



## A Multi-Label Classification-Based Approach for Dynamic Learner Grouping in Social Learning Environments

Noureddine Gouasmi<sup>1,2</sup> , Mahnane Lamia<sup>3</sup> , Yacine Lafifi<sup>3</sup> 

<sup>1</sup> Badji Mokhtar - Annaba University, Annaba 23000, Algeria

<sup>2</sup> Labstic, 08 Mai 45 University, B.P.401, Guelma 24000, Algeria

<sup>3</sup> LRS Laboratory, Badji Mokhtar - Annaba University, Annaba 23000, Algeria

Corresponding Author Email: [la.fifi.yacine@univ-guelma.dz](mailto:la.fifi.yacine@univ-guelma.dz)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310122>

### ABSTRACT

**Received:** 16 November 2025

**Revised:** 7 January 2026

**Accepted:** 20 January 2026

**Available online:** 31 January 2026

#### **Keywords:**

*collaborative learning, dynamic grouping, multi-label classification, K-nearest neighbors, social learning environments, learning analytics, student behaviour modeling*

Social networks have become integral to learners' daily activities, offering new opportunities for collaborative learning through online interaction and knowledge sharing. However, forming effective learner groups in such environments remains a complex and time-consuming task, particularly when considering diverse behavioural and interaction patterns. This study proposes a dynamic learner grouping approach based on multi-label classification within a social learning network. The method leverages behavioural attributes derived from learners' activities, including communication, engagement, and participation patterns, to assign multiple labels to each learner. A Multi-Label K-Nearest Neighbors (ML-KNN) algorithm is then employed to identify learner profiles and support the formation of heterogeneous groups that promote balanced collaboration. The proposed approach is implemented in a social learning system and evaluated through a controlled experiment involving undergraduate students. The performance of groups generated by the proposed method is compared with that of self-formed groups based on learner affinity. Experimental results demonstrate that the proposed grouping strategy leads to improved learning outcomes, as reflected in post-test performance, while also enhancing learners' motivation and interaction. These findings indicate that integrating multi-label classification with behavioural analytics provides an effective and scalable solution for dynamic group formation in social learning environments.

## 1. INTRODUCTION

Since their widespread appropriation, social networks have become indispensable in our daily lives. Learners often use them to search for documents, communicate and collaborate or share available information between them [1].

As well, for promoting collaborative learning, an interdisciplinary field, associating computer technology and educational sciences, is born [2]. This field, called Computer-Supported Collaborative Learning (CSCL), aims to improve cooperation and communication between learners through electronic means connected to networks. Thus, group formation is an important task because it determines the efficiency of collaboration between group members [3].

To obtain effective groups, two important aspects must be considered: the attributes used for grouping learners and the nature of the groups to form.

The attributes, that describe the learners to group, are of different natures, such as demographic variables (gender, age), grades, learning styles, learning activity, and personality [4]. Regarding groups nature, the groups are formed in three possible ways [5]:

- grouping learners with similar profile characteristics: Homogeneous grouping.

- grouping learners with different profile characteristics: Heterogeneous grouping.
- grouping learners with similar profile for some characteristics, but different for others: Mixed grouping.

Choosing the most effective method for forming groups is also an important and difficult task. Several methods have been proposed in this area, including evolutionary algorithms [6, 7], clustering [8, 9] and probabilistic algorithms [10].

These methods allow learners to be assigned to a group; each group being considered as a class to which the learner belongs. However, considering the learners, they may exhibit one or more facets during their learning, hence the importance of using a multi-label classification, which assigns each learner one, two, or all of these facets, depending on their learning activities.

Because of the above, this paper introduces a novel approach, based on multi-label classification, for forming high-performing learning groups, thereby improving collaboration efficiency. This approach is adopted by an SNL environment called: CLISON. The aim of this research is to answer the following question: What is the impact of the proposed approach on the performance of the learner?

The paper is structured in 5 sections. After the introduction,

Section 2 presents several relevant works. Section 3 provides a comprehensive explanation of the CLISON learning system, and mainly the proposed approach for grouping learners for collaborative activities, by presenting the multi-label classification and the ML-KNN algorithm used for the group formation. Section 4 proposes a detailed description of the experimentation of the proposed approach and an analysis of the results obtained. Lastly, the final section presents the study's findings and proposes future work to expand upon this research.

## 2. RELATED WORKS

Classification is a domain of data mining whose aim is to build a predictive model that decides how to classify or assign a new element to a class or label [11]. There are four types of classification methods [11, 12]:

1. **Binary classification:** The element to be classified may, or not, belong to a single, unique class.
2. **Multi-class classification:** The element to be classified can belong to only one class from a predefined set of classes.
3. **Multi-label classification:** The element to be classified can belong to one or more classes from a predefined set of classes (called labels).
4. **Multidimensional classification:** The element to be classified can be assigned a value, from a defined interval, for each of the labels to be assigned.

In recent years, multi-label classification has been used in several applications, such as: text classification [13-15], medical field [16], biology [17] and ornithology [18].

In the field of e-learning, multi-label classification has been used in various applications, including: assessing students engagement in an e-learning environment [19], automatic recognition of online learners' emotions [20], predicting e-learning resources quality [21] and predicting learning styles [12, 22]. However, as far as we know, there are no studies relating to the group formation in online learning systems.

There are several different attributes used to form groups, such as learning styles [6, 23, 24], personality traits [7, 25], learners' activities in the online learning system [8, 26]. Krouska and Virvou [6] presented a multi-characteristic solution based on three dimensions: academic (such as creating skill), cognitive (such as learning styles), and social (such as contributions badge), and Muuro et al. [27] proposed the collaborative competency level as a way to increase the interactions between group members.

Regarding grouping methods, the researches [6, 7, 23, 28] use a genetic algorithm to perform the group formation, while the authors [29] adopt a particle swarm optimization algorithm. The paper [9] applies two machine learning algorithms for grouping learners: SK-means and Expectation Maximization clustering algorithms. Further, the study [8] employs K-means clustering algorithm. Finally, the researches [24, 27, 26] introduce particular algorithms for forming groups of learners.

**Table 1.** Summary table on some group formation methods

Authors	Grouping Method	Groups Type	Grouping Attributes
Krouska and Virvou [6]	Genetic algorithm	Intra-heterogeneity and inter-homogeneity	Academic dimension, cognitive dimension, social dimension
Revelo-Sanchez et al. [7]	Genetic algorithm	Homogeneous	Big five personality traits
Shanmuganeethi et al. [8]	K-means clustering algorithm	Heterogeneous	Learners' activities
Maina et al. [9]	SK-means and Expectation Maximization clustering algorithms	Homogenous and heterogenous	Moodle forum
Sukstrienwong [23]	Genetic algorithm	Intra-heterogeneity and inter-homogeneity	Learning styles
Haq et al. [24]		Balanced groups of heterogeneous students	Learning styles and knowledge level
Bourkougou et al. [26]		Homogenous and heterogenous	Learners' activities and preferences
Muuro et al. [27]		Heterogeneous	Learners' collaborative competency level
Liang et al. [28]	Genetic algorithm	Homogenous and heterogenous	Engagement, Course Score, BookRoll Score, Moodle Score, Group Score, Friendship
Wang et al. [29]	Particle Swarm Optimization algorithm		Knowledge and interest of learners
Our approach	ML-KNN multi-label classification	Heterogeneous	Assiduity, engagement and collaboration skills

Table 1 summarizes some of the research works described above.

In this context, we present a social network-based learning system called: Collaborative Learning in Social Networks (CLISON), offering features and tools for collaborative learning, and proposing a new approach for group formation, considering multiple aspects of collaborative work: communication, interactions, and activities of the learner, and a novel grouping approach based on a multi-label classification algorithm.

## 3. A NEW METHOD FOR GROUPING LEARNERS BASED ON MULTI-LABEL CLASSIFICATION

The aim of our work is to propose a new group formation method, based on multi-label classification, to enhance learners' performances. This new grouping method is adapted by a Social Network Learning system called: CLISON.

Before we present our grouping method, we introduce the main features of the system, which are:

- To provide lessons as posts;
- To allow students and teachers to comment and like

posts;

- To support instant messaging between the system users;
- To offer collaborative learning tools;
- To answer MCQ individually and to solve exercises (individually or collaboratively).

Figures 1 and 2 show some screens captured from the system interface.

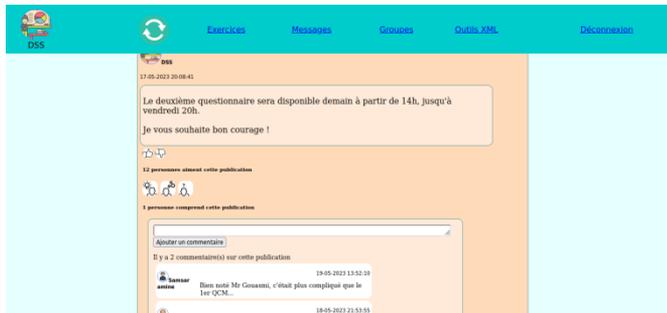


Figure 1. CLISON's home page



Figure 2. CLISON's collaborative activities page

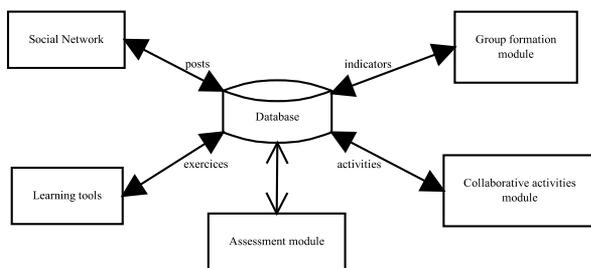


Figure 3. CLISON system architecture

The social network-based learning system CLISON consists of five important modules (see Figure 3):

- A social network offers the features of a social media, such as publications, comments, messages, etc. It allows students to communicate with each other and with the teacher, and it provides the teacher with the capability of sending to students, lessons and activities.
- A group formation module is used to form groups of students whose aim is to carry out collaborative activities. In this module, there are two supported grouping methods: an automatic grouping, using a multi-label classification, and a manual grouping, where the students join together by affinity. This module is adaptable. It can be supplemented by other grouping methods (for instance, random grouping, using genetic algorithm, etc.).
- A collaborative activities module allows students to complete the activities, requested by the teacher, in

groups (groups formed by the previous module).

- Learning tools which are tools proposed to help students solve the exercises and carry out the activities proposed on the system.
- An assessment module presents students with some exams, in the form of Multiple Choice Questionnaires (MCQs), and allows them to fulfill them.

In the following, we describe the multi-label classification method used to form the groups of students.

### 3.1 Multi-Label K-Nearest Neighbours, classification algorithm

A classification process is a systematic distribution of beings, things and concepts into classes or categories. This process is carried out according to several attributes describing them.

The set of attributes in a classification dataset is divided into two subsets: the first subset involves the attributes used to predict the class to which the object can be associated with, while the second subset contains the class or label assigned to each object. These two subsets are analyzed by a classification algorithm in order to identify the correlation between the input attribute and the output classes. Once a trained model is obtained, it can be used to process the attributes set of new data samples, obtaining a class prediction [30].

The multi-label classification is a predictive data mining task with several real-world applications, such as the automatic labeling of text, image, music, and video. The learning process can be accomplished through various approaches, such as data transformation, method adaptation, and the use of ensembles of classifiers [11]. Typically, the problem of multi-label classification resides on representing a population by a feature vector and then associating it with a set of labels [31].

One of the widely used multi-label classification algorithms is the Multi-Label K-Nearest Neighbors (ML-KNN), which is an adaptation of the KNN classification algorithm.

Listing 1 describes the ML-KNN algorithm, proposed by Zhang and Zhou [32].

#### Listing 1. ML-KNN algorithm

**Train** the classifier:

**Calculate** the priori probabilities of each label on the training set

**Calculate** the conditional probabilities of each label regarding the labels of its K-Nearest Neighbors

**Assign** labels to a new instance:

**Locate** the K-Nearest Neighbors of the new instance

**Calculate** the maximum posterior probabilities (MAPs) from the conditional probabilities obtained previously

**Generate** the labels of the new instance from the MAP probabilities.

The priori probabilities are the probability of appearance of a label in the dataset, measured for each label. Likewise, the conditional probabilities for a given label are the ratio of instances in the dataset, having this label, whose K-Nearest Neighbors have the same label. Finally, maximum posterior probabilities (MAPs) relate to the estimation of the most probable value of assigning a label to a new instance, considering both the a priori probability and the conditional

probability.

This algorithm was used to label learners in a class, in order to form heterogeneous groups where all labels are present.

### 3.2 Multi-label classification based learners' grouping

In our system, we form heterogeneous groups of learners. The grouping is based on each learner's activities on the learning system.

Each learner activity is represented by seven attributes:

- the number of posts published on the learning social network,
- the number of added comments,
- the number of added likes/dislikes,
- the number of messages written and send,
- the sentiments expressed in the messages,
- the grades in the individual activities, and,
- the number of completed individual activities.

The first five attributes address the identification of learners' abilities in relation to communication and collaboration. These two competences, along with critical thinking and creativity, constitute important soft skills that every learner must acquire [33].

Based on these attributes, we want to answer three important questions about learners: how often does the learner participate in activities on the site? Is their participation active? And finally, do they interact with their peers? The answers to these questions should allow us to form productive groups, based on the assumption that assiduous and engaged students will ensure the completion of collaborative activities, while collaborative students will ensure better communication within the group.

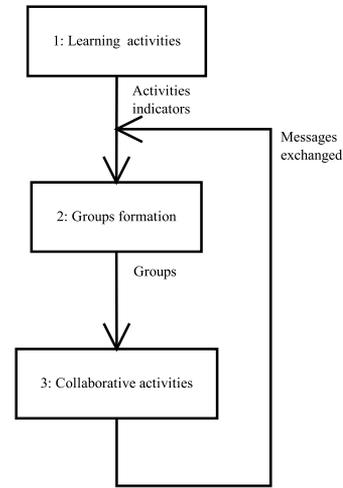
Therefore, using the above attributes, learners are tagged according to 3 labels:

- **Collaborative learner:** who actively communicates during the course and exchanges with his peers and the teachers.
- **Assiduous learner:** who regularly attends and participates in learning activities.
- **Engaged learner:** who actively participates in activities and often expresses positive feelings.

Groups are then formed, while trying to ensure that each of the labels appear at least once among the learners belonging to it. Figure 4 represents the conceptual structure of the group formation module. Hence, in this module, after collecting parameters relating to learner activities on the system, learners are labeled using the ML-KNN classifier. Then, according to the label classes, the learners are assigned to groups, ensuring that the labels are mixed within each group (see Listing 2).

**Listing 2.** Group formation algorithm

**Input:** Sets of labels  
**Output:** Sets of groups  
**While** all label classes are not empty  
    Calculate the cardinality of each non-empty label class  
    Sort the labels from fewest to most learners  
    **Repeat** for the fewest label class *i*  
        Assign a learner from label *i* to each group  
        Remove the learner from the other labels, if any  
    **Until** the class of the *i*<sup>th</sup> label is empty  
**End While**



**Figure 4.** Group formation module

The following example illustrates the proposed algorithm. Consider a set of nine learners {id1,...,id9} to be assigned to 3 groups.

The learners were classified as follows:

$$\begin{aligned}
 \text{engaged} &= \{\text{id1}, \text{id8}, \text{id9}\}, \\
 \text{collaborative} &= \{\text{id1}, \text{id2}, \text{id6}, \text{id7}\}, \\
 \text{assiduous} &= \{\text{id1}, \text{id2}, \text{id3}, \text{id4}, \text{id5}\}.
 \end{aligned}$$

In the first step of the group formation, we assign the learners belonging to the label class with the fewest members: engaged. Every learner from this class is assigned to each of the groups successively:

$$G1 = \{\text{id1}\}, G2 = \{\text{id8}\}, G3 = \{\text{id9}\}.$$

The affected learners are removed from the labels classes:

$$\begin{aligned}
 \text{engaged} &= \{\}, \\
 \text{collaborative} &= \{\text{id2}, \text{id6}, \text{id7}\}, \\
 \text{assiduous} &= \{\text{id2}, \text{id3}, \text{id4}, \text{id5}\}.
 \end{aligned}$$

Now, we assign the learners with the collaborative label:

$$G1 = \{\text{id1}, \text{id2}\}, G2 = \{\text{id8}, \text{id6}\}, G3 = \{\text{id9}, \text{id7}\}.$$

After assignment, the labels classes are:

$$\begin{aligned}
 \text{engaged} &= \{\}, \\
 \text{collaborative} &= \{\}, \\
 \text{assiduous} &= \{\text{id3}, \text{id4}, \text{id5}\}.
 \end{aligned}$$

Finally, we assign the assiduous learners:

$$G1 = \{\text{id1}, \text{id2}, \text{id3}\}, G2 = \{\text{id8}, \text{id6}, \text{id4}\}, G3 = \{\text{id9}, \text{id7}, \text{id5}\}.$$

Now, the label classes are empty, and all the learners are assigned to groups.

Indeed, to assign *N* learners to *L* groups, a group must have (*N*/*L*) learners, if *N* is a multiple of *L*. If not, (*N* modulo *L*) groups must have (*N*/*L*) + 1 learners, whereas other groups must have (*N*/*L*) learners.

This approach was implemented as a module in the CLISON learning system and experimented to demonstrate its validity.

## 4. EXPERIMENTATION RESULTS AND DISCUSSION

We led an experimentation of the proposed system at the University of Guelma (Algeria). The participants to the experiment were the third-year computer science degree students, and its goal was to provide the students with a social learning network for the Semi-Structured Data course, which main purpose is to learn the XML language.

The aim of the experiment is to address two research questions:

RQ1. Does the grouping method based on a multi-label classification of students improve their learning?

RQ2. What is the students' perception of learning using CLISON?

The experiment process is described in the following section.

### 4.1 Participants

As mentioned above, the students participating to the experimentation (40 graduates of the 3rd degree) were involved in the Semi-Structured Data course. The students had total access to the learning system, the courses, and the offered tools. However, the class was divided into two groups of 20 students to fulfill the collaborative exams.

The first group is the experimental group, composed of 20 students, whose grouping into 5 groups of 4 members is carried out with the proposed method. While for the second group, the control group, the 20 students were given the freedom to form groups of 4 members, according to their affinities.

### 4.2 Methodology

To validate the proposed approach, we present the results of our experiment of creating heterogeneous groups of students, using a multi-label classification. Then, we compare these results with those obtained by the grouping by affinities, and, finally we analyze and discuss the data collected (see section 4.4).

The experimentation process followed the steps below (see Figure 5):

**Phase 1.** During this phase, the platform was presented to the students by showing its objectives and tools. Students were also asked to complete a pre-test to determine their basic academic performance.

**Phase 2.** In this phase, lessons and exercises are regularly posted on the platform, and the students had the possibility to contact the teacher, or other students, for clarification regarding the concepts of the course.

**Phase 3.** In this step, the students were associated, in groups of 4 members, to respond collaboratively to graded exercises.

**Phase 4.** To finish the course, an individual MCQ test (a post-test) was submitted to students, in order to establish their final academic performance.

**Phase 5.** Finally, a satisfaction questionnaire was proposed to the students, to evaluate their feelings, benefits and dislikes perceived during the experimentation process.

The groups formed in the third phase are heterogeneous and composed of 4 members. The choice of four members per group proceeds from two important reasons: on one side, a group composed of a pair of students is too small and,

therefore, lacks diversity, and in the other side, it is hard to ensure equality between members in larger groups. Whereas, an odd number of group members can lead to one of the students to remain apart during collaborative activities.

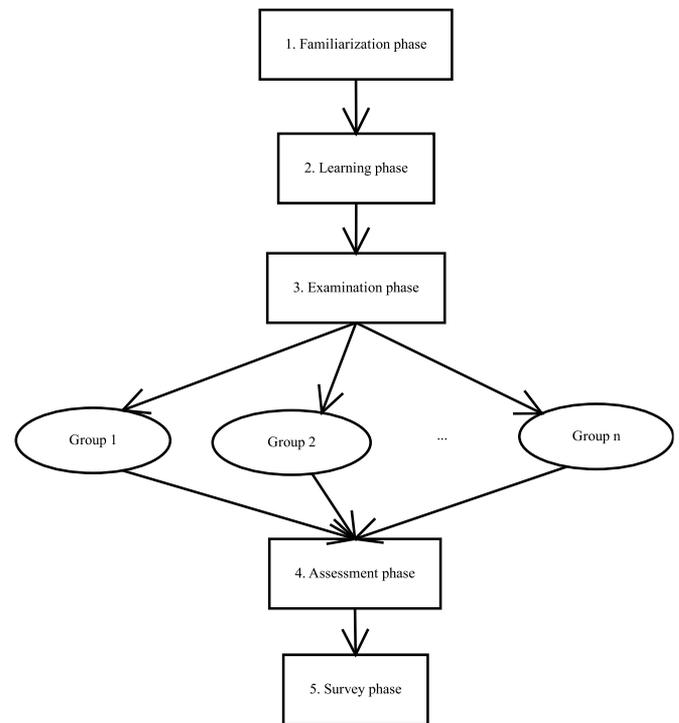


Figure 5. Experimentation process

Additionally, heterogeneous grouping is adopted to allow members with misconceptions to benefit from others more familiar with the topic, and furthermore to enable students with good communication skills to encourage others who are less sociable.

### 4.3 Comparative study of the proposed method

In the third phase of the experimentation process, where students are associated to solve exercises, the grouping was carried out with two methods: grouping using multi-label classification and affinity grouping. Therefore, to validate the efficiency of the proposed grouping method, we divided the third degree class into two subclasses:

- an experimental group: where students are grouped using the proposed multi-label classification, and
- a control group: where students join together according to their affinities.

All groups performed the collaborative activities of the third phase of the methodology, after which the entire class is requested to fulfill the post-test.

Finally, to validate the effectiveness of our grouping approach, the academic performances of both groups are compared.

The results of the comparative analysis are presented and discussed in the following.

### 4.4 Results and discussion

As stated in the precedent section, 40 students took part in the experiment, divided into two groups of 20 students each: an experimental group and a control group.

#### 4.4.1 Pre-test results

In this first step, in the experiment process, the students were asked to fulfill a questionnaire regarding the course. This pre-test questionnaire consisted of 20 multiple-choice questions, each one introducing five possible answers, of which one or more may be correct. The maximum score for the questionnaire is 100 points.

Regarding the participation of the students to the pre-test, in the experimental group, six students, out of 20, didn't answer the questionnaire, whereas, for the control group, 7 students did not respond. Therefore, the participation rate in the pre-test is: 70% for the experimental group, and 65% for the control group. So, 27 students fulfilled the test, getting grades between 0 and 100.

In order to compare the pre-test grades obtained by the students of each group, a t-test (also called Student's test) is used. It is a widely used statistical hypothesis test, whose aim is to compare the means of two groups to determine if there is a significant difference between them.

The test relies on a set of assumptions, which are:

- **Normality of data distribution:** We used the Shapiro-Wilk Test to check whether the data follow a normal distribution or not. For the experimental group, the Shapiro-Wilk test did not show a significant departure from normality ( $W(14) = 0.92, p = 0.218$ ). In the same way, the Shapiro-Wilk test did not show a significant departure from normality ( $W(13) = 0.94, p = 0.485$ ) for the control group.
- **Homogeneity of variance:** Levene's test is a statistic test used to check if two sets of data have equal variances. The result of the test is not significant at  $p < 0.05$  (f-ratio = 1.88,  $p = 0.182$ ). Therefore, the requirement of homogeneity is met.

Table 2 presents a summary of the T-test for the groups that took part in the experiment.

**Table 2.** T-Test: Pre-test results (experimental group students vs. control group students)

Parameters	Values
H0	Both groups are equal regarding the students' grades in the pre-test
H1	One group is superior than the other
Mean (experimental group)	71.29
Mean (control group)	72.46
Standard deviation (experimental group)	19.45
Standard deviation (control group)	55.77
Degree of freedom	25
t	-0.50267
P	0.619598

The T-test shows that there is no significant effect for grades,  $t(25) = 0.50267, p = 0.6196$  ( $p > 0.05$ ). We can conclude that the means of the two groups have no significant differences, and therefore the groups are equivalent in term of level of knowledge regarding the course.

#### 4.4.2 Post-test results

At the end of the experiment, and to finish the course, the students were asked to complete a post-test questionnaire.

In this stage, the participation rate was 85% for the

experimental group, and the same rate for the control group. In fact, there were 3 students in each group who did not participate in the test.

As for the pre-test, the post-test is a multiple-choice questionnaire composed of 20 questions. For this test, 34 students fulfilled the questionnaire, 17 from each group.

Before applying the T-test, we check the test assumptions:

- **Normality of data distribution:** For the experimental group, the Shapiro-Wilk test did not show a significant departure from normality ( $W(17) = 0.97, p = 0.862$ ). In the same way, the Shapiro-Wilk test did not show a significant departure from normality ( $W(17) = 0.9, p = 0.060$ ) for the control group.
- **Homogeneity of variance:** The result of the test is not significant at  $p < 0.05$  (f-ratio = 1.68,  $p = 0.203$ ). Therefore, the requirement of homogeneity is met.

Since the prerequisites of the t-test are met, we will use it to compare the means of the results obtained for the two groups involved in the experiment. A summary of the study is shown in the Table 3.

**Table 3.** T-Test: Post-test results (experimental group students vs. control group students)

Parameters	Values
H0	Both groups are equal regarding the students' grades in the post-test
H1	One group is superior than the other
Mean (experimental group)	78.88
Mean (control group)	73.06
Standard deviation (experimental group)	5.53
Standard deviation (control group)	8.99
Degree of freedom	23
t	2.27346
P	0.029849

The participants in the experimental group ( $M=78.88, SD=5.53$ ) compared to the participants in the control group ( $M=73.06, SD=8.99$ ) demonstrated significantly better grades,  $t(23)=2.27346, p=0.0298$  ( $p < 0.05$ ). This leads us to assume that our grouping method, based on a multi-label classification, is more efficient compared to learners self-grouping, and therefore brings great benefits to their members.

Despite this positive result, the proposed approach still needs to be tested on a larger sample, as the current experiment involved only 40 students, which does not allow for generalization. Another facet to consider is the management of dropouts during the learning process and its influence on the labeling of learners, and therefore their assignment to a group. Finally, it would be interesting to examine the possible relationship between the labels, particularly by considering what is the influence of the labels on group effectiveness, and also whether one label is more significant than the others.

In the next section, we present the results of a survey on the students' perception about the learning platform used for the experiment, and also toward the proposed collaborative activities.

#### 4.5 Satisfaction survey

After the end of the experimentation, we requested the 40

students to acknowledge anonymously an online satisfaction survey. The choice of anonymity was guided by the opportunity given to students to express their opinions freely.

The aim of the survey is to appreciate the students' perception toward the platform and the tools offered, and also to identify what influence the group had on their progress in the course.

The survey was completed only by 18 students, which represent 45% of the experimentation set.

The inquiry focused on the system usability and the group activities satisfaction. It took place in the form of four questions presented to the students; each question rated using a 3 points Likert scale. The questions are:

Q1. Do you agree or disagree with the statement: "the platform is easy to use and the courses are accessible without any difficulties"?

Q2. Do you agree or disagree with the statement: "the tools provided on the platform are helpful for improving my knowledge of the concepts learned"?

Q3. Do you agree or disagree with the statement: "the group activities increased my motivation to find solutions to the proposed exercises"?

Q4. Do you agree or disagree with the statement: "sharing knowledge between students during group activities helped me better assimilate the course concepts"?

Despite the small number of answers obtained in the survey, we will discuss the results (Figures 6, 7, 8 and 9), while being aware that they cannot be generalized.

The results of the inquiry are as follows:

- Generally speaking, there is a positive attitude towards the proposed method, as:
  - for Q1, 72.22% of the surveyed students agree with the proposition that the platform was not difficult to use, while 27.78% have a mixed opinion on the usability of the system (Figure 6).
  - for Q2, the surveyed students, at a rate of 72.22%, don't oppose the proposition that the tools provided were helpful (Figure 7). We also notice that 11.11% of the students find the proposed tools useless.
  - for Q3, 77.78% of the students agree with the hypothesis that group activities increased their motivation, while only one student disagree with the statement (Figure 8).
  - for Q4, 72.22% of the students consider that communication during group activities helped them better assimilate the provided courses (Figure 9). On this issue too, only one student disagreed with the statement.
- It is noticed that none of the students expressed their disagreement with the usability of the system (Q1), whereas two students, from the experimental group, found the proposed tools useless (Q2).
- Concerning Q3, the majority of the students agreed with the statement that performing collaborative activities motivates them, while only one student disagreed with this statement.
- Finally, for the fourth question, the same proportion of the students, as the first three questions, agreed that the group they belonged to helped them understand the courses, while, again, only one student disagreed.

Even though the number of replies for the inquiry is small, which does not allow us to generalize the findings, we can still notice that group activities were considered useful for

improving students' knowledge, and its impact on the student motivation is high. However, these results should be confirmed by a study involving a larger set of students.

Q1: The platform is easy to use and the courses are accessible without difficulty.

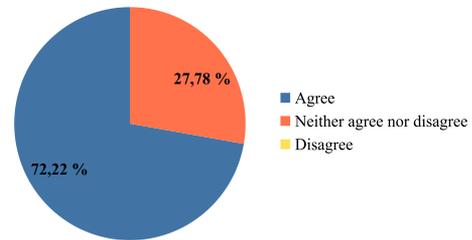


Figure 6. Students' perception of platform usability

Q2: The tools provided on the platform are helpful in improving my knowledge of the concepts inculcated.

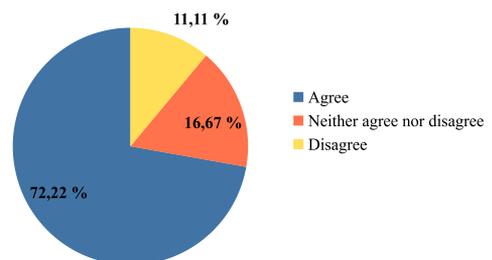


Figure 7. Students' perception of the usefulness of the proposed tools

Q3: The group activities increased my motivation to find solutions to the proposed exercises.

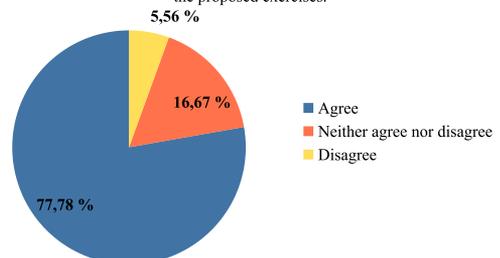


Figure 8. Students' perception of the contribution of the formed groups to motivation

Q4: Sharing knowledge between students during group activities helped me better assimilate the course concepts.

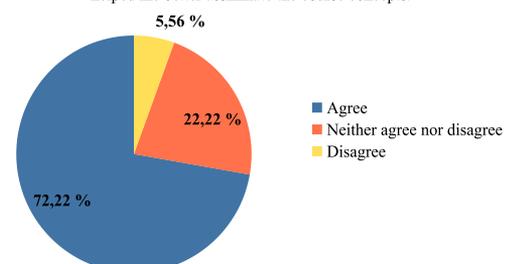


Figure 9. Students' perception of the collaboration benefits on learning

## 5. CONCLUSION AND FUTURE WORKS

In this paper, we exposed an efficient learning groups formation approach based on the classification of learners into multiple labels, using a multi-label ML-KNN classification algorithm. Groups are formed by selecting learners from different labels (collaborative, assiduous and engaged), after categorizing them using the ML-KNN multi-label classification algorithm. These labels allow us to select the learners who can be grouped together in order to form effective groups for collaborative learning. The groups formed are heterogeneous to ensure a diversity of profiles in each of the groups, promoting complementarities in the same group, regarding knowledge and sociability. The classification of learners according to the defined labels is based on five attributes describing the main characteristics of the learners relating to collaborative learning.

The efficiency of the proposed system was evaluated according to two important aspects: (i) the effectiveness of the groups formed using our approach, validated by an experiment where these groups were compared to self-formed groups according to learner's affinities, and (ii) the approval of the system by the students evaluated by the analysis of a satisfaction inquiry.

After analyzing the results of the experiment, we were able to assume that the grades obtained by the learners belonging to the groups formed with our approach were better than those of the self-formed groups. Moreover, in the satisfaction survey, 72.22% of the participants perceived that they better assimilate the courses provided by the system with the help of groups' activities. However, this result must be tempered by the small sample size. Generalizing the result will require experimentation on a larger scale.

Our main objective, for the continuation of this work, is to compare our approach with learner's grouping through a genetic algorithm, to better evaluate its effectiveness. In parallel, the proposed SNL system (CLISON) needs to be tested on larger learners' samples to validate its usability in an intensive use context.

## REFERENCES

- [1] Al-Rahmi, W.M., Yahaya, N., Alturki, U., Al-robai, A., Aldraiweesh, A.A., Omar Alsayed, A., Kamin, Y.B. (2022). Social media-based collaborative learning: The effect on learning success with the moderating role of cyberstalking and cyberbullying. *Interactive Learning Environments*, 30(8): 1434-1447. <https://doi.org/10.1080/10494820.2020.1728342>
- [2] Hmelo-Silver, C.E., Jeong, H. (2021). Benefits and challenges of interdisciplinarity in CSCL research: A view from the literature. *Frontiers in Psychology*, 11: 579986. <https://doi.org/10.3389/fpsyg.2020.579986>
- [3] Liang, C., Toyokawa, Y., Ogata, H. (2025). Optimizing group formation with a mixed genetic algorithm: An empirical study in active reading using marker data. *International Journal of Computer-Supported Collaborative Learning*, 20: 519-548. <https://doi.org/10.1007/s11412-025-09452-9>
- [4] Moreno, J., Sánchez, J.D., Pineda, A.F. (2021). A hybrid approach for composing groups in collaborative learning contexts. *Heliyon*, 7(6): e07249. <https://doi.org/10.1016/j.heliyon.2021.e07249>
- [5] Maqtary, N., Mohsen, A., Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, 24: 169-190. <https://doi.org/10.1007/s10758-017-9332-1>
- [6] Krouska, A., Virvou, M. (2019). An enhanced genetic algorithm for heterogeneous group formation based on multi-characteristics in social-networking-based learning. *IEEE Transactions on Learning Technologies*, 13(3): 465-476. <https://doi.org/10.1109/TLT.2019.2927914>
- [7] Revelo-Sánchez, O., Ordóñez, C.A.C., Duque, M.Á.R. (2021). Group formation in collaborative learning contexts based on personality traits: An empirical study in initial programming courses. *Interaction Design & Architecture(s) Journal*, 49: 29-45. <https://doi.org/10.55612/s-5002-049-002>
- [8] Shanmuganeethi, V., Muthuramalingam, S., Uma, K. (2020). Intelligent dynamic grouping for collaborative activities in learning management system. *Journal of Engineering Education Transformations*, 34(2): 108-116. <https://doi.org/10.16920/jeet/2020/v34i2/151590>
- [9] Maina, E.M., Oboko, R.O., Waiganjo, P.W. (2017). Using machine learning techniques to support group formation in an online collaborative learning environment. *International Journal of Intelligent Systems & Applications*, 9(3): 26-33. <https://doi.org/10.5815/ijisa.2017.03.04>
- [10] Sankaranarayanan, S., Dashti, C., Bogart, C., Wang, X., Sakr, M., Rosé, C.P. (2018). When optimal team formation is a choice-self-selection versus intelligent team formation strategies in a large online project-based course. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018. Lecture Notes in Computer Science*, 10947: 518-531. [https://doi.org/10.1007/978-3-319-93843-1\\_38](https://doi.org/10.1007/978-3-319-93843-1_38)
- [11] Herrera, F., Charte, F., Rivera, A.J., del Jesus, M.J. (2016). Multilabel classification. In *Multilabel Classification: Problem Analysis, Metrics and Techniques*, pp. 17-31. [https://doi.org/10.1007/978-3-319-41111-8\\_2](https://doi.org/10.1007/978-3-319-41111-8_2)
- [12] Goštautaitė, D., Sakalauskas, L. (2022). Multi-label classification and explanation methods for students' learning style prediction and interpretation. *Applied Sciences*, 12(11): 5396. <https://doi.org/10.3390/app12115396>
- [13] Surolia, A., Mehta, S., Kumaraguru, P. (2025). Deep learning and transfer learning to understand emotions: A PoliEMO dataset and multi-label classification in Indian elections. *International Journal of Data Science and Analytics*, 20: 4193-4207. <https://doi.org/10.1007/s41060-025-00738-7>
- [14] Sun, G., Cheng, Y., Dong, F., Wang, L., Zhao, D., Zhang, Z., Tong, X. (2024). Multi-label text classification model integrating label attention and historical attention. *Knowledge-Based Systems*, 296: 111878. <https://doi.org/10.1016/j.knosys.2024.111878>
- [15] Zhang, L., Ning, Y., Zhou, S. (2026). Automated multi-label classification of risk clauses in construction contracts using GPT-driven data augmentation. *Automation in Construction*, 181: 106599. <https://doi.org/10.1016/j.autcon.2025.106599>
- [16] Wen, W., Zhang, H., Wang, Z., Gao, X., Wu, P., Lin, J., Zeng, N. (2024). Enhanced multi-label cardiology

- diagnosis with channel-wise recurrent fusion. *Computers in Biology and Medicine*, 171: 108210. <https://doi.org/10.1016/j.compbiomed.2024.108210>
- [17] Joe, H., Kim, H.G. (2024). Multi-label classification with xgboost for metabolic pathway prediction. *BMC bioinformatics*, 25: 52. <https://doi.org/10.1186/s12859-024-05666-0>
- [18] Swaminathan, B., Jagadeesh, M., Vairavasundaram, S. (2024). Multi-label classification for acoustic bird species detection using transfer learning approach. *Ecological Informatics*, 80: 102471. <https://doi.org/10.1016/j.ecoinf.2024.102471>
- [19] Shiri, F.M., Perumal, T., Mustapha, N., Mohamed, R., Ahmadon, M.A.B., Yamaguchi, S. (2024). Recognition of student engagement and affective states using convnextlarge and ensemble GRU in E-learning. In 2024 12th International Conference on Information and Education Technology (ICIET), Yamaguchi, Japan, pp. 30-34. <https://doi.org/10.1109/ICIET60671.2024.10542707>
- [20] Makhoukhi, H., Roubi, S. (2024). Multi-label emotion classification of online learners' reviews using machine learning. In Proceedings of the 2024 5th International Conference on Education Development and Studies, Cambridge United Kingdom, pp. 59-64. <https://doi.org/10.1145/3669947.3669963>
- [21] Lu, Y., Ma, Y. (2024). Predicting e-learning resource quality based on multi-modal data. In 2024 13th International Conference on Educational and Information Technology (ICEIT), Chengdu, China, pp. 307-312. <https://doi.org/10.1109/ICEIT61397.2024.10540924>
- [22] Sarker, G.C., Hasan, M.M., Hoque, M.R., Ahmed, M. (2025). Predicting learning styles with AI: Toward adaptive and personalized education. In CONF-IRM 2025 Proceedings, 14.
- [23] Sukstrienwong, A. (2023). Anova as fitness function for genetic algorithm in group composition. *TEM Journal*, 12(1): 396-405. <https://doi.org/10.18421/TEM121-49>
- [24] Haq, I.U., Anwar, A., Rehman, I.U., Asif, W., Sobnath, D., Sherazi, H.H.R., Nasralla, M.M. (2021). Dynamic group formation with intelligent tutor collaborative learning: A novel approach for next generation collaboration. *IEEE Access*, 9: 143406-143422. <https://doi.org/10.1109/ACCESS.2021.3120557>
- [25] Putro, B.L., Rosmansyah, Y., Agustine, S.S. (2020). Intelligent agent to form heterogeneous group based on personality traits with genetic algorithm. In 2020 International Conference on Information Technology Systems and Innovation (ICITSI), Bandung, Indonesia, pp. 294-299. <https://doi.org/10.1109/ICITSI50517.2020.9264906>
- [26] Bourkougou, O., Bachari, E.E., Boustani, A.E. (2019). Building effective collaborative groups in e-learning environment. In International Conference on Advanced Intelligent Systems for Sustainable Development, pp. 107-117. [https://doi.org/10.1007/978-3-030-36653-7\\_11](https://doi.org/10.1007/978-3-030-36653-7_11)
- [27] Muuro, M.E., Oboko, R.O., Wagacha, W.P. (2016). Evaluation of intelligent grouping based on learners' collaboration competence level in online collaborative learning environment. *International Review of Research in Open and Distributed Learning*, 17(2): 40-64. <https://doi.org/10.19173/irrodl.v17i2.2066>
- [28] Liang, C., Majumdar, R., Ogata, H. (2021). Learning log-based automatic group formation: System design and classroom implementation study. *Research and Practice in Technology Enhanced Learning*, 16: 14. <https://doi.org/10.1186/s41039-021-00156-w>
- [29] Wang, Y., Wang, Q. (2022). A student grouping method for massive online collaborative learning. *International Journal of Emerging Technologies in Learning (IJET)*, 17(3): 18-33. <https://doi.org/10.3991/ijet.v17i03.29429>
- [30] Yang, J., Hu, S., Wang, Q., Fong, S. (2021). Discriminable multi-label attribute selection for pre-course student performance prediction. *Entropy*, 23(10): 1252. <https://doi.org/10.3390/e23101252>
- [31] Li, J., Li, P., Zou, Y., Hu, X. (2021). Multi-label learning with missing features. In 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, pp. 1-8. <https://doi.org/10.1109/IJCNN52387.2021.9533967>
- [32] Zhang, M.L., Zhou, Z.H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, 40(7): 2038-2048. <https://doi.org/10.1016/j.patcog.2006.12.019>
- [33] Jalinus, N., Sukardi, S., Wulansari, R.E., Heong, Y.M., Kiong, T.T. (2023). Teaching activities for supporting students' 4Cs skills development in vocational education. *Journal of Engineering Researcher and Lecturer*, 2(2): 70-79. <https://doi.org/10.58712/jerel.v2i2.95>