Pak. j. life soc. Sci. (2024), 22(2):16382-16390 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2024-22.2.001188

RESEARCH ARTICLE

Prediction of Student Decision-Making Behaviour based on Machine Learning Algorithms

Hao Luo*

Faculty of Big Data, University of BaoShan, BaoShan, 678000, YunNan Province, China

ARTICLE INFO

Received: Sep 17, 2024

Accepted: Nov 22, 2024

Keywords

Machine Learning Algorithms

Decision-Making Behaviour

Behavioural Prediction

Model Optimisation

***Corresponding Author:**

bsxyluohao@163.com

ABSTRACT

Understanding how students make decisions is a key focus in educational research, with significant consequences for academic success, retention, career development, and overall student well-being. As educational institutions, policymakers, and educators increasingly turn to data-driven approaches, the ability to identify patterns that shape students' academic and behavioural choices has become more critical. Recent strides in machine learning (ML) have provided robust tools that can analyse vast datasets, helping to uncover meaningful insights into student decision-making processes. This study applies several machines learning algorithms, including Decision Trees, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and ensemble methods like Random Forests, to predict students' decision-making behaviours across different educational contexts. The main goal of this research is to create predictive models that can accurately forecast student decisions—such as course selections, dropout risks, and career choices—by evaluating key factors. These factors range from demographic data (e.g., age, gender, socio-economic status) to academic performance metrics (e.g., grades, attendance), and psychological traits (e.g., motivation, stress levels). Through machine learning techniques, this study examines which algorithm delivers the highest accuracy and reliability in predicting student behaviour, while also evaluating the significance of each factor in the modelling process. The findings demonstrate that machine learning algorithms, especially ensemble methods, outperform traditional statistical models in predicting student decision-making. These methods significantly enhance both accuracy and generalizability. For instance, Random Forest models excel at predicting complex decision-making processes, thanks to their capacity to manage high-dimensional data and capture nonlinear relationships (Romero & Ventura, 2010). Additionally, this research delves into the ethical concerns surrounding the use of machine learning in educational settings, with a focus on issues such as data privacy, fairness in algorithms, and the potential biases that may be present in predictive models (Witten, Frank, Hall, & Pal, 2016). Tackling these ethical concerns is essential to prevent machine learning applications from reinforcing existing disparities or stigmatizing vulnerable student groups. The insights from this study are valuable for educators and administrators seeking to personalize interventions, detect at-risk students early, and create decision-support systems that are tailored to student needs. With data-driven recommendations, educational institutions can better allocate resources, improve retention strategies, and support students in making informed academic and career decisions. Future research should concentrate on integrating more granular data, such as real-time behavioural data from learning management systems (LMS), to further enhance the predictive accuracy of machine learning models in educational environments.

1. INTRODUCTION

The use of machine learning (ML) algorithms to predict student decision-making behaviour has become a central focus in educational data mining (EDM) and learning analytics. As educational institutions gather vast amounts of data from sources like student interactions with learning management systems (LMS), assessments, and extracurricular activities, the need to derive meaningful insights from this data continues to grow (Romero & Ventura, 2010). Such insights can help educators, administrators, and policymakers make informed decisions to promote student success, enhance academic outcomes, and customize educational programs to accommodate the diverse needs of learners. By applying ML techniques to student data, researchers can forecast a variety of behaviours and outcomes, such as academic performance, course selection trends, graduation probabilities, and even career paths. These predictions enable institutions to proactively support students throughout their educational journey (Dutt et al., 2017).

One of the significant advantages of machine learning over traditional analytical methods lies in its ability to process large, complex, and multi-dimensional datasets. Conventional statistical approaches often require predefined assumptions and linear relationships among variables, whereas ML algorithms can discover nonlinear patterns, correlations, and hidden trends in the data without such assumptions (Baker & Inventado, 2014). This flexibility is particularly valuable in education, where student behavior is influenced by a wide range of factors, including motivation, socio-economic background, engagement with instructional materials, and external life events. By capturing these complex interactions, ML models can produce highly accurate predictions that guide decisions at both individual and institutional levels. For instance, a predictive model might indicate that a student is at risk of dropping out due to reduced course engagement, low attendance, and external factors such as financial instability. In response, educators could offer targeted interventions like academic counselling, financial aid, or customized learning plans to prevent the student from prematurely leaving the educational system (Zhang et al., 2019).

The application of various ML algorithms for predicting student behaviour has shown promising results across different educational contexts. Algorithms such as decision trees, random forests, support vector machines (SVMs), and neural networks are commonly used in this domain. Decision trees, for example, provide a hierarchical structure that splits data based on key attributes at each node, making the results easy to interpret for educators and administrators (Romero & Ventura, 2020). Random forests, which combine multiple decision trees, improve accuracy by reducing overfitting and managing missing data more effectively. SVMs are particularly effective in classification tasks, creating a decision boundary that distinguishes between categories of student behaviour, such as the likelihood of passing or failing a course (Shalev-Shwartz & Ben-David, 2015). Neural networks, especially deep learning models, have garnered attention for their ability to process unstructured data—like text and images—and identify complex patterns that may be missed by other methods. For example, neural networks can analyse student essays or discussion posts to predict engagement levels, comprehension, and future academic success.

Despite the potential of machine learning to transform the understanding of student decision-making, several challenges need to be addressed to ensure predictions are reliable, actionable, and ethically responsible. One significant challenge is the interpretability of models. Many ML models, especially deep learning models, function as "black boxes," making it difficult to understand how specific predictions are generated (Shalev-Shwartz & Ben-David, 2015). In educational contexts, this lack of transparency can be problematic, as educators and administrators often need clear explanations to justify their interventions. To address this, researchers are increasingly investigating interpretable ML techniques, such as explainable AI (XAI), which aims to provide more transparency by explaining model decisions in ways that humans can understand.

Another critical challenge concerns data quality and preparing datasets for ML analysis. Educational data is frequently noisy, incomplete, and subject to biases based on factors such as socio-economic status, ethnicity, and previous academic performance (Binns, 2018). If these biases aren't addressed during data preprocessing, the resulting predictive models could reinforce or amplify existing inequalities. For instance, a model trained on biased data might disproportionately predict that students from marginalized groups are more likely to fail, leading to unjust interventions. Therefore, it's essential to employ data-cleaning techniques, feature selection, and balancing methods to ensure models are trained on high-quality, representative datasets (Dutt et al., 2017). Additionally, using fairness-aware ML algorithms can help mitigate bias by adjusting for disparities in the data.

Ethical considerations also play a central role in predicting student behaviour with ML. Collecting and using student data for predictions must comply with legal and ethical standards, particularly regarding privacy, consent, and data protection (Zawacki-Richter et al., 2019). Students should be informed about how their data will be used, and institutions must ensure robust safeguards are in place to protect sensitive information. Furthermore, there is ongoing debate about the use of predictive analytics in education, particularly when high-stakes decisions are at play. Some critics argue that over-reliance on algorithmic predictions could create a deterministic view of students, where decisions are based primarily on data-driven insights, potentially neglecting the human element of education. Therefore, educational institutions must strike a balance between using machine learning and maintaining a holistic, human-centred approach to student support.

As machine learning continues to advance, its capacity to predict and shape student decision-making behaviour is expected to expand. Future research may focus on more sophisticated techniques, such as reinforcement learning, where algorithms learn to optimize student outcomes through ongoing feedback loops. Additionally, integrating natural language processing (NLP) into ML models could allow for the analysis of unstructured data, like essays, discussion posts, and social media activity, offering deeper insights into student engagement, motivation, and emotional well-being (Zhang et al., 2019). Moreover, adaptive learning systems powered by ML algorithms could provide real-time personalized instruction, delivering tailored content, assignments, and feedback based on individual learning patterns. By embracing these innovations, educational institutions can create more responsive, inclusive, and effective learning environments that better address the diverse needs of today's students.

2. RESEARCH OBJECTIVES

The first objective is to identify the key factors that influence student decision-making behaviour by analysing academic performance, demographic information, and online engagement metrics through machine learning algorithms. This goal aims to reveal the primary variables impacting student choices, such as course selection, dropout risks, or participation in academic activities. By examining data points like grades, attendance, and time spent on learning platforms, the study seeks to provide insights into how these elements shape student behaviour. Understanding these factors can assist educational institutions in tailoring interventions that better support their students (Romero & Ventura, 2010). Machine learning algorithms efficiently handle large datasets, helping to identify the most significant influences on student decision-making.

Additionally, the second objective is to develop and evaluate machine learning models—including decision trees, random forests, and neural networks—to predict student decision-making behaviour in both academic and extracurricular contexts. This objective involves building predictive models with high accuracy using various machine learning techniques. By training these models on datasets containing academic, behavioural, and demographic information, the research will assess which algorithm performs best for predicting decisions such as course enrolment, extracurricular participation, and academic achievement (Hastie et al., 2013). This objective is crucial for identifying the most effective machine learning methods for application in educational settings.

Furthermore, the third objective is to evaluate the effectiveness of different machine learning algorithms in predicting specific student decisions, such as course selection, engagement on learning platforms, and academic outcomes. This objective focuses on comparing the accuracy and performance of machine learning models across various types of student decisions. The research will examine algorithmic performance in predicting outcomes like course choice, graduation rates, and engagement with online platforms. Metrics such as accuracy, precision, and recall will be used to assess the models' effectiveness (Fawcett, 2006). The goal here is to determine the best algorithm for each decision type, offering actionable insights for educational institutions.

3. METHODOLOGY

The study adopts a quantitative research approach to predict student decision-making behaviour using machine learning algorithms. This approach enables the systematic collection and analysis of numerical data to reveal patterns in student decisions. The methodology is structured into four key phases: data collection, data preprocessing, model selection, and evaluation.

3.1 Data collection

The data for this study was collected from educational institutions, including student demographic details, academic performance records, and behavioural data from online learning platforms. Ethical guidelines were strictly adhered to, with all personal information anonymized. The dataset was sufficiently large to support robust predictive modelling, aligning with the recommendation that larger datasets generally enhance machine learning model performance (Han et al., 2012).

3.2 Data preprocessing

Before applying machine learning algorithms, the collected data was pre-processed. This involved addressing missing values, normalizing the data, and encoding categorical variables. Data imputation techniques, such as mean replacement, were used to handle missing values, while one-hot encoding was applied to manage categorical variables (Lin et al., 2022). Additionally, feature scaling was performed to standardize the range of independent variables, ensuring that the algorithms would not be biased by differences in data scale (Jain, 2010).

3.3 Model selection

Several machine learning algorithms were evaluated to predict student decision-making behaviour, including decision trees, random forests, support vector machines (SVM), and neural networks. Decision trees were selected for their simplicity and ease of interpretation, while random forests were included for their ability to reduce overfitting by combining multiple decision trees (Breiman, 2001). SVM was chosen due to its effectiveness in handling high-dimensional spaces (Vapnik, 1998), and neural networks were applied to explore non-linear relationships within the data (Hastie et al., 2013).

3.4 Model evaluation

The models were evaluated using cross-validation to ensure the generalizability of the results. Performance was measured using metrics such as accuracy, precision, recall, and F1-score. Confusion matrices were created for each model to display the true positives, true negatives, false positives, and false negatives. Additionally, ROC (Receiver Operating Characteristic) curves were plotted to analyse the trade-off between true positive and false positive rates (Fawcett, 2006). Hyperparameter tuning was performed using grid search to optimize the performance of each model (Bergstra & Bengio, 2012).

4. LITERATURE REVIEW

Besides that, he article "Advanced Machine Learning Approaches to Personalize Learning: Learning Analytics and Decision Making" (Kurilovas, 2018) explores the application of machine learning techniques to enhance the personalization of educational experiences through detailed learning analytics and decision-making models. It promotes the use of methodologies that evaluate the acceptance and effectiveness of personalized learning units by incorporating multiple criteria decision analysis and models such as the Unified Theory on Acceptance and Use of Technology (UTAUT). The central concept is to adapt learning environments in real-time to meet individual learning styles and needs, using analytics to fine-tune educational content and strategies based on continuous data collection on student interactions and performance. " (Kurilovas, 2018) research indicates that these personalized learning approaches lead to significant improvements in educational outcomes, highlighting their potential for broader application in refining teaching methods and educational technologies. This comprehensive framework seeks to dynamically align educational practices with individual learner profiles, offering valuable contributions to the field of educational technology.

Moreover, the article "Analysis of Learning Behaviour Characteristics and Prediction of Learning Effect for Improving College Students' Information Literacy Based on Machine Learning" by Shi et al. (2023) explores how machine learning algorithms can predict and improve college students' information literacy. The study analysed the learning behaviours of 320 Chinese college students, using the Pearson algorithm to identify correlations between information literacy behaviours and learning outcomes. Several machines learning models, including Decision Tree, KNN, Naive Bayes, Neural Networks, and Random Forest, were used to predict the effectiveness of information literacy learning. The Random Forest model demonstrated the highest accuracy, making it the best option for classifying and predicting learning outcomes. The paper highlights the importance of incorporating machine learning into educational strategies to enhance information literacy, offering valuable insights into how information literacy education can be tailored to improve teaching quality and foster innovative talent.

Furthermore, the article "Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms" by Yağcı (2022) discusses the use and effectiveness of various machine learning algorithms to predict academic performance in higher education. Drawing on a dataset from 1854 students enrolled in the Turkish Language-I course at a state university in Turkey during the fall semester of 2019-2020, the study evaluates the predictive power of midterm grades, along with departmental and faculty data, to forecast students' final exam scores. The algorithms tested include Random Forests, k-Nearest Neighbours, Naïve Bayes, Neural Networks, Support Vector Machines, and Logistic Regression, achieving classification accuracy between 70% and 75%. This research is noteworthy for demonstrating the practical application of educational data mining in real-world settings, providing valuable insights into how data-driven methods can enhance educational outcomes and teaching strategies. By comparing various machine learning models, the study not only identifies the most effective algorithms for predicting student performance but also contributes to the advancement of predictive analytics in education. This approach highlights the potential for using academic data proactively to identify students who may need additional support, allowing institutions to tailor interventions more effectively and improve overall student success. The findings are particularly relevant for educational administrators and policymakers aiming to adopt data-driven strategies for educational planning and resource allocation.

On the other hand, the article "Guaranteeing Correctness of Machine Learning Based Decision Making at Higher Educational Institutions" by (Nauman et al., 2021) discusses enhancing decision-making processes in higher education through supervised machine learning, specifically using decision trees and formal verification with Coloured Petri Nets (CP-Nets). It highlights the challenges of machine learning outcomes, such as bias and mislabelling, and introduces a method to ensure the correctness of decision rules derived from machine learning algorithms. By applying hierarchical Coloured Petri Nets, the authors demonstrate how formalism can verify and validate machine learning outcomes to improve the accuracy of decisions in educational settings. The study, supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University, uses a dataset from academic administrations to test the methodology, showing significant improvements in decision-making accuracy through the application of CP-Nets to model, analyse, and ensure the correctness of machine learning-generated decision rules.

In addition, the article "IntelliDaM: A Machine Learning-Based Framework for Enhancing the Performance of Decision-Making Processes. A Case Study for Educational Data Mining" by (Czibula et al., 2022) introduces IntelliDaM, a flexible framework designed to improve decision-making in educational environments using machine learning and data mining. The research applies this framework to analyse educational data from Babeș-Bolyai University, with a focus on predicting and improving student performance in a Computer Science course. IntelliDaM integrates feature analysis, as well as both unsupervised and supervised learning, to effectively identify patterns and predict student outcomes. The study shows that the framework can significantly enhance decision-making by adapting educational programs to better suit students' needs, underscoring its potential utility in various academic and practical contexts.

5. DISCUSSION

The prediction of student decision-making behaviour through machine learning algorithms has attracted growing attention in educational data mining, largely due to the rapid increase in student-related data and the need to leverage this information to improve educational outcomes. This approach utilizes the power of artificial intelligence to reveal patterns and insights within student behaviour, offering valuable predictions for stakeholders—including educators, policymakers, and institutions—who can use these insights to inform decisions on curriculum design, resource allocation, and personalized learning strategies. By predicting student decisions, educational institutions can enhance engagement while proactively addressing issues like dropout rates, poor academic performance, and course failures (Romero & Ventura, 2020).

Machine learning algorithms such as decision trees, support vector machines (SVM), random forests, k-nearest neighbours (KNN), and neural networks have been widely used to predict student behaviours. Each algorithm has distinct strengths and limitations, depending on the dataset's complexity and the specific behaviour being predicted. For instance, decision trees are valued for their interpretability and ability to visually represent decision-making pathways, which helps educators understand the importance of factors like prior academic performance, demographics, engagement metrics, and social behaviours in shaping a student's decisions (Kotsiantis, 2011).

A commonly used machine learning algorithm in predicting student decision-making is the random forest. As an ensemble learning technique, random forests generate multiple decision trees to create a more robust prediction by averaging the outcomes of individual trees. This not only boosts predictive accuracy but also helps mitigate overfitting—a problem where a model performs well on training data but poorly on unseen data. The ensemble method ensures that the model generalizes effectively across different student cohorts and educational contexts, making it ideal for predicting decisions such as course selection, graduation outcomes, or dropout risks (Benjamín et al., 2022).

Support vector machines (SVM), on the other hand, excel when working with high-dimensional datasets, where the number of features—such as socio-economic factors, attendance rates, extracurricular participation, and academic history—may exceed the number of students. SVM is particularly adept at creating a hyperplane that optimally separates students into categories, such as those likely to succeed versus those at risk of underperformance. By accurately classifying students, SVM models can help institutions provide timely interventions like tutoring, mentoring, or counselling to at-risk students (Alamri et al., 2020).

Neural networks and deep learning models, known for their exceptional performance across various predictive tasks, can also model the non-linear and complex relationships found in student decision-making. These models can uncover subtle patterns that simpler algorithms might miss, making them particularly useful for understanding more intricate aspects of student behaviour. However, a key challenge with neural networks is their "black box" nature, as the reasoning behind predictions is not always transparent or easily interpretable. This lack of clarity can be a barrier to adopting neural networks in educational settings where actionable insights are crucial (García et al., 2019).

The use of machine learning to predict student decision-making behavior has seen success in various educational domains. In course recommendation systems, for example, machine learning models analyse historical student data—such as past performance, course preferences, and social factors—to predict which courses a student is likely to excel in. By aligning course recommendations with students' strengths and interests, these systems improve student satisfaction and academic success (Sweeney et al., 2016). Similarly, dropout prediction is another area where machine learning has made significant strides. These models use features like attendance, grades, socio-economic background, and engagement levels to identify students at risk of dropping out, allowing institutions to intervene with measures such as financial aid, academic counselling, or personalized learning support (Baker & Inventado, 2014).

However, the success of machine learning in predicting student decisions hinges on the quality, diversity, and representativeness of the data. Biases in the data—whether from incomplete records, over-representation of certain demographics, or historical inequalities—can lead to biased predictions that perpetuate existing disparities. For example, if a model is trained on data that overrepresents affluent students, it may make less accurate predictions or recommendations for students from lower socio-economic backgrounds (Eckman et al., 2024). It is therefore essential that training datasets be diverse, inclusive, and representative of the entire student population.

Ethical considerations are also critical when applying machine learning to student decision-making. Predictive models should be used to empower students, not to limit their opportunities or reinforce negative stereotypes. For instance, predicting that a student is likely to drop out should not lower expectations or reduce resources for that student; instead, it should trigger additional support to keep the student on track. Furthermore, safeguarding data privacy is crucial, as student information must be handled securely and in compliance with legal frameworks like the General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) (Slade & Prinsloo, 2013).

In addition to technical and ethical challenges, the integration of machine learning models into educational decision-making requires collaboration between educators, data scientists, and policymakers. Working together, these stakeholders can design models that are both technically sound and aligned with educational goals. This interdisciplinary approach ensures that predictive models are interpreted correctly and used to make informed, beneficial decisions that foster equitable educational environments (Romero & Ventura, 2020).

In conclusion, machine learning algorithms offer a powerful tool for predicting student decision-making behaviour and improving educational outcomes. From personalized learning paths to early interventions for at-risk students, machine learning enables data-driven decisions that optimize the learning experience. However, ensuring the success of these models requires attention to data quality, interpretability, ethical considerations, and collaboration across stakeholders. Addressing these challenges allows educational institutions to fully leverage the potential of machine learning to support and improve student outcomes.

6. LIMITATIONS AND FUTURE WORK

While machine learning algorithms have shown great potential in predicting student decision-making behaviour, several limitations must be addressed to maximize their effectiveness in educational contexts. One of the main challenges involves data quality and availability. Many machine learning models rely on large and diverse datasets to generate accurate predictions, but educational institutions often struggle with incomplete, inconsistent, or biased data. For instance, historical data may over-represent certain student groups while under-representing others, resulting in models that make less accurate predictions for minority or underprivileged students (Eckman et al., 2024). This bias not only reduces the overall effectiveness of the model but can also perpetuate educational inequalities if not properly addressed.

Another limitation is the "black box" nature of some machine learning algorithms, particularly deep learning models. While these algorithms excel at identifying complex patterns and non-linear relationships, they often lack transparency, making it difficult for educators and decision-makers to understand how predictions are generated (García et al., 2019). The lack of interpretability poses a significant barrier to adoption, as stakeholders may hesitate to rely on models they cannot fully explain or understand. This issue is especially relevant in education, where decisions based on predictive models can have lasting and profound impacts on students' academic paths and futures.

Additionally, the ethical implications of using machine learning to predict student decision-making behaviour must be carefully considered. Predictive models run the risk of being used in ways that limit students' opportunities or reinforce negative stereotypes. For example, if a model predicts that a student is likely to drop out, it may lead to lowered expectations or reduced resources for that student, potentially worsening the problem (Slade & Prinsloo, 2013). Moreover, data privacy and security are critical concerns, as student data must be handled according to legal standards like the General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) to protect students' rights and ensure ethical data usage.

Looking ahead, several areas for future research could enhance the predictive power and ethical application of machine learning algorithms in education. One promising avenue is the development of more interpretable models. Research into explainable AI (XAI) is advancing rapidly, and applying these techniques to educational data mining could bridge the gap between accuracy and interpretability, enabling educators to better understand predictions and act on them with greater confidence (Alamri et al., 2020).

Another important area for future research is integrating more diverse and comprehensive data sources. Currently, many predictive models rely heavily on academic indicators such as grades and attendance, but future studies could explore incorporating non-academic factors like social-emotional skills, mental health, and extracurricular involvement. These more holistic datasets could improve the accuracy and fairness of predictions by providing a fuller picture of each student's unique circumstances and needs (Baker & Inventado, 2014).

Furthermore, addressing biases in machine learning models should be a key focus of future research. Techniques for bias detection and mitigation need to be refined to ensure that predictive models do not inadvertently disadvantage specific groups of students based on socio-economic status, ethnicity, or other factors (Eckman et al., 2024). Ensuring that models are fair and inclusive is essential for making sure that the benefits of machine learning in education are distributed equitably.

Finally, collaboration between data scientists, educators, and policymakers will be critical for the continued development and responsible implementation of predictive models in education. Future research should prioritize interdisciplinary partnerships to ensure that these models are designed and used in ways that align with educational objectives and ethical standards. By fostering collaboration, stakeholders can create predictive tools that not only enhance educational outcomes but also empower students and protect their rights.

7. CONCLUSION

In conclusion, applying machine learning algorithms to predict student decision-making behaviour is a powerful tool within educational data mining. These technologies have the potential to greatly enhance educational outcomes by offering insights that allow educators and institutions to create personalized learning experiences, anticipate challenges such as student dropouts, and implement timely interventions. Algorithms like decision trees, random forests, support vector machines, and neural networks have demonstrated their ability to make predictions based on various factors, including academic performance, demographic background, and student engagement (Romero & Ventura, 2020). However, the effectiveness of these models depends on the quality and diversity of the data used to train them, as well as the interpretability of their results. Issues such as bias and ethical concerns, including the risk of reinforcing inequalities and the need to protect student privacy, are critical challenges that must be addressed (Eckman et al., 2024).

Looking ahead, the future of machine learning in educational decision-making should prioritize the development of more interpretable models, the integration of holistic datasets that consider non-academic factors, and the creation of strategies to mitigate bias (Baker & Inventado, 2014). Additionally, collaboration between educators, data scientists, and policymakers will be crucial to ensuring that these technologies are applied responsibly and equitably, ultimately supporting the academic success and well-being of all students. As these predictive models continue to evolve, they hold the potential to revolutionize the educational landscape by providing educators with invaluable tools to better understand and support their students.

Funding: This work was supported by Yunnan Province Local Universities Joint Special Youth Project(202101BA070001-270) and Teaching Quality and Teaching Reform Project of Baoshan University in 2022-2023(ZHP202344)

REFERENCES

Alamri, L. H., Almuslim, R. S., Alotibi, M. S., Alkadi, D. K., Khan, I. U., & Aslam, N. (2020). Predicting Student Academic Performance using Support Vector Machine and Random Forest. https://doi.org/10.1145/3446590.3446607

Benjamín, M. Q., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., Apaza-Huanca, J. M., & Sanchez-Ilabaca, J. (2022). A Predictive Model Implemented in KNIME Based on Learning Analytics for Timely Decision Making in Virtual Learning Environments. International Journal of Information and Education Technology, 12(2), 91–99. https://doi.org/10.18178/ijiet.2022.12.2.1591

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. 13(1), 281–305.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Journal of Machine Learning Research. https://ora.ox.ac.uk/objects/uuid:2ff2785b-b0d4-447a-8326-a1fcc4c80840
Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324
Dutt, A., Ismail, M. A., & Herawan, T. (2017). A Systematic Review on Educational Data Mining. In IEEE Access (Vol. 5, pp. 15991–16005). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2017.2654247
Eckman, S., Plank, B., & Kreuter, F. (2024). Position: Insights from Survey Methodology can Improve Training Data.
Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques. Choice Reviews Online, 49(06), 49–3305. https://doi.org/10.5860/choice.49-3305
Hastie, T., Tibshirani, R. J., & Friedman, J. (2013). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. http://catalog.lib.kyushu-u.ac.jp/ja/recordID/1416361
Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
Kotsiantis, S. B. (2011). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artificial Intelligence Review, 37(4), 331–344. https://doi.org/10.1007/s10462-011-9234-x
Kurilovas, E. (2018). Advanced machine learning approaches to personalise learning: learning analytics and decision making. Behaviour and Information Technology, 38(4), 410–421. https://doi.org/10.1080/0144929x.2018.1539517
Lin, W. C., Tsai, C. F., & Zhong, J. R. (2022). Deep learning for missing value imputation of continuous data and the effect of data discretization. Knowledge-Based Systems, 239. https://doi.org/10.1016/j.knosys.2021.108079
Nauman, M., Akhtar, N., Alhudhaif, A., & Alothaim, A. (2021). Guaranteeing Correctness of Machine Learning Based Decision Making at Higher Educational Institutions. IEEE Access, 9, 92864–92880. https://doi.org/10.1109/access.2021.3088901
Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. In IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews (Vol. 40, Issue 6, pp. 601–618). https://doi.org/10.1109/TSMCC.2010.2053532
Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, 10(3). https://doi.org/10.1002/widm.1355
Shalev-Shwartz, S., & Ben-David, S. (2015). Understanding Machine Learning: From theory to Algorithms. https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf
Shi, Y., Sun, F., Zuo, H., & Peng, F. (2023). Analysis of Learning Behavior Characteristics and Prediction of Learning Effect for Improving College Students' Information Literacy Based on Machine Learning. IEEE Access, 11, 50447–50461. https://doi.org/10.1109/access.2023.3278370
Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. American Behavioral Scientist, 57(10), 1510–1529. https://doi.org/10.1177/0002764213479366
Sweeney, M., Lester, J., Rangwala, H., & Johri, A. (2016). Next-Term Student Performance Prediction: A Recommender Systems Approach. Zenodo (CERN European Organization for Nuclear Research). https://doi.org/10.5281/zenodo.3554604
Vapnik, V. N. (1998). Statistical learning theory. Wiley.
Witten, I. H., Frank, E., & Hall, M. A. (2016). Data Mining : Practical Machine Learning Tools and Techniques Ed. 4. In Elsevier eBooks. https://unr-ra.scholarvox.com/catalog/book/88835090?_locale=en
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
Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(1). https://doi.org/10.1186/s41239-019-0171-0