12th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Čačak, 9-12 June 2025

An Overview of Supervised Learning Techniques in Contemporary Educational Technologies

Veljko Aleksić
Faculty of Technical Sciences Čačak
University of Kragujevac
Čačak, Serbia
veljko.aleksic@ftn.kg.ac.rs ORCID
0000-0003-2337-1288

***Abstract*—The rapid advancement of artificial intelligence has catalyzed significant transformations in educational technology, enabling more personalized, adaptive, and data-driven approaches to teaching and learning. The paper examines the theoretical foundations and practical applications of various supervised learning algorithms. The analysis reveals that these techniques offer powerful tools for predicting student performance, automating assessment processes, identifying at-risk students, and personalizing learning. However, their implementation faces challenges related to data quality, model interpretability, and algorithmic bias. The review contributes to the growing body of knowledge on AI in education by synthesizing current research, identifying implementation challenges, and proposing pathways for responsible and effective integration of supervised learning techniques in educational contexts.**

***Keywords*— supervised learning, educational technology, artificial intelligence, algorithms, students**

I. Introduction

Stimulated by the rapid advancement of artificial intelligence technologies and the growing recognition of their potential to transform pedagogical practices and learning experiences, the convergence of artificial intelligence (AI) and educational technologies has emerged as a critical research question in the last decade [1][2]. Let us recall the historical trajectory of artificial intelligence implementation in education, from the early days of Computer-Assisted Instruction (CAI) and Intelligent Tutoring Systems (ITS) [3][4], to the current era of adaptive personalized learning environments [5]. This historical perspective illuminates the dynamic development of the relationship between artificial intelligence and educational technology, highlighting some key transformations in the educational paradigm [6]. Artificial intelligence encompasses a broad spectrum of techniques and algorithms that enable the performance of tasks typically requiring human intelligence, such as learning, problem-solving, and decision-making [7].

The rest of the paper is organized as follows. Section 2 provides a brief overview of the supervised learning paradigm and its implementation in an educational context. The following two sections describe the main characteristics of the most commonly used classification and regression algorithms. The last section gives concluding remarks on the topic.

II. Supervised Learning

Supervised Learning represents a foundational machine learning paradigm wherein algorithms learn from labeled data in order to predict or make decisions regarding the new, previously unseen/unknown data [8]. In the context of educational technology, supervised learning techniques have been extensively used for various tasks, such as predicting student performance [9], automated essay scoring and assessment [10], forecasting student dropout [11], etc.

Supervised learning algorithms establish a mapping function from input characteristics to output variables based on a labeled training dataset [12]. The training data comprises input-output pairs, where properties or attributes of data instances are represented by the input characteristics, and the output variables represent target values or corresponding labels. The supervised learning objective is to establish a generalized mapping function that can accurately predict the outputs for new, previously unseen/unknown input data. This training/learning process involves minimizing a loss function which quantifies the discrepancies between predicted outputs and actual (achieved) training data outputs [13]. The algorithm iteratively adjusts parameters of the model in order to minimize the loss function, thus effectively learning underlying data patterns and relationships. Common loss functions include mean squared error (MSE) for regression tasks and cross-entropy for classification tasks [14]. Based on the previously stated, supervised learning algorithms can generally be dually categorized into classification and regression algorithms. Classification algorithms aim to predict discrete output variables based on the input characteristics (e.g., student dropout status or pass/fail student task) [15]. Typical classification algorithm examples are Logistic Regression, Decision Trees, Random Forest, Naive Bayes Classifiers, and Support Vector Machines (SVMs). Conversely, regression algorithms predict continuous output variables based on the input characteristics (e.g., student grades or test scores). Examples of regression algorithms are Linear Regression, Polynomial Regression, and Support Vector Regression (SVR). The k-Nearest Neighbors (k-NN) algorithm can function both as a classification and regression algorithm.

As implicitly indicated, there are numerous applications of supervised learning in educational domain enabling clear insights into teaching/learning processes and facilitating data-driven decision-making to support both student and teacher activities. A prominent application involves predicting student performance wherein supervised learning algorithms are utilized to forecast academic success (e.g., grade or test result). By leveraging longitudinal student data, including demographic characteristics, prior academic success and/or learning activity records, supervised learning algorithms can identify students at risk of poor performance and indirectly provide the teacher proactive intervention [16]. Another significant application is Automated Essay Scoring (AES) [17] that employ supervised learning algorithms to evaluate and grade essays based on

ISBN 978-86-6200-031-6 ©2025 AM & ETRAN

https://doi.org/10.69994/12Ic25014

various linguistic features and characteristics, such as grammar, vocabulary, coherence, and content relevance. These systems can provide immediate feedback to students and assist teachers in assessing large numbers of submissions, thereby conserving their time and effort [18]. However, the implementation of these systems has raised concerns regarding the reliability and fairness of automated assessment, emphasizing the necessity for human oversight and interpretation. Supervised learning is also applied for predicting tertiary education student dropout, particularly in online distance learning programs [19] through the analysis of demographic, academic, and behavioral student data. Early identification of high-risk students enables teachers and institutions to timely provide personalized support, tutoring, or academic counseling in order to assist them.

Despite offering powerful tools for educational application, supervised learning implementation presents several challenges that must be addressed in order to ensure its effective and responsible implementation. A primary concern involves the availability and quality of labeled data [20]. Supervised learning algorithms require substantial quantity of labeled training data to develop accurate and generalized models. However, obtaining/creating labeled data in educational environments is usually time-consuming, expensive, and subjected to privacy and ethical concerns, necessitating the synchronization among educational institutions, researchers, and education policy makers to develop standardized and safe/secure data collection and sharing practices. Complex models, such as Deep Neural Networks, may achieve high accuracy in prediction tasks but often lack satisfactory interpretability, making it difficult for students and teachers to understand the rationale behind specific decisions. The lack of transparency can readily undermine the confidence in any AI-based educational technology. To mitigate this issue, feature importance analysis, rule extraction, and visual explanation techniques are employed. Algorithmic bias usually occurs when models learn and perpetuate biases that existed in training data, and may result in unfair or discriminatory outcomes for certain student groups. Regular audits, bias mitigation techniques, and inclusive data collection are essential for ensuring fairness and equity in supervised learning educational application [21].

A potential future direction for supervised learning development is its integration with other AI techniques, such as unsupervised learning, reinforcement learning, and natural language processing, to create more comprehensive and adaptive educational tools and systems [22]. Combination of supervised learning for performance prediction with reinforcement learning for personalized content sequencing may enable truly individualized learning experiences [23]. Another alternative direction of further development is explainable artificial intelligence (XAI) technology that is specifically tailored for educational applications [24]. This technology aims providing comprehensible decision explanations and predictions made by AI models, promoting transparency and trust. Alternative approaches are also being explored, such as rule-based explanations, counterfactual explanations, and interactive visualizations, in order to make the supervised learning models more usable and interpretable. The increased availability of multimodal data in educational environments has provided new opportunities for supervised learning refinement. Multimodal supervised learning algorithms can leverage information contained in various data modalities (text, audio, image, video, sensory data, etc.) to gain a more comprehensive understanding of the learning process and enable innovative applications, such as recognizing students' affective states, predicting engagement levels, and utilizing intelligent tutoring systems [25].

III. CLASSIFICATION ALGORITHMS

A. Decision Trees and Random Forest

Decision Trees and Random Forest are powerful supervised learning methods that have found extensive application in educational data mining and learning analytics, facilitating the development of predictive models for student success, attrition and dropout forecast, and identifying factors that influence educational achievements [9].

Decision Trees constitute a versatile and intuitive non-parametric supervised learning method that recursively partitions the input space into homogeneous subregions based on the most informative characteristics. The algorithm constructs a tree-like model by selecting the optimal feature for data division at each node, with the objectives to maximize the information gain or to minimize the heterogeneity of the resulting subsets. The tree structure comprises internal nodes (representing features), branches (representing decision rules), and leaves (representing target values or class labels). The recursive partitioning process continues until a termination criterion is satisfied, such as reaching maximum depth or the minimum number of instances per leaf. Decision Trees can process both categorical and numerical characteristics and are capable of identifying non-linear relationships within the data [26]. Decision Trees algorithms are frequently employed for predicting student performance and identifying factors that influence academic achievement. The advantages of this approach are its interpretability, the ability to function with missing values, and resilience to outliers. Furthermore, Decision Trees can capture non-linear relationships and interactions between characteristics, making them suitable for modeling complex educational phenomena. Conversely, these algorithms are prone to overfitting, particularly when the tree becomes deep and complex [27]. This issue can be addressed through the Random Forest method, which combines multiple Decision Trees to enhance predictive performance and reduce overfitting.

The Random Forest method constructs a large number of Decision Trees and aggregates their predictions to improve accuracy and control overfitting [28]. This algorithm introduces randomness by utilizing sampling to create different training sets for each tree thereby reducing variance (bagging), or by selecting a random subset of characteristics at each split node (feature bagging). The training of a Random Forest model encompasses the following sequential steps [29]: (1) Creating bootstrap samples by randomly sampling instances and replacing them with the original training data; (2) Constructing a Decision Tree for each bootstrap sample through recursive partitioning based on a random subset of features at each node; (3) Repeating the previous two steps to generate an ensemble of Decision Trees. To perform the prediction, outputs of individual trees must be aggregated through majority voting (for classification tasks) or by calculating the mean value (for regression tasks). Random Forests offer several advantages compared to individual Decision Trees: they are more robust for overfitting, as the incorporated randomness helps reduce model variance; they can effectively process high-dimensional data and are less sensitive to hyperparameter selection (e.g., maximum depth or minimum number of instances per leaf); they provide a measure of feature importance which can be utilized to identify the most relevant predictors and facilitate informed characteristics selection. For example, by combining multiple weak learners, the Random Forest method can achieve high accuracy and generalization capability while simultaneously reducing the risk of overfitting. Beyond its predictive capabilities, Random Forest algorithm can provide valuable insights into the importance of various features through permutation importance. This information can assist educators and researchers in identifying key factors that contribute student success and in developing interventions based on validated data. The training data quality and representativeness can significantly influence the model performance and generalization capability. Furthermore, the interpretability of Random Forest results may be diminished compared to individual Decision Trees, as the model combines multiple trees predictions. Prior to its implementation, practitioners should carefully evaluate the fairness of both models and consider the ethical implications of their utilization in decision-making processes. Multiple researchers have conducted comparative assessments of Decision Trees and Random Forests in educational contexts. For instance, the performance of machine learning algorithms for predicting student dropout rates in Brazilian higher education were evaluated, discovering that Random Forest outperformed Decision Trees and other methods in terms of accuracy and robustness [30]. Decision Trees and Random Forest were also compared in predicting secondary vocational education student performance likewise concluding that Random Forest achieved superior results and provided more stable predictions [31].

B. *Naive Bayes Classifiers*

Naive Bayes Classifiers (NBCs) constitute a family of straightforward probabilistic classifiers based on the application of Bayes' theorem with pronounced "naive" assumptions regarding feature independence. The fundamental concept involves learning the joint probability distribution of input characteristics and output classes during the training, and afterwards calculating the posterior probability for each class based on input features for the new example. The underlying assumption posits that the characteristics are conditionally independent given the class, which simplifies the calculation of joint probability distribution and generally yields good performance in practice as it significantly streamlines the learning and inference processes since feature probabilities can be estimated independently for each class [32]. NBCs have demonstrated effectiveness in numerous real-world applications, particularly for text classification and spam filtering [33]. Their success in performing these tasks lies in the resilience to irrelevant feature identification, capacity to handle missing values, and efficiency in training and prediction processes. In the context of educational data, NBCs are usually applied for predicting student achievements, identifying learning styles, and analyzing feedback. For instance, NBCs were used for predicting academic success of university students based on their performance metrics [34]. The authors compared Naive Bayes with SVMs and found that it achieved superior prediction accuracy while requiring significantly less computational time. In another example, NBCs were used to detect learning styles through e-learning systems [35]. To identify learning styles (e.g., visual, auditory, kinesthetic), student preferences and behaviors were used as input features, which consequently allowed more personalized educational experience based on individual learning preferences. NBCs were successfully used for sentiment analysis of student feedback in online courses [36]. Researchers surveyed over 1000 students using the course evaluation questions and manually labeled responses as positive, negative, or neutral. NBCs achieved 90% accuracy in predicting sentiment in student comments and provided insights into the strengths and weaknesses of individual courses. These results facilitated improvements in instructional design and further development of early warning systems for student dissatisfaction. Despite the simplicity and efficiency, NBCs present certain limitations that warrant consideration when applying them in educational contexts. The naive independence assumption frequently does not hold in real-world scenarios, potentially leading to diminished performance. Additionally, NBCs are sensitive to the selection of prior probabilities and the presence of irrelevant or redundant features, making high-quality data processing and algorithm assumption validation essential for ensuring reliable and unbiased results. Although they can handle missing data by ignoring such instances or employing imputation techniques for missing values estimation, the selection of imputation methodology itself can significantly impact performance and interpretability. Furthermore, the interpretability is inherently limited, particularly for complex phenomena such as education [9]. Even though learned probabilities may provide insight into feature-class relationships, they may fail to recognize underlying causal mechanisms or domain knowledge. As research of NBCs educational implementation advances, these models are expected to play a significant role in data-driven educational decision-making processes.

C. *Support Vector Machines*

Support Vector Machines constitute a powerful class of supervised learning algorithms widely utilized for classification and regression due to their capacity to handle high-dimensional data, nonlinear relationships, and complex decision boundaries. The fundamental concept involves finding an optimal hyperplane that maximally separates different classes in a high-dimensional feature space, while minimizing the classification error and maximizing the geometric margin. By maximizing this distance, SVMs achieve strong generalization and reduce the risk of overfitting. The optimization process can be formulated as a quadratic programming problem that can efficiently be solved via sequential minimal optimization (SMO) or the Kernel-Adatron algorithm techniques [37]. SVMs offer several advantages that make them an attractive choice for educational data mining and learning analytics. First, the method demonstrates robustness when working with high-dimensional data and can efficiently manage large feature spaces [38]. Second, SVMs exhibit good generalization performance and are less prone to overtraining, especially when the number of characteristics exceeds the number of samples [39]. Third, it is possible to model complex nonlinear decision boundaries using kernel functions, making this model suitable for a wide spectrum of educational tasks. In educational contexts, SVMs are used for predicting student performance, detecting unproductive learning behaviors, and automating assessment tasks. It can also be employed for automated assessment of programming assignments by extracting features from the program code, such as syntax errors and/or complexity. Compared to teacher assessment grades, the model has demonstrated remarkably similar predictive accuracy [40]. It is worth emphasizing that SVMs can efficiently address both linear and nonlinear classification tasks.

Beyond improper hyperparameter selection, which significantly impacts performance, SVMs demonstrate sensitivity to noise and gaps in training data, potentially degrading generalization capability substantially [41]. Kernel functions may complicate the understanding of decision boundaries and the significance of individual characteristics. Feature selection techniques, model visualization, and rule extraction can enhance the interpretability of this method in educational context. The computation time for training and prediction increases quadratically with the number of examples, rendering this method impractical for real-time analysis of large datasets [42]. Stochastic gradient descent, parallel computing, and approximation methods can improve model scalability and efficiency in these situations. Model performance may also exhibit sensitivity to kernel function selection, regularization parameters, and class imbalances within the data. Model selection techniques and hyperparameter tuning methods are used in these instances (e.g., cross-validation and grid search) to achieve optimal performance outcomes.

D. *Logistic Regression*

Logistic regression is a statistical method for modeling the probability of a binary outcome (e.g., passed/failed, dropped out/did not drop out) as a function of input characteristics. It employs a sigmoid logistic function that maps a linear combination of input features to a probability within the interval between 0 and 1 [43]. Logistic regression is frequently applied in educational data mining and learning analytics to predict learning outcomes and identify students at risk of failure. For instance, logistic regression was used to predict student dropout rates in secondary schools and achieved high accuracy in identifying students who were at high risk of discontinuing their education [19]. Logistic regression can also be used for analyzing factors influencing student satisfaction with online learning environments [16]. Researchers employed a survey dataset encompassing student demographic characteristics, course features, and learning experiences as input factors. The results demonstrated that instructor/teacher support, course design, and perceived usefulness were significant predictors of student satisfaction, thus proving the validity of this approach.

IV. Regression algorithms

A. *Linear Regression*

Linear and logistic regression represent two fundamental statistical methods for modeling relationships between input characteristics and output variables. While the logistic regression is utilized for modeling binary or categorical results, linear regression is used for predicting continuous variables. Linear regression is a parametric method that assumes a linear relationship between input characteristics and a continuous output variable. In educational contexts, linear regression can be used for predicting student grades and to analyze factors influencing academic success [44]. Using multiple linear regressions can even assess dropout rates, but the prediction is sensitive to the quality of processed data, particularly when it contains missing values. Linear regression was utilized to identify key predictors of student success in their first year of studies using a combination of academic, demographic, and behavioral variables as input features [45]. It was concluded that class attendance, homework completion, and midterm examination grades were the most significant predictors of final grades, thereby demonstrating the linear regression as a valid technique for early identification of students at risk of failure and attrition. The primary limitation of linear regression is its assumption of a linear relationship between input features and the output variable, which often does not hold true in complex educational environments. Conversely, logistic regression assumes a linear relationship between input features and the probability of a binary outcome, thereby failing to identify nonlinear interactions. Both models demonstrate sensitivity to multicollinearity, outliers, and missing data, affecting model stability and interpretability [46]. To ensure reliable and valid results, data must be carefully processed, relevant features properly selected, and model assumptions validated. Mitigating the limitations and improving generalization of linear and logistic regression models can be accomplished through feature scaling, regularization, and cross-validation [47]. These techniques can help addressing the inherent constraints of beforementioned statistical methods while enhancing their applicability to educational data analysis.

B. *Polynomial Regression*

Polynomial regression is a statistical technique that extends linear regression by including higher-order terms of predictor variables, enabling researchers to model curved relationships between defined inputs and outcomes. As educational data exhibits non-linear patterns that cannot be adequately captured by simple linear models, polynomial regression offers a valuable methodological alternative, particularly in modeling longitudinal student learning trajectories. While linear models assume constant rates of learning, polynomial functions can capture acceleration or deceleration in knowledge acquisition using cubic polynomial functions to model achievement grades, revealing distinct phases of acceleration and eventually a plateau [48]. Determining appropriate polynomial degree represents a critical decision in educational application as educational data often contains nested structures and multiple covariates that must be considered during model specification. Research employing polynomial regression in educational contexts points toward its integration with multilevel modeling frameworks that enables researchers to account for classroom and school-level effects while modeling non-linear individual learning trajectories [49].

C. *Support Vector Regression*

Building upon the structural foundations of SVMs, SVR extends the approach to regression problems by estimating continuous values rather than categorical classifications [50]. SVR is particularly robust to outliers through its dual optimization approach and the incorporation of kernel functions that enable modeling of non-linear relationships without requiring explicit feature transformations. This technique has demonstrated effectiveness in predicting student academic performance across various educational levels finding that SVR outperformed other machine learning approaches when dealing with imbalanced grade distributions [51]. SVR models are specifically useful in online learning environments analysis because they can incorporate clickstream data, engagement metrics, and assignment completion patterns to predict course outcomes. The non-linear modeling capabilities of SVR proved especially beneficial in capturing the complex relationship between student proficiency and performance in adaptive testing and personalized assessment. The integration of SVR with multimodal learning analytics represents a promising model in which affective states are modeled during learning that demonstrated its capacity to integrate heterogeneous data sources while capturing complex non-linear relationships between physiological signals and learning states [52].

D. k-Nearest Neighbors

The *k*-Nearest Neighbors (*k*-NN) algorithm is a straightforward and effective non-parametric method for addressing classification and regression tasks. The fundamental concept involves predicting a class label or continuous value for a new sample based on majority voting or the average value of its k-nearest neighbors in the feature space. This instance-based learning algorithm requires no explicit training phase and can naturally process higher multi-class problems [53]. Selecting an appropriate *k* value is critical for algorithm performance. A small value may lead to overfitting and noise sensitivity, while a large value results in underfitting and loss of local information. In practice, the optimal value is frequently determined through cross-validation or domain knowledge application. The *k*-NN algorithm can process both numerical and categorical data, as well as missing values, by employing appropriate distance metrics and techniques. Furthermore, the algorithm can adapt to changes in training data without requiring retraining, making it suitable for dynamic educational environments [54]. The *k*-NN algorithm is often applied for predicting student success, recommending educational resources, and detecting plagiarism in written essays.

V. Conclusion

The AI integration and particularly supervised learning algorithms has transformed traditional educational paradigms by enabling data-driven decision-making, personalized learning experiences, and improved educational outcomes. Through the analysis of various supervised learning approaches several consistent themes and insights have emerged. First, supervised learning techniques demonstrated considerable potential for addressing critical educational challenges such as predicting student performance, identifying at-risk students, automating assessment processes, and personalized learning. The ability of presented algorithms to process large educational datasets and extract meaningful patterns provide educators and institutions with the powerful tools to enhance learning and teaching process. For instance, Decision Trees and Random Forest have shown particular efficacy in identifying key factors influencing academic achievement and predicting student dropout rates, while SVMs have proven valuable in automated assessment tasks. Supervised learning techniques also face significant implementation challenges in educational contexts due to the problems with availability and quality of labeled data. The interpretability of complex models is also challenging for educational stakeholders who require transparent decision-making processes. The “black box” nature of certain algorithms (e.g., SVMs) may undermine confidence in AI-based educational technologies and limit their adoption. As highlighted, supervised learning models can inadvertently perpetuate bias from training data, potentially leading to unfair or discriminatory outcomes for certain student groups. Regular audits, bias mitigation techniques, and inclusive data collection practices are essential for ensuring equity in the application of these technologies. Several promising directions for future research and development emerged. The integration of supervised learning with other AI techniques, such as unsupervised learning, reinforcement learning, and natural language processing, opens possibilities for more comprehensive and adaptive educational systems. XAI approaches specifically tailored for educational applications represent another vital research direction, as they can enhance transparency and build trust among stakeholders. The increasing availability of multimodal data in educational environments also presents opportunities for developing more sophisticated models that can leverage information from various data modalities to gain deeper insights into learning processes. Supervised learning techniques offer powerful tools for addressing contemporary educational challenges, but their effective implementation requires careful consideration.

Acknowledgment

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-136/2025-03/200132 with University of Kragujevac, Faculty of Technical Sciences Čačak.

References

[1] E. Glikson and A. W. Woolley, “Human Trust in Artificial Intelligence: Review of Empirical Research,” Academy of Management Annals, vol. 14, no. 2, pp. 627–660, Jul. 2020.

[2] F. Kamalov, D. Santandreu Calonge, and I. Gurrib, “New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution,” Sustainability, vol. 15, no. 16, p. 12451, Aug. 2023.

[3] L. Chen, P. Chen, and Z. Lin, “Artificial Intelligence in Education: A Review,” IEEE Access, vol. 8, pp. 75264–75278, 2020.

[4] E. Mousavinasab, N. Zarifsanaiey, S. R. Niakan Kalhori, M. Rakhshan, L. Keikha, and M. Ghazi Saeedi, “Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods,” Interactive Learning Environments, vol. 29, no. 1, pp. 142–163, Dec. 2018.

[5] H. Peng, S. Ma, and J. M. Spector, “Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment,” Smart Learning Environments, vol. 6, no. 1, pp. 1-14, Sep. 2019.

[6] D. Ginting, D. Sabudu, Y. Barella, A. Madkur, R. Woods, and M. K. Sari, “Student-centered learning in the digital age: in-class adaptive instruction and best practices,” International Journal of Evaluation and Research in Education (IJERE), vol. 13, no. 3, p. 2006, Jun. 2024.

[7] T. B. N. Da Silveira and H. S. Lopes, “Intelligence across humans and machines: a joint perspective,” Frontiers in Psychology, vol. 14, p. 1209761, Aug. 2023.

[8] T. Jiang, J. L. Gradus, and A. J. Rosellini, “Supervised Machine Learning: A Brief Primer,” Behavior Therapy, vol. 51, no. 5, pp. 675–687, Sep. 2020.

[9] C. Romero and S. Ventura, “Educational data mining and learning analytics: An updated survey,” WIREs Data Mining and Knowledge Discovery, vol. 10, no. 3, p. e1355, Jan. 2020.

[10] D. Ifenthaler, “Automated Essay Scoring Systems,” Handbook of Open, Distance and Digital Education, pp. 1–15, 2022.

[11] B. Perez, C. Castellanos, and D. Correal, “Applying Data Mining Techniques to Predict Student Dropout: A Case Study,” 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI), pp. 1–6, May 2018.

[12] S. Vieira, W. H. Lopez Pinaya, and A. Mechelli, “Introduction to machine learning,” Machine Learning, Academic Press, pp. 1–20, 2020.

[13] S. Nagel, “Supervised Learning,” Machine Learning in Asset Pricing, Princeton University Press, pp. 11–30, May 2021.

[14] U. Braga-Neto, “Physics-Informed Machine Learning,” Fundamentals of Pattern Recognition and Machine Learning, pp. 293–324, Cham: Springer International Publishing, 2024.

[15] G. James, D. Witten, T. Hastie, and R. Tibshirani, “Statistical Learning,” An Introduction to Statistical Learning, vol. 112, no. 1, pp. 15–57, New York: Springer, 2013.

[16] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, “Student Engagement Predictions in an e-Learning System and Their Impact on Student Course

Assessment Scores," Computational Intelligence and Neuroscience, vol. 2018, no. 1, pp. 1–21, Oct. 2018.

[17] P. Lagakis and S. Demetriadis, "Automated essay scoring: A review of the field," 2021 International Conference on Computer, Information and Telecommunication Systems (CITS), pp. 1–6, Nov. 2021.

[18] K. Zupanc and Z. Bosnić, "Automated essay evaluation with semantic analysis," Knowledge-Based Systems, vol. 120, pp. 118–132, Mar. 2017.

[19] C. Márquez‐Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. Mousa Fardoun, and S. Ventura, "Early dropout prediction using data mining: a case study with high school students," Expert Systems, vol. 33, no. 1, pp. 107‐124, Nov. 2015.

[20] R. S. Baker and A. Hawn, "Algorithmic Bias in Education," International Journal of Artificial Intelligence in Education, vol. 32, no. 4, pp. 1052–1092, Nov. 2021.

[21] E. Ntoutsi, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanikis, I. Kompatsiaris, K. Kinder-Kurlanda, C. Wagner, F. Karimi, M. Fernandez, H. Alani, B. Berendt, T. Kruegel, C. Heinze, K. Broelemann, G. Kasneci, T. Tiropanis, and S. Staab, "Bias in data‐driven artificial intelligence systems−An introductory survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 3, p. e1356, Feb. 2020.

[22] O.Zawacki-Richter, V.I.Marín, M.Bond, and F.Gouverneur, "Systematic review of research on artificial intelligence applications in higher education-where are the educators?," International Journal of Educational Technology in Higher Education, vol. 16, no. 1, pp. 1-27, Oct. 2019.

[23] R. M. Martins and C. Gresse Von Wangenheim, "Findings on Teaching Machine Learning in High School: A Ten - Year Systematic Literature Review," Informatics in Education, vol. 22, no. 3, pp. 421-440, Sep. 2022.

[24] R. Dwivedi, D. Dave, H. Naik, S. Singhal, O. Rana, P. Patel, B., Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core Ideas, Techniques, and Solutions," ACM Computing Surveys, vol. 55, no. 9, pp. 1–33, Jan. 2023.

[25] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, J. Bohg, "Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks," 2019 International Conference on Robotics and Automation (ICRA), pp. 8943–8950, May 2019.

[26] S. Chen, and X. Lin, "Application of Decision Tree Algorithm in Educational Data Mining," Curriculum and Teaching Methodology, vol. 6, no. 8, pp. 120-127, 2023.

[27] H. H. Patel and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," International Journal of Computer Sciences and Engineering, vol. 6, no. 10, pp. 74–78, Oct. 2018.

[28] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018, pp. 758–763, Dec. 2018.

[29] J. Han, J. Pei, and H. Tong, "Data, measurements, and data preprocessing," Data Mining, Elsevier Science & Technology, pp. 23–84, 2023.

[30] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. V. Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," Journal of Business Research, vol. 94, pp. 335–343, Jan. 2019.

[31] S. Sultana, S. Khan, and M. A. Abbas, "Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts," International Journal of Electrical Engineering & Education, vol. 54, no. 2, pp. 105–118, Jan. 2017.

[32] F.-J. Yang, "An Implementation of Naive Bayes Classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 301–306, Dec. 2018.

[33] T. M. Ma, K. Yamamori, and A. Thida, "A Comparative Approach to Naïve Bayes Classifier and Support Vector Machine for Email Spam Classification," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), pp. 324–326, Oct. 2020.

[34] A. Tripathi, S. Yadav, and R. Rajan, "Naive Bayes Classification Model for the Student Performance Prediction," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), pp. 1548–1553, Jul. 2019.

[35] S. T. Sianturi and U. L. Yuhana, "Student Behaviour Analysis To Detect Learning Styles Using Decision Tree, Naïve Bayes, And K-Nearest Neighbor Method In Moodle Learning Management System," IPTEK The Journal for Technology and Science, vol. 33, no. 2, p. 94, Aug. 2022.

[36] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms," Adaptive and Intelligent Systems, pp. 40–49, Cham: Springer International Publishing, 2014.

[37] S. Abe, "Fuzzy support vector machines for multilabel classification," Pattern Recognition, vol. 48, no. 6, pp. 2110–2117, Jun. 2015.

[38] A. A. Popov and A. S. Sautin, "Feature Spaces Combination in Face Recognition Task using Support Vector Machines," 2006 8th International Conference on Actual Problems of Electronic Instrument Engineering, pp. 289–289, Sep. 2006.

[39] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: a review," Artificial Intelligence Review, vol. 52, no. 2, pp. 857–900, Jan. 2018.

[40] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," Neurocomputing, vol. 408, pp. 189–215, Sep. 2020.

[41] J. Hays, A. Bevan, and T. J. Stevenson, "Support Vector Machines and generalisation in HEP," Proceedings of 38th International Conference on High Energy Physics — PoS(ICHEP2016), p. 976, Feb. 2017.

[42] D. Horn, A. Demircioğlu, B. Bischl, T. Glasmachers, and C. Weihs, "A comparative study on large scale kernelized support vector machines," Advances in Data Analysis and Classification, vol. 12, no. 4, pp. 867–883, Jul. 2016.

[43] D. G. Kleinbaum, M. Klein, "Introduction to Logistic Regression," Logistic Regression: a self-learning text, pp. 1–39, Springer Science+Business Media, 2010.

[44] A. Collins, "Predicting Student Success at a Large State University using Multiple Linear Regression and Hierarchical Clustering," Iowa State University, Aug. 2022.

[45] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," Computers & Education, vol. 103, pp. 1–15, Dec. 2016.

[46] R. Little and D. Rubin, "Statistical Analysis with Missing Data, Third Edition," Wiley Series in Probability and Statistics, Apr. 2019.

[47] M. Kuhn and K. Johnson, "Feature Engineering and Selection," Chapman and Hall/CRC, 2019.

[48] L. I. S. Giel, G. Noordzij, L. Noordegraaf-Eelens, and S. Denktaş, "Fear of failure: a polynomial regression analysis of the joint impact of the perceived learning environment and personal achievement goal orientation," Anxiety, Stress, & Coping, vol. 33, no. 2, pp. 123–139, Nov. 2019.

[49] A. Paletta, G. Alimehmeti, G. Mazzetti, and D. Guglielmi, "Educational leadership and innovative teaching practices: a polynomial regression and response surface analysis," International Journal of Educational Management, vol. 35, no. 4, pp. 897–908, Apr. 2021.

[50] F. Zhang and L. J. O'Donnell, "Support vector regression," Machine Learning, pp. 123–140, Academic Press, 2020.

[51] C. López-Martín, R. L. Ulloa-Cazarez, and A. García-Floriano, "Support vector regression for predicting the productivity of higher education graduate students from individually developed software projects," IET Software, vol. 11, no. 5, pp. 265–270, Oct. 2017.

[52] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova, "Supervised machine learning in multimodal learning analytics for estimating success in project‐based learning," Journal of Computer Assisted Learning, vol. 34, no. 4, pp. 366‐377, May 2018.

[53] O. Kramer, "K-Nearest Neighbors," Dimensionality Reduction with Unsupervised Nearest Neighbors, vol 51, pp.13-23, Berlin:Springer, 2013.

[54] N. Nuswantari, Y. F. Rachman, P. W. D. Setiawan, and W. D. Prakoso, "The Application of the K-Nearest Neighbors Method as A Recommendation for The Selection of Departments in Higher Education Based on The Results of Multiple Intelligence Tests," Journal of Physics: Conference Series, vol. 1464, no. 1, p. 012024, Feb. 2020.