

Intelligent dynamic grouping for collaborative activities in Learning Management System

Shanmuganeethi V, Muthuramalingam S, Uma K V

¹Department of Computer Science and Engineering,
National Institute of Technical Teachers Training and Research (NITTTR), Taramani, Chennai.

^{2,3} Department of Information Technology, Thiagarajar College of Engineering, Madurai.

¹shanneethi@nitttr.ac.in

²smrit@tce.edu

³kvuit@tce.edu

Abstract : In a sustainable education environment, Learning is fixed everyone and learning time may vary with individual capability. To provide additional learning space, Learning Management System (LMS) is a perfect solution. All kind of learners can make use of the additional learning space and attain the objectives of the content. Learning Management System is an integral component of the e-Learning design and development, and to manage the learning process. LMS would provide pre-defined templates for development as well as to deploy e-content, integrate multimedia content such as audio, video, images, graphics and animation, organise assessments, evaluation procedures and analyze the learner style by including the learning analytics. Most of the Learning Management Systems offer convenient tools that allow to create custom courses and thus implement educational, teaching, training and management processes in an organization. Today learners are more towards active and team-oriented learning. LMS would improve active and collaborative learning by providing varieties of activities and group projects. While giving the group assignments and projects to the learners, an effective and dynamic grouping mechanism need to be integrated. Based on the group size, the group should consist of mixed kind of learners to learn from group

activity and to lead a group activity. The formation of the group must be dynamic for every activity. The groups should be dynamic in nature and every learner of the course must have active learning with each other in the same course. To prepare an intelligent grouping mechanism, a framework was developed and deployed in the MOODLE LMS. This intelligent dynamic grouping mechanism improve the overall learning activities in the LMS environment.

Keywords: Active learning, Collaborative learning, Learning Management System, Learning Analytics, Learning space

1. Introduction

Teaching and Learning (TL) paradigm changing its shape to facilitate active and collaborative learning methods. TL in e-Learning environment has been improved drastically to meet common features of face-face learning. In the face to face traditional learning environment, most of the experience of the teacher in the teaching context and the experience of the learner in the learning context are not documented. The evolution of the e-learning system never leaves an experience without documentation. This documentation details are very much useful to understand the learners. This e-Learning platform leads to personalized learning practices where the teacher could address the individual learner. To address the individual learning preferences and engage with different learning activities, the Learning Management System (LMS) is a perfect solution. It brings the teacher and learner under a single umbrella.

Shanmuganeethi V

Department of Computer Science and Engineering,
National Institute of Technical Teachers Training and Research
(NITTTR), Taramani, Chennai.
shanneethi@nitttr.ac.in

It is a common e-learning application widely used in all kind of education and training sectors. The LMS application minimize the risk of losing data and provide the simplified way of creating courses and activities. In the LMS environment, the course materials are available for unlimited access and it provide the ability to access the learner performance. Most of the learning management systems supporting the learning analytics tools which provides the meta cognition of the learners. These analytics gives the at-risk information to the course developer. In the Web2.0 era, LMS integrated with social learning applications such as Facebook, Twitter pages, LinkedIn groups and etc. This integration attract new learners in the e-learning environment.

Experiences of the Learners in Learning Management System

Diversity of student experiences are significant in Learning Management System. The learners are started their learning with more number activities and resources. The learners are rate the consistency of teacher response and communication as important component in LMS. This influence their online learning experience. Specifically, students value the prompt and informative feedback on their work and clarity of faculty expectations of their work. Mason (2001) has proclaimed that “time is the new distance”, aslack of time, rather than long distance, has become one of the primary reasons that students withdraw from courses. Learning management system is considered as most important supporting teaching platform for many Higher Education (HE) institutions.

All kind of LMS are supporting lectures, activities, practices and project work. The lectures are in the form of PDF documents, e-books and etc. The LMS platform is also a right place to address the individual learner and collect their individual experience, performance and feedback. This enables to capture the students' responses while enabling anonymity, since no user credentials are stored. So, the students' participation is satisfaction and their responses often very realistic. In general, the learners are responding based on the clarity and understandability of materials and activities, cooperation with learners and to what extent the teachers help and guide learners in their course. In the same manner, the learners are assessed on for completing the exercises and projects, cooperation with students and to what extent the learners helps and guide other learners in their

course. The learner experiences lead to do analytics about their progress and provide the space for improve the course in the LMS.

Literature Review

Analytics and big data have a significant role to play in the future of higher education. In education the value of analytics and big data can be found in: (1) their role in guiding reform activities in higher education; and (2) how they can assist educators in improving teaching and learning. Siemens et.al proposed that the Learning analytics can penetrate the fog of uncertainty around how to allocate resources, develop competitive advantages, and most important, improve the quality and value of the learning experience. A genre classification system, called GCS was proposed to facilitate the automatic coding process. It examines the coding coherence between GCS and experts' judgment and discusses how to adjust the parameters of the GCS to improve the coherence. The empirical evaluation of the classified genres from a repository of postings in an online course on earth science in a senior high school shows that GCS can effectively facilitate the coding process. These results imply that GCS based on the cascade model can perform as an automatic posting coding system. Saulius et.al analyzed the learners behavior using data mining method. BlackBoard vista Distance learning platform was used to observe about 528 students learning behavior in 15 distance learning courses. About three different learners style of students using virtual learning environment (VLE) were identified using Clustering method. Normally VLE was used to uploading learning materials and also can be used as discussion forum. VLE was used to track the student's activities. In order to consolidate and analyze the students data ,a data tracking tool called as PowerSight Kit (VistaPowerSightKit 2008). Marcos Wander Rodrigues et.al collected about 525 articles which are related to data mining and education, in which about 453 articles were discarded such that they did not emphasize the improvement of the teaching and learning process. From the analysis of the bibliographic material, it was possible to observe the different trends of research given below. Activities that reduce the gap between the teacher and students, recommending the different teaching mode that are effective, identifying the similar kind of learners and also the unusual behavior of student, improving personalized learning of the students from the data obtained through social networks. From the different articles it was concluded that the more research going on to identify the learning process of

students, identify the factors that influence the students learning and improving the students learning skills and methods for identifying and overcoming their deficiencies.

Leah P. Macfadyenet.al done an exploratory study on students variables collected through UBC's BB Vista™ production server .The data was extracted and analyzed for a particular online course. Regression model was built by using the students tracking variables. The model identified the students with a risk of failure with 70.3% accuracy, at risk with 80.9% accuracy. The past studies focused on the student's time in online as a measure of engaging the students. But the current study make this attribute as weak that total time spent online did not act as a significant predictor of student grade. The important predictor variable observed in this study was mail tool. The tool was used for communication between the instructor and the individual students. The data related to mail messages was identified as a measure of attention of the students in their learning. The study suggests that a separate predictive model must be designed for each course instead of having the same predictive model for the entire course. Because, the learning behavior of students will vary from one subject to other. Hence this suggestion was made. Data visualization resources will make the instructor to identify the students learning behavior while using learning management system.HananKhalil et.al analyzed the attrition rates of the distance education students on MOOC initiatives. Investigation was done on to analyze why students dropped or failed in MOOC initiatives. By analyzing ,strategies was made to increase the retention rate of the students in the MOOCs and also to increase the satisfactory index of the students. Many factors such as Lack of time, less motivation among learners, isolation and lack of interaction among students, less skills and hidden costs contribute to high dropouts in MOOCs. The solution suggested to increase high retention rate in MOOC was to accommodation of students by using different time tables by making the students participate in peer grading of assignment with synchronization, promoting the completion of activities by students by providing certificates, enhancing the students-students interaction and students –instructors interaction with help of teaching assistants. Wolfgang Grellerand, Hendrik Drachsler discussed a framework that support learning analytics in education. It also acts as learner guidance in assuring the quality in the curriculum, improving the effectiveness of teaching. Learning analytics considers the data collected through Learning

Management Systems and other systems. The proposed framework consider six dimensions which further divided into several instantiations. The framework determines the connections between these six dimensions. Aderibigbe Israel Adekitan and OdunayoSalau carried out predictive analysis to determine the extent to which the fifth year and final Cumulative Grade Point Average (CGPA) of engineering students in a Nigerian University can be determined using the program of study, the year of entry and the Grade Point Average (GPA) for the first three years of study as inputs into a Konstanz Information Miner (KNIME) based data mining model. A predictive model was constructed using the KNIME analytics platform .The model will predict the grade of the final year students. The data used was the GPA obtained by the students in the previous years. The model was tested for about 1,841 engineering students across seven departments. The features were analyzed using the forward feature selection Meta node on the KNIME server. It was observed that Logistic Regression predictor provided the highest accuracy of 89.15%. John Watson and Butch Gemin discussed the growth of online programs .It focused on at-risk students or credit recovery that has redefined how educational technology can be used to address the needs of all students, from advanced students in search of Advanced Placement or dual-credit courses, to at-risk students trying to find the right instructional mix to fit their learning styles.As online programs increasingly focus on at-risk students and credit recovery, educators are finding that reaching these students presents a specific set of issues. Students will earn a credit on passing an exam. If the students failed in a subject, credit recovery can be done by attempting the exam as arrear. The characteristics of risk students who apply for credit recovery need to be analyzed to determine their learning behavior. And provide them personalized learning such that they cope up with the subjects. Ryan S.J.D. Baker and Kalina Yecéf reviewed the history and current trends in the field of Educational Data Mining (EDM). The methodological profiles of research in the early years of EDM were compared to in 2008 and 2009 and, trends and shifts in the research conducted by this community were discussed. Also, the emergence of work using existing models to make scientific discoveries (“discovery with models”), and the reduction in the frequency of relationship mining within the EDM community were discussed. Historically, relationship mining methods of various types have been the most prominent category in EDM research. In Romero & Ventura's survey of EDM

research from 1995 to 2005, 60 papers were reported that utilized EDM methods to answer research questions of applied interest (according to a post-hoc analysis conducted for the current article), 26 of those papers (43%) involved relationship mining methods and 17 more papers (28%) involved prediction methods of various types. Srećko Natek and Moti Zwilling proposed the use of REP tree model in weka tool for identifying the key influencer for students final grade attribute. The following attributes related to students such as study year , student number, gender, student year of birth , employment , status, e.g. sport etc. (no, yes), registration , type of study , exam condition , activities points , exam points , final points and predictable final grade were considered. Segun I. Popoola et.al illustrated the differences and trends in the GPA data of engineering undergraduates of various disciplines. The box plot diagram depict the proportions of engineering students that graduated with First Academic Affairs. Descriptive statistics and frequency distributions of the academic performance data are presented in tables and graphs for easy data interpretations. In addition, one-way Analysis of Variance (ANOVA) and multiple comparison post-hoc tests are performed to determine whether the variations in the academic performances are significant across the seven engineering programs.

Proposed work

Intelligent model for dynamic grouping of learners

The proposed model is collecting the data from different activities performed by the learner, use the resources by the learner in the LMS, access the resources from open educational resources during the period of learning and crawling data. In an online learning different style of learners such as auditory, visual, read and kinesthetic types and different kind of learners such as fast learners, average learners and slow learners. All kind of learners need to be addressed in the online platform. When a group activity is assigned to the learners, how the groups are to be prepared so that collaborative learning is more successful. This model, bring the learners in common platform and prepare a group with a proper blending of learners. At the end of the learning of a course, a learner should have an activity with minimum of 80% of the learners of a same course. The overall system design is shown in figure 1.

Step 1: Tracking of interaction data

In online learning platform, collecting the learner

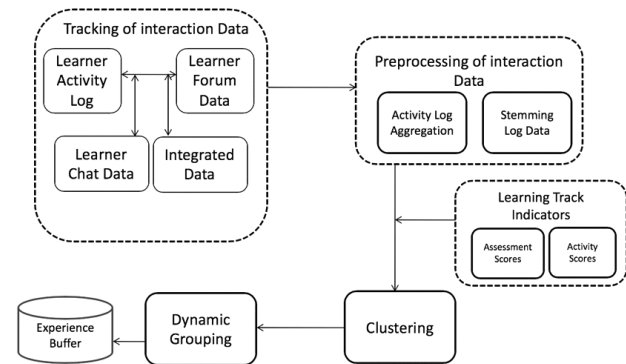


Fig. 1: Intelligent model for dynamic grouping of learners

The proposed Intelligent model is split into five phases. They are

1. Tracking of Interaction data
2. Preprocessing of Interaction Data
3. Clustering

log data such as discussion forum, activities, interactions and downloads which reveals lot of academic and technological information to improve course on learning management system and to attract the learners towards active and e-learning platforms.

Activity log data analytics

Activity Log data Analysis of the learner will identify the present state of learning and their activities. The log data can be maintained as Activity Web Log which consists of the data about the activities involved by the learner. In addition, it generates complete activity report based on the data collected from the learner. Value Point Indicators (VPI) are defined to identify the key components of log to do the analysis the data. The VPI information of the log data are to be investigated to find out the activity details of the learner. The common type of indicators are activities such as discussion forum entry, any kind of webinar through tools like bigbluebutton, prctest with tools such as hotpotatoes, most viewed resources in the course, and etc., When the VPI indicators are applied to a learning management system, it provides very important and useful datasets which are input for clustering the learners in the learning management system.

Integrated data analytics

The learner may do different courses in the same learning management system. If a learner registered for more than one course, it is necessary to find out the course progress comparison with other courses. The

VPI are help to discover the compared details, technology adoption in different courses, activities configured and its comparison among the courses registered by a particular learner. These details provide the reasons for a learner doing good learning in a course and average learning in another course. Integrated data analytics is also used to establish the learning resources such as e-content, video materials, e-books and assessments are accessed and used by leaners throughout the courses. It is an important analysis of the data analytics, as the key challenge behind developing an efficient learning management system. Also, it may provide useful information on presentation logics such as user friendliness different facets of the LMS are from the perspective of various user groups.

Crawling log Analytics

Crawling log analytics is more complex when compared other kind of traditional web analysis methods, as it cannot be interpreted without considering beyond learning management system usage and understanding. When a learner undergoing a particular concept, whether the resources and activities are enough to understand the concept or the learner is crawling other web sites, open educational resources or any other e-learning sites. If a user crawling from the course LMS to any other link, the link needs to be analysed whether the site is providing the same content as in the course LMS or the site is consisting higher grade content than the content which is undergoing by the learner. The higher-grade content can be used to do online exercises programme which grade the programme for the learners. These analytics details are indicators for categorize the course and cluster the learners.

Step 2:Preprocessing of log data

Data preparation, known also as data pre-processing, is focused mainly on two issues: firstly, the data must be organized into a proper form for data mining algorithms, and, secondly, the data sets used must lead to the best performance and quality for the models obtained by data mining operations [5]. The data collected from various activity log, integrated log and crawling log are considered as input for pre-processing module. The data collected from the logs are incomplete, noisy and not appropriate for mining with respect to identifying the pattern of usage or to classify the learner. So, in the pre-processing stage data removal, data cleaning and data filtering are done

because the source of logs data causes are combined with inappropriate information.

Algorithm

1. Read each word,
 - a. Remove the less useful parts of text through stopword removal, dealing with capitalization and characters.
 - b. Remove annotations from the log data
 - c. Change to small case
 - d. Remove stop-words compared with stop-word list
 - e. Tokenization through lexical analysis
 - f. Translation of terms in the scheme or linguistic reductions through Stemming operations, Lemmization and other forms of standardization.
2. Prepare the set of words to map the VPI
3. Statistically probing the dataset for feature analysis.

Step 3: Clustering of the learner

The dataset received from the pre-processing module is the input of the clustering module. The clustering module performs the clustering of the learners based on the pre-processed output. Clustering is an unsupervised machine learning method to divide a set of objects into clusters, so that objects in the same cluster are similar to each other, and/or objects in different clusters are dissimilar. This proposed model working with k-means clustering algorithm. As per the k-means, cluster centroid is to be defined. The cluster centroid is random and the clusters are identified by using Euclidean distance.

In Cartesian coordinates, if $p=(p_1,p_2..p_n)$ and $q=(q_1,q_2..q_n)$ are the two different points in Euclidean n-space, the distance from p to q is defined as given in equation 1.

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} \quad 1$$

In order to identify K clusters, the k points are initially considered as cluster centers. K-means finds out the best separated clusters by trying to minimize a distortion function. The distortion function is given as in equation 2.

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (||X_i - V_j||)^2 \quad 2$$

where, $||x_i - v_j||$ is the Euclidean distance between x_i and v_j . c_i is the number of data points in i th cluster and c is the number of cluster centers. The distortion function is a non-convex function and thus implies that K-means is susceptible to getting stuck in local optima. This means that when K-means is run with random cluster centroids; we might not reach the best solution possible. To reduce the chances of getting a sub-optimal clustering we restarted K-means 200 times with random initialization. For each of the cluster identified, a linear regression model is trained, such that if there are K clusters, then K linear regression model is built. All of these K cluster models taken together to make predictions on the entire test set. These K cluster models together can be thought to form a more complex Prediction model. The prediction model PMK, with the subscript K identifying the number of cluster models. The value of K can be varied from 1 to K to obtain K prediction models. For example: if $K = 1, 2$ and 3 , there would be three prediction models - PM1 having a single cluster model ($K=1$), PM2 having two different cluster models ($K=2$) and PM3, that is the prediction model with three different cluster models ($K=3$). It is noteworthy that the cluster models in different prediction models would be different. If K prediction models are constructed from the data, there would be a set of K different predictions on the test data. These predictions are compared with those obtained on PM1, i.e. a linear regression model fit over the entire data-set to see if there is an improvement in prediction accuracy. An improvement would indicate a strong result that dynamic assessment indeed gives a much better assessment of student learning.

Results and Discussion

The Information technology student undergoes the course Computer networks in third semester. There are about 141 students in this batch. The course is started in Moodle to post the learning materials, for conducting quiz, submitting assignment, discussion forum for discussing on scenarios.

Activity Log Analytics

Initially, the students' present state of learning and their activities need to be identified. The participation of students in the activities like Preconcept Quiz, Introductory Video and viewing the text material are considered. The participation is analysed through number metrics and ratio metrics.

Number Metrics

The total number of students in the class is 141. The various activities and its view are given in table 1.

Table 1: Various initial activities conducted for the course and its views

Activities	Number Of Views	Number of Students Participated
Preconcept Test	106	94
Discussion Forum	98	45
Introductory Text Material	120	120

Ratio Metrics

The ratio metrics of preconcept test is given in Table 2.

Table 2: Preconcept Test

	Number of Questions	Average Mark Obtained
Preconcept Test	5	8.36

The average score obtained in each of the question is given in figure 2.

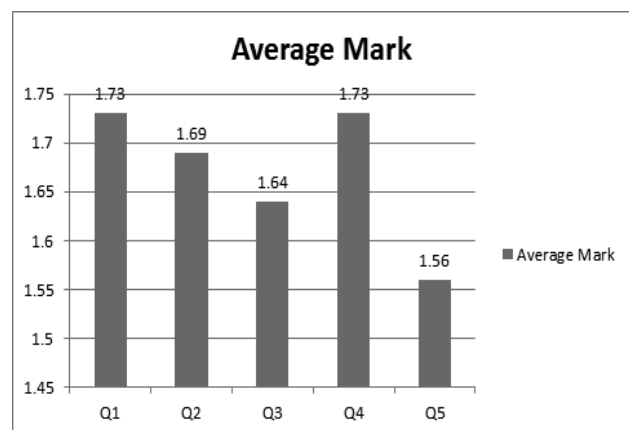


Fig. 2: Average mark obtained in each Question by the Students in Preconcept test

The different statistical measure obtained for Preconcept test is given in Table 3.

Table 3: Statistical measure for Preconcept Test

Measure	Percentage
Average grade of first attempts	82.00%
Average grade of all attempts	83.56%
Average grade of last attempts	85.50%
Average grade of highest graded attempts	85.50%
Median grade (for highest graded attempt)	100.00%
Standard deviation (for highest graded attempt)	20.75%
Score distribution skewness (for highest graded attempt)	-1.1714
Score distribution kurtosis (for highest graded attempt)	0.0345
Coefficient of internal consistency (for highest graded attempt)	52.41%
Error ratio (for highest graded attempt)	68.98%
Standard error (for highest graded attempt)	14.31%

Integrated Data Analytics

The learner has enrolled in another course in the same semester in the same learning management system. The students' progress in that course is taken for comparison purpose.

Number Metrics

The total number of students in the class is 145. The number metrics of the same students participation in other course is given in Table 4.

Table 4: Students participation in other course

Activities	Number Of Views	Number of Students Participated
Preconcept Test	96	68
Discussion Forum	98	45
Introductory Text Material	110	120

Crawling Log Analytics

In order to identify whether the chapter material posted in Moodle is sufficient or not, the crawling data of the student is captured and analyzed. The crawled data is given in table 5

Table 5 : Crawled data

Topic	Number of Relevant URLs crawled	Number of Irrelevant URLs crawled
DataLink Layer	5	130
Application Layer	8	128

From the table it is inferred that the material posted is more relevant.

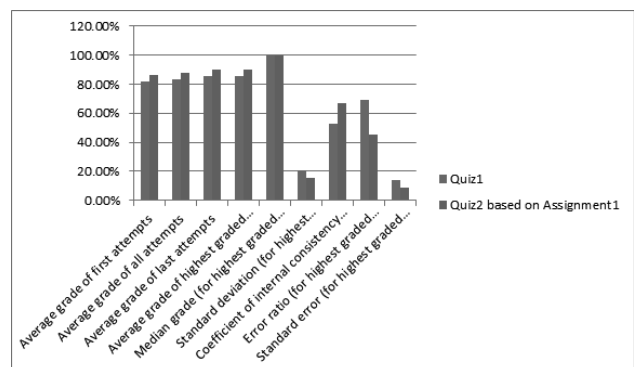
Forming Groups:

Initially the students were given an option of making a group and given an Team Assignment on Exploratory Network analysis of different labs in the department. Then Quiz was conducted for the Assignment. The performance of students in the Quiz is given in Table 6.

Table 6 : Performance of Students in Quiz based on Assignment1

Measure	Percentage
Sum of Average grade of first attempts	86.34%
Sum of Average grade of all attempts	87.56%
Sum of Average grade of last attempts	89.75%
Sum of Average grade of highest graded attempts	89.75%
Sum of Median grade (for all attempts)	100.00%
Sum of Standard deviation (for all attempts)	15.6%
Sum of Score distribution skewness (for all attempts)	-2.6936
Sum of Score distribution kurtosis (for all attempts)	1.19
Sum of Coefficient of internal consistency (for all attempts)	67.25%
Sum of Error ratio (for all attempts)	45.08%
Sum of Standard error (for all attempts)	8.78%

The performance comparison of Preconcept Quiz and Assignment Quiz is given in Figure 3.

**Fig. 3: Performance comparison of Preconcept Quiz and Assignment Quiz**

Preprocessing of Log Data

The learner's data such as their chat data, Activity log are collected. The VPI of the student are identified. Based on it the students are investigated and clustered.

k-means clustering technique is implemented to

form the clusters by assigning number of clusters as 7. The clusters are named as excellent, very good, good, Satisfied, need to improve, special attention is needed, not involved in the discussion, test or any activity. Final cluster centroids are given in Table 7 .

Table 7: Final Cluster Centroids

Cluster No	Full Data	1	2	3	4	5	6	7
Quiz	38.617	45.9375	31.5417	46.0769	46.9048	45.4	26.15	17.6364
Assignment	38.7021	28.75	46.9167	50.5128	25.6667	49.1	26.35	31.2727
Puzzle	47.5603	37.815	45.667	53.2564	53.5238	36.6	46.45	46.2727
Discussion	1.383	1	1	2	1.7619	2	1	0

The number of instances Clustered in each group is given in Table 8.

Table 8: Number of Instances in each Cluster

Cluster No	Number of Instances
1	16(11%)
2	24(17%)
3	39(28%)
4	21(15%)
5	10(7%)
6	20(14%)
7	11(8%)

Within cluster sum of squared errors obtained is 11.44. Then the each individual is marked with the cluster number. Then the new team is formed by

Table 9: Performance of Individuals in Quiz after the new assignment of Team

Measure	Percentage
Sum of Average grade of first attempts	94.54%
Sum of Average grade of all attempts	94.54%
Sum of Average grade of last attempts	94.54%
Sum of Average grade of highest graded attempts	94.54%
Sum of Median grade (for all attempts)	100%
Sum of Standard deviation (for all attempts)	12.15%
Sum of Score distribution skewness (for all attempts)	-3.6936
Sum of Score distribution kurtosis (for all attempts)	17.19
Sum of Coefficient of internal consistency (for all attempts)	73.26%
Sum of Error ratio (for all attempts)	51.71%
Sum of Standard error (for all attempts)	6.28%

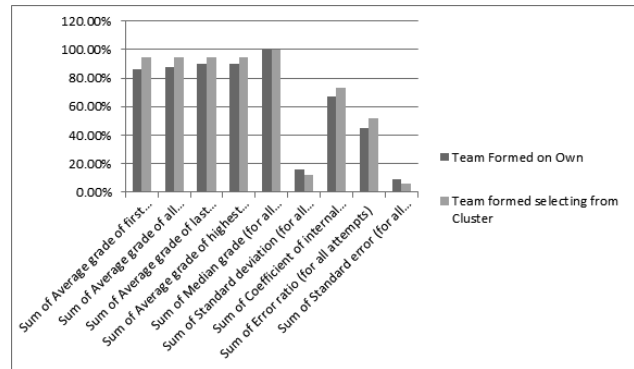


Fig. 5 Performance comparison of student of their own assigned team and newly assigned team.

selecting an individual from each of the seven groups. Hence the new team comprises of all categories of learners .The new team formed is then assigned with different activities such as Quiz, Assignment, project and discussion. Performance of students in Quiz conducted based on Assignment2 is given in Table 9.

Figure 5 shows the performance comparison of student of their own assigned team and newly assigned team.

We observe that the performance of individual has been improved in terms of participation and marks scored in various activities.

Conclusion

Learning analytics is an emerging learning domain which can be used for map the curriculum, predict the learning style and behavior, design and develop teaching – learning interventions and reflect the learning experiences. The learning behavior of the students is analyzed in detail by using the activity log of the students in Moodle login. Based on that, students are grouped into seven clusters. Again a new team is formed by selecting an individual from each cluster such that a team comprises of different levels of learners. The performance of the new team is better than the old team.

References

[1] Siemens, George; Long, Phil,” Discovering genres of online discussion threads via text mining-Using data mining as a strategy for assessing asynchronous discussion forums”,Penetrating the Fog: Analytics in Learning and Education,pp.30-40,2011

- [2] SauliusPreidys& Leonidas Sakalauskas ,”Analysis of students' study activities in virtuallearning environments using data mining methods”,*Technological and Economic Development of Economy*,vol.16(1),pp. 94-108,2010
- [3] Marcos Wander Rodrigues, Seiji Isotani, Luiz Enrique Zarate,” Educational Data Mining: A review of evaluation process in the e-learning”,*Telematics and Informatics*,vol.35,pp.1701-1717,2018
- [4] Leah P. Macfadyen, Shane Dawson,” Mining LMS data to develop an "early warning system" for educators: A proof of concept”,*Computers & Education*,vol.54,pp. 588–599,2010.
- [5] Hanan Khalil and Martin Ebner,” MOOCs Completion Rates and Possible Methods to Improve Retention- A Literature Review”. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*,pp.1236-1244, 2014
- [6] Wolfgang Greller and HendrikDrachsler,” Translating Learning into Numbers: A Generic Framework for LearningAnalytics”,*Translating Learning into Numbers: A Generic Framework for Learning Analytics*,vol.15(3),pp.42-57,2012
- [7] Aderibigbe Israel Adekitan and OdunayoSalau ,”The impact of engineering students' performance in the first three years on theirgraduation result usingeducational data mining”,*Heliyon* 5,2019
- [8] John Watson and Butch Gemin ,'Using Online Learning for At-Risk Students and Credit Recovery”,*PROMISING PRACTICESIN ONLINE LEARNING*. (North American Council for Learning(NACOL)),2008
- [9] RYAN S.J.D. BAKER,KALINA YECEF,” The State of Educational Data Mining in 2009: A Review and Future Visions”, 2009
- [10] Srec̃koNatek , MotiZwilling,” Student data mining solution–knowledge management system relatedto higher education institutions”,*Expert systems with Application*,vol.41,pp. 6400- 6407, 2014.
- [11] HananAldowaha, Hosam Al-Samarraie, Wan MohamadFauzy, “Educational data mining and learning analytics for 21st century higher education: A review and synthesis”,*Telematics and Informatics* ,vol-37,pp.13 -49, 2019.
- [12] SegunI.Popoola ,AderemiA.Atayero, JokeA. Badejo, TemitopeM. John, JonathanA. Odukoya b , DavidO.Omole.”Learning analytics for smart campus: Data on academic performances of engineering undergraduates in Nigerian private university”, *Data in brief*,vol.17,pp. 76 -94 ,2018
- [13] Mason, R. (2001), *Time is the New Distance? An Inaugural Lecture*. The Open University, Milton Keynes, 14 February. Webcast available at: http://kmi.open.ac.uk/projects/stadium/live/berrill/robin_mason.html