024.
[19] A.M. Rabelo, “A model for predicting dropout of higher education students,” *Procedia Computer Science*, vol. 266, pp. 34–41, 2025.
[20] V. Zhang, “A Machine Learning Approach for Predicting Student Progress in Online Programming Education,” *International Journal of Artificial Intelligence in Education*, vol. 35, pp. 123–135, 2025

DOI: https://doi.org/10.65126/jocosir.v3i1
Journal of Computer Science Research (JoCoSiR) with CC BY NC SA license