

## Data Analytics for Effectiveness Evaluation of Islamic Higher Education using K-Means Algorithm

Ara Hidayat<sup>1</sup>, Dindin Jamaluddin<sup>2</sup>, Dian Sa'adillah Maylawati<sup>3</sup>

<sup>1</sup>Faculty of Education and Teacher Training, UIN Sunan Gunung Djati Bandung, Indonesia. ORCID ID: 0000-0002-4782-1442

<sup>2</sup>Faculty of Education and Teacher Training, UIN Sunan Gunung Djati Bandung, Indonesia. ORCID ID: 0000-0003-2926-1434

<sup>3</sup>Department of Informatics, UIN Sunan Gunung Djati Bandung, Indonesia.

<sup>3</sup>Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia. ORCID ID: 0000-0002-1193-3370

<sup>1</sup>arahidayat@uinsgd.ac.id, <sup>2</sup>din2jamaluddin@uinsgd.ac.id, <sup>3</sup>diansm@uinsgd.ac.id

### Abstract

The aim of this research is to utilize data analytics technology in evaluating the development of Indonesian national curriculum based on Indonesian National Qualification Framework, especially in universities. This research uses Exploratory Data Analysis (EDA) and several clusterization method, among others K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++. The result of this research is not to measure the accuracy of clusterization result, but to discover the insight and interpretation information from data collections that related with national curriculum in Indonesia. Based on the EDA and clusterization methods with 30 variables of questions and 67 students as respondent, MiniBatch K-Means with 2 cluster has the best pattern that reliable with highest Silhouette Coefficient value. However, on average K-Means++ has better interpretation than the others, with the average of Silhouette Coefficient value is highest. From that result, this research found that generally around 77,67% students can understand and feel the application of the Indonesian national curriculum well, but specifically only about 19.4% of students really understand and feel the impact of the curriculum very well. This is important to be evaluated by curriculum users in this case students and tertiary educational institution to improve the quality of academic services in the application of the Indonesian national qualification network.

**Keywords:** Clustering, Curriculum, Education, Exploratory Data Analysis, Islamic Higher Education, Qualification Framework

### 1. Introduction

Every student is expected to be able to master the knowledge and skills in accordance with their respective fields in order to compete globally. Likewise, students graduating from Islamic Higher Education, are expected to have adequate competence in order to compete with the advancement of information technology. Some of these competencies include: the ability to communicate, think clearly, critically, have intelligence, responsibility and innovation in the environment and have the ability to live in a globalized society. All of these competencies are expected to be an antidote to the various negative behaviors that arise in the community such as drug abuse, corruption, test cheating (plagiarism) and other negative behaviors.

Many aspects are involved to produce competent graduates in order to compete in the current global era. Every college has a variety of challenges in an effort to prepare graduates who are qualified and in accordance with international standards. The problem of implementing higher education is required to always evaluate all aspects of education, both the system and available resources. One factor that is a major concern is curriculum

issues. Curriculum innovation is needed to harmonize the curriculum with the demands of the times. Curriculum innovation is defined in various ways with the following characteristics [1]: (1) targeted at improving learning and development that is managed in a planned manner; (2) implemented through the development and implementation of educational products that reflect new values and ideologies; (3) involves possible changes in stakeholders' beliefs and behaviors; and (4) promulgated in a multi-dimensional socio-cultural system. One way that can be taken is to create a standardized curriculum. In Indonesia, these efforts are implemented by the Indonesian National Qualification Framework for curriculum. This curriculum is a set of academic documents resulting from the fusion of the scientific ontology of the study program (deductive mindset) with the competence of the business world and the industrial world (inductive mindset). Indonesian National Qualification Framework was created in the hope that it could anticipate the needs of every graduate with jobs they would later work on. In its development this curriculum focuses on aspects of attitudes, knowledge and skills of students through learning experience in the form of instructional effects and natural effects. By implementing this curriculum, we want to address that higher education has taken a step further in trying to anticipate the needs of graduates who are ready to face competition at the global level.

Indonesian National Qualification Framework as a product of curriculum innovation is sure to have an effect from its implementation. Curriculum innovation always has an effect [2], this is because the curriculum is the heart of education and curriculum change is part of educational change [3], [4]. Many studies have been conducted relating to curriculum change or innovation [1]–[3], [5]–[9], in general the emphasis lies in how the actors implement and perceive the curriculum changes. The curriculum based on Indonesian National Qualification Framework, being a curriculum innovation for Universities in Indonesia, in its implementation the perception of prospective teacher students or prospective education practitioners as stakeholders of Higher Education is very important. In line with that, this study aims to analyze the perceptions of Islamic Higher Education students on the implementation of the Indonesian National Qualification Framework curriculum and its relevance to market demands in this case the fulfillment of work needs as an effort to evaluate the application of the Indonesian National Qualification Framework curriculum in Islamic Higher Education in Indonesia.

In the digital and technological era, data can be processed more than statistics, but can be processed and analyzed using data analysis techniques known as Exploratory Data Analysis (EDA) [10], [11]. EDA is also the first step or pre-process stage in preparing data before the data mining process is carried out, which can be in the form of classification and clusterization [12][13]. With EDA the data will be cleaner and can improve the quality of data mining result. This research, besides using EDA also use clusterization method as one of data mining technique. Data mining itself is a technique to find the insight knowledge from data where computer learn from the data collection, either supervised learning (classification), unsupervised learning (clusterization), or semi supervised learning (both classification and clusterization) [14]–[17]. Many algorithm or methods in data mining technique that can be used [18], such as Artificial Neural Network [19], K-Means [20]–[22], K-Nearest Neighbor [23], Fuzzy Logic [24]–[26], Decision Tree [27], and so on. Actually, every algorithm has different function and different case that can be solved. Clustering is an unsupervised learning method in data mining that not always to find the accuracy of the cluster, and not for predicting something. However, clustering is a method that can produce many insight of information from the pattern or cluster that is produced. This research aims to use EDA and several clusterization method to evaluate and discover the important information (the best pattern/cluster) about curriculum based on Indonesian National

Qualification Framework. The clustering methods that are used among others: K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++.

## 2. Research Method

This research uses quantitative method and utilize the Data Mining technology with clustering method. Generally, the method of this research that provided in Figure 1 begin from data collecting from the questionnaire, EDA, clustering (using K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++), evaluate each cluster using Silhouette Coefficient, and evaluate each cluster result. Then, evaluate each cluster to discover the insight of knowledge/infomation. The subjects of this research are collage students with ages were between 19-23 years old from Islamic Education Management department.

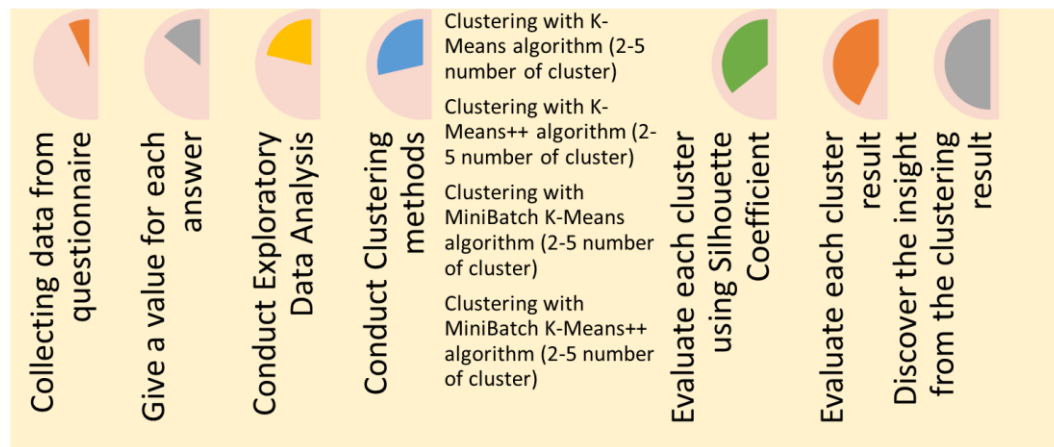


Figure 1 Research Methodology

## 3. Result and Discussion

### 3.1. Exploratory Data Analysis (EDA)

EDA is introduced as a procedures to analyze data an technique to interpret the result of data analysis easier, more acurate, more precise, and the result is mathematical (statistic) that is processed by machine automatically and can be applied for analyzing data [28], [29]. Basically, EDA provides summary of numerical data such as mean, median, maximum value, minimum value, an the quartiles. EDA aims to suggest hypotheses about the causes of observed phenomena, to assess the assumptions on which statistical inference will be based, to support the selection of appropriate statistical techniques, and to provide a basis for further data collection. The result of EDA usually visualized with graphical technique, such as box plot, histogram, pareto chart, scatter plot, multidimensional scaling, principal component analysis, and interactive version of plots. In data mining or machine learning technique, EDA usually used in pre-precossing process to visualize, find missing, and also to look for corellation between data or variables. Because pre-process phase is important for data inegration, data selection, data cleaning to improve quality, data transformation, and data reduction to run efficient mining process.

### 3.2. K-Means Algorithm

K-Means is one of widely used clustering algorithm that find the minimize distance value in the same cluster [30]–[32]. K-Means is a simple algorithm with fast time process and produce the optimal cluster. The algorithm of K-Means is as follows:

1. Determining number of cluster
2. Initializing the centroid value for each cluster ( $\mu_1, \mu_2, \dots, \mu_K \in \mathbb{R}$ ) randomly

3. Repeat the following formula (1) and (2) until convergence:

a. For every  $i$ , set

$$c^{(i)} := \arg \min_j \|x^{(i)} - \mu_j\|^2 \quad (1)$$

b. For every  $j$ , set

$$\mu_j := \frac{\sum_{i=1}^m 1\{c^{(i)}=j\}x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)}=j\}} \quad (2)$$

### 3.3. MiniBatch K-Means Algorithm

MiniBatch K-Means is also the development of K-Means algorithm for clustering with massive datasets [33], [34]. This algorithm reduce the number of distance computation with not use all the dataset, only the sample of the dataset [35]. The pseudocode of MiniBatch K-Means is as follow [36]:

```

Given: k, mini-batch size b, iterations t, data set X
Initialize each  $c \in C$  with an  $x$  picked randomly from X
 $v \leftarrow 0$ 
for  $i = 1$  to  $t$  do
     $M \leftarrow b$  examples picked randomly from X
    for  $x \in M$  do
         $d[x] \leftarrow f(C, x)$  //cache the center nearest to x
    end for
    for  $x \in M$  do
         $c \leftarrow d[x]$  //get cached center for this x
         $v[c] \leftarrow v[c] + 1$  //update per-center counts
         $\eta \leftarrow 1/v[c]$  //get per-center learning rate
         $c \leftarrow (1-\eta)c + \eta x$  //take gradient step
    end for
end for

```

### 3.4. MiniBatch K-Means Algorithm

MiniBatch K-Means++ is a combination of K-Means++ and MiniBatch K-Means algorithm [36]. So that, MiniBatch K-Means++ more robust because not randomly in giving the initialization centroid value. Besides, MiniBatch K-Means++ is better for massive datasets because use the sample data from all datasets, so that can reduce the computation process.

### 3.5. Silhouette Coefficient

The result of clustering must be measured to make sure that the pattern that is produced is well enough. There are internal measures and external measures. External measure such as Jaccard Coefficient [37], Purity [38], Precision and Recall [39], F-Measure [40], and so on. While, internal measures such as Z-Score Index [41], Gamma and Somer's Gamma [42], also Silhouette Coefficient [43], and so on. Silhouette coefficient is widely used to evaluate the result of clusterization. Silhouette Coefficient is a metrics that measures the separation and compactness of the cluster at the same time [43]–[46]. Formula (3) is used to calculate the average distances within a cluster and minimum distance between an object to other cluster.

$$SC = \frac{1}{N} \sum_{i=1}^N \frac{\beta_i - \alpha_i}{\max\{\alpha_i, \beta_i\}} \quad (3)$$

Where  $\alpha_i$  is the average distances of the objects in a cluster, that is (formula (4)):

$$\alpha_i = \frac{\sum_{j \neq i, x_j \in c_i} |x_i - x_j|}{|c_i|} \quad (4)$$

and  $\beta_i$  is the distance between an object  $x_i$  with the closest cluster centroid  $w_j$ .  $\beta_i$  is calculated with formula (5) below:

$$\beta_i = \min\{|x_i - w_j|, j = 1, 2, \dots, k, j \neq 1\} \quad (5)$$

An Silhouette Coefficient value can range from 1 to -1 ( $-1 \leq \text{Silhouette Coefficient} \leq 1$ ), where 1 means the clustering solution is “correct”

and -1 means the clustering solution is “incorrect”. However, in accordance with clusterization result, it offers no accuracy guarantees, but many interpretation from the clusterization result. So that, it is not guarantee if the Silhouette Coefficient value approaches 1 always has the right cluster and many interpretations, and vice versa.

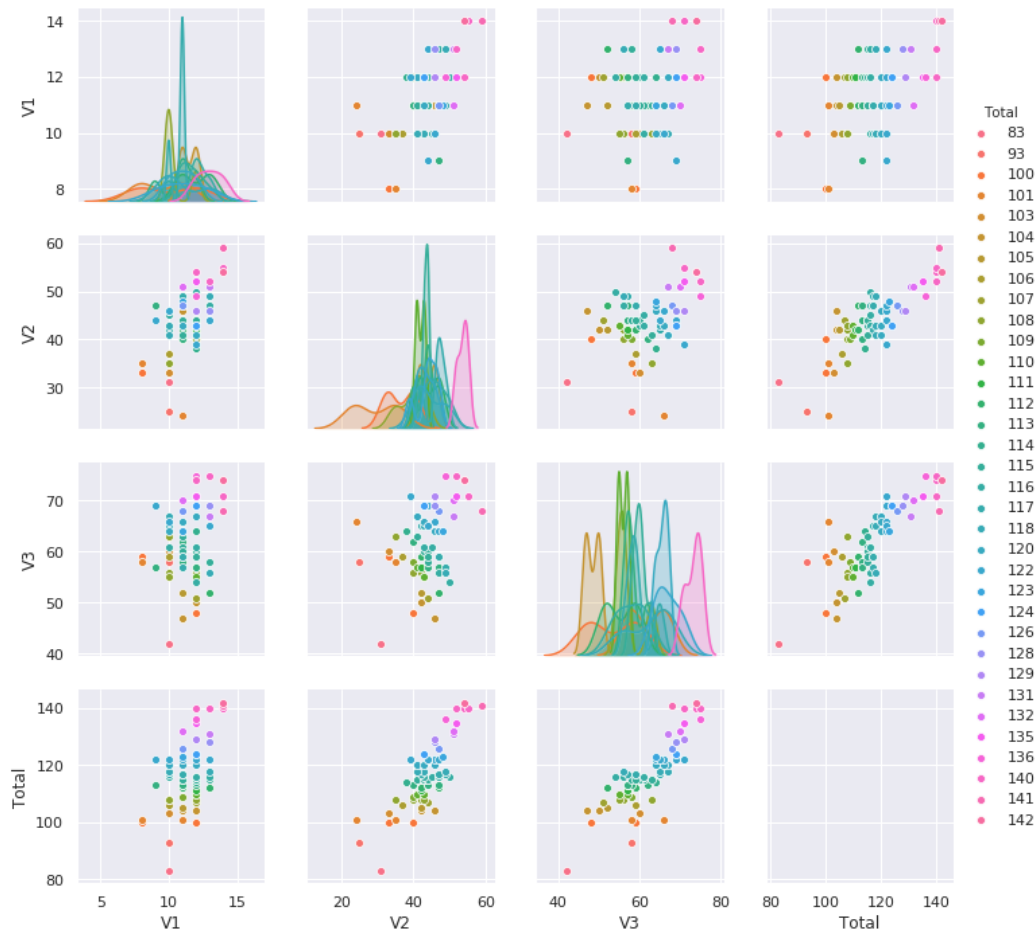
### 3.6. Exploratory Data Analysis (EDA) Result

Datasets are collected from questionnaire that distributed to collage student with ages were between 19-23 years old from Islamic education management department. Regarding the ethical issues, all participants in this study were asked to read the details of a consent form before signing off it as a sign that their participation in this study was voluntary, and all data were kept confidential. The questions of the questionnaire are related with curriculum that implemented based on Indonesian National Qualification Framework (provide in Table 1). Each question have a value from 1 until 5, where 1 for strongly disagree, 2 for disagree, 3 for neutral, 4 for agree, and 5 for strongly agree.

**Table 1 Content of Questions**

ID Question	Question
Q1	Am I able to explain theories according to the department / study program?
Q2	Am I able to use information and communication technology in education?
Q3	Am I able to adapt in different situation, society changing either local or global, demands for progress with applying my expertise?
Q4	Do I master the theoretical concepts and scientific foundation?
Q5	Am I able to think logically, critically, and innovative in the context of science and technology development or implementation that according with social value?
Q6	Am I able to show that I independent, qualified, and measured in work?
Q7	Am I able to examine the implications of the development or implementation of technological science that pays attention to and applies the values of the humanities in accordance with their expertise?
Q8	Am I able to make decisions appropriately in the context of problem solving in the field of expertise, based on the results of data and information analysis?
Q9	Do I contribute in improving the quality of life in the community, nation, state, and progress of civilization based on Pancasila (basic state of the Republic of Indonesia)?
Q10	Do I serve as a citizen who is proud and loves the country, has nationalism and a sense of responsibility in the nation and the State?
Q11	Do I appreciate the diversity of cultures, views, religions, and beliefs and original opinions or findings of others?
Q12	Am I aware of the level of qualifications in the Indonesian National Qualification Framework currently being pursued?
Q13	Do I understand the minimum abilities that must be possessed in accordance with the qualification level?
Q14	Do Graduates Competency Standars refer to the description of learning outcomes of Indonesian National Curriculum graduates and has equality with qualification levels?
Q15	I the courses that are presented refer to and support the qualification level of the Indoensian National Curriculum?
Q16	Is study program or department has established a Graduate Profile which is used as the basis for establishing Graduates' Competency Standards?
Q17	Is study program or department has set of Graduate Competency Standards or minimum criteria regarding the qualifications of graduates' abilities which include attitudes, knowledge and skills stated in the formulation of graduate learning outcomes?
Q18	Can I internalizing the spirit of independence, struggle and entrepreneurship?
Q19	Can I demonstrate an attitude of responsibility for work in his area of expertise independently?
Q20	Am I able to be responsible for achieving the results of group work and supervising and evaluating the completion of the work assigned to workers who are under his responsibility?
Q21	Am I able to carry out a self-evaluation process on work groups that are under their

ID Question	Question
	responsibility, and able to manage learning independently?
Q22	Do I have special skills that support self-development?
Q23	Do I have decision making skill?
Q24	Do I have a good work commitment?
Q25	Do My program study or department has set Learning Process Standards which are the minimum criteria for implementing learning in study programs to obtain graduate learning outcomes?
Q26	Do My department has a set of Learning Assessment Standards which are the minimum criteria for evaluating student learning processes and outcomes in order to fulfill graduate learning outcomes?
Q27	Do My department has a set of Learning Management Standards which are the minimum criteria regarding planning, implementation, monitoring and evaluation, control, and reporting of learning activities at the level of the study program?
Q28	Am I able to communicate both verbally and in writing by using Arabic and English in the development of the academic world and the world of work?
Q29	Am I able to follow the development of discourses and contemporary issues in the field of management of educational institutions for enrichment in the implementation of management of educational institutions that are creative, innovative and productive?
Q30	Am I able to follow and compete in the development of the world of work at the global level?



**Figure 2 The example pairplot visualization of EDA process**

Then, EDA is conducted using Python programming language and the result is shown in Table 2. The example visualization of EDA is shown in Figure 2. Figure 2 describes the relation between questions using pairplot visualization, if the questions is divided into 3 categories of similar questions. That example of visualization is not interpreted as true or false value, but many insight that can be

evaluated from the result. After EDA has been conducted, then clusterization method is implemented to find more insight of information that can be used to evaluate curriculum based on Indonesian National Qualification Framework. The source code implementation is shown in Figure 3, while the result of cluster from K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++ algorithm is described in Figure 4 using bar chart and scatterplot to see the accuracy and reliability of cluster result. The interpretation and evaluation of EDA, clusterization methods, and Silhouette Coefficient value (provided in Figure 5) will be explained in the next section.

**Table 2 Statistical Result as Part of EDA of Indonesian National Curriculum**

Variable	Count	Mean	Std.	Min	25%	50%	75%	Max
Q1	67	3,47761194	0,682202515	2	3	3	4	5
Q2	67	3,895522388	0,654456248	2	4	4	4	5
Q3	67	3,970149254	0,626924897	3	4	4	4	5
Q4	67	3,626865672	0,735173055	2	3	4	4	5
Q5	67	3,835820896	0,750696947	2	3,5	4	4	5
Q6	67	3,71641791	0,646809657	2	3	4	4	5
Q7	67	3,462686567	0,78486359	2	3	4	4	5
Q8	67	3,358208955	0,847482771	1	3	3	4	5
Q9	67	3,417910448	0,761761391	1	3	3	4	5
Q10	67	3,686567164	0,762947935	1	3	4	4	5
Q11	67	3,537313433	0,876085616	2	3	4	4	5
Q12	67	3,805970149	0,679545436	1	4	4	4	5
Q13	67	3,552238806	0,744343968	2	3	4	4	5
Q14	67	3,402985075	0,888643868	1	3	4	4	5
Q15	67	4,208955224	0,769441226	1	4	4	5	5
Q16	67	4,388059701	0,650296526	2	4	4	5	5
Q17	67	4,28358209	0,754902565	2	4	4	5	5
Q18	67	4,179104478	0,796305354	2	4	4	5	5
Q19	67	4,119402985	0,769147266	2	4	4	5	5
Q20	67	4,343283582	0,616742145	3	4	4	5	5
Q21	67	4,089552239	0,668134973	2	4	4	4,5	5
Q22	67	4	0,778498944	2	4	4	4,5	5
Q23	67	3,985074627	0,929198699	1	4	4	5	5
Q24	67	4,28358209	0,598127638	3	4	4	5	5
Q25	67	4,164179104	0,665081789	3	4	4	5	5
Q26	67	3,955223881	0,61380176	2	4	4	4	5
Q27	67	3,970149254	0,834305177	2	4	4	4	5
Q28	67	3,970149254	0,777917758	1	4	4	4	5
Q29	67	3,910447761	0,668134973	2	4	4	4	5
Q30	67	3,910447761	0,792604898	1	4	4	4	5

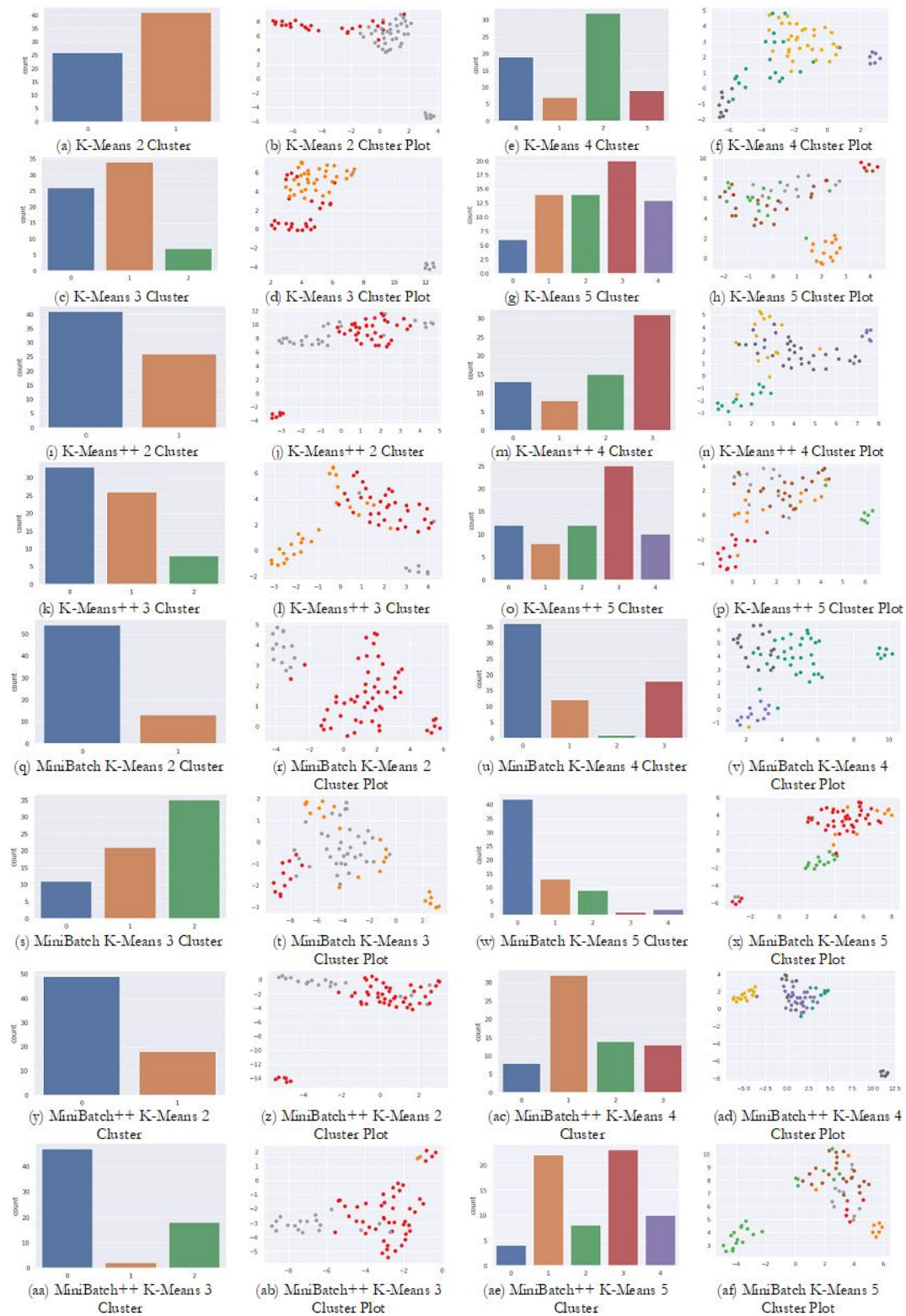
```
#Clustering with K-Means Algorithm
X2 = kkni_qa[['Q1','Q2','Q3','Q4','Q5','Q6','Q7','Q8','Q9','Q10',
              'Q11','Q12','Q13','Q14','Q15','Q16','Q17','Q18','Q19','Q20',
              'Q21','Q22','Q23','Q24','Q25','Q26','Q27','Q28','Q29','Q30']].values

kmX2_ = cluster.KMeans(n_clusters=2, init='random', max_iter=300, tol=0.0001, n_jobs=-1, random_state=9999)
kmX2_.fit(X2)

C_kmX2_ = kmX2_.predict(X2)
print("K-Means++ Result with 2 Clusters = ", C_kmX2_)
```

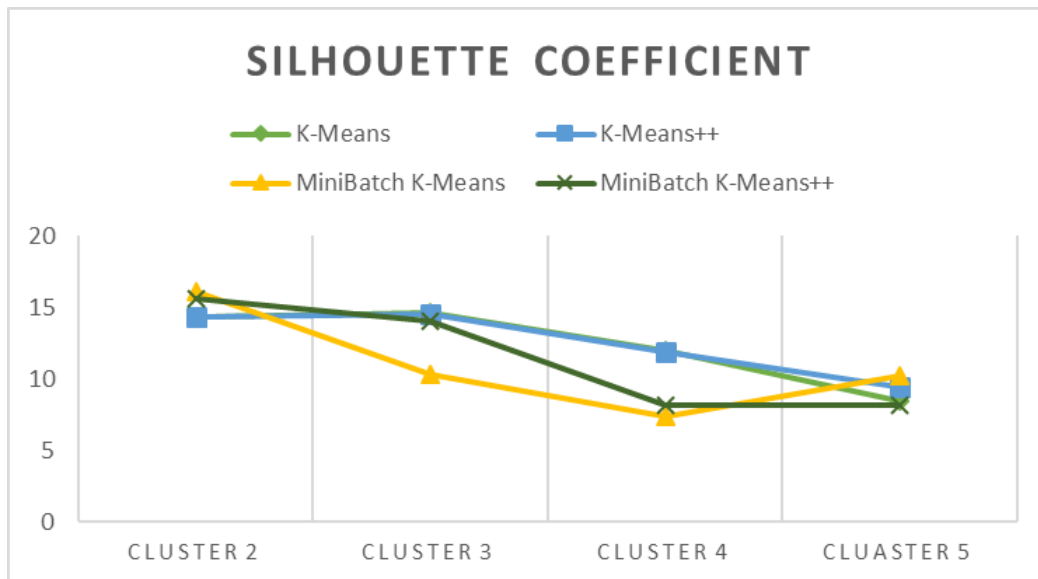
**Figure 3 The example of source code for clustering process**





**Figure 4 Result of Clustering Experiment**





**Figure 5 Silhouette Coefficient Value for the Result of K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++ to Evaluate Indonesian National Curriculum for Islamic Higher Education**

### 3.7. The Analysis and Interpretation of the EDA and Clustering Result

From the EDA process, clustering process, and evaluation the cluster result using Silhouette Coefficient, the analysis, evaluation, and interpretation of this research, among others:

1. Generally, more than 50% from 67 of students have been understand and implemented the curriculum well, both for personal and learning activities. This is evident from all the questions 50-75% on average they answered "agree" and "strongly agree", only a few questions were partially answered "neutral". These results also indicate that the curriculum based on Indonesian National Qualification Framework can be felt and used well by students and study programs/departments. Of course this result in line with curriculum innovation characteristic [1], that curriculum can be improve student understanding, either about the lecture material, the objective of their study programs, or understanding their skill that related with study program.
2. The manual calculation result shows that the highest score is 142 and the lowest score is 83 with an average of 116.51 and standard deviation of 11.617. The ideal score of this variable that can be obtained in the study is  $5 \times 30 \times 67 = 10050$ . After calculating the number of scores this variable is equal to 7806. Thus the variable value of student perceptions of the application of Indonesian National Qualification Framework displayed is  $(7806/10050) \times 100\% = 77.67\%$  of the expected 100%. Based on the criteria that have been made and the results of data analysis, the criteria for student perceptions of the application of the Indonesian National Qualification Framework are generally good.
3. From EDA process is found that there are several questions that have 1 value. It means still any students that not understand or strongly disagree about the content or the impact of current curriculum based on Indonesian National Qualification Framework, especially in Islamic Higher University. Based on analysis of the EDA result, there are any student that still disagree that curriculum give an impact on ability in making decision and problem solving, and also they doubt that curriculum can improve the quality of nasionalism and contribute to the nation. And also there are any student that non confident and not sure that the graduate quality from current curriculum based on Indonesian National Qualification Framework can be able to

compete in the era of globalization, both because of limited language and understanding of topic interest that is still lacking.

4. From the clustering process using K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++ with each algorithm is conducted for 2, 3, 4, and 5 cluster produce various cluster result that can be interpreted. However, based on the evaluation using Silhouette Coefficient value, in generally K-Means++ has the highest average Silhouette Coefficient value than the others. It is means that K-Means++ has many interpretation, although the distribution of cluster members formed is less reliable and quite a lot of outliers are produced. However, from all cluster result, MiniBatch K-Means algorithm for 2 cluster has the highest Silhouette Coefficient value. Cluster result of MiniBatch K-Means is accurate and reliable enough. Even tough there are 2 outlier member of one of cluster (red color), and 6 member of red color cluster which converge rather far away from the cluster. The analysis of 2 clusters that resulted from MiniBatch K-Means is first cluster is a group of students that has total score below 120, it means that in average student understand and get the impact of the current curriculum based on Indonesian National Qualification Framework well enough. However, the second cluster (grey color) that is a group of student that has total score more than 120 point show that only 19,4% students that very understanding and agree with current curriculum based on Indonesian National Qualification Framework, and they can be implemented and get the impact from current curriculum very well, either for personal or for the study program/department.

The results from point 4 indicate that there is still a need for socialization and other preparations by the study program, so that students' understanding of the Indonesian National Qualification Framework is not wrong and is sufficiently prepared to face it, because students are the main users of the Indonesian National Qualification Framework curriculum. The results of filling out this closed questionnaire also illustrate that knowledge, skills, quality manifestations, qualification levels, levels of learning achievement, self competence, educational standardization, and the dynamics of global development of students are still not satisfactory. This is indicated by their answers to the items given in the questionnaire, which shows their lack of understanding of the Indonesian National Qualification Framework which results in their perceptions as well. If it is confirmed by the existing theory, an explanation can be given that the Indonesian National Qualification Framework is a competency qualification framework that can juxtapose, equalize and integrate between the field of education and the field of job training and work experience in order to recognize work competencies in accordance with the work structure in various sectors. But on the other hand the results obtained in this study show that students as respondents to the study have not shown readiness as expected. This is indicated by their response to several questionnaire items that are still low.

#### 4. Conclusion

EDA and clustering methods using K-Means, K-Means++, Minibatch K-Means, and MiniBatch K-Means++ can produce many pattern that can be interpreted (Arthur & Vassilvitskii, 2007; Feizollah et al., 2015; Jiawei et al., 2012), and those pattern can be used to evaluate the curriculum based on Indonesian National Qualification Framework. The conclusion that can be drawn from the results of this research is that students' perceptions of the application of the Indonesian National Qualification Framework curriculum are good, but still need improvements and adjustments that continue to be sought so that academic services are maximized. Such efforts include providing courses to support other competencies in lectures, for example entrepreneurship is provided as an elective course for students who are interested in becoming entrepreneurs, it is also

necessary to provide advanced research and statistical methods for students who are interested in becoming researchers in the field of education. The provision of such elective courses also needs to be supported by adequate human resources, reliable lecturers and equipped with relevant and adequate learning resources and facilities. And field practice needs to be given more portions and the selection of internships for students also needs to be made more varied, not only in schools or educational institutions, so that students have extensive experience.

This research is important to be evaluated by the users and implementers of the curriculum in this case students and tertiary educational institutions, as an input to improve the quality of academic services in the application of the Indonesian National Qualification Framework-based curriculum. Because the curriculum must meet industry and institutional requirements (Chugh, Ledger, & Shields, 2017). The application of the Indonesian National Qualification Framework curriculum at Islamic Higher Education does not only depend on the graduates who are going to be printed by the department or study program, but also need a serious study related to the superior qualifications to answer the demands of users of education services. However, this research is still has the lack, such as the objects in this study are only limited to students' perceptions of the application of the Indonesian National Qualification Framework-based curriculum even though it can actually be linked to various other aspects such as the achievement of graduate learning. And the research should need to be supplemented by lecturers' perceptions of the Indonesian National Qualification Framework-based curriculum and innovations that lecturers do to keep up with curriculum changes that occur.

Last, for the further works, it needs more thorough case study needs to be carried out regarding the implementation of the Indonesian National Qualification Framework-based curriculum that involves more stakeholders (more data collection) so as to make a better contribution to the effective implementation of the Indonesian National Qualification Framework curriculum both at the institution of state and private universities. It is necessary to examine the effectiveness and the applicability of the Indonesian National Qualification Framework-based curriculum, especially in Islamic Higher Education. Then, it can implement another technique in data mining besides clusterization and another algorithm besides K-Means, K-Means++, MiniBatch K-Means, and MiniBatch K-Means++ to get better pattern.

## References

- [1] Y. Zhu and D. Shu, "Implementing foreign language curriculum innovation in a Chinese secondary school: An ethnographic study on teacher cognition and classroom practices," *System*, 2017.
- [2] A. Kostoulas and J. Stelma, "Understanding curriculum change in an ELT school in Greece," *ELT J.*, 2017.
- [3] A. Yildirim and K. Kasapoglu, "Teachers' perceptions of constructivist curriculum change as a predictor of their perceptions of the implementation of constructivist teaching-learning activities," *Asia Pacific Educ. Rev.*, 2015.
- [4] A. Musika, P. Pounsuk, N. Siriwan, and S. Intorrathed, "Relationship between curriculum and instructional materials of non-formal agricultural higher education," *J. Educ. Gift. Young Sci.*, 2019.
- [5] H. K. Altinyelken, "Curriculum change in Uganda: Teacher perspectives on the new thematic curriculum," *Int. J. Educ. Dev.*, 2010.
- [6] C. Yan and C. He, "Bridging the implementation gap: An ethnographic study of English teachers' implementation of the curriculum reform in China," *Ethnogr. Educ.*, 2012.
- [7] C. Yan, "'We can only change in a small way': A study of secondary English teachers' implementation of curriculum reform in China," *J. Educ. Chang.*, 2012.
- [8] M. Park and Y. K. Sung, "Teachers' perceptions of the recent curriculum reforms and their implementation: What can we learn from the case of Korean elementary teachers?," *Asia Pacific J. Educ.*, 2013.
- [9] Y. Kirkgöz, "A case study of teachers' implementation of curriculum innovation in English language teaching in Turkish primary education," *Teach. Teach. Educ.*, 2008.
- [10] J. T. Behrens, "Principles and Procedures of Exploratory Data Analysis," *Psychol. Methods*, 1997.
- [11] M. Stuart, D. C. Hoaglin, F. Mosteller, and J. W. Tukey, "Understanding Robust and Exploratory Data Analysis," *Stat.*, 2006.

- [12] H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [13] X. Yan, J. Han, and R. Afshar, "CloSpan: Mining Closed Sequential Patterns in Large Datasets," *Conf. Proc. Third SIAM Int. Conf. Data Mining, San Fr. CA, USA*, no. 0, 2003.
- [14] X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, 2014.
- [15] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth Edi. United States of America: Morgan Kaufman, 2012.
- [16] F. A. Hermawati, *Data Mining*. Yogyakarta: Andi, 2013.
- [17] A. Lamani, B. Erraha, M. Elkyal, and A. Sair, "Data mining techniques application for prediction in OLAP cube," *Int. J. Electr. Comput. Eng.*, 2019.
- [18] X. Wu *et al.*, "Top 10 algorithms in data mining," *Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1–37, 2008.
- [19] D. S. Maylawati, M. A. Ramdhani, W. B. Zulfikar, I. Taufik, and W. Darmalaksana, "Expert system for predicting the early pregnancy with disorders using artificial neural network," in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017.
- [20] H. Sulastri and A. I. Gufroni, "Penerapan Data Mining dalam Pengelompokan Penderita Thalassaemia," *Teknol. dan Sist. Inf.*, vol. 03, no. 02, pp. 299–305, 2017.
- [21] A. Taofik, N. Ismail, Y. A. Gerhana, K. Komarujaman, and M. A. Ramdhani, "Design of Smart System to Detect Ripeness of Tomato and Chili with New Approach in Data Acquisition," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 288, no. 1, p. 012018.
- [22] I. N. Anida and A. R. Salisa, "Driving cycle development for Kuala Terengganu city using k-means method," *Int. J. Electr. Comput. Eng.*, 2019.
- [23] D. Syahid, Jumadi, and D. Nursantika, "Sistem Klasifikasi Jenis Tanaman Hias Daun Philodendron Menggunakan Metode K-Nearest Neighbor (KNN) Berdasarkan Nilai Hue Saturation Value (HSC)," *J. Online Inform.*, vol. 1, no. 1, pp. 20–23, 2016.
- [24] Y. A. Gerhana, W. B. Zulfikar, Y. Nurrokhman, C. Slamet, and M. A. Ramdhani, "Decision support system for football player's position with tsukamoto fuzzy inference system," *MATEC Web Conf.*, vol. 197, p. 03014, 2018.
- [25] I. Septiana, M. Irfan, and A. R. Atmadja, "Sistem Pendukung Keputusan Penentu Dosen Penguji Dan Pembimbing Tugas Akhir Menggunakan Fuzzy Multiple Attribute Decision Making dengan Simple Additive Weighting (Studi Kasus: Jurusan Teknik Informatika Uin Sgd Bandung)," *J. Online Inform.*, vol. 1, no. 1, pp. 43–50, 2016.
- [26] K. V. Rajkumar, A. Yesubabu, and K. Subrahmanyam, "Fuzzy clustering and Fuzzy C-Means partition cluster analysis and validation studies on a subset of CiteScore dataset," *Int. J. Electr. Comput. Eng.*, 2019.
- [27] W. B. Zulfikar, M. Irfan, C. N. Alam, and M. Indra, "The comparison of text mining with Naive Bayes classifier, nearest neighbor, and decision tree to detect Indonesian swear words on Twitter," in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017.
- [28] J. W. Tukey, "The Future of Data Analysis," *Ann. Math. Stat.*, 1962.
- [29] D. C. Hoaglin, "John W. Tukey and Data Analysis," *Stat. Sci.*, 2004.
- [30] A. Hartigan and M. A. Wong, "A K-Means Clustering Algorithm," *J. R. Stat. Soc.*, 1979.
- [31] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *Appl. Stat.*, 2006.
- [32] C. Slamet, A. Rahman, M. A. Ramdhani, and W. Dharmalaksana, "Clustering the verses of the holy qur'an using K-means algorithm," *Asian J. Inf. Technol.*, 2016.
- [33] A. Feizollah, N. B. Anuar, R. Salleh, and F. Amalina, "Comparative study of k-means and mini batch k-means clustering algorithms in android malware detection using network traffic analysis," in *Proceedings - 2014 International Symposium on Biometrics and Security Technologies, ISBAST 2014*, 2015.
- [34] S. R. Fitriyani and H. Murfi, "The K-means with mini batch algorithm for topics detection on online news," in *2016 4th International Conference on Information and Communication Technology, ICoICT 2016*, 2016.
- [35] J. Béjar Alonso, "K-means vs Mini Batch K-means: a comparison," pp. 1–12, 2013.
- [36] D. Sculley, "Web-scale k-means clustering," 2010.
- [37] J. Santisteban and J. L. Tejada Carcamo, "Unilateral Jaccard similarity coefficient," in *CEUR Workshop Proceedings*, 2015.
- [38] S. C. Sripatha, "Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C Means Clustering," *Indian J. Comput. Sci. Eng.*, 2011.
- [39] Jumadi, D. S. A. Maylawati, B. Subaeki, and T. Ridwan, "Opinion mining on Twitter microblogging using Support Vector Machine: Public opinion about State Islamic University of Bandung," in *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*, 2016.
- [40] M. Azhari and Y. Jaya Kumar, "Improving text summarization using neuro-fuzzy approach," *J. Inf. Telecommun.*, vol. 1, no. 4, pp. 367–379, 2017.
- [41] V. Kathiresan and P. Sumathi, "An efficient clustering algorithm based on Z-Score ranking method," *Proc. 2012 Int. Conf. Comput. Commun. Informatics*, pp. 1–4, 2012.

- [42] E. H. Ahmad *et al.*, "Relationship of Work Stress to the Performance of Intensive Care Unit Nurses in Makassar," *Am. J. Public Heal. Res.*, 2018.
- [43] P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach," in *International Conference on Electrical, Electronics, and Optimization Techniques, ICEEOT 2016*, 2016.
- [44] M. Anggara, H. Sujiani, and N. Helfi, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness," *J. Sist. dan Teknol. Inf.*, 2016.
- [45] S. Aranganayagi and K. Thangavel, "Clustering categorical data using silhouette coefficient as a relocating measure," in *Proceedings - International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2007*, 2008.
- [46] S. Luan, X. Kong, B. Wang, Y. Guo, and X. You, "Silhouette coefficient based approach on cell-phone classification for unknown source images," in *IEEE International Conference on Communications*, 2012.

## Authors



**Ara Hidayat** is an associate professor at Department of Islamic Education, in Faculty of Tarbiyah and Education, UIN Sunan Gunung Djati Bandung. His current research interests focus on Education Management and Islamic Education. ORCID ID: 0000-0002-4782-1442



**Dindin Jamaluddin**, is an associate professor at Department of Islamic Education, UIN Sunan Gunung Djati Bandung. His current research interests focus on Education Quality Assurance and Islamic Education. ORCID ID: 0000-0003-2926-1434. SOPUS ID: 57205639503



**Dian Sa'adillah Maylawati**, is a lecturer in Department of Informatics at the UIN Sunan Gunung Djati Bandung, Indonesia. Her current research interests focus on Software Engineering, Expert System, Text Mining, and Natural Language Processing. She takes Ph.D degree of Information and Communication Technology in Universiti Teknikal Malaysia Melaka (UTeM), Malaysia. ORCID ID: 0000-0002-1193-3370. SCOPUS ID: 57200569961