International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

The Application of K-Means Clustering Algorithm for Initial Analysis of Students Online Learning

Johanes Eka Priyatma[1], Haris Sriwindono[2], Paulina H. Prima Rosa[3], Agnes Maria Polina[4*]

[1,2,3,4] *Informatics Department, Sanata Dharma University, Yogyakarta, Indonesia*

Date of Submission: 25-06-2024 Date of Acceptance: 05-07-2024

ABSTRACT: Despite its sudden onset due to the Covid-19 pandemic, the two years of online learning provided invaluable experiences. The success of students in online learning is not solely determined by their access to and proficiency with various information technology tools, such as Learning Management Systems, video communication platforms, and social media. Online learning necessitates significant changes in attitudes compared to traditional classroom settings. This paper employs the K-Means clustering algorithm on data from 7,000 students to explore the various factors influencing differences in student comprehension and attitudes towards online learning. Initial analysis using the Silhouette Score reveals two distinct clusters of students in terms of their understanding and attitudes. Further box plot analysis of each data attribute identifies the key factors contributing to these two clusters.

KEYWORDS: Covid-19,K-Means clustering, online learning.

I. INTRODUCTION

Online learning has become a necessity due to the COVID-19 pandemic, which necessitates indirect interaction and physical distancing. Online learning, however, is not a new concept; it has long been utilized for remote teaching to reach students who live far from educational facilities, effectively increasing access to education. Despite its benefits, the sudden transition from conventional to online learning presents challenges. Students may struggle to adapt quickly, as online learning significantly differs from face-to-face instruction, which allows for direct physical and psychological contact between lecturers and students, as well as among peers (Lalima &Dangwal, 2017).

Student readiness for online learning is influenced by several factors, including computer/internet self-efficacy, motivation, online communication self-efficacy, learner control, and self-directed learning (Hung, 2010). Successful online learning is characterized by: (1) students who are capable of independent learning and highly motivated; (2) lecturers proficient in online teaching technologies; (3) learning strategies that promote interaction among students, lecturers, and content; (4) clear and straightforward learning content; (5) short-duration video media; and (6) institutional support through online libraries, learning management systems (LMS), and lecturer training. Attention must be given to enhancing students' capacity for self-directed learning (Mulyatiningsih et al., 2020).

This initial study analyzed data from 7,000 students, examining attributes such as Learning Model, Self-Evaluation, Learning Ecosystem, Intrinsic Motivation, Social Factor, Planning, Input, Objective, Organizing, Evaluating, Self-Directed Learning, and Readiness. Using a data mining approach, specifically the K-Means clustering method, the study provides an initial analysis of students' attitudes and understanding of their sudden transition to online learning over the last three year.

The results of this study can be used by all parties involved in the development of education so that online learning can provide its optimal quality in term of its learning outcomes. Because online learning is believed to be the main mode of learning in the near future for it is in line with progress and the availability of internet access for the community, the contextual development of online learning will be very important and useful for everyone.

In particular, this study is useful for mapping the characteristics of attitudes and understanding of students undergoing online

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

learning so that appropriate programs can be developed to prepare them to enter a new era of modern education that will rely heavily on online interaction.

II. LITERATURES REVIEW

There are numerous definitions of online learning, as reviewed by Moore et al. (2010). Most of these definitions emphasize various attributes of digital technology as fundamental to the essence of online learning. However, Tavangarian, Leypold, Nölting, Röser, and Voigt (2004) and Triacca, Bolchini, Botturi, and Inversini (2004) argue that defining online learning solely through technological attributes is insufficient. From a constructivist perspective, Tavangarian et al. (2004) assert that online learning involves more than procedural aspects; it necessitates the transformation of individual experiences to facilitate the reconstruction of personal knowledge. Furthermore, Ellis (2004) and Triacca et al. (2004) contend that an effective definition of online learning must encompass a certain level of interaction to adequately describe the learning experience.

Expanding on this framework, numerous authors describe online learning as access to experiential learning enabled by technology (Benson, 2002; Carliner, 2004; Conrad, 2002). Benson (2002) and Conrad (2002) characterize online learning as the pinnacle of distance education, significantly improving the accessibility of educational opportunities. Other researchers highlight further aspects, such as connectivity, flexibility, and the ability to create various models of learning interactions (Hiltz & Turoff, 2005; Oblinger & Oblinger, 2005).

The online learning model has garnered extensive review from various stakeholders, not only for its capacity to address the challenges posed by the COVID-19 pandemic but also for its numerous advantages. Key benefits include enhanced learning effectiveness, its applicability in professional development, cost efficiency, facilitation of credit transfers between universities, and the substantial potential to provide world-class education to anyone with sufficient internet access (Koller & Ng, 2014; Lorenzetti, 2013).

The online learning model has been critically reviewed by numerous stakeholders, not only for its ability to address the challenges of the COVID-19 pandemic but also for its numerous benefits. Key advantages include enhanced learning effectiveness, utility in professional development, cost efficiency, facilitation of credit transfers across universities, and the significant potential to offer world-class education to anyone with adequate internet access (Koller & Ng, 2014; Lorenzetti, 2013).

Over the past two decades, data mining techniques have become increasingly popular in research methodologies, applied across various studies. One area where data mining has been extensively utilized is education. With the rise of online learning during the COVID-19 pandemic, the availability of learning data has increased, allowing learning data mining techniques to significantly contribute to our understanding of contemporary education (Baker &Inventado, 2014).

Despite this, the application of data mining techniques to examine the diversity of student understandings and responses to the sudden shift to online learning remains limited. This inquiry is crucial, as online learning is not merely the application of technology to distance education but also involves providing meaningful learning experiences as defined above.

In this study, researchers aimed to explore differences in student perceptions and attitudes towards online learning, focusing on three main factors:

1. **Paradigm level**: Investigating students' conceptual understanding and mental attitudes towards online learning.
2. **Managerial level**: Examining students' self-management experiences in an online learning environment.
3. **Technical level**: Assessing students' ability to manage and address technical aspects that support online learning.

III. METHODOLOGY

Building on the definition of online learning as a means of gaining access to learning for the constructive development of knowledge and skills, as proposed by Tavangarian et al. (2004), this study collected data across three levels from students' online learning experiences over the course of a year. These levels are paradigm, managerial, and technical.

At the paradigm level, this research aims to explore students' experiences with online learning from their conceptual standpoints and mental attitudes. Data were collected on students' enthusiasm for learning, learning attitudes, and their understanding of the different interaction models between online and offline learning.

At the managerial level, this study investigates students' experiences in self-management within the online learning environment. Key data points include the benefits, conveniences, and adaptive capacities of students

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

participating in online learning through the Learning Management System (LMS) available at http://belajar.usd.ac.id. Additionally, the study examines how well LMS management meets student expectations for optimal learning outcomes.

At the technical level, the research focuses on students' experiences in managing technical aspects that support online learning, such as computer infrastructure, internet connectivity, and their readiness to engage in online interaction models.

While each level addresses various attributes that support the success of online learning as mentioned in the introduction, the specific contributions of these constructs were not the primary focus of this study. Based on these three levels, the researcher developed 27 statements: 8 for the paradigm level, 15 for the managerial level, and 4 for the technical level. Each statement had four possible responses: Strongly Agree (SA/4), Agree (A/3), Disagree (DA/2), and Strongly Disagree (SDA/1). The four response options were designed to simplify the answering process for respondents. A complete list of these statements is provided in Appendix I.

These 27 statements were presented digitally to students at Yogyakarta Sanata Dharma University in Indonesia through the Academic Information System athttps://mahasiswa.usd.ac.id. Active students were free to respond to or ignore these statements each time they accessed the SIA. Ultimately, 7,000 responses were collected from students across all study programs and academic years at Sanata Dharma University.

The collected data were supplemented with supporting demographic information, including gender, study program and faculty, class, and region of origin, to provide adequate context for data interpretation. After ensuring the completeness of the responses, data analysis was conducted using data mining techniques. The analysis aimed to identify the number of clusters that best reflect students' responses to their online learning experiences from January to December 2020.

In general, the data mining analysis follows the flow as presented in Figure 1. Using this flow, the data mining analysis that is carried out is clustering analysis. Using clustering analysis, this study will map out how many clusters of student experiences and attitudes are formed while undergoing online learning.

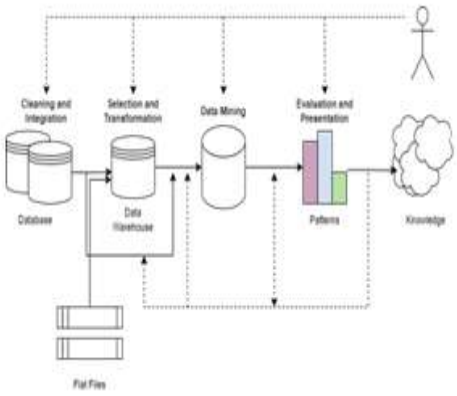

Figure 1. Clusterization Analysis Data Mining Flow

IV. FINDING AND DISCUSSION

For the dataset containing the answers of 7000 respondents to 27 questions with answers in the form of a scale of 1 to 4, clustering was carried out using the K-Means clustering algorithm. Experiments were carried out by comparing the results of clustering with different numbers of clusters ranging from 1 to 10. The experimental results are shown in Figure 2 below.

```
{2: 0.2001126395647186,
 3: 0.18008543661972695,
 4: 0.14777423156163272,
 5: 0.1218763304056742,
 6: 0.116218653655128,
 7: 0.11781247283039664,
 8: 0.11417235808242375,
 9: 0.10245499643494534}
```

Figure 2. Sillhoutte Coefficient for Various Numbers of Clusters

The experiments revealed that the optimal number of clusters, as indicated by the highest Silhouette Coefficient (SC), is 2. Consequently, the dataset is best grouped into two clusters, indicating that the survey responses naturally divide into two distinct groups. The proportion of members in each group is illustrated in the following Figure 3.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

Figure 3. Proportion of Number of Cluster Members

Based on the Silhouette Coefficient (SC) value of less than 0.25, the structure of the formed groups does not exhibit strong cohesion, falling more into the category of lacking distinct structure (Kaufmann &Rousseeauw, 1990). This is primarily due to the narrow range of values in the answers to each question (ranging from 1 to 5), where responses of 2 and 3 predominate, leading to significant overlap between the groups formed. To gain a more detailed understanding of the questions that differentiate the two groups, a box plot analysis was conducted for each question. Table 1 below summarizes the minimum, maximum, and median values for each response. Meanwhile, the ten questions with different medians are presented in Table 2 below.

Table 1. The minimum, maximum, median values for each question

	C1				C2			
	min	max	median	mean	min	max	median	mean
j1	1	4	3	2.885366	1	4	3	3.19201
j2	1	4	3	2.625784	1	4	3	3.153027
j3	1	4	3	2.603136	1	4	3	3.217191
j4	1	4	3	2.514286	1	4	3	3.121308
j5	1	4	2	2.238676	1	4	3	3.023487
j6	1	4	3	2.590592	1	4	3	3.052058
j7	1	4	3	2.718467	1	4	3	3.125666
j8	1	4	4	3.490244	1	4	3	3.360048
j9	1	4	3	2.872822	1	4	3	3.269734
j10	1	4	3	2.572822	1	4	3	3.132688
j11	1	4	3	3.039024	1	4	3	3.339709
j12	1	4	3	2.609756	1	4	3	3.178935
j13	1	4	3	2.612195	1	4	3	3.171429
j14	1	4	3	2.852613	1	4	3	3.221792
j15	1	4	2	2.344948	1	4	3	3.067312
j16	1	4	3	2.665157	1	4	3	3.162712
j17	1	4	3	2.569338	1	4	3	3.145036
j18	1	4	3	2.81324	1	4	3	3.21138
j19	1	4	3	2.55993	1	4	3	3.061743
j20	1	4	2	2.268641	1	4	3	3.074092
j21	1	4	2	1.884669	1	4	3	2.918886
j22	1	4	2	2.194077	1	4	3	3.005811
j23	1	4	2	1.91115	1	4	3	2.805811
j24	1	4	2	2.266551	1	4	3	2.937772
j25	1	4	2	2.256446	1	4	3	2.93632
j26	1	4	2	2.274216	1	4	3	3.004358
j27	1	4	3	2.660976	1	4	3	3.137772

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

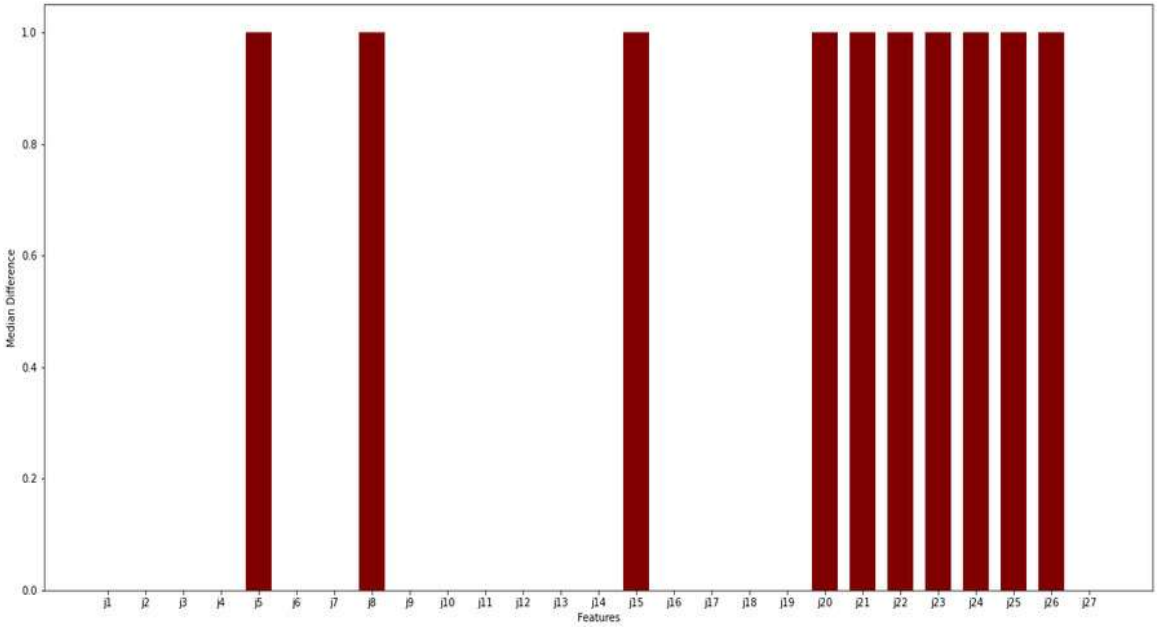

Figure 4. Median Difference for Each Question in the Two Clusters

Table 2. List of Questions with Median Difference 1

Question No.	Question/Statement	Median difference	Question Category
5	I enjoy studying off campus during online learning	1	Learning Ecosystem
8	I feel I will be more excited if I can physically interact with lecturers and college friends	1	Social Aspect
15	I can communicate and consult easily with lecturers during online lectures	1	Organizing
20	I can manage my study activities better when I study online.	1	Self-Directed
21	Online lectures make me more enthusiastic/enthusiastic about learning.	1	Self-Directed
22	Online lectures make me more free to express opinions and ask lecturers	1	Self-Directed
23	I didn't encounter any significant problems when I had to discuss or work in groups with friends.	1	Organizing
24	I have no significant problems in operating the LMS to follow and complete lecture assignments.	1	Readiness
25	During online learning I could access references easily.	1	Input
26	Online learning made me more confident and actives in discussion.	1	Self Directed

Upon comparing the medians of both groups for each question, it was observed that 17 questions from the questionnaire exhibited identical median values of 3 across both clusters, while 10 questions showed a median difference of 1, as depicted in Figure 4 above.

Examining the results of the data analysis and the clusters formed yields several conclusions. The low SC value for the 2 clusters indicates a lack of distinct structure, suggesting that the dataset may effectively represent a single cluster. This implies a general homogeneity in students' attitudes and perceptions towards online learning. Such uniformity facilitates the formulation of targeted policies concerning online learning infrastructure,

methodologies, student engagement, and faculty development.

Further exploration of the 10 questions that exhibited differing responses between the clusters reveals insights into specific areas. Among the paradigmatic questions, six out of eight categories showed consistent median values. For instance, in question number 5—pertaining to students' enjoyment of off-campus study during online learning—Cluster 1 had a median of 2, whereas Cluster 2 had a median of 3. This discrepancy indicates varying levels of enjoyment among students in engaging in off-campus learning. Similarly, in question number 8, which explores enthusiasm for physical interaction with lecturers and peers, Cluster 1 demonstrated a median of 4, compared to Cluster 2's median of 3, indicating differing levels of enthusiasm.

In the learning management and LMS categories, 10 out of 15 questions received a median value of 3, indicating widespread satisfaction with the provided LMS. However, for the remaining five technical questions (numbers 15 and 20 to 26), Cluster 1 had a median of 2, while Cluster 2 had a median of 3. These technical questions typically relate to challenges with learning and the LMS. These variations align with the findings from the paradigmatic questions, where respondents in Cluster 1 faced more technical issues and exhibited lower satisfaction with online learning compared to Cluster 2.

In summary, while respondents' answers were generally homogeneous, differences in responses across 10 questions delineate two clusters. Cluster 2 generally encountered fewer technical issues and exhibited higher satisfaction with online learning compared to Cluster 1, which faced technical challenges. Overcoming these technical barriers could potentially align responses in Cluster 1 with those in Cluster 2, highlighting the potential benefits of online learning over traditional offline methods.

Future research could delve into demographic characteristics such as geographic origin, gender, and socio-economic factors to further delineate each cluster. Additionally, exploring correlations between responses to different questions could yield deeper insights from the questionnaire data.

V. CONCLUSION

In line with the increasing availability of digital learning data as a result of the adoption of online learning during the Covid-19 pandemic, data mining analysis techniques are increasingly finding relevance in the world of education. This study has provided a form of using the clustering model in learning data mining. Based on the results of the clustering carried out on quite a lot of data, namely 7000 students, it was found that in general Sanata Dharma University students have a homogeneous attitude and understanding in undergoing online learning for a year. From the analysis of the median differences of the 2 clusters formed, it can also be concluded that students' attitudes and understanding tend to be positive towards online learning. This is an interesting and useful finding because even though online learning takes place suddenly with limited preparations, it turns out that students as a generation who are accustomed to online life do not experience serious resistance.

From an analysis of the contribution of attitudes and understandings that contribute significantly to the formation of clusters, it can be concluded that even though students have a positive attitude towards online learning, they still want to enjoy direct physical interaction which will of course further enrich their learning experience. However, in terms of learning management and LMS, some students have several problems that are not at the paradigmatic level related to online learning but tend to be related to technical matters regarding learning and LMS. If these technical problems can be overcome, there is a possibility that respondents will no longer be constrained by technical matters in online learning. Based on these interesting findings, an initial study of online learning using data mining using clustering analysis can provide relevant and useful results. This kind of study will certainly have stronger results and benefits if the available data sets come from more than one institution.

Acknowledgment

Thanks to Vitus Dama Jivanov as an assistant in this research.

REFERENCES

[1]. Baker, R. S., &Inventado, P. S. (2014). Educational data mining and learning analytics. In R. S. Baker & P. S. Inventado (Eds.), Learning analytics (pp. 61–75). New York: Springer.

[2]. Benson, A. (2002). Using online learning to meet workforce demand: A case study of stakeholder influence. Quarterly Review of Distance Education, 3(4), 443−452.

[3]. Bowen, W. G. (2013). Higher education in the digital age. Princeton University Press.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

[4]. Carliner, S. (2004). An overview of online learning (2nd ed.). Armherst, MA: Huma Resource Development Press.
[5]. E. Mulyatiningsih, K.Komariah, B. Lastariwati, M.G. Kartika, R. Restiana. (2020). EKSPLORASI FAKTOR-FAKTOR PENENTU KEBERHASILAN PEMBELAJARAN DARING DI ERA REVOLUSI INDUSTRI 4.0. LaporanPenelitian. https://simppm.lppm.uny.ac.id/uploads/8729/laporan_akhir/laporan-akhir-8729-20201217-193903.pdf
[6]. Ellis, R. (2004). Down with boring e-learning! Interview with e-learning guru Dr.Michael W. Allen. Learning circuits. Retrieved from. http://www.astd.org/LC/2004/0704_allen.htm.
[7]. Fisher, D. (2012, November 6). Warming Up to MOOC's. The Chronicle of Higher Education Blogs: ProfHacker.
[8]. Hiltz, S. R., & Turoff, M. (2005). Education goes digital: The evolution of online learning and the revolution in higher education. Communications of the ACM, 48(10), 59−64, doi:10.1145/1089107.1089139.
[9]. Hung, M.-L., Chou, C., Chen, C.-H., & Own, Z.-Y. (2010). Learner readiness for online learning: Scale development and student perceptions. Computers & Education, 55(3), 1080–1090. doi:10.1016/j.compedu.2010.05.004
[10]. Lalima., &Dangwal, K. L. (2017). Blended learning: An innovative approach. Universal Journal of Educational Research, 5(1), 129-136. doi:10.13189/ujer.2017.050116
[11]. Lewin, T. (2012, July 18). Anant Agarwal Discusses Free Online Courses Offered by a Harvard/M.I.T.
[12]. Lorenzetti, J. (2013.). Academic Administration - Running a MOOC: Secrets of the World's Largest Distance Education Classes - Magna Publications.
[13]. Moore, J.L., et al., e-Learning, online learning, and distance learning environments: Are they the same?, Internet and Higher Education (2010), doi:10.1016/j.iheduc.2010.10.001
[14]. Oblinger, D. G., & Oblinger, J. L. (2005). Educating the net generation. EDUCAUSE. Retrieved from. http://net.educause.edu/ir/library/pdf/pub7101.pdf.
[15]. Partnership. The New York Times. Retrieved from http://www.nytimes.com/2012/07/20/education/edlife/anant-agarwal-discusses-free-onlinecourses-offered-by-a-harvard-mit-partnership.html
[16]. Tavangarian, D., Leypold, M. E., Nölting, K., Röser, M., & Voigt, D. (2004). Is e-Learning the solution for individual learning? Electronic Journal of e-Learning, 2(2), 273−280.
[17]. Triacca, L., Bolchini, D., Botturi, L., &Inversini, A. (2004). Mile: Systematic usability evaluation for e-Learning web applications. AACE Journal, 12(4).

Appendix 1.
ONLINE LEARNING QUESTIONNAIRE
1. Identity
a. Student ID :
b. Study program :
c. Active cellphone :
d. Location of online learning: (village/district, city/regency, big city)
2. Questionnaire Give a tick (√) according to your opinion / experience.
SA : Strongly Agree A : Agreed DA : Disagree SDA : Strongly Disagree

#	Code	PernyataanUntukMahasiswa	Answer			
			SA	A	DA	SDA
Paradigm						
1.	LM	I feel the lecturer uses a method that demands my activeness				
2.	SE	I feel that I can be more independent in studying lecture material online				
3.	SE	I am still enthusiastic about attending lectures and diligently doing assignments from lecturers even though it is online.				
4.	SE	Online learning helps me better understand my				

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

		strengths and weaknesses in learning				
5.	LE	I enjoy studying off campus during online learning				
6.	IM	I provide independent study time outside of online lecture schedules.				
7.	MI	I explore new knowledge according to my interests during online learning from home				
8.	SA	I feel more excited when I can interact physically with lecturers and fellow students				
Managerial						
9.	PL	The LMS for the course I am taking is equipped with learning objectives, learning materials, and instructions along with clear learning activities.				
0.	IN	The learning materials presented in the LMS are easy to understand and help me master a topic.				
1.	IN	The learning materials presented in the LMS that I participated in varied types (text, ppt, video, etc.).				
2.	TJ	The learning material presented in the LMS makes it easier for me to study independently.				
3.	TJ	The learning activities presented in the LMS make it easier for me to study independently in order to achieve learning goals.				
4.	OR	The learning activities that I participate in vary in form and type.				
5.	OR	I can communicate and consult easily with lecturers during online lectures				
6.	OR	The features used in learning through the LMS make it easier for me to learn.				
7.	EV	Quizzes, tests, and assignments given by the lecturer are sufficient and relevant and help me study well.				
8.	EV	The grades I get from UTS, Quiz, UAS, or other assignments are in accordance with the effort I put in				
9.	EV	I get work results, tests, or feedback from lecturers within a reasonable timeframe.				
0.	SD	I can better organize my learning activities during online lectures.				
1.	SD	Online lectures make me more enthusiastic/enthusiastic in learning.				
2.	SD	Online lectures make me more enthusiastic/enthusiastic in learning.				
3.	OR	Online lectures allow me more freedom to express opinions and ask questions of lecturers				
Technical						
4.	RD	I have no significant problems in operating the LMS to follow and complete lecture assignments.				
5.	IN	During online learning I can easily access reference books.				
6.	SD	Online learning makes me more confident and active in discussions.				
7.	RD	The university provided adequate training for me so that I could optimize the LMS provided.				

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 07 July 2024, pp: 67-75 www.ijaem.net ISSN: 2395-5252

Coding:

1.	**LM**	**: Learning Model**
2.	**SE**	**: Self Evaluation**
3.	**LE**	**: Learning Ecosystem**
4.	**IM**	**: Internal Motivation**
5.	**SA**	**: Social Aspect**
6.	**PL**	**: Planning**
7.	**IN**	**: Input**
8.	**TJ**	**: Tujuan**
9.	**OR**	**: Organizing**
10.	**EV**	**: Evaluating**
11.	**SD**	**: Self Directed**
12.	**RD**	**: Readiness**