J. Electrical Systems 20-4s (2024): 984-993

[1]Ibnu Nur Akhsan

[2]Basuki Wibawa

[3]Robinson Situmorang

[4]Eveline Siregar

Fulfilling K-12 Students' Mathematics Needs: Near Peer Teaching and Scientific Intelligent Clustering Tutor Approach

Abstract: - This investigation delves into the pressing need for curriculum reform in Indonesia, specifically examining the "Merdeka Belajar: Kampus Merdeka (MBKM)" policy, with a focus on the Near Peer Teaching (NPT) model. Previous studies have flagged NPT's shortcomings, attributing them to inconsistent tutor feedback rooted in relational challenges. Despite positive anecdotal evidence of student transformation through NPT, such accounts often lack objectivity. This research strategically surveys K-12 vocational schools in Kuningan, honing in on challenges in Mathematics to inform responsive teaching strategies. Noteworthy is the persistence of selecting peer tutors based on final exam scores, a practice upheld despite the initial randomness in NPT tutor selection, creating hurdles in gauging effectiveness. Paradoxically, empirical data suggests that third-semester students make better tutors, yet the fixation on final exam scores persists. To propel the NPT model forward, the study advocates for clustering tutors based on scientific intelligence, integrating the innovative application of machine learning algorithm K-Means. This comprehensive approach melds quantitative data science with qualitative Deep Interviews, aiming to refine and optimize the identification of suitable peer tutors. The crux of the findings revolves around the imperative to refine the selection process for peer tutors, considering factors such as interest, motivation, and academic achievement, to significantly amplify the efficacy of NPT within the learning environment..

Keywords: Machine Learning, Near Peer Teaching, Aledu, K-12.

I. INTRODUCTION

Recent educational reforms around the world have been actively promoting science, technology, engineering, and mathematics (STEM) in vocational schools [1]. This is reflected in the analysis of 675 articles and proceedings from five editions spanning 2012-2021, showing that K-12 vocational schools/SMKs have a strong engagement with a high percentage in the fields of teaching, teachers, learning, and vocational education [2]. This certainly indicates an interest in teaching and learning in K-12 vocational education [3], [4]. In line with this, in Indonesia, the “Kurikulum Merdeka” has been established as the current national curriculum policy, focusing on student needs [5].

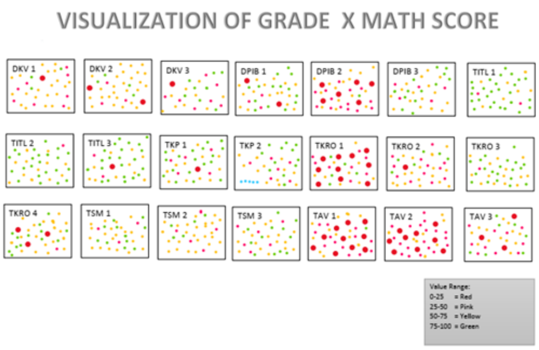

Figure 1. Visualization of Clustering of X Grade Students' Mathematics Final Exam Scores with NPT Learning

In a field study case related to student needs, a problem was found with the average final exam scores in

[1] State University of Jakarta,

[2]State University of Jakarta,

[3]State University of Jakarta,

[4]State University of Jakarta

[1] ibnu.nur.akhsan@mhs.unj.ac.id,[2]bwibawa@unj.ac.id, [3]rsitumorang@unj.ac.id,[3]esiregar@unj.ac.id

mathematics at a K-12 vocational school in Kuningan. After visualization (as seen in Figure 1), a significant disparity in scores was noted. After surveying students and interviewing teachers, students felt difficulties and experienced “jumping logic” in learning mathematics with the Near Peer Teaching Model. In response to student feedback, teachers wanted to provide appropriate training designs based on tutor ability clustering. Generally, teachers only use final exam scores as the main parameter, which ultimately poses challenges in measuring the effectiveness of Near Peer Teaching due to various biases and difficulties in accurate measurement. Paradoxically, teachers agree that academic factors are not the only main parameter in selecting peer tutors.

The initial problem in another NPT study case was claimed to be less effective due to the feedback given by tutors being partly correct and partly incorrect [6]. This could be caused by relational challenges in communication and interpersonal aspects [7]. However, there is abundant literature on the dynamics of peer tutor-student relationships, peer tutor evaluation performance, and student satisfaction with NPT [8], [9]. The core issue is that although many publications report positive changes in students using the NPT model, most are based on subjective opinions. Few studies investigate the effectiveness of NPT with measurable, objective, and prospective techniques [10]. Therefore, it is important to cluster tutors to obtain measurable learning model effectiveness and prospective tutors based on variable considerations such as motivation, interest, achievement, learning outcomes, and educational track record [11].

With the complex characteristics of potential NPT tutors, it is very possible to conduct an in-depth analysis using artificial intelligence machine learning technology to discover patterns of analysis for accurate information. The current artificial intelligence algorithms used to predict the prospects of potential tutors and classify tutors include the Naïve Bayes algorithm, k-Nearest Neighbors, Decision Tree-based algorithms [12], Random Forest, Support Vector Machines (SVM), Neural Networks [13], RA, J48, JPip, AdaBoost, BN Yahya Logistic Regression, Linear Regression, etc [14].

II. LITERATURE REVIEW

Near Peer Teaching (NPT) is an educational approach where students actively serve as peer tutors/instructors, typically chosen from the same learning environment or one academic year ahead [15], [16]. NPT offers a more individualized learning approach, where peer tutors can bridge the knowledge gaps that teacher-centered learning fails to address [17]. In developing NPT to meet student learning needs, complex and interrelated variables are necessary for the clustering process. The accuracy of tutor clustering can be achieved by leveraging technological advancements.

Technological progress benefits society [18]. The paradigm shift in the use of digital technology in the business world has a significant impact on the field of education, which plays a crucial role. The expanding trend of using Machine Learning (ML) is now a novel social panorama, not only for its ability to collect social interaction data but also for its evolved role in analyzing and extracting information to offer insights into the emotional and psychological aspects of learners [19].

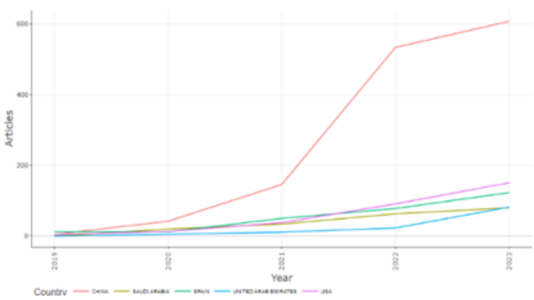

Figure 2. Countries Production Overtime

Machine Learning (ML) is a part of Artificial Intelligence (AI) that simulates human intelligence in machines or systems. An analysis of 613 articles from 2013-2023 shows that the implementation of AI in education has proven beneficial. Countries like China, America, Spain, Saudi Arabia, and the United Arab Emirates dominate the contribution to publications related to the use of AI in development, utilization, management, assessment, and design in the education sector (see Figure 2). Machine learning ranks third in AI education research topics (see Figure 3).

Figure 3. Word Cloud

Based on the average number of publications, citation counts, and publication year, the topic of machine learning scores the highest (see Figure 4). There are 43 recent articles with 53 distributions across different multidisciplinary topics and a significant number of citations. The analysis of classification and clustering data dominates the topic of machine learning. This mapping has been aligned with most researchers' practices by considering the number of occurrence compromise, choosing 3 to select the best 111 keywords out of 980 to create a better "helicopter view" visualization on Vosviewer. This visualization has undergone a "Thesaurus" identification to find different but synonymous keywords.

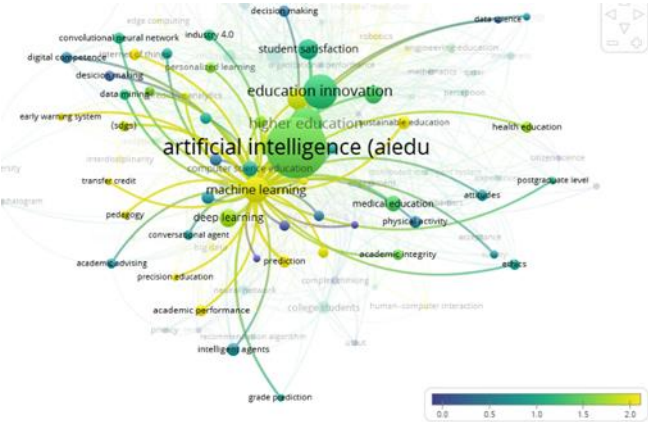

Figure 4. Overplay Visualization of Machine Learning

In the thematic mapping of AI in education, with settings of 390 number of words, cluster frequency value of 9, and 2 number of labels, key influences have emerged covering various fields, along with the relevance of using data mining. The emphasis on the application of ML in the peer tutor model is essential, especially considering the complexity of the "learner effect," which has been extensively explored but with slow evolution. Increasingly urgent is the development and further publication on the topic of "performance evaluation," which is evolving and requires more attention. While the topic of "learning outcomes" has been a primary focus, continuously involving basic thematic areas adds complexity to the analysis (see Figure 5), which includes a strategic diagram and thematic network. This mapping has been processed with a comprehensive science map.

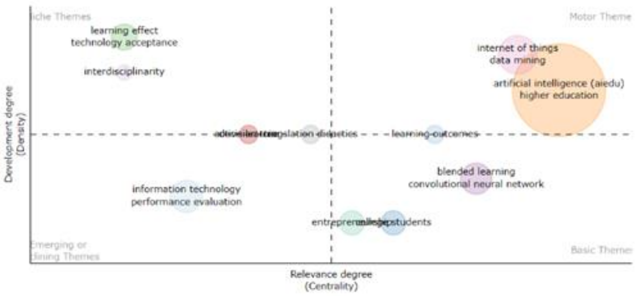

Figure 5. Thematic Map

There are 40 publications on ML innovations in the classification and clustering of student scores and performance, aiding teachers in making appropriate treatment decisions based on clusters. ML applications include predicting scores based on student performance similarities [20], using clustering techniques in EDM data science analysis with naïve Bayes and random forest algorithms in a bi-level learning framework [21]. ML can be used to make student performance model decisions by analyzing student performance, correlation levels, and feature rankings with ridge regression methods [22]. The application of neural networks (ANN) is suitable for predicting performance at educational levels, proving that the used predictors have a significant impact on outcome variables [23]. The use of grouping, classification, and regression algorithms like decision trees, SVM, and artificial neural networks on variables such as online student performance, face-to-face student performance, and blended learning outcomes can reflect the achievement of good performance [24]. The effectiveness of curriculum characteristics and learning performance prediction models can be obtained by spiking feedforward neural algorithms [25]. The application of feature engineering and instance engineering techniques can detect over 72% of students at risk of dropping out [26]. Data mining processed by ML can be a new process and innovation to improve student and teacher learning performance, serving as a tool to collect, analyze, measure, and report student performance aimed at understanding and optimizing learning [27]. With the sophistication of ML, it is possible to analyze the complex characteristics of potential NPT tutors for application in education through the diffusion of educational innovation, although the global distribution of its role in education is still unknown. One current challenge is the outdated education system and the learners' ability to adapt to social and mental model changes [28]. Therefore, designing better quality education is one of the main goals and solutions for the academic community today [29]. According to researchers' findings, the K-Means algorithm has not been widely used for improving student quality and performance. Its application through tutor clustering can help in assessing the measurable effectiveness of NPT and prospective tutors.

III. METHODOLOGY

This preliminary study employed a literature review method as a conceptual reinforcement, using the VOSviewer application for mapping previous research [30], Bibliometix for analyzing the strategic diagram of previous research [31], and the machine learning application Rapidminer for the clustering of peer tutors [32]. This was followed by designing its application so that it can be used by teachers for tutor clustering, referring to the correlated student variable data. The research stages included: formulating the research problem, reviewing previous research on ML in improving educational quality, identifying the K-means algorithm for clustering peer tutors, and interpreting the findings.

IV. APPLICATION OF K-MEANS ML ALGORITHM IN NEAR PEER TEACHING

A. Classification and Clustering of Student Data

Classification involves grouping data based on label similarity. Students in each class have different abilities. (see Figure 6).

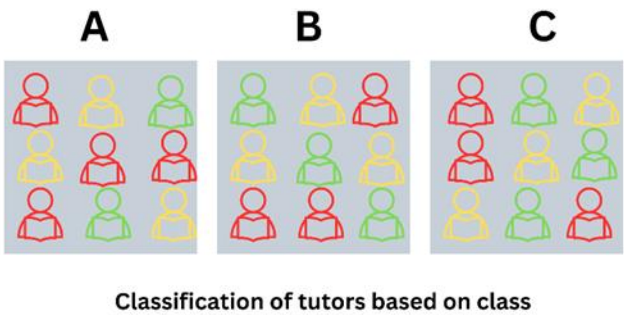

Figure 6. Classification of Tutors in Class

Clustering is similar to classification in that it groups data, but clustering groups data based on data similarity. In the case of peer tutor learning, clustering is necessary to group tutors based on the potential and abilities of students (motivation, interest, achievement, learning outcomes). This is very useful as a guideline for teachers to determine training interventions that match the abilities of tutors. (see Figure 7).

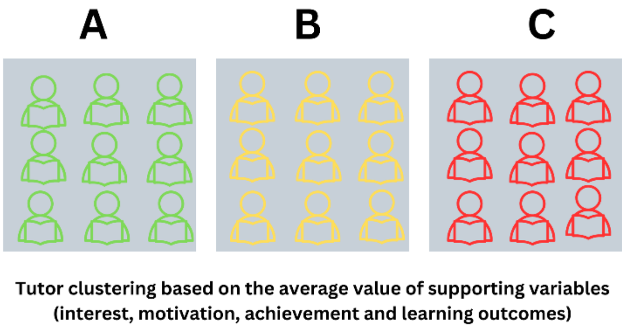

Figure 7. Clustering of Tutors Based on Student Abilities

The benefits of clustering include helping with data segmentation, which is very practical in dividing and grouping peer tutor data with a variety of variables. Clustering can label data for use in the classification process. Euclidean Distance is used to measure data similarity.

B. Measuring Data Distance

Data similarity can be measured by using the Euclidean Distance equation, as follows:

$$d = \sqrt{\sum_{N}(xi - yi)^2}$$

d : Distance Between Data

Σ : Summation of Squared Differences of xi and yi

x dan y : Attributes

i : Index of Attribute

Example of student data illustration:

Table 1.Example Illustration of Student Data

Achievement	Learning Outcome	Interest	Motivation
70	70	50	75
80	90	85	90
65	60	80	80

Then there is new data:

Table 2. Example of New Student Data

Achievement	Learning Outcome	Interest	Motivation
70	70	85	80

To measure the distance of values, Euclidean Distance measurement can be used as follows:

$d_1 = \sqrt{(70 - 70)^2 + (70 - 70)^2 + (85 - 50)^2 + (80 - 75)^2}$

$d_1 = 35.355$

$d_2 = \sqrt{(70 - 80)^2 + (70 - 90)^2 + (85 - 85)^2 + (80 - 90)^2}$

d_2 = 24.494

$d_3 = \sqrt{(70-65)^2 + (70-60)^2 + (85-80)^2 + (80-80)^2}$

d_3 = 12.247

It can be seen that the new data is closely related to the third data (d3) with a distance of 12.247, indicating a strong similarity with the third data set. Conversely, the distance of the new data from d1 is 35.355 and from d2 is 24.494, which is significantly farther, being more than twice the distance of the new data from d3. If illustrated by looking at the variables of motivation and interest on an x,y curve (see Figure 8), it is evident that the new data is very close to the third set.

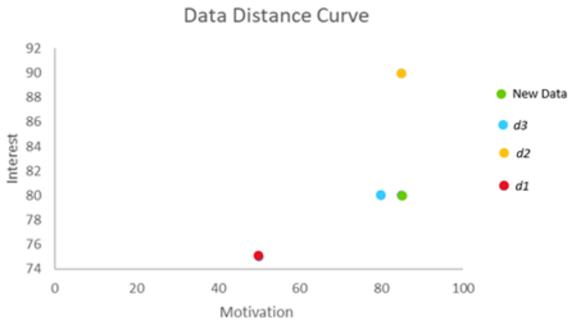

Figure 8. Data Distance Curve

C. K-Means Algorithm

K-Means is an algorithm for data clustering based on the measurement of data distance to the nearest cluster center. Each data point is grouped according to the nearest cluster center. (see Figure 9). In 3 cluster centroid centers.

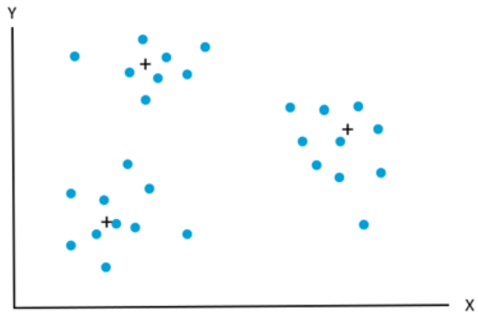

Figure 9. Illustration of 3 Centroids and Data Distribution

The plus sign indicates the cluster center. Each cluster center forms a cluster based on the nearest data, resulting in 3 different clusters from the 3 centroid cluster centers. Each data point is grouped based on its proximity to each data centroid. (see Figure 10).

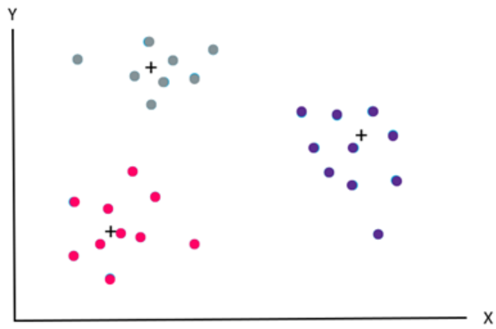

Figure 10. Formation of 3 Clusters Based on Distance to Centroid

D. How to Determine the Cluster Center (Centroid)?

The initial step is to determine the number of cluster centers to be used. Initially, cluster centers are placed randomly, and then each will gather its data. The cluster centers are then adjusted by calculating the average of each data point. The process of moving cluster centers stops when the significance of the data center movement becomes relatively small.

E. Implementation of the K-Means Algorithm in Rapidminer..

Teachers can prepare data on the interests, motivations, achievements, and learning outcomes of tutors (see Figure 11). The K-Means algorithm can then be selected from the “operator” menu under the name clustering. Setting the “operator” parameters is necessary, which include the number of clusters determined by the teacher, the “max run” feature to determine the number of clustering processes, and the “mixed measures” option in the “measure types” feature to facilitate the Euclidean Distance clustering process with different attributes.

Figure 11. Rapidminer Tutor Clustering Design

F. How to Determine the Correct Number of Clusters?

The appropriate number of clusters can be determined by adding the “cluster distance performance” analysis feature to the clustering design as seen in (see Figure 12). Teachers can measure the quality of clusters based on data distance. It should be noted that “cluster distance performance” will not process if there are polynomial attribute data such as the “status_type” attribute. The solution is to drop this attribute so that analysis can be processed. However, if the teacher needs polynomial attributes, they can be converted to numeric.

Figure 12. Adding Cluster Distance Performance Feature to Tutor Cluster Design

The design results will show the average data distance of each cluster, allowing teachers to analyze and compare the average data distances of each cluster. Then, teachers can choose the appropriate number of clusters using the “elbow method”. The average data distances of each cluster are visualized using an x,y curve, and teachers can select the number of clusters based on the most distinct “elbow” on the line.

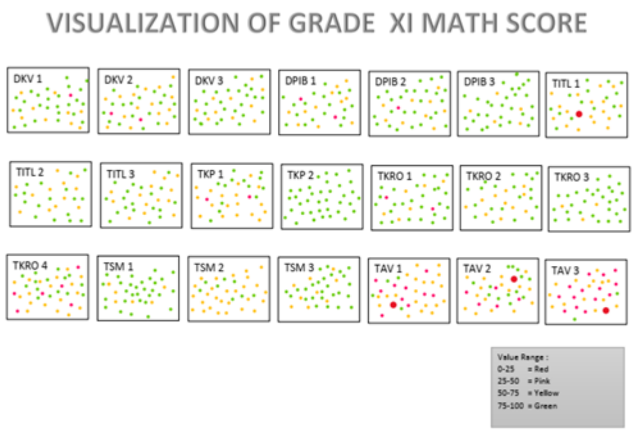

Figure 13. Visualization of Clustering of XI Grade Students' Mathematics Final Exam Scores with NPT

Learning

The implementation of machine learning can be applied to 3rd-semester students (see Figure 13) who have a higher average score and more homogeneous score mapping to become prospective tutors. After the clustering process using machine learning with the K-Means algorithm, teachers can conduct deep interviews by sampling some students from each cluster. This can determine the appropriate tutor training design, thus minimizing the chances of tutors making errors in delivering teaching materials to other students.

V. CONCLUSION

The K-Means algorithm in ML offers great potential in clustering tutors based on variables such as interest, motivation, achievement, and student learning outcomes accurately. By utilizing this method, teachers can identify complex patterns and hidden relationships between these correlated tutor ability factors, allowing for personalized teaching approaches and more effective tutor training. Grouping tutors based on these characteristics, the system can dynamically adjust learning strategies for each student group, maximizing teaching potential and enhancing academic achievement. Thus, the application of the K-Means algorithm in tutor clustering promises not only efficiency in the teaching process but also can have a significant positive impact on students' learning progress.

VI. FUTURE RESEARCH

Facing the urgent demand for quality improvement in education, future research can focus on implementing tutor clustering using the K-Means algorithm as an innovative solution. By leveraging the strength of this algorithm in grouping tutors based on interests, motivation, achievements, and student learning outcomes, such research can design a recommendation system that optimizes teaching effectiveness. This approach not only has the potential to improve academic success rates but also paves the way for more adaptive and personalized learning approaches, leading the world of education towards a more dynamic and responsive era to individual student needs.

VII. ACKNOWLEDGEMENTS

This research was supported by the Ministry of Education and Culture of the Republic of Indonesia through the PMDSU Program. We express our gratitude for the financial support that made this research possible. Thank you to everyone who has contributed and guided throughout this research.

REFERENCES

[1] Rehaf Madani and in Dubai Sufian Forawi, “Teacher Perceptions of the New Mathematics and Science Curriculum: A Step Towards STEM Implementation in Saudi Arabia أل والعلوم : بيقت ط تصورات المدرسين للمنهج المضمن حديثا ً لمادتي الرياضيات النماذج من برنامج حد " STEM بالمملكة العربية السعودية "," 2017.

[2] R. Santos, D. Anderson, and M. Milner-Bolotin, “Research trends in international science, technology, engineering, and mathematics education conference series: An analysis of a decade of proceedings,” Front Educ (Lausanne), vol. 7, Jan. 2023, doi: 10.3389/feduc.2022.1099658.

[3] Y. Li, J. E. Froyd, and K. Wang, “Learning about research and readership development in STEM education: a systematic analysis of the journal’s publications from 2014 to 2018,” International Journal of STEM Education, vol. 6, no. 1. Springer, Dec. 01, 2019. doi: 10.1186/s40594-019-0176-1.

[4] Y. Li, K. Wang, Y. Xiao, and J. E. Froyd, “Research and trends in STEM education: a systematic review of journal publications,” International Journal of STEM Education, vol. 7, no. 1. Springer, Dec. 01, 2020. doi: 10.1186/s40594-020-00207-6.

[5] R. Vhalery, A. M. Setyastanto, and A. W. Leksono, “KURIKULUM MERDEKA BELAJAR KAMPUS MERDEKA: SEBUAH KAJIAN LITERATUR,” Research and Development Journal of Education, vol. 8, no. 1, p. 185, Apr. 2022, doi: 10.30998/rdje.v8i1.11718.

[6] O. Trabelsi, A. Gharbi, L. Masmoudi, and M. Mrayeh, “Enhancing female adolescents’ engagement in physical education classes through video-based peer feedback,” Acta Gymnica, vol. 50, no. 3, pp. 93–104, 2020, doi: 10.5507/ag.2020.014.

[7] M. S. F. Mohd Shafiaai, A. Kadirvelu, and N. Pamidi, "Peer mentoring experience on becoming a good doctor: student perspectives," BMC Med Educ, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12909-020-02408-7.

[8] E. A. Ensher and S. E. Murphy, "The Mentoring Relationship Challenges Scale: The impact of mentoring stage, type, and gender," J Vocat Behav, vol. 79, no. 1, pp. 253–266, Aug. 2011, doi: 10.1016/j.jvb.2010.11.008.

[9] K. Overeem and P. & L. GVO drukkers en vormgevers, Doctor performance assessment : development and impact of a new system. s.n.], 2011.

[10] J. Nicky Hudson and A. L. Tonkin, "Clinical skills education: Outcomes of relationships between junior medical students, senior peers and simulated patients," Med Educ, vol. 42, no. 9, pp. 901–908, Sep. 2008, doi: 10.1111/j.1365-2923.2008.03107.x.

[11] D. Engels, E. Kraus, B. Obirei, and K. Dethleffsen, "Peer teaching beyond the formal medical curriculum," Adv Physiol Educ, vol. 42, pp. 439–448, 2018, doi: 10.1152/advan.00188.2017.-Peer-assisted.

[12] K. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 2, p. 26, 2018, doi: 10.9781/ijimai.2018.02.004.

[13] J. Xu, K. H. Moon, and M. Van Der Schaar, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE Journal on Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742–753, Aug. 2017, doi: 10.1109/JSTSP.2017.2692560.

[14] P. Kathiroli, V. Vijayalakshmi, K. Panimalar, and S. Janarthanan, "Predicting the performance of instructors using Machine learning algorithms", doi: 10.37896/HTL26.12/2315.

[15] Bulte, A. Betts, K. Garner, and S. Durning, "Student teaching: Views of student near-peer teachers and learners," Med Teach, vol. 29, no. 6, pp. 583–590, Jul. 2007, doi: 10.1080/01421590701583824.

[16] O. Ten Ten Cate and S. Durning, "Dimensions and psychology of peer teaching in medical education," Med Teach, vol. 29, no. 6, pp. 546–552, Jul. 2007, doi: 10.1080/01421590701583816.

[17] E. T. Murtisari, D. Puspitasari, and A. A. Setiamunadi, "Peer teaching to assist tertiary EFL grammar learning: Indonesian tutors' perceptions of challenges and strategies," Journal of Language and Education, vol. 6, no. 4, pp. 204–215, 2020, doi: 10.17323/jle.2020.10763.

[18] X. Chen, D. Zou, G. Cheng, and H. Xie, "Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education," Comput Educ, vol. 151, Jul. 2020, doi: 10.1016/j.compedu.2020.103855.

[19] H. Guruler, A. Istanbullu, and M. Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases," Comput Educ, vol. 55, no. 1, pp. 247–254, Aug. 2010, doi: 10.1016/j.compedu.2010.01.010.

[20] L. Barik, O. Barukab, and A. Ali Ahmed, "Employing artificial intelligence techniques for student performance evaluation and teaching strategy enrichment: An innovative approach," International Journal of Advanced and Applied Sciences, vol. 7, no. 11, pp. 10–24, Nov. 2020, doi: 10.21833/ijaas.2020.11.002.

[21] L. Nanglae, N. Iam-On, T. Boongoen, K. Kaewchay, and J. Mullaney, "Determining patterns of student graduation using a bi-level learning framework," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 4, pp. 2201–2211, Aug. 2021, doi: 10.11591/EEI.V10I4.2502.

[22] Pérez-Suay, R. Ferrís-Castell, S. Van Vaerenbergh, and A. B. Pascual-Venteo, "Assessing the Relevance of Information Sources for Modelling Student Performance in a Higher Mathematics Education Course," Educ Sci (Basel), vol. 13, no. 6, Jun. 2023, doi: 10.3390/educsci13060555.

[23] Y. Baashar et al., "Evaluation of postgraduate academic performance using artificial intelligence models," Alexandria Engineering Journal, vol. 61, no. 12, pp. 9867–9878, Dec. 2022, doi: 10.1016/j.aej.2022.03.021.

[24] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis," Front Psychol, vol. 12, Dec. 2021, doi: 10.3389/fpsyg.2021.698490.

[25] Liu, H. Wang, and Z. Yuan, "A Method for Predicting the Academic Performances of College Students Based on Education System Data," Mathematics, vol. 10, no. 20, Oct. 2022, doi: 10.3390/math10203737.

[26] J. Fernandez-Garcia, J. C. Preciado, F. Melchor, R. Rodriguez-Echeverria, J. M. Conejero, and F. Sanchez-Figueroa, "A real-life machine learning experience for predicting university dropout at different stages using academic data," IEEE Access, vol. 9, pp. 133076–133090, 2021, doi: 10.1109/ACCESS.2021.3115851.

[27] Wibawa, J. S. Siregar, D. A. Asrorie, and H. Syakdiyah, "Learning analytic and educational data mining for learning science and technology," in AIP Conference Proceedings, American Institute of Physics Inc., Apr. 2021. doi: 10.1063/5.0041844.

[28] H. A. M. Gerlache, P. M. Ger, and L. de la F. Valentín, "Towards the Grade's Prediction. A Study of Different Machine Learning Approaches to Predict Grades from Student Interaction Data," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 7, no. 4, pp. 196–204, 2022, doi: 10.9781/ijimai.2021.11.007.

[29] O. D. Castrillón, W. Sarache, and S. Ruiz-Herrera, "Prediction of academic performance using artificial intelligence techniques," Formacion Universitaria, vol. 13, no. 1, pp. 93–102, Feb. 2020, doi: 10.4067/S0718-50062020000100093.

[30] N. Jan van Eck and L. Waltman, "VOSviewer Manual."

[31] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," J Informetr, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.

[32] V. Kotu and B. Deshpande, Predictive analytics and data mining : concepts and practice with RapidMiner.