



Machine Learning for Early Detection of Dropout Risks and Academic Excellence: A Stacked Classifier Approach

Teuku Rizky Noviandy ¹, Zahriah Zahriah ², Erkata Yandri ³, Zulkarnain Jalil ⁴, Muhammad Yusuf ⁵, Nur Intan Saidaah Mohamed Yusof ⁶, Andi Lala ⁷ and Rinaldi Idroes ⁷

- ¹ Interdisciplinary Innovation Research Unit, Graha Primera Saintifika, Aceh Besar 23771, Indonesia; trizkynoviandy@gmail.com (T.R.N.)
- ² Department of Architecture and Urban Planning, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; zahriah@usk.ac.id (Z.Z.)
- ³ Graduate School of Renewable Energy, Darma Persada University, Jl. Radin Inten 2, Pondok Kelapa, East Jakarta 13450, Indonesia; erkata@gmail.com (E.Y.)
- ⁴ Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; zjalil@usk.ac.id (Z.J.)
- ⁵ Department of Pharmacy, STIKES Assyifa Aceh, Aceh 23242, Indonesia; iam Yusufibrahim@gmail.com (M.Y.)
- ⁶ Faculty of Pharmacy, Universiti Teknologi MARA Selangor, Bandar Puncak Alam 42300, Malaysia; intannis@yahoo.com (N.I.S.M.Y.)
- ⁷ School of Mathematics and Applied Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; andi_lala@usk.ac.id (A.L.); rinaldi.idroes@usk.ac.id (R.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 12 March 2024
Revised 9 May 2024
Accepted 15 May 2024
Available Online 24 May 2024

Keywords:

Machine learning
Academic excellence
Dropout risks
Stacked classifier
Educational data

Abstract

Education is important for societal advancement and individual empowerment, providing opportunities, developing essential skills, and breaking cycles of poverty. Nonetheless, the path to educational success is marred by challenges such as achieving academic excellence and preventing student dropouts. Early identification of students at risk of dropping out or those likely to excel academically can significantly enhance educational outcomes through tailored interventions. Traditional methods often fall short in precision and foresight for effective early detection. While previous studies have utilized machine learning to predict student performance, the potential for more sophisticated ensemble methods, such as stacked classifiers, remains largely untapped in educational contexts. This study develops a stacked classifier integrating the predictive strengths of LightGBM, Random Forest, and logistic regression. The model achieved an accuracy of 80.23%, with precision, recall, and F1-score of 79.09%, 80.23%, and 79.20%, respectively, surpassing the performance of the individual models tested. These results underscore the stacked classifier's enhanced predictive capability and transformative potential in educational settings. By accurately identifying students at risk and those likely to achieve academic excellence early, educational institutions can better allocate resources and design targeted interventions. This approach optimizes educational outcomes and supports informed policymaking, fostering environments conducive to student success.



Copyright: © 2024 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License. (<https://creativecommons.org/licenses/by-nc/4.0/>)

1. Introduction

Education is a fundamental human right and a critical driver of personal, social, and economic development [1]. It empowers individuals with the knowledge, skills, and values necessary to lead fulfilling lives, contribute to their communities, and navigate the complexities of the modern world. Education enhances cognitive abilities and fosters creativity, critical thinking, and problem-solving skills [2, 3]. It opens doors to opportunities, promotes social mobility, and helps break the cycle of poverty. Moreover, education plays a vital role in shaping responsible citizens who can actively participate in democratic processes and contribute to the betterment of society [4]. Given its transformative power, ensuring access to quality education remains a top priority for governments, organizations, and communities worldwide [5].

However, the education system faces significant challenges in ensuring all students achieve academic excellence and successfully complete their educational journey [6]. Academic excellence refers to the highest level of academic achievement, characterized by outstanding performance, a deep understanding of the subject matter, and the ability to apply knowledge effectively [7, 8]. Students who excel academically often demonstrate strong motivation, disciplinary skills, and a passion for learning [9, 10]. They will likely pursue higher education, secure prestigious scholarships, and embark on successful careers.

On the other hand, dropout risks refer to the factors that increase the likelihood of students leaving school before completing their education [11]. Students at risk of dropping out may struggle with academic difficulties, lack engagement, face personal or family challenges, or experience social and emotional issues [12]. Dropping out of school can have severe consequences, limiting future educational and employment opportunities, increasing the risk of poverty, and negatively impacting mental health and well-being [13]. Therefore, identifying and addressing academic excellence and dropout risks is crucial for ensuring all students can reach their full potential and lead successful lives.

Early detection of academic excellence and dropout risks is important for timely student support. By identifying students who are likely to excel or struggle early on, educators can take proactive measures to nurture their talents or address their challenges [14]. Early detection allows personalized attention, targeted resources, and tailored learning experiences that cater to individual needs [15]. For high-achieving students, early identification can lead to opportunities for advanced

coursework and enrichment programs. For students at risk of dropping out, early intervention can address underlying factors, such as academic struggles or personal challenges, and provide necessary support to keep them on track [16]. Educators can create a supportive learning environment that promotes student success by focusing on early identification and intervention.

In recent years, machine learning has emerged as a powerful tool for analyzing educational data and making predictions [17-19]. By training computer algorithms to learn patterns from data, machine learning can uncover insights that may not be apparent through traditional methods [20]. These algorithms can identify factors contributing to academic excellence or dropout risks, enabling educators to make data-driven decisions. Machine learning models can predict a student's likelihood of success based on performance, engagement, and other variables, helping teachers provide targeted support and personalized learning plans. Similarly, machine learning can identify students at risk of dropping out by analyzing risk factors like attendance, grades, and socio-economic indicators. This allows educators to intervene early and offer support services. Machine learning's potential in education lies in optimizing learning outcomes, reducing achievement gaps, and ensuring every student has the opportunity to succeed.

Several studies have explored the application of machine learning in education, demonstrating its potential to enhance student outcomes and support success. Researchers have employed various algorithms and models to predict student performance [21, 22], identify at-risk students [23, 24], and personalize learning experiences [25]. However, despite this progress, there remains a research gap. Existing studies have mainly focused on individual algorithms, but to address the complexities in educational data more accurately, advanced methods like stacked classifiers are needed. These classifiers can integrate multiple algorithms' strengths and hold promise for early dropout risk detection and promoting academic excellence.

The primary aim of this study is to develop a machine learning-based approach for early detection of dropout risks and academic excellence with a stacked classifier. By leveraging a stacked classifier model, we seek to accurately identify students who are at risk of dropping out and those who are likely to excel academically. This research aims to provide educators with a powerful tool to support student success and inform targeted interventions. The findings of this study can contribute to the growing body of knowledge on machine learning

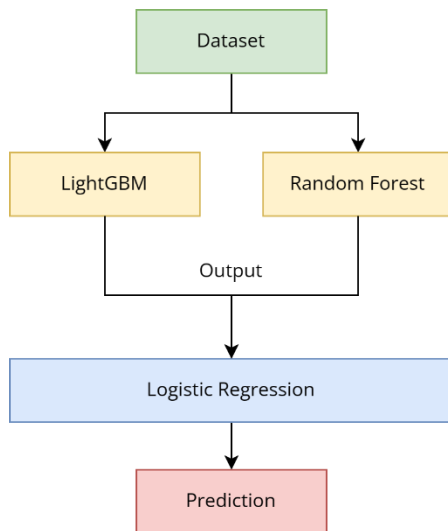


Figure 1. Stacked classifier approach used in this study.

applications in education and have practical implications for educational institutions and policymakers.

2. Materials and Methods

2.1. Dataset

The dataset used in this study was retrieved from the University of California, Irvine (UCI) machine learning repository [26, 27]. It encompasses information available during student enrollment, comprising academic path, demographics, and socio-economic factors. Before its utilization, the dataset underwent preprocessing to address anomalies, unexplainable outliers, and missing values, ensuring its quality and reliability for analysis.

The dataset consists of 4424 instances and comprises 36 features, each representing a distinct student profile. Detailed information for each feature can be accessed through the data repository [27]. These features encompass various aspects of students' academic, demographic, and socio-economic backgrounds. The dataset was employed for a multiclass classification task, aiming to categorize students into one of three classes: dropout, enrolled, or graduate. A student is classified as a graduate if they obtained their degree within the expected timeframe, as enrolled if they took up to three extra years to get the degree, and as a dropout if they took more than three extra years or failed to attain the degree. There were 1421 instances of dropouts, 794 enrolled students, and 2209 graduates in the dataset.

Next, the dataset was split into 80% training and 20% testing sets. The training set was used for model training, wherein the model learns patterns and relationships from the data. Meanwhile, the testing set was reserved for evaluating the trained model's performance on unseen data, providing an unbiased estimate of its

generalization ability. This separation ensures the model's performance can be accurately assessed and helps prevent overfitting the training data [28].

2.2. Stacked Classifier

In this study, we employed a stacked classifier approach to predict dropout risks and academic excellence. A stacked classifier is an ensemble learning method that combines the predictions of multiple base classifiers using a meta-classifier [29]. The goal is to leverage the strengths of different algorithms and create a more robust and accurate predictive model [30]. The diagram illustrating this approach is shown in Figure 1.

The stacked classifier model was implemented using Python, utilizing the versatile capabilities of the scikit-learn library [31]. This choice of programming language and library ensured flexibility and accessibility in model development and experimentation. To maintain simplicity and avoid unnecessary complexity, default hyperparameters were utilized for each constituent model within the stacked classifier [32]. This decision was made to establish a baseline performance and streamline the initial setup process. Additionally, adhering to default settings helped maintain transparency and reproducibility in the model's construction.

We first selected two base classifiers to build our stacked classifier: LightGBM and Random Forest model. The rationale behind the choice of base classifiers lies in their respective strengths and suitability for the task at hand. LightGBM is a gradient-boosting framework that uses tree-based learning algorithms. It is known for its fast training speed and high efficiency, making it suitable for large-scale datasets [20]. Random Forest is an ensemble learning method that constructs multiple decision trees and combines their predictions. It is effective in handling high-dimensional data and reducing overfitting [33].

We trained each base classifier on the training dataset, allowing them to learn patterns and relationships from the data. After training, we used these classifiers to predict the same training data. The predictions from each base classifier and the true labels formed a new dataset. These predictions were then fed into the meta-classifier, which generated the final prediction. This stacked classifier approach allowed us to leverage the diverse perspectives of multiple algorithms and create a more accurate and reliable predictive model.

We employed a logistic regression model as the meta-classifier. The logistic regression model is a simple but powerful algorithm for classification tasks [34]. The meta-classifier was trained on the new dataset, learning how to

Table 1. Performance of the machine learning models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LightGBM	79.77	78.68	79.77	78.74
Random Forest	79.44	78.26	79.44	78.11
Logistic Regression	77.85	76.27	77.85	75.99
Stacked Classifier	80.23	79.09	80.23	79.20

combine the base classifiers' predictions optimally. By using the predictions of the base classifiers as input features, the meta-classifier could capture the strengths and weaknesses of each model and determine the best way to aggregate their outputs.

The stacked classifier offers several advantages over using a single classifier. Combining the predictions of multiple models can capture complex patterns and relationships in the data that individual classifiers may miss. Additionally, the meta-classifier can learn to assign weights to the base classifiers based on their performance, effectively leveraging their strengths and mitigating their weaknesses. This approach also reduces the risk of overfitting, as the meta-classifier learns to generalize well to new data.

2.3. Performance Evaluation

To evaluate the performance of our stacked classifier, we employed several widely used metrics: accuracy, precision, recall, and F1-score. These metrics comprehensively assess the model's predictive capabilities and help us understand its strengths and limitations [35]. To compute the evaluation metrics, we applied our trained stacked classifier to the testing dataset previously unseen by the model. We compared the predicted class labels with the true labels and calculated each class's accuracy, precision, recall, and F1 scores. Finally, we aggregated these metrics using a weighted average to provide an overall assessment of the model's performance on new data and to gauge its generalization ability.

Accuracy represents the proportion of correctly classified instances, providing an overall measure of model prediction performance. Precision measures the model's ability to correctly identify instances of a specific class, which is crucial in scenarios with high costs for false positives. Recall quantifies the model's ability to identify all instances of a specific class, which is important when false negatives carry significant consequences. The F1-score, the harmonic mean of precision and recall, offers a balanced performance measure, particularly useful in uneven class distributions where both metrics are vital, ensuring effective identification of positive instances while minimizing false positives.

We generated a confusion matrix to visualize the model's performance, providing a tabular summary of its predictions, including true positives, true negatives, false positives, and false negatives. This matrix helps identify errors and provides insights into strengths and weaknesses. Additionally, we compare our stacked classifier's performance with that of three individual models: Logistic Regression, Random Forest, and LightGBM. Evaluating these models alongside our stacked classifier offers insights into the effectiveness of ensemble learning versus individual algorithms.

3. Results and Discussion

The performance metrics for different machine learning models used in this study, including Logistic Regression, Random Forest, LightGBM, and a stacked classifier, are presented in Table 1. The evaluation criteria include Accuracy, Precision, Recall, and F1-Score, providing a comprehensive view of each model's ability to predict the target classes effectively.

The results show that the stacked classifier achieved the highest performance across all metrics. With an accuracy of 80.23%, it outperformed the other models' overall prediction capability. Its high precision (79.09%) indicates its ability to correctly identify instances of the positive class while minimizing false positives. This is crucial when false positives can lead to significant negative consequences. The stacked classifier's high recall (80.23%) reflects its effectiveness in identifying all instances of the target class, minimizing false negatives. Its balanced F1-Score of 79.20% emphasizes the model's ability to perform well across precision and recall, particularly important in cases where the positive class distribution is uneven.

LightGBM also delivered strong results, with accuracy close to the stacked classifier's, indicating its robust prediction power. The Random Forest model showed marginally lower metrics but remained competitive. While exhibiting lower metrics than the other models, Logistic Regression still showed reasonable performance in terms of precision and recall, which could be beneficial in simpler classification tasks.

Figure 2 illustrates the confusion matrices for the different models evaluated in this study. The stacked classifier emerges as the best performer among the

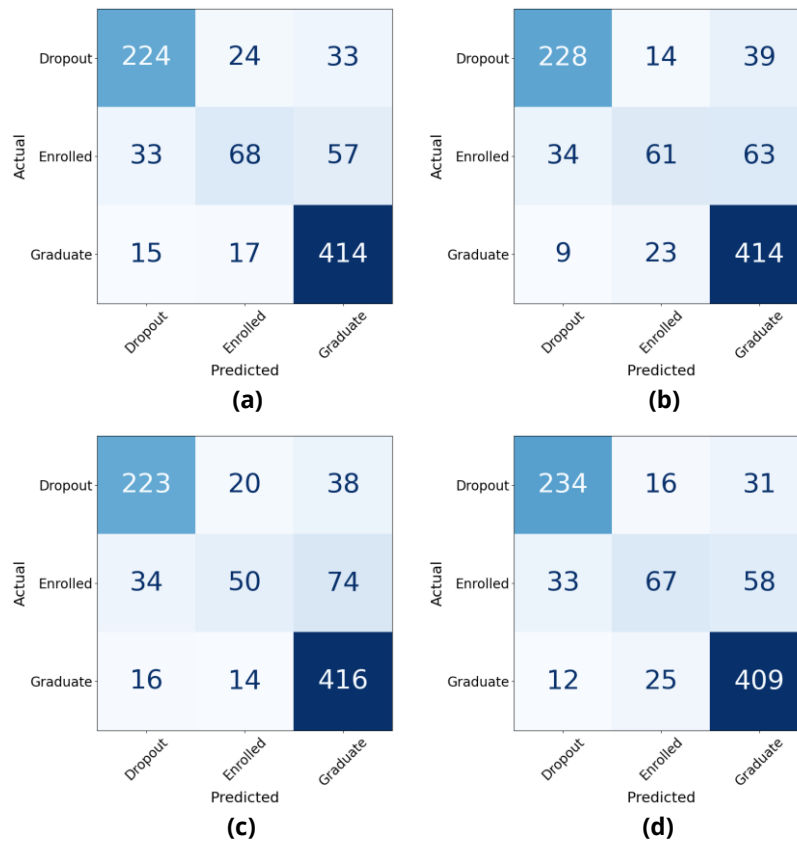


Figure 2. Confusion matrix of the machine learning models, (a) LightGBM, (b) Random Forest, (c) Logistic Regression, (d) Stacked Classifier.

evaluated models. This model demonstrates superior predictive power across all target classes, particularly excelling in the "Graduate" class, where it correctly classified 414 instances. Its misclassification rates for the other classes are notably lower than those of the other models, indicating its ability to distinguish between different categories effectively.

The second-best model is LightGBM. Although it also performs well, it falls slightly behind the stacked classifier in accuracy. The model accurately classified 416 instances in the "Graduate" class and performed reasonably well in the other classes, but its misclassification rates were somewhat higher than those of the stacked Classifier.

On the other hand, the Logistic Regression model had the lowest performance among the evaluated models. Although it accurately classified many instances, it had higher misclassification rates, especially in differentiating between the "Dropout" and "Enrolled" categories. This model's performance demonstrates the limitations of simpler models compared to more advanced ensemble techniques like stacking.

This study underscores the significant potential of machine learning, particularly through stacked classifiers, in advancing educational management by enabling the early identification of students at risk of dropping out and

recognizing those likely to excel academically. By outperforming individual models, the stacked classifier effectively demonstrates how ensemble learning can uncover intricate patterns in educational data, leading to more accurate predictions.

The classifier's superior precision and recall metrics make it especially effective in detecting students at risk of dropping out early, which is critical for timely intervention. This early detection capability allows schools to design targeted support programs, optimize resource allocation, and provide personalized interventions to reduce dropout rates and help students complete their education. Additionally, the model's ability to identify academically promising students enables educators to nurture their talents by providing tailored enrichment programs, advanced coursework, and scholarship opportunities that align with their needs. This fosters academic excellence and nurtures future leaders by offering suitable learning pathways for high-potential students.

For educational institutions, which often face resource constraints, the accurate predictions offered by machine learning models can facilitate more efficient resource allocation. By pinpointing which students are at risk or have the potential to excel, institutions can channel

support services to those needing them most, optimizing funding and staff time. Finally, the findings from this study provide policymakers with valuable insights into the effectiveness of machine learning in educational settings. With accurate predictive models, policymakers can design data-driven policies that address systemic issues contributing to dropout rates while promoting academic excellence. Such policies can encompass targeted intervention funding, improved access to advanced learning opportunities, and comprehensive support for at-risk students.

Despite the promising results, this study has several limitations that should be acknowledged. First, the dataset used is specific to one educational institution and might not reflect the diversity of educational environments and demographics. The machine learning model's performance could vary significantly when applied to different educational systems with distinct characteristics. Moreover, the study focused on academic and socio-economic features available during enrollment, which might not capture dynamic changes throughout the student's educational journey. This limitation means that predictions may not account for significant student circumstances shifts, potentially impacting the model's accuracy over time.

Future studies could address these limitations by incorporating more diverse datasets representing various educational systems to test the model's generalizability. Additionally, exploring methods to simplify the stacked classifier while maintaining predictive accuracy could help make it more accessible for school implementation. Efforts to improve interpretability are also necessary to ensure that educational practitioners can understand and act on the insights provided by the model.

4. Conclusions

The study presented a comprehensive analysis of machine learning for the early detection of dropout risks and academic excellence among students. By leveraging a stacked classifier model combining LightGBM, Random Forest, and logistic regression algorithms, this approach offers superior predictive accuracy compared to standalone classifiers. The model achieved an accuracy of 80.23%, with notable performance in precision, recall, and F1-score, making it a valuable tool for educational institutions aiming to support student success. The findings demonstrate the immense potential of machine learning in transforming educational practices. By accurately predicting student outcomes, schools can efficiently allocate resources to provide targeted interventions, ensuring that support reaches those who

need it most. Furthermore, these predictive insights can inform policymakers to design data-driven strategies that address systemic challenges and promote academic excellence.

Author Contributions: Conceptualization, T.R.N., A.L. and R.I.; methodology, T.R.N., Z.Z. and E.Y.; software, T.R.N. and Z.J.; validation, N.I.S.M.Y., A.L. and R.I.; formal analysis, T.R.N., Z.Z. and E.Y.; investigation, T.R.N. and Z.Y.; resources, M.Y., A.L. and R.I.; data curation, E.Y. and R.I.; writing—original draft preparation, T.R.N., Z.Z., Z.J. and M.Y.; writing—review and editing, N.I.S.M.Y., A.L. and R.I.; visualization, T.R.N.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study does not receive external funding.

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized in this study are available from: <https://doi.org/10.24432/C5MC89>.

Conflicts of Interest: All the authors declare no conflicts of interest.

References

- Chankseliani, M., Qoraboyev, I., and Gimranova, D. (2021). Higher Education Contributing to Local, National, and Global Development: New Empirical and Conceptual Insights, *Higher Education*, Vol. 81, No. 1, 109–127. doi:10.1007/s10734-020-00565-8.
- Calavia, M. B., Blanco, T., and Casas, R. (2021). Fostering Creativity as a Problem-Solving Competence through Design: Think-Create-Learn, a Tool for Teachers, *Thinking Skills and Creativity*, Vol. 39, 100761. doi:10.1016/j.tsc.2020.100761.
- Chen, X. M. (2021). Integration of Creative Thinking and Critical Thinking to Improve Geosciences Education, *The Geography Teacher*, Vol. 18, No. 1, 19–23. doi:10.1080/19338341.2021.1875256.
- Zelenkovski, K., Prodanova, J., and Kocarev, L. (2024). Exploring Citizen Perceptions and Values for a Responsible Society, *Social Science Quarterly*, Vol. 105, No. 2, 296–310. doi:10.1111/ssqu.13353.
- Daniel, J. (2010). *Mega-Schools, Technology and Teachers: Achieving Education for All*, Routledge.
- Sadler, D. R. (2017). Academic Achievement Standards and Quality Assurance, *Quality in Higher Education*, Vol. 23, No. 2, 81–99. doi:10.1080/13538322.2017.1356614.
- Abdallah, A. K., and Abdallah, R. K. (2023). Achieving Academic Excellence: The Intersection of Teacher Development, Quality Education, and Entrepreneurship, *Innovations in Teacher Development, Personalized Learning, and Upskilling the Workforce*, IGI Global, 136–158.
- Nir, A. E., and Zilberstein-Levy, R. (2006). Planning for Academic Excellence: Tenure and Professional Considerations, *Studies in Higher Education*, Vol. 31, No. 5, 537–554. doi:10.1080/03075070600922725.
- Brophy, J. (2004). *Motivating Students to Learn*, Routledge.
- Siegle, D., Rubenstein, L. D., and Mitchell, M. S. (2014). Honors Students' Perceptions of Their High School Experiences, *Gifted Child Quarterly*, Vol. 58, No. 1, 35–50. doi:10.1177/0016986213513496.

11. De Witte, K., Cabus, S., Thyssen, G., Groot, W., and van den Brink, H. M. (2013). A Critical Review of the Literature on School Dropout, *Educational Research Review*, Vol. 10, 13–28. doi:10.1016/j.edurev.2013.05.002.
12. Del Bonifro, F., Gabbriellini, M., Lisanti, G., and Zingaro, S. P. (2020). Student Dropout Prediction, 129–140. doi:10.1007/978-3-030-52237-7_11.
13. Samuel, R., and Burger, K. (2020). Negative Life Events, Self-Efficacy, and Social Support: Risk and Protective Factors for School Dropout Intentions and Dropout., *Journal of Educational Psychology*, Vol. 112, No. 5, 973–986. doi:10.1037/edu0000406.
14. Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A.-E., and Cortadas, P. (2023). An Early Warning System to Identify and Intervene Online Dropout Learners, *International Journal of Educational Technology in Higher Education*, Vol. 20, No. 1, 3. doi:10.1186/s41239-022-00371-5.
15. Bañeres, D., Rodríguez, M. E., Guerrero-Roldán, A. E., and Karadeniz, A. (2020). An Early Warning System to Detect At-Risk Students in Online Higher Education, *Applied Sciences*, Vol. 10, No. 13, 4427. doi:10.3390/app10134427.
16. de Vasconcelos, A. N., Freires, L. A., Loureto, G. D. L., Fortes, G., da Costa, J. C. A., Torres, L. F. F., Bittencourt, I. I., Cordeiro, T. D., and Isotani, S. (2023). Advancing School Dropout Early Warning Systems: The IAFREE Relational Model for Identifying At-Risk Students, *Frontiers in Psychology*, Vol. 14. doi:10.3389/fpsyg.2023.1189283.
17. Maulana, A., Noviandy, T. R., Sasmita, N. R., Paristiowati, M., Suhendra, R., Yandri, E., Satrio, J., and Idroes, R. (2023). Optimizing University Admissions: A Machine Learning Perspective, *Journal of Educational Management and Learning*, Vol. 1, No. 1, 1–7. doi:10.60084/jeml.v1i1.46.
18. Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., and Phasinam, K. (2023). Classification and Prediction of Student Performance Data Using Various Machine Learning Algorithms, *Materials Today: Proceedings*, Vol. 80, 3782–3785. doi:10.1016/j.matpr.2021.07.382.
19. Zeineddine, H., Braendle, U., and Farah, A. (2021). Enhancing Prediction of Student Success: Automated Machine Learning Approach, *Computers & Electrical Engineering*, Vol. 89, 106903. doi:10.1016/j.compeleceng.2020.106903.
20. Noviandy, T. R., Nainggolan, S. I., Raihan, R., Firmansyah, I., and Idroes, R. (2023). Maternal Health Risk Detection Using Light Gradient Boosting Machine Approach, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 48–55. doi:10.60084/ijds.v1i2.123.
21. Maulana, A., Idroes, G. M., Kemala, P., Maulydia, N. B., Sasmita, N. R., Tallei, T. E., Sofyan, H., and Rusyana, A. (2023). Leveraging Artificial Intelligence to Predict Student Performance: A Comparative Machine Learning Approach, *Journal of Educational Management and Learning*, Vol. 1, No. 2, 64–70. doi:10.60084/jeml.v1i2.132.
22. Yağcı, M. (2022). Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms, *Smart Learning Environments*, Vol. 9, No. 1, 11. doi:10.1186/s40561-022-00192-z.
23. Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., and Addison, K. L. (2015). A Machine Learning Framework to Identify Students at Risk of Adverse Academic Outcomes, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, 1909–1918. doi:10.1145/2783258.2788620.
24. Marwaha, A., and Singla, A. (2020). A Study of Factors to Predict At-Risk Students Based on Machine Learning Techniques, 133–141. doi:10.1007/978-981-13-8618-3_15.
25. Muñoz-Carpio, J. C., Jan, Z., and Saavedra, A. (2021). Machine Learning for Learning Personalization to Enhance Student Academic Performance., *LALA*, 88–99.
26. Martins, M. V., Tolledo, D., Machado, J., Baptista, L. M. T., and Realinho, V. (2021). Early Prediction of student's Performance in Higher Education: A Case Study, 166–175. doi:10.1007/978-3-030-72657-7_16.
27. Realinho, V., Vieira Martins, M., Machado, J., and Baptista, L. (2021). Predict Students' Dropout and Academic Success, *UCI Machine Learning Repository*. doi:10.24432/C5MC89.
28. Noviandy, T. R., Idroes, G. M., and Hardi, I. (2024). Enhancing Loan Approval Decision-Making: An Interpretable Machine Learning Approach Using LightGBM for Digital Economy Development, *Malaysian Journal of Computing (MJOC)*, Vol. 9, No. 1, 1734–1745. doi:10.24191/mjoc.v9i1.25691.
29. Noviandy, T. R., Maulana, A., Idroes, G. M., Irvanizam, I., Subianto, M., and Idroes, R. (2023). QSAR-Based Stacked Ensemble Classifier for Hepatitis C NS5B Inhibitor Prediction, *2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE)*, IEEE, 220–225. doi:10.1109/COSITE60233.2023.10250039.
30. Supriatna, D. J. I., Saputra, H., and Hasan, K. (2023). Enhancing the Red Wine Quality Classification Using Ensemble Voting Classifiers, *Infolitika Journal of Data Science*, Vol. 1, No. 2, 42–47. doi:10.60084/ijds.v1i2.95.
31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python, *The Journal of Machine Learning Research*, Vol. 12, 2825–2830.
32. Idroes, G. M., Noviandy, T. R., Maulana, A., Zahriah, Z., Suhendrayatna, S., Suhartono, E., Khairan, K., Kusumo, F., Helwani, Z., and Abd Rahman, S. (2023). Urban Air Quality Classification Using Machine Learning Approach to Enhance Environmental Monitoring, *Leuser Journal of Environmental Studies*, Vol. 1, No. 2, 62–68. doi:10.60084/ljes.v1i2.99.
33. Noviandy, T. R., Maulana, A., Idroes, G. M., Emran, T. B., Tallei, T. E., Helwani, Z., and Idroes, R. (2023). Ensemble Machine Learning Approach for Quantitative Structure Activity Relationship Based Drug Discovery: A Review, *Infolitika Journal of Data Science*, Vol. 1, No. 1, 32–41. doi:10.60084/ijds.v1i1.91.
34. Shah, K., Patel, H., Sanghvi, D., and Shah, M. (2020). A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification, *Augmented Human Research*, Vol. 5, No. 1, 12. doi:10.1007/s41133-020-00032-0.
35. Carvalho, D. V., Pereira, E. M., and Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics, *Electronics*, Vol. 8, No. 8, 832.