G-Tech: Jurnal Teknologi Terapan
Vol. 9, No. 1, 2025, p. 530-539
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Feature Management Approaches and Predictive Methods in Machine Learning and Deep Learning for Student Performance Prediction: A Systematic Review

Desty Yani[1]✉, Kusrini[2]

[1, 2] Magister Teknik Informatika, Universitas Amikom Yogyakarta, Indonesia

✉ **Corresponding Author:** Desty Yani (e-mail: destyyani@students.amikom.ac.id)

Article Information

Article History

Received : January 18, 2025
Revised : January 26, 2025
Published : January 30, 2025

***Keywords*:**
Deep Learning;
Feature Management Approach;
Machine Learning;
Student Performance Prediction;

ABSTRACT

Research on student performance prediction has advanced rapidly in recent years, driven by the increasing volume of educational data generated by digital learning platforms. This data can be analyzed using Machine Learning (ML) and Deep Learning (DL) techniques, integrated with feature management strategies tailored to specific needs. However, selecting the most relevant features and optimizing predictive models remain significant challenges. Different studies apply various feature selection and engineering techniques, leading to inconsistent results and limited generalizability. This study conducts as a Systematic Literature Review (SLR) to explore ML and DL approaches for student performance prediction, emphasizing their relationship with feature management techniques. The reviewed studies span publications from 2019 to 2024. This SLR aims to assist researchers in identifying effective strategies for predicting student performance, including the selection of methods, datasets, or feature management techniques. Most studies utilized publicly available datasets due to their accessibility and ease of use. Among ML methods, Random Forest emerged as the most frequently applied, achieving an F-measure of 99.5% integration of filter-based and wrapper-based feature selection techniques. Among DL approaches, the ANN-PCACSN model, employing Principal Component Analysis (PCA) for dimensionality reduction, achieved the highest accuracy of 99.32%. These findings highlight the importance of aligning preprocessing strategies with dataset properties and algorithm capabilities to enhance predictive performances.

INTRODUCTION

In recent years, many educational institutions have adopted learning systems that deliver data through various platforms, such as Learning Management Systems (LMS), Student Information Systems (SIS), and Massive Open Online Courses (MOOCs). This transition has significantly influenced the volume, variety, and complexity of data available in the education sector (Nawang et al., 2021). Activity-related data, including students' frequency of accessing course materials, participation in examinations, login history, and the maximum duration spent on the platform, offer valuable insights into student behavior and academic performance. When processed and analyzed effectively, these data points can reveal patterns, evaluate the learning process (Alnasyan et al., 2024), and predict students' academic outcomes (Abubakari, 2020).

Extensive research on predicting student performance has been conducted with diverse focuses, ranging from secondary education levels (Nachouki et al., 2023) to higher education (Yağcı, 2022), MOOCs (Al-azazi & Ghurab, 2023), academic achievements (Bujang et al., 2021), and even identifying students at risk of low performance (Liao et al., 2019). The availability of vast

This is an open access article under the CC BY license

amounts of educational data and varying requirements poses significant challenges in accurately assessing student performance. Researchers have identified relationships between variables and factors that influence predictive models of student performance (B. Guo et al., 2015). These findings have contributed to the development of more comprehensive and integrative models, particularly with the application of Machine Learning (ML) and Deep Learning (DL) approaches (Alyahyan & Düştegör, 2020).

ML and DL are commonly applied approaches for processing educational data. ML demonstrates the ability to analyze and process data efficiently while offering high accuracy in predicting student performance. Techniques such as Random Forest, Decision Tree, and Support Vector Machine are frequently utilized for this purpose (Mengash, 2020). With the increasing complexity of data, advancements in ML have led to the development of DL, which has gained popularity in recent years for predicting student performance (Haris et al., 2021). Leveraging complex neural networks, DL enables deeper analysis and provides highly accurate predictions. Furthermore, DL can process various data types, including text, images, and sequential data, which opens new possibilities for analyzing educational datasets (Alturki et al., 2022).

The effectiveness of predictive models largely depends on preprocessing techniques, particularly on feature management strategies, including Feature Selection (FS) and Feature Engineering (FE). FS focuses on identifying and selecting the most relevant features from datasets, which reduces noise, improves computational efficiency, and enhances model accuracy (Syed Mustapha, 2023), For instance, (Bujang et al., 2021) employed a combination filter-based and wrapper based for FS. In contrast, FE involves creating or transforming features to better represent the underlying data, enabling the model to uncover complex relationships and improve its predictive capabilities (Gupta & Gueneau, n.d.). For example, (Ali et al., 2023) applied Principal Component Analysis (PCA) to reduce dimensionality, thereby optimizing the predictive performance.

Several researchers have reviewed and summarized studies related to educational data analysis, focusing on predicting student performance through Systematic Literature Review (SLR). SLR is a structured method for collecting, critically analyzing, integrating, and presenting research from various sources (Pati & Lorusso, 2018). For example, (Alnasyan et al., 2024) conducted an SLR explaining ML and DL techniques, (Nawang et al., 2021) discussed general approaches to predicting student performance, while (Haziq Muhammad & Chen Jen Chwen, 2022) focused on Educational Data Mining (EDM) from 2015 to 2021.

This study aims to conduct a systematic literature review focusing on research related to predicting student performance, specifically utilizing ML and DL technologies, as well as their relationship with feature management. The reviewed studies span publications from 2019 to 2024. This SLR is expected to assist researchers in determining effective strategies for predicting student performance, including the choice of methods, datasets, or feature management techniques.

RESEARCH METHODS

This research employed a SLR approach which was structured into three main steps: planning, review processing, and reporting. Planning and review processing are discussed in this section, while reporting is presented in the results section.

Planning

The planning phase consists of five key steps: defining the research question, identifying keywords and sources, determining inclusion and exclusion criteria and finally deciding on the data extraction strategy. For defining the research question, this study adopts the PICOC criteria, as outlined in Table 1.

Based on these criteria, the Research Questions (RQ) are identified as follows:
RQ1: What datasets are commonly used for student performance prediction?
RQ2: What are the main approaches for implementing feature management in datasets?
RQ3: What methods are commonly used for student performance prediction?

RQ4: How do feature management approaches impact the performance of predictive algorithms in student performance prediction?

Table 1. PICOC criteria

Population	Intervention	Outcomes	Context
High school and University students as well as participants on digital learning platform	The implementation of ML and DL algorithm in prediction methods	The dataset used and the effectiveness of the algorithm in making prediction	Educational institutions and digital learning platforms

After defining the research questions (RQs), the next step is to identify the keywords. Keywords include terms such as "student performance prediction" and "student performance prediction method." The third step focuses on identifying sources and retrieving for journal articles from online databases including ScienceDirect, IEEE, ACM, Springer, and platforms like Google Scholar, Connected Papers, and Perplexity. Following the search results, the inclusion and exclusion criteria are established to filter journal articles according based on the specific research objectives. Table 2 illustrates the criteria applied in this SLR.

Table 2. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
High school and University students as well as participants on digital learning platform	Journal articles focusing on the implementation of ML and DL criteria unrelated to prediction methods.
Journal Articles published between 2019–2024	Journal Article published before 2019
Journal Articles with content methods, datasets, and prediction performance results	Journal articles with insufficient content regarding methods, datasets, or prediction performance, as required in the inclusion criteria
Journal Articles based on relevant keywords	Duplication Journal Article from difference resource

The final phase in the planning section involves deciding on the extraction strategy for selecting the relevant journal articles. These articles are organized into a table with the following columns, as defined in Table 3:

Table 3. Data Layout

Column	Description
Id Artikel	An Id assigned to facilitate easy access and review
Title	The Title of the journal article
Year	The year the journal article was published
Publisher	The name of Publisher on the journal article
Method	The algorithm applied in the research
Performance	The result achieved using the applied methods
Technology	Categorization of algorithms into ML or DL
Amount datasets	The total number of datasets used
Source datasets	Classification of the source of the datasets

Review Process

The review process in this SLR involves multiple phases meticulously designed to collect, evaluate, and synthesize knowledge from the selected journal articles. These phases encompass the identification and selection of relevant articles. During the identification phase, journal articles were filtered according to the criteria specified in Table 2. As a result, 50 journal articles were identified, with their distribution illustrated in Figure 1

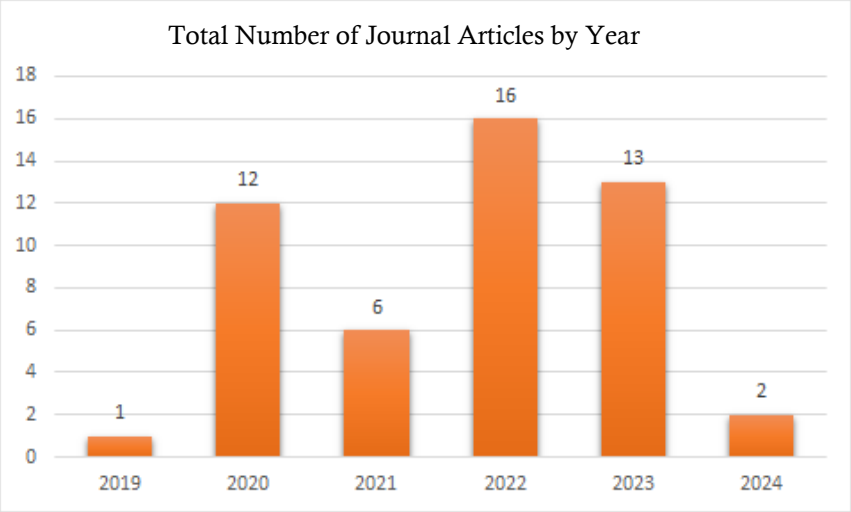

Figure 1. Total number of journal articles by year

Figure 1 illustrates the distribution of articles based on the year of publication. The majority of articles were published in 2022 (16 articles), followed by 2023 (13 articles) and 2020 (12 articles). A smaller number of articles were published in 2021 (6 articles), and 2019 (1 article). This distribution highlights the increasing research focus on student performance prediction in recent years, with a peak in 2022. The next step in this research involves the selection and review of articles based on the inclusion and exclusion criteria previously discussed. The data layout outlined in Table 3 is employed to systematically manage and present the extracted information. This framework facilitates the organization and evaluation of each article, ensuring its relevance and adherence to the criteria. Out of the 50 articles evaluated, 35 articles satisfied the criteria and were selected for further analysis. The selected articles are detailed in Table 4.

Table 4. The Selected Articles

Id Article	Researcher	Id Article	Researcher
J1	(Nachouki et al., 2023)	J19	(Ali et al., 2023)
J2	(Mengash, 2020)	J20	(Al-Zawqari et al., 2022)
J3	(Alturki et al., 2022)	J21	(Beckham et al., 2022)
J4	(Al-azazi & Ghurab, 2023)	J22	(Bujang et al., 2021)
J5	(Feng et al., 2022)	J23	(Chen et al., 2023)
J6	(Liao et al., 2019)	J24	(Gámez-Granados et al., 2023)
J7	(Ünal, 2020)	J25	(Alshdaifat, 2020)
J8	(Khairy et al., 2024)	J26	(Dien et al., 2020)
J9	(Alhazmi & Sheneamer, 2023)	J27	(Brahim, 2022)
J10	(Yağcı & Olpak, 2021)	J28	(El-Keiey et al., n.d.)
J11	(Orji Fidelia A. & Vassileva Julita, 2022)	J29	(Ghosh et al., 2022)
J12	(Dervenis et al., 2022)	J30	(Liu et al., 2020)
J13	(Mahmood AB & Khalfan AM, 2020)	J31	(Priyambada et al., 2023)
J14	(Ojajuni et al., 2021)	J32	(Pérez Sánchez et al., 2022)
J15	(Nasser Alsubaie, 2023)	J33	(Resce et al., 2022)
J16	(Namraiza et al., 2023)	J34	(Ben Said et al., 2024)
J17	(Mastour et al., 2023)	J35	(Jiao, 2022)
J18	(Santos & Henriques, 2023)		

RESULTS AND DISCUSSION

The third step in this SLR is reporting. In this stage, the analysis results from the selected papers are summarized and documented based on the research questions.

RQ1: What Datasets are Commonly Used for Student Performance Prediction?

From table 6, it concluded that public datasets are the most commonly used, accounting for 40% of the total research. These datasets are typically freely accessible through websites such as Kaggle, UCI *Machine learning* and *Web of Science* (WoS). Research utilizing datasets from universities and academic institutions accounts for 34%. Furthermore 23% research employs *Open Education Dataset* such as *Open University Learning Analytics Dataset* (OULAD), only 3% of the research used a combination of public datasets and institutional datasets.

Table 5. Datasets used in student performance prediction

Type of Dataset	Reference	Count	Percent (%)
Dataset Publik	J7, J8, J9, J10, J12, J13, J14, J16, J17, J18, J25, J26, 731, J32	14	40%
Institutional Education	J1, J2, J3, J5, J6, J11, J15, J21, J22, J23, J28, J29	12	34%
Open Education Dataset	J4, J19, J20, J24, J27, J30, J34, J35	8	23%
Data Publik & Institutional	J33	1	3%

RQ2: What are the main approaches for feature management in datasets?

Based on a comprehensive review of the literature, nine primary approaches to feature management for datasets have been identified. These include Embedded Approaches (EA), Wrapper-Based (WB), Filter-Based (FB), Combination (CMB) (e.g., integrating filter-wrapper or wrapper-embedded methods), Dimensionality Reduction (DR), Hybrid Optimization (HO), Graph-Based (GB), Fuzzy Rule-Based Optimization (FRB), and the No Feature Selection (NFS) approach. The selection of an appropriate strategy depends on contextual factors such as dataset characteristics and the complexity of the predictive task.

Among these, the EA is the most frequently utilized, representing 29% of the studies reviewed. Following EA, FB and DR approaches are equally prominent, each applied in 17% of the studies. The WB approach, employed in 9% of the studies, evaluates feature subsets by iteratively training and testing models on various feature combinations. Additionally, some studies employ hybrid methods by combining approaches, such as FB and WB techniques, to leverage the strengths of each. Less commonly applied methods, including HO, GB, FRB, and EBL, are observed in 3% to 6% of the studies.

Table 6. Feature Management Approaches

Feature Management Approaches	Reference	Count	Percent (%)
Embedded (EA)	J3, J11, J12, J14, J17, J18, J20, J27, J29, J32	10	29%
Dimensionality Reduction (DR)	J4, J16, J19, J26 J30, J34, J35	7	20%
Filter-Based (FB)	J2, J6, J8, J13, J25, J28	6	17%
Wrapper-Based (WB)	J1, J5, J7	3	9%
Combination (CMB)	J9, J21, J22	3	9%
Graph-Based (GB)	J23, J33	2	6%
Hybrid Optimization (HO)	J15	1	3%
Fuzzy Rule-Based (FRB)	J24	1	3%
No Feature Selection (NFS)	J10	1	3%
Ensemble-Based Learning (EBL)	J31	1	3%

RQ3: What are The Methods Commonly Used for Student Performance Prediction?

The results of the commonly applied methods in this research are summarized in Table 5, categorized into ML and DL technologies. ML methods include common algorithms such as Random Forest (RF), Linear Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Extreme Gradient Boosting (XGBOOST), and Ensemble-Based Progressive Prediction (EPP), along with other notable methods like the FlexNSLVOrd algorithm, Artificial Neural Network (ANN), and Multilayer Perceptron (MP). Among these, RF emerges as the most frequently used ML algorithm. ML algorithms collectively cited in 14 studies, accounting for 74% of the total references. Additionally, these algorithms have been applied using various feature management approaches, as shown in Table 7.

In contrast, DL techniques encompass advanced algorithms, including Artificial Neural Network (ANN), Artificial Neural Network–Long Short Term Memory (ANN-LSTM) Convolutional Neural Network (CNN) and their variants such as, Artificial Neural Network–Principal Component Analysis Convolutional Spiking Neural Network (ANN-PCACSN), Recurrent Multilayer Hopfield Neural Network (RMHNN), Recurrent Neural Network-), Long Short Term Memory (RNN-LSTM), Gramian Angular Field–Convolutional Neural Network (GAF-CNN), and Principal Component Analysis–Factorization Deep Product Neural Network (PCA-FDPN). These DL algorithms employ a variety of feature management approaches, including FB, DR, WB, HO, and GB approaches. Out of the 35 references, DL algorithms are cited in 9 studies, accounting for 26% of the total references.

Table 7. Methods used in student performance prediction

Technology	Algorithm	Reference	Feature Management Approaches	Count	Percent (%)
Machine Learning	RF	J1, J3, J7, J9, J11, J12, J17, J18, J20, J22, J27, J29, J32, J33	WB, EA, WB, FB, EA, NFS, EA, EA, EA, FB, EA, EA, EA, GB	14	74%
	SVM	J6, J8, J15, J25	FB, FB, HO, FB	4	
	KNN	J10	NFS	1	
	SVC	J13	FB	1	
	XGBOOST	J14	EA	1	
	MLP	J21	CMB	1	
	FlexNSLVOrd	J24	FR	1	
	DT	J28	FB	1	
	EPP	J31	EA	1	
	NN	J16	DR	1	
Deep Learning	ANN	J2	FB	1	26%
	ANN-LSTM	J4	DR	1	
	K-Means-CNN	J5	WB	1	
	ANN-PCACSN	J19	HO	1	
	RMHNN	J23	GB	1	
	CNN-LSTM	J26	DR	1	
	RNN-LSTM	J30	DR	1	
	GAF-CNN	J34	DR	1	
	PCA-FDPN	J35	DR	1	

RQ4: How do feature management approaches impact the performance of predictive algorithms in student performance prediction?

The analysis results presented in Table 8 highlight the critical connection between algorithms, feature management approaches, and evaluation outcomes in enhancing the accuracy of student performance prediction. Four key algorithms with outstanding performance are XGBOOST, ANN-PCACSN, Random Forest, and RNN-LSTM, each employing distinct feature management approaches. XGBOOST (J14) achieved an accuracy of 97.12% and a cross-validation score of 35.67%, leveraging an embedded approach through the plot_importance function. Random Forest (J22) demonstrated exceptional performance with an F-measure of 99.5%, combining filter-based and wrapper-based feature selection methods and optimizing the model using SMOTE.

Additionally, ANN-PCACSN (J19) applied a dimensionality reduction approach using PCA and optimization via the Cuckoo Search Algorithm, achieving an accuracy of 99.32%. Meanwhile, RNN-LSTM (J30) utilized a dimensionality reduction technique combined with the Attention Mechanism Enhanced Deep Knowledge Tracing (MFA-DKT) framework, resulting in an accuracy of 97% and an AUC of 98%. These findings underscore the importance of selecting the appropriate algorithm and integrating effective feature management approaches to significantly enhance prediction performance in educational contexts.

Table 8. Performance of algorithms with their evaluation results and feature management approaches

Algorithm	Reference	Evaluation Result	Feature Management Approaches
XGBOOST	J14	Accuracy: 97.12% Cross-validation: 35,67%	Embedded Approach
Random Forest	J22	f-measure: 99.5%	Combination Filter-Based and Wrapper-Based
ANN-PCACSN	J19	Accuracy: 99.32%	Dimensionality Reduction
RNN-LSTM	J30	Accuracy: 97% AUC: 98%	Dimensionality Reduction

CONCLUSION

This study conducted a Systematic Literature Review (SLR) to explore the application of ML and DL algorithms integrated with feature management strategies in predicting student performance. The analysis highlights the critical role of methods like the embedded approach and dimensionality reduction in improving model accuracy, particularly for large, heterogeneous, and complex datasets. Public and institutional datasets, frequently utilized in these studies, demonstrated substantial improvements in predictive performance when paired with appropriate feature optimization techniques, underscoring the important connection between dataset characteristics and preprocessing strategies.

The impact of these strategies is evident, as Random Forest (RF), the most frequently used ML algorithm in this SLR, achieved an exceptional F-measure of 99.5% through the combined use of filter-based and wrapper-based feature selection. For DL algorithms, which are typically applied to high-complexity and large datasets, advanced processing models are essential. From the analysis, the ANN-PCACSN model, employing dimensionality reduction techniques like PCA, achieved the highest accuracy of 99.32%. These findings highlight the necessity of aligning preprocessing strategies with dataset properties and algorithm capabilities to maximize predictive performance. Future research should focus on integrating optimization methods to further address challenges posed by increasingly complex educational datasets.

ACKNOWLEDGEMENTS

The author expresses gratitude for the guidance and support provided by the Supervisors and fellow students of the Magister of Informatics Engineering at Universitas Amikom Yogyakarta.

REFERENCES

Abubakari, M. S. (2020). *Educational Data Mining to Predict Students Performance Based on Deep Learning Neural Network*.

Mahmood AB, & Khalfan AM. (2020). *Machine Learning Based Predicting Student Academic Success*. IEEE.

Al-azazi, F. A., & Ghurab, M. (2023). ANN-LSTM: A deep learning model for early student performance prediction in MOOC. *Heliyon*, *9*(4). https://doi.org/10.1016/j.heliyon.2023.e15382

Alhazmi, E., & Sheneamer, A. (2023). Early Predicting of Students Performance in Higher Education. *IEEE Access*, *11*, 27579–27589. https://doi.org/10.1109/ACCESS.2023.3250702

Ali, M., Liaquat, M. D., Atta, M. N., Khan, A., Lashari, S. A., & Ramli, D. A. (2023). Improving Student Performance Prediction Using a PCA-based Cuckoo Search Neural Network Algorithm. *Procedia Computer Science*, *225*, 4598–4610. https://doi.org/10.1016/j.procs.2023.10.458

Alnasyan, B., Basheri, M., Alassaa, M., & Alassafi, M. (2024). *Deep Learning Techniques for Predicting Student's Academic Performance on Virtual Learning Environments: A Review*. https://doi.org/10.21203/rs.3.rs-3888441/v1

Alshdaifat, E. (2020). The Impact of Data Normalization on Predicting Student Performance: A Case Study from Hashemite University. *International Journal of Advanced Trends in Computer Science and Engineering*, *9*(4), 4580–4588. https://doi.org/10.30534/ijatcse/2020/57942020

Alturki, S., Cohausz, L., & Stuckenschmidt, H. (2022). Predicting Master's students' academic performance: an empirical study in Germany. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00220-y

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. In *International Journal of Educational Technology in Higher Education* (Vol. 17, Issue 1). Springer. https://doi.org/10.1186/s41239-020-0177-7

Al-Zawqari, A., Peumans, D., & Vandersteen, G. (2022). A flexible feature selection approach for predicting students' academic performance in online courses. *Computers and Education: Artificial Intelligence*, *3*. https://doi.org/10.1016/j.caeai.2022.100103

Beckham, N. R., Akeh, L. J., Mitaart, G. N. P., & Moniaga, J. V. (2022). Determining factors that affect student performance using various machine learning methods. *Procedia Computer Science*, *216*, 597–603. https://doi.org/10.1016/j.procs.2022.12.174

Ben Said, A., Abdel-Salam, A. S. G., & Hazaa, K. A. (2024). Performance prediction in online academic course: a deep learning approach with time series imaging. *Multimedia Tools and Applications*, *83*(18), 55427–55445. https://doi.org/10.1007/s11042-023-17596-9

Brahim, G. Ben. (2022). Predicting Student Performance from Online Engagement Activities Using Novel Statistical Features. *Arabian Journal for Science and Engineering*, *47*(8), 10225–10243. https://doi.org/10.1007/s13369-021-06548-w

Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, *9*, 95608–95621. https://doi.org/10.1109/ACCESS.2021.3093563

Chen, Z., Cen, G., Wei, Y., & Li, Z. (2023). Student Performance Prediction Approach Based on Educational Data Mining. *IEEE Access*, *11*, 131260–131272. https://doi.org/10.1109/ACCESS.2023.3335985

Dervenis, C., Kyriatzis, V., Stoufis, S., & Fitsilis, P. (2022, September 16). Predicting Students' Performance Using Machine Learning Algorithms. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3564982.3564990

Dien, T. T., Luu, S. H., Thanh-Hai, N., & Thai-Nghe, N. (2020). Deep Learning with Data Transformation and Factor Analysis for Student Performance Prediction. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 11, Issue 8). www.ijacsa.thesai.org

El-Keiey, S., Elmenshawy, D., & Hassanein, E. (n.d.). Student's Performance Prediction based on Personality Traits and Intelligence Quotient using Machine Learning. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 13, Issue 9). www.ijacsa.thesai.org

Feng, G., Fan, M., & Chen, Y. (2022). Analysis and Prediction of Students' Academic Performance Based on Educational Data Mining. *IEEE Access*, *10*, 19558–19571. https://doi.org/10.1109/ACCESS.2022.3151652

Gámez-Granados, J. C., Esteban, A., Rodriguez-Lozano, F. J., & Zafra, A. (2023). An algorithm based on fuzzy ordinal classification to predict students' academic performance. *Applied Intelligence*, *53*(22), 27537–27559. https://doi.org/10.1007/s10489-023-04810-2

Ghosh, S. K., Janan, F., & Ahmad, I. (2022). Application of the Classification Algorithms on the Prediction of Student's Academic Performance. *Trends in Sciences*, *19*(14). https://doi.org/10.48048/tis.2022.5070

Gupta, R., & Gueneau, C. (n.d.). *Feature Correlation with Student Education Performance*. www.JSR.org

Haris, M., Pustaka, T., Diponegoro, M. H., Kusumawardani, S., & Hidayah, I. (2021). Tinjauan Pustaka Sistematis: Implementasi Metode Deep Learning pada Prediksi Kinerja Murid (Implementation of Deep Learning Methods in Predicting Student Performance: A Systematic Literature Review). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 10, Issue 2).

Haziq Muhammad, & Chen Jen Chwen. (2022). 9.View of Educational Data Mining for Student Performance Prediction_ A Systematic Literature Review (2015-2021). *IJET*, *Vol. 17*. https://doi.org/https://doi.org/10.3991/ijet.v17i05.27685

Jiao, X. (2022). A Factorization Deep Product Neural Network for Student Physical Performance Prediction. *Computational Intelligence and Neuroscience*, *2022*, 1–8. https://doi.org/10.1155/2022/4221254

Khairy, D., Alharbi, N., Amasha, M. A., Areed, M. F., Alkhalaf, S., & Abougalala, R. A. (2024). Prediction of student exam performance using data mining classification algorithms. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-12619-w

Liu, D., Zhang, Y., Zhang, J., Li, Q., Zhang, C., & Yin, Y. (2020). Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction. *IEEE Access*, *8*, 194894–194903. https://doi.org/10.1109/ACCESS.2020.3033200

Mastour, H., Dehghani, T., Moradi, E., & Eslami, S. (2023). Early prediction of medical students' performance in high-stakes examinations using machine learning approaches. *Heliyon*, *9*(7). https://doi.org/10.1016/j.heliyon.2023.e18248

Mengash, H. A. (2020). Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*, *8*, 55462–55470. https://doi.org/10.1109/ACCESS.2020.2981905

Nachouki, M., Mohamed, E. A., Mehdi, R., & Abou Naaj, M. (2023). Student course grade prediction using the random forest algorithm: Analysis of predictors' importance. *Trends in Neuroscience and Education*, *33*. https://doi.org/10.1016/j.tine.2023.100214

Namraiza, Abid, K., Aslam, N., Fuzail, M., Maqbool, M. S., & Kainat Sajid. (2023). An Efficient Deep Learning Approach for Prediction of Student Performance Using Neural Network. *VFAST Transactions on Software Engineering*, *11*(4), 67–79. https://doi.org/10.21015/vtse.v11i4.1647

Nasser Alsubaie, M. (2023). Predicting student performance using machine learning to enhance the quality assurance of online training via Maharat platform. *Alexandria Engineering Journal*, *69*, 323–339. https://doi.org/10.1016/j.aej.2023.02.004

Nawang, H., Makhtar, M., & Hamzah, W. M. A. F. W. (2021). A systematic literature review on student performance predictions. In *International Journal of Advanced Technology and Engineering Exploration* (Vol. 8, Issue 84, pp. 1441–1453). Accent Social and Welfare Society. https://doi.org/10.19101/IJATEE.2021.874521

Ojajuni, O., Ayeni, F., Akodu, O., Ekanoye, F., Adewole, S., Ayo, T., Misra, S., & Mbarika, V. (2021). Predicting Student Academic Performance Using Machine Learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12957 LNCS*, 481–491. https://doi.org/10.1007/978-3-030-87013-3_36

Orji Fidelia A., & Vassileva Julita. (2022). *Machine Learning Approach for Predicting Students Academic Performance and Study Strategies based on their Motivation*. https://doi.org/https://doi.org/10.48550/arXiv.2210.08186

Pérez Sánchez, C. J., Calle-Alonso, F., & Vega-Rodríguez, M. A. (2022). Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform. *Education and Information Technologies*, *27*(9), 12913–12938. https://doi.org/10.1007/s10639-022-11128-y

Priyambada, S. A., Usagawa, T., & ER, M. (2023). Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100149

Resce, G., Zinilli, A., & Cerulli, G. (2022). Machine learning prediction of academic collaboration networks. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-26531-1

Santos, R. M., & Henriques, R. (2023). Accurate, timely, and portable: Course-agnostic early prediction of student performance from LMS logs. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100175

Syed Mustapha, S. M. F. D. (2023). Predictive Analysis of Students' Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods. *Applied System Innovation*, *6*(5). https://doi.org/10.3390/asi6050086

Ünal, F. (2020). *Data Mining for Student Performance Prediction in Education*. www.intechopen.com

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1). https://doi.org/10.1186/s40561-022-00192-z

Yağcı, M., & Olpak, Y. Z. (2021). *Predicting Academic Failure: Estimating Semester Grade Point Average with Data Mining Methods*. https://doi.org/10.21203/rs.3.rs-698647/v1