Classification of Student Learning Styles Using Artificial Neural Networks on Imbalanced Data

Fikri Baharuddin*[1], Ahmad Miftah Fajrin[2], Felix Handani[3]

[1,2,3]Department of Informatics, Faculty of Engineering, University of Surabaya, Surabaya, East Java, Indonesia
Email: [1*]fikribaharuddin@staff.ubaya.ac.id, [2]ahmadmiftah@staff.ubaya.ac.id, [3]felix.handani@staff.ubaya.ac.id

(Received: 20 Sep 2024, revised: 11 Oct 2024, accepted: 14 Oct 2024)

Abstract

The transformation of learning activities towards digital form since the COVID-19 pandemic can affect students' learning process. One of the factors that can affect this learning process is the learning style owned by each student. Learning patterns that are not in line with students' learning styles can influence their learning process. This study aims to identify students' learning styles based on data extracted from the Moodle Learning Management System (LMS). The research methods applied in this study include data collection by extracting data from Moodle LMS logs and classifying student learning styles using the Artificial Neural Network (ANN) algorithm. This study uses 310 log extraction data on the Moodle platform. The Isolation Forest algorithm was applied to this study to detect anomalies or outliers in the dataset. The data used in this study also has an unbalanced distribution of data per class. To prevent the performance degradation of the classifier model caused by the imbalance of data distribution, this study uses the SMOTE algorithm which can generate new synthetic data on minority class. This study combines three algorithms consisting of the Isolation Forest Algorithm for dataset management, the SMOTE Algorithm to solve the problem of data imbalance, and the ANN Algorithm to build a classification model. The model evaluation is carried out by considering the values of accuracy, precision, recall, and F1-Score to identify the reliability level of the produced model. Based on the research, this study produced a classifying model with an accuracy of 96%. The model produced in this study can be used to identify students' learning styles and as a reference for improving the quality of the teaching and learning process.

Keywords: Machine Learning, Imbalanced Data, Learning Style, Artificial Neural Network, SMOTE.

I. INTRODUCTION

Education has undergone significant changes in recent years, along with the development of digital technology. The COVID-19 pandemic is one factor that has accelerated the transition to online distance learning, so more and more educational institutions are using the Learning Management System (LMS) platform. The transformation of learning to digital form has caused changes in learning styles that need to be considered so that the implementation of learning can continue to run optimally. Identifying a student's learning style is important to maximize the learning process. Learning patterns contrary to a person's learning style can increase the risk of learning failure. This learning failure can have several negative impacts on a person, such as decreased interest in learning, decreased achievement index, and the risk of stopping the learning process [1].

Moodle is one of the Learning Management System (LMS) platforms that educational institutions worldwide widely use. Moodle provides a flexible platform and can be designed according to the academic institution's needs. Moodle can collect student behavior and interaction data when using the LMS platform. The data stored includes information on the time of student access to the system, activities carried out, and the results of the learning evaluation achieved. This data can be extracted for later analysis to identify students' preferences and learning styles. To identify student learning styles, in this study, the Felder-Silverman Learning Style Model (FSLSM) was used with four main learning style domains consisting of input, understanding, preprocessing, and perception. As in research, FSLSM has been used to detect a person's learning style. The domain in FSLSM can be used to identify student learning styles through the data engineering process on student learning data stored in Moodle LMS logs [1].

The Artificial Neural Networks machine learning algorithm is used to classify student learning styles. The use of this machine learning algorithm aims to automate the classification process. By developing an ANN classification model with good performance, the model can automatically predict students' learning styles with good accuracy. The ANN algorithm selection is based on ANN's ability to understand complex data patterns [2], [3] which is better than similar algorithms such as SVM, k-NN, and Logistic Regression. In addition, ANN has a flexible architecture, making it easy to adjust to the case study you want to complete.

Several previous studies have attempted to understand how students interact with Moodle LMS and study the factors that can affect learning effectiveness through this platform. The research [4] analyzed students' study habits using several classification algorithms such as Decision Tree, Naïve Bayes, and K-Nearest Neighbors. The use of Moodle LMS data logs has also been used to analyze learning patterns, as discussed in the study [5]. Several studies [6], [7], [8], [9] FSLSM is an essential reference for mapping student learning styles. Some studies propose improving the methods used to analyze student learning patterns.

This research can help adjust learning more effectively in educational institutions that utilize the Moodle platform. It is an innovative product for education and information technology. In addition, the output of this research can be used as a reference for similar studies in the future to develop technical solutions to improve the quality of learning implementation through the learning management system.

II. RESEARCH METHOD

The method used in this study is a machine learning model development method, as can be observed in Figure 1. Referring to the flow diagram, four main stages dataset load, data pre-processing, model training, and model evaluation need to be passed in this study.

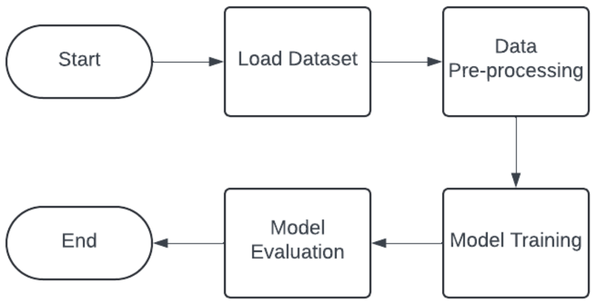

Figure 1. Machine Learning Model Development Workflow

A. Moodle E-Learning

Moodle is one of the most widely used Learning Management Systems today and has been studied and accepted by all parties through the Acceptance Model [10]. Moodle provides a comprehensive educational process through complete learning management, such as material creation, user monitoring, and the assessment process of material [11]. However, learning materials created on Moodle must be well-designed so that users are interested and continue to use the system [12].

The use of Moodle in an institution is essential with an analysis related to technological developments and system usage behavior by users [13]. User behavior on Moodle can be seen in a log. The log can track the behavior patterns of users who have certain tendencies such as discipline in opening material in the form of text, video to how to communicate on forums [14]. Teachers such as lecturers who use Moodle need to pay attention to student interaction when using Moodle because students are valuable assets [15].

B. Felder and Silverman Learning Styles Model (FSLSM)

The FSLM model is a modeling framework initiated by [16]. The FSLMS model was developed by considering factors that affect the suitability of the delivery of learning materials and the level of student understanding. Learning patterns are measured by evaluating the effectiveness of students in understanding the learning material. Meanwhile, the teaching pattern is developed by considering factors affecting the success rate in conveying the core of learning to students. FSLSM can describe in detail its characteristics, can be easily translated into a framework, and can be used as a framework for research [17].

Improving the success of the student learning process can be done by identifying their learning style. One way to identify student learning styles is to use Felder and Silverman Learning Styles (FSLSM) [6], [7], [8]. FSLSM can be analyzed using logs obtained on Moodle. The learning style domains contained in the study [16] can be observed in Table **1**.

Table 1. FSLSM Learning Style Domain

Preferred Learning Style	
sensory intuitive	perception
visual auditory	input
inductive deductive	organization
active reflective	processing
sequential global	understanding

C. Dataset

The dataset used in this study is obtained from the data extraction results on the Moodle E-Learning platform used at "X" University. The data used is students' learning log data from several courses under the informatics department. The data used in this study amounted to 327 data. The data used has attributes, as can be observed in Table 2. The dataset labeling was based on research [18] using the FSLSM learning style domain, where data labeling based on FSLSM domains is carried out by considering student statistical data in accessing learning materials, quizzes, and other assessments on the Moodle LMS platform. This research dataset is public and can be accessed through [19]

Table 2. Features of the Dataset

Feature	Description
RESULT ASSIGNMENT	An attribute to show the total assignment that a student has done.
VALUES ASSIGNMENT	An attribute to show the percentage of a student's completion of a student's assignment to the total assignment given by the teacher.
RESULT QUIZ	An attribute to show the total *quiz* that a student has done.
VALUES QUIZ	An attribute to show the percentage of a student's completion of a student's assignment against the total *quiz* given by the teacher.
RESULT PDF	An attribute to show the total PDF material that a student has accessed.
VALUES PDF	An attribute that shows the percentage of a student's activity toward the total PDF material provided by the teacher.
RESULT PPT	An attribute to show the total PPT material that a student has accessed.
VALUES PPT	An attribute to show the percentage of activity of a student towards the total of the PPT material given by the teacher.
CLASS	Labels to categorize students' learning styles based on aspects of FSLSM

D. Data Pre-processing

The pre-processing methods used in this study include handling missing values, handling outliers, class balancing, standard scaling, and split data. Handling missing values eliminates data that contains "null" or empty data from the dataset. Data containing this "null" element can degrade the model's performance, so it needs to be cleaned from the dataset to be used. The handling outlier aims to check for anomalies in the dataset to be used. Anomalies in datasets, also called outliers, can affect the performance of a classification model. Therefore, it is crucial to manage outliers properly [20]. This study uses the Isolation Forest algorithm to recognize anomalies in the dataset. Referring to research [21] (Liu et al., 2, the Isolation Forest (iForest) algorithm works by performing three processes. The process begins with a random selection of data features to identify whether there are significant character differences. Identification of data that has anomalies can be made based on the results of calculating the anomalous score from the data on other data. The iForest algorithm can be used to detect outlier and noise data [22], [23], [24].

E. Synthetic Minority Over-sampling Technique (SMOTE)

The SMOTE method is applied to overcome the problem of data imbalance between classes, which can cause a decrease in the model's performance. SMOTE is a pre-processing method that is widely used to balance datasets with an unbalanced distribution of data per class, as in research [25], [26], [27], [28], [29], [30].

Referring to the SMOTE algorithm that can be seen in Table **3**, the SMOTE method works by creating synthesis data obtained from the results of the feature distance calculation using the formula (1). The result of the distance calculation is then added with bias(λ) to produce new data that is not the same as observable in equation (2). The newly generated synthetic data is combined with the initial dataset to create a balanced dataset. Thus, the total data in the minority class will be equalized with the total data in the minority class. This balanced data will then be used in the model training stage.

Table 3. SMOTE Algorithms

SMOTE Algorithm		
Input	-	Dataset X (Original unbalanced dataset)
	-	N (generated synthesis data)
	-	K (number of nearest neighbors)
Output	-	Dataset D' (Balanced Dataset)
Algorithm		

1. Separating datasets into minority and majority-class data
2. **For x in minority ():**
3. Looking for neighbors as many as K;
4. selectedNeighbors = K nearest neighbors from randomly selected x data;
5. **End For**
6. **For neighbor in selectedNeighbors:**
7. $$distance = \sqrt{\sum_{d=1}^{n}\left(x_i^d - x_{ij}^d\right)^2} \quad (1)$$
8. $$x_{new} = x_i + \lambda\left(x_{ij} - x_i\right) \quad (2)$$
9. $|D'| = |D| + x_{new}$
10. **End For**
11. return $|D'|$

F. Artificial Neural Networks

Artificial Neural Network (ANN) is one of the machine learning methods. ANNs work using interconnected neurons and send signals to simulate the human neural learning process using complex algorithms [2]. The ANN algorithm is used to obtain learning prediction results based on the input features that have been described previously. The ANN algorithm can be observed in Table **4**.

Table 4. ANN Algorithm

Artificial Neural Networks Algorithm

1. Determination of learning rate (α) and number of iterations (epoch)
2. Initialization of the weight value (W) and the refractive value (b)
3. **For each iteration(epoch):**
4. **For each data training(X,y)**
5. Linear output calculation
6. Calculation of activation functions using weights and bias

```
Artificial Neural Networks Algorithm
7.          Prediction output calculation
8.          Loss/error rate calculation
9.          Calculation of error value changes to weights
        and bias
10.         Update weight values and bias
11.     End For
12.     If(stop condition == true)
13.         break
14.         Save the best model
15.     End If
16.   End For
17.   Return best model
```

One of the uses of ANN is to predict learning styles for eLearning applications [3]. The use of ANN for learning style prediction is carried out by using interactive videos and can recognize learning types from teachers and students [31]. Predicting learning styles and providing feedback that can be processed are essential parameters in the eLearning domain.

A survey conducted on the use of ANN for learning styles resulted in the conclusion that student behavior analysis is very important so that it can be prevented, such as dropout problems in students [32]. ANN has a high prediction to assess student graduation based on the correlation of viewing learning videos in online courses [33]. ANN is also used to predict student categories, namely bad, medium, and good for their academic achievement [34].

G. Model Training

The Artificial Neural Network (ANN) algorithm creates the classifier model. The model was trained using the hyperparameter configuration, as can be observed in Table 5.

Table 5. Hyperparameter Configuration in the ANN Model

Hyperparameter	*Value*
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Loss Metric	Sparse Categorical Cross Entropy
Epoch	30
Hidden layer	2 layers
1st Hidden Layer	128 neurons
2nd Hidden Layer	64 neurons
Validation split	0.4

H. Model Evaluation

The model is evaluated using evaluation metrics consisting of accuracy, precision, recall, and F-Score metrics. Metric evaluation involves several attributes, namely True Positive(TP), True Negative(TN), False Positive(FP), and False Negative(FN). The calculation formula for each metric can be observed in Table **6**.

Table 6. Model Evaluation Metric

Metrics	*Formula*
Accuracy	$Accuracy = \frac{TP + \boldsymbol{TN}}{TP + TN + FP + FN}$
Precision	$Precision = \frac{TP}{TP + FN}$
Recall	$Recall = \frac{TP}{TP + FN}$
F-Score	$F - Score = 2 \times \frac{Precision}{Recall}$

III. RESULTS AND DISCUSSION

The dataset used in this study was first filtered using the Isolation Forest algorithm, which aims to separate the noise or outlier from the data used in the modeling process. Implementing the Isolation Forest algorithm successfully detected as many as 17 data classified as outliers. The data classified as outliers is omitted from the dataset. The dataset used in this study also has an unbalanced class composition, as can be observed in Figure 2.

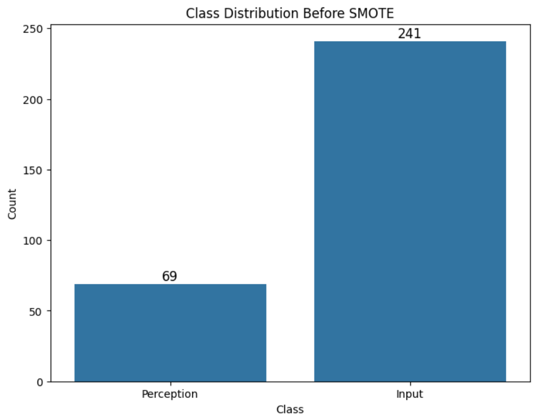

Figure 2. Amount of Data per Class

After applying the SMOTE method, the dataset was successfully balanced with an output of 241 data for each class, as shown in Figure 3.

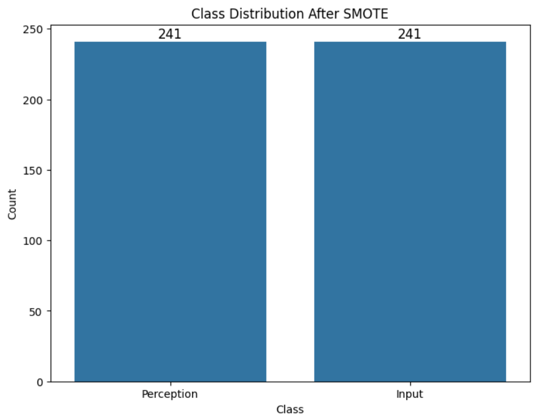

Figure 3. Amount of Data per Class after SMOTE

The balanced dataset is then transformed into a more standard form using the standard scaling method. The data processed through several methods is then separated into two partitions: the training data partition and the test data partition. The data division was done by applying a 60:40 composition, with 60% used for training data. Of the total 482 data in the dataset, 289 were used as training data, and 193 were used as test data.

The model is created according to the hyperparameter configuration described earlier. The model was trained in 30 epochs as described in the hyperparameter configuration and produced the best model results, as can be observed in Table 7. The callback implementation stores the best model state during training. In addition, early stopping is implemented to stop the training automatically if accuracy increases or loss levels do not decrease three times.

Table 7. Details of the Best Models Produced

Aspect	Value
Epoch	29
Accuracy	0.988439
Loss	0.031415
Validation Accuracy	0.974138
Validation Loss	0.056495
Training Time	45s

Based on Table 4, the best model produced in this study was obtained in the 29th epoch training. The metrics used to assess the model's performance are validation accuracy and loss values. In the 29th epoch, the model showed a validation accuracy value of 0.97 (97%) with a validation loss value of 0.06. By paying attention to the visualization of the comparison between the training loss and validation level, there is no significant indication that leads to overfitting. Comparative data between training accuracy and validation also did not indicate the occurrence of overfitting. The visualization of the comparison of accuracy and loss during the training process can be observed in Figure 4 and Figure 5.

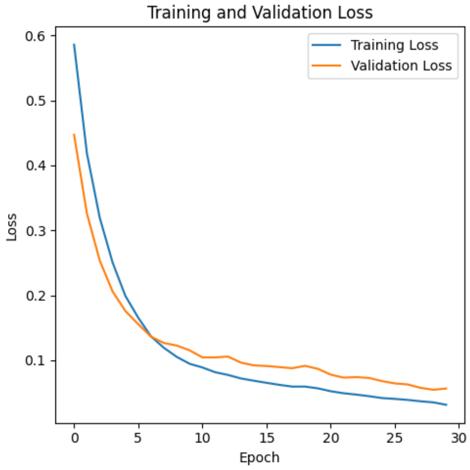

Figure 4. Loss Comparison Graph

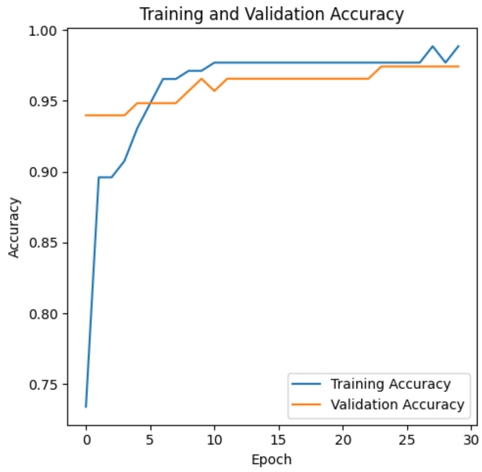

Figure 5. Graph Comparison Accuracy

The model generated at the training model stage is then evaluated with several evaluation metrics to identify the model's performance by testing the model to conduct classification tests on the test data. The test data used came from data partitions of 40% of the total dataset. The test was carried out by paying attention to accuracy, precision, recall, and F-Score metrics. Based on the evaluation carried out. The model can classify test data with excellent performance with an accuracy level of 97%, accuracy of 98%, recall of 97%, and F-Score of 97%, as observed in Table 8.

The classification test conducted on the test data shows that 84 data have been properly classified as Input classes. Five data should have been included in the Input class but were incorrectly classified as class perception, 104 data were successfully classified as class perception, and no data class perception was incorrectly classified as an input class. The confusion matrix of the evaluation results can be observed in Figure 6.

Table 8. Classification Test Results Report

Aspect	Precision	Recall	F1-Score
Input Class	1.00	0.94	0.97
Perception Class	0.95	1.00	0.98
Accuracy	**0.97**		
Macro Average	**0.98**	**0.97**	**0.97**
Weighted Average	0.98	0.97	0.97

 DOI: 10.34148/teknika.v13i3.1075

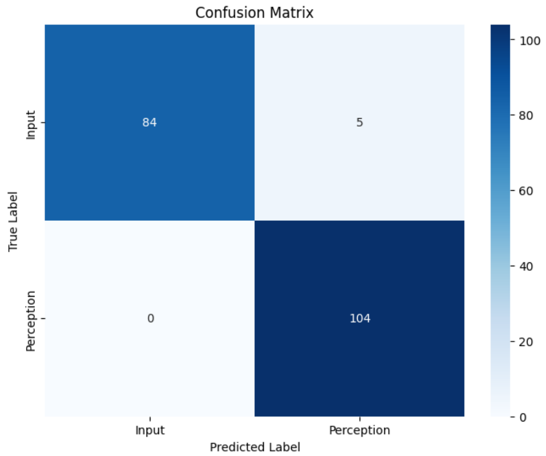

Figure 6. Confusion Matrix Model Evaluation Results

IV. CONCLUSION

Based on the research results, the Artificial Neural Network Algorithm can be used to classify a person's learning style based on the Felder and Silverman Learning Styles Model (FSLSM) domain. This is evidenced by the model's evaluation results, which showed an accuracy of 98%, *a recall* of 97%, and an F-Score of 97%. The resulting model shows good performance by paying attention to the balance of accuracy, recall, precision, and F-score metrics. High recall and precision values indicate the consistency of the model in making accurate predictions. The low number of classification errors also shows that the resulting model is reliable.

The application of the Isolation Forest method in this study succeeded in eliminating some outlier data that could affect the performance of the classifier model. The SMOTE method can solve the data imbalance problem for each class, which can influence the model training process. The dataset used in the study only consists of two labels, namely input and perception. This is due to the lack of quantity of datasets that can undoubtedly be added to future studies to cover other FSLSM domains. The dataset used to build the model also only consists of one study program and does not cover all courses. This has the potential to generate bias when the model is implemented in real-time, so in the future, it needs to be developed with a dataset that includes more data variants to get more accurate and reliable results. In addition, optimization can be performed on the model's hyperparameters to obtain the best performance.

ACKNOWLEDGMENT

The author expresses deepest gratitude to the Research and Community Service Institution of the University of Surabaya for funding this research through internal funding for a quality research publication scheme.

REFERENCES

[1] T. Sheeba and R. Krishnan, “Prediction of student learning style using modified decision tree algorithm in e-learning system,” in *Proceedings of the 2018 International Conference on Data Science and Information Technology*, 2018, pp. 85–90.

[2] J. Cuauhtemoc *et al.*, “AI-Based Prediction of Capital Structure: Performance Comparison of ANN SVM and LR Models,” *Wiley Online LibraryJC Tellez Gaytan, K Ateeq, A Rafiuddin, HM Alzoubi, TM Ghazal, TA Ahanger, S ChaudharyComputational intelligence and neuroscience, 2022•Wiley Online Library*, vol. 2022, 2022, doi: 10.1155/2022/8334927.

[3] K. Alhumaid, M. Habes, and S. A. Salloum, “Examining the Factors Influencing the Mobile Learning Usage during COVID-19 Pandemic: An Integrated SEM-ANN Method,” *IEEE Access*, vol. 9, pp. 102567–102578, 2021, doi: 10.1109/ACCESS.2021.3097753.

[4] S. Tiodora, S. | Umi, L. Yuhana, * Umi, and P. Address, “Student behaviour analysis to detect learning styles using decision tree, Naïve Bayes, And K-Nearest Neighbor method in moodle learning management system,” *iptek.its.ac.idST Sianturi, UL YuhanaIPTEK The Journal for Technology and Science, 2022•iptek.its.ac.id*, vol. 33, no. 2, pp. 853–4098, doi: 10.12962/j20882033.v33i2.13665.

[5] K. Dobashi, C. P. Ho, C. P. Fulford, M. F. Grace Lin, and C. Higa, “Learning pattern classification using moodle logs and the visualization of browsing processes by time-series cross-section,” *Computers and Education: Artificial Intelligence*, vol. 3, Jan. 2022, doi: 10.1016/J.CAEAI.2022.100105.

[6] Y. Ikawati, M. U. H. Al Rasyid, and I. Winarno, “Student Behavior Analysis to Detect Learning Styles in Moodle Learning Management System,” *IES 2020 - International Electronics Symposium: The Role of Autonomous and Intelligent Systems for Human Life and Comfort*, pp. 501–506, Sep. 2020, doi: 10.1109/IES50839.2020.9231567.

[7] M. Liyanage, … K. G.-… I. J. on, and undefined 2014, “Using learning styles to enhance learning management systems,” *journal.icter.orgMPP Liyanage, KSL Gunawardena, M HirakawaThe International Journal on Advances in ICT for Emerging Regions, 2014•journal.icter.org*, Accessed: Sep. 18, 2024. [Online]. Available: https://journal.icter.org/index.php/ICTer/article/view/179

[8] A. S. Sprock, “Inclusion of the FuzzyILS method in MOODLE for creating effective courses,” *International Journal of Learning, Teaching and Educational Research*, vol. 19, no. 10, pp. 32–59, Oct. 2020, doi: 10.26803/IJLTER.19.10.3.

[9] I. Azzi, A. Jeghal, A. Radouane, A. Yahyaouy, and H. Tairi, “A robust classification to predict learning styles in adaptive E-learning systems,” *Educ Inf Technol*

(Dordr), vol. 25, no. 1, pp. 437–448, Jan. 2020, doi: 10.1007/S10639-019-09956-6.

[10] G. G. Murillo, P. Novoa-Hernández, H. Hernández, and R. Serrano Rodríguez, "Technology Acceptance Model and Moodle: A systematic mapping study," *journals.sagepub.comGG Murillo, P Novoa-Hernández, RS RodriguezInformation Development, 2021•journals.sagepub.com*, vol. 37, no. 4, pp. 617–632, Nov. 2021, doi: 10.1177/0266666920959367.

[11] M. Zabolotniaia, Z. Cheng, E. Dorozhkin, and A. Lyzhin, "Use of the LMS Moodle for an effective implementation of an innovative policy in higher educational institutions.," *pdfs.semanticscholar.orgM Zabolotniaia, Z Cheng, E Dorozhkin, A LyzhinInternational Journal of Emerging Technologies in Learning, 2020•pdfs.semanticscholar.org*, vol. 15, no. 13, pp. 172–189, Jan. 2020, doi: 10.3991/ijet.v15i13.14945.

[12] D. Keržič, A. Aristovnik, N. Tomaževič, and L. Umek, "An assessment of the effectiveness of Moodle e-learning system for undergraduate public administration education," *International Journal of Innovation and Learning*, vol. 21, no. 2, p. 165, 2017, doi: 10.1504/IJIL.2017.10002132.

[13] B. Bervell, I. U.-E. J. of Mathematics, S. and, and undefined 2017, "A decade of LMS acceptance and adoption research in Sub-Sahara African higher education: A systematic review of models, methodologies, milestones and main," *ejmste.comB Bervell, IN UmarEurasia Journal of Mathematics, Science and Technology Education, 2017•ejmste.com*, vol. 13, no. 11, pp. 7269–7286, 2017, doi: 10.12973/ejmste/79444.

[14] N. Kadoic and D. Oreski, "Analysis of student behavior and success based on logs in Moodle," *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, pp. 654–659, Jun. 2018, doi: 10.23919/MIPRO.2018.8400123.

[15] F. Amankwah, F. K. Sarfo, M. O. Aboagye, D. Konin, and R. K. Dzakpasu, "Concerns of university teachers about the adoption of the Moodle learning management system in a Ghanaian University campus," *Taylor & FrancisF Amankwah, FK Sarfo, MO Aboagye, D Konin, RK DzakpasuEducation inquiry, 2024•Taylor & Francis*, vol. 15, no. 3, pp. 312–332, 2022, doi: 10.1080/20004508.2022.2109849.

[16] R. Felder, L. S.-E. education, and undefined 1988, "Learning and teaching styles in engineering education," *academia.eduRM Felder, LK SilvermanEngineering education, 1988•academia.edu*, Accessed: Sep. 18, 2024. [Online]. Available: https://www.academia.edu/download/31039406/LS-1988.pdf

[17] M. Dominic, B. Anthony Xavier, and S. Francis, "A Framework to Formulate Adaptivity for Adaptive e-Learning System Using User Response Theory," *International Journal of Modern Education and Computer Science*, vol. 7, no. 1, pp. 23–30, Jan. 2015, doi: 10.5815/IJMECS.2015.01.04.

[18] T. Sheeba and R. Krishnan, "Prediction of student learning style using modified decision tree algorithm in e-learning system," in *Proceedings of the 2018 International Conference on Data Science and Information Technology*, 2018, pp. 85–90.

[19] F. Baharuddin, "Dataset Log Data Moodle for SLSA," Sep. 2024, *Zenodo*. doi: 10.5281/zenodo.13819870.

[20] T. Nyitrai and M. Virág, "The effects of handling outliers on the performance of bankruptcy prediction models," *Socioecon Plann Sci*, vol. 67, pp. 34–42, Sep. 2019, doi: 10.1016/J.SEPS.2018.08.004.

[21] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[22] W. R. Chen, Y. H. Yun, M. Wen, H. M. Lu, Z. M. Zhang, and Y. Z. Liang, "Representative subset selection and outlier detection: Via isolation forest," *Analytical Methods*, vol. 8, no. 39, pp. 7225–7231, Oct. 2016, doi: 10.1039/C6AY01574C.

[23] Z. Cheng, C. Zou, and J. Dong, "Outlier detection using isolation forest and local outlier factor," in *Proceedings of the Conference on Research in Adaptive and Convergent Systems*, in RACS '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 161–168. doi: 10.1145/3338840.3355641.

[24] R. C. Ripan *et al.*, "An Isolation Forest Learning Based Outlier Detection Approach for Effectively Classifying Cyber Anomalies," in *Hybrid Intelligent Systems*, A. Abraham, T. Hanne, O. Castillo, N. Gandhi, T. Nogueira Rios, and T.-P. Hong, Eds., Cham: Springer International Publishing, 2021, pp. 270–279.

[25] S. Bujang, A. Selamat, O. Krejcar, … F. M.-I., and undefined 2022, "Imbalanced classification methods for student grade prediction: a systematic literature review," *ieeexplore.ieee.org*, Accessed: Sep. 18, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9965398/

[26] A. Yaqin, M. Rahardi, F. A.-I. J. of, and undefined 2022, "Accuracy Enhancement of Prediction Method using SMOTE for Early Prediction Student's Graduation in XYZ University," *researchgate.net*, vol. 13, no. 6, p. 2022, 2022, doi: 10.14569/IJACSA.2022.0130652.

[27] M. Shrinidhi, … T. K. J.-A. in S., and undefined 2023, "Classification of Imbalanced Datasets Using Various Techniques along with Variants of SMOTE Oversampling and ANN," *Trans Tech Publ*, Accessed: Sep. 18, 2024. [Online]. Available: https://www.scientific.net/AST.124.504

[28] T. Riston, S. Suherman, … Y. Y.-… S. and I., and undefined 2023, "Oversampling Methods for Handling Imbalance Data in Binary Classification," *Springer*, Accessed: Sep. 18, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-37108-0_1

[29] A. Bokhare, A. Bhagat, R. B.-S. C. Science, and undefined 2023, "Multi-layer perceptron for heart failure detection using SMOTE technique," *Springer*, Accessed: Sep. 18, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s42979-022-01596-x

[30] C. Kaope, Y. P.-M. Jurnal, and undefined 2023, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," *journal.universitasbumigora.ac.id*, vol. 22, no. 2, pp. 227–238, 2023, doi: 10.30812/matrik.v22i2.2515.

[31] A. Jalal and M. Mahmood, "Students' behavior mining in e-learning environment using cognitive processes with information technologies," *Educ Inf Technol (Dordr)*, vol. 24, no. 5, pp. 2797–2821, Sep. 2019, doi: 10.1007/S10639-019-09892-5.

[32] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Applied Sciences (Switzerland)*, vol. 10, no. 3, Feb. 2020, doi: 10.3390/APP10031042.

[33] C. Yu, J. Wu, A. L.-E. sciences, and undefined 2019, "Predicting learning outcomes with MOOC clickstreams," *mdpi.comCH Yu, J Wu, AC LiuEducation sciences, 2019•mdpi.com*, vol. 9, no. 2, Jun. 2019, doi: 10.3390/educsci9020104.

[34] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. M. S. Islam, "Mining educational data to predict students performance: A comparative study of data mining techniques," *Educ Inf Technol (Dordr)*, vol. 26, no. 5, pp. 6051–6067, Sep. 2021, doi: 10.1007/S10639-021-10575-3.