International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 05, 2020
ISSN: 1475-7192

# Longitudinal Tracking of Students' Academic Progress using Model-based Clustering

Norliza Adnan and Norhaiza Ahmad*

***Abstract---*** *Clustering methods have often been used to group similar students' performance for tracking their academic progress. A typical approach is to group the students by treating the observations as accumulated average marks on several assessed subjects within a period of time. Here, similar characteristics of students are identified based on the overall variation of marks between the subjects but ignores the temporal aspect of students performance even though the assessments are carried out at different time points. Alternatively, such characteristics in the observations could be treated as a set of longitudinal data since the measurements consider time-spaced and repeated events. This paper aims to compare the output between these two different treatments of observations using Model-based Clustering (MBC). Specifically, the average observations are applied to the classical MBC whereas the longitudinal observations require some adjustment to the covariance matrix to cater for its longitudinal data structure. A synthetic data set generated based on some pre-university students' marks on four science subjects from three series of continuous assessments are applied to the methods. The results show that the longitudinal observations on adjusted MBC produce a greater number of clusters that could characterise students' progress with a better internal and external cluster quality compared to the average observations on classical MBC.*

***Keywords---*** *Model-based Clustering, Longitudinal Tracking, Academic Progress.*

## I. INTRODUCTION

Many academic institutions use performance-based assessments to track and evaluate students' academic progress as part of their learning process [1,2]. One strategy to provide more focus towards improving students' performances is by identifying homogeneous groups of students with similar academic progress through a series of assessments in different subjects simultaneously using cluster analysis. Non-parametric clustering methods such as hierarchical and K-means clustering algorithms are widely used to discover distinct patterns in groups of students [3-5]. These clustering methods differentiate groups of subjects based on certain dissimilarity measures. and assumes that the series of students assessments observed are independent. Alternatively, another type of clustering approach that partitions observations of interests is model based clustering (MBC) method. By assuming that the finite mixture of distributions generated the data vectors, the method is optimized by an Expectation-Maximization (EM) algorithm. This mixture based model is particularly appealing when clustering students' academic performance because it allows the identified groups to overlap and produce additional information in the form of probabilities and visual trajectory progress [5,6]. However often times, monitoring students' academic progress [7-,9] are based on a series of repeated type of observed assessments by following the students development overtime within an academic period. This include monitoring students achieved marks for a series of tests and examinations.

*Norliza Adnan, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, UTM Skudai, Johor, Malaysia*

*Norhaiza Ahmad*, Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, UTM Skudai, Johor, Malaysia.*

*E-mail: norhaiza@utm.my*

**DOI: 10.37200/IJPR/V24I5/PR201714**
**Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020** 

These observations can be translated as a longitudinal study since the data measurement considers time-spaced and repeated events. Thus applying distance based non-parametric clustering methods may not be appropriate. In addition, clustering such observations with dependence feature directly using the classical MBC may not be suitable because it will change the repeated events into univariate form by averaging them into the variables. Therefore, the longitudinal structure with time-spaced characteristics [10] that can give resulting insight into behaviour over time will be violated when they lost the variation of score marks amongst subjects. This paper aims to track similar students' academic progress by clustering them based on three repeated continuous assessment carried out within three months equal time-spaced for four different subjects in the academic period. We compare two different MBC based methods using data with different treatments. Firstly, the repeated continuous assessment are treated into an averaged univariate form and analyzed using classical MBC approach. Secondly, the observations with time-spaced characteristics are treated as longitudinal form data and analyzed using MBC approach with modified Cholesky decompositions. This adjustment to the classical MBC is necessary in order to deal with the covariance structure between measurements at different time points in the longitudinal observations. The clustering results for both approaches are compared in terms of appropriate number of clusters and the index quality of group membership (internal and external).

## II. METHODOLOGY

### *2.1 Model-based Clustering*

In a general finite mixture, the density for a random variable *y* takes the form

$$f(\mathbf{y}\,|\,\boldsymbol{\theta}) = \pi_1 f_1(\mathbf{y}\,|\,\boldsymbol{\theta}_1) + \ldots + \pi_K f_K(\mathbf{y}\,|\,\boldsymbol{\theta}_K) \tag{1}$$

where $f_k$ and $\boldsymbol{\theta}_k$ are the density and parameters of the *k*th component and $\pi_k$ is the probability an observation belongs to the *k*th component $\left(\pi_k > 0; \sum_k^K \pi_k = 1\right)$[11]. The full parameter vector $\boldsymbol{\theta}$ includes the prior probabilities $\pi_k$ and the component parameters $\boldsymbol{\theta}_k$. In many situations, the component densities are assumed as multivariate Gaussian parameterized with a mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$

$$f(\mathbf{y}\,|\,\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_k)}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}-\boldsymbol{\mu}_k)\right) \tag{2}$$

If the mean outcome values are thought to depend on explanatory variables, the mean vectors are replaced with $\boldsymbol{\mu}_k = \mathbf{x}\boldsymbol{\beta}_k$ such that **x** is a design matrix based on those variables that impact the mean. The covariance matrix can be simplified by assuming a structure such as independence $\left(\Sigma_k = \sigma_k^2 \mathbf{I}\right)$, with compound symmetry of $\left(\boldsymbol{\Sigma}_k = \sigma_k^2\left(\rho_k \mathbf{1}\mathbf{1}^T + (1-\rho_k)\mathbf{I}\right)\right)$, or exponential covariance $\left([\boldsymbol{\Sigma}_k]_{jl} = \sigma_k^2 \exp\left(-\left|t_{ij} - t_{il}\right| / r_k\right)\right)$. On the other hand, the covariance structure can be parameterized through the eigenvalue decomposition of the form

$$\boldsymbol{\Sigma}_k = \lambda_k \mathbf{D}_k \mathbf{A}_k \mathbf{D}_k^T \tag{3}$$

**DOI: 10.37200/IJPR/V24I5/PR201714**
**Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020** 

where $\mathbf{D}_k$ is the orthogonal matrix of eigenvectors, $\mathbf{A}_k$ is a diagonal matrix whose elements are proportional to the eigenvalues, and $\lambda_k$ is a proportional constant [12].

Previous work on mean-covariance models showed that, $\mathbf{T\Sigma T'} = \mathbf{D}$ is the relation of the decomposition of a covariance matrix $\mathbf{\Sigma}$ for a random variable, where **T** is a unique lower triangular matrix. Their diagonal elements **1** and **D** is a unique diagonal matrix which only allowed positive entries. This relation is known as the modified Cholesky decomposition.

The Gaussian distribution can be reparameterized to accommodate longitudinal data and to reduce the number of parameters. For the longitudinal data, the framework of Model-based clustering have been developed using Gaussian mixture models where the constraints were applied into modified Cholesky decomposition of the group covariance matrices in order to give parsimonious models [12]. The EM algorithm are used to fit the mixture models [13].

Gaussian mixture model, with a modified Cholesky-decomposed covariance structure, are assumed for each mixture component. Therefore, the density of an observation $x_i$ in group *k* is given by

$$f(\mathbf{x}_i \mid \boldsymbol{\mu}_k, \mathbf{T}_k, \mathbf{D}_k) = \frac{1}{\sqrt{(2\pi)^p |\mathbf{D}_k|}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_k)^T \mathbf{T}_k' \mathbf{D}_k^{-1} \mathbf{T}_k (\mathbf{x}_i - \boldsymbol{\mu}_k)\right) \quad (4)$$

Where $\mathbf{T}_k$ is the $p \times p$ lower triangular matrix and $\mathbf{D}_k$ is the $p \times p$ diagonal matrix that follow from the modified Cholesky decomposition of $\mathbf{\Sigma}_k$.

Now, there is the option to constraint the $\mathbf{T}_k$ or the $\mathbf{D}_k$ to be equal across groups and there is also the option to impose the isotropic constraint $\mathbf{D}_k = \delta_k \mathbf{I}_p$, which leads to a family of eight Gaussian mixture models. Each member of this family, along with their respective nomenclature and number of covariance parameters, is given in Table 1. The nomenclature is quite intuitive; for example, the VEA model has variable autoregressive structure and equal, anisotropic noise across groups.

The modified Cholesky decomposition are expressed in the form of $\mathbf{\Sigma}^{-1} = \mathbf{T'D}^{-1}\mathbf{T}$, when modelling the covariance of a multivariate Gaussian distribution. The values of **T** and **D** have interpretations as generalized autoregressive parameters and innovation variances, respectively [12] so that the linear least-squares predictor of $\mathbf{Y}_t$, based on $\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_t$, can be written

$$\hat{\mathbf{Y}}_t = \mu_t + \sum_{s=1}^{t-1} (-\varphi_{ts})(\mathbf{Y}_s - \mu_s) + \sqrt{d_t} \in_t, \quad (5)$$

Where $\in_t \sim N(0,1)$, while the $\varphi_{ts}$ are the (sub-diagonal) elements of **T** and the $d_t$ are the diagonal elements of **D** [12].

**DOI: 10.37200/IJPR/V24I5/PR201714**
**Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020**

Table 1: The Nomenclature, Covariance Structure and Number of Covariance Parameters for Each Model

| **Model** | $\mathbf{T}_k$ | $\mathbf{T}_k$ | $\mathbf{T}_k$ | **No of Cov Parameters** |
|---|---|---|---|---|
| EEA | Equal | Equal | Anisotropic | $p(p-1)/2+p$ |
| VVA | Variable | Variable | Anisotropic | $G[p(p-1)/2]+Gp$ |
| VEA | Variable | Equal | Anisotropic | $G[p(p-1)/2]+p$ |
| EVA | Equal | Variable | Anisotropic | $p(p-1)/2+Gp$ |
| VVI | Variable | Variable | Isotropic | $G[p(p-1)/2]+G$ |
| VEI | Variable | Equal | Isotropic | $G[p(p-1)/2]+1$ |
| EVI | Equal | Variable | Isotropic | $p(p-1)/2+G$ |
| EEI | Equal | Equal | Isotropic | $p(p-1)/2+1$ |

The $\mathbf{T}_k$ are constrained to be equal across groups, which suggests that the correlation structure of the longitudinally recorded data values is the same for all of the groups. In this context, the autoregressive relationship between time points are reflected the correlation structure as outlined in (5). The variability at each time point is taken to be the same for each group in the constraint $\mathbf{D}_k$, while the variability suggests that is the same at each time point in group *k*, imposing the isotropic constraint $\mathbf{D}_k = \delta_k \mathbf{I}_p$. Therefore, for each given data set, it is appropriate for any of the eight combinations of these constraints given in Table 1.

### *2.2 Assessing the Performance of Clustering Results*

The clustering performance for both averaged and longitudinal datasets are compared for the classical and adjusted model-based clustering approaches respectively. The clustering performance are evaluated based on: (a) the number of cluster groups produced such that cluster groups with a higher number of clusters are deemed preferable as it tends to show more variety in groups of students' with similar characteristics. (b) the evaluation of internal (Dunn Index & Sillhouette Index) and external (Adjusted Rand Index) cluster qualities.

Dunn index value lies between zero and $\infty$ and measures the ratio of the smallest distance between observations not in the same cluster to the largest intra-cluster distance. Thus, a high value of Dunn Index indicates that the points are well clustered. Another internal measure is the Sillhouette index. This index measure the degree of confidence in the clustering assignment of a particular observation [14,15]. Here, an index with values close to one indicates that the points are well-clustered. On the other hand, the external cluster quality is measured using the adjusted rand index (ARI) for the agreement between two partitions. A higher index value of ARI indicates better partition accuracy [16].

## III. DATA GENERATION AND ARRANGEMENTS

### *3.1 Design of Synthetic Data*

The synthetic data for this study are simulated based on the results of actual tests taken by 250 students from a pre-University institution in Malaysia.

**DOI: 10.37200/IJPR/V24I5/PR201714**
**Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020** 

Table 2: The Details of Data Representations

| | **Details** |
|---|---|
| **Subjects** | Mathematics, Chemistry<br>Physics, Biology |
| **Number of tests** | Test 1 (*Time 1*)<br>Test 2 (*Time 2*)<br>Test 3 (*Time 3*) |
| **Number of students** | 250 students |
| **Grade level** | Excellent (80 – 100)<br>Credit (60 – 79)<br>Pass (40 – 59)<br>Fail (0 – 39) |

The tests involved four science subjects, ie; Mathematics, Chemistry, Physics and Biology, conducted in the first semester of an academic year at three time points. The details are shown in Table 2. The marks obtained for Test 1 are represented by Time 1, conducted at the end of the first month in the first semester of the academic year. Similarly Test 2 (Time 2) and Test 3 (Time 3) were conducted in the following consecutive months. The simulated data was generated using normal distribution to follow the model assumptions with parameters mimicking the actual tests. In particular, data are generated based on four different levels of grades following a typical guideline by a pre-university institution.

### *3.2 Data Treatment: Average vs Longitudinal form*

The generated data produced in section 3.1 are then arranged into two sets of observations. For the first set of observations, the students marks are calculated as average marks over the three tests. Thus the data matrix to be clustered consist of a two-dimensional data of students and subjects. For the second set of observations, the students marks are arranged in a longitudinal form. Thus the data matrix to be clustered consist of a three-dimensional data of students, tests and subjects.

A total of 100 sets of data are generated for each form. Each set is randomised prior to applying the MBC clustering methods to ensure non-biasedness in the analysis. We have used *R*-programming software : *mclust* and *longclust* packages for this analysis.

## IV. DATA ANALYSIS

### *4.1 Cluster of Students based on Their Academic Progress*

Monitoring students' academic progress needs a series of repeated type of observed assessments by following students achievement over time within certain academic period. By clustering them into group with same characteristics, the specific remedial approach in teaching and learning can be developed based on the group performances and characteristics.

Model-based clustering are applied on data that treated in two ways, i.e; classical MBC on average data and

DOI: 10.37200/IJPR/V24I5/PR201714
Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020 

MBC with modified Cholesky on longitudinal data. The performance of both approaches can be summarized and compared based on the number of clusters and the measures of the clusters index. Results indicate the model-based clustering that treat the data using longitudinal form provides greater number of cluster group compared to an average form as presented in Table 3(a). The greater number of clusters implies that the method are able to show more variations of the student characteristics.

Table 3(a): The Cluster Group for the Model Fitted by MBC for Averaged and Longitudinal Data

| **Data** | **Cluster group** | | | | | | |
|---|---|---|---|---|---|---|---|
| **Averaged** | **1** | **2** | **3** | | | | |
| | 86 | 94 | 70 | | | | |
| **Longitudinal** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| | 30 | 37 | 38 | 72 | 37 | 20 | 16 |

Table 3(b): The Model Fitted by MBC for Averaged and Longitudinal Data

| **Data** | **Model** | **BIC** | **Df** |
|---|---|---|---|
| **Averaged** | VVE | –6533.20 | 32 |
| **Longitudinal** | VVI | –20506.55 | 174 |

Since the clustering scenario is simulated by taking three tests from three consecutive months, all the models are run and used Bayesian Information Criterion (BIC) to choose the model where the model that gives the minimum BIC score can be selected as the best model [12]. The BIC for averaged data (BIC : –6533.20, Table 3(b)) selects a VVE (diagonal, varying volume and shape) model (Table 1) that gives reasonable clustering performance (ARI : 0.428, Table 4). On the other hand, the BIC for longitudinal data (BIC : –20506.55, Table 3) selects a VVI model that gives slightly better clustering performance (ARI : 0.669, Table 4). The degrees of freedom (df) for the cluster groups of longitudinal data are high (df : 174, Table 3(b)) compared to averaged data (df : 32, Table 3), reflecting a fit that is very close to a Gaussian mixture.

Table 4. Cluster validity indices based on averaged and longitudinal data using MBC

| **Data** | **Internal** | | **External** |
|---|---|---|---|
| | **Dunn** | **Silhouette** | **Adj Rand Index (ARI)** |
| Averaged | 0.126 | 0.519 | 0.428 |
| Longitudinal | **2.805** | **0.993** | **0.669** |

From Table 4, the Internal cluster quality obtained from the longitudinal data under the adjusted MBC indicate that the Dunn Index is higher (2.805) and Sillhouette Index (0.993) is closer to one compared to the average data.

This implies that the points are well-clustered when the temporal information are retained. The external cluster quality also show that the ARI (higher index of 0.669) with longitudinal data also show better partitional accuracy than the averaged data. These comparison clearly indicates that clustering the repeated measurements data in the longitudinal structure best capture the data cluster as compared to the averaged data.

**DOI: 10.37200/IJPR/V24I5/PR201714**
**Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020** 

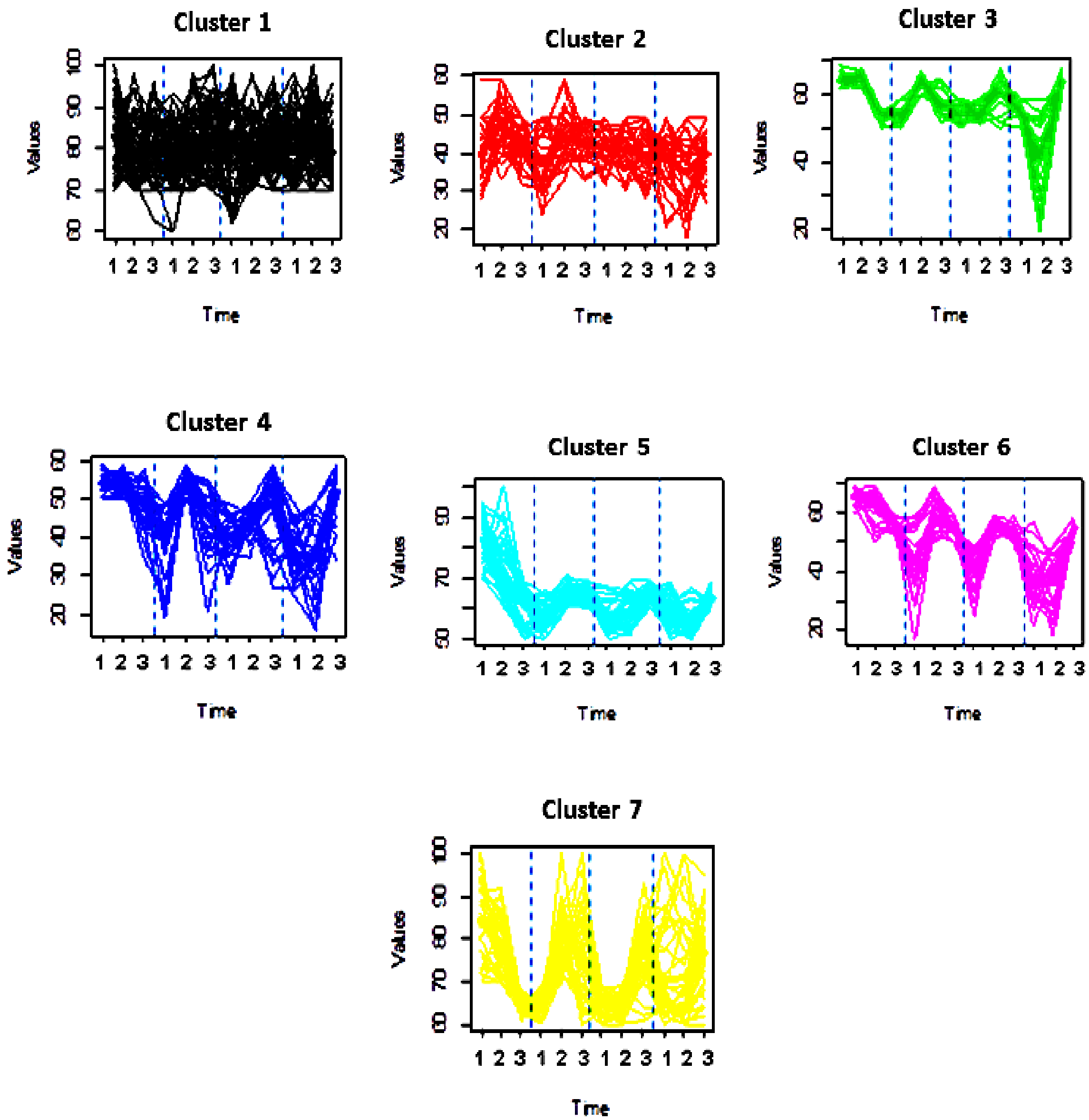


Figure 1: Clustering Pattern of Each Group for Longitudinal Data

The seven clusters (Table 3(a)) obtained from the longitudinal data with the adjusted MBC are shown in Figure 1. Each cluster show different characteristics and variation from the other. Specifically, we can see the progressive patterns of the students represented in the seven cluster groups for the three tests conducted across four subjects (distinct by dotted lines): starting with Mathematics, followed by Chemistry, Physics and Biology, respectively.

## V. Conclusion

In this paper, a series of tests from three consecutive months for four science subjects has been treated as average type of data and longitudinal type with time point. Classical MBC and modified MBC with Cholesky decomposition have been used to cluster both type of data, respectively. The purpose is to introduce the data with time point due to

DOI: 10.37200/IJPR/V24I5/PR201714
Received: 08 Feb 2020 | Revised: 03 Mar 2020 | Accepted: 18 Mar 2020 

the temporal aspect of students performance and treated as longitudinal form for tracking their academic progress. From the results, the clustering model of MBC with modified Cholesky for longitudinal data is superior to classical MBC for averaged data, which are proven by the validity indices to indicate better internal, external and relative cluster quality. This study also shows that the proposed data treatment, clustered by MBC with modified Cholesky is the best approach for representing longitudinal tracking of students' academic progress. As a conclusion, intervention activities that suits the group characteristics can be develop, hence, the objective to provide best approach in teaching and learning process can be achieved. The result also offers a good anticipation of the students' capabilities across subjects in order for the teachers to assist students.

## ACKNOWLEDGEMENTS

The authors are grateful to the Ministry of Higher Education, Malaysia (MOHE) and Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM) for the financial support under grant UTMFR REF NO : PY/2019/02100.

## REFERENCES

[1] J. O'Flaherty, J. Gleeson. Irish student teachers' levels of moral reasoning : Context, comparisons, and contributing influences. *Teachers and Teaching: Theory and Practice*, Vol. 23, No. 1, pp. 59–77, 2017.

[2] E. Kim, R.P. Corcoran. How engaged are pre-service teachers in the United States?. *Teaching and Teacher Education*, Vol. 66, pp. 12–23, 2017.

[3] M. de Morais, M. F. R. Araujo, E. B. Costa, Monitoring Student Performance Using Data Clustering and Predictive Modelling. IEEE, 2014.

[4] M. Singh, H. Nagar, A. Sant, K-mean and EM Clustering algorithm using attendance performance improvement Primary school Student. *International Journal of Scientific Research in Science, Engineering and Technology*, 2016.

[5] P. Jalote, B.N. Jain, S. Sopory. Classification for Research Universities in India. *Higher Education*, 2019.

[6] W. Hämäläinen, V. Kumpulainen, M. Mozgovoy, Evaluation of clustering methods for adaptive learning systems. In : Artificial Intelligence Applications in Distance Education. *IGI Global Book series Advances in Mobile and Distance Learning*, 2015.

[7] J. Xu, K.H. Moon, M. van der Schaar. A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs. *IEEE*, 2020.

[8] J. Leppink. When Negative Turns Positive and Vice Versa: The Case of Repeated Measurements. *Health Professions Education*, Vol. 5, pp. 76 – 81, 2019.

[9] R. Vilanova, M. Dominguez, J. Vicario, M.A. Prada, M. Barbu, M.J. Varanda, P. Alves, M. Podpora, U. Spagpnolini, A. Paganoni. Data-driven Tool for Monitoring of Students Performance. *IFAC PapersOnline*, Vol. 52, No. 9, pp. 165 – 170, 2019.

[10] O. Kwok, M. H. Lai, F. Tong, R. Lara-Alecio, B. Irby, M. Yoon, Y. Yeh. Analyzing Complex Longitudinal Data in Educational Research : *A Demonstration With Project English Language and Literacy Acquisition (ELLA) Data Using xxM. Front Psychol*. Vol. 9, pp. 790, 2018.

[11] M.S. Yang, S-J. Chang-Chien, Y. Nataliani. Unsupervised Fuzzy Model-based Gaussian Clustering. *Information Sciences*, Vol. 481, pp. 1–23, 2019.

[12] P.D. McNicholas, Model-Based Clustering. *Journal of Classification*. Vol. 33, pp. 331- 373, 2016.

[13] W.L. Wang, L.M. Castro, V.H. Lachos, T. Lin. Model-based Clustering of Censored Data via Mixtures of Factor Analyzers. *Computational Statistics & Data Analysis*, Vol. 140, pp. 104–121, 2019.

[14] M. Hoffman, D. Steinley, M.J. Brusco. A Note on Using the Adjusted Rand Index for Link Prediction in Networks. *Social Networks*, Vol. 42, pp. 72–79, 2015.

[15] S. Manochandar, M. Punniyamoorthy, R.K. Jeyachitra. Development of New Seed with Modified Validity Measures for K-Means Clustering. *Computers & Industrial Engineering*, Vol. 141, 2020.

[16] A.J. Gates, Y. Ahn. The Impact of Random Models on Clustering Similarity. *Journal of Machine Learning Research*, Vol. 18, pp. 1–28, 2017.