Application of Clustering Algorithms to enhance Personalized Learning through Recommendation Model

Mhd Saeed Sharif
Intelligent Technologies Research Group, School of Architecture, Computing and Engineering, UEL
London, United Kingdom
s.sharif@uel.ac.uk

Madhav Theeng Tamang
Intelligent Technologies Research Group, School of Architecture, Computing and Engineering, UEL
London, United Kingdom
u1430774@uel.ac.uk

Dyuthi Mani
Intelligent Technologies Research Group, School of Architecture, Computing and Engineering, UEL
London, United Kingdom
u2517230@uel.ac.uk

Wael Elmedany
College of Information Technology
University of Bahrain
Kingdom of Bahrain
welmedany@uob.edu.bh

Abstract*— Technology and education have recently changed student involvement with learning resources and educational experiences. It is very important to have a platform which can provide personalized learning resources and group students by learning patterns. The research develops recommendation systems and grouping algorithms to help students with their learning patterns and performance measures. To record multidimensional student learning experiences, demographics, learning activities, problem-solving behaviours, and performance measures are extracted and processed. This system makes personalised student suggestions via collaborative filtering, especially Alternating Least Squares (ALS). The model predicts learning materials and problem sets based on student preferences and ability levels by analysing historical student interactions with learning resources. The research also uses clustering methods like K-means clustering to group students with similar learning and performance patterns. Clustering analysis lets instructors discover student group traits and tailor interventions and support techniques. This research also examines temporal relationships in student learning sequences using sequential models. Sequential learning activities and problem-solving behaviour of students help recurrent neural networks (RNNs), or sequential pattern mining algorithms predict their next activities. This research used a huge dataset of over sixteen million exercise logs and applied collaborative filtering (ALS) and K-means clustering to find learning patterns. The ALS-based recommendation system achieved a Mean Squared Error (MSE) of 0.23, showed excellent predictive accuracy. Clustering grouped students by learning behaviours, enabling targeted interventions. This research improves personalised learning using machine learning and data analytics in education. The recommendation system can help instructors tailor training, give individualised support, and boost academic success for various pupils by revealing student learning patterns.

Keywords—personalised learning methods, clustering algorithms, recommendation system, machine learning

I. INTRODUCTION

Recently, a lot more online learning tools have come out. Some main reasons are the rising demand for higher education, the lack of teachers, improvements in technology and artificial intelligence (AI), and the new effects of COVID-19. Massive open online course sites have recently been added to the education systems of several schools. This integration aims to improve standard classrooms, get around time and location problems, and support justice by giving more people access to high-quality education [1]. Some schools have also added virtual laboratories to their lessons, which let students do studies, especially those who can't get to real labs. Still, the current systems for online training have some big problems. Personalised education has mostly been limited to a certain kind of recommendation system, but it has more potential than just offering appropriate online classes to each user. The goal of current suggestion systems and personalised schooling is different. The first one focuses on user involvement as a way to make the most money, with a focus on the system and measurable data. The second approach, on the other hand, focuses on learning results and needs help describing them [2].

Since technology and education have come together in recent years, there has been a big change in how students interact with learning tools and school activities. One place where this change is very clear is in the field of personalised learning. Static, one-size-fits-all ways of teaching are being replaced by dynamic, adaptive systems that tailor lessons to each student's hobbies, skills, and weaknesses [3]. New analytics techniques like machine learning (ML) and others are helping to change this because they let teachers use huge amounts of data to improve and tailor their students' learning.

Machine learning and smart data are very important for implementing personalised learning. Algorithms capable of processing and analysing vast volumes of education data can teach teachers much about their pupils' habits, interests, and classroom development [4]. These results can help make adaptable learning systems that instantly change course materials, student growth, and teacher methods. Advanced analytics techniques can also be used to find trends, relationships, patterns, and patterns in educational data. This can give teachers useful feedback for making decisions about content, teaching methods, and policy.

II. Literature Review

Learning analytics refers to the use of technology for machine learning in schools. The idea is to personalise the teaching method to each student's requirements and abilities. This could include providing feedback and teaching resources to at-risk kids. It employs approaches from semantics [5], machine learning, data representation, and learning research [6]. With AI-based competence learning, educational institutions may be proactive by gaining valuable insights and predictions about the skills in which students may excel. In addition to competency-based learning, learning analytics uses AI's vast learning ability. AI can analyse a wide range of dropout rate-related characteristics to classify incoming students according to their likelihood of leaving school [5]. This procedure yields valuable data for educational institutions and aids in developing early warning systems. The next challenge for educational analytics is to expand its scope to include a wider range of subjects, including social skills, the arts, literature, and other subjects. This adds another complexity to measuring and evaluating learning outcomes or capacities [6]. The challenge with learning analytics is that they need to be flexible enough to be used to different courses and organisations specific to their learning environments.

Multiple studies indicate that data mining has significant promise for improving our understanding of student behaviour and academic achievement. Teachers and regulators can tailor interventions to their students' needs using advanced analytics to gain insightful knowledge about the variables affecting their achievement. Academics, teachers, and data scientists need to work together across fields right away to fill in the gaps and solve the problems that have been found. To move the field of predictive analytics in education forward, we need standardised methods, strong comparison studies, and continuous analyses [7]. Finding clear causal links is key to improving educational outcomes. Adaptive education systems use AI to automatically identify and group students by learning styles (LSs), offering more accurate and flexible insights than traditional surveys [8]. Machine learning algorithms are used in these methods to make e-learning more personalised [9]. To do this, they instantly and flexibly match each student's behaviour traits to a specific learning style (LS). The goal is to improve the e-learning experience and help each person learn better. The success of personalised adaptive education systems depends on how well they identify and collect data on each student's learning style based on their needs and traits and how this data is used to create an intelligent and adaptable learning environment. So, flexible learning systems can use LS knowledge to provide precise customisation if they correctly classify learners' LSs. Traditionally, students' LSs are evaluated by having them fill out a form.

Teachers and schools can acquire valuable lessons from vast amounts of data by using data analytics, an effective tool in education. Many kinds of information are accessible in the data sources, including demographic information, student success statistics, attendance records, and social and emotional indicators. Giving teachers evidence-based information on how kids are performing, how they learn, and where they need to improve is the primary objective of data analytics in education: to improve results. As an outcome of this analysis method, teachers can better understand the learning process and adapt their methods and strategies to suit each student's unique requirements better [10]. In the past, teachers used random notes and regular test scores to judge how well their students were doing and decide how to teach them. This approach has changed since better data-gathering technology and powerful analysis tools emerged.

How well personalised adaptive education systems work depends on how they sort and collect information about each student's learning style based on their needs and traits and how that information is used to make a smart and flexible learning environment [7]. So, flexible learning systems can use LS knowledge to provide precise customisation if they correctly classify learners' LSs. Traditionally, students' LSs are evaluated by having them fill out a form. However, there are some big problems with this method. At first, filling out surveys is a process that takes a lot of time. In addition, the poll data may need to be more accurate when determining how satisfied students are. This is because students only sometimes know or care about how satisfied they are, which makes them give answers that could be more helpful. LSs are also always changing, which is different from poll data, which stays the same. To get around these problems, individual adaptive education systems use AI to automatically determine how people learn best. Automated identification of learning styles to put students into groups based on how they like to learn is better than using surveys because it works more quickly. This method is flexible and can be changed depending on how the students behave. Customising e-learning is done with these methods, which use techniques from the field of machine learning (ML). To do this, they instantly and constantly match each student's behaviour traits to a certain learning style (LS). The goal is to improve the e-learning experience and help each person learn better [11].

Personalised customisable education systems that use AI have changed how people learn, especially when it comes to getting around the problems that come with set learning styles (LSs). These systems use machine learning techniques to ensure students have a personalised learning experience. They look at how students act to get the best learning results and match that with specific learning methods. A study discusses a unique development towards automatically identifying Learning Systems using artificial neural network techniques. This indicates a developing field in adaptive e-learning environments driven by machine learning. However, the absence of comparative studies of deep learning methods indicates a significant lacuna in the current literature that requires further empirical research. This indicates that additional research is required to demonstrate and compare the effectiveness of various deep-learning algorithms in classifying LSs. This will enhance the flexibility and utility of personalised learning options [12].

A study designed a Mamdani Fuzzy Set Optimisation framework to give each student personalised feedback [13]. Many studies have been done on backward analysis, which means looking at data from the past to guess how students will do in the future. Even though these attempts have led to useful discoveries, only some ongoing studies track student progress over long periods. Longitudinal studies could better explain how students' learning paths change over time and how well predictive models work. If we understood the basic factors affecting student success better, we could design programmes and support systems to help students do better in school.

III. METHODOLOGY

This research utilizes a data-driven approach to improve personalized learning experiences for students through the development of recommendation systems and clustering algorithms. To capture the multifaceted nature of student learning, the system extracts and processes data including student demographics, learning activities, problem-solving behaviors, and performance measures. The recommendation system is built using collaborative filtering, specifically the Alternating Least Squares (ALS) algorithm, which analyzes historical student interactions with learning resources to predict suitable learning materials and problem sets based on student preferences and ability levels. A Kaggle dataset of Junyi Academy Foundation which is an educational platform based in Taiwan. It holds more than sixteen million exercise attempt logs of over 72,000 students that were logged in a year (from August 2018 to July 2019). It consists of three main tables in the dataset: Log_Problem.csv, which holds 16,217,311 logs of problem tried by 72,630 students; Info_Content.csv, which provides metadata on the exercises, each consisting of multiple problems; and Info_UserData.csv, which contains metadata on the selected registered students. This rich dataset is what we use for our research for personalized learning through machine learning techniques as show in the various stages of this research as illustrated in Fig. 1 below.

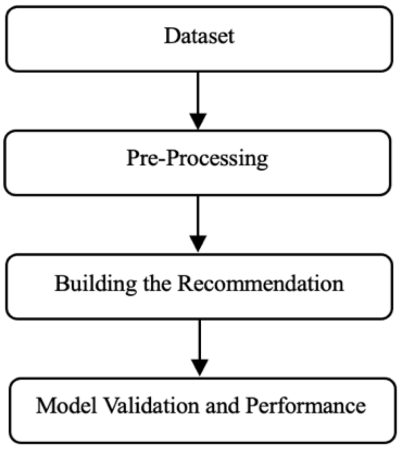

Fig. 1. Flowchart diagram on developing a personalized learning system

A. Data Pre-processing

The data is pre-processed, and the features are designed at the beginning of the study. By looking at how students interact with an educational platform, such as through problem-solving tasks, session length, and different success measures, the dataset gives us a lot of information about how each student learns, what they like, and what problems they face. The suggestion system and grouping analysis are based on a deep understanding of how engaged students are and how well they are doing in school. This research uses data from a learning management system (LMS). The dataset presents the interactions of student with an educational platform, including problem-solving activities, session time, and performance measures, to provide insights into individual learning behaviours, preferences, and obstacles. The dataset is crucial for completing our research goals, which involve developing personalised recommendation algorithms to suggest learning resources and problems tailored to each student's needs and interests. The dataset enables us to apply clustering algorithms to categorise students by learning patterns and performance profiles. We can gain practical insights and suggestions to enhance the engagement and student learning through data analysis [14].

Clustering models are crucial when studying educational data because they enable us to identify patterns, comprehend student behaviour, and customise instruction. By allocating students to groups according to their connections, clustering allows for more customised guidance through targeted activities and interventions [15]. It also helps to find students who are at risk and need extra help by looking at the traits and actions of groups. This proactive method helps teachers step in early to make sure students do well in school. Clustering helps institutions plan their strategies and decide how to use their resources by showing which programmes are strong and which are weak. Clustering gives us information that helps us make decisions about hiring, offering specialised help, and spending money on learning materials [16]. A common method called K-means clustering divides data into groups based on how similar the data is. Student performance analysis sorts of students into groups based on things like tries, time spent, correctness rates, and tips. The algorithm changes the centres over and over to make stable groups, which gives us useful information about how engaged and skilled the students are. Then, teachers can change the methods to fit the needs of each group of students, which improves learning overall.

MSE is used because it directly measures the squared differences between actual and predicted values (in collaborative filtering) or between points and their centroids (in K-means clustering), providing a smooth optimisation function. RMSE provides a more interpretable version of MSE, offering the same units as the original data, making it easier to evaluate the model's performance in real-world terms. Thus, MSE and RMSE are widely used in these contexts because they align with the optimization objectives and provide interpretable and meaningful ways to assess the model's performance.

IV. IMPLEMENTATION

A. Analysis of Student interaction with the exercises

The Table I below show the number of times students attempted each problem across different levels. It can be observed that easier problems, typically found at the beginning levels, receive a higher number of attempts from students. As the difficulty increases, the number of attempts on the more advanced problems gradually decreases.

TABLE I. NUMBER OF PROBLEMS ATTEMPTED PER LEVEL

Level	Number of Problems
0	11809119
1	2352668
2	996819
3	751424
4	307281

Likewise, Fig. 2 displays the number of problems attempted at each level, while Fig. 3 compares the number of problems solved versus those not solved by students.

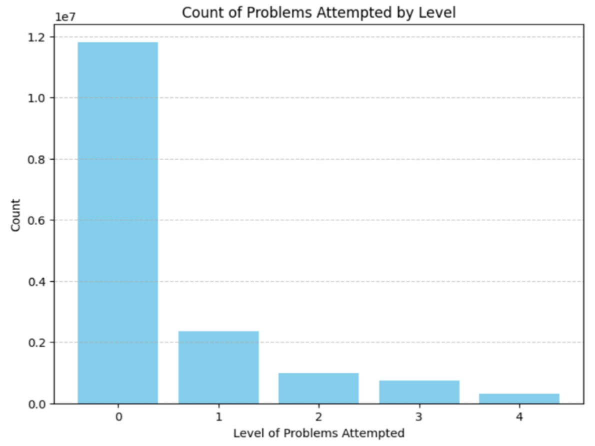

Fig. 2. Number of problems attempted by students in different levels

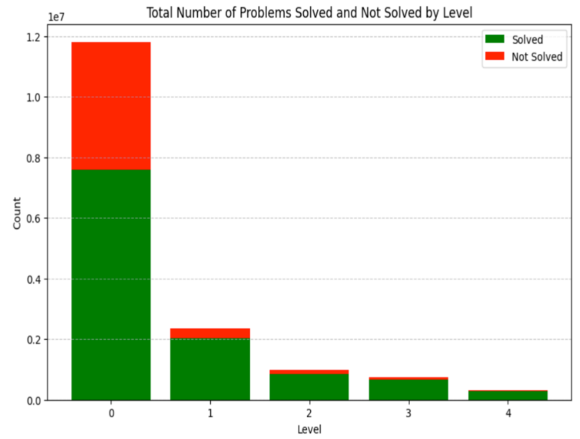

Fig. 3. Comparison of problems solved and not solved by students.

Similarly, Fig. 4 illustrates the average time students spend on each problem level. It shows the average total seconds taken for problems at each difficulty level.

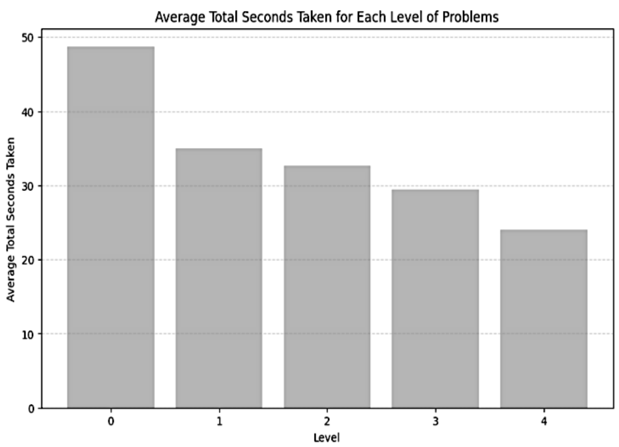

Fig. 4. Average total seconds taken for each level.

B. K-Means Clustering

This is an unsupervised clustering approach that organises data by similarity. It's most efficient when the cluster count is known or estimated. K-means clustering may classify pupils by behaviour, engagement, and performance on an educational platform. K-means clustering can divide pupils by interaction, performance, or other educational data. Students can be categorised by several sessions, exercise duration, accuracy rates, or hints used. Instructors can learn about different learning styles by clustering pupils by conduct or involvement. These profiles can inform instructional design, personalised learning, and intervention tactics. Regression and classification algorithms can employ K-means clusters as input variables. Clusters of students with different engagement or performance can indicate future performance, dropout risk, or learning outcomes. Clustering analysis helps identify children with special learning needs or obstacles. This information may help build a curriculum, allocate resources, and provide interventions for children's various learning needs. K-means clustering helps compare student groupings. To determine what causes student success or failure, educators can compare group performance, engagement, and conduct. The elbow approach is employed to find the optimal number of clusters for K-means clustering. Fig. 5 displays the optimal K value in this method.

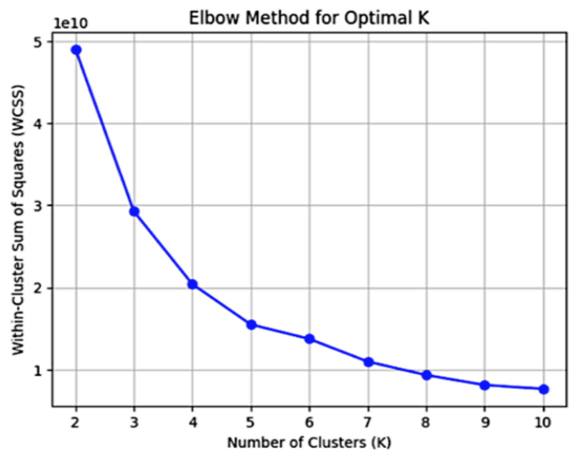

Fig. 5. Application of Elbow method to determine the optimal K.

C. Student Personalisation Problem recommendation

A recommendation engine, often called a recommender system, is a type of information filtering system that makes predictions or offers suggestions for items that the user would find interesting based on past user behavior, preferences, and similarities with other users. A recommendation engine can play a key role in customizing students' learning experiences on educational platforms like the one used in this study. It helps them find relevant exercises, materials, and other resources that meet their individual needs and interests.

D. Collaborative Methods Recommendations

Collaborative filtering is essential for personalising student experiences by providing customised problem recommendations that are based on the collective knowledge of the user community. Collaborative filtering models utilise interaction data from multiple students to offer valuable recommendations that improve learning outcomes, engagement, and satisfaction. Integrating collaborative filtering-based recommendation systems into educational platforms can greatly enhance the effectiveness of student personalization initiatives. Similarly, ALS (Alternating Least Squares) algorithm is also used for collaborative filtering-based recommendation. It imports necessary modules, including ALS for recommendation.

V. Model Evaluation

A. K-means Clustering Model Evaluation

Table II displays an analysis of user performance depending on each cluster. The clusters that the K-means algorithm produced represent different user groups based on the behaviours and attributes of the users in the dataset.

Table II. USER'S PERFORMANCE BASED ON EACH CLUSTER

Level	No of users	Average sessions	Average number of problems attempted	Average Solved	Average total attempts
1	83	5654.72	3892.96	3813.02	15004.25
2	837	2666.42	2064.59	1817.72	6194.30
4	3195	1326.56	1105.1	924.93	2663.41
3	10056	523.85	452.38	365.63	932.59
0	58587	68.45	62.37	50.23	103.91

In Cluster 1, 83 users have a high level of involvement with the platform, as evidenced by the average number of sessions and issues tried as shown in Fig 6. below. While these engaged individuals are dedicated to solving problems, the platform can offer challenging or advanced issue sets to maintain their attention. Encouragement comes in the form of tailored remarks and rewards for learning.

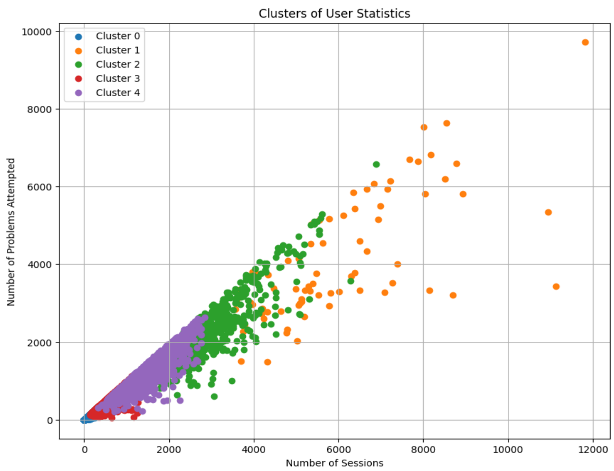

Figure 5. K-means clustering to group student based on number of problems attempend.

Similarly, cluster 2 has 837 moderately engaged users based on their average sessions and issues attempted. Recommending issues that match this cluster's users' skills and interests can improve learning. They might be able to improve with the assistance of a progress-tracking system, related resources, or practice activities.

Likewise, cluster 3 is enormous, but unlike the other clusters, it has a lower level of engagement, with fewer sessions and issues being attempted. Initiatives for individualised participation are required for this cluster. An example would be sending personalised emails or reminders about interesting topics, providing incentives for completing assignments or providing further assistance to overcome learning challenges.

Furthermore, cluster 0 showed the lowest engagement despite being the latest cluster. Very few sessions and issues were attempted. The re-engagement method can focus on the users to increase their interest in the site. This could comprise personalised re-engagement initiatives, rewards or challenges to encourage participation, and user interface improvements that enhance appeal and usability. Additionally, cluster 4 has moderately engaging users who completed several sessions and tasks. This cluster can benefit from personalised recommendations and focused feedback.

B. Recommendation of Problems Based on the Previous Interactions Using the ALS Model

In this study, we created a recommendation model that showed a mean squared error (MSE) of 0.2327, a milestone proving its efficacy and accuracy. The Mean Squared Error (MSE) illustrates the level of similarity between predicted values and actual values, with lower values indicating higher performance. Our successful methodology demonstrates the efficacy of machine learning in education, specifically collaborative filtering utilising Alternating Least Squares (ALS). Learning outcomes can be enhanced by assessing students' interactions with learning materials and providing tailored recommendations. Our technology can comprehend and adjust to individuals' unique learning patterns by properly forecasting the problems a learner will solve based on previous experiences. This customised recommendation system enhances efficiency for students by directing them towards pertinent problems and fosters engagement by adapting learning experiences to their needs. Our model demonstrates resilience and reliability due to its ability to effectively generalize new and unknown data and its low mean squared error (MSE). This generalisation is crucial for practical applications as the model should be able to manage novel student interactions effectively.

The value of our model goes beyond its technological achievements. It increases the effectiveness of education by providing instructors and students with an effective tool for personalised learning. Data and machine learning may develop new possibilities for learning that address student's various needs and interests. The collaborative filtering paradigm suggests promising innovative problems. Based on previous interactions and problem-solving habits, the model forecasts the user's chances of solving a problem. This feature improves personalised education.

The programme suggests issues matched to each user's skill level and interests in order to provide personalised learning interventions. User engagement in challenges appropriate for their skill level creates a dynamic yet achievable learning experience. These principles also assist instructors and platform administrators in selecting problem sets that meet the different needs of their pupils. The model's precise problem prediction improves user engagement and happiness. The suggestion system assigns tasks that fit their skills, increasing their confidence and engagement with the learning platform. To summarise, new challenges improve personalised training and educational platforms.

VI. CONCLUSION AND FUTURE WORK

In this study, we employed several machine learning techniques to develop a recommendation system, particularly for students, to enhance their learning experience. This research aims to identify students based on their academic achievement and provide them with challenging opportunities to improve their learning. By employing clustering techniques and collaborative filtering, we can enhance student performance. Initially, we employed K-means clustering to categorise students based on the associated academic performance. By employing the analysis of problem-solving time, ability level and the number of attempts, we have identified clusters that accurately depict students' distinct learning habits and abilities. Clustering facilitated our comprehension of the varied learning attributes exhibited by our students. This data provides valuable information for tailored treatments and support.

In addition, we implemented a recommendation system based on collaborative filtering, utilising the Alternating Least Squares (ALS) algorithm. Using student-problem interactions, we generated tailored concepts for novel challenges that students would find appealing and advantageous. This strategy increases student engagement and fosters continued learning by exposing them to challenging but appropriate assignments. The Mean Squared Error (MSE) score of 0.23 shows that our recommendation algorithm predicts student preferences and conduct adequately. The low Mean Squared Error (MSE) figure shows a significant connection between projected ratings and student replies, proving our recommendation engine's reliability and usefulness.

Overall, this research advances student-centred teaching. Machine learning has given us a solid mechanism for classifying pupils by similarity and offering personalised educational materials. We want to add features, sequential learning models, and novel recommendation methods to our approach. Our fundamental goal is to give each student customised educational experiences that ignite interest, foster development, and maximise potential.

REFERENCES

[1] Regmi, K. and Jones, L., 2020. A systematic review of the factors–enablers and barriers–affecting e-learning in health sciences education. BMC medical education, 20, pp.1-18.

[2] Muhammad, B.A., Qi, C., Wu, Z. and Ahmad, H.K., 2024. An evolving learning style detection approach for online education using bipartite graph embedding. Applied Soft Computing, 152, p.111230.

[3] Butnaru, G.I., Niță, V., Anichiti, A. and Brînză, G., 2021. The effectiveness of online education during covid 19 pandemic—a comparative analysis between the perceptions of academic students and high school students from romania. *Sustainability*, *13*(9), p.5311.

[4] Pelima, L.R., Sukmana, Y. and Rosmansyah, Y., 2024. Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review. *IEEE Access*.

[5] Ciolacu, M., Tehrani, A.F., Binder, L. and Svasta, P.M., 2018, October. Education 4.0-Artificial Intelligence assisted higher education: early recognition system with machine learning to support students' success. In *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging(SIITME)* (pp. 23-30). IEEE.

[6] Ong, V.K., 2015, July. Big data and its research implications for higher education: Cases from UK higher education institutions. In 2015 IIAI 4th International Congress on Advanced Applied Informatics (pp. 487-491). IEEE.

[7] J. Knox, "Artificial intelligence and education in China," Learning, Media and Technology, vol. 45, no. 3, pp. 1–14, 2020

[8] Kavipriya, P., 2016. A review on predicting students' academic performance earlier, using data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 6(12), pp.101-105.

[9] Yadav, S.K., Bharadwaj, B. and Pal, S., 2012. Data mining applications: A comparative study for predicting student's performance. *arXiv preprint arXiv:1202.4815*.

[10] Jalota, C. and Agrawal, R., 2019, February. Analysis of educational data mining using classification. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 243-247). IEEE.

[11] Colchester, K., Hagras, H., Alghazzawi, D. and Aldabbagh, G., 2017. A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. *Journal of Artificial Intelligence and Soft Computing Research*, *7*(1), pp.47-64.

[12] Chen, L., Chen, P. and Lin, Z., 2020. Artificial intelligence in education: A review. Ieee Access, 8, pp.75264-75278.

[13] SCOTT, M. (2020). *Kaggle*. Retrieved from https://www.kaggle.com/code/kianwee/eda-students-academic-performance-dataset

[14] Murtaza, M., Ahmed, Y., Shamsi, J.A., Sherwani, F. and Usman, M., 2022. AI-based personalized e-learning systems: Issues, challenges, and solutions. *IEEE Access*, *10*, pp.81323-81342.

[15] Tahiri, M., Ismaîli, B.A. and Bakkali, S., 2023, December. Applications of AI in higher education: a review of the literature engineering education from developing countries. In *2023 7th IEEE Congress on Information Science and Technology (CiSt)* (pp. 456-461). IEEE.

[16] Hilali, K., Chergui, M. and Ammoumou, A., 2023, November. Adaptive Learning Systems: A Comprehensive Overview and Identification of Challenges. In *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)* (pp. 1-7). IEEE.