



## Grouping of Success Levels in E-Learning Learning Factors: Approaches with Machine Learning Algorithm

Elin Panca Saputra<sup>1</sup>, Sugiono<sup>2</sup>, Indriyanti<sup>3</sup>, Supriatiningsih<sup>4</sup>, Hafis Nurdin<sup>5</sup>

<sup>1234</sup>Fakultas Teknik dan Informasi, Universitas Bina Sarana Informatika, Jakarta Indonesia

<sup>5</sup> Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta Indonesia

E-mail: <sup>1</sup>elin.epa@bsi.ac.id, <sup>2</sup>sugiono.sgx@bsi.ac.id, <sup>3</sup>indriyanti.iyt@bsi.ac.id, <sup>4</sup>supriatiningsih.stq@bsi.ac.id, <sup>5</sup>hafis.nnr@nusamandiri.ac.id

### ARTICLE INFO

#### Article history:

Received: 10/03/2021

Revised: 20/03/2021

Accepted: 21/03/2021

#### Keywords:

Grouping, E-Learning, K-means, SVM

### ABSTRACT

The purpose of this study is to obtain the results of the modeling process on grouping the results of student learning, and to produce student success rates, while to find the results of the accuracy level of student learning success based on E-Learning with the Support Vector Machine (SVM) method. In this grouping, there are 5 clusters. Furthermore, the process of counting can be as many as 2 iterations, namely getting the final result in the form of Cluster-1 with a total of 10 students, cluster-2 with a total of 45 students, cluster-3 with a total of 22 students, cluster 4 with a total of 13 students, and the next is cluster-5 with a total of 19 students. Then the results of the resulting process with a total of 5 types of clusters, namely from students who get the highest results to the lowest. In addition, this study also looks for the level of accuracy in e-learning student learning processes using the Support Vector Machine (SVM) method, the accuracy results obtained are 90.91%, while the AUC results are 82.81%. then the value of the calculated accuracy rate can be classified as accuracy with the predicate result that is good.

Copyright © 2021 Jurnal Mantik.  
All rights reserved.

## 1. Introduction

The Covid-19 epidemic affects all areas of life, one of which includes learning activities organized by educational institutions in this part of the world. Therefore, online methods or online learning are a method that is more effective in learning today and is widely used by universities (Learning & Approach, 2020). eLearning is a type of learning method used in distance education that allows synchronous and asynchronous exchange of resources with communication networks. (Khan, 1998). but there are problems faced by higher education institutions that are not ready to face e-learning based learning systems because they do not have experienced resources in online or online learning. E-Learning, which is the latest technology, is becoming more affordable in higher education but poses major constraints in the development costs of its resources (Muniasamy et al., n.d., 2020).

Online or online learning refers to an electronic learning environment, unlike classroom learning, or face-to-face learning, and there is freedom of space and time when learning. However, online learning makes learning more flexible and provides an alternative to those who cannot attend face-to-face classes for whatever reason. (Baber, 2020). The use of new technology in educational institutions has resulted in many trends such as VLEs or Massive Open Online Courses (MOOC). As a result, students no longer have to be in an educational institution to receive learning services. Although there are many VLEs that offer MOOCs, such as Absorb, Schoology or Blackboard, Moodle is the most popular and used LMS. (Rivas et al., 2021). Due to the features, design and implementation of E-learning systems have grown exponentially in recent years (Wu et al., 2012).

Although, e-learning has limitations for lecturers and students who must have the ability and skills, on certain knowledge and experience in technology and pedagogical skills to complete their learning, the online system infrastructure must be synchronous, efficient, and secure. it is necessary to support teaching staff and student interaction, storing data, and evaluating the effectiveness of learning (Saini & Salim Al-Mamri, 2019). Therefore, the problem is how to have a good e-learning learning system. One of them is with an interesting approach, namely evaluating the learning system based on the factors that influence it (Mahmodi, 2017). In this study, it has a contribution to identify and investigate what factors influence the attitudes of students towards adoption and acceptance, e-learning systems can provide broader insights into external



factors and provide valuable recommendations for making policies on system implementation. e-learning learning (Kanwal & Rehman, 2017).

In previous research, factors that can influence the use of e-learning are technology and interface characteristics, experience content area, student roles and instructional assignments. Effective online learning requires interdependence with one another (Vonderwell & Zachariah, 2005). The factors that can influence include disclosing that factors related to technology (ease of use, accessibility of speed and service delivery), organizational factors (training support and diversity), environmental factors (user attitudes) and impact-related factors (learning experience, skills development), academic performance, and level of involvement) influence the adoption of e-learning facilities by students (Eze et al., 2020).

Evaluation of e-learning based learning systems is very important to determine the level of learning success (Saputra et al., 2020). This research was conducted to divide data into several clusters. Mka thus, can help managers of e-learning learning in evaluating and comparing factors that can influence the system to make the right decisions and build a better system. (Learning & Approach, 2020). One more interesting approach is the use of data mining methods, sometimes called Knowledge Discovery in Databases- (KDD), which is an effective way of using new technologies with high potential to help focus on information In education, predictions of student achievement, detection of Inappropriate learning behavior, and development of student profiles can be considered as e-learning problems and therefore data mining is successful in solving them (IONITA, 2016).

So in the discussion of the description above, identifying the results of student learning using e-learning, we made a classification model for forecasting the learning medium. which will help us in calculating student learning outcomes. In short, to solve two different problems. First, evaluate for grouping factors and features based on the clustering method using the K-mean algorithm. The reason for using the k-means clustering algorithm is that this method is the oldest and best known in cluster analysis. It has been studied extensively in various extensions and is applied in various substantive fields (Yang & Sinaga, 2019). Clustering method can be categorized as a method that can group the system by partition (Sugiono et al., 2019). One of the most commonly used partitioning algorithms is k-means, which is often used for grouping text due to its ability to converge to local optimals even for a very large sparse matrix. (Naeem & Wumaier, 2018).

Second, to get the level of prediction, namely by using classification techniques to support students in e-learning learning. Experiments collected data samples as many as 109 students studying at the University. We Tried to use a Support vector Machine (SVM) algorithm, which is a highly non-linear network and a single layer of higher generalizability; it can classify invisible patterns properly (Cristianini, N., n.d., 2013). By using a Support vector Machine (SVM) classifier can minimize structural risk, not empirical risk like in other classifiers. This maximizes the distance between the pattern and the class which separates the hyperplane together to distinguish patterns belonging to different classes (Khamparia & Pandey, 2018).

## 2. Research Method

### 2.1 Research Review

The stages in conducting this research consist of several steps, namely:

- a. Identify and determine the problems and objectives of the research to be carried out. The problems and objectives of this study are to evaluate the measurement of the influence of factors and features of e-learning learning.
- b. Build a predictive model using classification techniques to support students in choosing appropriate learning methods using the Support Vectore Machine (SVM) algorithm.
- c. Collecting Using Data Sets We obtained data by distributing questionnaires and logs of student teaching and learning activities using the Learning Management System (LMS) that has been provided by educational institutions.
- d. Doing clustering based on the method used, the most appropriate in this study is clustering, namely using the K-mean algorithm.
- e. Conducting an evaluation, namely an evaluation that is carried out to identify the factors that can affect the e-learning-based learning system.

### 2.2 Stages of the K-Means Algorithm

The K-Means algorithm is one of the most popular clustering techniques that relies on measuring distances (Banerjee et al., 2016). grouping can be defined as: Given a set of examples, organize them into groups. In a data set,  $X$ , consists of  $n$  data points in attribute space,  $A$ , composed attributes  $d$ , each data point can be represented as  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$ , Where,  $i = 1, 2, \dots, n$ . The grouping problem is partitioning the points in such a way that:

$$X = C_1 \cup \dots \cup C_k C_{outliers} \quad (1)$$

where  $C_1 \dots C_k$  is the  $k$  cluster and  $C_{outliers}$  contains outliers. If the clusters do not overlap, then:

$$C_1 \cap C_j = \emptyset, i \neq j \quad (2)$$

a. K-Means

K-Means Algorithm (Banerjee et al., 2016), seen in TABLE I, is the most commonly used partition algorithm. Distance The measure commonly used is the Euclidean distance, which is given by:

$$\sqrt{\sum_{i=1}^d (x_i - y_i)^2} \quad (3)$$

where  $x_i$  and  $y_i$  are two points in Euclidean - $d$  space dimension. The objective function to be minimized is called the Sum of Squared Error (SSE), given by the following equation:

$$SSE = \sum_{i=1}^d \sum_{x_i \in C_k} (x_i - c_k)^2 \quad (4)$$

The cluster centroid  $c_k$  can be updated as:

$$c_k = \frac{\sum_{x_i \in C_k} x_i}{C_k} \quad (5)$$

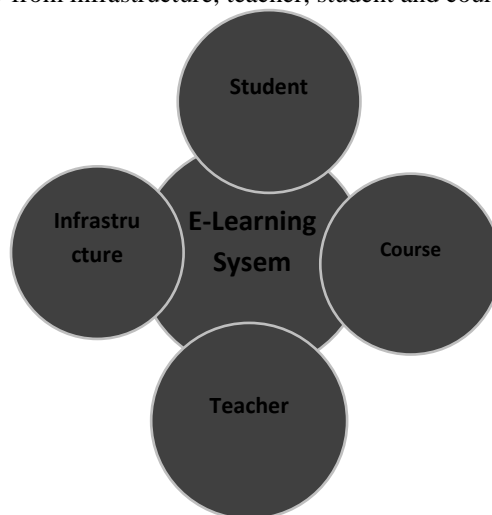
**Table 1.**  
Algoritma K-Means

| No | Nama   |
|----|--|
|    | Generate $k$ random numbers and corresponding $k$ k                        |
| 1. | Centroid group,  |
| 2. | For Each Example,<br>Set Instance to group with the nearest future center, |
| 3. | Repeat,  |
| 4. | Renew centroid,  |
| 5. | Reassign Instances to cluster,   |

source: (Banerjee et al., 2016)

b. Influencing Factor Model

In Fig 1, we can see several factors that influence the evaluation of learning activities using an e-learning system, namely from infrastructure, teacher, student and course on content.



**Fig 1.** Influence Factors [1]

c. K-MEANS Clustering Flow Model.

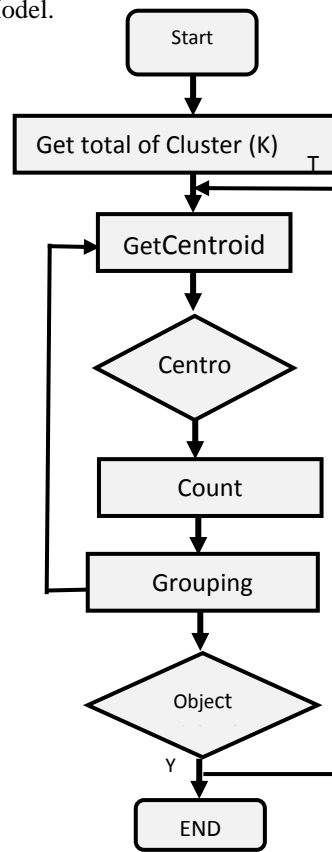


Fig 2. K-MEANS clustering flow

In Fig 2, the model we proposed 3 shows a process carried out in the modeling stage to complete the results of clustering. Fig 2 explains that in a process for clustering, the first step to clustering is to apply the number of clusters to be made. Next The next step is to determine the centroid or center of the cluster by selecting the next data that will be held as the centroid of each specified cluster. The next step is to calculate the distance between each data and the predetermined centroid cluster. Next, the data is grouped with the closest centroid cluster.

### 2.3 Support Vector Machine-(SVM)

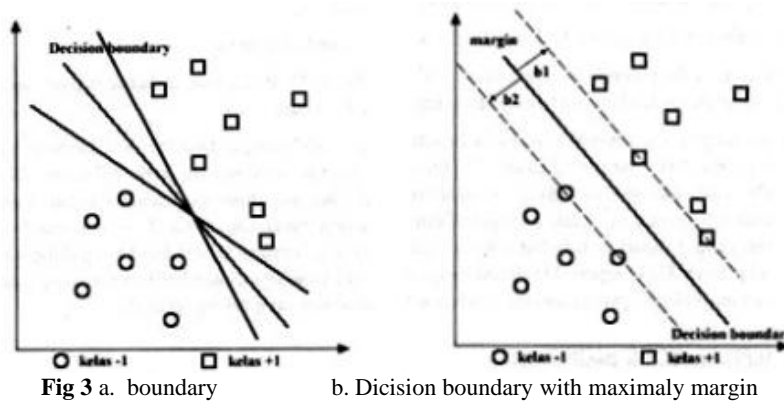
Support Vector Machine algorithm is a more precise classification, the classification shows better application performance, it has generalization performance, which can avoid the selection of local minimum structures and problems for small sample cases on machine learning, and is suitable for solving nonlinear and dimensional problems high (Riyanto et al., 2019). The Support Vector Machine can also be used in personalized e-learning applications and provides solid evidence of feasibility and feasibility (Gong & Wang, 2011).

$$-\frac{1}{2} \sum_{i,j=1} \alpha_i \alpha_j \langle x_i x_j \rangle - \varepsilon \sum_{i=1} \alpha_i + \sum_{i=1} y_i \alpha_i$$

Where  $\alpha_i = \beta_i - \beta_i^*$  and  $\beta_i, \beta_i^*$  obtained by completing a quadratic program and a Lagrange multiplier, that is, which handles problems from non-linear programs, the parameter vector  $w$  is obtained by:

$$w = \sum_{i=1} (\beta_i^* - \beta_i) \varphi(X_i)$$

Then the Support Vector Machine (SVM) algorithm can maximize from the hyperplane limit or (maximal-margin-hyperplane), as shown in Fig 2 (a) there are a number of possible hyperplane options for the data set, and 2 (b) is hyperplane with the maximum margin. Even though 2 (a) should be able to use an arbitrary hyperplane, the hyperplane at the maximizing margin will give a better generalization to a classification.



### 3. Results and Discussion

#### 3.1 Test-Results Using the K-Means Method

Basically, the K-Means method in sequence is as follows as in Fig 2 below which is a flow chart of the KMeans method used in clustering between the number of assignments, grades and the number of student activity logs. The following is the data that will be processed into the cluster grouping, in the form of the results of e-learning lectures, namely the number of tasks, values and the number of activity logs.

**Table 2**  
Determine Cluster

| CODE | Task | Total Grade | LOG |
|------|------|-------------|-----|
| M1   | 4    | 81,00       | 767 |
| M2   | 4    | 87,62       | 505 |
| M3   | 4    | 88,24       | 139 |
| M4   | 4    | 88,24       | 225 |
| M5   | 2    | 25,31       | 341 |
| M6   | 4    | 73,43       | 449 |
| M7   | 4    | 78,33       | 161 |
| M8   | 4    | 82,17       | 298 |
| M9   | 4    | 87,98       | 500 |
| M10  | 4    | 80,00       | 330 |
| M11  | 4    | 80,00       | 245 |
| M12  | 4    | 76,00       | 460 |
| M13  | 4    | 80,00       | 456 |
| M14  | 4    | 92,04       | 321 |
| M15  | 4    | 73,00       | 367 |
| M16  | 4    | 94,90       | 356 |
| M17  | 4    | 94,14       | 334 |
| M18  | 4    | 86,74       | 376 |
| M19  | 3    | 60,00       | 345 |
| M20  | 4    | 98,10       | 378 |
| M21  | 4    | 95,24       | 455 |
| M22  | 4    | 96,05       | 398 |
| M23  | 4    | 100,00      | 345 |
| M24  | 3    | 66,43       | 387 |
| M25  | 4    | 93,19       | 272 |
| M26  | 4    | 80,34       | 343 |
| M27  | 4    | 100,00      | 165 |
| M28  | 1    | 12,57       | 367 |
| M29  | 4    | 60,52       | 374 |
| M30  | 4    | 66,43       | 332 |



| CODE | Task | Total Grade | LOG |
|------|------|-------------|-----|
| M31  | 2    | 59,79       | 325 |
| M32  | 4    | 85,38       | 376 |
| M33  | 2    | 46,03       | 367 |
| M34  | 4    | 95,10       | 417 |
| M35  | 4    | 94,70       | 411 |
| M36  | 4    | 97,90       | 765 |
| M37  | 4    | 81,91       | 415 |
| M38  | 4    | 95,46       | 112 |
| M39  | 4    | 100,00      | 322 |
| M40  | 4    | 93,92       | 344 |

(Experimen Results, 2021)

### 3.1 Cluster Initial Determination

**Table 3.**

*Cluster Start Center Point*

| C1(M7) | C2(M33) | C3(M58) | C4(M74) | C5(M103) |
|--------|---------|---------|---------|----------|
| 4,00   | 3,00    | 4,00    | 4,00    | 2,00     |
| 78,33  | 95,46   | 89,34   | 75,31   | 65,00    |
| 161    | 367     | 429     | 457     | 263      |

To calculate the distance between the data and the initial center of the cluster using the euclidean Distance equation in the calculation below:

### 3.2 The 1st Data Distance to the ClusterCenter

$$C1 = \sqrt{(4-4)^2 + (4-4)^2 + (81-78,33)^2 + (81-78,33)^2 + (767-161)^2 + (767-161)^2} = 606,01$$

$$C2 = \sqrt{(4-2)^2 + (4-2)^2 + (81-46,03)^2 + (81-46,03)^2 + (767-367)^2 + (767-367)^2} = 401,53$$

$$C3 = \sqrt{(4-4)^2 + (4-4)^2 + (81-89,38)^2 + (81-89,38)^2 + (767-429)^2 + (767-429)^2} = 338,10$$

$$C4 = \sqrt{(4-4)^2 + (4-4)^2 + (81-75,31)^2 + (81-75,31)^2 + (767-457)^2 + (767-457)^2} = 310,05$$

$$C5 = \sqrt{(4-4)^2 + (4-4)^2 + (81-65)^2 + (81-65)^2 + (767-263)^2 + (767-263)^2} = 504,25$$

### 3.3 The 1st Data Distance to the ClusterCenter

$$C1 = \sqrt{(4-3,6)^2 + (4-3,6)^2 + (81-76,37)^2 + (81-76,37)^2 + (767-138,8)^2 + (767-138,8)^2} = 628,22$$

$$C2 = \sqrt{(4-3,4)^2 + (4-3,4)^2 + (81-68,93)^2 + (81-68,93)^2 + (767-363)^2 + (767-363,7)^2} = 403,45$$

$$C3 = \sqrt{(4-3,95)^2 + (4-3,95)^2 + (81-87,8)^2 + (81-87,8)^2 + (767-411,5)^2 + (767-411,5)^2} = 3,55,57$$

$$C4 = \sqrt{\frac{(4 - 4)^2 + (4 - 4)^2 + (81 - 84,38)^2 + (81 - 84,38)^2}{(767 - 522,2)^2 + (767 - 522,2)^2}} = 244,79$$

$$C5 = \sqrt{\frac{(4 - 3,73)^2 + (4 - 3,73)^2 + (81 - 78,35)^2 + (81 - 78,35)^2}{(767 - 284)^2 + (767 - 284)^2}} = 482,85$$

**3.4 Final Cluster Determination**

Furthermore, it has been found in the final data in determining a cluster, from each of these clusters it can be determined for the final centroid of each of a cluster is:

**Table 3.**

the Last Centroid Cluster

| C1     | C2     | C3     | C4     | C5     |
|--------|--------|--------|--------|--------|
| 628,22 | 403,45 | 355,57 | 244,79 | 482,85 |
| 366,37 | 142,50 | 93,50  | 17,53  | 221,04 |
| 11,88  | 225,56 | 272,50 | 383,25 | 145,49 |

**3.5 Results Using the Support Vector Machine (SVM) Method**

The data we compute is from e-learning learning activities. To get the results of testing the accuracy level for calculating our data using the support vector machine (SVM) algorithm that we do, namely using the Rapidminer software, are as follows:

Table 4.

| Testing Results and Data Processing |                    |          |
|-------------------------------------|--------------------|----------|
| Classification                      | Predicted <> Class |          |
|                                     | Class =1           | Class =2 |
| Class =1                            | 69                 | 10       |
| Class =2                            | 0                  | 30       |

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{69 + 30}{69 + 30 + 0 + 10} = 0,9082$$

$$Sensitivity = \frac{TP}{TP + FN} = \frac{69}{69 + 30} = 0,6969$$

$$Specificity = \frac{TN}{TN + FP} = \frac{30}{30 + 0} = 1$$

$$PPF = \frac{TP}{TP + FP} = \frac{69}{698 + 24} = 1$$

$$NPV = \frac{TN}{TN + FN} = \frac{30}{30 + 10} = 0,75$$

After our data performs calculations with the Support Vector Machine (SVM) method of the confusion matrix, it can be seen in table 1. The results of the True Positive (TP) value have a total of 69 which is classified as 1 according to predictions made with the Support Vector Machine (SVM), and the results of False Negative (FN) have 10 data predicted as 1 but it turns out to be 2, then the True Negative (TN) result has a value of 30 data as 2 according to predictions, and False Positive (FP) has a score of 0 predicted data 2 turns out that 1. The accuracy level produced by the Support Vector Machine (SVM) algorithm is 90.91%, with the AUC value the accuracy value is 82.81% and we can calculate it to get the value of accuracy, sensitivity, specificity, ppv, and npv can be seen in table 3.

**4. Conclusion**

After we carried out the clustering process using the k-means method, it was obtained a number of success rates for E-learning students, divided into 5 clusters. The attributes that we use are students with the number of doing assignments, score scores, the number of activities log student activities using LMS e-Learning. Students with a large number of activity logs will get high scores and while students with a small number of activities will get low scores. To find out the patterns and behaviors carried out by other Elearning actors, it is necessary to carry out further research by processing other data generated from the Elearning learning process. While the results that have been calculated in the test to get the results of the success rate



using the Support vector machine method with the number of attributes as many as 3 predictor variables obtained 90.91% accuracy value, and the AUC value is 82.81%. Then with the result that the level of accuracy is included in the very good classification category or (very good classification).

## 5. References

- [1] Baber, H. (2020). Determinants of Students ' Perceived Learning Outcome and Satisfaction in Online Learning during the Pandemic of COVID19. 7(3), 285–292. <https://doi.org/10.20448/journal.509.2020.73.285.292>
- [2] Banerjee, S., Choudhary, A., & Pal, S. (2016). Empirical evaluation of K-Means, Bisecting K-Means, Fuzzy C-Means and Genetic K-Means clustering algorithms. 2015 IEEE International WIE Conference on Electrical and Computer Engineering, WIECON-ECE 2015, 168–172. <https://doi.org/10.1109/WIECON-ECE.2015.7443889>
- [3] Cristianini, N., & Taylor, J.S. (2013). An Introduction to support vector machines and other kernel based learning methods. Cambridge University Press. doi:10.1017/CBO9780511801389.
- [4] Eze, S. C., Chinedu-Eze, V. C. A., Okike, C. K., & Bello, A. O. (2020). Factors influencing the use of e-learning facilities by students in a private Higher Education Institution (HEI) in a developing economy. *Humanities and Social Sciences Communications*, 7(1), 1–15. <https://doi.org/10.1057/s41599-020-00624-6>
- [5] Gong, W., & Wang, W. (2011). Application research of support vector machine in E-learning for personality. CCIS2011 - Proceedings: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems, 638–642. <https://doi.org/10.1109/CCIS.2011.6045147>
- [6] IONITA, I. (2016). Data mining technique for e-learning. *Journal of Applied Computer Science & Mathematics*, 10(2), 26–31. <https://doi.org/10.4316/jacsm.201602004>
- [7] Kanwal, F., & Rehman, M. (2017). Factors Affecting E-Learning Adoption in Developing Countries-Empirical Evidence from Pakistan's Higher Education Sector. *IEEE Access*, 5(c), 10968–10978. <https://doi.org/10.1109/ACCESS.2017.2714379>
- [8] Khamparia, A., & Pandey, B. (2018). SVM and PCA Based Learning Feature Classification Approaches for E-Learning System. *International Journal of Web-Based Learning and Teaching Technologies*, 13(2), 32–45. <https://doi.org/10.4018/IJWLTT.2018040103>
- [9] Khan, B. H. (1998). Web-based instruction (wbi): An introduction. *International Journal of Phytoremediation*, 21(1), 63–71. <https://doi.org/10.1080/0952398980350202>
- [10] Learning, A. M., & Approach, A. (2020). education sciences The Factors Affecting Acceptance of E-Learning: Mahmodi, M. (2017). The Analysis of the Factors Affecting the Acceptance of E-learning in Higher Education. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 8(1), 1–9. <https://doi.org/10.5812/ijvlms.11158>
- [11] Muniyasamy, A., Alasiry, A., & Arabia, S. (n.d.). Deep Learning : The Impact on Future eLearning. 188–199.
- [12] Naeem, S., & Wumaier, A. (2018). Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K. *International Journal of Computer Applications*, 182(31), 7–14. <https://doi.org/10.5120/ijca2018918234>
- [13] Rivas, A., González-Briones, A., Hernández, G., Prieto, J., & Chamoso, P. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing*, 423(xxxx), 713–720. <https://doi.org/10.1016/j.neucom.2020.02.125>
- [14] Riyanto, V., Hamid, A., & Ridwansyah, R. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(1), 1–9. <https://doi.org/10.24014/ijaidm.v2i1.6424>
- [15] Saini, D. K., & Salim Al-Mamri, M. R. (2019). Investigation of Technological Tools used in Education System in Oman. *Social Sciences & Humanities Open*, 1(1), 100003. <https://doi.org/10.1016/j.ssaho.2019.100003>
- [16] Saputra, E. P., Supriatiningsih, Indriyanti, & Sugiono. (2020). Prediction of Evaluation Result of E-learning Success Based on Student Activity Logs with Selection of Neural Network Attributes Base on PSO. *Journal of Physics: Conference Series*, 1641(1). <https://doi.org/10.1088/1742-6596/1641/1/012074>
- [17] Sugiono, Nurdiani, S., Linawati, S., Safitri, R. A., & Saputra, E. P. (2019). Pengelompokan Perilaku Mahasiswa Pada Perkuliahan E-Learning dengan K-Means Clustering. *Jurnal Kajian Ilmiah Universitas Bhayangkara Jakarta Raya*, 19(2), 126–133.
- [18] Vonderwell, S., & Zachariah, S. (2005). Factors that influence participation in online learning. *Journal of Research on Technology in Education*, 38(2), 213–230. <https://doi.org/10.1080/15391523.2005.10782457>
- [19] Wu, Y., Nian, Q., & Gu, S. (2012). An improved Learning Evaluation system based on SVM for E-learning. 2012 IEEE 5th International Conference on Advanced Computational Intelligence, ICACI 2012, 527–529. <https://doi.org/10.1109/ICACI.2012.6463219>
- [20] Yang, M. S., & Sinaga, K. P. (2019). A feature-reduction multi-view k-means clustering algorithm. *IEEE Access*, 7, 114472–114486. <https://doi.org/10.1109/ACCESS.2019.2934179>

