Egyptian Informatics Journal 26 (2024) 100469

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Predict student learning styles and suitable assessment methods using click stream

Ahmed Rashad Sayed [a,b], Mohamed Helmy Khafagy [c], Mostafa Ali [b], Marwa Hussien Mohamed [a,*]

[a] Information System Department, Information System and Computer Science Faculty, October 6 University, 6th of October 12573, Egypt
[b] Information System Department, Faculty of Computers and Artificial Intelligence, Fayoum University, Faiyum 63514, Egypt
[c] Computer Science Department, Faculty of Computers and Artificial Intelligence, Fayoum University, Faiyum 63514, Egypt

ARTICLE INFO

Keywords:
Adaptive learning
Machine learning
Classification
Learning styles

ABSTRACT

Adaptive learning, which aims to give each learner engaging, effective learning experiences, is one method of offering modified education. Adaptive learning seeks to consider the student's unique characteristics by personalizing the learning course materials and evaluation procedures. To determine the student's preferred learning strategies, we first ascertain their attributes utilizing VAK learning styles. In this study, we developed an integrated model to classify learners based on their learning activity clicks by combining machine learning algorithms like K-Nearest Neighbor (KNN), random forest (RF), and support vector machine (SVM) and Logistic regression (LR) with semantic association, which is used to help us map learning activity with VAK learning style. This enables us to classify learners, determine their preferred methods of learning, and offer the most suitable; as a result, we were able to group pupils according to their learning styles and provide the best evaluation technique or strategies. To assess the effectiveness of the suggested model, several tests were executed on the actual dataset (Open University Learning Analytics Dataset, or OULAD). According to studies, using a Random Forest algorithm, the suggested model can predict which evaluation strategy or strategies will be most effective for each student and can classify individuals with the highest degree of accuracy—98%.

1. Introduction

The development of the educational system because of the advancement of Information Technology, particularly the Internet, creates prospects for improving the information services provided by educational institutions. E-learning is an educational method that involves technological tools. One of the e-learning skills is the Learning Management System (LMS). Students can enter lecture data, discussion boards and chat rooms and access lecture assignments given by lecturers [1]. Through the Learning Management System. The popular learning management system that is now in use is Moodle. Students' skills and actions during online learning can be recorded by Moodle and kept in the logs of Moodle.

The problematic with e-learning is that undergraduates are less engaged and leave the classroom more frequently; therefore, it is necessary to understand students' preferences in the learning process by considering the learning style of each situation student. where there are different levels of knowledge among students during the study process. The concept of learning style is the way students choose to study efficiently. Students will understand their wishes during the learning process if they are familiar with their favourite learning method. Since Moodle is unable to detect students' learning preferences automatically. Student behaviour should be analyzed based on the number of times they access e-learning in Moodle, and completing a questionnaire on learning styles is required to determine the best learning style for each student [2].

Students can access video lectures and educational resources on the leading MOOC platforms. In addition to successful completion of the course, students will also be awarded a certificate. The MOOC platform ensures that the material is free from grammar, punctuation, and spelling mistakes. Although MOOCs are very popular, the high dropout rate of students and poor performance measures are frustrating, as less than 13 % of pupils [3]. As a result, significant contributors to students' lack of interest in courses include their lack of participation and their

* Corresponding author.
E-mail addresses: Ar2348@fayoum.edu.eg, Ahmed.rashad.csis@o6u.edu.eg (A. Rashad Sayed), Mhk00@fayoum.edu.eg (M. Helmy Khafagy), mam16@fayoum.edu.eg (M. Ali), Marwa.hussien.csis@o6u.edu.eg, Eng_maroo1@yahoo.com (M. Hussien Mohamed).

https://doi.org/10.1016/j.eij.2024.100469
Received 28 December 2023; Received in revised form 14 March 2024; Accepted 28 March 2024

difficulties in tracking resources and activities for assessments [4].

Therefore, in online education systems, student participation is a crucial element in the success of the course. Although a virtual learning environment with a certification of achievement is known as a MOOC, the online learning platform is assumed to be more of a hybrid [5]. Student satisfaction and the calibre of the educational experience are mainly related to student participation. To reduce the dropout rate, it is of great importance to realize how students interact with the interests of continuing professional education.

The improved learning environment helps to promote self-control and motivation, which helps students stay on task and perform better. Understanding students, learning, and interacting with virtual learning environments are essential to creating better learning environments [6]. The perfect learning models and the customization of learning environments depend greatly on student participation.

Research groups use learning analytics to predict student performance due to the evolution of educational data. The current study uses datasets from the Open University Learning Analytics dataset collected by Flea at the Open University in the UK to predict student participation.

Compared to MOOC platforms, virtual learning environments (VLEs) enable instructors to monitor student actions and assessments because they were designed to move passive learners toward active learning. As a result of the VLE's recording of student interaction, the instructor can better understand how students behave in these learning environments. The data of these learning platforms have made it possible to make decisions that are consistent with the data, and the clickstream data of these platforms can even predict student engagement at an early stage. The data set includes information on demographics and student interactions with the VLE. It is laborious for anyone to use conventional approaches to examine and produce valuable information because the data is kept in the Open University's extensive data repository [7].

Research goal:

- This work used classification approaches, such as machine learning (ML) techniques, to construct a model to analyze student interaction, predict learning styles, and appropriate evaluation methods in VLE courses.
- Learning analytics frequently employs classification approaches. The aim is to investigate how VLE activities affect students' participation.
- The novel of this work is to create a machine learning model that can detect a learner's involvement in their studies at an early point of the course based on the number of clicks activity in the course.
- The main idea is to increase student involvement to help instructors develop appropriate strategies. In this work, we explore an ideal collection of variables that reliably predict the engagement of the learners.
- Based on the students' clicks on their learning activities, our suggested model can classify students with an average accuracy of 98 % and predict the best evaluation techniques for each student.

In this paper, we have organized our discussion into several sections. Section 2 elaborates on the learning style, while Section 3 lists machine learning techniques. In Section Four, we present some related work, followed by a discussion of student modelling in Section Five. Section Six outlines our research question, while Section Seven proposes a model and methodology. In Section Eight, the experimental results, and in Section Nine, we recommend an assessment method. Finally, we conclude with a summary of our findings and outline areas for future work.

2. Learning style

Since a few decades ago, personalization has been a feature of computing, and all current systems provide users with a personalized experience. From direct-programmed training and examinations to adaptable virtual environments, e-learning systems have significantly developed [8]. Learner personality, which are the traits of the learner that influence the learning process [9], is a more general term that includes learning varieties. In the context of learning, the learning style is a fascinating characteristic. The chosen method of employing one's capacity for learning is their learning style. Different psychologists have presented us with several learning style models.

There were four types of learners: activists, theorists, pragmatists, and reflectors. The learning modalities describe the various ways that students can learn. The four modalities that are typically considered are tactile (touching), kinesthetic (moving), auditory (hearing), and visual (seeing). Visual aid learners frequently learn by viewing things, whereas auditory learners love reading or listening. The tangible learns best when they touch items and experience their configuration and sensation, while the kinesthetic learns best by doing rather than merely watching [10].

Linda Silverman and Richard Felder created the Felder-Silverman Index of Learning Styles (ILS) in 1988. This model analyses learning styles and makes four-dimensional preference distinctions. visual/verbal, sensing/intuitive, active/reflective, and sequential/global. These four dimensions correspond to four psychological processes: input, perception, processing, and understanding.

We can implement the VAK (Visual, Auditory, Kinesthetic) mapping process with teachers in classrooms by detecting the students' learning styles by using various methods like surveys, questionnaires, and observations. We will apply this step first of the course teaching this part 1 then we will prepare the material and discuss it with the students in the classroom based on the learner style we determine.

For visual students, we will add posters, charts, graphs, and multimedia presentations to demonstrate concepts, processes, and relationships. For auditory students, we will add lectures, conversations, audio recordings, and oral presentations. Also, encourage students to engage in group discussions, debates, and presentations to improve their auditory learning experiences.

For kinesthetic students, we will provide hands-on learning experiences, such as experiments, demonstrations, simulations, and role-playing. Table 1 shows the relation between a list of tools and their learning style orientations (Visual, Auditory, Kinesthetic).

Teachers can develop inclusive learning environments that accommodate students' different learning styles and preferences by incorporating the VAK mapping process into their classroom instruction. This strategy encourages student involvement, motivation, and academic performance by personalizing instruction to individual needs and interests.

3. Machine learning

Prediction and classification involve machine learning models and algorithms. Machine learning is the convergence of statistics and computer science, in which machines learn to improve performance from previous experience, just as humans. The only difference is that

Table 1
Relation between a list of tools and their learning style orientations (Visual, Auditory, Kinesthetic).

Tool	Learning Style Category
posters	Visual
charts	
Graphs	
multimedia presentations	
lectures	Auditory
oral presentations	
audio recordings	
conversations	
experiments	Kinesthetic
demonstrations	
simulations	
role-playing	

computers learn from data, while humans learn through events and experiences. Machine learning includes three types of learning algorithms: supervised, unsupervised, and reinforcement.

Supervised learning occurs when data, along with labels, are sent to a computer as input. The machine learns patterns from it and tries to predict the label of the newly provided data. For example, the computer will estimate whether it will rain today or not based on temperature, humidity, precipitation, and other weather factors, after learning the data [11].

In unsupervised learning without labels, the computer must examine patterns in the data and classify them based on the properties of the data. For example, tweets are automatically tagged according to different topic categories [12].

Classification in machine learning is used to create forecasts after the machine has been educated from the data. There are many classification applications, from serious applications, such as disease prediction and traffic monitoring, to simple applications, such as identifying spam and games. A naive Bayes classifier is a set of classification methods based on Bayes theorem. The algorithms share common elements such that each pair of classified attributes is self-dependent on each other [13].

Support vector machine (SVM) is based on the idea of defining a hyperplane that best divides a data set when displayed on a graph into two classes. The K-Nearest Neighbors (KNN) algorithm uses "similarity" to predict the value of new data points, meaning that the new data point will be assigned a value based on its similarity.

Using AI in education to determine VAK learning styles can provide various advantages, such as personalized learning experiences, tailored instructional tactics, and increased student engagement. Here's how AI can recognize VAK learning patterns in education: AI algorithms may analyze a range of data, including student performance, interactions with learning materials, and assessment results.

AI can predict visual, auditory, and kinesthetic learning styles by analyzing patterns in student behaviour and learning outcomes. AI-powered interactive learning platforms can tailor information and activities to individual learning methods.

Using AI technology to detect VAK learning styles in education allows instructors to develop more personalized and engaging learning experiences that match the different needs of their students.

4. Related work

In [14], to verify the effect of engagement on pupil performance, machine learning (ML) algorithms were used in this study to detect little-engagement learners in social science classes at the Open University (OU)., The highest educational level, outcomes, assessment score and the number of clicks on virtual learning environment (VLE) activities, such as a glossary, homepage, forum, collaborate, content, resources, subpages, data plus and URL during the first-course assessment, were among the study's input variables. The degree of student participation in the various activities was the output variable [15]. Some ML algorithms were applied to the data set to predict low-engagement students. These techniques were used to create training models which were then compared for accuracy and kappa values. The results showed that the J48 algorithm consumed an accuracy of 88.5 %, the decision tree gained an accuracy of 85.9 %, JRIP had an accuracy of 83.27 %, and gradient-boosted classifiers had an accuracy of 86.43 %.

In [7] this study, data collected over multiple years are used to predict student participation in the initial stages of a virtual learning environment (VLE). A portion of the Open University Learning Analytics Dataset, published by the British Open University (OU), is given to make predictive models using a machine learning approach. The study's data set includes 7,775 undergraduates who assumed social science classes over numerous evaluation years. Experiments are carried out with a set of restricted characteristics to determine whether students are actively participating in their courses at a high or low level. The most critical factors in predictive analysis are the features that show how students engage with the VLE, their scores, and the outcomes. A reduced feature vector is created using these variables as the foundation. The linear regression model served as the baseline of the study. The Random Forest classification technique produced the best findings of the model, which were 95 % accuracy, 95 % exact, and 98 % appropriate. The click activities that serve as an efficient interface between students and the VLE are a subset of click activities that are pertinent to early prediction.

In [16], this study creates a robust prediction model is created and essential variables are identified that significantly impact student effects in an e-learning environment. The significant contribution consists of two parts: highlighting some experimental visions under the impact of a collection of factors using feature selection techniques and proposing a prediction model using the most important characteristics while using the K-fold cross-validation approach. Using student data from the Learning Management System, it is thoroughly examined how various variables affect model performance and how input and target output correlations relate to each other. The most common machine learning techniques are then evaluated against the suggested approach. The findings showed that while less engaged students in the course tend to interact less frequently, they perform much better when using the e-learning system [17]. Additionally, according to the study findings, some prediction systems, like the Random Forest method, have significant advantages in predicting student performance and can do so with an accuracy rate of up to 80 %. There is also a discussion of other student features that could be useful in the e-learning system.

In [18], this study collects VLE data and processes using many pre-processing algorithms, such as removing missing values, normalization, encoding, and outlier identification. Several machine learning (ML) classification algorithms were applied to our data, and each algorithm's performance was evaluated using cross-validation techniques and a variety of useful indicators. Metrics, including precision accuracy, recall, and AUC (Area Under Curve) scores, are used to assess the model's performance. The findings indicate that the CAT (Computerized Adaptive Testing) Boost model is more accurate than the others. When comparing this new model to earlier research, it performed better in all respects.

According to the Results section of this work, the CAT Boost model has an AUC score of 96.24 %, an accuracy of roughly 92.23 %, a precision of 94.40 % and a recall rate of 100 %. Overall performance was similar for the multilayer perceptron, random forest, and Boost predictive models. We tested our model to the AISAR (Artificial Intelligence-Based Student Assessment and Recommendation System) model and found that our results were more accurate, with a 94.64 % accuracy compared to the AISAR model's 91 % accuracy. Compared to our models, which had a recall of about 92 %, the AISAR model had only about 50 %.

5. Student modeling

It can be challenging to predict how students will behave in learning situations. Starting with raw OULAD data, the data preparation step produced a refined and condensed data set that was used to create a prediction model. The classification techniques used in this work are KNN, Random Forest, and Decision Tree. The adoption of these classification techniques led to their selection.

Researchers in educational data mining have successfully applied them and have obtained an extreme performance model utilizing many methods [19]. The decision tree is a classification approach that can be used to extract student behavior patterns from educational data. The Decision Tree methodology needs less data cleansing and is unaffected by missing values and outliers. The consequence of the logistic regression method decides the predicted likelihood of a collaborative exclusive event coming (to) based on multiple external inputs.

An ensemble classifier, such as Random Forest, combines the findings of various decision trees. According to KNN, related things are always close together, but Random Forest is slower than decision trees and requires a lot more computing power. This explains why different algorithms were picked to develop the predictive model. These algorithms have been developed by researchers using Python libraries. Pandas, NumPy, and Sklearn are the packages used. The student training dataset, which makes up about 80 % of the dataset, is available for the classifiers to learn from.

6. Research question

In this research, to predict student learning styles in a virtual learning environment, we compare various machine-learning approaches. Instructors can use the results for the prediction and accuracy of our model at an early stage in studying students for the course. Every student has many characteristics that make it difficult for the models to make compelling predictions. This research was used to find the best model to classify the learning style of the students using data from the VLE log. We will answer two questions:

RQ1: which is the best model for good performance to predict the learner style?

RQ2: Which is the best examination (assessment) model for students based on their learning style?

7. Methodology

The suggested methodology for advising students on a suitable learning style and evaluation strategy based on their learning preferences is described in this section. The architecture comprises three essential elements, as shown in Fig. 1. The first step is to use various mapping techniques to enter student data in the VAK form. Based on their data, students will be classified as interested in any of the three VAK learning modes in the second half. Fig. 1 explains how the representation splits into three main parts.

Various mapping techniques primarily map Students' data into the VAK form. Based on their data, the next component aims to categorize the students into one of the three (visual, auditory, or kinesthetic) VAK study methods. The third and final component forecasts the most effective assessment methods based on the learners' preferred learning styles.

The shift from a four-modality approach (tactile, kinesthetic, auditory, and visual) to a three-modality approach (kinesthetic, auditory, and visual) can be explained by shared traits and conceptual clarity. Here are some of the causes behind this transformation: Tactile and kinesthetic modalities frequently overlap in practice. Tactile learning uses the sense of touch and is closely related to kinesthetic learning, which involves physical motion and interaction. Many tactile-based activities include kinesthetic movement. As a result, merging both modalities into a single (kinesthetic) simplifies classification and eliminates redundancy.

7.1. Using the VAK learning style to map student learning activities

In [20], this stage entails applying a map between student data and the VAK form. The mapping process makes use of semantic technology. In the suggested model, the calculation of the semantic relationship with student input and VAK classifications makes use of Wordnet, a reputable source for linguistic comprehension, as well as the benefit of semantic similarity methods. As a result, the mapping procedure is broken down into two parts listed below. Building a corpus of VAK study methods for learners is the first stage.

The student's current activities are then mapped out. In this work, an example study of data from Open University is provided. However, it is essential to note that the mapping technique is typically applicable to all student activities.

7.1.1. Constructing corpus of VAK studying model

In [21] claims that beginning in the 1920s, psychologists and educational experts created the first VAK principles. The VAK model is now offered by elaborating on a user's preference regarding a particular text description. As a result, the tokens that stand for nouns are extracted from the VAK text and any unnecessary words are removed.

This procedure creates three vectors, each representing one of the three learning styles and listing the keywords that best describe it. Then, a learning-style dictionary was created. Following the tokenization procedure, WordNet was used to create a dictionary with synonyms and

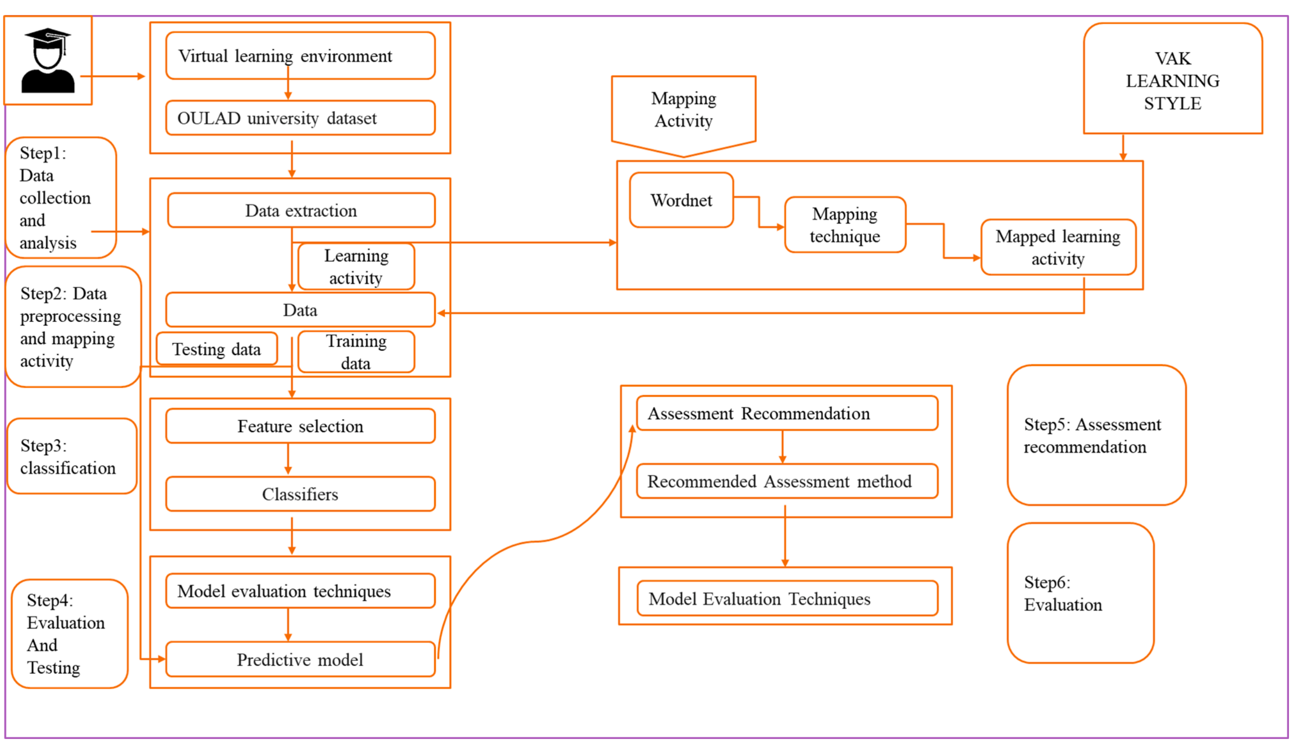

Fig. 1. Proposed Model.

related words for the newly created learning style keywords. In this step, WordNet capabilities to find synonyms and associated phrases are utilized. Constructed on the semantic similarity relationship involving models in an identical context, the get related words function gets associated terms. At the end of this step, the vector for the individual education style grouping would then be a template that includes all the alternative expressions and related words that best describe that type of study category.

This stage makes use of WordNet's ability to locate synonyms and related terms. The obtain-associated-words function returns phrases associated with the models and is built on the semantic similarity relationship between models in the same context. After this process, a matrix that includes all alternative expressions and related words that best describe that category of studying type would then serve as the vector for the individual education stylishness grouping.

7.1.2. Mapping process

With this strategy, educational activities for learners are mapped to VAK learning-type groups. There are two fundamental categories into which the mapping process can be split: full matching and semi-matching. This stage creates the mapping between educational actions in the dataset and the VAK studying type dictionary using semi- and full matching. From the individual study activities in the collection, the noun-representing tokens are first taken out. All learning activity results are kept as a vector that can be compared to the dictionary of education type classifications and the mapping.

7.1.2.1. Full matching category. In [22], the index terms in this category were sorted into a list of synonyms using careful matching based on WordNet synonym relations. In this mapping type, synonyms for learning style traits also include descriptions of learning activities. This establishes a strong correlation connecting the interest and the subsequent studying from 100 % of mapping, producing an exact match when using a dictionary-based mapping. Each token from the education activity's description vector is contrasted with the Python learning style dictionary throughout the comparison step.

There are two circumstances. Most of the learning activity tokens in the first are grouped under a single learning style category. This activity is suited to the pattern categorization study in this situation. As of the second in this case, this activity is linked explicitly to the pattern categorization study. If the educational activity tokens are present in two or more categories with an equal quantity of tokens and the activity is mapped to both categories in the second scenario.

7.1.2.2. Semi-matching category. In [22], Semantic similarity served as the basis for creating this category. Semantic similarity is a way of describing the link between two models that have similar attributes. Using a specific threshold, this category would then include ideas loosely related to the original idea. The topological parameters edge and information content (I.C.) inside the taxonomy enable us to measure or evaluate the semantic similarity of concepts in accordance with Tversky's cognitive psychology hypothesis at this level.

A few examples are the hierarchy created by "is-a" relations, the member/member-of relations with the taxonomic importance, and features based on the has-part connection. Therefore, if there is a "is-a," "has-part," or "member of" relationship between the activity and a learning style category at this level in the WordNet ontology, then the activity is related to that learning style category.

7.1.2.3. Semantic similarity measure calculations. In [22], if none of the learning activities perfectly fits into any of the categories of learning styles. The semantic similarity measure would be applied during the similarity mapping process to establish how closely two concepts within the same information content (I.C.) are related to one another. Depending on how much the learning activity resembles a particular VAK category, it will be put into that category in this case.

In other words, the result of the semantic similarity assessment will represent how closely a given learning activity reflects a given learning style. The learning activity will be mapped to the learning style group using the semantic similarity metric and a predetermined threshold.

We use a few different equations to figure out how closely two different terms are related to semantic similarity. We use I.C., path and depth, and a hybrid measure to determine their proximity. We also factor in the LCA of both senses and how deep they are in the taxonomy.

$$Similarity\,(C1,\ C2) = (2 \times N3) / (N1 + N2 + (2 \times N3)) \tag{1}$$

Where:

- N1, the minimal collective super notion of C1 and C2, is the length specified as several nodes on the path from C1 to C3.
- N2 represents the number of nodes along the path from C2 to C3 in terms of length.
- N3 represents the hierarchy's overall depth and acts as a scaling factor.

Wu and Palmer Similarity (Wu and Palmer) Measure (WPM) is used in the proposed Semantic Sentence (SST) similarity calculation methodology [23]. The WPM quantifies the similarity between two words based on a score. It is based on the complexity of the two senses within the taxonomy, as well as the complexity of their (LCS). (Farthest specified ancestor node). Note that the shortest path connecting these two senses is not usually the least common subsumer (LCS), which is the deepest common ancestor within the taxonomy. The node closest to these two perceptions is not it. When there are multiple candidates for the LCS (the LCS has more than one path to the root), the longest path is used to determine similarity. Wu and Palmer calculated the degree of semantic similarity between concepts C1 and C2 using Eq. (1).

The suggested approach uses the Wu- and Palmer-based WordNet functions to assess how much two concepts within the same information content (I.C.) resemble each other. See the following section for further information on the Wu and Palmer-based WordNet function, which determines the degree of similarity between two words. The similarity score in our system indicates how closely connected the activity description and the learning style keywords are to one another.

The similarity score was obtained using the formula:

Wordnet.wup_similratiyScore("synset('keywords.pos.nn')", "words.pos.nn").

The OULAD datasets have many activities Table 2 lists some activities on the VLE logs datasets and its description.

Following of the mapping technique, each learning activity that the learner starts is assigned to one or more of the VAK learning styles based on the activity belonging to full matching or semi-matching. This mapping identifies a particular VAK learning style group in which this educational activity fits.

We found that the three activities shared_subpages, subpage, and ou_elluminate have a semi-matching in our coding we selected the highest percentage as shown in the results in Table 3. As a result, it

Table 2
Activity description.

Learning Activity	Activity Description
Forming	Use of discussion forums, where learners and teachers can engage in conversations and facilitate the exchange of ideas.
Dualpane	Split the display into two sections (left and right) to allow learners to choose the type of exercise or material they want to add.
Ouwiki	The educational wiki is a web page in HTML where learners and teachers can contribute remarks or alter material to improve interaction and cooperation in the learning environment.
Ou collaborates	video conferencing
Ou elluminate	Audio-only conferencing

Table 3
Result of mapping between VAK learning style categories and learning activities.

Activity Name	Visual Score	Auditory Score	Kienthatic Score
glossary	0	0	1
dualPane	0	0	1
dataplus	0	0	1
glossary	0	0	1
shared_subpages	0.66	0.63	0.83
subpage	0.66	0.53	0.71
questionnaire	1	0	0
ou_collaborate	1	0	0
ou_content	1	0	0
repeat_activity	1	0	0
resource	1	0	0
URL	1	0	0
HTML activity	0	1	0
ou_wiki	0	1	0
forming	0	1	0
htmlactivity	0	1	0
ou_elluminate	0.72	0.8	0.6

allows for the modelling of students based on their VAK (variables associated with learning).

8. Experiment result

In this section, we will show the content of the datasets and the preprocessing steps to the data by our model to prepare the datasets with the mapping process then discuss the experimental results.

8.1. Data collection

Studies evaluated the efficacy of the suggested strategy using the Open University Learning Analytics Dataset (OU, 2017). The data set is unique because it includes stream data that have been aggregated from student interactions with Moodle-based learning activities in a virtual learning environment (VLE). This analyzes pupils' behaviour as it appears in their activities. Twenty other activity types were also used to better categorize user engagement with the VLE. Various behaviours, such as downloading or watching lectures, reading course materials, or taking tests, were described by each activity category. The data set contains data on 7 courses, 32.593 students, test scores, and VLE interactions (10.655.280 entries per day). It also has the results of their test quizzes and homework.

- Student demographics: includes information on the student's age, gender, and educational level. Course Enrollment: Information on the courses that students have enrolled in, such as course codes, titles, and start/end dates.
- Learning Activities: Records of how students interact with learning materials, such as accessing course information, submitting assignments, participating in conversations, and using multimedia tools.
- Assessment Data: Details regarding how students performed on quizzes, exams, assignments, and other assessments.
- Clickstream Data: Student information clicks activities in the online learning environment, such as navigation patterns, time spent on different pages, and interactions with course tools and features.
- Style data can be recorded as attributes in individual student profiles or independent records related to student identifiers.
- Style data could comprise categorical variables that reflect students' preferred learning styles. Analysing user clicks based on log data might reveal insights into students' learning behaviours and preferences, potentially indicating their preferred learning approaches.
- Clickstream data can be analyzed to uncover interaction patterns associated with learning styles. For example, students who prefer visual learning may spend more time studying multimedia presentations or diagrams, whereas auditory learners may spend more time listening to audio lectures or debates.

The OULAD dataset offers a variety of educational data, such as user click behaviours, style information, and other essential characteristics. Analysing the relationship between user click actions and style data can provide useful insights into students' learning preferences and behaviours, influencing instructional design, personalized learning activities, and student support techniques in online education.

The original datasets didn't contain a learning style. We make a preprocessing step to enhance the datasets based on the student's activity on VLE to map between the student's activity and VAK learning style this shown in the previous section.

As an example, ou_content, questionnaire and repeat_activity belong to visual and htmlactivity and forming belongs to Auditory.

8.2. Preprocessing of data

Seven relational tables make up the OULAD. Any machine learning algorithm cannot directly receive these data as input. Firstly, we converted the relational tables to a dataset that can be processed by machine learning. The input features for this study are collected from many tables and combined into a recorded dataset.

A portion of the dataset with The code_presentations 2013 J and 2014 J is chosen for additional testing. The generated dataset rows correspond to each student's ID, and their columns to each student's features. The missing attribute values are filled with zeros if the student has not logged in to such activities. It does not include students who dropped out of classes to obtain more precise and insightful results.

The data set that resulted from such interactions only contains 22,593 students, their test results, and log data of their VLE connections, which are represented by daily summaries of student clicks. We removed duplicate data and integrated the findings of the mapping phase with the original data.

When the attributes are examined, it is discovered that several of them are detrimental to this study because they do not have a bearing on our objective. Early preprocessing eliminated these unnecessary features, ultimately making the development of the predictive model easier.

The current research aims to discover the variables that affect how learners' styles are predicted in a virtual learning environment. An extremely involved student will dynamically participate in the learning environment, according to studies. Based on student participation, the various types of activities in OULAD are quantified. The attribute total_number_of_VLE_clicks, which tracks the learner's overall VLE access, oversees that. However, it is possible that the student will click on interfaces other than the VLE throughout the session. Therefore, the total of clicks alone cannot be used to determine a learner's learning preferences, so we will do a series of steps to get the best shape of the data that helps us to achieve our goal.

1. We will merge the learning styles with the activity type and student_VLE_Data, which contains the number of clicks of each activity and the id_site, Table 4 shows the relations between tools and activity (clicks) and their learning style orientations ((Visual, Auditory, Kinesthetic) which identifies the number for the VLE material. After doing that, we will select the max value of the sum of clicks that have been done based on the id_site. We will get all the data features that we need to start working, as shown in Fig. 2. The activity type is the activity done by the students on Moodle, and the learning style is the results of his preferred study type of data to get one of the VAK students learning.
2. Following the data merger, we encoded it since this method enables us to turn categorical variables into numerical values that can be quickly fitted to a machine learning model. After encoding the data, we noticed a large gap between the values in the data, so we

Table 4
The relations between tools and activity (clicks) and their learning style orientations (visual, auditory, kinesthetic).

Activity Type	Learning Style	Number of clicks
ou_collaborate	Visual	1217
ou_content		16,661
questionnaire		314
repeat_activity		8
Resource		46,480
URL		4398
htmlactivity	Auditory	45
Forming		9317
ou_elluminate		385
ou_wiki		3018
shared_subpages	Kinesthetic	13
Subpage		16,661
Glossary		1456
dualPane		258
Dataplus		512

	id_site	code_module	code_presentation	id_student	date	sum_click	activity_type	study_type
0	526721	FFF	2013B	2691740	240	4098	homepage	Visual
1	526735	FFF	2013B	2691740	240	20	forumng	Auditory
2	526737	FFF	2013B	2691740	240	110	forumng	Auditory
3	526738	FFF	2013B	2691740	240	82	forumng	Auditory
4	526739	FFF	2013B	2691740	240	46	forumng	Auditory
...	...	...	...	...	...	...	...	...
6263	1042376	CCC	2014J	2691267	262	10	resource	Visual
6264	1046237	CCC	2014J	2691267	261	7	resource	Visual
6265	1046812	DDD	2014J	2432655	242	170	glossary	Kinesthetic
6266	1046866	CCC	2014J	2668106	266	17	resource	Visual
6267	1049562	CCC	2014J	2326041	240	2	oucontent	Visual

6268 rows × 8 columns

Fig. 2. Data after the first step (merging).

performed data scaling shown in Fig. 3 to facilitate model detection and understanding of the problem. Scaling is a technique of reducing distances to bring them closer together. When we run machine learning algorithms on the dataset, scaling the data is one of the preprocessing procedures. Most supervised and unsupervised learning methods, as we all know, base their inferences on the data sets that are applied to them, and many times, algorithms calculate the distance between data points to draw more accurate inferences from the data.

As we all know, many supervised and unsupervised learning methods rely on their conclusions on the data sets that are applied to them, and frequently algorithms measure the distance between data points to conclude the more accurate data.

	id_site	code_module	code_presentation	id_student	date	sum_click	activity_type	study_type
0	-1	1	-1	0	0	21	-2	2
1	-1	1	-1	0	0	0	-2	0
2	-1	1	-1	0	0	0	-2	0
3	-1	1	-1	0	0	0	-2	0
4	-1	1	-1	0	0	0	-2	0
...	...	...	...	...	...	...	...	...
6263	2	0	1	0	0	0	0	2
6264	2	0	1	0	0	0	0	2
6265	2	0	1	0	0	0	-2	1
6266	2	0	1	0	0	0	0	2
6267	2	0	1	0	0	0	-1	2

6268 rows × 8 columns

Fig. 3. Data after encoding and scaling steps.

8.3. Result and discussion

It is challenging to see how students will learn in different contexts. The data preparation process, which started with raw OULAD, produced a refined and condensed data set that was used to create a prediction model. In this study, the support vector machine, K-nearest neighbour, Random Forest and Logistic regression were used as classification methods.

Both linear and non-linear solutions are supported by SVM. Additionally, KNN outperforms linear regression when the signal-to-noise ratio SNR of the data is high, and it is assumed that the objects are always close to each other. Decision trees are less effective and accurate than Random Forest. This clarifies the assortment of calculations utilized for predictive demonstration creation since it requires training data and runs slower than decision trees. Where do we typically use Python libraries in these algorithms? Sklearn, NumPy, and Pandas are the packages used. We ran two experiments using 90 % to 80 % of the data for each identified student training data set. The data are tested using the remaining 10 % and 20 % of the data, and all three assessment metrics are used to assess them. Accuracy, recall, and outcome. To respond to research questions, experiments are carried out.

8.3.1. Performance-based classification algorithm

This Subsection tries to address research question one, "Which classifier offers the best performance for predicting a student's learner style in the VLE?" Predictive models have been developed using a variety of machine-learning approaches and tests to offer a solution to this research problem. The predicted variable in this scenario is each student's preferred learning method, and the characteristics were VLE activity clicks in a VLE cycle. A 10 K-fold approval approach is used to decide the adequacy of the categorization models. Tables 5 and 6, as well as Figs. 4-7, include a list of evaluation metrics for the various models.

The random forest calculation crosses the other classification strategies in this test setup. Unlike the other models, the Random Forest show is a gathering model; each tree within the Random Forest produces comes about, and the course with the most votes decides the model's estimate.

The accuracy of the Random Forest classification technique is 98 %. The classifier has a recall of 99 % and a precision of 97 %. The classifier's recall performance is essential for solving the current issue of identifying students' learning preferences.

We have made a few important classifications. On the off chance that the learning strategies of most understudies can be recognized at an early stage, at that point, the teaching/learning preparation will be an advantage. The random forest classifier is so popular that 98 % of the overall students were predicted by their learning styles within the current exploratory setting. The execution of a random forest is a marker that the highlight under study has great prescient control. The Distribution of student learner style of OULAD Dataset results in Fig. 8 the highest percentage was 75.90 % for the visual learner style.

We applied our new algorithm in this research on our students in our faculty on general course for the first year which all students registered for this course (580 students).

We have a Moodle learning management system in our faculty, firstly we prepare the material for this course to let users engaged on the

Table 5
Shows the algorithm performance using the training model of 90 and the test model of 10.

Algorithm	Accuracy	Precision	Recall	F1-Score
Support vector machine (SVM)	**84 %**	**81 %**	**99 %**	**89 %**
K-Nearest Neighbor (KNN)	**83 %**	**80 %**	**99 %**	**89 %**
Random Forest (RFR)	**98 %**	**97 %**	**99 %**	**99 %**
Logistic regression	**86 %**	**91 %**	**82 %**	**83 %**

Table 6
The result of algorithm performance when training mode 80 and test model 20.

Algorithm	Accuracy	Precision	Recall	F1-Score
Support vector machine (SVM)	**96 %**	**81 %**	**99 %**	**89 %**
K-Nearest Neighbor (KNN)	**73 %**	**87 %**	**91 %**	**89 %**
Random Forest (RFR)	**95 %**	**86 %**	**91 %**	**90 %**
Logistic regression	**87 %**	**91 %**	**83 %**	**81 %**

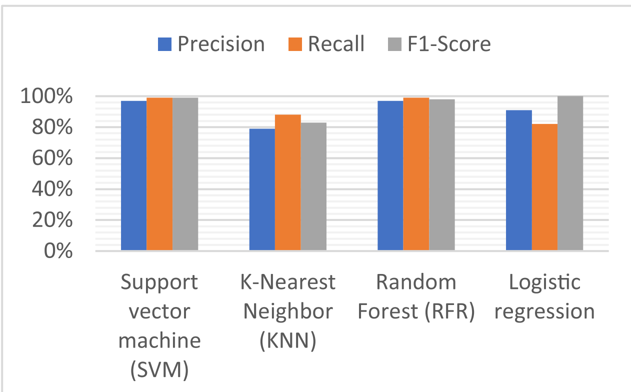

Fig. 4. Comparison between the performance of each algorithm.

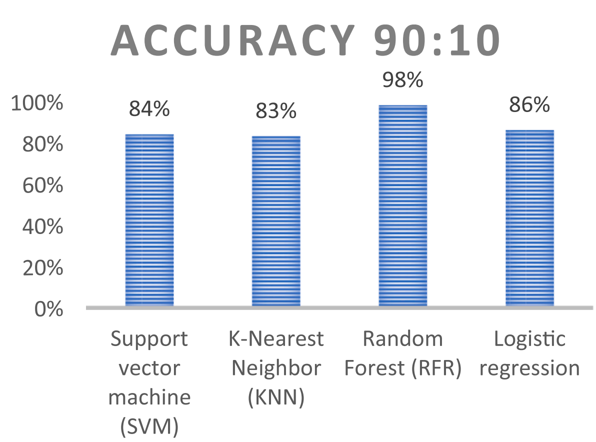

Fig. 5. Comparison of the performance of each algorithm based on precision.

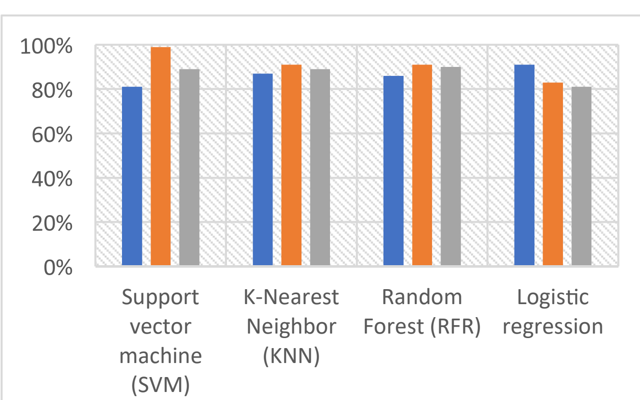

Fig. 6. Comparison of the performance of each algorithm.

system with VAK learning styles, secondly, all students registered on the site with no learning style we select the suitable style after following their activity on the system after a period time based on logs and number

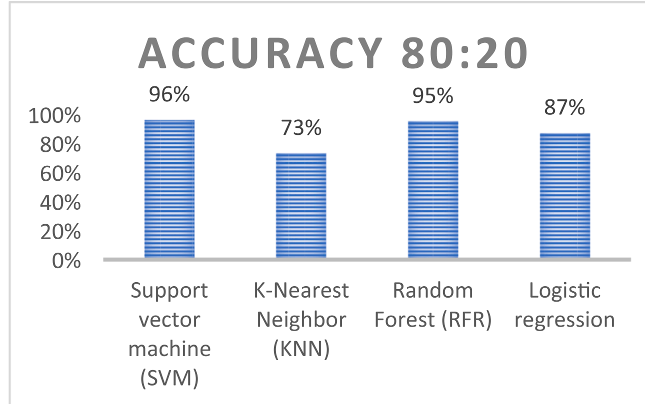

Fig. 7. Comparison between the performance of each algorithm based on accuracy.

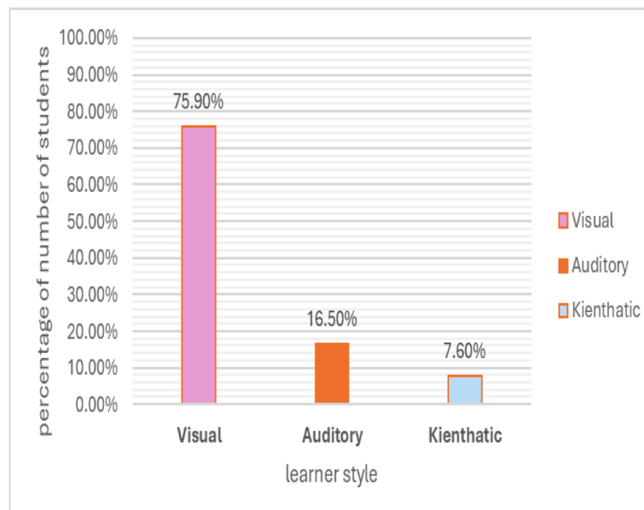

Fig. 8. Distribution of student learner style of OULAD Dataset.

of clicks activity and recommend the material based on their learning styles selected from the system and apply different assessment methods to find the final results.

The distribution of student learner style in faculty shown in Fig. 9 that visual has the highest learner style with 60 % of the total number of students and Table 7 shows the accuracy of different machine learning models to measure the effectiveness of the new algorithm and accuracy. We split the data into 90 % training and 10 % testing.

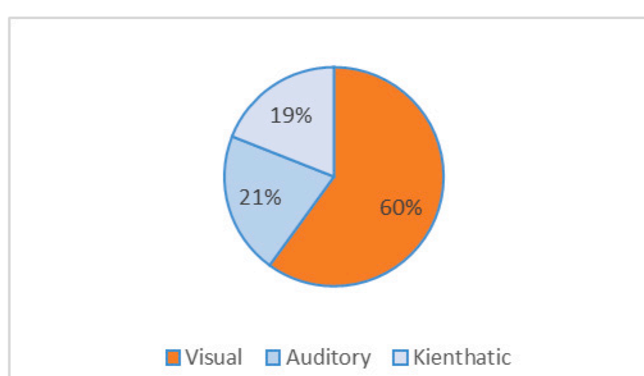

Fig. 9. Distribution of student learner style in faculty.

Table 7
Shows the algorithm performance using the training model on the faculty students.

Algorithm	Accuracy	Precision	Recall	F1-Score
Support vector machine (SVM)	87 %	92 %	97 %	94 %
K-Nearest Neighbor (KNN)	95 %	94 %	99 %	96 %
Random Forest (RFR)	96 %	98 %	98 %	97 %
Logistic regression	92 %	98 %	98 %	97 %

9. Recommended assessment method

Each type of VAK learning style is associated with evaluation techniques that students can improve. In this phase, an assessment approach type is suggested based on the learned style of the learner. he states that the VAK learning style assessment method types were applied, with auditory learners favouring subjective evaluation techniques. Visual learners embrace subjective assessment techniques instead of kinesthetic learners who favour objective assessment techniques. Long essay questions and case studies are two assessment methods that might be classified as tutor-marked assessment (TMA) types but do not have a definite right or wrong answer.

Objective assessment includes various assessment methods even if computer-marked assessments (CMAs) of multiple-choice, true-or-false, and other forms only need one correct answer. Based on each student's previously identified learning style and performance in each course, we will choose the most appropriate assessment techniques for them shown in Table 8. The VLE_data, Study_type, Assessment_data, and Student Assessment Data were combined in the preprocessing step to help us detect suitable assessment methods.

Performance-based classification algorithm This section seeks a solution to the second RQ2 by focusing on which classifier performs the best in predicting a student's appropriate evaluation technique in the VLE. Predictive models have been developed using a variety of machine-learning approaches and tests to offer a solution to this research problem. The anticipated variable in this scenario is each student's assessment technique, and the features included VILE VLE_data, Studey_type, assessment_data, and student assessment data in a VLE cycle. The evaluation metrics for the different models are in Fig. 10.

Content learning: The primary purpose of education is to give students an in-depth understanding of the material being taught. Mastery of knowledge means that students can understand and apply concepts regardless of how they were taught.

Full mastery of knowledge involves the ability to learn independently, regardless of the learning mode. Educators should seek to establish learning environments that promote deep comprehension, critical thinking, and the ability to apply knowledge in a variety of circumstances, regardless of learning style.

Learners may choose the method of assessment as a suggestion to their teachers, which may influence the performance rating for the final results of the test.

10. Conclusion and future work

Using student models and learning-type predictions, this research created an adaptable approach for evaluation ideas based on detected learning styles. The following understudy learning exercises that use both semantic and machine learning innovation allowed the identification of the VAK learning style of the understudy.

Tests were conducted utilizing the Open College Learning Analytics Dataset to show the system's capacity to recognize student learning preferences and recommend assessment techniques. The productivity of the proposal framework is surveyed using precision, accuracy, recall, and score. According to studies, the system can determine a student's learning preferences with an accuracy of more than 98 %, while the F1 test results and the evaluation recommendation accuracy both average 98 %.

Table 8
Result of the Prediction of the Student Assessment Method.

Algorithm	Accuracy	Precision	Recall	F1-Score
Support vector machine (SVM)	96 %	97 %	96 %	97 %
K-Nearest Neighbor (KNN)	98 %	99 %	98next %	98 %
Random Forest (RFR)	96 %	97 %	96 %	97 %
Logistic regression	95 %	96 %	97 %	97 %

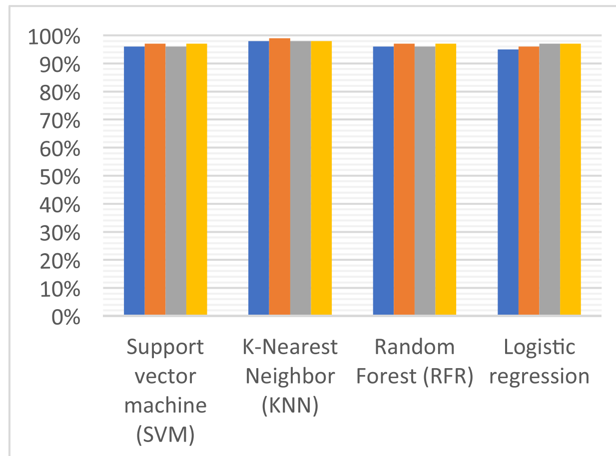

Fig. 10. Result of the students' assessment method prediction.

We will make quick alterations to learning aspects, such as learning contents, evaluation procedures, and more, in our subsequent work while considering the student's current learning activities.

In the future, we can propose hybrid learning techniques, and establishing engaging and effective models for self-regulated learning requires capturing the most recent technological breakthroughs and educational research.

Investigate the use of emerging technologies such as artificial intelligence, augmented reality, virtual reality, and immersive simulations to develop engaging and interactive learning environments.

Using AI-powered adaptive learning platforms and intelligent tutoring systems to provide students with tailored feedback, recommendations, and support based on their unique learning requirements and preferences.

CRediT authorship contribution statement

Ahmed Rashad Sayed: Methodology, Software, Data curation, Validation. **Mohamed Helmy Khafagy:** Data curation, Investigation, Validation. **Mostafa Ali:** Validation, Visualization, Investigation. **Marwa Hussien Mohamed:** Conceptualization, Methodology, Validation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Essa SG, Celik T, Human-Hendricks NE. Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: a systematic literature review. IEEE Access 2023;11:48392–409. https://doi.org/10.1109/ACCESS.2023.3276439.

[2] Muñoz JLR, et al. Systematic review of adaptive learning technology for learning in higher education. Eurasian J Educ Res 2022;2022(98):221–33. https://doi.org/10.14689/ejer.2022.98.014.
[3] Hussain M, Zhu W, Zhang W, Abidi SMR. Student engagement predictions in an e-learning system and their impact on student course assessment scores. Comput Intell Neurosci 2018;2018. https://doi.org/10.1155/2018/6347186.
[4] N. Holmes, Engaging with assessment: increasing student engagement through continuous assessment. [Online]. Available: http://shura.shu.ac.uk/11578/.
[5] Amantha Kumar J, Al-Samarraie H. An investigation of novice pre-university students' views towards MOOCs: the case of Malaysia. Reference Librarian Apr. 2019;60(2):134–47. https://doi.org/10.1080/02763877.2019.1572572.
[6] Hassan SU, Waheed H, Aljohani NR, Ali M, Ventura S, Herrera F. Virtual learning environment to predict withdrawal by leveraging deep learning. Int J Intell Syst Aug. 2019;34(8):1935–52. https://doi.org/10.1002/int.22129.
[7] N.R. Aljohani, A. Fayoumi, S.U. Hassan, Predicting at-risk students using clickstream data in the virtual learning environment, Sustainability (Switzerland), vol. 11, no. 24, Dec. 2019, doi: 10.3390/su11247238.
[8] Ayar Z. The systematic review of studies about web-based learning in language teaching context. e-Kafkas Eğitim Araştırmaları Dergisi Aug. 2021;8(2):123–36. https://doi.org/10.30900/kafkasegt.957362.
[9] Del Valle R, Duffy TM. Online learning: Learner characteristics and their approaches to managing learning. Instr Sci Mar. 2009;37(2):129–49. https://doi.org/10.1007/s11251-007-9039-0.
[10] N. Alajlani, M. Crabb, I. Murray, A systematic review in understanding stakeholders' role in developing adaptive learning systems, Journal of Computers in Education, Jun. 2023, doi: 10.1007/s40692-023-00283-x.
[11] A.M. Ali, A. Abdelhafeez, DeepHAR-Net: a novel machine intelligence approach for human activity recognition from inertial sensors 2, doi: 10.61185/SMIJ.
[12] Kotsiantis S. Supervised machine learning: a review of classification techniques. Informatica (slovenia) 2007;31:249–68.
[13] Ikawati Y, Al Rasyid MUH, Winarno I. Student behavior analysis to predict Learning styles based felder silverman model using ensemble tree method. EMITTER Int J Eng Technol Mar. 2021;9(1):92–106. https://doi.org/10.24003/emitter.v9i1.590.
[14] A.M. Ali, A. Abdelhafeez, Breaking the silence: convolutional neural networks for sign 1 language recognition in the deaf community, doi: 10.61185/SMIJ.
[15] Alzahrani NA, Abdullah MA. Student engagement effectiveness in E-learning system. Biosci Biotechnol Res Commun Feb. 2019;12(1):208–18. https://doi.org/10.21786/bbrc/12.1/24.
[16] A. Abdelhafeez, A.S. Aziz, N.A. Khalil, Building a sustainable social feedback loop: a machine intel-1 ligence approach for twitter opinion mining, doi: 10.61185/SMIJ.
[17] N. Alruwais, M. Zakariah, Student-engagement detection in classroom using machine learning algorithm, Electronics (Switzerland), vol. 12, no. 3, Feb. 2023, doi: 10.3390/electronics12030731.
[18] Zhang Y, Yun Y, An R, Cui J, Dai H, Shang X. Educational data mining techniques for student performance prediction: method review and comparison analysis. Front Psychol Dec. 2021;12. https://doi.org/10.3389/fpsyg.2021.698490.
[19] Zhu X, Yang X, Huang Y, Guo Q, Zhang B. Measuring similarity and relatedness using multiple semantic relations in WordNet. Knowl Inf Syst Apr. 2020;62(4):1539–69. https://doi.org/10.1007/s10115-019-01387-6.
[20] W. Samanthi, H. Sunesh. (2017). Relationship among students' learning styles, assessment methods and students' performances. Third International Conference on Education and Distance Learning – 28th April 2017, Sri Lanka.
[21] D. Soergel, WordNet. An Electronic Lexical Database New ways of learning for a changed population of learners preparing for a changing world of work. Technology-enabled knowledge support makes it possible View project DL.org-Coordination Action on Digital Library Interoperability, Best Practices, and Modelling Foundations View project," 1998. [Online]. Available: www.let.uva.nl/~ewn.
[22] D.S. Sharma, R. Sangal, J.D. Pawar, M.A. Tayal, M.M. Raghuwanshi, L. Malik, Word net based Method for Determining Semantic Sentence Similarity through various Word Senses, NLPAI, 2014.
[23] Wu, Zhibiao & Palmer, Martha. (2002). Verbs Semantics and Lexical Selection. Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics. 133-138. 10.3115/981732.981751.