algorithms

MDPI

Article

Learning Analytics in a Non-Linear Virtual Course

Jhon Mercado [1], Carlos Mendoza-Cardenas [2], Luis Fletscher [1,*] and Natalia Gaviria-Gomez [1]

[1] Faculty of Engineering, Universidad de Antioquia, Medellin 050010, Colombia; jfredy.mercado@udea.edu.co (J.M.); natalia.gaviria@udea.edu.co (N.G.-G.)
[2] Twitch Interactive Inc., San Francisco, CA 94104, USA; cmendoza@udel.edu
* Correspondence: luis.fletscher@udea.edu.co

Abstract: Researchers have extensively explored learning analytics in online courses, primarily focusing on linear course structures where students progress sequentially through lessons and assessments. However, non-linear courses, which allow students to complete tasks in any order, present unique challenges for learning analytics due to the variability in course progression among students. This study proposes a method for applying learning analytics to non-linear, self-paced MOOC-style courses, addressing early performance prediction and online learning pattern detection. The novelty of our approach lies in introducing a personalized feature aggregation that adapts to each student's progress rather than being defined at fixed timelines. We evaluated three types of features—engagement, behavior, and performance—using data from a non-linear large-scale Moodle course designed to prepare high school students for a public university entrance exam. Our approach predicted early student performance, achieving an F1-score of 0.73 at a 20% cumulative weight assessment. Feature importance analysis revealed that performance and behavior were the strongest predictors, while engagement features, such as time spent on educational resources, also played a significant role. In addition to performance prediction, we conducted a clustering analysis that identified four distinct online learning patterns recurring across various cumulative weight assessments. These patterns provide valuable insights into student behavior and performance and have practical implications, enabling educators to deliver more personalized feedback and targeted interventions to meet individual student needs.

Keywords: learning analytics; MOOCs; non-linear courses; self-paced learning; student performance prediction; cluster analysis; moodle; learning management systems (LMS); educational data mining; machine learning in education

check for updates

Academic Editor: Francesc Pozo

Received: 2 April 2025
Revised: 2 May 2025
Accepted: 8 May 2025
Published: 13 May 2025

Citation: Mercado, J.; Mendoza-Cardenas, C.; Fletscher, L.; Gaviria-Gomez, N. Learning Analytics in a Non-Linear Virtual Course. *Algorithms* **2025**, *18*, 284. https://doi.org/10.3390/a18050284

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The adoption of information and communication technologies in educational settings has steadily increased. Among these technologies, Learning Management Systems (LMS) have gained significant importance [1]. These platforms serve as versatile tools that aid in delivering, documenting, monitoring, and administering educational courses in virtual settings [2]. They play a crucial role in transforming virtual and face-to-face courses by offering tools for content management, communication, assignment submissions and assessments, online quizzes, and student grading [3].

Within the diverse applications of LMS platforms, a growing number of educators are utilizing them to provide Massive Open Online Courses (MOOCs). These courses attract large audiences through open access and asynchronous delivery, allow learners to enroll freely and progress at their own pace, and generally involve minimal or no direct

interaction with instructors. Despite their adaptability and general outreach, MOOCs and other extensive virtual courses regularly face challenges with low completion rates. Research indicates that fewer than 13% of students finish these courses, with only 2–10% achieving the course objectives and receiving certificates [4]. The issue of student dropout has collected significant research interest, leading to various interpretations and proposed solutions. Some researchers attribute dropout to students' lack of motivation, commitment, and initial intent [5]. Conversely, others emphasize the necessity for high autonomy, which can promote feelings of isolation among participants [6]. Researchers have proposed numerous adaptive, personalized, and recommendation strategies to reduce dropout rates by considering learning pathways [7], prior knowledge [8], skills [9], learning styles [10], and language support [11]. Additionally, innovative pedagogical approaches such as enquiry-based learning have demonstrated potential in enhancing student engagement and persistence in online learning environments [12].

Educators can implement retention and academic support strategies more effectively when they identify students at risk of dropping out early in an online course. Although researchers have introduced various methods for predicting student performance [13–21], these studies typically use data from courses that follow a strict sequence and schedule for lessons and assessments. Conversely, our research concentrates on what we refer to as "non-linear" virtual courses, also identified as "learner-paced", "self-paced", or "self-study" [22], where students can navigate through the materials at their speed and in any preferred order. Assignments may be completed at any time and in any sequence. Given that not all course assessments carry the same weight, these non-linear virtual courses present a unique challenge in the early prediction of student performance, as the percentage of assessments completed at any specific moment can vary significantly among students.

Studying non-linear or self-paced courses presents challenges different from those in traditional instructor-paced environments. In these courses, students control the sequence and timing of their activities, leading to highly variable engagement patterns. While beneficial for accommodating diverse needs, this flexibility makes monitoring progress and predicting outcomes more difficult. As self-paced learning becomes more common, developing methods to understand and support student success in these settings is increasingly critical.

Given this context, our contributions and findings are as follows:

- An adaptive feature extraction method tailored for interaction data from non-linear virtual courses. We introduce the concept of a Feature Aggregation Time Point (FATP), where features are extracted based on the student's Cumulative Weight Assessment (CWA).
- Feature engineering from the interaction logs of 663 students in a non-linear Moodle course. We define engagement, behavior, and performance features.
- Multiple experiments on student performance prediction that show the predictive power of our features and the effectiveness of our feature extraction method.
- A clustering and correlation analysis of students' online learning patterns that provide insights into the relationship between students' interactions with the online course and students' performance.
- A feature importance analysis that reveals key interaction features that significantly impact the model's accuracy in predicting student performance, and that can improve the identification of interaction patterns through clustering analysis.

2. Related Works

Improving students' academic performance remains a persistent challenge in academia. Accurate prediction of student outcomes and analysis of their digital footprints in online

courses can inform timely and targeted interventions, enhancing the learning process. Two complementary research areas in the literature address these goals: early student performance prediction, which focuses on forecasting academic outcomes from interaction data, and online learning pattern analysis, which seeks to uncover the engagement behaviors underlying different learner profiles. This section reviews both areas, leveraging findings from studies based on Learning Management System (LMS) course records.

2.1. Students' Performance Prediction

Numerous researchers have explored student performance prediction, yet no standardized set of features exists for solving this problem. The features used often depend on the specifics of the course, the LMS platform, or additional available student data. Tomasevic et al. [15] addressed both regression and classification problems for predicting student performance using various machine learning methods, such as K-Nearest Neighbor (K-NN), Support Vector Machine (SVM), Logistic Regression (LR), and others. They employed different feature combinations, including demographic, performance, and engagement data, using the Open University Learning Analytics Dataset (OULAD) from two courses [23]. The classification problem involved predicting whether a student would pass or fail the course after each of the six mid-term assessments. As the assessments progressed, the F1-score of the models improved, starting at 78% for the first assessment and reaching 94.9% for the sixth. The researchers found that demographic data did not significantly impact prediction accuracy. Similarly, a more recent study [17] also utilizing OULAD and other datasets confirmed that demographic features offer little added value when performance or activity data is available.

Researchers in early student performance prediction have utilized various methods to establish the Prediction Time Point (PTP), the precise moment a prediction is generated. For instance, Riestra-González et al. [18] defined the PTP at 10%, 25%, 33%, and 50% of the course duration, using the first and last login timestamps on the platform as reference points to calculate these percentages across the student cohort. Similarly, Conijn et al. and Waheed et al. [13,16] divided the course into weekly segments, updating features each week. Conversely, Tomasevic et al., Hoq et al., and Adnan et al. [14,15,19] based the PTP on specific assignments, updating features as assignments became available.

Defining the PTP by specific assignments is impractical for non-linear courses, as all assignments are continuously available and can be completed in any order or repeated. Similarly, using course duration or weekly intervals for the PTP is unsuitable. While the course has set start and end dates, students can progress at their own pace. Some may take the entire course duration to finish, while others may complete it in a few weeks.

Although Waheed et al. [16] claim to perform early predictions in self-paced education by defining the PTP through weekly course divisions, this approach is actually more suitable for linear courses. The OULAD dataset used in their study features course modules released sequentially over time, implying a linear progression [23]. This inherent structure justifies dividing the course into weekly segments for analysis, as student interactions are expected to follow a linear pattern throughout the course duration.

2.2. Online Learning Pattern Analysis

An important aspect of analyzing student learning patterns is identifying the most influential features that affect academic performance. To quantify the impact of these features, researchers can employ techniques such as SHapley Additive exPlanations (SHAP) or examine the model's parameters. For example, Hoq et al. [19] and Rohani et al. [20] utilized SHAP to reveal that engagement frequency and submission times, and interaction timing and quiz scores, respectively, are the most significant features driving academic

performance. Similarly, studies analyzing the parameters of models like Random Forest and Logistic Regression have shown that second-semester grades and exam registration counts are important in predicting performance [24].

In addition to feature importance, clustering techniques have been used to categorize students into distinct interaction profiles. Riestra et al. [18] applied agglomerative clustering to identify various student groups, from high performers to minimally engaged learners. Similarly, Cenka et al. [25] showed that clustering based on student engagement with course materials could highlight different performance levels across groups.

Other research has demonstrated strong correlations between clustered learning profiles and student performance. Bessadok et al. [26] identified three distinct clusters where students with higher levels of interaction achieved better academic outcomes. Similarly, other research has reinforced this relationship, showing that students who fall into specific clusters, such as those with frequent early submissions, tend to outperform others [20].

While many studies highlight influential features affecting academic performance, it is necessary to address the process of selecting these features beyond explanations. For example, although Hoq et al. [19] and Rohani et al. [20] utilize SHAP to identify significant attributes, they do not retrain their models to validate these findings. In contrast, Riestra et al. [18] effectively reduce features through agglomerative clustering, strengthening the reliability of their student interaction profiles. This gap highlights the need for future research to combine feature selection with model retraining for a more robust analysis of student learning patterns.

In summary, while prior research has made significant advances in predicting student performance and analyzing learning behaviors, several gaps remain. First, there is no standardized approach to feature selection or validation across studies, limiting the generalizability of findings. Second, most methodologies assume a linear course structure, making them less applicable to non-linear, self-paced environments. Finally, the literature often treats performance prediction and pattern analysis as separate areas, overlooking potential synergies between these perspectives. Addressing these shortcomings is crucial for developing more robust and generalizable models of online student behavior.

3. Methodology

We present in Figure 1 a high-level overview of our methodology, from the data preprocessing steps to the prediction of the student performance and analysis of their online learning patterns. In the next subsections, we provide more details about our methodology.

3.1. Moodle Sessions

Moodle, a popular open-source LMS, logs a wide range of user activity, referred to as clickstream data. This data captures interactions within the platform, including page visits, resource views, forum posts, and various other actions [27]. However, these raw data are not directly suitable for predictive modeling and must first be transformed into a more structured format for analysis [28]. Below, we provide a brief overview of the key Moodle database tables that store this user activity data.

The primary table for storing clickstream data is `mdl_logstore_standard_log`, which logs user activities through several key columns, like the name of the event (e.g., `\core\event\user_loggedin`), the associated component (e.g., `core`, `mod_quiz`), the user ID, and the event timestamp.

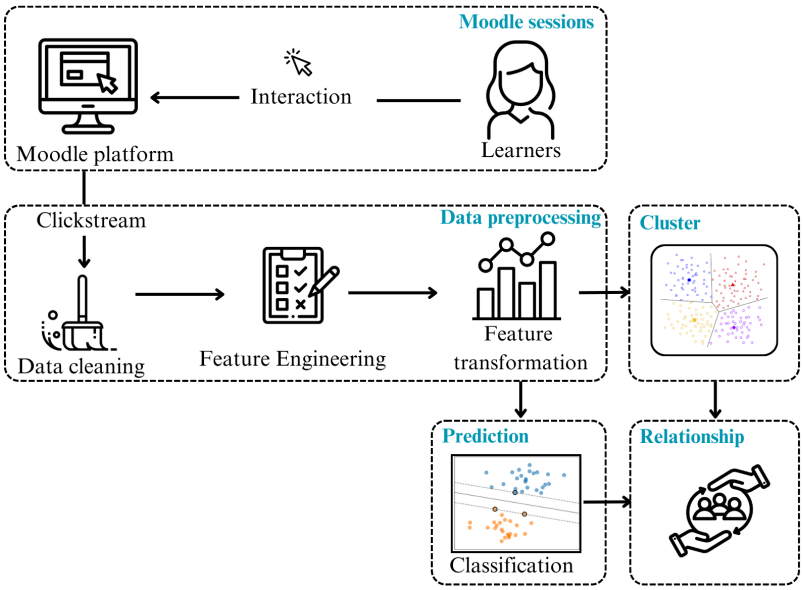


Figure 1. Overview of the proposed method. The icons used in this figure are sourced from Flaticon, designed by various authors, and licensed under Creative Commons (CC BY 3.0, accessed on 15 April 2025, https://www.flaticon.com).

Quiz and lesson data in Moodle are stored across various tables. `mdl_quiz_attempts` table logs quiz attempts, including details such as user ID, quiz ID, attempt number, completion time, and grade. The `mdl_lesson_timer` table tracks lesson durations; recording start time and time spent on lessons. Finally, the `mdl_lesson_grades` table stores lesson grades, capturing user ID, lesson ID, grade, and completion status.

3.2. Data Preprocessing

As noted in Section 3.1, the raw data stored in Moodle is not immediately suitable for predictive modeling due to its high volume, complexity, and lack of a structured format. Consequently, preprocessing is essential to transform the data into a matrix with dimensions $M \times N$, where M represents the number of students and N is the number of features for each student. This preprocessing comprises data cleaning and feature extraction steps, which are detailed below.

3.2.1. Data Cleaning

Data cleaning involves the selection of relevant tables and columns from the Moodle database and then extracting valuable records while filtering out irrelevant ones. Initially, we focused on the primary tables that record students' interactions within the platform, selecting four key tables (e.g., `logstore_standard_log`). We extracted columns related to student interactions and grades from these tables. For example, in the `logstore_standard_log` table, we excluded the `objecttable` column, as it did not contribute to our analysis. This step ensured that only essential data necessary for predictive modeling was retained, optimizing the data's volume and relevance.

3.2.2. Feature Engineering

Extracting and computing features from Moodle logs relevant to student performance is necessary to facilitate predictive modeling and clustering [15]. For the M students in our dataset, we derived N descriptors for each student.

We define two key concepts: (1) the Cumulative Weight Assessment (CWA) as the running total of the percentages of the assessments that a student has taken at any given

point in time; and (2) the Feature Aggregation Time Point (FATP) as the specific point in time where the activity of the user is aggregated to compute the features.

Given the flexible nature of the course, where students progress at their own pace, retake assessments, and complete lessons in any order, students may reach varying levels of CWA at a given time. To handle this variability, we define the FATP of each student as the point in time where the student's CWA remains below or equal to CWA_U, a given upper bound on the CWA. In our analysis, we use different values of CWA_U.

For better clarity, Figure 2 provides an illustrative example of the CWA progression for several students. By the end of the course, there is variability in student completion rates, with some students finishing all assessments (100% completion) and others not, as well as differences in certification outcomes, with some students receiving a certificate of participation and others not. In this example, CWA_U is set to 50%, and each student's FATP is determined individually based on when their CWA reaches or remains below this threshold. For instance, Student 3 's FATP is set at t_3 when their CWA first reaches 50%. In contrast, for Student 4, the FATP is calculated at 40% CWA because their next evaluative assessment pushes their CWA beyond the 50% threshold.

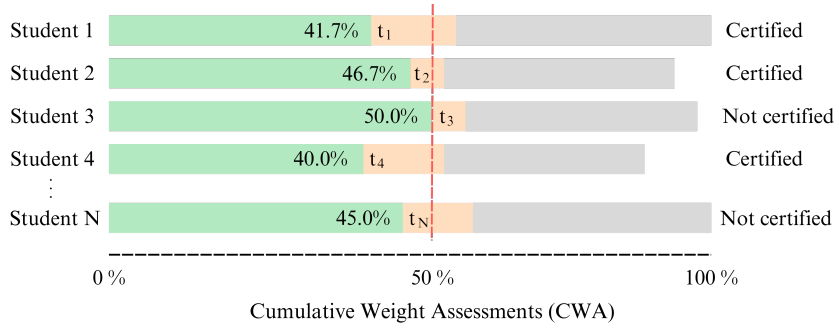


Figure 2. Example of the adaptive feature update strategy. The green bars represent the Cumulative Weighted Assessment (CWA) of each student before reaching the CWA_U threshold, with the corresponding Feature Aggregation Time Point (FATP), t_i, indicated on the right. The orange bar shows the following assessment completed by the student, demonstrating that the CWA will exceed the threshold. The grey bar displays the final CWA after course completion and the student's participation certificate status [29].

3.2.3. Feature Transformation

After computing the FATP for each student for a given CWA_U, we applied z-score normalization to ensure that all features have zero mean and unit variance. This standardization helps stabilize the learning process by creating a more symmetric loss surface, accelerating model convergence during training [30]. Once the dataset is prepared and normalized, it is ready for both predictive modeling and clustering analysis.

3.3. Prediction

The prediction task is formulated as a binary classification problem. $\mathcal{D} = \{\mathbf{x}_i, y_i\}_{i=1}^{M}$ represent a dataset of M students, where $\mathbf{x}_i \in \mathbb{R}^N$ denotes the vector of N normalized features of the i-th student. The label $y_i \in \{0,1\}$ is set to 1 if the student's final grade (on a scale of 0 to 5) is below 3.5, and 0 otherwise, with 3.5 being the minimum final grade needed to obtain a certificate of participation. Our objective is to train a binary classifier $f_\theta : \mathbb{R}^N \to \{0,1\}$ such that the predicted label $\hat{y} = f_\theta(\mathbf{x})$ correctly classifies student performance based on the feature vector $\mathbf{x}$.

3.4. Clustering

To complement our performance prediction task, we conducted a clustering analysis using the same engagement, behavior, and performance features to uncover distinct groups of students with similar learning patterns and interactions [31]. The goal of this analysis is not only to validate the usefulness of these features in student performance prediction but also to provide actionable insights for course personalization. Each cluster represents a unique learning profile, which can help educators design targeted interventions, such as offering additional guidance to students showing signs of disengagement, adapting content formats to match learner preferences, or providing enrichment activities for students who progress rapidly through the course [18].

3.5. Relationship Between Student Patterns and Performance

By uncovering patterns in student interactions with course materials and relating them to academic performance, we can examine these relationships using statistical methods like ANOVA [32]. This analysis helps identify behavioral factors that significantly influence academic outcomes. Such insights are instrumental in designing targeted interventions to enhance educational effectiveness, offering valuable support for educators and institutions striving to improve student success [18].

3.6. Data

We utilize data from the "Prepárate para la Vida Universitaria" (PPVU) course, hosted on the Moodle LMS and developed by the Universidad de Antioquia. This eight-week course focuses on logical-mathematical reasoning and problem-solving strategies to support students preparing for university entrance exams. It is open to learners worldwide, although we believe most participants are based in or near the Antioquia region in Colombia. The course requires an estimated 48 h of study, with a recommended pace of six hours per week. It is structured into five units containing explanatory videos and interactive modules. Modules are delivered either through Moodle's lesson activity or as multimedia content. These formats are treated separately when defining features (see Tables 1 and 2), and includes 12 graded assessments—six workshops (30% of the final grade) and six entrance-exam-style simulations (70% of the final grade). Two additional diagnostic exams are provided to assess students' prior knowledge, though they do not contribute to the final grade. A distinctive feature of the course is its self-paced design: students may complete activities in any order, repeat them as needed, and retain only their highest score. Those achieving a final average grade of 3.5 or higher (on a 0–5 scale) receive a certificate of participation. We anonymized and collected all data by the institution's data policy, which students accepted upon enrollment. We did not access any personal information, and we securely stored the data on institutional servers with access restricted to authorized researchers.

We quantify the degree of linearity of the student's learning path through a Linearity Score computed over the 12 evaluative activities of the course. This score measures how closely a student's assessment sequence follows an ascending order. A Linearity Score of 1 indicates perfect linearity, where a student completes the assessments in strictly ascending order from 1 to 12 or with repetitions while maintaining the ascending sequence (e.g., 1, 2, 2, 2, 3, 4, 5, 5, 5). In contrast, scores closer to 0 represent a lack of linearity, where the order of assessments appears disordered (e.g., 2, 5, 3, 6, 5). Figure 3 shows the distribution of the Linearity Score across students. We can see the diversity in how students resolve the course assessment, ranging from highly linear patterns (scores near 1) to highly disordered ones (scores near 0). This variability in assessment order reflects the self-paced, flexible nature of the course and highlights differences in how students

approach the learning material [33,34]. The algorithm used to compute the Linearity Score is detailed in Algorithm A1.

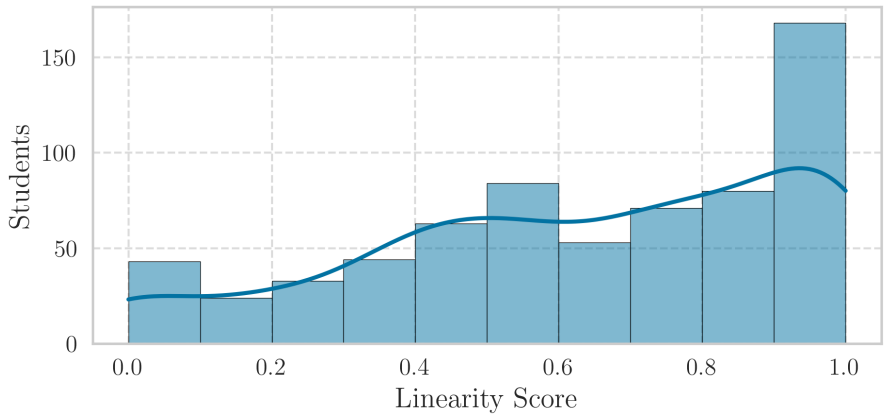


Figure 3. Distribution of Linearity Scores among students. A score of 1 indicates perfect linearity in assessment completion, while scores near 0 represent disordered sequences.

Figure 4 shows the distribution of the students according to the percentage of the overall course assessment they completed, highlighting a steep dropout rate from the beginning of the course. In this context, "early stages" refers to the first completed evaluative activities, since students can follow the course in any order. Many students disengage before reaching 5% of the Cumulative Weight Assessment (CWA); among those who reach 5%, more than half drop out before completing 25% of the assessments, with a particularly sharp decline observed between the 5% and 20% CWA. This early dropout pattern emphasizes the importance of developing predictive models to accurately identify at-risk students in the initial phases of the course.

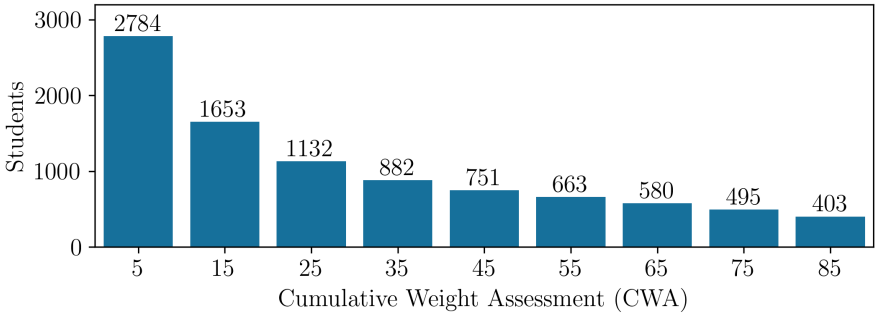


Figure 4. Distribution of students by cumulative weight assessment.

Table 1. Engagement features extracted from Moodle.

Feature	Description	Feature	Description
course_navigation	Number of accesses to the course modules	mean_multimedia_views	Mean of multimedia views
forum_viewed	Number of access to the forum	std_multimedia_views	Std of multimedia views
post_created	Number of post created	mean_multimedia_time	Mean of time spent in multimedias
discussion_viewed	Number of access to the discussion	std_multimedia_time	Std of time spent in multimedias
discussion_created	Number of discussion created	mean_video_views	Mean of video views
forum_searched	Number of searches in the forum	std_video_views	Std of multimedia views
outline_viewed	Number of outline viewed	mean_video_time	Mean of time spent in videos
week	Number of interactions per week	std_video_time	Std of time spent in videos
weekend	Number of interactions per weekend	mean_lessons_attempts	Mean of lessons attempts
days	Number of days on platform	std_lessons_attempts	Std of lessons attempts
mean_interactions	Mean of interactions generated per day	mean_lessons_time	Mean of time spent in lessons
std_interactions	Std of interactions generated per day	std_lessons_time	Std of time spent in lessons

Table 2. Behavior features extracted from Moodle.

Feature	Description	Feature	Description
mean_prior_knowledge_time	Mean time spent in prior knowledge	std_workshops_attempts	Std workshops attempts
std_prior_knowledge_time	Std time spent in prior knowledge	mean_simulations_time	Mean time spent in simulations
mean_workshops_time	Mean time spent in workshops	std_simulations_time	Std time spent in simulations
std_workshops_time	Std time spent in workshops	mean_simulations_attempts	Mean simulations attempts
mean_workshops_attempts	Mean workshops attempts	std_simulations_attempts	Std simulations attempts

3.6.1. Features

We selected data from four key Moodle tables—*lesson grades*, *lesson timer*, *activity logs*, and *quiz attempts*—corresponding to `mdl_lesson_grades`, `mdl_lesson_timer`, `mdl_logstore_standard_log`, and `mdl_quiz_attempts`, respectively, as described in Section 3.1.

We cleaned the raw data from these tables to prepare the data for model training. The features are categorized into three types: Engagement (E), Behavior (B), and Performance (P).

- The engagement features capture student interactions with various aspects of the course, such as forum participation, the number of module accesses, and time spent on different learning materials, including lessons, videos, and multimedia resources. The multimedia resources were developed with the Genially platform and include topic explanations, short videos, and illustrative examples. Engagement features also track the total number of interactions on a daily, weekly, and weekend basis. There are 24 engagement features (see Table 1).
- The behavior features describe how students interact with evaluative activities, including the time spent and the number of attempts made on assessments. These features provide insights into the students' interaction with quizzes, lessons, and assignments. There are 10 behavior features (see Table 2).
- The performance feature `weighted_average` summarizes students' academic performance by calculating the weighted average of their highest grades across workshops and simulations.

Figure 5 shows an example of feature aggregation for `Student 1`, who engaged in various platform activities. Interactions such as `course_navigation`, `forum_viewed`, and `post_created` are aggregated as the total number of interactions. Additionally, `Student 1` accessed multimedia resources, but outlier data points, such as unusually long or short viewing times, were removed. The remaining data points were used to calculate the mean time spent on multimedia resources, representing the student's productive engagement with educational content.

Student	Event type	Timestamp	Elapsed time
Student 1	course_navigation	15:40:30	0:00:10
Student 1	forum_viewed	15:40:40	0:01:10
Student 1	post_created	15:41:50	0:03:40
Student 1	multimedia_viewed	15:45:30	0:24:45
Student 1	multimedia_viewed	16:10:15	0:40:15
Student 1	multimedia_viewed	16:50:30	0:00:50
Student 1	course_navigation	16:51:20	0:01:34
Student 1	multimedia_viewed	16:52:54	0:00:36
Student 1	multimedia_viewed	16:53:30	3:28:43
Student 1	course_navigation	20:22:13	

⇩

Student	course_navigation	forum_viewed	post_created	mean_multimedia_time (s)
Student 1	3	1	1	1950

Figure 5. Example of feature extraction. Orange represents filtered logs (very long or short duration), green shows logs used to calculate the mean time spent on the resource in seconds, and uncolored logs are aggregated by count.

As noted in Section 3.2.2, we updated the features until a student's CWA_U was less than or equal to a given threshold. The distribution of each feature evolves as the CWA_U threshold increases. For example, Figure 6 shows the distribution of the `mean_multimedia_views` feature for two different CWA_U thresholds. As expected, increasing the threshold from 20% to 60% resulted in a broader distribution with higher mean and variance values.

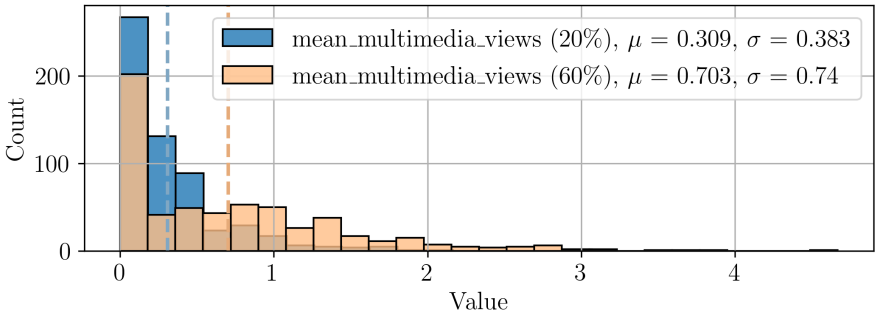


Figure 6. Distribution of the mean multimedia views for CWA_U thresholds of 20% and 60%.

3.6.2. Target Variable

The course awards a certificate of participation to students who achieve a final grade of 3.5 or higher on a 0 to 5 scale. The final grade is computed as a weighted average, using the highest grade achieved in each evaluative activity. For the binary classification model, we define the target variable y as 1 if the student does not achieve the certificate of participation and 0 otherwise.

For clarity, we provide a sample of the training set in Table A1, which includes data from five students. This example presents all the features used to describe student interactions on the platform and their values utilized for model training.

4. Experiments and Results

This section presents experiments on the data described in Section 3.6. Since we are interested in early student performance prediction, we use four distinct Cumulative Weight Assessment upper thresholds (CWA_U): 20%, 40%, 60%, and 80% in all of our experiments, for both classification and clustering. These thresholds divide the course into progressive stages, helping us better interpret how student performance and learning patterns evolve. The 20% and 40% thresholds represent early engagement, critical for identifying students needing timely support. The 60% and 80% thresholds reflect later stages of the course, allowing us to evaluate how predictive performance improves as more data becomes available. All code was implemented in Python 3.10.9 using the scikit-learn library [35].

4.1. Sample Size

For the analysis, we defined CWA_L as the lower bound for selecting students. In other words, the final dataset used in our experiments included students who achieved a CWA equal to or greater than CWA_L. Figure 7 shows how the total number of students and the class distribution (certified vs. not certified) changed with CWA_L. With a CWA_L threshold of 5%, 2437 students did not achieve certification ($y = 1$), and 347 did ($y = 0$), resulting in an imbalanced dataset. Using that threshold would have included students with limited interaction on Moodle, which could have led to noisy data. In contrast, when we increased CWA_L to 75%, the dataset contained students with substantial interactions but fewer samples: 148 students who did not achieve certification ($y = 1$) and 347 who did ($y = 0$), maintaining imbalance and potentially filtering out important early-stage dropout patterns crucial for identifying at-risk students.

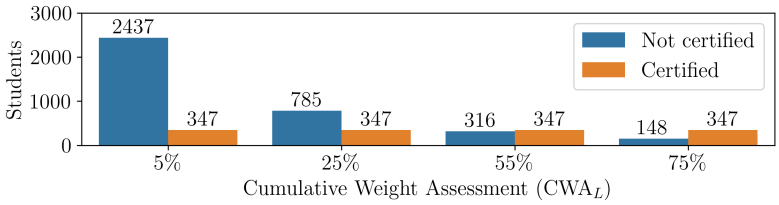


Figure 7. Class distribution in different CWA$_L$ thresholds.

To select the value of CWA$_L$ and obtain the final dataset, we evaluated the performance of the random forest classifier in predicting whether a student would achieve the certificate of participation at four different values of CWA$_L$: 5%, 25%, 55%, and 75%. We used a CWA$_U$ threshold of 20% for all sample sizes to determine each student's Feature Aggregation Time Point (FATP). Figure 8 shows the confusion matrix of the model for each value of CWA$_L$. We observed a bias toward the positive class at lower thresholds, while higher thresholds favored the negative class. A threshold of 55% offered the best trade-off between the true positive and true negative rates. Consequently, we selected the 55% CWA$_L$ threshold, which corresponded to a dataset with 663 samples: 316 students who did not achieve a participation certificate ($y = 1$) and 347 who did ($y = 0$).

4.2. Prediction

This section presents the experiments and results on early student performance prediction.

4.2.1. Feature Set Ablation

Recalling that we have in total 35 features and that there are three feature categories: Engagement (E), Behavior (B), and Performance (P) (see Section 3.6.1). We assessed model performance using each feature category (E, B, P) and their combinations (E + B, E + P, B + P, E + B + P).

4.2.2. Machine Learning Models and Training Procedure

We evaluated three machine learning algorithms for classification: Logistic Regression (LR) [36], Random Forest (RF) [36], and Support Vector Machine (SVM) [36]. These algorithms were selected because they provide insights into feature importance during prediction [37]. We employed nested cross-validation for model training and evaluation, using stratified 4-fold cross-validation in the inner loop and stratified 10-fold cross-validation in the outer loop. We use the inner loop to optimize hyperparameters, while the outer loop estimates the models' generalization error [38]. Table 3 shows the hyperparameter ranges used for each algorithm.

Table 3. Hyperparameters optimized using the Grid-Search technique for classification.

Classifier	Hyperparameters	Values' Interval
LR	penalty	[None, 'l1', 'l2']
	C	[0.01, 0.1, 1, 10, 100]
	solver	['liblinear', 'lbfgs', 'newton-cg', 'sag', 'saga']
	max_iter	[100, 200, 300]
SVM	C	[0.01, 0.1, 1, 10, 100]
	kernel	['linear', 'rbf']
	gamma	[0.01, 0.1, 1, 10, 100]

Table 3. *Cont.*

Classifier	Hyperparameters	Values' Interval
RF	n_estimators	[200, 250, 300, 350, 400, 450]
	max_depth	[None, 10, 20]
	min_samples_split	[10, 20, 30]
	min_samples_leaf	[1, 2, 5]
	max_features	['sqrt', 'log2', None]

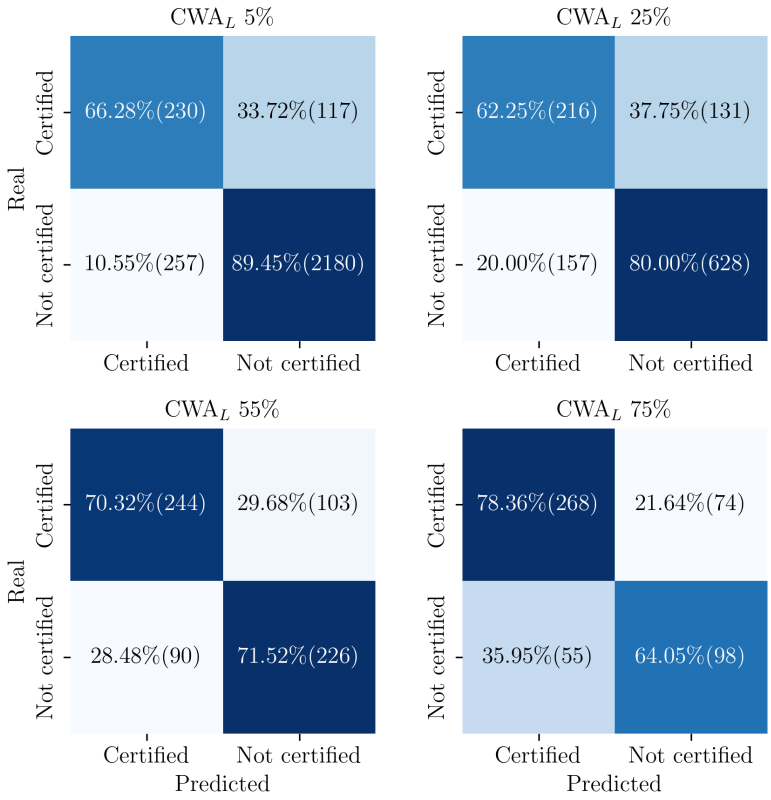


Figure 8. Confusion matrix of a random forest model at different CWA$_L$ thresholds.

4.2.3. Performance Measures

In this context, a *positive* example refers to a student who did not receive a certificate of participation ($y = 1$). In contrast, a *negative* example refers to a student who did not receive a certificate ($y = 0$). Since minimizing misclassifications of positive examples is crucial, we primarily focus on the F1-score to evaluate model performance.

Precision measures the proportion of correctly classified positive examples (true positives) among all examples predicted as positive (true and false positives). It indicates the model's accuracy in predicting students who will not receive a certificate. Recall measures the model's ability to correctly identify students at risk of not receiving a certificate (true positives, overall actual positives). The F1-score, which is the harmonic mean of precision and recall, balances these two metrics, providing a comprehensive measure of model performance.

4.2.4. Heuristic Baseline

To provide a point of comparison, we implement a simple, non-machine-learning baseline that uses only the performance (P) feature. Specifically, we utilize the student's weighted average grade at a given CWA$_U$ threshold. The baseline predicts that a student will not receive a certificate of participation ($\hat{y} = 1$) if their weighted average grade is below 3.5 and will receive one ($\hat{y} = 0$) otherwise. We apply this heuristic to the same 10-fold used

in the outer loop of the nested cross-validation process described in Section 4.2.2 to ensure a fair comparison with the machine learning models. We then compute the mean (μ) and standard deviation (σ) of the F1-score, as shown in Table 4.

Table 4. Performance results (F1-score) of baseline at different CWA_U.

Model	CWA_U	F1-Score ($\mu \pm \sigma$)
Baseline	20%	0.63 ± 0.06
	40%	0.70 ± 0.08
	60%	0.72 ± 0.06
	80%	0.73 ± 0.07

As illustrated in Table 4, the baseline model's F1-score improves as the CWA_U threshold increases, underscoring the relevance of the performance feature in predicting whether a student will achieve the certificate of participation. The baseline's competitive performance, even at the 20% CWA_U threshold, provides a valuable benchmark for evaluating the machine learning model results.

4.2.5. Individual Models

We train and validate the classifiers using the same CWA_U for this experiment. The results in Table 5 reveal a consistent trend across all models: an incremental improvement in F1-score as CWA_U increases from 20% to 80%. This result aligns with expectations, as higher CWA_U values capture more comprehensive student grade and behavioral data.

Table 5. Performance comparison (F1-score) of different models at different CWA_U and features (E—Engagement, B—Behavior, P—Performance data).

Model	CWA_U	F1-Score ($\mu \pm \sigma$)					
		E	**B**	**E + B**	**E + P**	**B + P**	**E + B + P**
LR	20%	0.59 ± 0.05	0.60 ± 0.05	0.66 ± 0.07	0.68 ± 0.04	0.71 ± 0.06	0.71 ± 0.05
	40%	0.63 ± 0.05	0.62 ± 0.08	0.69 ± 0.06	0.75 ± 0.06	**0.76 ± 0.05**	0.75 ± 0.06
	60%	0.64 ± 0.07	0.64 ± 0.09	0.69 ± 0.05	0.77 ± 0.04	**0.79 ± 0.05**	0.79 ± 0.06
	80%	0.65 ± 0.07	0.65 ± 0.07	0.75 ± 0.04	0.78 ± 0.06	0.82 ± 0.06	0.80 ± 0.04
RF	20%	0.63 ± 0.04	0.57 ± 0.07	0.68 ± 0.05	0.70 ± 0.07	0.66 ± 0.05	0.70 ± 0.06
	40%	0.65 ± 0.05	0.64 ± 0.06	0.70 ± 0.05	0.75 ± 0.07	0.73 ± 0.07	0.76 ± 0.06
	60%	0.63 ± 0.07	0.64 ± 0.06	0.71 ± 0.04	0.78 ± 0.04	0.77 ± 0.09	0.79 ± 0.05
	80%	0.66 ± 0.08	0.70 ± 0.08	0.74 ± 0.04	0.81 ± 0.04	0.80 ± 0.06	0.82 ± 0.05
SVM	20%	0.64 ± 0.03	0.58 ± 0.05	0.66 ± 0.05	0.70 ± 0.05	**0.73 ± 0.06**	0.72 ± 0.06
	40%	0.68 ± 0.04	0.60 ± 0.07	0.69 ± 0.05	0.74 ± 0.07	0.74 ± 0.05	0.74 ± 0.06
	60%	0.66 ± 0.04	0.64 ± 0.07	0.70 ± 0.05	0.76 ± 0.02	0.78 ± 0.04	0.77 ± 0.04
	80%	0.67 ± 0.03	0.64 ± 0.07	0.72 ± 0.06	0.79 ± 0.04	**0.82 ± 0.05**	0.82 ± 0.05

Note: Bold values indicate the best F1-score for each CWA_U threshold.

Tables 4 and 5 show that the baseline outperforms the machine learning models when trained solely on engagement or behavior features. Moreover, combining engagement with behavior or performance features does not substantially improve F1-scores, suggesting that engagement alone is less predictive of whether a student will earn a participation certificate, which aligns with the feature importance analysis presented in Section 4.3.

When engagement and behavior features are combined, the model's performance becomes comparable to the baseline. Although these two features are not directly related to student performance, they provide valuable insights. These insights enable more accurate predictions regarding participation certificate achievement.

By incorporating performance-related features, we observe an improvement, highlighting the role of engagement and behavior in predictive modeling. The combination of behavior and performance features, which are linked to evaluative activities, leads to better prediction performance. This is expected, given their close association with student grades. The similarity between the results from behavior + performance and the three-feature combination underscores the direct impact of behavior and performance on performance outcomes.

While LR, RF, and SVM models show comparable results, RF performs the best. However, RF is computationally the most intensive—the mean and standard deviation of RF's execution time for nested cross-validation was 655.57 ± 54.34 s, while other models required 2.94 ± 0.31 s; these times may vary based on system hardware and configuration. LR is the least computationally expensive model, and it predicts student performance early, achieving an F1-score of 0.73 at just 20% CWA_U. This insight can help educators identify at-risk students early, allowing them to implement targeted interventions.

4.2.6. Single Model

Given the higher computational costs of training multiple models than a single model, we investigate using a single model trained at a specific CWA_U threshold for making predictions at different thresholds. We trained a logistic regression model using features at the 40% CWA_U threshold, referred to as *SingleModel@40*. We then compared its performance with models trained for specific CWA_U thresholds, referred to as *IndividualModel@k*, where *k* represents the percentage threshold.

To ensure a fair comparison, we selected the fold with the highest performance from the outer loop in the SingleModel@40's cross-validation, and used this fold to compare models across all the CWA_U thresholds. Features were computed for each fold at the 20%, 40%, 60%, and 80% thresholds, with predictions made using SingleModel@40. For IndividualModel@k, a separate model was trained at the k% CWA_U threshold and tested on the same data. Note that in this context, a fold is a subset of the students. By definition, the data itself changes as we change the CWA_U.

Table 6 summarizes the results. As expected, using a single model across different CWA_U thresholds leads to a general decrease in F1-scores compared to models explicitly trained for each threshold.

Table 6. Comparative results of the single model against individual models at different CWA_U.

Approach	CWA_U	F1-Score
		B + P
SingleModel@40	20%	0.73
	40%	0.86
	60%	0.85
	80%	0.83
IndividualModel@k	20%	0.78
	40%	0.86
	60%	0.89
	80%	0.92

4.3. Feature Importance

We calculated the importance of features using three models: Logistic Regression (LR), Support Vector Machine (SVM) with a linear kernel, and Random Forest (RF). Each model provided a different method for estimating feature importance.

In logistic regression, we computed feature importance based on the coefficients associated with each feature. The model predicts the probability of the target class using the following expression:

$$P(y = 1|\mathbf{x}) = \frac{1}{1 + e^{-(w_0 + \sum_{i=1}^{N} w_i x_i)}} \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ are the input features, w_0 is the intercept, and w_i represents the coefficient (weight) for feature x_i. The feature importance corresponds to the absolute value of each coefficient, $|w_i|$, with larger magnitudes indicating greater feature influence.

We derived the feature importance from the weight vector $\mathbf{w}$, which defines the decision boundary for the support vector machine with a linear kernel. We expressed the decision function as:

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0 = w_0 + \sum_{i=1}^{N} w_i x_i \tag{2}$$

where w_i is the weight assigned to each feature x_i, and w_0 is the bias term. The importance of each feature is given by the absolute value of the corresponding weight, $|w_i|$, where larger values denote a more significant impact on the model's decision-making.

For the random forest model, we computed feature importance by measuring the mean decrease in impurity (Gini importance). We split the data at different features to reduce impurity (e.g., Gini index or entropy) at each node. We based the importance of each feature on the average reduction in impurity when that feature was used for splitting across all trees in the forest:

$$I_i = \sum_{t=1}^{T} \frac{\Delta I_t(x_i)}{T} \tag{3}$$

where I_i is the importance of feature x_i, $\Delta I_t(x_i)$ is the decrease in impurity for feature x_i in tree t, and T is the total number of trees.

To ensure the robustness of our results, we calculated the importance of features using 10-fold cross-validation. We compute feature importance for all models for each fold, then average the values across the ten folds to obtain a stable estimate of feature importance. Finally, to compare the relative importance of each feature, we used a min-max normalization, assigning the most important feature a value of 1 and the least important a value of 0. This normalization allows us to compare feature importance across different features and models more efficiently, ensuring that the results are interpretable and comparable across various modeling approaches.

The feature importance analyses across LR, SVM with a linear kernel, and RF are shown in Figures 9, 10 and 11, respectively, revealing consistent trends in how different types of features relate to student performance. Features associated with student engagement, such as course navigation, forum participation, and navigation activity, consistently exhibited low predictive value across all CWA_U thresholds. In contrast, some engagement features, such as multimedia and video resources viewed, including mean views and time spent, became increasingly important at higher CWA_U thresholds, indicating that active interaction with these materials is more predictive of student success as they progress through the course.

Behavior-based features, such as the frequency of attempts and time spent on assessments, emerged as significant predictors of student performance, suggesting that how students interact with evaluative activities directly correlates with their expected grades. Interestingly, while workshop metrics were important at lower thresholds, their relevance diminished over time. Conversely, simulation metrics gained importance as thresholds increased. This shift can be attributed to the higher weighted average of simulation activi-

ties within the overall course structure, meaning simulations contribute more heavily to students' final grades in the later stages.

Figures 9 and 10 show that LR's feature importance values were lower than SVM's. This difference can be attributed to the regularization techniques employed in LR during grid search, which included both L1 and L2 penalties. L1 regularization tends to let some coefficients be zero, excluding less important features, while L2 regularization in SVM allows all features to contribute their importance.

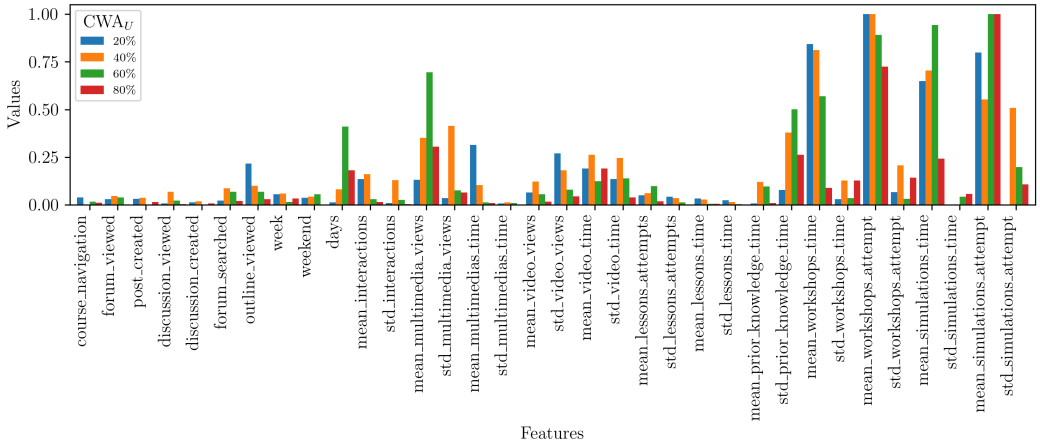


Figure 9. Feature importance in logistic regression.

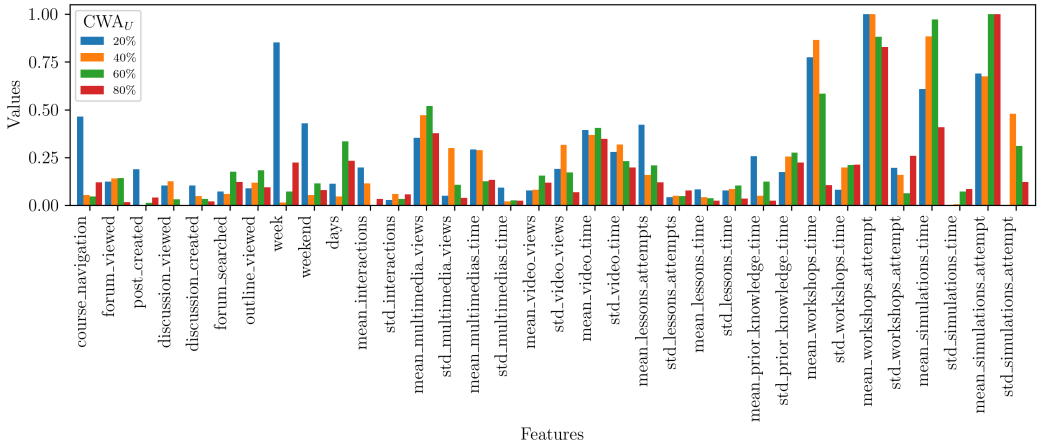


Figure 10. Feature importance SVM.

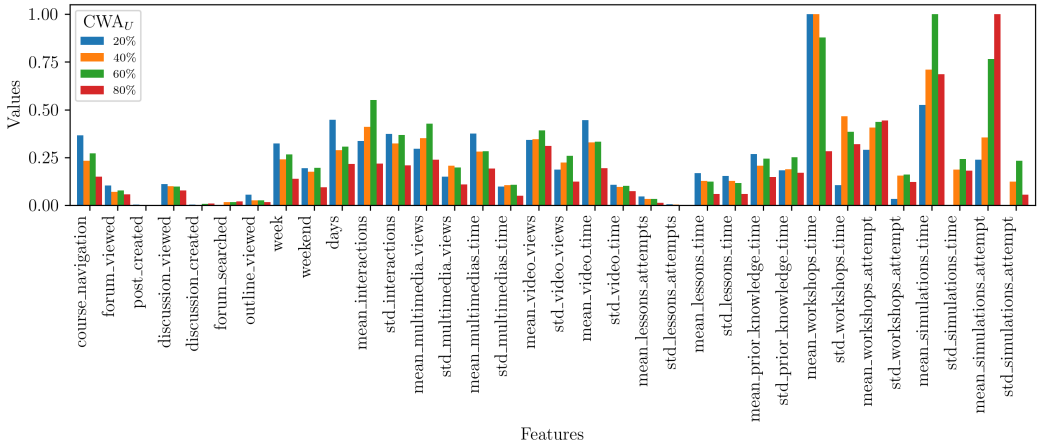


Figure 11. Feature importance random forest.

4.4. Clustering

We performed clustering analysis to identify distinct learning behavior patterns among students based on their interactions with the course, using the features presented in Section 3.6. All variables were z-score normalized ($\mu = 0$ and $\sigma = 1$). We created four student groupings, one for each CWA_U threshold, using the k-means algorithm. Since k-means require a predetermined number of clusters, we utilized the Elbow and Silhouette methods to determine the optimal number of clusters [39].

We found that the optimal number of clusters for each CWA_U threshold is $k = 4$. ANOVA revealed significant differences among the four groups (p-value < 0.05) for all features across every CWA_U prediction threshold, except for `forum_searched`. Table 7 presents the mean and standard deviation of a subset of the features at the centroid of each cluster and for different CWA_U values. The index of the clusters grows with the weighted average, i.e., Cluster 0 and Cluster 3 correspond to the clusters where the centroid has the smallest and largest weighted average grade, respectively.

Table 7. Number of students per group (N), feature average (μ), standard deviation (σ) values per cluster, for the four different CWA_U (20%, 40%, 60%, and 80%).

CWA_U	Features	Cluster 0 (N = 249)	Cluster 1 (N = 286)	Cluster 2 (N = 109)	Cluster 3 (N = 19)
		$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
20%	mean_multimedias_time	10.37 ± 6.58	2.11 ± 4.64	9.78 ± 7.01	12.67 ± 6.02
	mean_video_time	13.89 ± 7.48	1.59 ± 4.22	10.33 ± 7.13	12.05 ± 6.22
	mean_simulations_time	14.30 ± 7.24	4.57 ± 5.96	0.00 ± 0.00	7.43 ± 8.99
	mean_workshops_time	9.82 ± 5.73	4.62 ± 5.09	26.06 ± 16.00	21.63 ± 12.83
	weighted_average	3.40 ± 1.14	3.72 ± 1.18	4.00 ± 0.83	4.05 ± 0.82
		Cluster 0 (N = 224)	Cluster 1 (N = 136)	Cluster 2 (N = 290)	Cluster 3 (N = 13)
40%	mean_multimedias_time	12.23 ± 6.10	11.62 ± 5.08	2.32 ± 5.12	14.75 ± 3.93
	mean_video_time	13.05 ± 5.69	12.72 ± 4.94	1.43 ± 3.91	13.67 ± 2.76
	mean_simulations_time	26.13 ± 14.44	21.34 ± 14.75	7.43 ± 8.50	20.13 ± 11.81
	mean_workshops_time	27.61 ± 15.22	30.75 ± 19.03	9.22 ± 9.59	40.89 ± 18.65
	weighted_average	3.35 ± 0.99	3.78 ± 0.97	4.03 ± 0.98	4.10 ± 0.54
		Cluster 0 (N = 234)	Cluster 1 (N = 114)	Cluster 2 (N = 15)	Cluster 3 (N = 300)
60%	mean_multimedias_time	12.66 ± 5.49	12.37 ± 4.38	13.46 ± 4.92	2.47 ± 5.07
	mean_video_time	12.91 ± 4.92	12.93 ± 4.54	14.65 ± 3.67	1.60 ± 3.91
	mean_simulations_time	41.08 ± 20.21	35.39 ± 19.19	36.80 ± 16.76	11.02 ± 11.61
	mean_workshops_time	37.18 ± 19.37	36.93 ± 20.38	44.04 ± 19.40	11.17 ± 12.23
	weighted_average	3.40 ± 0.97	3.83 ± 0.96	3.95 ± 0.60	4.11 ± 0.91
		Cluster 0 (N = 215)	Cluster 1 (N = 135)	Cluster 2 (N = 16)	Cluster 3 (N = 297)
80%	mean_multimedias_time	12.70 ± 5.73	12.59 ± 4.26	13.09 ± 4.80	2.24 ± 4.25
	mean_video_time	12.85 ± 5.12	12.93 ± 4.16	13.93 ± 3.38	1.59 ± 3.87
	mean_simulations_time	52.25 ± 25.91	47.88 ± 25.18	53.82 ± 27.19	13.74 ± 14.48
	mean_workshops_time	39.91 ± 21.32	41.64 ± 21.66	45.55 ± 18.48	12.08 ± 12.91
	weighted_average	3.33 ± 0.96	3.84 ± 0.97	3.97 ± 0.58	4.19 ± 0.86

The clustering analysis across different CWA_U thresholds reveals distinct patterns in student engagement with multimedia, video, simulations, and workshops, which correlate with their performance, as measured by the weighted average (see Table 7). These findings align with the previous feature importance analysis, highlighting the critical role of early interaction with multimedia and video resources in predicting student success.

Students in higher-performing clusters consistently engage more with multimedia and video resources at early CWA_U stages, suggesting that early and frequent interaction with these materials improves outcomes. As students progress to higher CWA_U thresholds, the importance of simulations and workshops becomes more pronounced, particularly in

high-performing clusters. Increased engagement with evaluative activities in these stages is associated with better performance, indicating the value of taking the necessary time to solve those evaluative activities. These findings suggest two potential interventions:

1. Encouraging early and sustained engagement with multimedia and video materials may improve outcomes for lower-performing students.
2. Additional support for simulations and workshops at later stages could help students maximize their learning outcomes as they approach higher CWA_U thresholds.

4.5. Correlation Between Student Clusters and Performance

We now examine whether a relationship exists between the clusters identified in the previous section and students' final grades. An ANOVA test confirmed this relationship, revealing statistical significance (p-value < 0.05). To further explore these differences, we conducted Tukey's post hoc tests to determine significant performance variations between pairs of clusters.

Figure 12 shows the distribution of final grades across different clusters at varying CWA_U thresholds. Tukey's post hoc test highlights the significant differences between clusters as shown in Table 8; the bold values indicate which clusters these differences are statistically significant.

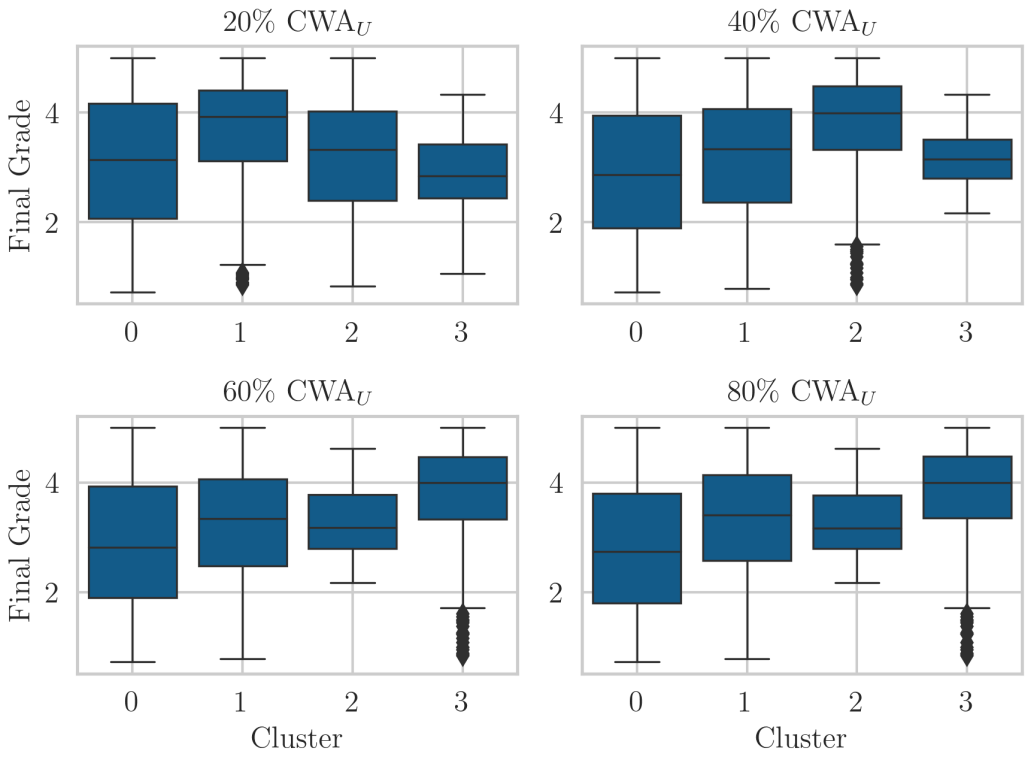


Figure 12. Students' final grade for each cluster at different CWA_U thresholds.

Tukey's post hoc results indicate notable performance differences between clusters at different CWA_U thresholds (see Figure 12 and Table 8). Three key patterns consistently emerge across all thresholds:

1. Effective Resource Engagement: Clusters with higher weighted averages, particularly Clusters 2 and 3 at all CWA_U thresholds, show a more robust and more strategic engagement with educative resources. These clusters spent the most time on multimedia, simulations, and workshops, likely contributing to their better academic outcomes.
2. Strategic Use of Key Resources: There are notable differences in the amount of time spent on engagement across clusters, indicating that higher academic achievement is

associated with allocating time and strategic focus on specific resources. Clusters that spent more time in simulations and workshops at higher CWA_U thresholds performed better than those that focused on other educative resources.

3. Small Clusters, High Performance: Smaller clusters, such as Cluster 3 at 20% and 40% and Cluster 2 at 60% and 80% CWA_U thresholds, respectively, consistently exhibited higher weighted averages and more time spent with evaluative resources like simulations and workshops, suggesting that smaller groups of students who focused their study efforts efficiently on evaluative resources were able to achieve better performance.

Table 8. p-values result for each cluster at different CWA_U thresholds. Bold values indicate significant clusters ($p \leq 0.05$).

CWA_U	Cluster	p-Value		
		Cluster 1	**Cluster 2**	**Cluster 3**
20%	Cluster 0	**0.001**	0.765	0.749
	Cluster 1		**0.005**	**0.016**
	Cluster 2			0.531
40%	Cluster 0	0.063	**0.001**	0.900
	Cluster 1		**0.001**	0.900
	Cluster 2			0.116
60%	Cluster 0	**0.031**	0.551	**0.001**
	Cluster 1		0.900	**0.001**
	Cluster 2			0.256
80%	Cluster 0	**0.001**	0.378	**0.001**
	Cluster 1		0.900	**0.001**
	Cluster 2			0.143

Our findings emphasize the importance of strategic engagement and focused study habits in shaping academic success. Specific clusters consistently outperform others, primarily due to their effective use of educational resources.

We conducted an analysis comparing the weighted average with the Linearity Score. As shown in Table 9, students with higher weighted averages tend to have lower linearity scores, suggesting that these students do not follow a strictly sequential or predetermined path when engaging with learning resources and assessments. Instead, they exhibit a more adaptive and exploratory approach, often revisiting previous materials, iterating on assessments, or engaging in self-regulated learning strategies to reinforce their understanding. Note that from Tables 7 and 9, and Figure 13, clusters 3 and 2 seem to have swapped; this was indeed the case: 12 of the 13 students in cluster 3 at $CWA_U = 40\%$ moved to cluster 2 and stayed there for $CWA_U \in \{60\%, 80\%\}$

This behavior aligns with self-regulated learning (SRL) principles, where students actively plan, monitor, and regulate their learning process to optimize their academic performance [33]. Rather than simply following a linear progression through course content, these students take strategic paths, reviewing previous topics, selectively repeating assessments to improve their grades, or navigating resources to best suit their individual learning needs. Research on SRL in online learning environments supports this finding. It highlights how high-performing students engage in metacognitive monitoring, set learning goals, and adjust their strategies based on feedback and performance data [34]. These insights suggest that self-learning practices, emphasizing reflection, revision, and iterative assessment, contribute significantly to students' academic success in non-linear, self-paced learning environments.

Table 9. Weighted Average and Linearity Score comparison ($\mu \pm \sigma$) values per cluster, for the four different CWA_U thresholds (20%, 40%, 60% and 80%).

CWA_U	Feature	$\mu \pm \sigma$		
		Cluster 1	**Cluster 2**	**Cluster 3**
20%	Weighted Average	3.72 ± 1.18	4.00 ± 0.83	4.05 ± 0.82
	Linearity Score	0.58 ± 0.28	0.43 ± 0.28	0.39 ± 0.36
40%	Weighted Average	3.78 ± 0.97	4.03 ± 0.98	4.10 ± 0.54
	Linearity Score	0.65 ± 0.33	0.58 ± 0.27	0.30 ± 0.25
60%	Weighted Average	3.83 ± 0.96	3.95 ± 0.60	4.11 ± 0.91
	Linearity Score	0.66 ± 0.31	0.37 ± 0.30	0.59 ± 0.27
80%	Weighted Average	3.84 ± 0.97	3.97 ± 0.58	4.19 ± 0.86
	Linearity Score	0.67 ± 0.30	0.37 ± 0.30	0.59 ± 0.28

We analyzed students' active forum participation, including posting and responding to discussions. While the feature importance analysis does not indicate a high predictive value for forum participation, Figure 13 reveals that clusters with fewer students and higher weighted averages tend to exhibit greater engagement in forum discussions, suggesting that students in these high-performing clusters actively seek help, clarifying doubts, discussing concepts, and leveraging peer interactions to enhance their understanding. Notably, many of their messages demonstrate continuity within discussions, indicating that they initiate and sustain engagement by responding to peers, following up on advisors' answers, and contributing to ongoing conversations.

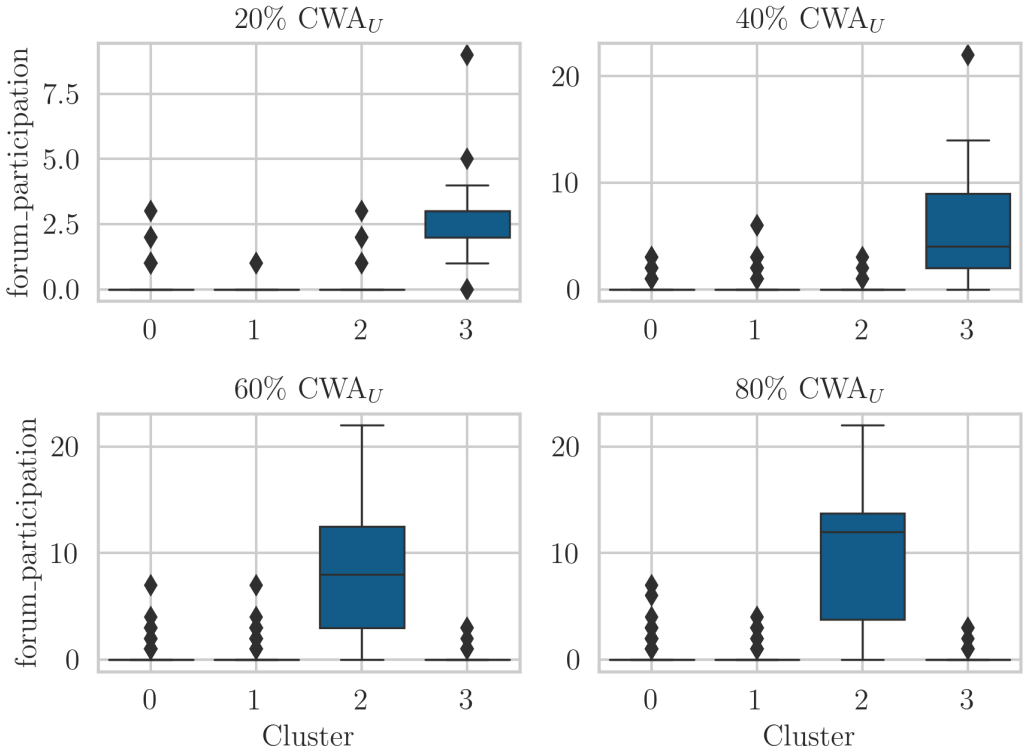


Figure 13. Students' forum participation cluster at different CWA_U thresholds.

This behavior aligns with SRL strategies, particularly the help-seeking dimension, which Zimmerman [33] identifies as a key component of effective self-regulation. Rather than passively struggling with difficulties, these students actively approach their learning by reaching out to peers or instructors, demonstrating metacognitive awareness of their

knowledge gaps and a willingness to use available resources to overcome challenges. In online learning environments, research has shown that students who engage in interactive, feedback-driven learning, such as forum discussions, tend to develop deeper conceptual understanding and achieve better academic outcomes [34]. Furthermore, help-seeking in online learning environments is an indicator of self-directed learning, where students navigate their educational journey by combining independent study with strategic social interactions. The findings suggest that while forum participation alone may not directly predict performance for all students, it plays a crucial role for those who exhibit a high degree of self-regulated learning behaviors. Therefore, encouraging structured peer discussions and fostering a collaborative learning culture could benefit students who actively engage in self-regulation strategies.

4.6. Clustering: Top Feature Importance

We utilized all the features in the clustering analysis presented in Section 4.4. In this section, we evaluate how much clustering changes when we use a subset of the features, leveraging the feature importance analysis from Section 4.3. For each classification method used in the binary prediction (LR, RF, and SVM) and each CWA_U (20%, 40%, 60%, and 80%), we selected the top 5 most important features. We then identified the features that appeared most frequently. Finally, the following five features emerged as the most significant for the analysis: `mean_simulations_attempt`, `mean_simulations_time`, `mean_workshops_attempt`, `mean_workshops_time`, and `weighted_average`. The Elbow and Silhouette analyses using these features indicate that the optimal k value is $k = 5$ at 20% and 60% CWA_U and $k = 4$ at 40% and 80% CWA_U. We selected $k = 4$ for k-means clustering across all CWA_U thresholds to guide our analysis for consistency and better interpretation.

Table 10 shows the centroids of all features used for the clustering analysis. Each cluster contains more similar samples across all CWA_U thresholds.

The following general patterns were observed across all thresholds:

1. Students in clusters who engage intensively with simulations and workshops tend to have higher weighted averages, indicating better performance.
2. Clusters with fewer students often show more focused and efficient study habits, achieving higher weighted averages despite lower overall engagement.
3. The balance and variation in study habits play a crucial role in academic performance, as evidenced by the higher weighted averages of students who manage their study time effectively across different resources.

We also ran ANOVA tests, showing that such a relationship exists (p-value < 0.05). Then, we conducted Tukey's post hoc tests to see if there were significant differences in student performance for each pair of clusters.

The consistent significant p-value in the post hoc analysis highlights the strength of the clustering approach in identifying distinct groups of students based on their behavior and performance patterns (as we can see from Figure 14 and Table 11, where the clustering results are more distinguishable than the saw in Section 4.5). This insight can help design targeted educational strategies and support mechanisms to enhance student success. Our main conclusions are that higher engagement in simulations and workshops corresponds to better academic outcomes, and the distinct clusters represent different levels of student performance and engagement. These findings underline the importance of active participation in simulations and workshops, which can inform targeted interventions to support students effectively.

Additionally, analyzing the top features improves clustering by increasing the number of paired clusters with significant differences compared to clustering with all features,

leading to more meaningful insights into student performance patterns. However, despite these improvements, our analysis of the Linearity Score using top features revealed no clear relationship between linearity and weighted average. This suggests that the selected feature set may not fully capture the factors influencing a student's navigation path.

Table 10. Number of students per group (N), average (μ), and standard deviation (σ) for top feature importance values per cluster for the four different CWA_U (20%, 40%, 60%, and 80%).

CWA_U	Features	Cluster 0 (N = 107)	Cluster 1 (N = 176)	Cluster 2 (N = 156)	Cluster 3 (N = 224)
		$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$	$\mu \pm \sigma$
20%	mean_simulations_attempt	0.16 ± 0.06	0.19 ± 0.07	0.00 ± 0.00	0.26 ± 0.09
	mean_simulations_time	5.87 ± 5.10	19.06 ± 4.53	0.00 ± 0.00	4.59 ± 4.17
	mean_workshops_attempt	0.24 ± 0.13	0.20 ± 0.07	0.68 ± 0.30	0.26 ± 0.12
	mean_workshops_time	5.85 ± 4.93	11.62 ± 5.87	21.74 ± 16.34	4.26 ± 4.20
	weighted_average	1.95 ± 0.63	3.15 ± 0.78	4.12 ± 0.72	4.54 ± 0.46
		Cluster 0 (N = 96)	Cluster 1 (N = 193)	Cluster 2 (N = 92)	Cluster 3 (N = 282)
40%	mean_simulations_attempt	0.36 ± 0.11	0.37 ± 0.10	0.15 ± 0.12	0.52 ± 0.16
	mean_simulations_time	11.23 ± 9.35	35.73 ± 9.61	8.67 ± 8.79	8.52 ± 7.49
	mean_workshops_attempt	0.56 ± 0.21	0.58 ± 0.19	1.26 ± 0.37	0.72 ± 0.27
	mean_workshops_time	12.86 ± 9.93	32.71 ± 11.50	35.82 ± 24.77	9.67 ± 8.36
	weighted_average	2.19 ± 0.70	3.29 ± 0.77	4.13 ± 0.63	4.47 ± 0.46
		Cluster 0 (N = 85)	Cluster 1 (N = 204)	Cluster 2 (N = 168)	Cluster 3 (N = 206)
60%	mean_simulations_attempt	0.55 ± 0.11	0.54 ± 0.12	0.81 ± 0.26	0.74 ± 0.22
	mean_simulations_time	20.70 ± 15.81	52.54 ± 13.07	19.02 ± 12.58	8.89 ± 9.04
	mean_workshops_attempt	0.74 ± 0.29	0.85 ± 0.26	1.46 ± 0.31	0.75 ± 0.22
	mean_workshops_time	18.55 ± 13.51	48.21 ± 18.43	21.71 ± 12.80	9.04 ± 9.58
	weighted_average	2.05 ± 0.64	3.47 ± 0.68	4.33 ± 0.55	4.43 ± 0.50
		Cluster 0 (N = 79)	Cluster 1 (N = 216)	Cluster 2 (N = 160)	Cluster 3 (N = 208)
80%	mean_simulations_attempt	0.76 ± 0.16	0.73 ± 0.20	1.18 ± 0.31	1.06 ± 0.34
	mean_simulations_time	24.73 ± 20.12	66.78 ± 16.80	25.73 ± 17.25	10.31 ± 10.17
	mean_workshops_attempt	0.85 ± 0.36	0.99 ± 0.29	1.70 ± 0.35	0.89 ± 0.26
	mean_workshops_time	18.35 ± 13.71	51.95 ± 19.11	24.28 ± 13.89	9.45 ± 10.16
	weighted_average	2.01 ± 0.61	3.49 ± 0.68	4.34 ± 0.55	4.49 ± 0.44

Table 11. p-values result for each cluster at different CWA_U thresholds with top features. Bold values indicate significant clusters ($p \leq 0.05$).

CWA_U	Cluster	p-Value		
		Cluster 1	Cluster 2	Cluster 3
20%	Cluster 0	**0.004**	**0.001**	**0.001**
	Cluster 1		**0.001**	**0.001**
	Cluster 2			**0.001**
40%	Cluster 0	**0.005**	**0.001**	**0.001**
	Cluster 1		**0.001**	**0.001**
	Cluster 2			**0.001**
60%	Cluster 0	**0.001**	**0.001**	**0.001**
	Cluster 1		**0.001**	**0.001**
	Cluster 2			0.900
80%	Cluster 0	**0.001**	**0.001**	**0.001**
	Cluster 1		**0.001**	**0.001**
	Cluster 2			0.477

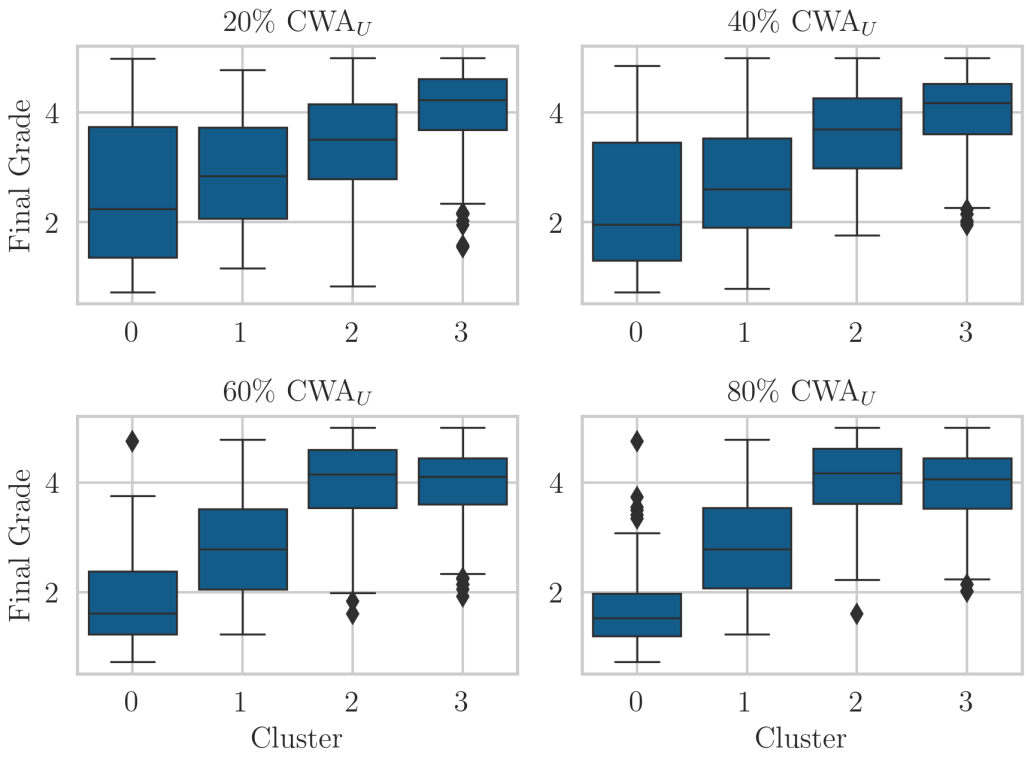


Figure 14. Students' final grade performance for each cluster at different CWA$_U$ thresholds with top features.

5. Discussions and Conclusions

We explored machine learning techniques to predict student performance in the non-linear logic reasoning PPVU course at the Universidad de Antioquia, focusing on its flexible learning path. This non-linear structure requires dynamically updating the features at the moment of prediction to capture relevant student interactions accurately. To address this, we introduced the concept of a Feature Aggregation Time Point (FATP), where features are extracted based on the student's Cumulative Weight Assessment (CWA), ensuring the feature set reflects student activity up to a specific performance threshold.

Since no standard set of features exists for predicting performance in non-linear courses, we extracted engagement, behavior, and performance features. We used these features to train a binary classifier for early student performance prediction and compared it against a non-machine-learning baseline. Our analysis confirmed that when updated at the FATP, these features effectively represent and predict student performance, significantly improving classification accuracy.

We adopted a three-phase methodology: (a) binary classification to predict whether a student would achieve a certificate-granting grade, (b) cluster analysis to identify online learning patterns among students, and (c) a correlation analysis between student performance and the behavioral clusters, including a feature importance analysis through the selected machine learning models. The non-linear nature of the PPVU course required selecting the optimal lower Cumulative Weight Assessment (CWA$_L$) to ensure balanced data samples and setting the FATP based on the student's progress relative to the upper threshold (CWA$_U$).

Our analysis of students who completed at least 55% of the course revealed that models significantly outperformed the baseline, with Logistic Regression demonstrating strong performance across all CWA$_U$ thresholds. Although Random Forest performed better at specific thresholds, Logistic Regression offered a favorable balance of accuracy, computational efficiency, and interpretability. The combination of behavior and perfor-

mance features yielded the best results, showing that while engagement features contribute to predictions, their importance diminishes when combined with other feature types. Our approach enabled early predictions with an F1-score of 0.73 at 20% CWA_U and 0.76 at 40% CWA_U for binary classification.

Our feature importance analysis showed that certain engagement features, such as course navigation and forum participation, had relatively low importance across all CWA_U thresholds. However, engagement features related to time spent on educational resources were consistently more influential. Our clustering analysis revealed how students interact with different resources (multimedia, video, simulations, and workshops) and how these behaviors relate to performance at varying CWA_U thresholds. Early and sustained engagement with multimedia and video was key to success, especially in high-performing clusters. As students progressed to higher CWA_U levels, simulations and workshops became more influential, highlighting their role in assessing later-stage performance. We also identified distinct self-regulated learning (SRL) patterns: students who strategically engaged with simulations and workshops consistently outperformed others across all thresholds. These high-performing clusters demonstrated strong SRL practices, such as time management and resource prioritization. They also showed lower linearity scores and higher forum participation, reflecting an adaptive learning approach that favored exploration over rigid sequences. This flexible behavior highlights the value of diverse engagement strategies tailored to individual learning needs.

From these findings, two actionable interventions arise: (1) encouraging early engagement with multimedia resources to support lower-performing students and (2) providing targeted assistance with simulations and workshops as students progress through the course. These strategies could help improve learning outcomes and support students in achieving better performance as they advance.

Clustering students based on top features resulted in distinct groups with identifiable engagement and performance patterns. The consistent results from Tukey's post hoc tests validate the robustness of our clustering approach in capturing meaningful differences in student performance, and highlight the value of feature importance analysis in learning analytics.

Overall, our study demonstrates the effectiveness of machine learning techniques in predicting student performance in non-linear Moodle courses. The insights gained from our analysis, including using FATP and CWA thresholds, offer a framework for future research and for designing targeted intervention strategies that enhance student engagement and academic outcomes. The recurring patterns identified across various CWA_U thresholds provide educators with a structured approach to guiding students toward improved performance.

These results also address several gaps identified in prior research. By introducing FATP based on students' progress rather than course time or assignment availability, we provide a practical solution for performance prediction in non-linear, self-paced courses, an issue overlooked in existing PTP approaches. Moreover, our integration of feature importance analysis with clustering and model retraining responds to calls in the literature for more robust validation of selected features. Finally, our work demonstrates the synergy between these areas by combining performance prediction and pattern analysis within a unified framework. It extends current learning analytics methodologies to better support adaptive interventions. Despite these contributions, our study has limitations. We focused on a single self-paced course, which may affect generalizability, and the sample size of the final dataset may reduce stability.

6. Future Work

Future research can focus on several key areas to enhance the predictive models and their applicability. One priority is expanding and refining features by investigating additional data points, such as more detailed student interactions with course components, which can lead to improved predictive accuracy. Incorporating self-regulated learning (SRL) indicators, such as goal-setting, time management, and self-monitoring behaviors, help-seeking tendencies, where students proactively explore and engage with diverse learning resources, could enhance the feature set and provide deeper insights into student performance. Additionally, developing real-time predictive models that provide immediate feedback and support to students will allow for timely interventions, enhancing their learning experience.

Another important direction is the design of personalized intervention strategies based on clustering analysis insights. These tailored strategies can address the specific needs of different student clusters, making support mechanisms more effective and aligning these interventions with SRL principles to encourage students to develop stronger self-regulation, help-seeking behaviors, and adaptive learning strategies. Integrating these predictive models into educational platforms like Moodle can automate insights and recommendations, providing valuable feedback directly within the learning environment.

Testing the scalability and generalizing the proposed methodology across different courses, disciplines, and institutions will ensure broader applicability and robustness. Future research can build on our findings by addressing these areas to create more robust, scalable, and effective predictive models that support student success in various learning environments.

Author Contributions: Conceptualization, J.M. and C.M.-C.; methodology, J.M. and C.M.-C.; data curation, J.M.; writing—original draft preparation, J.M.; writing—review and editing, C.M.-C.; supervision, L.F., N.G.-G. and C.M.-C.; project administration, L.F. and N.G.-G.; funding acquisition, L.F. and N.G.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the General Royalty System of Colombia (SGR—Sistema General de Regalías) BPIN-2021000100186. We thank Ude@ and the EDidactica research group for providing the Moodle backup that made this research possible.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Students who participated in the course were informed about the study, and by enrolling in the course, they agreed to the terms of participation.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare that they have no conflicts of interest. Although the author Carlos Mendoza-Cardenas was employed by Twitch Interactive Inc. at the time of working on this research, he declares that there was no commercial or financial interest. The funders had no role in the design of the study; in the collection, analysis or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1 presents a sample of the dataset used for analysis, showing features updated at 80% CWA_U. It includes the hashed IDs of five students, the final grade as the target feature, the weighted average as the performance feature, 24 engagement features, and 10 behavior features.

Table A1. Sample of the dataset used.

Userid	**00554…**	**0068d…**	**01bb1…**	**01bcd…**	**03340…**
final_grade	4.53	2.91	3.69	2.05	4.17
weighted_average	4.57	4.36	4.18	2.85	4.34
days	101	42	5	105	68
mean_interactions	49.5	22.18	95.33	39.41	53.62
std_interactions	39.12	17.71	54.24	34.25	63.61
week	416	233	137	584	408
weekend	178	11	149	86	21
course_navigation	111	62	33	91	12
forum_viewed	28	0	0	69	0
post_created	0	0	0	0	0
discussion_viewed	119	0	0	141	0
discussion_created	0	0	0	0	0
forum_searched	1	0	0	1	0
outline_viewed	2	0	0	1	0
mean_multimedia_views	0.8	0.07	0.0	1.07	0.07
std_multimedia_views	0.89	0.0	0.0	0.93	0.0
mean_video_views	1.0	0.0	0.0	0.83	0.0
std_video_views	2.19	0.0	0.0	0.5	0.0
mean_video_time	14.03	0.0	0.0	20.94	0.0
std_video_time	8.42	0.0	0.0	1.29	0.0
mean_multimedias_time	12.15	3.18	0.0	13.12	0.0
std_multimedias_time	7.56	0.0	0.0	0.0	0.0
mean_simulations_grade	3.06	2.85	1.98	1.34	1.91
std_simulations_grade	0.59	0.19	0.19	2.28	0.65
mean_simulations_time	52.57	80.54	8.4	12.96	1.72
std_simulations_time	11.96	20.62	10.93	22.92	0.42
mean_simulations_attempt	0.67	0.67	1.33	0.83	1.83
std_simulations_attempt	0.0	0.0	0.0	0.5	0.96
mean_workshops_grade	3.78	3.03	2.83	2.4	2.57
std_workshops_grade	0.51	0.32	0.69	1.84	0.69
mean_workshops_time	28.03	63.34	19.61	38.86	3.81
std_workshops_time	7.24	36.18	12.83	47.24	3.64
mean_workshops_attempt	0.83	0.67	1.5	1.0	1.83
std_workshops_attempt	0.0	0.0	0.55	0.45	0.84
mean_prior_knowledge_grade	2.75	1.5	0.0	1.75	0.75
std_prior_knowledge_grade	1.77	0.0	0.0	1.06	0.35
mean_prior_knowledge_time	41.52	13.92	1.56	26.77	11.07
std_prior_knowledge_time	26.87	0.0	0.0	24.89	7.66
mean_lessons_time	63.82	0.0	0.0	14.24	0.0
std_lessons_time	0.0	0.0	0.0	0.0	0.0
mean_lessons_attempts	0.5	0.0	0.0	0.5	0.0
std_lessons_attempts	0.0	0.0	0.0	0.0	0.0

Appendix B

The following algorithm calculates the linearity score for a given sequence of assessments. This score measures how closely a student's progression adheres to a linear sequence:

Algorithm A1 Calculate Linearity.

Require: A sequence of assessment steps, `path`
Ensure: The linearity score

$c \leftarrow 0$
if `path`$[0] \neq 1$ **then**
  $c \leftarrow c - 1$
end if
for $i = 1$ **to** `len(path)` $- 1$ **do**
  `current_unit` $\leftarrow$ `path`$[i]$
  `previous_unit` $\leftarrow$ `path`$[i-1]$
  if `current_unit` $=$ `previous_unit` **or** `current_unit` $-$ `previous_unit` $= 1$ **then**
    $c \leftarrow c + 1$
  end if
end for
if `len(path)` > 1 **then**
  return $\max\left(\frac{c}{\texttt{len(path)}-1}, 0\right)$
else
  return 1
end if

This function works as follows:

- Initialize a counter, c, to track linear steps in the sequence.
- If the first assessment is not unit 1, c is decremented to penalize the sequence.
- For each subsequent unit, c is incremented if the unit is the same as the previous unit or follows it sequentially.
- The final score is normalized by the total number of transitions in the sequence, ensuring a range between 0 and 1.
- If the sequence has only one assessment, a perfect linearity score of 1 is returned.

This algorithm computes the Linearity Score, which reflects the orderliness of a student's progression through assessments.

References

1. Li, R.; Singh, J.; Bunk, J. Technology tools in distance education: A review of faculty adoption. In *EdMedia+ Innovate Learning*; Association for the Advancement of Computing in Education (AACE): Amsterdam, The Netherlands, 2018; pp. 1982–1987.
2. Ellis, R.K. *Learning Management Systems*; American Society for Training & Development (ASTD): Alexandria, VI, USA, 2009.
3. Llamas, M.; Caeiro, M.; Castro, M.; Plaza, I.; Tovar, E. Use of LMS functionalities in engineering education. In Proceedings of the 2011 Frontiers in Education Conference (FIE), Rapid City, SD, USA, 12–15 October 2011; p. S1G-1.
4. Kloft, M.; Stiehler, F.; Zheng, Z.; Pinkwart, N. 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, Doha, Qatar, 25 October 2014; pp. 60–65.
5. Seo, J.T.; Park, B.N.; Kim, Y.g.; Yeon, K.W. Analysis of LMS Data of Distance Lifelong Learning Center Learners and Drop-out Prediction. *J. Hum.-Centric Sci. Technol. Innov.* **2021**, *1*, 23–32. [CrossRef]
6. Nordin, N.; Norman, H.; Embi, M.A. Technology acceptance of massive open online courses in Malaysia. *Malays. J. Distance Educ.* **2015**, *17*, 1–16. [CrossRef]
7. Nabizadeh, A.H.; Goncalves, D.; Gama, S.; Jorge, J.; Rafsanjani, H.N. Adaptive learning path recommender approach using auxiliary learning objects. *Comput. Educ.* **2020**, *147*, 103777. [CrossRef]
8. Alsadoon, E. The impact of an adaptive e-course on students' achievements based on the students' prior knowledge. *Educ. Inf. Technol.* **2020**, *25*, 3541–3551. [CrossRef]

9. Chen, Y.; Li, X.; Liu, J.; Ying, Z. Recommendation system for adaptive learning. *Appl. Psychol. Meas.* **2018**, *42*, 24–41. [CrossRef]
10. Mercado, J.; Mendoza, C.H.; Ramirez-Salazar, D.A.; Valderrama, A.; Gaviria-Gomez, N.; Botero, J.F.; Fletscher, L. Work in progress: A didactic strategy based on Machine Learning for adaptive learning in virtual environments. In Proceedings of the 2023 IEEE World Engineering Education Conference (EDUNINE), Bogota, Colombia, 12–15 March 2023; pp. 1–4.
11. Vasquez Diaz, K.R.; Iqbal, J. Challenges Faced by International Students in Understanding British Accents and Their Mitigation Strategies—A Mixed Methods Study. *Educ. Sci.* **2024**, *14*, 784. [CrossRef]
12. Walker, A.; Diaz, K.R.V.; McKie, D.; Iqbal, J. Enquiry-Based Learning Pedagogy—Design, Development and Delivery of a Reproducible Robotics Framework. In Proceedings of the International Congress on Information and Communication Technology, London, UK, 19–22 February 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 363–374.
13. Conijn, R.; Snijders, C.; Kleingeld, A.; Matzat, U. Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Trans. Learn. Technol.* **2016**, *10*, 17–29. [CrossRef]
14. Adnan, M.; Habib, A.; Ashraf, J.; Mussadiq, S.; Raza, A.A.; Abid, M.; Bashir, M.; Khan, S.U. Predicting at-risk students at different percentages of course length for early intervention using machine learning models. *IEEE Access* **2021**, *9*, 7519–7539. [CrossRef]
15. Tomasevic, N.; Gvozdenovic, N.; Vranes, S. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Comput. Educ.* **2020**, *143*, 103676. [CrossRef]
16. Waheed, H.; Hassan, S.U.; Nawaz, R.; Aljohani, N.R.; Chen, G.; Gasevic, D. Early prediction of learners at risk in self-paced education: A neural network approach. *Expert Syst. Appl.* **2023**, *213*, 118868. [CrossRef]
17. Cohausz, L.; Tschalzev, A.; Bartelt, C.; Stuckenschmidt, H. Investigating the Importance of Demographic Features for EDM-Predictions. In Proceedings of the 16th International Conference on Educational Data Mining, Bengaluru, India, 11–14 July 2023; International Educational Data Mining Society: Bengaluru, India, 2023.
18. Riestra-González, M.; del Puerto Paule-Ruíz, M.; Ortin, F. Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Comput. Educ.* **2021**, *163*, 104108. [CrossRef]
19. Hoq, M.; Brusilovsky, P.; Akram, B. Analysis of an Explainable Student Performance Prediction Model in an Introductory Programming Course. In Proceedings of the 16th International Conference on Educational Data Mining, Bengaluru, India, 11–14 July 2023; International Educational Data Mining Society: Bengaluru, India, 2023.
20. Rohani, N.; Gal, K.; Gallagher, M.; Manataki, A. Early prediction of student performance in a health data science MOOC. In Proceedings of the 16th International Conference on Educational Data Mining, Bengaluru, India, 11–14 July 2023; International Educational Data Mining Society: Bengaluru, India, 2023.
21. Mao, Y.; Khoshnevisan, F.; Price, T.; Barnes, T.; Chi, M. Cross-lingual adversarial domain adaptation for novice programming. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2022; Volume 36, pp. 7682–7690.
22. Anderson, T.; Annand, D.; Wark, N. The search for learning community in learner paced distance education: Or, 'Having your cake and eating it, too!'. *Australas. J. Educ. Technol.* **2005**, *21*, 222–241. [CrossRef]
23. Kuzilek, J.; Hlosta, M.; Zdrahal, Z. Open university learning analytics dataset. *Sci. Data* **2017**, *4*, 1–8. [CrossRef]
24. Alturki, S.; Cohausz, L.; Stuckenschmidt, H. Predicting Master's students' academic performance: An empirical study in Germany. *Smart Learn. Environ.* **2022**, *9*, 38. [CrossRef]
25. Cenka, B.A.N.; Santoso, H.B.; Junus, K. Analysing student behaviour in a learning management system using a process mining approach. *Knowl. Manag. E-Learn.* **2022**, *14*, 62–80.
26. Bessadok, A.; Abouzinadah, E.; Rabie, O. Exploring students digital activities and performances through their activities logged in learning management system using educational data mining approach. *Interact. Technol. Smart Educ.* **2023**, *20*, 58–72. [CrossRef]
27. Moodle Docs. *Database Schema Introduction*. Available online: https://moodledev.io/docs/apis/core/dml/database-schema (accessed on 5 April 2025).
28. Siemens, G.; Baker, R.S.d. Learning analytics and educational data mining: Towards communication and collaboration. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, 29 April–2 May 2012; pp. 252–254.
29. Mercado, J.; Mendoza-Cardenas, C.; Gaviria-Gomez, N.; Botero, J.F.; Fletscher, L. Predicting Student Performance in a Non-linear Self-Paced Moodle Course. In Proceedings of the 22nd Latin American and Caribbean Conference for Engineering and Technology (LACCEI), Zaragoza, Spain, 22–23 October 2024.
30. LeCun, Y.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient backprop. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 9–50.
31. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.
32. Kim, T.K. Understanding one-way ANOVA using conceptual figures. *Korean J. Anesthesiol.* **2017**, *70*, 22. [CrossRef]
33. Zimmerman, B.J. Becoming a self-regulated learner: An overview. *Theory Into Practice* **2002**, *41*, 64–70. [CrossRef]
34. Viberg, O.; Khalil, M.; Baars, M. Self-regulated learning and learning analytics in online learning environments: A review of empirical research. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20), Frankfurt, Germany, 23–27 March 2020; pp. 524–533.

35. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
36. Joshi, P. *Python Machine Learning Cookbook*; Packt Publishing Ltd.: Birmingham, UK, 2016.
37. Saarela, M.; Jauhiainen, S. Comparison of feature importance measures as explanations for classification models. *SN Appl. Sci.* **2021**, *3*, 272. [CrossRef]
38. Cawley, G.C.; Talbot, N.L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
39. Trivedi, S.; Pardos, Z.A.; Heffernan, N.T. Clustering students to generate an ensemble to improve standard test score predictions. In Proceedings of the Artificial Intelligence in Education: 15th International Conference, AIED 2011, Auckland, New Zealand, 28 June–2 July 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 377–384.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.