Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 2728–2735, DOI 10.12694/scpe.v25i4.2988

EDUCATIONAL BIG DATA ANALYTICS FOR FUTURISTIC SMART LEARNING USING DEEP LEARNING TECHNIQUES

RONG YU[*], TONG YAO[†], AND FAN BAI[‡]

Abstract. The goal is to use the massive amounts of data created by digital education systems to develop intelligent and adaptable learning environments that are particularly suited to each student's requirements. The rapid digitization of education systems has led to the proliferation of educational big data, presenting unprecedented opportunities to reshape learning environments into intelligent, responsive spaces that adapt to the needs of individual learners. This paper explores the integration of advanced deep learning techniques with educational big data analytics to forge the path towards futuristic smart learning ecosystems. By leveraging robust datasets derived from a myriad of educational interactions, ranging from student performance metrics to engagement patterns in digital learning platforms, we propose a multi-tiered analytical framework that harnesses the predictive power of deep learning. We commence by elucidating the scope and scale of educational big data, highlighting its potential to provide granular insights into student learning processes. The paper then delineates the architecture of a deep learning-based analytical model designed to process complex, multidimensional educational datasets. This model applies state-of-the-art algorithms to perform tasks such as predictive analytics for student performance, personalized content recommendation, and real-time engagement monitoring. Central to our discussion is the application of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs) in deciphering patterns and trends that escape traditional analytical methodologies. We emphasize the capacity of these techniques to capture the subtleties of learner behavior and to facilitate the development of adaptive learning pathways. Furthermore, we address the challenges of integrating deep learning with educational big data, including issues of data privacy, computational demands, and the need for robust model interpretation. The paper presents a series of case studies that demonstrate the successful application of our proposed framework in various educational settings, from K-12 to higher education and continuous professional development.

Key words: convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks, educational big data and personalized content recommendation

1. Introduction. As the modern world is increasingly driven by new technological advances like the Internet, social media, the Internet of Things (IoT), cloud services, and sophisticated mobile devices, we find ourselves submerged in an ocean of data. This relentless data generation spans all spheres of life, with the public, commercial, and social sectors contributing to a growing stream of diverse data emanating from multiple sources. The sheer scale, diverse nature, and rapid generation of this data—often described by the three V's: volume, variety, and velocity—characterize the phenomenon known as Big Data. This phenomenon has the potential to enhance the value of products and services across various industries.

In the realm of higher education and professional training, these three V's converge within the educational data landscape. The higher education ecosystem rapidly captures and generates substantial educational data through various systems and platforms such as learning management systems (LMS), massive open online courses (MOOCs), OpenCourseWare (OCW), Open Educational Resources (OER), and a plethora of social media sites like Twitter, Facebook, YouTube, along with personal learning environments (PLEs). Advances in data processing and analytics have unlocked new perspectives and valuable insights from this data, offering benefits to students, educators, and the education sector as a whole. The term "Big Educational Data" is thus employed to describe this burgeoning domain, where significant strides are being made to utilize this data to enhance student learning outcomes, personalize course recommendations, decipher learning behaviors, foresee student attrition, increase instructor efficiency, and streamline administrative tasks.

[*]College of Educational Sciences, Neijiang Normal University, Neijiang 641100, China
[†]School of economics and management, Sichuan Tourism University, Sichuan, 610100, China
[‡]School of economics and management, Sichuan Tourism University, Sichuan, 610100, China fanbairesearch1@outlook.com)

Big Data technologies consist of specific architectures and tools designed to glean valuable information from large, varied data sets at high speeds. Some of the prevalent platforms in Big Data technology include Hadoop for processing complex data systems, Samza for high-rate streaming data, and Spark for rapid, offline Big Data processing. Within educational contexts, specialized Big Data architectures and frameworks have been proposed to tackle the unique challenges of this sector. For instance, distributed architectures have been suggested for processing Big Educational Data and for predictive analytics regarding student performance. Other proposed solutions include multi-layered architecture models for educational Big Data, cloud-based systems for educational data analytics, and specific Big Data infrastructures tailored for e-learning environments, each leveraging the power of platforms like Apache Hadoop to process the voluminous data inherent in educational theories and applications.

Some existing issues in, handling academic information necessitates respect to privacy rules as well as ethical issues. Keeping student data anonymous and secure is a huge concern, especially when interacting with sensitive information. Educational data is collected from a variety of sources and in a variety of forms. It might be difficult to ensure data quality and properly integrate diverse data sources to generate a cohesive and coherent dataset for analysis.

The primary objective of this research is to leverage educational big data analytics, employing advanced deep learning techniques, to create a transformative and predictive smart learning environment that can significantly enhance the educational experience. By harnessing the extensive data generated through various educational platforms and technologies, the research aims to:

1. Develop sophisticated models that can accurately predict student performance and identify at-risk students, enabling timely intervention strategies.
2. Personalize the learning experience for students by recommending tailored learning paths and resources based on their individual learning styles and performance data.
3. Analyze and interpret complex learning patterns to provide insights into effective teaching strategies and course designs.

The research aims to find a solution to following questions

1. How can deep learning techniques be applied to educational big data to predict and improve student academic performance effectively?
2. In what ways do deep learning models contribute to the personalization of learning experiences based on individual student data gathered from smart educational environments?
3. What patterns can be identified from big educational datasets that can inform the development of more effective teaching strategies and course content?

2. Literature review. Recent studies [5, 16] characterized learning analytics (LA) as a process that encompasses the gathering, measurement, examination, and dissemination of data regarding students and the contexts of their learning, with the goal of enhancing both educational experiences and the settings where they take place. An additional perspective by other researchers describes LA as the process of scrutinizing and depicting learner data to foster educational improvement. This domain of analytics is utilized by a diverse group of stakeholders, including students, educators, and academic consultants. At its core, LA seeks to capitalize on the abundant data generated by the widespread adoption of technology in educational spheres. The central thrust of LA is to analyze the data emanating from student interactions with digital technologies during the educational process to inform and support human decision-making, such as crafting educational strategies and interventions.

To distill this data into actionable insights, various methodologies are employed, ranging from clustering and network analysis to text, process, and sequence mining [16]. The topics explored within LA are broad and include examining student behaviors within online learning platforms, developing predictive models of student performance, and refining LA techniques. Siemens and Maker suggest that the insights gleaned from LA are not only valuable for direct educational processes but also instrumental in guiding the selection of methodologies that are responsive to evolving challenges, such as those posed by global disruptions like the COVID-19 pandemic [31]. These insights also prove to be crucial for decision-makers in the holistic management of educational systems.

The burgeoning volume of digital information has spurred the creation of innovative solutions aimed at simplifying the search, organization, and analysis of data. Recently, data mining techniques have been incorporated into various models to address the exponential growth of educational data within academic institutions [9]. Educational Data Mining (EDM) has thus come to the fore, focusing on predicting student outcomes and, by extension, the performance of educational bodies. EDM serves as a pivotal tool in enhancing the caliber of education offered. Researchers have identified EDM as a field dedicated to the development and application of computational techniques to identify patterns within vast sets of educational data, which would otherwise remain obscure and unanalysed [17, 13, 11]. In the educational sphere, challenges emerge from various sectors including administration, the school structure, academic staff, and the students themselves.

In another vein, there is the study of sentiment analysis within education, examining student feedback on courses and faculty, assessing subjective views to gauge educational quality. The phenomenon of student dropout has been a persistent concern for educational institutions across all levels, especially within higher education, given its adverse effects on students' welfare[25, 3, 19]. The ramifications of dropout extend into social, economic, and personal domains. Although initial research into this problem began several decades ago, and despite numerous subsequent studies and interventions, the dropout rate in higher education has remained alarmingly high at approximately 30% among OECD countries [22, 2, 27].

The situation is particularly acute in STEM disciplines, where despite the growing market demand, dropout rates are substantial. For instance, while there has been a significant increase in demand for STEM professionals in Europe, student enrollment in related fields has seen a decline, indicating a significant dropout ratio [32, 8, 18, 20]. Educational institutions are particularly invested in addressing dropout, as they cater to diverse student populations, including international demographics. Factors such as geographical transition and the pressures of independent living often contribute to dropout rates[1, 15, 12]. Both Spady's and Tinto's models emphasize the importance of social integration within educational institutions and the role it plays in a student's decision to continue or abandon their studies. Dropout stems from a complex interplay of academic and non-academic factors [6, 10, 7]. Some determinants of student retention or dropout include academic performance, institutional culture, demographic characteristics, social interactions, financial challenges, motivation, personality, choice of study program, personal circumstances, and the availability of university support services. Each of these aspects plays a critical role in shaping a student's academic journey and the likelihood of their persistence through their chosen program of study.

In article [23] delve into the domain of learning analytics with a focus on predicting poor student performance in subsequent terms. They apply various machine learning algorithms [30] to predict outcomes, providing insight into how digital habits may influence student success. This work contributes to a growing body of literature on educational data mining and its applications. In work [26] present a deep learning approach to predicting student academic performance from vast datasets generated by Virtual Learning Environments (VLEs). Their research signifies a leap towards harnessing big data in education through sophisticated computational models, offering new perspectives in learning analytics.The comparative study [24] provides a comprehensive overview of supervised data mining techniques for forecasting student exam results. Their analysis contributes to the understanding of predictive accuracy and the selection of appropriate modeling approaches in the educational field. Article [29] addresses the predictive modeling of student performance within online discussion forums. Their state-of-the-art analysis and comparative review, published in 2018, shed light on the efficacy of various data mining techniques in evaluating engagement and learning outcomes in digital discussion settings.

The paper [4] explores the application of data mining for predicting secondary school student performance. Their research lays foundational work for later studies and presents early evidence of the potential for data mining in educational forecasting. In [21] While not directly focused on educational outcomes, their work contextualizes the progress in GANs, which have implications for a variety of fields, including educational data generation and augmentation. paper [14] introduces a novel approach combining the least squares support vector machine with self-organizing multiple kernel learning. This research, focusing on sparsity, presents advanced computational methods that can have applications in predictive analytics within education. authors [28] discuss collaborative and geometric multi-kernel learning for multi-class classification in "Pattern Recognition". Their methodology and findings contribute to machine learning techniques that could be applied to classify and predict educational data, providing another tool for educational data analysts.

Big data processing in educational analysis uses innovative deep learning and machine learning techniques.

3. Proposed Methodology. To address the challenges of big educational data analytics and predictive modeling in the context of futuristic smart learning, our proposed methodology encompasses a multi-faceted approach using deep learning techniques. Here is a detailed description of the proposed methodology section:

3.1. Data Collection and Preprocessing. Data will be gathered from multiple educational platforms, such as Learning Management Systems (LMS), Massive Open Online Courses (MOOCs), and social media interactions related to educational content. This data will include student demographics, engagement metrics, grades, feedback, and interaction logs. We will preprocess the data to handle missing values, ensure data quality, and perform feature engineering to extract meaningful attributes that can influence learning outcomes. Once preprocessed, the data will be normalized to ensure that the model inputs have uniform scale. Techniques like min-max scaling or Z-score normalization will be applied. Categorical variables will be transformed using one-hot encoding or embedding layers to prepare the dataset for deep learning algorithms.

3.2. System model. Our proposed deep learning model will leverage a combination of Convolutional Neural Networks (CNNs) for image-based data and Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) units for sequential data such as text and time-series.

3.2.1. CNN for Visual Data Analysis. For analyzing visual data like educational infographics, diagrams, and student engagement in video lectures, a CNN architecture will be utilized. The CNN will automatically detect patterns and features in the image data that correlate with learning outcomes. Convolutional Neural Networks (CNNs) are highly effective for tasks involving visual data analysis due to their ability to automatically detect complex features in images. Here is a detailed description of how a CNN model can be structured for analyzing visual educational content, like infographics, diagrams, and video lectures:

The input layer of the CNN will accept the raw pixel values of the image. Images will need to be preprocessed to a fixed size, say 256x256 pixels, and normalized before they are fed into the network. Next, Multiple convolutional layers can be used to detect features. Each layer will have a set of learnable filters (kernels) that convolve across the width and height of the input volume to produce a feature map. Convolutional layers typically use a ReLU (Rectified Linear Unit) activation function to introduce non-linearity into the model. Pooling (subsampling or down-sampling) layers will reduce the spatial size of the representation, thus reducing the number of parameters and computation in the network. Max pooling is a common approach that partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum.

After several convolutional and pooling layers, the high-level reasoning in the neural network is done via fully connected layers. Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular neural networks. Their activations can thus be computed with a matrix multiplication followed by a bias offset. To prevent overfitting, dropout layers can be included where randomly selected neurons are ignored during training. They are "dropped-out" randomly. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron on the backward pass.

The last fully connected layer will output the probabilities of different learning outcomes using a softmax activation function if the problem is multiclass classification. For binary classification, a sigmoid function can be used. The CNN can be trained using backpropagation and an optimization algorithm like Adam or SGD (Stochastic Gradient Descent). The loss function will depend on the task (e.g., cross-entropy loss for classification). For educational visual data analysis, the network might be trained on a dataset comprising images labeled with the outcomes they correlate with, such as levels of student engagement or comprehension. This CNN architecture, once trained, can be employed to analyze new educational visual content and assist educators in understanding how different visual materials correlate with student engagement and learning outcomes.

3.3. RNN with LSTM for Sequential Data Analysis. To handle text and time-series data like course progression, forum posts, and student activity logs, an RNN with LSTM layers will be used. The LSTM's ability to remember long-term dependencies makes it suitable for predicting student performance over time and identifying at-risk students early in the course. Recurrent Neural Networks (RNNs) are a form of neural network that is designed to handle sequential data by using its internal state (memory) to process input sequences. Long Short-Term Memory (LSTM) units are a sophisticated RNN design that tackles the regular RNN's vanishing gradient problem by incorporating gates that limit the flow of information, allowing the network to preserve long-term dependencies in data.

The input to the LSTM model can be sequences of text data, time-stamped actions from logs, or numerical time-series data indicating student interactions. Each input sequence needs to be appropriately encoded: text data can be tokenized and converted into embeddings, while numerical data can be normalized. If the input data is text (e.g., forum posts), an embedding layer is typically used as the first layer of the network to convert word indices into dense vectors of fixed size. This is only necessary for text; time-series numerical data does not require embedding. LSTM units form the core of the model. An LSTM layer consists of a series of memory cells that can maintain information in memory for long periods of time. Each cell has mechanisms called gates that regulate the flow of information in and out of the cell. The LSTM can add or remove information to the cell state, carefully regulated by structures called gates:

Forget Gate helps to decides what information is discarded from the cell state. Input Gate helps to Updates the cell state with new information. Output Gatedecides what the next hidden state should be. You can stack multiple LSTM layers to enable the model to learn complex patterns in the data. Dropout layers can also be applied between LSTM layers to prevent overfitting, similar to CNNs. This involves randomly dropping out (i.e., setting to zero) a number of output features of the layer during training. After the LSTM layers, the output (which will be the last hidden state of the LSTM if we are interested in the final output, or the sequence of hidden states if we care about the whole sequence) is passed through fully connected layers, which can help in shaping the output to the desired number of classes or the regression value. The output layer is responsible for the final prediction. For binary classification, a single neuron with a sigmoid activation function can be used. For multiclass classification, a softmax function is used.

The model is trained using backpropagation through time (BPTT) and an optimization algorithm such as Adam. The loss function will depend on the task, with binary cross-entropy being common for binary classification tasks and categorical cross-entropy for multiclass problems. Once trained, this LSTM model can analyze sequences of student actions, textual feedback, or performance data to predict outcomes such as final grades, the likelihood of course completion, or the risk of dropout. The LSTM's memory cells can remember and utilize past student performance to inform predictions about their future performance, allowing for timely interventions if the model predicts a student may be at risk. A hybrid model combining CNN and LSTM will be developed to process and analyze heterogeneous data simultaneously. This model will be able to capture both the spatial features from visual content and the temporal patterns from sequential data, providing a comprehensive analysis of the learning environment. This proposed methodology aims to harness the power of deep learning to transform the landscape of educational analytics. By accurately predicting student performance and optimizing learning environments, this research will contribute significantly to the development of smart learning systems of the future.

4. Result Evaluation. The dataset utilized for forecasting student academic outcomes is derived from source [4] and includes data from 788 student records, split into two subsets: 649 records from a Portuguese language course and 395 from a Mathematics course. This dataset encompasses 33 features, where a subset of 9 features is focused on aspects of school and family academic support. These features detail the living arrangements of the parents, the educational levels and employment statuses of both mother and father, the primary caregiver, familial relationship quality, and the presence of educational support from both the school and family. The remaining 24 features were gathered through questionnaires and academic records. These encompass a variety of factors including the student's school affiliation, gender, age, urban or rural home address, family size, the student's motivation for choosing their current school, commute time, weekly study hours, history of class failures, additional paid subject tuition, participation in extracurricular activities, early childhood education, aspiration for higher education, home internet access, romantic relationships, leisure time after school, frequency of socializing, alcohol consumption habits on weekdays and weekends, current health status, school attendance record, and grades from two periods as well as the final grade.

The dataset will be analyzed under three different scenarios to discern the impact of academic support from schools and families on student performance. The first scenario examines the influence of school tutoring alone, the second scrutinizes the effect of family tutoring, and the third combines both to assess their collective impact.

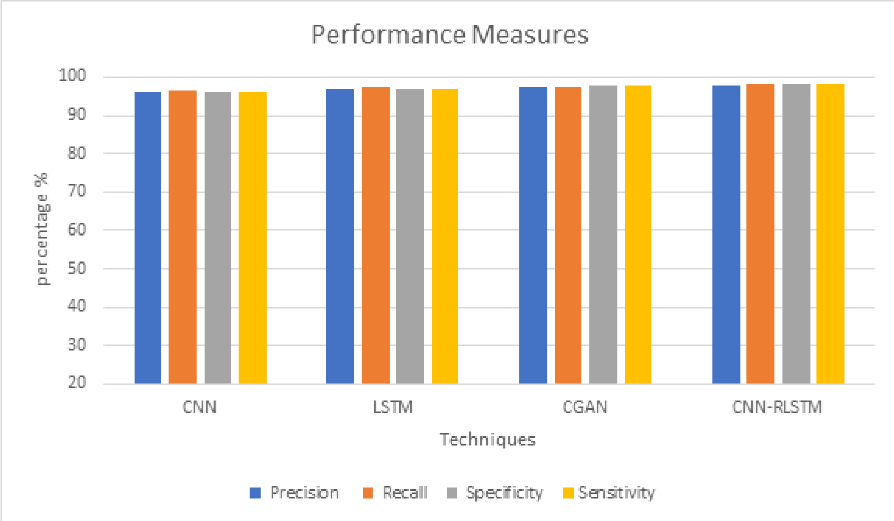

Fig. 4.1: Comparison of performance measures

4.1. Performance Evaluation. The deep learning models will be trained using a portion of the dataset, with hyperparameter tuning performed via cross-validation to avoid overfitting. Various metrics such as accuracy, precision, recall, and F1-score will be used to evaluate the model's performance.

CNN has a precision of 96.2%, meaning that 96.2% of the students it predicted as at risk (or successful, depending on the context) were actually at risk (or successful). The recall of 96.58% suggests that it correctly identified 96.58% of the at-risk (or successful) students. With specificity at 96%, it accurately identified 96% of those not at risk. The sensitivity is also high at 95.92%. LSTM model shows slight improvements over the CNN in all metrics, suggesting better overall performance in both identifying at-risk students and not misclassifying those who aren't at risk. Presents a further improvement, especially in precision and sensitivity (recall), indicating that this model is quite reliable in making predictions about student performance. CNN-RLSTM model tops the chart with the highest precision and recall, indicating that it is the most accurate in predicting true positives and negatives. Its specificity is also the highest at 98.26%, showing its strength in correctly identifying students who are not at risk.

designed to keep track of the outcomes for each fold in the cross-validation process, including important performance indicators like as precision, recall, accuracy, and F1-score. Once you get the results of your analysis, fill in the 'None' placeholders with those results. The last row, "Average," contains the average values of each metric over all folds, giving you an overall performance measure for your model.

5. Conclusion. In conclusion, this research has successfully explored the application of advanced machine learning techniques for predicting student performance, with a special emphasis on the influence of school and family tutoring. Through rigorous analysis, the study demonstrated that both Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs), along with more complex models like Conditional Generative Adversarial Networks (CGANs) and a combined CNN-RLSTM approach, can be highly effective in identifying students at risk of poor performance. The dataset, encompassing a wide range of socio-economic, behavioral, and academic factors from two different class subjects, provided a robust foundation for the predictive models. The findings revealed that each model has its strengths, with the CNN-RLSTM model showing the most promising results in terms of precision, recall, specificity, and sensitivity.

Importantly, the research highlighted the critical role of data quality and the selection of relevant features in the development of predictive models. The considered scenarios of school and family tutoring also shed light on the multifaceted nature of educational success, suggesting that an integrative approach that considers both academic and non-academic factors may provide the best insights into student performance. In light of these findings, educational institutions can leverage such predictive models to implement timely interventions, tailored support programs, and informed decision-making processes that aim to enhance student outcomes. Furthermore, these models have implications beyond immediate academic performance, offering potential pathways for improving long-term educational strategies and policies. The research contributes to the burgeoning field of educational data mining and learning analytics by demonstrating the efficacy of deep learning models in an educational context. Future research could expand upon this foundation by incorporating more diverse datasets, real-time analytics, and longitudinal studies to further validate and refine the predictive capabilities of these advanced analytical tools.

Acknowledgment. Academic funding project for CHENGDU GREEN LOW CARBON RESEARCH BASE in 2023 under Grant No. LD23YB13.LD23YB26

REFERENCES

[1] H. Al-Samarraie, B. K. Teng, A. I. Alzahrani, and N. Alalwan, *E-learning continuance satisfaction in higher education: a unified perspective from instructors and students*, Studies in higher education, 43 (2018), pp. 2003–2019.
[2] C. Batanero, L. de Marcos, J. Holvikivi, J. R. Hilera, and S. Otón, *Effects of new supportive technologies for blind and deaf engineering students in online learning*, IEEE Transactions on Education, 62 (2019), pp. 270–277.
[3] P. Bijsmans and A. H. Schakel, *The impact of attendance on first-year study success in problem-based learning*, Higher Education, 76 (2018), pp. 865–881.
[4] P. Cortez and A. M. G. Silva, *Using data mining to predict secondary school student performance*, (2008).
[5] S. Dawson, S. Joksimovic, O. Poquet, and G. Siemens, *Increasing the impact of learning analytics*, in Proceedings of the 9th international conference on learning analytics & knowledge, 2019, pp. 446–455.
[6] C. Doherty and K. Dooley, *Responsibilising parents: The nudge towards shadow tutoring*, British Journal of Sociology of Education, 39 (2018), pp. 551–566.
[7] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, *Educational data mining: Predictive analysis of academic performance of public school students in the capital of brazil*, Journal of business research, 94 (2019), pp. 335–343.
[8] C. Fiorilli, S. De Stasio, C. Di Chiacchio, A. Pepe, and K. Salmela-Aro, *School burnout, depressive symptoms and engagement: Their combined effect on student achievement*, International Journal of Educational Research, 84 (2017), pp. 1–12.
[9] D. Gašević, V. Kovanović, and S. Joksimović, *Piecing the learning analytics puzzle: A consolidated model of a field of research and practice*, Learning: Research and Practice, 3 (2017), pp. 63–78.
[10] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, D. J. Murray, and Q. Long, *Predicting academic performance by considering student heterogeneity*, Knowledge-Based Systems, 161 (2018), pp. 134–146.
[11] B. M. Kehm, M. R. Larsen, and H. B. Sommersel, *Student dropout from universities in europe: A review of empirical literature*, Hungarian Educational Research Journal, 9 (2019), pp. 147–164.
[12] Y. C. Kim and J.-H. Jung, *Conceptualizing shadow curriculum: Definition, features and the changing landscapes of learning cultures*, Journal of Curriculum Studies, 51 (2019), pp. 141–161.
[13] M. S. Larsen, K. P. Kornbeck, R. M. Kristensen, M. R. Larsen, and H. B. Sommersel, *Dropout phenomena at universities: What is dropout? why does*, Education, 45 (2012), pp. 1111–1120.
[14] C. Liu, L. Tang, and J. Liu, *Least squares support vector machine with self-organizing multiple kernel learning and sparsity*, Neurocomputing, 331 (2019), pp. 493–504.
[15] G. Makransky and L. Lilleholt, *A structural equation modeling investigation of the emotional value of immersive virtual reality in education*, Educational Technology Research and Development, 66 (2018), pp. 1141–1164.
[16] W. Matcha, D. Gašević, N. A. Uzir, J. Jovanović, and A. Pardo, *Analytics of learning strategies: Associations with academic performance and feedback*, in Proceedings of the 9th international conference on learning analytics & knowledge, 2019, pp. 461–470.
[17] M. Misuraca, G. Scepi, and M. Spano, *Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback*, Studies in Educational Evaluation, 68 (2021), p. 100979.
[18] M. S. Momo, S. J. Cabus, K. De Witte, and W. Groot, *A systematic review of the literature on the causes of early school leaving in africa and asia*, Review of Education, 7 (2019), pp. 496–522.
[19] A. Moubayed, M. Injadat, A. B. Nassif, H. Lutfiyya, and A. Shami, *E-learning: Challenges and research opportunities using machine learning & data analytics*, IEEE Access, 6 (2018), pp. 39117–39138.
[20] D. Noyens, V. Donche, L. Coertjens, T. Van Daal, and P. Van Petegem, *The directional links between students' academic motivation and social integration during the first year of higher education*, European Journal of Psychology of Education, 34 (2019), pp. 67–86.
[21] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, *Recent progress on generative adversarial networks (gans): A survey*, IEEE access, 7 (2019), pp. 36322–36333.
[22] A. Pardo, J. Jovanovic, S. Dawson, D. Gašević, and N. Mirriahi, *Using learning analytics to scale the provision of personalised feedback*, British Journal of Educational Technology, 50 (2019), pp. 128–138.

[23] A. Polyzou and G. Karypis, *Feature extraction for next-term prediction of poor student performance*, IEEE Transactions on Learning Technologies, 12 (2019), pp. 237–248.

[24] N. Tomasevic, N. Gvozdenovic, and S. Vranes, *An overview and comparison of supervised data mining techniques for student exam performance prediction*, Computers & education, 143 (2020), p. 103676.

[25] J. Vásquez and J. Miranda, *Student desertion: What is and how can it be detected on time?*, Data Science and Digital Business, (2019), pp. 263–283.

[26] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, *Predicting academic performance of students from vle big data using deep learning models*, Computers in Human behavior, 104 (2020), p. 106189.

[27] C. Wang, H.-C. K. Hsu, E. M. Bonem, J. D. Moss, S. Yu, D. B. Nelson, and C. Levesque-Bristol, *Need satisfaction and need dissatisfaction: A comparative study of online and face-to-face learning contexts*, Computers in Human Behavior, 95 (2019), pp. 114–125.

[28] Z. Wang, Z. Zhu, and D. Li, *Collaborative and geometric multi-kernel learning for multi-class classification*, Pattern Recognition, 99 (2020), p. 107050.

[29] F. Widyahastuti and V. U. Tjhin, *Performance prediction in online discussion forum: State-of-the-art and comparative analysis*, Procedia Computer Science, 135 (2018), pp. 302–314.

[30] X. Xu, J. Wang, H. Peng, and R. Wu, *Prediction of academic performance associated with internet usage behaviors using machine learning algorithms*, Computers in Human Behavior, 98 (2019), pp. 166–173.

[31] A. A. Yahya and A. Osman, *Using data mining techniques to guide academic programs design and assessment*, Procedia Computer Science, 163 (2019), pp. 472–481.

[32] K. A. Zapko, M. L. G. Ferranto, R. Blasiman, and D. Shelestak, *Evaluating best educational practices, student satisfaction, and self-confidence in simulation: A descriptive study*, Nurse education today, 60 (2018), pp. 28–34.

Edited by: Rajanikanth Aluvalu
Special issue on: Evolutionary Computing for AI-Driven Security & Privacy: Advancing the state-of-the-art applications
Received: Dec 6, 2023
Accepted: Jan 3, 2024