Open Access

Unlocking teachers' potential: MOOCLS, a visualization tool for enhancing MOOC teaching



Brahim Hmedna^{1*}, Aicha Bakki², Ali El Mezouary³ and Omar Baz³

*Correspondence: brahim.hmedna@edu.uiz.ac.ma

¹ IMIS Laboratory, FSA-AM | ENCGA, Ibn Zohr University, Agadir, Morocco ² Univ Angers, LERIA, SFR MATHSTIC, 49000 Angers, France ³ IRF-SIC Laboratory, FSA | ESTA, Agadir, Morocco

Abstract

Massive Open Online Courses (MOOCs) are revolutionizing online education and have become a popular teaching platform. However, traditional MOOCs often overlook learners' individual needs and preferences when designing learning materials and activities, resulting in suboptimal learning experiences. To address this issue, this paper proposes an approach to identify learners' preferences for different learning styles by analyzing their traces in MOOC environments. The Felder–Silverman Learning Style Model is adopted as it is one of the most widely used models in technologyenhanced learning. This research focuses on developing a reliable predictive model that can accurately identify learning styles. Based on insights gained from our model implementation, we propose MOOCLS (MOOC Learning Styles), an intuitive visualization tool. MOOCLS can help teachers and instructional designers to gain significant insight into the diversity of learning styles within their MOOCs. This will allow them to design activities and content that better support the learning styles of their learners, which can lead to higher learning engagement, improved performance, and reduction in time to learn.

Keywords: MOOCs, Learning styles, Learning behaviors, Machine learning, Visualization tool

Introduction

Over the past decade, Massive open online courses (MOOCs) have offered an innovative way of providing open education through distance learning (Yousef et al., 2015). These courses can enhance the autonomy of learners and allow institutions to share high-quality educational resources (Brown, 2013). Despite the advantages and benefits that have been highlighted by participants in these environments, including researchers, students, event teachers and instructors, there are some criticisms of MOOCs that still need to be considered and managed, in order to improve them as open model of learning. These limitations relate to various aspects of courses, such as teaching and learning methods, the learning content, and the varying needs of learners, among others (Fasihuddin et al., 2014).



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativeCommons.org/licenses/by/4.0/.

One of the main limitations of MOOCs is their static organizational structure (Durand et al., 2011), in which all learners are generally provided with the same teaching method, regardless of their individual learning styles, knowledge levels, and personal preferences. A second limitation is due to the massive and open nature of MOOCs that leads to a high degree of learner diversity within these contexts (Chuang & Ho, 2016; Koller et al., 2013).

Many research has been conducted to address this issue by exploring ways to personalize and adapt learning experiences in MOOCs (Assami et al., 2018; El Mawas et al., 2018; Williams et al., 2017). The integration of recommender systems further enhances this personalization by providing learners with targeted recommendations of courses, resources, and activities based on their interests and previous learning experiences. These combined efforts may contribute to a more learner-centric and effective MOOC ecosystem.

However, issues related to the massiveness and openness of MOOCs have raised further relevant research concerns and present ongoing research challenges. Indeed, understanding the learning diversity among MOOC learners and framing effective learning strategies based on the spectrum of pedagogical approaches of learning styles (Felder & Brent, 2005) is undoubtedly challenging. This is due to the massive numbers of learners involved which makes it difficult for teaching teams to observe the behavior of each learner through direct, face-to-face, interactions.

Overcoming these obstacles requires innovative solutions that leverage technology and data analysis to gain a deeper understanding of learners' needs and preferences, paving the way for more effective and personalized learning experiences in MOOCs. In this regard, several research studies have been carried out of various aspects of MOOCs, such as the motivation, intentions, self-regulation, competence and learning styles of the students (Bakki et al., 2015, 2016; Graf & Liu, 2009; Kizilcec et al., 2016; Koller et al., 2013). Among these, the current work focuses on the exploration of learning styles as a way to design an effective learning environment that is effectively tailored to the needs and characteristics of each learner, rather than delivering the same resources to all learners in the same way (Blagojević & Milosević, 2013).We believe that, by recognizing and accommodating diverse learning styles, MOOCs can create a more inclusive and engaging learning experience that maximizes the potential of each learner.

A learning style refers to an individual learner's preferred approach to acquire new information and ideas, as well as the confidence in processing and using this information (Coffield et al., 2004). It is a holistic model that provides a wide range of directions for learning that cater to individual preferences. This model recognizes and makes the same teaching strategy can be beloved by some learners and hated by others (Oxford, 2003). Consequently, understanding and accommodating various learning styles is essential for creating an inclusive and effective learning environment that can engage a wide range of learners. Indeed, embracing different directions for learning, may allow educators to customize their instructional methods to suit the preferences and strengths of each individual, enhancing the overall learning experience.

The main goal of this paper is to design a model that can automatically identify the learning styles of learners by applying machine-learning algorithms to the large collections of click logs associated with MOOCs. This work makes the following contributions:

(i) we define a list of features associated with each learning style; (ii) we investigate the most appropriate clustering algorithms for partitioning learners into homogenous groups according to their learning styles; (iii) we investigate the most appropriate classification algorithms for predicting learning styles; and (iv) we present MOOCLS, a visualization tool that aims to help teachers and instructors to design teaching materials in the best possible way and to ensure effective learning for all types of learners.

The remaining sections of the paper are structured as follows: Section "Literature review" provides a brief overview of learning styles and provides detailed information on approaches to the detection of learning styles. Section "Decoding Learners' preferences: methodological approach to identifying learning styles in MOOCs" outlines our proposed methodology, including data pre-processing, unsupervised modeling, aggregation and supervised modeling. Section "From events to visual insights: leveraging MOOCLS for learning style visualization"4 gives an overview of MOOCLS, a tool for the visualization of learning styles, and presents the main results of an experimental study that aims to evaluate the utility and usability of this tool. Finally, section "Discussion" offers the main conclusions drawn from the research conducted.

Literature review

Learning styles

The term "style" began to be used in educational psychology in 1950 (Cid et al., 2018). It refers to the behavioral traits adopted by individuals in a specific context that distinguish them from others (Fischer & Fischer, 1979). In the educational context, learners do not all perceive a learning situation in the same way, each having a personal style for processing and organizing information. This concept is denoted as "learning styles" in both pedagogy and psychology (Felder, 1996).

Learning styles are a category under a broader of differentiated pedagogy, first introduced by Herb Thelen in 1954 (Petty, 2004). In the light of academic discussion, there's often confusion between the terms 'learning style', 'learning strategy', and 'cognitive style'. While they may seem interchangeable, they represent distinct concepts. Cognitive styles are the preferred, consistent, individual characteristics in organizing and processing information (Ford & Chen, 2001; Messick, 1984). In contrast, learning strategies are set of tactics that learners employ to control their learning process. It's a set of action taken by learners in regulating their learning process. According to Slack and Norwich, (2007), learning strategies are more varied since learners might choose a distinct strategy for each task. Meanwhile, learning styles are more stable and can be viewed as personality traits. It is essential to emphasize that learners' learning styles should not be treated as a static and stable concept (Kolb, 1976). Rather, they ought to be recognized as a dynamic characteristic that can evolve over time.

The concept has gained significant attention in academic circles, leading to a plethora of proposed definitions. However, a universally accepted definition of learning style is still lacking in the literature. Chevrier et al., (2000) delineated the definitions into three main frameworks. Consequently, learning styles can be understood as:

(i) Learning-centered: In the light of this, we can reference the definition proposed by Keefe (1979) " *Learning styles refer to cognitive, affective and psychological behaviors that indicate how learners perceive, interact with and respond to the learning*

environment". According to the Dunn et al. (1979) "learning style is the way each person begins to concentrate on, process, internalize, and retain new and difficult academic information".

- (ii) Cognitive-centered: In this context, Reinert (1976) states that "an individual's learning style is the way in which that person is programmed to learn most effectively, i.e., to receive, understand, remember, and be able to use new information.". Oxford (2001) believe that learning styles is "the general approaches that learners prefer to employ when acquiring knowledge, learning a new language, or solving problems".
- (iii) Personality-centered: Within this category, Barbe et al. (1988) states that "learning style describes an individual's relative ability to perform an academic task according to the main perceptual modalities". According to the Grasha (1984)"Learning styles are personal dispositions that influence a student's ability to acquire information, to interact with peers and teachers, and to participate in a learning experience".

This variety of definitions and approaches has led to the development of multiple models for identifying learning styles over the years. Coffield et al. (2004) identified 71 different learning styles models in the literature, reflecting the diversity of approaches. Among these models, the Kolb's learning styles inventory (Kolb, 1976), Dunn and Dunn learning styles model (Dunn et al., 1979), and FSLSM (R. M. Felder & Silverman, 1988) are widely known and frequently utilized. These models offer valuable frameworks for understanding how learners process information and provide insights into their preferences and strengths.

Based on the extensive range of learning styles models identified by Coffield et al. (2004), the FSLSM (Felder-Silverman Learning Style Model) stands out as a prominent and widely utilized framework (Essa et al., 2023; Raleiras et al., 2022). The FSLSM was developed by Richard Felder and Linda Silverman in 1988. It was initially introduced as a model for engineering learners, to capture learner learning preferences with regards to perception, input, processing and understanding through four dimensions (Fig. 1).

Felder and Silverman (Felder & Silverman, 1988) defined learning styles based on answering the four following questions:

(i) What type of information does the learner preferentially perceive?

Perception (Sensing/Intuitive): Sensing learners tend to be patient with details and good at remembering facts. Intuitive learners prefer to grasp new concepts, and prefer to discover possibilities and relationships; they are more comfortable with abstractions, theories and mathematical formulations.



(ii) What type of sensory information is most effectively perceived?

Input (Visual/Verbal): Visual learners prefer teaching material when it is presented in form of pictures, diagrams, flow charts or videos. Verbal learners get more out of words: through written and spoken explanations.

(iii) How does the learner prefer to process information?

Processing (Active/Reflective): Active learners prefer to retain and understand information by doing something active with it, and prefer to work in groups, as this allows them to discuss and explain the information they have received. Reflective learners prefer to think and absorb the information individually or in small groups.

(iv) How does the learner characteristically progress toward understanding?

Understanding (Sequential/Global): Sequential learners prefer information to be provided in a logical progression of incremental steps, and tend to make small steps through learning material by clicking the "next/previous" buttons. Global learners have a tendency to acquire knowledge through substantial leaps and often gain an understanding of the overarching concepts before delving into the finer details of a subject.

Each dimension consists of two opposite poles representing different learning styles. The learning style for an individual is generated by merging the poles of each dimension (Hasibuan et al., 2016). For example, in the FSLSM, the input dimension encompasses two distinct poles or learning styles: visual and verbal. These contrasting learning styles reflect the diverse ways in which the learner prefers to receive information. Figure 2 provides a visual representation of these two poles within the four dimensions.

To identify the degree of preference of learners for each dimension, the index of learning styles (ILS) questionnaire can be used (Felder & Solomon, 2006). This questionnaire, developed by Richard Felder and Barbara Soloman in 1991, consists of a 44-item (11 items for each dimension) forced-choice instrument. Participants respond to the questionnaire by selecting their preferred option from each pair of items, indicating their inclination towards a particular learning style within each dimension. The responses are then compared, and an odd score value ranging from -11 to 11 is assigned to indicate the degree of preference. Score -3 to 3 represents balanced style on the two poles of the



Fig. 2 Scales of learning styles

dimension. Score 5 or 7 represents moderate preference for one pole of that dimension, and 9 or 11 strong preference for one pole of that dimension (Fig. 2).

Each pole has its own singularity and characteristics, indicating how learners prefer to interact with specific resources and perform specific activities.

The use of the FSLSM in this study is motivated by several compelling reasons. Firstly, the model has demonstrated its applicability in addressing fundamental scientific issues, making it well-suited for investigating learning styles in our context (Özpolat & Akar, 2009). Secondly, the FSLSM has been widely recognized as the most appropriate choice for hypermedia courseware, indicating its compatibility with modern educational technologies (Carver et al., 1999; Essa et al., 2023; Graf & Kinshuk, 2007; Raleiras et al., 2022; Zhang et al., 2020). Thirdly, the dimensions of the FSLSM are distinct and independent, allowing for a comprehensive and nuanced description of learning styles (R. M. Felder & Silverman, 1988). Additionally, the model represents each dimension on a scale from -11 to +11, enabling a more precise depiction of learners' preferences at a granular level (Graf & Kinshuk, 2007). fourthly, the Index of Learning Styles (ILS) questionnaire has been proven valid and reliable for assessing learning styles (Marosan et al., 2022).

The relevance of learning styles in education

Educational researchers acknowledge the inherent uniqueness of individual learners: no two learners are the same as regards their way to learn (Wood, 2009). Recognizing this variability is essential to designing effective instructional strategies that meet the diverse needs and preferences of learners. In order to reduce attrition and improve skill development, instructional materials should be tailored to meet the needs of learners. This concept, known as the "meshing hypothesis" according to Pashler (2008), emphasizes the importance of aligning instructional strategies with learners' preferences. An example of these differences is seen in learners' preferences for information presentation. While some learners are comfortable with information presented in visual format (graphical representations), others prefer verbal explanations and have a stronger inclination towards retaining information through reading. Supporting this perspective, Claxton and Murrell (1987) emphasize the need for instructors to recognize and respect the differences that learners bring in the classroom. To foster effective learning, they advocate for systematically crafted learning experiences tailored to match individual student's learning styles. The more thoroughly the instructors understand the differences, the better chance they have of meeting the diverse learning needs of all their learners (Felder & Brent, 2005).

This alignment of instructional methods with learning styles has tangible benefits. Many researchers assert that a recognizing and integrating learning styles in the classroom can facilitate effective learning (Graf & Liu, 2009). By tailoring curricula around learning styles, academic achievement as well as the learners' self-confidence can be improved (Reid, 2005; Sadeghi et al., 2012). It enhances learner satisfaction, augments students' motivation and engagement, and even reduces the time required for learning (efficiency) (Dağhan & Akkoyunlu, 2012; Graf & Liu, 2009; Reid, 2005). According to Peacock, (2001) 81% of teachers believe that a mismatch between teaching and learning styles leads to learning failures, frustration and lack of motivation among learners. However, the idea of learning styles hasn't been free from controversy, Numerous researchers and educators have expressed skepticism about the theory of learning styles. Some even argue that they might be a myth without solid scientific evidence (Dekker et al., 2012; Kirschner, 2017; LeBlanc, 2018; Pashler et al., 2008; Riener & Willingham, 2010). According to Coffield et al. (2004), most of the criticism arises from the lack of a unified framework, given the diverse models in literature, causing confusion in their application. Pashler et al. (2008)stated that the practice of tailoring educational resources and activities based on learners' learning styles lacked scientific validation. Popescu et al., (2007) argued that learning's complexity surpasses mere learning styles. Moreover, they question whether learning styles have ever significantly influenced learning. They advocate for harnessing various models and their unique attributes for a holistic approach. Newton and Miah (2017) highlighted out the risks of applying learning styles in education, emphasizing the danger of confining learners. For instance, a "visual learner" might feel discouraged tackling subjects that don't align with their identified learning style, such as studying music.

Despite the numerous criticisms regarding learning styles, no alternatives have emerged. The plethora of models in literature underscores a widely held belief in the concept's utility, stating that learners do have traits affecting how they learn (Essa et al., 2023; Graf et al., 2012; Li et al., 2019; Suganya & Sheshasaayee, 2022).

Approaches to identify learning styles

There are two approaches to identify learning styles (Fig. 3), namely collaborative and automatic (Brusilovsky, 1996). The collaborative approach involves learners actively participating in the process by completing a questionnaire that aligns with a specific learning style model. However, this approach is static in nature and has certain limitations. When the questionnaire is excessively lengthy, learners may provide answers hazardly, compromising the accuracy of the results. Additionally, learners may not be fully aware of the significance of the questionnaire, leading to a potential lack of commitment and attentiveness during its completion (Jackson, 1990).



Fig. 3 Collaborative and automatic detection of learning styles

Furthermore, it is important to note that learning styles can vary and change depending on the learning pressure or the specific situation within the learning process (Coffield et al., 2004; Pashler et al., 2008). As a result, the static collaborative approach may not accurately capture the flexible and stable personal characteristics of learners (Kolb, 1976). Learners' preferences and styles may evolve over time, and relying on a static model may limit our understanding of their adaptive learning behaviors.

To overcome the limitations of using such questionnaires, researchers have proposed two main approaches for automatically identifying learning styles: the literature-based approach and the data-driven approach. These approaches are focused on collecting and analyzing traces that learners leave behind in order to identify learners' learning styles.

Graf et al. (2008) first introduced a literature-based approach that uses the learner's traces to get patterns about their learning styles This approach involves identifying patterns in the data using simple rules in the form of "if...then...else" instructions defined to calculate learners' learning styles (Ahmad et al., 2013; Latham et al., 2012; Scott et al., 2014). The main advantage of this approach is both generic and applicable to data collected from any courses (Dung & Florea, 2012). One limitation of this approach is the reliance on predefined rules in exhaustive way. The approach may also struggle with handling large volumes of data, requiring efficient data processing and analysis techniques to identify meaningful patterns.

The data-driven approach aims to build a model that imitates a learning style questionnaire (Feldman et al., 2015). Instead of relying on explicit questionnaires, this approach utilizes artificial intelligence algorithms that take learners' behaviors as input and generate their learning styles as output.

In this sense, many algorithms have been used in the literature, such as: (i) Bayesian technique (Halawa et al., 2015; Maraza-Quispe et al., 2019; Rasheed & Wahid, 2021), (ii) Neural network (Bajaj & Sharma, 2018; Ferreira et al., 2018; Kolekar et al., 2017), (iii) Decision tree method (Crockett et al., 2017; Karagiannis & Satratzemi, 2018; Sheeba & Krishnan, 2018), (iv) Naïve bayes (L. X. Li & Abdul Rahman, 2018; Maraza-Quispe et al., 2019), (v) deep learning algorithms (Alshmrany, 2022; Anantharaman et al., 2018; Mubarak et al., 2022).

One of the primary advantages of the data-driven approach is its utilization of real data to classify learners. By analyzing learners' actual behaviors and interactions within the learning environment, the approach can provide accurate and real-time tracking of learners' learning styles. This dynamic nature enables the detection of changes in learning styles as learners progress and adapt to different learning situations and context (Benabbes et al., 2023).

Educational data science and dashboard: leveraging data for enhanced education

The field of educational data science (EDS) holds immense promise in its ability to drive meaningful educational outcomes by using advanced data analysis techniques. Through the application of machine learning, data mining, and statistical analysis to educational data, EDS aims to enhance student learning performance, evaluate teacher effectiveness, address student retention and success, and inform educational policy and decision-making (Romero & Ventura, 2017).

One key mains of EDS is to enhance student learning performance. By analyzing student records and academic data, EDS can identify patterns and trends that help identify factors influencing student achievement (Pratsri et al., 2022). For instance, it can uncover the impact of different teaching methodologies, classroom environments, or even socioeconomic factors on student outcomes (Suganya & Sheshasaayee, 2022). This knowledge enables educators to make data-driven decisions and tailor instructional strategies to meet individual student needs.

EDS also focuses on evaluating and improving teacher effectiveness (Swai et al., 2023). Through the analysis of teacher evaluations, classroom observations, and student feedback, EDS can provide valuable insights into instructional practices that lead to improved student engagement and academic growth (Wengrowicz et al., 2022). By identifying effective teaching strategies, EDS can assist in teacher professional development efforts, ultimately enhancing overall educational quality.

Furthermore, EDS plays a role in addressing student retention and success. By examining data related to student attendance, participation, and socio-emotional factors, EDS can identify early warning signs of student disengagement or potential dropout risks. This information enables educators and administrators to intervene and provide targeted support, thereby increasing student retention rates and fostering a more inclusive and supportive learning environment (Schofield, 2021).

Another area where EDS proves valuable is in the field of educational policy and decision-making. By analyzing large-scale educational data, such as national or international assessments, EDS can provide policymakers and educational leaders with evidencebased insights. These insights can inform the development of targeted interventions, curriculum improvements, and resource allocation strategies to enhance overall educational quality and equity (Fig. 4).



Fig. 4 CRIPS Model (Kelleher et al., 2020; Martínez-Plumed et al., 2019)

The Education Data Science lifecycle can be divided into six main steps:

Problem identification and understanding The first step is to identify and understand the problem that you are trying to solve. This involves understanding the business need, the different specifications, and the requirements.

Data collection The next step is to collect and understand the data that you will use to train your model. This data can come from a variety of sources, such as web server logs, social media data, data streams extracted from web APIs, databases, etc. Once you have collected the data, you need to perform exploratory analysis to better understand it.

Data preprocessing The third step is to prepare the data for use in model training. This involves data cleaning and feature extraction. Data cleaning is the process of removing errors and inconsistencies from the data. Feature extraction is the process of identifying the most important features in the data that will be used to train the model.

Modeling The fourth step is the modeling phase. This involves selecting a learning algorithm and its hyperparameters. A learning algorithm is a method for training a model. Hyperparameters are the settings of the learning algorithm.

Evaluation The fifth step is the evaluation phase. This involves testing and verifying the model and its parameters to ensure that it meets the objectives formulated at the beginning of the analysis. This phase is where the decision is made whether the model is robust enough and ready for deployment or if it needs reconfiguring.

Deployment The final step in the lifecycle is to operationalize and integrate the solution into the company in order to solve the original problem. This involves deploying the model into production and making it available to users.

The process of analyzing learning traces is a complex and iterative process, but it can be a valuable tool for improving learning. The main study in this article is a practical example of how the data science lifecycle (crips model) can be applied in the MOOC environment. This study used the lifecycle to develop a model that could predict learners' learning styles. This model was then used to provide valuable support to teachers and instructors in gaining deep understanding of their learners' preferences.

During the deployment phase of the data science lifecycle, one effective way to support teachers and instructors is through the use of dashboards. These dashboards serve as a means to communicate the insights generated throughout the data analysis process. By presenting key findings and visually representing relevant data, dashboards offer a user-friendly interface for educators to access and interpret the insights derived from the data analysis process. Our study demonstrates the significant impact of dashboarding, or learning analytics, on enhancing learners' educational experiences. user-friendly interface for educators to access and interpret the insights generated through the data analysis process. The upcoming section of the article will delve into a comprehensive exploration of these aspects.

This focus on using dashboards and learning analytics aligns with the field of Learning Analytics, which aims to address questions pertaining to improving learners' educational experiences. Analytics provide insights at various levels to enable informed decision-making. Descriptive analytics, for instance, offer snapshots of variables, indicating trends and current status, while predictive analytics utilize machine learning algorithms to forecast future outcomes based on patterns derived from past data (Susnjak et al., 2022). By seamlessly integrating the data science lifecycle with learning analytics, educators gain a deeper understanding of their learners' needs and preferences, enabling them to make informed decisions and implement targeted strategies that enhance the overall learning experience.

Learning analytics, as described by Elias (2011) and Larusson and White (2014), provides valuable insights to users, such as teachers, about what transpires in a class, regardless of the activity type. In other words, learning analytics is defined as the effort to enhance learning through targeted data analysis. The goal is to collect and analyze learners' interaction data, identify at-risk students early on, and improve the quality of the educational experience (Susnjak et al., 2022).

To effectively present the analyzed data, learning dashboards are commonly used. These dashboards integrate various indicators about the learner, learning process, and contexts, using visualization methods. In the field of education, data visualization plays a critical role in understanding, analyzing, and communicating learning-related information. Visual representations, such as graphs and charts, simplify complex concepts and facilitate the establishment of connections between ideas (Ramaswami et al., 2022).

Educational dashboards serve multiple stakeholders, providing real-time monitoring of student engagement, particularly beneficial in online environments with physical separation. They offer students greater visibility into their online learning behaviors, empowering them to make more informed study-related decisions. Additionally, personalized metrics facilitate self-reflection and encourage positive behavioral adjustments. For instructors, educational dashboards equipped with predictive analytics help identify at-risk students and enable timely interventions to improve course outcomes (Akçapınar et al., 2019; Queiroga et al., 2020).

Aligned with these concepts, our research focuses on the application of the educational data science process. Our article delves into the entire process of educational data science, elucidating how the integration of data analysis and visualization can effectively transform teaching practices and drive meaningful educational outcomes. Through our research, we aim to contribute to the advancement of teaching methodologies in the realm of MOOCs, equipping educators with the tools they need to unlock their full potential and create engaging learning experiences. By harnessing the power of data analytics and dashboards, educators can gain valuable insights, make informed decisions, and ultimately improve the overall learning experience for their students.

Decoding learners' preferences: methodological approach to identifying learning styles in MOOCs

In this section, we provide an outline of our proposed approach for identifying learning styles in MOOCs. We first present an overview of our solution and then delve into the detailed phases of our methodology.

The main aim of our work is to propose a predictive model for learning styles. Our process is composed of three stages, as shown in Fig. 5. The first phase of our process consists of data collection (dataset). The data that are gathered are related to the behavioral traces of the learners during the learning process. To ensure the reliability and usefulness of the collected data, a pre-processing step that involves cleaning and filtering the raw data was performed. This step helps eliminate any inconsistencies or noise that



Fig. 5 Stages of the methodology

may affect the accuracy of our model. Additionally, we perform a feature selection step to identify the most relevant and informative attributes from the collected data.

In the second phase of our process, we construct a predictive model that can identify learning styles based on the processed data. To achieve this, we used clustering and classification techniques and performed a series of experiments and an evaluation of the model.

In the third phase, we create a visualization tool called MOOCLS that can be used to give further recommendations to teachers and instructors. This tool utilizes the output of our predictive model to offer insights regarding the learning style preferences of individual learners and recommendations about the learning content of the course.

In our previous work (Hmedna et al., 2019, 2020) we presented an overview of our learning styles prediction model. Nevertheless, in order to ensure a comprehensive understanding of the current paper and to improve comprehension of the entire process employed we will delve into several important steps of our learning styles prediction model in subsequent sections. That being said, the current paper is centered on the steps that are visually highlighted in blue, as depicted in Fig. 5.

Dataset and data preparation

Dataset description

The dataset used in this research work, was collected from edX course "Statistical Learning" (session Winter 2015 and 2016), via a data-sharing agreement with the Stanford University. The course emphasized the instruction of both supervised and unsupervised learning algorithms, along with the underlying theoretical concepts. This comprehensive curriculum aimed to equip learners with the knowledge and skills necessary to understand and apply these algorithms in practice. The course spanned nine weekly sessions, each consisting of six to eleven lecture videos, readings, and quizzes. To receive a statement of accomplishment, learners have to get an overall course grade of 50% or higher. Table 1 provides an overview of the dataset, highlighting the number of enrolled learners and the events recorded from these learners.

Whenever a learner interacts with the MOOC platform, a clickstream event is generated. These events capture various actions and activities performed by learners while engaging with the course content and platform. Each event is associated with specific attributes that provide valuable information about the nature and context of the

Session	# Learners	# Events	Launch date	End date
Winter 2015	32,209	18 475 724	2015-01-19	2015-04-06
Winter 2016	20,526	12 363 613	2016-01-12	2016-04-06

 Table 1
 A description of MOOC dataset used in this study

interaction (Benabbes et al., 2023). These events capture different types of interactions, including reading outlines, viewing video lectures, attempting graded quizzes, and participating in forum discussions. Each event is described by several attributes, such as the type of interaction (event_type), the date and time of the interaction (timestamp_event) as well as the learner identifier (anon_screen_name).

Before using the data to train the model, data preprocessing was performed. Raw data contains of noisy data, this noise affects the learning ability of the machine-learning model, Therefore, before using the data to train our model, it must be cleaned, formatted, and restructured — this is typically known as preprocessing (Mazzola & Mazza, 2009). Fortunately, for our dataset, we are not dealing with missing values, however, there are some qualities about certain features that must be adjusted. In this context, we reduced the number of "event_type" from 7550 distinct event_type to 55 after preprocessing.

The database was analyzed in order to detect outliers corresponding to learners with anomalous behaviors (values far out of the typical range). These outliers can have a negative impact on the results of our analysis (the model can be trained more slower, augment the complexity of the model, higher risk of overfitting) (Fayyad et al., 1996). Concerning the technical part, we tested two unsupervised algorithms for anomaly detection, isolation forest or Iforest (Liu et al., 2008) and Local Outlier Factor (LOF) (Breunig et al., 2000). The isolation forest algorithm was chosen for the results it offers and its simplicity of implementation.

Isolation Forests are based on the principle that anomalies are few and different from normal instances. They utilize the aggregation of many trees called isolation trees, where each tree uses a random sample of observations and selects variables and thresholds for division randomly.

To implement this algorithm, we represented each learner by a set of characteristics, namely:

- "# events": The total number of events each learner was generated.
- "# weeks": The total number of weeks to which a learner has accessed.
- "certified": Indicates if the learner received the certificate of completion.

As a result, 166 learners have been identified as anomalies. An exploration of the traces of these learners showed that the main cause of these anomalies was initially due to a bug in the application (openEdx) that iteratively sends the same request, causing the contamination of the traces of these learner's traces.

Feature engineering: characterizing learners' learning styles

In the feature selection phase, the goal is to determine which features from the raw data are most relevant for creating a robust model. Selecting relevant features associated with each

learning style is difficult, time-consuming and requires expert multidisciplinary knowledge. Most of the methods proposed in the literature use the linear correlation between the features (Beal, 2015). Linear correlation is a statistical measure that quantifies the strength and direction of the relationship between two variables. By examining the correlation between features and the target variables of interest, researchers can identify the most influential features for prediction or analysis. High correlation suggests that changes in one feature correspond to predictable changes in another, making it a useful criterion for feature selection.

In our study, the feature selection process is guided by prior research conducted by Garcia et al. (2007), Graf (2007), Latham et al. (2012), and Villaverde et al. (2006), as shown in Table 2. One of the commonly used methods in addressing this phase of process is Backward Feature Elimination (BFE) (Kohavi & John, 1997). It is an iterative process that progressively eliminates non-significant features. The BFE process begins with fitting a model that incorporates the full set of features. Subsequently, a certain proportion of the less discriminating features are removed based on a set criterion. With each iteration of feature removal, we assessed cluster compactness and separation. Compactness measures intracluster closeness, while separation measures inter-cluster distinction. By evaluating these metrics in each iteration, we aimed to find a balance between removing irrelevant features and maintaining optimal clustering quality.

Table 2 provides an overview of the features considered in our study, based on the aforementioned research methods. By following this approach, we aim to identify the subset of features that best contribute to the predictive power and interpretability of our model, thereby enhancing the effectiveness of our analysis in the context of learning styles and education.

In addition to feature selection, another critical aspect of data preprocessing is feature scaling, as highlighted by Rebala et al., (2019). Scaling is particularly important for machine learning algorithms, as their performance can be sensitive to the magnitude of feature values. If features are not scaled, certain features with larger values can disproportionately influence the model's output, potentially leading to biased or inaccurate results. To address this issue and ensure a uniform distribution of feature values, we apply the MinMax scaling method (Eq. 1). This method involves normalizing numerical values to a consistent scale, typically ranging from -1 to 1.

$$x' = \frac{x - x_{\min}}{x_{\min} - x_{\max}} \tag{1}$$

where x_{min} and x_{max} are the minimum and maximum values, respectively, of the feature x in the dataset. To carry out this normalization, the MinMaxScaler function provided by sklearn was used.

In summary, our data preprocessing pipeline encompasses both feature selection, guided by prior research, and feature scaling using the MinMax method. By selecting relevant features and ensuring uniformity in their values, we aim to enhance the performance and interpretability of our machine learning model in the context of this study.

Table 2 Features set used for the analysis

	Features Description		Study				
			Villaverde et al., (2006)	Latham et al., (2012),	Garcia et al., (2007)	Graf, (2007)	
Processing							
Active	# pause_video	Number of pauses videos		*			
	# problem_vis- its	Number of visits of problems	£ *		*	*	
	Problem_stay	Total time spent or problems	ז*		*	×	
	# problem_sub mit	-Number of assign- ment problems submited	×			*	
	∆ problem_ submit	Aver- age # of assign- ment problems submited	*			*	
	# forum_post	Number of forum posting	*	*	*	*	
Reflective	# show_answer	Number of visits of the answers of problems		×	*	*	
	Show_answer _stay	Total time spent reading the answers problems		*	*	*	
	Δ show_answe	r Average time spent reading the answers problems		*	*	*	
	# outline_visit	Number of visits of outlines	f			*	
	# view_post	Number of viewing posts in forum	9		*	*	
Perception							
Sensing	# example_visit	Number of visits of examples	*		*	*	
	example_stay	Total time spent or examples	ז*	*	*	*	
	# problem_visit	Number of visits of problems	F *		*	*	
	problem_stay	Total time spent or problems	ז*		*	*	
	# show_answei	Number of visits of the answers of problems		*	*	*	
	Show_answer_ stay	Total time spent reading the answers problems		*	*	*	
	# seq_prev	Number of time the learner navigates to the previous unit in a sequence	*		*		
	# concret_ material	Number of visits of concret material	£ *	*	*		

Table 2 (continued)

	Features Description		Study				
			Villaverde et al., (2006)	Latham et al., (2012),	Garcia et al., (<mark>2007</mark>)	Graf, (2007)	
Intuitive	# seq_visit	Number of visits of units				*	
	seq_stay	Total time spent on units				*	
	# forum_post	Number of forum posting	*	*	*	*	
	# view_post	Number of viewing posts in forum	g		*	*	
	# outline_visit	Number of visits or outlines	f			*	
Input							
Visual	# video_visit	Num- ber of viewed vid- eos		×			
	video_stay	Total time spent or viewing videos	n	*			
	# play_video	Number of play videos		*			
	# hide_tran- script	Number of hiding video transcripts		*			
Verbal	# show_tran- script	Number of showing video transcripts		×			
	# speed_ change_video	Number of chang- ing video speed		*			
	# view_post	Number of viewing posts in forum	g		*	*	
	# forum_post	Number of forum posting		*	*	*	
	# forum_stay	Total time spent reading forum posts					
	# outline	Number of visits or outlines	f			*	
Understanding							
Sequential	# seq_next	Number of time the learner navi- gates to the next unit in a sequence	*		*		
	# seq_prev	Number of time the learner navigates to the previous unit in a sequence	*		*		
	∆ sequential_ navig	= (# seq_next + # seq_prev) / (# seq_next + # seq_ prev + # seq_goto	*		*		

Table 2 (continued)

	Features	Description	Study			
			Villaverde et al., (2006)	Latham et al., (2012),	Garcia et al., (2007)	Graf, (<mark>2007</mark>)
	# dist_unit_visi	t Number of distinct unit (pages) visted	t			*
	# show_answe	r Number of visits of the answers of problems		*	*	*
	# page_close	Number of closed pages				*
Global	# progress_ show	Number of visits of progress page	f			*
	# seek_video	Number of seek in video		*		
	# seq_goto	Number of time the learner skip to specific units	*		*	×
	∆ global_navi- gation	=# seq_goto / (# seq_next + # seq_prev + # seq_goto)	*		×	*
	# outline	Number of visits of outlines	f			*

From events logs to insights: a methodology through machine learning for learning style detection

Unsupervised modeling: uncovering clusters of learning styles

In this phase, we rely on the features obtained from the previous step of feature selection to perform clustering analysis. The main objective is to group learners based on their level of preference for each learning style, with the aim of identifying clusters where learners within the same group exhibit higher similarity to each other compared to learners in other clusters (Li et al., 2020). To achieve this, we employ the widely used K-means algorithm, which is known for its simplicity and effectiveness in clustering tasks (Hartigan & Wong, 1979).

A major challenge in configuring the k-means algorithm is the choice of the optimal value of clusters (k). Based on the existing solutions in the literature, to estimate the number k, we have chosen to use the elbow method (Kodinariya & Makwana, 2013). This method remains simple and offers good results, even though it requires a subjective judgement as to the location of the elbow.

This method involves plotting a graph with the number of clusters (K) on the x-axis and the sum of squared distances of samples to the nearest cluster center (SSE) on the y-axis. The resulting curve typically exhibits an "elbow" shape, and the point of inflection signifies a trade-off between cluster compactness and separation.

By applying the elbow method, we have determined that the optimal number of clusters for each session and learning style is four. The assignment of cluster labels is based on the average feature values associated with each learning style. Learners with a strong preference for a specific learning style exhibit higher average feature values and are more likely to be grouped together in the same cluster. Cluster analysis provides insights into learner preferences based on the average values of features associated with each learning style (Song & Wang, 2023). This analysis allows us to assign cluster labels to learners based on their level of preference for a specific learning style. The cluster label is defined based on the average value of the features associated with each learning style. Learners who have a strong preference for a particular learning style have a higher average value for all features and are more likely to interact with the platform than learners in other clusters. Their learning style preference aligns closely with the features associated with that particular style. On the other hand, learners in other clusters may have weaker preferences or exhibit a balance of multiple learning styles.

Our clustering analysis (Fig. 6) has resulted in the following labels for the identified clusters: Cluster 1 represents learners with a "very weak preference," Cluster 2 represents learners with a "weak preference," Cluster 3 represents learners with a "moderate preference," and Cluster 4 represents learners with a "strong preference."

Interestingly, the largest number of learners are grouped into the "very weak" preference clusters. This observation can be attributed to the high drop-out rates commonly observed in Massive Open Online Courses (MOOCs). Previous studies (Kloft et al., 2014; Li et al., 2017; Onah et al., 2014) have highlighted the challenges and factors contributing to drop-out rates in MOOCs. The prevalence of learners with a "very weak preference" suggests that a significant portion of learners may disengage or struggle to align their learning styles with the course content and delivery.

After having clustered learners according to their preferences for each learning style, the quality of this clustering should be assessed, despite the fact that in unsupervised learning it is difficult to evaluate the performance of a clustering model, especially when there are no reference labels (Shutaywi & Kachouie, 2021). To overcome the absence of reference labels, internal evaluation criteria are utilized to assess the quality of the obtained clusters based on two main criteria: cluster compactness, which measures the similarity among samples within a cluster, and cluster separability, which quantifies the distinction between a cluster and others (Liu et al., 2010). Various indices exist in the literature to quantify these criteria. In this study, we employ the Calinski-Harabasz (CH) index (Caliński & Harabasz, 1974) and the Silhouette (SI) index (Rousseeuw, 1987) for evaluation.



Fig. 6 Clustering results of the "Winter 2015" session: **a** active; **b** reflective; **c** visual; **d** verbal; **e** sequential; **f** global; **g** intuitive and **h** sensing learning style

Table 3	Computed	Calinski-Harabasz	& Silhouette index	(best results bolded)
---------	----------	-------------------	--------------------	-----------------------

Validation index	n Calinski-Harabasz (CH)			Silhouette (SI)				
Learning styles	K-means	MiniBatch	Birch	Agglomerative	K-means	MiniBatch	Birch	Agglomerative
Session 20	15							
Active	182,350.04	182,096.07	87,462.00	164,592.11	0.80	0.80	0.74	0.79
Reflec- tive	81,988.04	81,734.32	29,448.13	66,986.25	0.76	0.76	0.71	0.75
Sensing	73,109.54	71,040.24	64,871.41	57,600.90	0.73	0.76	0.77	0.75
Intuitive	27,146.78	25,250.56	14,583.55	22,774.94	0.68	0.72	0.77	0.61
Visual	123,760.92	114,153.13	71,621.93	111,571.95	0.67	0.72	0.78	0.65
Verbal	33,756.19	31,552.79	22,630.88	30,930.14	0.70	0.60	0.77	0.64
Sequen- tial	23,9064.22	239,022.98	218,663.28	232,308.86	0.81	0.80	0.80	0.80
Global	26,101.06	24,182.84	18,263.43	22,722.22	0.82	0.74	0.80	0.80
Session 20	16							
Active	102,957.48	93,535.52	53,426.37	73,219.40	0.77	0.77	0.71	0.75
Reflec- tive	52,268.05	51,947.40	32,443.41	43,995.80	0.76	0.76	0.73	0.73
Sensing	44,779.28	44,256.98	39,460	38,913.64	0.72	0.72	0.75	0.74
Intuitive	16,893.7	14,158.7	9,936.57	15,121.62	0.70	0.43	0.78	0.74
Visual	92,224.74	91,347.66	47,077.40	81,402.06	0.68	0.70	0.76	0.75
Verbal	15,951.95	9,562.31	4,123.77	15,085.92	0.76	0.74	0.94	0.75
Sequen- tial	177,096.72	175,840.89	154,747.88	170,035.39	0.81	0.80	0.80	0.80
Global	18,222.55	15,414.20	14,687.96	14,329.18	0.83	0.81	0.82	0.80

The CH index calculates a normalized ratio of inter-cluster dispersion (separability) to intra-cluster dispersion (compactness). A higher value of the CH index indicates better clustering quality. The Silhouette Index (SI) is a widely-used internal validation index that measures both cluster separation and compactness. The SI index ranges from -1 to 1, with values closer to 1 indicating better clustering quality, indicating that the points are closer to each other within the same cluster and farther away from other clusters.

Based on our evaluation of the CH and SI indexes across different clustering algorithms (k-means, MiniBatch, Birch, Agglomerative), our results indicate that k-means demonstrated superior performance compared to the other algorithms (Table 3). This finding aligns with the results from a systematic literature review by Essa et al. (2023). Their comprehensive review, spanning from 2015 to 2022, highlighted the predominant use of k-means in the detection of learning styles in various studies.

Aggregation process

In the Felder–Silverman Learning Style Model (FSLSM), the poles of each dimension are opposed (Felder & Silverman, 1988; Graf et al., 2008), as shown in Fig. 2. Therefore, when there is a strong occurrence of a specific behavior that indicates one pole, a low occurrence of the same behavior indicates the other pole, and vice versa (Graf & Liu, 2009). For example, learners with a strong active learning style also show a weak preference for



Fig. 7 Balance of learning styles

Table 4	Balance	of	learning	styl	es: inp	ut c	lime	ension
			J					

	Visual LS						
	Very weak	Weak	Moderate	Strong			
Verbal LS							
Very weak	Balanced	Moderate visual	Strong visual	Strong visual			
Weak	Moderate verbal	Balanced	Moderate visual	Strong visual			
Moderate	Strong verbal	Moderate verbal	Balanced	Moderate visual			
Strong	Strong verbal	Strong verbal	Moderate verbal	Balanced			

a reflective learning style. If the preferences are similar (either strong or weak), the learning style is balanced.

In order to label our dataset according to this scale of preferences (*strong*_{pole1}, *mod*erate_{pole1}, *balanced*, *moderate*_{pole2}, *strong*_{pole2}), we propose a grid called the "*balance of learning styles*" (Fig. 7), which consists of adding the two degrees of preferences (*prefer*ence_{pole1}, *preference*_{pole2}) relative to each dimension.

Table 4 presents an example that demonstrates the application of our balance for the input dimension. It showcases how the degrees of preferences (preferencepole1 and preferencepole2) are combined to determine the corresponding learning style label. This table serves as an illustration of how the balance of learning styles approach is employed in our study for labeling purposes.

Each learner is represented as a feature vector. After merging the two pole feature vectors using the balance of learning styles grid, we obtain a global feature vector dimension. Figure 8 provides a visual representation of this merging process.



Fig. 8 The aggregation process

Strong	Moderate	Balanced	Moderate	Strong	
2171	7297	40,128	2849	124	Reflective
3411	8888	39,759	486	25	Verbal
761	4566	45,650	1564	28	Intuitive
3753	18,333	30,213	263	7	Global
	Strong 2171 3411 761 3753	Strong Moderate 2171 7297 3411 8888 761 4566 3753 18,333	StrongModerateBalanced2171729740,1283411888839,759761456645,650375318,33330,213	StrongModerateBalancedModerate2171729740,12828493411888839,759486761456645,6501564375318,33330,213263	StrongModerateBalancedModerateStrong2171729740,12828491243411888839,75948625761456645,650156428375318,33330,2132637

 Table 5
 The distribution of classes in each datasets

The aggregation phase allows us to generate four labeled datasets each of which is related to a given dimension. Table 5 shows the distribution of classes. It is important to note that the datasets are multi-class and unbalanced, meaning that the data points are not evenly distributed among the different classes. These datasets will be used as input to the predictive models, as described in the next section.

Supervised modeling: predictive models for identifying learners' learning styles

In this paper, we compare several classifiers to assess their performance for the identification of learning styles: decision tree (DT), random forest (RF), k-nearest neighbor (KN) and neural network (NN). These classifiers were chosen because of their large use and good performance in similar research papers.

To conduct the evaluation, the dataset was divided into four separate datasets, each corresponding to a dimension of the Felder-Silverman Learning Style Model (FSLSM). For each dataset, a training set and a testing set were created.

In the context of machine learning algorithms, the selection of appropriate hyperparameters is crucial for achieving optimal model performance (Duong, 2019). Hyperparameters are considered as properties of an algorithm that need to be defined prior to training a model. To determine the best combination of hyperparameters from a set of possibilities, a grid search approach was utilized, coupled with tenfold cross-validation.

By performing a grid search, we systematically explored various parameter combinations to identify the configuration that yielded the highest classifier performance. The inclusion of cross-validation ensured that the models were validated on multiple subsets of the data, reducing the risk of overfitting and increasing the generalizability of the results.

The outcomes of the grid search and cross-validation process are summarized in Table 6, which presents the performance metrics for each classifier model. These results provide valuable insights into the effectiveness of different parameter settings and assist in selecting the most appropriate configuration for future applications.

After the development of the classification models, the subsequent step involved evaluating their performance on unseen test data. To assess the models' effectiveness, we employed four commonly used evaluation metrics: accuracy, precision, recall, and F1-score.

Considering the highly unbalanced nature of the dataset, it is crucial to employ evaluation metrics that can effectively handle class imbalance. In this context, the use of macro-precision and micro-precision metrics proves to be valuable for assessing the performance of our models (Tharwat, 2018). These metrics specifically account

Model	Hyper-parameters	Datasets			
		A/R	S/I	V/V	S/G
DT	Max_depth	8	7	8	7
	Criterion	Entropy	Entropy	Entropy	Entropy
	Max_features	None	None	None	None
RF	N_estimator	7	10	7	7
	Max_depth	8	7	8	8
KNN	N_neighbors	50	50	50	50
	Leaf_size	1	1	1	1
	Algorithm	Auto	Auto	Ball_tree	Auto
NN	Hidden_layer_size	7	7	8	8
	Max_iter	500	500	500	500

Table 6	Configuration	of algorithms	adopted in	each datasets
---------	---------------	---------------	------------	---------------

Table 7	Evaluation	metrics
lable /	Evaluation	methes

Metric	Formula	Description
Accuracy	$ACC = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FN_i}$	The percentage of predictions that are correct
Precision	$Pr \ ecision_i = \frac{TP_i}{TP_i + FP_i}$	The percentage of positive predictions that are correct
Recall	$\operatorname{Re} call_i = \frac{TP_i}{TP_i + FN_i}$	The percentage of positive cases that were predicted as positive
F1-score	$F1-score_i = \frac{2 \times P \times R}{P+R}$	The Harmonic mean of precision and recall. P represents precision and R represents recall
Micro-precision	$Micro_{precision} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + FP_i}$	The average precision of all classes
Macro-precision	$Macro_{precision} = \frac{\sum_{i=1}^{n} precision_i}{n}$	The sum of each class precision divided by the number of classes

for the imbalanced distribution of classes in the data and provide a comprehensive evaluation of the models' predictive capabilities.

Having explored the methodology and evaluation process, we now turn our attention to the core of our study: predicting the learning styles of students in a MOOC. In the subsequent analysis, we will delve into the performance of the developed models across the four datasets.

Assessing the performance of the developed models across all four datasets, it is evident that they let to favorable results (Table 7).

However, upon close examination of the performance metrics showcased in Tables 8, 9, 10 and 11, it is noted that among all the developed models, the decision tree (DT) classifier consistently demonstrated the highest level of performance.

This finding strongly emphasizes the relevance and suitability of the DT classifier for our specific study objectives. Its superior performance underscores its efficacy in accurately predicting and understanding students' learning styles within the context of our research.

As mentioned previously, the four datasets are highly unbalanced. We therefore used macro and micro-precision indexes to assess the quality of our classifiers. The results reveal that the DT handles unbalanced data better than the other classifiers. These

Algo	Avg. accuracy	Avg. precision	Avg. recall	Avg. f1-score	Avg. micro- precision	Avg. macro- precision
DT	0.97	0.97	0.97	0.97	0.97	0.88
RF	0.94	0.94	0.94	0.93	0.94	0.71
NN	0.97	0.97	0.97	0.97	0.96	0.91
KNN	0.90	0.90	0.91	0.89	0.90	0.60

 Table 8
 Results of prediction models—Perception dataset (best predicted result of the model are bolded)

 Table 9
 Results of prediction models—Processing dataset (best predicted result of the model are bolded)

Algo	Avg. accuracy	Avg. precision	Avg. recall	Avg. f1-score	Avg. micro- precision	Avg. macro- precision
DT	0.990	0.99	0.99	0.99	0.99	0.98
RF	0.988	0.99	0.99	0.99	0.99	0.98
NN	0.979	0.98	0.98	0.98	0.98	0.95
KNN	0.809	0.78	0.81	0.87	0.81	0.54

 Table 10
 Results of prediction models—Input dataset (best predicted result of the model are bolded)

Algo	Avg. accuracy	Avg. precision	Avg. recall	Avg. f1-score	Avg. micro- precision	Avg. macro- precision
DT	0.984	0.98	0.98	0.98	0.98	0.96
RF	0.961	0.96	0.96	0.96	0.96	0.93
NN	0.953	0.95	0.95	0.95	0.95	0.67
KNN	0.881	0.88	0.88	0.88	0.88	0.64

 Table 11
 Results of prediction models—Understanding dataset (best predicted result of the model are bolded)

Algo	Avg. accuracy	Avg. precision	Avg. recall	Avg. f1-score	Avg. micro- precision	Avg. macro- precision
DT	0.984	0.98	0.98	0.98	0.98	0.96
RF	0.961	0.96	0.96	0.96	0.96	0.93
NN	0.953	0.95	0.95	0.95	0.95	0.67
KNN	0.881	0.88	0.88	0.88	0.88	0.64

results support our finding that the DT is the most suitable and accurate classifier for predicting learning styles.

The outcomes obtained from our proposed model offer valuable applications in various contexts. One significant application involves assisting learners in comprehending their learning styles, enabling them to identify their strengths and weaknesses in the learning process. By emphasizing their weaknesses, learners can adopt targeted strategies to enhance their learning efficiency and effectiveness (Li & Zhou, 2018). This self-awareness empowers learners to tailor their approaches and allocate resources more effectively, leading to optimized learning outcomes.

Additionally, the utilization of the proposed model, which is built upon the traces generated by learners' interactions with MOOC platforms, opens up possibilities for personalized learning path recommendations. By leveraging the identified learning styles of individuals, the model can suggest specific learning paths that align with their preferences and needs. These recommendations can include tailored instructions, activities, and resources that cater to the unique learning style of each individual learner. This personalized approach not only enhances learner engagement but also promotes effective learning by ensuring that the content and activities are relevant and meaningful to the learner's preferred style of learning (Li & Zhou, 2018).

Moreover, the introduction of the proposed model may provide valuable support to teachers and instructors in gaining a deep understanding of their learners' preferences. It enables them to customize their instructions and deliver content in a format that aligns with the specific learning styles of individual students, rather than employing a one-size-fits-all approach. As emphasized by Claxton and Murrel, (1987), instructors should be attuned to the diverse needs of learners and strategically design learning experiences that cater to their unique learning styles, ultimately fostering effective learning outcomes. By utilizing the insights generated by the model, teachers can gain a comprehensive view of their students' preferred learning styles. Armed with this knowledge, they can thoughtfully and systematically craft learning experiences that are tailored to meet the specific needs of each student. This personalized approach not only enhances student engagement but also optimizes the learning process by presenting content in a manner that resonates with the individual learning preferences of each student.

In the subsequent sections of this work, our focus will revolve around addressing the needs of teachers and instructors. Specifically, we aim to provide a solution that can assist educators in leveraging learning analytics to personalize the learning experiences of their students. To achieve this, we propose of a tool called MOOCLS (MOOC Learning Styles), that we explore in the following section.

From events to visual insights: leveraging MOOCLS for learning style visualization

Empowering stakeholders with actionable insights: overview

Visualization and dashboard tools have emerged as essential components in the field of education, offering a multitude of benefits and advantages (Susnjak et al., 2022). These tools enable stakeholders (students, educators, instructional designers, administrators) to gain a holistic understanding of student progress and performance, allowing for the identification of their strengths and weaknesses. By visualizing data such as assessment scores, attendance records, and engagement metrics, teachers can discern patterns and trends, leading to targeted interventions and personalized support (Martin & Sherin, 2013). This proactive approach to monitoring and analysis enhances the effectiveness of teaching strategies, ultimately fostering student success and achievement.

As discussed in section "Approaches to identify learning styles" visualization tools have the potential to significantly contribute to educational research by providing stakeholders with valuable insights and data-backed evidence for decision-making. By visualizing key performance indicators (KPI) and trends, stakeholders can identify systemic challenges, evaluate the effectiveness of interventions. The visual representation of data simplifies complex information, enabling stakeholders to comprehend and communicate the impact of their actions more effectively.

In line with leveraging visualization for educational research, MOOCLS (which stands for "MOOC Learning Styles") was developed based on an approach to the automatic identification of learning styles. In this section, we present the underlying motivation behind MOOCLS, its technical architecture, and describe the functionalities of our scheme and its validation from a user perspective.

The motivation behind MOOCLS stems from the recognition that understanding learners' preferences in learning styles can assist teachers in tailoring their pedagogical materials. Instead of assuming that all learners are alike and providing the same resources to everyone, identifying learning style preferences offers valuable information about individual needs. Learners who have less interest in using technology or who receive resources that are not aligned with their learning styles may feel frustrated in such an environment, increasing the risk of course dropout (Chang et al., 2015). Thus, personalized and adaptive learning can be an effective solution to maintain learners' interest.

The primary objective of MOOCLS is to empower teachers by providing them with a visualization tool that predicts and identifies participants' learning styles. This information enables teachers to intervene appropriately during the current or upcoming session, ensuring effective course improvement and the design of learning resources that best suit each learner's needs and characteristics.

The technical architecture of MOOCLS involves leveraging machine learning algorithms and data analysis techniques (Sect. "Decoding Learners' preferences: methodological approach to identifying learning styles in MOOCs"). By collecting and analyzing data on learners' interactions with the online course platform, MOOCLS generates predictions about their learning styles. These predictions are then presented to teachers through a user-friendly visualization tool.

The functionalities of MOOCLS include learner profile analysis, learning style prediction, and recommendations for adapting teaching strategies and resources. Teachers can gain insights into learners' preferences, strengths, and weaknesses, allowing them to make informed decisions regarding instructional approaches.

User interaction with MOOCLS

MOOCLS is a dynamic web application that is accessible via a browser and does not require prior installation. The overall functions of the platform are realized via PHP and MySQL, and we also used Highcharts JavaScript library, which is a tool for developing interactive charts. MySQL was used as a relational database engine to store the features of the learners and their predicted learning styles. This section presents a brief description of the features of MOOCLS in terms of user interaction.

After logging into MOOCLS, users are greeted with a page displaying a comprehensive list of available courses (Fig. 9). This empowers teachers and instructors to choose a particular course they are interested in and explore the distribution of learning styles among the enrolled learners. By selecting a specific course, they gain access to visualizations that depict how the learners' learning styles are distributed within that course.

The home page of MOOCLS provides users with an informative overview of the MOOC (Massive Open Online Course) data (Fig. 10). It presents KPIs such as the total number of weeks the course spans and the number of learners enrolled in the course. Additionally, the home page offers visualization capabilities to showcase the distribution of learners across each dimension of the FSLSM. This allows teachers to gain insights into how learners are distributed in terms of their learning preferences and styles. Furthermore, the home page provides valuable information regarding the number of certified learners, giving educators a clear understanding of the achievement levels within the course.

To delve into the data related to each dimension in greater detail, the teacher can access individual pages dedicated to each dimension. Let's take the example of the "Processing" dimension. By clicking on the "Processing" link in the side menu, the teacher can navigate to the page specifically designed for this dimension.

Once loaded, the page shows the distribution of learners is visually presented in relation to the features associated with the "Processing" dimension (Fig. 11). This visualization helps the teacher understand how learners are distributed across different aspects of their processing styles.

Moreover, the page allows for additional filtering options to display only the distribution of certified learners, providing insights into the performance of those who have obtained certification. In the visualization, each dot on the chart represents an individual learner and is uniquely identified by their learner ID. The learner ID serves as a reference point for further analysis and discussion. This level of granularity allows the teacher to examine individual learners' behaviors and characteristics within the "Processing" dimension, aiding in personalized instruction and support.

In addition to the aforementioned features, the page offers advanced customization options for the chart, allowing users to finely tailor the visualization to their specific needs. One such option is the ability to modify the X and Y axes, enabling the display of various metrics related to the learner. With this enhanced flexibility, educators can choose from a range of parameters to represent on the axes. This allows for analyzing how learners behave and position themselves in relation to specific variables and

All courses					CREATE COURSE
Courses	Sessions	From	То	# Weeks	
Statistical Learnin	g				
	Winter 2015	19 Jan. 2015	06 Apr. 2015	9	
Fig. 9 List of co	Winter 2016 DUITS	12 Jan. 2016	06 Apr. 2016	9	



Fig. 10 Statistics relating to each dimension

pedagogical materials, which can provide valuable insights for teaching and pedagogical adaptation.

By clicking on a node representing a particular learner, the teacher gains access to more detailed information about that learner within the specific dimension being analyzed. This enables a closer examination of the learner's behaviors, characteristics, and performance metrics in relation to that dimension (Preprocesing here).

The page is also providing teachers and instructional designers with a valuable set of recommendations aimed at improving their course design and teaching methods (Fig. 12). These recommendations are the result of a thorough review of pertinent literature, encompassing evidence-based practices within the realm of education.

These recommendations may encompass various aspects, including instructional strategies, assessment techniques, content delivery, and learner engagement. For instance, educators may be encouraged to incorporate visual hands-on activities for learners with a preference for visual processing. It is important to note that, as the field of education continually evolves, these recommendations are subject to ongoing refinement in the medium term. As new research emerges and practical insights are gathered, the recommendations will be further improved to ensure their effectiveness and relevance in optimizing course design and teaching methodologies.

For a more comprehensive view of an individual learner, clicking on the learner ID provides a global perspective (Fig. 13). However, it is important to note that all learner data is anonymized to ensure data privacy and protect sensitive information.



Fig. 11 Statistics relating to each dimension

Upon clicking the learner ID, a subsequent page would display a comprehensive overview of the selected learner's learning styles and features. This page would provide detailed insights into how the learner behaves, engages with the learning materials, and utilizes various resources.

Within this page, teachers can explore specific information such as the learner's preferred learning styles. It highlights whether the learner benefits more from visual/ verbal, active/reflective, sequential/global, sensitive/intuitive poles. This understanding allows teachers to adapt their instructional methods to better suit the learner's preferences. Additionally, the subsequent page reveals the learner's interactions with different types of content. It showcases how the learner engages with various learning



Fig. 12 Insight about learner's processing dimension

materials, such as videos, articles, quizzes, etc. This information helps educators identify the most effective resources and formats for engaging the learner.

Usability and utility evaluation of MOOCLS

The aim of this section is to evaluate the utility and usability of the MOOCLS as a learning styles visualization tool. In this part, we present the setup used for evaluation and discuss the results of this evaluation.



Fig. 13 Insight learners' learning styles

Evaluation setup

To assess our proposed tool, a survey study was performed, involving participants from various backgrounds: individuals with no prior experience in designing online courses, those who had previously created e-learning courses, and individuals who had participated in the development of Massive Open Online Courses (MOOCs).

The evaluation process focused on two main criteria. The first criterion was usability, which refers to the level of effectiveness, efficiency, and satisfaction experienced by users while using a product to accomplish specific goals (Tricot et al., 2003). Usability was evaluated based on five quality components: learnability, efficiency, memorability, errors, and satisfaction, following the framework established by Nielsen, (2003).

The second criterion was utility, which aimed to determine the tool's capability to provide a concise overview of essential information (Tricot et al., 2003). Utility focuses on the tool's ability to present key information in a clear and useful manner.

To gather participants for the study, a call for participation was circulated through multiple mailing lists, including info-ic, bull-i3, ATIEF, and AUF. Additionally, emails were specifically sent to teachers and researchers affiliated with the IRF-SIC Laboratory. Furthermore, we shared the invitation in relevant Facebook groups such as Analytics/Machine Learning/Data Mining and Business Intelligence and Big Data Maroc.

Despite the relatively small number of participants, consisting of 21 individuals, this sample size was considered adequate to identify any significant usability issues, following the guidelines proposed by Virzi (1992). Each participant was provided with a link to access the MOOCLS platform and was requested to complete a question-naire to assess the tool's performance.

After the participants had interacted with MOOCLS, they were asked to fill out a questionnaire containing 16 items, using an online Google Form and consisted of 16 items divided into different sections. The first section (Q1 to Q6) of the questionnaire collected personal data about the participants, such as their age, gender, educational background, and prior experience in pedagogical design for online courses. This section aimed to gather information about the participants' profiles and their levels of expertise in the subject matter.

The second section (Q7 to Q12) focused on assessing the utility of the MOOCLS tool. Participants were presented with specific questions aimed at evaluating the tool's ability to provide a concise overview of key information. This section aimed to measure participants' perceptions of the tool's effectiveness in presenting relevant information in a clear and concise manner.

The third section (Q13) evaluated the usability of the tool using the System Usability Scale (SUS) questionnaire developed by Bangor et al. (2009). The SUS is a widely used and validated scale that measures the participants' subjective perceptions of usability.

The last section (Q14 to Q16) consisted of open-ended questions where participants were given the opportunity to provide additional feedback. They were encouraged to share any recommendations for improvements and to report any bugs or issues they encountered while using the MOOCLS application. This section aimed to gather qualitative insights and suggestions from the participants to further enhance the tool's functionality and address any identified problems.

Evaluation results

The usability of the MOOCLS tool was evaluated using the System Usability Scale (SUS) questionnaire. The SUS questionnaire consists of 10 items, each rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

To calculate the SUS score, a specific formula was followed. For odd-numbered items (1, 3, 5, 7, 9), the score contribution was determined by subtracting one from the scale value. For even-numbered items (2, 4, 6, 8, 10), the score contribution was obtained by subtracting the scale value from five. The score contributions from all the items were then added together.

To obtain the final SUS score, the sum of the score contributions was multiplied by 2.5. This calculation transformed the overall score into a range from zero to 100, providing a standardized measure of usability for the MOOCLS tool.

To summarize, the SUS score is formulated as follows:

$$SUS_{score} = \left(\sum_{i=1}^{5} (5 - S_{2i}) + (S_{2i-1} - 1)\right) \times 2.5$$

The calculation of the SUS score can be influenced by misunderstandings of the negative statements in the questionnaire. Sauro and Lewis (2011) reported that13% of SUS questionnaires are prone to contain errors. In order to mitigate this issue, we implemented a preprocessing step for the participants' responses, following the guidelines presented by Mclellan et al. (2012). According to McLellan's grid, any response higher than three for negative statements was considered to be an error and was thus treated accordingly. This preprocessing step aimed to minimize the impact of misinterpretations and enhance the accuracy of the SUS score calculation.

Based on the obtained findings, it can be observed that the average SUS (System Usability Scale) score for the MOOCLS tool was 75.9, with a range of scores varying from 55 to 100. The standard deviation for the scores was calculated to be 14.2 (Fig. 14).

Referring to the guidelines provided by (Bangor et al., 2009) for interpreting SUS scores, it can be inferred that the MOOCLS tool falls into the category of being "acceptable" in terms of usability. Systems scoring below 50 are generally considered unacceptable, scores between 50 and 70 are marginally acceptable, and scores exceeding 70 are acceptable. This means that it has achieved a level of usability that is considered satisfactory. Additionally, with a SUS score above 70, the tool can be regarded as performing well, indicating a positive user experience.

This means that it has achieved a level of usability that is considered satisfactory. Additionally, with a SUS score above 70, the tool can be regarded as performing well, indicating a positive user experience.



An in-depth analysis of participant responses to questions Q7 to Q12 reveals highly positive feedback regarding the usefulness of the MOOCLS tool. All participants unanimously agreed that MOOCLS effectively aided in acquiring a profound understanding of learners' learning styles. They also acknowledged its value in identifying suitable activities and learning resources that could be recommended to learners, thereby enhancing their learning outcomes.

Furthermore, a significant majority of participants, specifically 19 out of 21, expressed that MOOCLS would facilitate the decision-making process related to their pedagogical strategies. This indicates that the tool offers practical insights and guidance to educators, enabling them to make informed choices regarding instructional approaches and techniques.

The participants found the information displayed in the charts to be necessary and relevant. This agreement can be attributed to the careful selection and organization of graphics, which were specifically tailored to the proposed case study. The choice and arrangement of visuals effectively conveyed pertinent information, enabling participants to comprehend and interpret the data more easily.

The overall positive feedback on MOOCLS' usefulness, its contribution to decisionmaking, and the relevance of the displayed information demonstrates the effectiveness of the tool in supporting educators' understanding of learners' needs and assisting them in making informed pedagogical choices. The alignment between participants' perceptions and the design of the tool affirms its value in the context of the proposed case study.

It is worth noting that the SUS score provides a quantitative measure of usability, but it is essential to complement it with qualitative feedback and further user evaluations to gain a comprehensive understanding of the tool's strengths and areas for improvement.

The results of our study indicated that the feedback received regarding the usability and utility of MOOCLS was predominantly positive and constructive. However, it is important to note that the tool is still in the experimental phase, and there were suggestions for improvement provided by the participants.

One valuable suggestion put forward by participants was the inclusion of lists of resources that would have the least impact on each learning style. This suggestion aims to provide educators with guidance on selecting resources that may be less influential for specific learning styles. Additionally, participants also recommended providing information on resources that have a higher priority for improvement, catering to the learning styles with the greatest need.

By incorporating these suggestions, MOOCLS can enhance its functionality and provide more targeted recommendations to educators. These improvements would allow for a more tailored and effective approach when it comes to resource selection and improvement strategies, aligning better with the diverse learning styles of individual students.

Discussion

Unlike traditional classrooms where learners often come from similar backgrounds, MOOCs attract learners from various geographical, cultural, educational, and socioeconomic backgrounds. Given this heterogeneity, tailoring teaching methodologies to each learner's unique characteristics becomes challenging. This make the one-size-fitsall approach less effective for MOOCs, and underscores the importance of designing tailored instructional strategies for such massive courses.

To meet this challenge, with a particular focus on enhancing MOOC teaching, and we have leveraged the theory of learning styles and the widely used FSLSM model. In this paper we analyzed digital traces from participants' interactions on the EDX platform, specifically focusing on the "Statistical Learning Stat" course offered by Stanford University during two sessions. Our study is distinctive in the research field, as few works analyzed MOOC databases, especially with such a large number of learners (Essa et al., 2023; Raleiras et al., 2022).

To achieve this goal, we proceeded with data pre-processing and feature selection, followed by grouping learners into clusters based on their preferences for each learning style using unsupervised machine learning techniques. Subsequently, to create labeled datasets representing each dimension of learning styles, a unique approach termed the "balance of learning styles" was employed. This method involved merging the two poles (learning styles) within each dimension to quantify the degree of dominance exhibited by each style. By employing this approach, we aimed to establish a comprehensive understanding of learners' preferences and their relative strengths within different learning style dimensions. Our study reveals that a predominant number of learners exhibited active, visual, sequential, and sensory learning styles. These findings align with the works of Felder and Silverman (1988).

We conducted a comparative analysis of four supervised machine learning algorithms: decision tree, random forest, k-nearest neighbor, and a neural network. The objective was to assess their performance in predicting learning styles based on the collected data. The results of the analysis revealed that the decision tree model exhibited exceptional accuracy, surpassing a threshold of 98% across all four datasets. This finding highlights the effectiveness of the decision tree algorithm in accurately predicting learners' learning styles based on the provided features. The high accuracy achieved by the decision tree model underscores its potential as a valuable tool for educators seeking to understand and cater to the diverse learning preferences of their students. Many studies in the literature support our finding, reinforcing the credibility of decision trees as a robust method for predicting diverse learning styles (Essa et al., 2023; Raleiras et al., 2022).

In our context, in order to implement the obtained model, we have developed MOOCLS, a dedicated visualization tool designed for MOOCs. MOOCLS serves as a powerful resource for teachers, enabling them to gain a comprehensive understanding of the diverse range of participants in a MOOC in terms of their learning styles. By utilizing this tool, educators can access valuable insights into their learners' preferences, thus facilitating the design and delivery of more customized course content that aligns with their specific learning styles.

There are a number of interesting directions to further extend our model in the future. Learners can benefit from model-driven recommendations ensuring they engage with materials and methods best suited to their individual styles. The model can also facilitate



Fig. 15 Connecting data stream with machine learning model

the personalization of learning contents and resources, allowing for more targeted and effective instruction. Furthermore, our model cas offers a promising direction for enhancing the MOOC experience through personalized learning paths.

Finally, we highlight that the data analyzed in this research was specifically obtained from the edX course "Statistical Learning," which focuses on Supervised Machine Learning and targets learners with a scientific background. While the findings and insights derived from this dataset are valuable, it's important to acknowledge that the underlying methodology was designed with genericity in mind, it can thus be easily applied to analyze the log of many other courses to enable deep understanding of student behaviors. To realize these benefits, the Fig. 15 illustrates a generic solution based on the proposed model. This solution can seamlessly integrate with diverse MOOC datasets, facilitating educational research and assisting instructors in improving course design.

To ensure the broader applicability of MOOCLS, it would be beneficial to apply the solution to datasets from MOOCs covering various themes, such as art & culture, literature, medicine, and more. By examining diverse datasets, we can gain a better understanding of how the tool performs across different subject areas and learner populations.

Conclusion

In this work, we have proposed a generic approach for predicting learners' learning styles based on their interactions with the MOOC platform. It relies in on tree major steps. This approach encompasses three major steps. In the first step, we aimed to extract and select features aligned with the Felder–Silverman Learning Style Model (FSLSM), as it one of the most widely adopted models in technology-enhanced learning (TEL). Using these features, an unsupervised clustering technique was applied to cluster learners according to their preferences for each learning style. In the second step, We evaluate four machine learning algorithms: decision tree, random forest, K-nearest neighbors, and a neural network. Our findings indicate that the decision tree algorithm achieves a high accuracy of over 98%. In final step, to operationalize these results, we developed MOOCLS, a visualization tool for MOOCs, which allows teachers to better understand the variety and diversity of the participants in a MOOC in terms of their learning styles. Using this tool, teachers and instructors can gain significant insight into their learners' preferences, allowing them to design more customized course content that matches their learners' learning styles.

Acknowledgements

We are grateful to CAROL(the center for advanced research through online learning), university of Stanford, for providing the Datasets necessary for accomplishing this research.

Author contributions

As Corresponding Author, I confirm that the manuscript has been read and approved for submission by all the named authors.

Funding

Not applicable.

Availability of data and materials

Preprocessed data that support the findings of this study are available from the corresponding author upon request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 2 June 2023 Accepted: 25 October 2023

Published online: 06 November 2023

References

- Ahmad, N., Tasir, Z., Kasim, J., & Sahat, H. (2013). Automatic detection of learning styles in learning management systems by using literature-based method. *Procedia - Social and Behavioral Sciences*, 103, 181–189. https://doi.org/10.1016/j. sbspro.2013.10.324
- Akçapınar, G., Altun, A., & Aşkar, P. (2019). Using learning analytics to develop early-warning system for at-risk students. International Journal of Educational Technology in Higher Education, 16(1), 1–20.
- Alshmrany, S. (2022). Adaptive learning style prediction in e-learning environment using levy flight distribution based CNN model. *Cluster Computing*, *25*(1), 523–536.
- Anantharaman, H., Mubarak, A., & Shobana, B.T. (2018). Modelling an adaptive e-learning system using LSTM and random forest classification. 2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (pp. 29–34).
- Assami, S., Daoudi, N., & Ajhoun, R. (2018). Personalization criteria for enhancing learner engagement in MOOC platforms. IEEE Global Engineering Education Conference (EDUCON), 2018, 1265–1272.
- Bajaj, R., & Sharma, V. (2018). Smart Education with artificial intelligence based determination of learning styles. *Procedia Computer Science*, 132, 834–842.
- Bakki, A., Oubahssi, L., Cherkaoui, C., & George, S. (2015). Motivation and engagement in MOOCs : How to increase learning motivation by adapting pedagogical scenarios? In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Éds.), Design for Teaching and Learning in a Networked World : 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15–18, 2015, Proceedings (pp. 556–559). Springer International Publishing. https://doi.org/10.1007/978-3-319-24258-3_58
- Bakki, A., Oubahssi, L., Cherkaoui, C., & George, S. (2016). cMOOC : How to assist teachers in integrating motivational aspects in pedagogical scenarios? *Stakeholders and Information Technology in Education: IFIP TC 3 International Conference, SaITE 2016, Guimarães, Portugal, July 5-8, 2016* (pp. 72–881). Revised Selected Papers 1.
- Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies, 4(3), 114–123.
- Barbe, W. B., Milone, M. N., & Swassing, R. H. (1988). *Teaching through modality strengths: Concepts and practices*. Zaner-Bloser.
- Beal, A. (2015). Description et sélection de données en grande dimension [PhD Thesis].
- Benabbes, K., Housni, K., Hmedna, B., Zellou, A., & Mezouary, A. E. (2023). Explore the influence of contextual characteristics on the learning understanding on LMS. *Education and Information Technologies*, 1–39.
- Blagojević, M., & Milosević, M. (2013). Collaboration and learning styles in pure online courses: An action research. *Journal of Universal Computer Science*, *19*(7), 984–1002.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (pp. 93–104).

Brown, S. (2013). Back to the future with MOOCs. ICICTE 2013 Proceedings, 3, 237-246.

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction, 6(2–3), 87–129. https://doi.org/10.1007/BF00143964

- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3(1), 1–27.
- Carver, C. A., Howard, R. A., & Lane, W. D. (1999). Enhancing student learning through hypermedia courseware and incor-
- poration of student learning styles. *IEEE Transactions on Education*, *42*(1), 33–38. https://doi.org/10.1109/13.746332 Chevrier, J., Fortin, G., Théberge, M., & Le Blanc, R. (2000). Le style d'apprentissage: Une perspective historique. *Le style d'apprentissage*, *28*(1).

Chuang, I., & Ho, A. D. (2016). HarvardX and MITx: Four Years of Open Online Courses - Fall 2012-Summer 2016.

- Cid, F. M., Ferro, E. F., Muñoz, H. D., & Contreras, L. V. (2018). Learning styles in physical education. Advanced Learning and Teaching Environments: Innovation, Contents and Methods, 243.
- Claxton, C. S., & Murrell, P. H. (1987). *Learning styles : Implications for improving educational practices*. ASHE-ERIC Higher Education Report No. 4, 1987. ERIC.

Coffield, F., Moseley, D., Hall, E., Ecclestone, K., & others. (2004). Learning styles and pedagogy in post-16 learning: A systematic and critical review. Learning and Skills Research Centre London. http://www.voced.edu.au/content/ngv:13692

Crockett, K., Latham, A., & Whitton, N. (2017). On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees. *International Journal of Human-Computer Studies*, *97*, 98–115.

Dağhan, G., & Akkoyunlu, B. (2012). An examination through conjoint analysis of the preferences of students concerning online learning environments according to their learning styles. *International Education Studies, 5*(4). https://avesis. hacettepe.edu.tr/yayin/0608384c-2f90-49f7-a52d-0e93d5976cfa/an-examination-through-conjoint-analysis-of-the-preferences-of-students-concerning-online-learning-environments-according-to-their-learning-styles

- Dekker, S., Lee, N. C., Howard-Jones, P., & Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, *3*, 429.
- Dung, P. Q., & Florea, A. M. (2012). An approach for detecting learning styles in learning management systems based on learners' behaviours. *International Conference on Education and Management Innovation, 30*, 171–177.

Dunn, R. S., Dunn, K. J., & Price, G. E. (1979). *Learning style inventory*. Price Systems Lawrence.

Duong, M. K. (2019). Automated architecture-modeling for convolutional neural networks. *BTW 2019-Workshopband*. Durand, G., Laplante, F., & Kop, R. (2011). A learning design recommendation system based on markov decision processes.

- KDD-2011: 17th ACM SIGKDD conference on knowledge discovery and data mining.
- Elias, T. (2011). Learning analytics: Definitions, processes and potential.
- Essa, S. G., Celik, T., & Human-Hendricks, N. (2023). Personalised adaptive learning technologies based on machine learning techniques to identify learning styles : A systematic literature review. *IEEE Access*. https://ieeexplore.ieee.org/ abstract/document/10124747/
- Fasihuddin, H., Skinner, G., & Athauda, R. (2014). Towards an adaptive model to personalise open learning environments using learning styles. In Proceedings of International Conference on Information, Communication Technology and System (ICTS) 2014 (pp. 183–188).
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *Al Magazine*, 17(3), 37–37.

Felder, R., & Solomon, B. (2006). Index of Learning Styles Questionnaire (2009). North Carolina State University. Online at: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS-a.htm.

- Felder, R. M. (1996). Matters of style. ASEE Prism, 6(4), 18–23.
- Felder, R. M., & Brent, R. (2005). Understanding student differences. Journal of Engineering Education, 94(1), 57–72.
- Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education*, 78(7), 674–681.
- Feldman, J., Monteserin, A., & Amandi, A. (2015). Automatic detection of learning styles: State of the art. Artificial Intelligence Review, 44(2), 157–186.
- Ferreira, L. D., Spadon, G., Carvalho, A. C., & Rodrigues, J. F. (2018). A comparative analysis of the automatic modeling of learning styles through machine learning techniques. *IEEE Frontiers in Education Conference (FIE), 2018*, 1–8.
 Fischer, B. B., & Fischer, L. (1979). Styles in teaching and learning. *Educational Leadership, 36*(4), 245–254.
- Ford, N., & Chen, S. Y. (2001). Matching/mismatching revisited: An empirical study of learning and teaching styles. British Journal of Educational Technology, 32(1), 5–22. https://doi.org/10.1111/1467-8535.00173
- Garcia, P., Amandi, A., Schiaffino, S., & Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*, 49(3), 794–808.
- Graf, S., & Kinshuk, K. (2007). Providing adaptive courses in learning management systems with respect to learning styles. E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (pp. 2576–2583).
- Graf, S. (2007). Adaptivity in learning management systems focussing on learning styles. na.
- Graf, S., Liu, T.-C., & others. (2008). Identifying learning styles in learning management systems by using indications from students' behaviour. Advanced Learning Technologies, 2008. Eighth IEEE International Conference on ICALT'08 (pp. 482–486). http://ieeexplore.ieee.org/abstract/document/4561743/
- Graf, S., Kinshuk, Zhang, Q., Maguire, P., & Shtern, V. (2012). Facilitating learning through dynamic student modelling of learning styles. In *Towards learning and instruction in Web 3.0* (pp. 3–16). Springer. https://doi.org/10.1007/978-1-4614-1539-8_1
- Graf, S., & Liu, T.-C. (2009). Supporting teachers in identifying students' learning styles in learning management systems : An automatic student modelling approach. *Journal of Educational Technology & Society, 12*(4), 3.
- Grasha, A. F. (1984). Learning styles: The journey from Greenwich Observatory (1796) to the college classroom (1984). Improving College and University Teaching, 32(1), 46–53.
- Halawa, M. S., Shehab, M. E., & Hamed, E. M. R. (2015). Predicting student personality based on a data-driven model from student behavior on LMS and social networks. *Fifth International Conference on Digital Information Processing and Communications (ICDIPC), 2015,* 294–299.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society Series C (applied Statistics), 28(1), 100–108.
- Hasibuan, M. S., Nugroho, L. E., Santosa, P. I., & Kusumawardani, S. S. (2016). A proposed model for detecting learning styles based on agent learning. *International Journal of Emerging Technologies in Learning (IJET)*, *11*(10), 65. https://doi.org/10.3991/ijet.v11i10.5781

Hmedna, B., El Mezouary, A., & Baz, O. (2019). How does learners' prefer to process information in MOOCs ? A data-driven study. *Procedia Computer Science*, 148, 371–379.

Hmedna, B., El Mezouary, A., & Baz, O. (2020). A predictive model for the identification of learning styles in MOOC environments. *Cluster Computing*, 23(2), 1303–1328.

Jackson, G. A. (1990). Evaluating learning technology: Methods, strategies, and examples in higher education. *The Journal of Higher Education*, 61(3), 294–311.

Karagiannis, I., & Satratzemi, M. (2018). An adaptive mechanism for Moodle based on automatic detection of learning styles. Education and Information Technologies, 23, 1331–1357.

Keefe, J. W. (1979). Learning style: An overview. Student Learning Styles: Diagnosing and Prescribing Programs, 1, 1–17. Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies. MIT Press.

Kirschner, P. A. (2017). Stop propagating the learning styles myth. Computers & Education, 106, 166–171.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Recommending self-regulated learning strategies does not improve performance in a MOOC. In *Proceedings of the third (2016) ACM conference on learning@ scale* (pp. 101–104).

Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs (pp. 60–65).

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in K-means clustering. *International Journal*, 1(6), 90–95.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. Kolb, D. A. (1976). *Learning style inventory technical manual*. McBer.

Kolekar, S. V., Pai, R. M., & MM, M. P. (2017). Prediction of learner's profile based on learning styles in adaptive E-learning system. International Journal of Emerging Technologies in Learning, 12(6).

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3), 62–63.

Larusson, J. A., & White, B. (2014). Learning analytics: From research to practice (Vol. 13). Springer.

Latham, A., Crockett, K., McLean, D., & Edmonds, B. (2012). A conversational intelligent tutoring system to automatically predict learning styles. *Computers & Education*, *59*(1), 95–109. https://doi.org/10.1016/j.compedu.2011.11.001

LeBlanc, T. R. (2018). Learning styles: Academic fact or urban myth? A recent review of the literature. https://digital.library.txsta te.edu/handle/10877/7914

Li, Y., Yu, M., Xu, M., Yang, J., Sha, D., Liu, Q., & Yang, C. (2020). Big Data and cloud computing. In *Manual of digital earth* (pp. 325-355). Springer.

Li, C., & Zhou, H. (2018). Enhancing the efficiency of massive online learning by integrating intelligent analysis into MOOCs with an application to education of sustainability. *Sustainability*, *10*(2), 468.

Li, J., Han, S., & Fu, S. (2019). Exploring the relationship between students' learning styles and learning outcome in engineering laboratory education. *Journal of Further and Higher Education*, 43(8), 1064–1078. https://doi.org/10.1080/ 0309877X.2018.1449818

- Li, L. X., & Abdul Rahman, S. S. (2018). Students' learning style detection using tree augmented naive Bayes. *Royal Society* Open Science, 5(7), 172108.
- Li, Y., Fu, C., & Zhang, Y. (2017). When and who at risk? International Educational Data Mining Society.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. Eighth IEEE International Conference on Data Mining, 2008, 413–422.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. *IEEE International Conference on Data Mining*, 2010, 911–916.

Maraza-Quispe, B., Alejandro-Oviedo, O., Cisneros-Chavez, B., Cuentas-Toledo, M., Cuadros-Paz, L., Fernandez-Gambarini, W., Quispe-Flores, L., & Caytuiro-Silva, N. (2019). Model to personalize the teaching-learning process in virtual environments using case-based reasoning. In *Proceedings of the 2019 11th International Conference on Education Technology and Computers* (pp. 105–110).

Marosan, Z., Savic, N., Klasnja-Milicevic, A., Ivanovic, M., & Vesin, B. (2022). Students' perceptions of ils as a learning-styleidentification tool in e-learning environments. *Sustainability*, 14(8), 4426.

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning : An introduction to the special issue. *Journal of the Learning Sciences*, *22*(4), 511–520.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2019). CRISP-DM twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061.

El Mawas, N., Gilliot, J.-M., Garlatti, S., Euler, R., & Pascual, S. (2018). Towards personalized content in massive open online courses. *10th International Conference on Computer Supported Education*.

Mazzola, L., & Mazza, R. (2009). Toward adaptive presentations of student models in eLearning Environments. *AIED*, 761–762.

Mclellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on system usability scale ratings. *Journal of Usability Studies*, 7(2), 56–67.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist*, 19(2), 59–74.

Mubarak, A. A., Cao, H., Hezam, I. M., & Hao, F. (2022). Modeling students' performance using graph convolutional networks. Complex & Intelligent Systems, 8(3), 2183–2201.

Newton, P. M., & Miah, M. (2017). Evidence-based higher education-is the learning styles 'myth' important? *Frontiers in Psychology*, *8*, 444.

Nielsen, J. (2003). Introduction to usability. Retrieved November, 14, 2014.

Onah, D. F., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioural patterns. EDU-LEARN14 Proceedings, 1, 5825–5834.

Oxford, R. (2001). Language learning styles and strategies. *Teaching English as a Second or Foreign Language*.

- Oxford, R. (2003). Language learning styles and strategies: Concepts and relationships. *IRAL International Review of Applied Linguistics in Language Teaching*, 41(4), 271–278. https://doi.org/10.1515/iral.2003.012
- Özpolat, E., & Akar, G. B. (2009). Automatic detection of learning styles for an e-learning system. *Computers & Education*, 53(2), 355–367.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest*, *9*(3), 105–119. https://doi.org/10.1111/j.1539-6053.2009.01038.x

Peacock, M. (2001). Match or mismatch? Learning styles and teaching styles in EFL. International Journal of Applied Linguistics, 11(1), 1–20. https://doi.org/10.1111/1473-4192.00001

Petty, L. M. (2004). An exploratory study of learning styles and environmental preferences of accelerated undergraduate and non-accelerated graduate adult students in a degree program at a small Catholic liberal arts college in the mid-Atlantic region [PhD Thesis]. Wilmington College (Delaware).

Popescu, E., Trigano, P., & Badica, C. (2007). Adaptive educational hypermedia systems: A focus on learning styles. *EURO-CON 2007-The International Conference on*" *Computer as a Tool*" (pp. 2473–2478).

Pratsri, S., Nilsook, P., & Wannapiroon, P. (2022). Synthesis of data science competency for higher education students. International Journal of Education and Information Technologies, 16, 101–109.

Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., Villarroel, R., & Cechinel, C. (2020). A learning analytics approach to identify students at risk of dropout: A case study with a technical distance education course. *Applied Sciences*, 10(11), 3998.

Raleiras, M., Nabizadeh, A. H., & Costa, F. A. (2022). Automatic learning styles prediction: A survey of the State-of-the-Art (2006–2021). *Journal of Computers in Education, 9*(4), 587–679.

Ramaswami, G., Susnjak, T., Mathrani, A., & Umer, R. (2022). Use of predictive analytics within learning analytics dashboards: A review of case studies. *Technology, Knowledge and Learning*, 1–22.

Rasheed, F., & Wahid, A. (2021). Learning style detection in E-learning systems using machine learning techniques. *Expert Systems with Applications, 174*, 114774.

Rebala, G., Ravi, A., & Churiwala, S. (2019). Clustering. In G. Rebala, A. Ravi, & S. Churiwala (Éds.), *An introduction to machine learning* (pp. 67–76). Springer International Publishing. https://doi.org/10.1007/978-3-030-15729-6_6
 Reid, G. (2005). Learning styles and inclusion. *Learning Styles and Inclusion*, 1–192.

Reinert, H. (1976). One picture is worth a thousand words? Not necessarily! *Modern Language Journal*, 160–168. Riener, C., & Willingham, D. (2010). The myth of learning styles. *Change: the Magazine of Higher Learning*, 42(5), 32–35. Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews:*

Data Mining and Knowledge Discovery, 7(1), e1187.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.

- Sadeghi, N., Kasim, Z. M., Tan, B. H., & Abdullah, F. S. (2012). Learning styles, personality types and reading comprehension performance. *English Language Teaching*, *5*(4), 116–123.
- Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2215–2224).
- Schofield, M. (2021). Exploring datafication for teaching and learning development: A higher education perspective. In Fostering Communication and Learning With Underutilized Technologies in Higher Education (pp. 79-92). IGI Global.

Scott, E., Rodríguez, G., Soria, Á., & Campo, M. (2014). Are learning styles useful indicators to discover how students use Scrum for the first time? *Computers in Human Behavior*, *36*, 56–64.

- Sheeba, T., & Krishnan, R. (2018). Prediction of student learning style using modified decision tree algorithm in e-learning system. In Proceedings of the 2018 International Conference on Data Science and Information Technology (pp. 85–90).
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 759.

Slack, N., & Norwich, B. (2007). Evaluating the reliability and validity of a learning styles inventory: A classroom-based study. Educational Research, 49(1), 51–63.

Song, W., & Wang, Z. (2023). Improved Clustering Strategies for Learning Style Identification in Massive Open Online Courses. Data Mining and Big Data: 7th International Conference, DMBD 2022, Beijing, China, November 21–24, 2022, Proceedings, Part I (pp. 240–254).

Suganya, A., & Sheshasaayee, A. (2022). An analysis of the influential learning styles in teaching and learning inclined towards learners. *International Journal of Early Childhood Special Education*, 14(6).

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education, 19*(1), 12.

Swai, C. T., Liu, Q., & Wu, L. (2023). Teachers' professional growth on SCT strategies: Mwalimu hub MOOC analytics. Interactive Learning Environments, 1–15.

Tharwat, A. (2018). Classification assessment methods. Applied Computing and Informatics.

Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité: Interpréter les relations entre trois dimensions de l'évaluation des EIAH.

Villaverde, J. E., Godoy, D., & Amandi, A. (2006). Learning styles' recognition in e-learning environments with feed-forward neural networks. *Journal of Computer Assisted Learning*, 22(3), 197–206. https://doi.org/10.1111/j.1365-2729.2006. 00169.x

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, 34(4), 457–468.

Wengrowicz, N., Lavi, R., Kohen, H., & Dori, D. (2022). Modeling with real-time informative feedback: Implementing and evaluating a new massive open online course component. *Journal of Science Education and Technology*, 1–14.

- Williams, J. J., Rafferty, A. N., Maldonado, S., Ang, A., Tingley, D., & Kim, J. (2017). MOOClets : A framework for dynamic experimentation and personalization. In Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale (pp. 287–290). https://doi.org/10.1145/3051457.3054006
- Wood, W. B. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental*, 25, 93–112.
- Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M., & Jakobs, H. (2015). The state of MOOCs from 2008 to 2014 : A critical analysis and future visions. In S. Zvacek, M. T. Restivo, J. Uhomoibhi, & M. Helfert (Éds.), Computer supported education (pp. 305–327). Springer International Publishing. https://doi.org/10.1007/978-3-319-25768-6_20
- Zhang, H., Huang, T., Liu, S., Yin, H., Li, J., Yang, H., & Xia, Y. (2020). A learning style classification approach based on deep belief network for large-scale online education. *Journal of Cloud Computing*, 9(1), 26. https://doi.org/10.1186/ s13677-020-00165-y

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- ► Rigorous peer review
- Open access: articles freely available online
- ► High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at > springeropen.com