Analysis of the Effectiveness of Online Learning Using Eda Data Science and Machine Learning

Halim Maulana[1]*, Al-Khowarizmi[2]
[1][2]Universitas Muhammadiyah Sumatera Utara, Indonesia
[1]halimmaulana@umsu.ac.id

Abstract: In educational institutions, it is important to evaluate the performance of lecturers. One aspect that can be used as a reference in assessing the performance of lecturers is the assessment in terms of teaching (learning - teaching process). Value data in the learning evaluation process can be further processed to classify or map the learning evaluation results. In this study, the authors used Data Science and Exploratory Data Analysis (EDA) techniques and applied the K-Means++ clustering algorithm using the Java programming language to identify group learning evaluation data. To determine the number of clusters to be formed, the elbow method is applied. Meanwhile, to measure the quality of each cluster formed, the silhouette coefficient method is applied. The elbow graph can visualize the Sum of Square Error (SSE) value as a way to determine the best k value in the clustering process. From the tests carried out 12 times, using 6 datasets with variations in 4 areas of assessment, 10 tests resulted in the value of k = 3, 10 tests resulted in the value of k = 4, and 1 test resulted in the value of k = 5 as the best number of clusters. In the cluster quality measurement process, 13 tests have good clustering quality, because on average they produce clusters that have a good structure with a Global Silhouette Coefficient > 0.65.

Keywords: Data Science, K-Means++,Exploratory Data Analysis, Clustering

INTRODUCTION

In a company or agency, the performance of employees is an important concern. Therefore, the process or activity of evaluating employee performance is a common thing today. Performance appraisal is used by leaders or authorities in an agency to determine the quality of employee performance, determine whether an employee is doing his job in accordance with his duties and responsibilities. This performance appraisal is widely applied in various types or fields of agencies, including educational institutions, such as universities.In the scope of higher education, performance appraisal activities can be applied to assess the performance of lecturers, staff/employees, and also students. In relation to the object of assessment in educational institutions, an assessment of the performance of lecturers is important to do. One aspect that can be used as a reference in assessing the performance of lecturers is the assessment in terms of teaching (learning - teaching process). In carrying out the assessment, students give a value for each subject taught by a particular lecturer based on predetermined criteria or statements. To be able to further assist the assessment process, it is necessary to have good data management and processing. The data processing in question is the data owned (the results of the assessment) can be processed with certain techniques or ways, so that in the end the data provide new information or knowledge. The university needs to know the results of the evaluation of the learning process, from the results of the evaluation it will be known the qualifications of the lecturer's success in the learning process(Maulana, 2016).With enough data, it is necessary to make groups which are qualifications from the results of the learning evaluation to map the results of the lecturer's performance assessment in teaching and learning activities. To group the data in question, a certain technique is needed to help or make it easier to find new information or knowledge. In this case, the science of data mining (data mining) can be applied. In data mining, to group data into certain groups, clustering techniques or methods are used. Clustering techniques or methods work by dividing or grouping data based on the similarity of characteristics of a data with other data. In the clustering technique, there are several algorithms that can be applied, one of the algorithms that is quite familiar is the K-Means algorithm. In a study conducted by Syahputra, et al

*name of corresponding author

(2018) which used the K-Means algorithm for clustering lecturer assessments based on the Student Satisfaction Index, from the results of the study it was concluded that it was necessary to increase the clustering accuracy of the K-Means algorithm. The K-Means algorithm has several variations of development, one of the results of this development is the K-Means++ algorithm. In another study conducted by Shen, et al (2020), according to the researcher, the K-Means++ algorithm was used to overcome problems in the speed and accuracy of the K-Means algorithm. To solve this problem, the K-Means++ algorithm applies a centroid spread (cluster center). Therefore, in this study the author will apply the K-Means ++ algorithm to classify the data on the results of the learning evaluation and it is hoped that this research can map the results of the learning evaluation well (Maulana, 2017).In this study, an analysis of the results of online learning conducted in the Information Systems Study Program, Faculty of Computer Science and Information Technology, UMSU will be carried out. Based on this description, the title of this study was drawn "Analysis of the Effectiveness of Online Learning Using Eda Data Science and Machine Learning" So that it is hoped that by applying this it will produce a data insight that can be used to improve the quality of the learning process at Universitas Muhammadiyah Sumatera Utara in particular the Information Systems Study Program. In this study, an analysis of the results of online learning conducted in the Information Systems Study Program, Faculty of Computer Science and Information Technology, UMSU will be carried out.

LITERATURE REVIEW

Data Mining

Data mining (data mining) is a term used to describe the discovery of knowledge in databases. Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various large databases (Dharmarajan, et al. 2019).

Knowledge Discovery in Database

The terms data mining and knowledge discovery in databases (KDD) are often used interchangeably to describe the process of extracting hidden information in a large database. The two terms have different concepts. According to Han, et al (2006) the process of knowledge discovery in databases (KDD) consists of 7 stages which are carried out sequentially, namely(Subbulaxmi, 2020) :
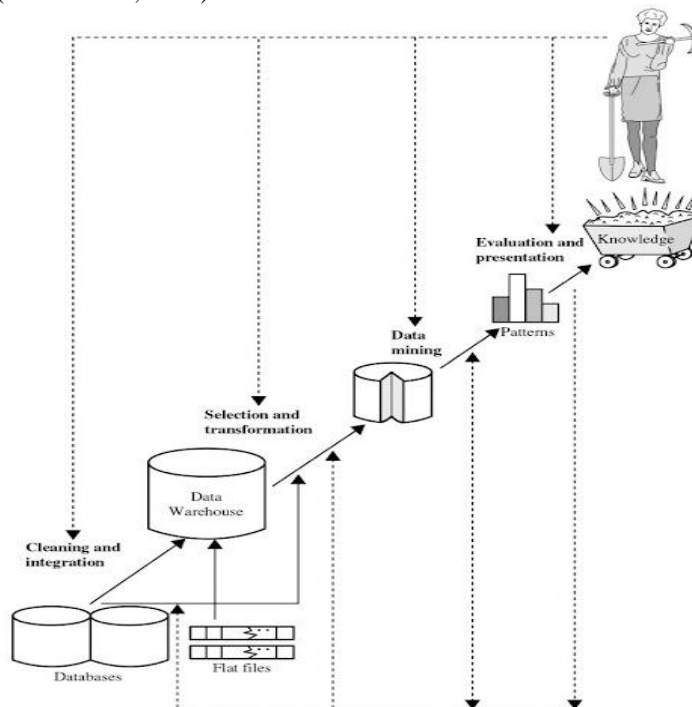

Fig 1. *Knowledge Discovery in Database*

1. Data cleaning (data cleaning)
   This process is used to remove noise and inconsistent data.
2. Data integration (data integration)
   In this process, it is possible to combine several data sources.
3. Data selection (data selection)

*name of corresponding author

Relevant data / according to the analysis task is taken from the database.
4. Data transformation (data transformation)
   The data is converted into a form suitable for the mining process by performing an aggregation operation (summary).
5. Data mining (data mining)
   In important process where intelligent methods are applied to extract data patterns.
6. Pattern evaluation
   This process identifies really interesting patterns that represent knowledge by a certain measure.
7. Knowledge presentation
   The process of presenting or exposing the knowledge of the results of the mined data to the user.

Student Learning Evaluation
Student learning evaluation is an assessment made by students on the performance of a lecturer in the teaching process. The results of this evaluation are very useful for the development of learning and the need for accreditation of study programs(Maulana, 2020). The assessment or evaluation process is carried out by considering several statements or criteria that have been set, and each statement or criterion will be rated on a scale of 1 - 7. There are six aspects of assessment in the evaluation of the learning process, as shown in table 1.

Tab1e 1 Aspek Penilaian Evaluasi Pembelajaran

No	Aspek Penilaian	Nomor Pernyataan
1	Overall field	1-2
2	Course management	3-7
3	Student engagement	8-10
4	Lecturer-student interaction	11-13
5	Content and evaluation of learning	14-17
6	Integration in life	18-20

Clustering
According to Han, et al (2006), clustering is the process of grouping data sets into several groups so that objects in one group have many similarities and have many differences with objects in other groups. The differences and similarities are usually based on the attribute values of the object and can also be in the form of distance calculations. Clustering itself is also called unsupervised classification, the clustering arena is more to be studied and paid attention to. Cluster analysis is the process of partitioning a set of data objects into subsets. Each subset is a cluster, so the objects in the cluster are similar to each other, and have differences from objects from other clusters. Partitioning is not done manually but with a clustering algorithm. Therefore, clustering is very useful and can find unknown groups in the data.

K-Means algorithm
K-Means is a non-hierarchical clustering method that tries to partition existing data into one or more clusters. This method partitions data into clusters so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into other clusters. K-Means was discovered by several people, namely Lloyd (1957, 1982), Forgey (1965), Friedman and Rubin (1967), and McQueen (1967). The idea of clustering was first discovered by Lloyd in 1957, but it was only published in 1982. In 1965, Forgey also published the same technique so that it is sometimes known as Lloyd-Forgy in some sources.
The stages of the K-Means algorithm (Han et al, 2012)(Mohammed et al, 2020):

1. Determine the number of clusters to be formed.
2. Choose an object randomly from the data set to be the center of the cluster (centroid).
3. Assign an object to a cluster whose objects are more similar (has the closest distance to the centroid).
4. Recalculate the centroid of each cluster that is formed to update the new centroid.
5. Repeat steps 3 – 5 until there is no change in cluster members (objects).

*name of corresponding author

K-Means ++ . Algorithm

The K-Means++ algorithm is a variation or development of the K-Means algorithm. The K-Means++ algorithm is considered to be able to overcome problems in the speed and accuracy of the K-Means algorithm. According to the journal K-Means++: The Advantages of Careful Seeding, the use of the K-Means algorithm can be improved in terms of accuracy by using a randomized seeding technique(Syakur et al, 2018).

The process of forming clusters in K-Means is very dependent on the value of the centroid at the beginning of the calculation. By using or applying the randomized seeding technique, it will determine the centroid value at the beginning of the calculation.

Sillhouette Coefficient

Sillhouette Coefficient is a method used to measure the quality of a cluster that combines the values of cohesion and separation. In calculating the sillhouette coefficient value, there are two components, namely ai and bi. The ai component is the average distance of the i data to all other data in one cluster, while bi is the result of calculating the average distance of the i data to all other data not in the same cluster with the i data, then the value is taken smallest. The value of ai measures how similar the data is to the cluster it follows, a smaller value indicates the more precise the data in question is in a particular cluster. A large bi value indicates how bad the data is against other clusters. The sillhouette coefficient value obtained is in the range of -1 to 1. The sillhouette coefficient value that is close to 1 indicates that the data is more precise in the cluster in question.

METHOD

This study aims to conduct a study and/or study on the analysis of online learning outcomes using Data Scicene and EDA techniques and choose K-Means++ as the algorithm for clustering. To achieve the objectives set out above, this research will use a qualitative approach using triangulation (Amei,et al (25011)) from the study of documents and related literature, data-sets, as well as observing the research object. Based on the data and information obtained then analyzed and described descriptively to get the expected conclusions.

Parameters of Measurement and Observation

This research conducts studies and/or studies on the analysis of online learning outcomes using Data Science and EDA techniques and chooses K-Means++ as the algorithm for clustering.

Research Model

The research method used is a literature study. The deepening of the concept of a proposition by collecting literature related to the method using basic/pure types of research.

Data Collection and Analysis Techniques

The data collection method used to produce a study or analysis is the library research method through books related to the software that will be built. Analyze existing data so that the information obtained provides a clear picture of the materials needed in making the basic framework for research-related design and development.

Data

In this study, the data used is learning evaluation data obtained from the E-Learning System of the Muhammadiyah University of North Sumatra. The data is the result of the 2020-2021 learning evaluation for the Faculty of Computer Science and Information Technology. The data obtained consisted of: name of lecturer, course, class, study program, faculty, number of course participants, number of evaluation respondents, average evaluation score, and assessment data for each statement/criteria given a total of 20. The total data obtained is a total of approximately 280 data.

Knowledge Discovery in Database (KDD)

The Knowledge Discovery in Database stage aims to explore important information in a database. This stage consists of several processes, namely: data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, and the last is knowledge presentation. The cleaning and data integration process is done manually with Microsoft Excel software. The next process is data selection, data transformation, data mining, pattern evaluation, and knowledge presentation carried out on the system created. In the system created, the development uses a linear sequential development model or the waterfall method. The waterfall method is a sequential software development process through the phases of planning (analysis), modeling (design), implementation, and testing.

*name of corresponding author

The explanation is as follows:
1. Planning (analysis) At this stage the analysis process will be carried out on all software requirements.
2. Modeling (design) At this stage the software design is carried out according to the needs that have been set. The modeling process focuses on data structures, software architecture, interface design, and algorithm details.
3. Implementation At this stage, the translation of the modeling process is carried out into a form that can be understood by the machine using a programming language.
4. Testing This stage aims to examine all possible errors and check whether the results of software development are as desired.

Clustering Results Analysis and Discussion

At this stage, the analysis process is carried out on the best/ideal number of clusters that can be used in the clustering process using the Elbow method. After the data mining process is carried out, the results will be evaluated and visualized so that users can easily understand the results of the data mining process based on data science and machine learning. The process of evaluating the results of clustering uses the Silhouette Coefficient method(Ding et al, 2018).

Spesification
Hardware
The computer hardware used in making the system are as follows

1.	*Proccesor*	: Intel Core i5-2410M CPU @ 2.30 GHz
2.	*RAM*	:4GB
3.	*Harddisk*	: 500 GB

Software
The computer software used in making the system is as follows:

1.	Sistem Operasi	: Windows 10
2.	*Compiler / IDE*	: Python 3.0

RESULT

Initial Processing
Data Cleaning
This process is used to remove noise and inconsistent data. At this stage, empty data (no value) in an attribute will be removed from the table..

Data Integration
The data integration process is the process of combining several data sources. The data in this study were obtained from one source, so there is no need to combine data from other sources. It's just that the author performs the process of merging data on data that allows it to be merged from the same source. Data that has the same value in the Name, Course, Study Program, and Faculty attribute, but differs in the value of the Class attribute will be combined into one.

Data Selection
At this stage, a selection of the level of filling in the evaluation of a course and data attributes is carried out. Data with a filling rate by respondents who are less than 50% and attributes that are not relevant in the process of evaluating learning outcomes will be deleted. The following are the attributes of the initial learning evaluation data listed in table 4.1, and the selected attributes of learning evaluation data listed in table 4.2, and figure 4.1 is an example of learning evaluation data.

Tabel 5.1 Initial Attributes of Learning Evaluation Data

No	Atribute	Detail
1	Nama	Lecturer name
2	Matakuliah	Course name
3	Kelas	Class name of a course

*name of corresponding author

4	Program Studi	Study program name
5	Fakultas	Architectural Name
6	Participant	Number of participants of a course
7	Respondens	Number of course participants who do an evaluation
8	Average	Average from the value of P1 to P20
9	P1	Value Statement 1
10	P2	Statement value 2
11	P3	Statement value 3
12	P4	Statement value 4
13	P5	Statement value 5
14	P6	Statement value 6
15	P7	Statement value 7
16	P8	Statement value 8
17	P9	Statement value 9
18	P10	Statement value 10

Kode	Mata Kuliah	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
001	Algoritma dan Pemrogaman	6,44	6,56	4,96	5,26	5,08	5,04	5,15	4,85	5,15	5,02	4,90	5,03	5,23	4,94	5,02	5,05	4,90	5,05	5,12	5,06
002	Kalkulus Aljabar Vektor dan Matriks	6,27	6,31	5,86	6,15	5,98	5,97	6,15	6,02	5,82	5,78	5,91	5,91	5,86	5,69	6,02	5,86	6,01	5,96	5,97	5,99
003	Komputasi Numerik dan Pemrograman	6,25	5,88	6,08	6,12	6,07	6,00	6,01	6,00	6,02	5,88	5,72	5,72	5,99	5,98	6,01	6,09	6,00	5,96	5,97	5,99
5,74004	Struktur Data	5,80	5,64	6,33	6,41	6,41	5,91	6,44	6,19	6,24	6,22	5,78	5,97	6,36	5,92	6,17	6,19	5,83	6,24	6,00	6,39
005	Kriptografi dan Sekuriti Sistem	4,45	4,00	6,28	6,35	6,38	6,07	6,31	5,99	6,19	6,36	6,15	6,04	5,95	5,95	6,36	5,95	6,23	5,93	6,26	6,14
006	Prak. Struktur Data	3,83	3,56	5,25	5,50	5,47	5,31	5,50	5,31	5,36	5,42	5,03	5,22	5,42	5,25	5,31	5,44	5,39	5,19	5,36	5,53

Data Transformation
At this stage, the data is converted into a form suitable for the mining process, i.e. the data will be converted into an array, so that the data is ready for the clustering process in the software that was built..

Evaluasi Hasil Pengujian
The results of software testing can be concluded that the software has been running according to the design that has been made. All functions can run well as expected.

Comparison of Test Results With Manual Counting Results
To test or validate the results of the calculation of the algorithm, a test is carried out by comparing the results of manual calculations with the results of the software. The test was carried out using 15 data of Learning Evaluation data. The data used is listed in the table below.

Tabel 5.3 Test Dataset

ID	Subjects	P1	P2
001	Algoritma dan Pemrogaman	5,22	5,11
002	Kalkulus Aljabar Vektor dan Matriks	6,43	6,33
003	Komputasi Numerik dan Pemrograman	4,84	4,38
004	Struktur Data	6,33	6,40
005	Kriptografi dan Sekuriti Sistem	3,25	3,33
006	Prak. Struktur Data	6,40	6,40

Evaluation of Manual Counting Results Testing and Software Results
The test results show that the clustering process in manual calculations and software has the same results, so that it can ensure that the software built is as expected.

Testing Using Datasets
Analysis of Test Results based on Elbow Method
Testing with the elbow method uses a value of k = 2 to k = 15, and uses six types of learning evaluation assessment areas. The following are the results of testing the elbow method on software using four datasets:
1. Odd Semester Learning Evaluation Data in 2021 Based on the results of testing the elbow method on evaluation data for odd semester learning in 2021 with a total of 86 data, it was found that in the areas of: Overall Field, Student Involvement, Lecturer-Student Interaction, and Content and Learning Evaluation had a value The best k = 3. While in the field of Course Management and Integration in Life, the best k value = 4. Table 5.4 below is the SSE value data in the odd semester learning evaluation data in 2021.
2. Learning Evaluation Data for Even Semester 2021 Based on the results of testing the elbow method on evaluation data for even semester 2021 learning with a total of 54 data, it was found that in the field of Integration in Life has the best k value = 3. In the fields of Overall Field, Course Management, Engagement Students, Student Lecturer Interaction, have the best k value = 4. While in the field of Content and Learning Evaluation the best k value = 5. Table 5.5 below is the data on SSE scores in the 2021 even semester learning evaluation data.
3.

Tabel 5.4 SSE Value of Odd Semester Learning Evaluation Data in 2021

	SSE					
	Bidang 1	Bidang 2	Bidang 3	Bidang 4	Bidang 5	Bidang 6
1	37,267	80,6257	59,1602	53,8385	77,0929	49,6610
2	12,774	33,2154	23,9451	24,7516	30,2677	20,0242
3	8,0498	23,3161	12,7799	15,7657	15,8026	12,9226
4	5,8776	15,1409	9,8526	12,2913	12,1351	9,0543
5	4,5772	11,5709	7,0966	11,1009	10,8350	7,45888
6	2,90329	9,74829	5,61957	9,08352	8,66040	5,32581
7	2,02025	8,76729	4,61129	7,66600	8,32102	4,23437
8	1,77186	8,48521	3,91734	7,26813	7,86489	3,97844
9	1,47461	7,75998	3,42172	6,82580	6,75018	2,89896
10	1,44537	6,97185	3,28463	6,36682	6,86224	2,70113
11	1,37255	6,75399	3,03411	4,73659	6,14908	2,59964
12	1,15726	6,46997	2,76367	4,48757	5,80296	2,37034
13	0,94165	6,30598	2,46487	3,82253	5,61192	2,25269
14	0,87727	4,93028	2,34421	3,44508	4,23030	2,04871
15	0,76264	4,62051	2,10171	3,21725	3,88802	1,97000

*name of corresponding author

Figures 5.9 to 5.14 below show an elbow graph for evaluation data for odd semester learning in 2021. In Figure 5.9 there is a significant decrease in SSE scores when k = 1 to k = 2, then followed by a decrease again when k = 2 to k = From k = 3 to k = 4 and so on, the graph begins to show the stability of the SSE value or does not experience a significant decrease in the SSE value.

DISCUSSIONS

Cluster Result Analysis
1. Based on the best k value from the elbow method testing process for four different datasets, an analysis of the clusters formed in the clustering process was carried out to find the characteristics of each cluster.
2. Odd Semester Learning Evaluation Data in 2021. In the tests conducted on the 2016 odd semester learning evaluation data, it was found that the resulting clusters on average had a cluster structure with a weak category, which also had an impact on the global silhouette coefficient value for each clustering process in each – each type of assessment field. In other words, the clustering process carried out for the 2016 odd semester learning evaluation data has a not very good quality, because it has a global silhouette coefficient value which is classified in the weak category. The characteristic data for each cluster of test results using the 2016 odd semester learning evaluation data are listed in tables 5.9 to 5.14 below.

Tabel 5.9 Characteristics Odd semester learning evaluation data in 2021 Overall Field

Cluster	Rata - rata		Anggota	Silhouette Coefficient	Kategori Struktur
	P1	P2			
1	5,95	5,97	43	0,62213	Good
2	4,84	4,5	7	0,23399	Bad
3	5,42	5,23	36	0,48872	Weak
Silhouette Coefficient Global				0,44828	Lemah

Tabel 5.10 Characteristics Odd semester learning evaluation data in 2021 Overall Field

Cluster	Rata – rata					Anggota	Silhouette Coeffiient	Kategori Struktur
	P3	P4	P5	P6	P7			
1	5,92	5,93	5,89	5,93	5,86	36	0,43639	Weak
2	4,12	4	4,25	4,25	4,17	2	0,12037	Bad
3	4,99	5,25	5,14	4,99	5,06	20	0,42981	Weak
4	5,41	5,55	5,51	5,38	5,48	28	0,36974	Weak
Silhouette Coefficient Global							0,33908	Lemah

Tabel 5.11 Clusters Data evaluation of odd semester 2016 learning – Student active involvement

Cluster	Rata - rata			Anggota	Silhouette Coefficient	Kategori Struktur
	P8	P9	P10			
1	5,44	5,56	5,46	43	0,48325	Lemah
2	4,71	5,05	4,79	21	0,43446	Lemah
3	6,04	6,12	5,96	22	0,5089	Baik
Silhouette Coefficient Global					0,47554	Lemah

*name of corresponding author

Tabel 5.12 Cluster Characteristics of Odd Semester Learning Evaluation Data in 2021 – Lecturer and Student Interaction

Cluster	Rata - rata			Member	Silhouette Coefficient	Category Structure
	P11	P12	P13			
1	5,8	5,96	6,03	22	0,3711	Lemah
2	4,74	4,79	5,11	23	0,40776	Lemah
3	5,28	5,46	5,57	41	0,40826	Lemah
Silhouette Coefficient Global					0,39571	Lemah

Tabel 5.13 Cluster Characteristics of the 2016 odd semester learning evaluation data – Content and Learning Evaluation

Cluster	Rata - rata				Member	Silhouette Coefficient	Category Structure
	P14	P15	P16	P17			
1	5,25	5,37	5,35	5,31	32	0,49113	Lemah
2	4,58	4,83	4,75	4,66	18	0,49182	Lemah
3	5,79	5,82	5,84	5,82	36	0,46093	Lemah
Silhouette Coefficient Global						0,48129	Lemah

CONCLUSION

Based on the results and discussion above, the authors can draw the following conclusions:

1. The elbow graph can visualize the Sum of Square Error (SSE) value as a way to determine the best k value in the clustering process.
2. From the tests carried out 24 times, using 4 datasets with a variation of 6 areas of assessment, 13 tests resulted in the value of k = 3, 10 tests resulted in the value of k = 4, and 1 test resulted in the value of k = 5 as the best number of clusters. In the cluster quality measurement process, 13 tests have good clustering quality, because on average they produce clusters with good structures (Global Silhouette Coefficient > 0.50).

REFERENCES

A. Dharmarajan and T. Velmurugan, "Performance analysis on K-means and fuzzy C-means clustering algorithms using CT-DICOM images of lung cancer," J. Adv. Res. Dyn. Control Syst., 2019 https://doi.org/10.5373/JARDCS/V11/20192597.

Amei, W., Huailin, D., Qingfeng, W., & Ling, L. (2011). A survey of application-level protocol identification based on machine learning. 2011 International Conference on Information Management, Innovation Management and Industrial Engineering, 3, 201–204.

*name of corresponding author

D. Ding, J. Li, H. Wang, and Z. Liang, "Student Behavior Clustering Method Based on Campus Big Data," in Proceedings - 13th International Conference on Computational Intelligence and Security, CIS 2017, Feb. 2018, vol. 2018-January, pp. 500–503, doi: 10.1109/CIS.2017.00116.

Han, Jung Yeon, et al. "Expression and RNA interference-induced silencing of the dammarenediol synthase gene in Panax ginseng." Plant and cell physiology 47.12 (2006): 1653-1662.

Han, Jiawei & Kamber, Micheline & Pei, Jian. (2012). Data Mining: Concepts and Techniques. 10.1016/C2009-0-61819-5.

Maulana, H. (2016). Analisis Dan Perancangan Sistem Replikasi Database Mysql Dengan Menggunakan Vmware Pada Sistem Operasi open source. InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan), 1(1), 32–37.

Maulana H. and Syahputra E. R. 2017 Analysis of Multiple Data Hiding Combined Coloured Visual Cryptography and LSB Journal of Physics: Conference Series 930

Maulana, H., & Al-Khowarizmi, A.-K. (2020). Analysis and Design of Village Management Information Systems (VMIS) based on MVC and E-Government in Indonesia. Journal of Computer Science, Information Technology and Telecommunication Engineering, 1(2), 63–69.

M.A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," in IOP Conference Series: Materials Science and Engineering, 2018 https://doi.org/10.1088/1757-899X/336/1/012017.

M.Mohammed, M. B. Khan, and E. B. M. Bashier, "k-Means Clustering," in Machine Learning, 2020.

Shen W., Yang C. and Gao L. Address business crisis caused by COVID-19 with collaborative intelligent manufacturing technologies IET Collaborative Intelligent Manufacturing 2 96-99 6 2020

Engineering Technology 7 58-61

S.S Subbulaxmi and G. Arumugam, "K-Means Cluster-Based Oversampling Algorithm for Imbalanced Data Classification," no. 5, pp. 3436–3440, 2020 https://doi.org/10.35940/ijrte.E6535.018520.

Syahputra E. R., Sembiring B. O., Dewi A. R., Hasdiana H. and Maulana H. 2018 Ads block management system using open Virtual Private Network on Ubuntu operating system International Journal of

*name of corresponding author