CLUSTERIZATION OF LECTURER’S PROFILE IN ONLINE LEARNING DURING THE COVID-19 PANDEMIC

Agnes Maria Polina¹, Christiyanti Aprinastuti², and Hari Suparwito³
¹,²,³ Sanata Dharma University, Indonesia
¹a.m.polina@usd.ac.id, ²christiyantia@usd.ac.id, and ³shirsj@jesuits.net
*correspondence: a.m.polina@usd.ac.id
https://doi.org/10.24071/ijiet.v7i2.6495
received 3 June 2023; accepted 10 July 2023

Abstract
The learning process changed from classroom to online learning during the COVID-19 pandemic. One of the things that must be done is to analyze the readiness of lecturers in facing online learning. The purpose of this study is to cluster the profiles of lecturers dealing with online learning. The clustering method uses a Machine Learning approach with the K-means algorithm. Data were taken from 274 lecturers who returned questionnaires during April–June 2022. The questionnaire consisted of 27 questions on a Likert scale (1–4). The Boruta technique is used to determine the five most significant variables (Variable Importance) in the clustering. The results of the clustering show that the lecturers are divided into 2 large groups with the following criteria: focus on learning methods, learning materials, student independence, exploration of new knowledge, and online learning evaluation tools.

Keywords: Boruta, clustering, K-means, online learning, variable importance

Introduction
The COVID-19 pandemic hit and almost paralyzed all countries in almost all aspects of life (Vo & Tran, 2021). To control the spread of COVID-19, the Indonesian government has imposed warnings and prohibitions on leaving the house, working, and going to school. Several new terms have emerged, including “working from home” and “studying from home”. In the field of education, the term “e-learning” or “online learning” became well-known to the wider community during the pandemic. With e-learning or online learning, the learning process utilizes information technology in teaching and learning, where learning is managed using an electronic or computer system to support the learning process. It is hoped that the teaching and learning process will not be interrupted due to the existence of a lockdown regulation to keep the COVID-19 virus from spreading more widely (Maatuk et al. 2022).

A shift in the learning paradigm is the challenge that stakeholders in education face. Normally, learning takes place in the classroom, but during a pandemic, learning shifts to online learning using resources supported by information technology and the internet (Turnbull et al., 2021). In Indonesia, this challenge is
getting bigger due to the uneven distribution of technological facilities and the high cost of bandwidth (Al-Ansi et al., 2021). But more than that, the most important thing is the challenges that must be faced by teachers and lecturers in preparing teaching materials for students and how to do a good assessment in the online learning process (Almazova et al., 2020).

Problems with educational media and interactions between students and teachers are another issue that requires addressing. If teachers and students can engage directly in the classroom, then online learning requires that these interactions be altered and new approaches developed to ensure that the teaching and learning process is successful. The Learning Management System (LMS) is a popular online learning tool. As long as the user is connected to the system via the Internet, this system offers a learning platform that can enable interactive learning whenever and wherever the user is. In addition to offering instructional materials, well-designed e-learning platforms frequently make it easier to complete other tasks, including quizzes, written tests, and discussion boards (Dobre, 2015; Cavus, 2015).

Specifically for lecturers, several things must be worked on, and the mindset must be changed if they are conducting online tutoring. Lecturers are senior intellectuals, the key force that determines the quality of learning through guidance, the transmission of ideas, orientation, knowledge, and good life values to students. In addition, it is hoped that lecturers can also motivate themselves and students, bring positive energy to students, and contribute to changes in the learning paradigm if they are associated with changes in the teaching and learning process from a classroom system to online (Guri-Rosenblit, 2018).

Several previous studies that focused on lecturer profiles and relationships with online learning (e-learning) can be mentioned as follows: analysis of readiness and use of technology by lecturers (Sulisworo et al., 2020); readiness of lecturers in implementing synchronous and asynchronous learning systems (Sunarto, 2021); lecturer and student interactions in online learning systems (Davidovitch & Wadmany, 2021); digital literacy issues for lecturers (Guri-Rosenblit, 2018). In general, it can be seen that the role of lecturers as mentors in the learning process is important and therefore must be well-prepared by stakeholders in education (Purbojo, 2018).

The process of online teaching and learning can be argued to have been sparked and accelerated by the COVID-19 pandemic crisis. It is impossible to isolate the impact of the fourth industrial revolution from this online teaching and learning process. This concept is known as "education 4.0" in the educational community (Hussin, 2018). Education is established and developed based on lecturer performance and student perspectives, utilizing data gathered from their everyday teaching and learning activities, which is one of the hallmarks of Education 4.0. Data mining is the practice of processing and analyzing stored data in the realm of information technology to produce knowledge or information (Han et al. 2012). Data mining in education is the analysis of data relevant to the world of education and the application of the findings to inform the development of teaching and learning processes (Cope & Kalantzis, 2016; Suhirman et al., 2014).

The purpose of this study is to analyze the profile of lecturers in dealing with changes in learning methods from classroom learning to online learning. What factors influence the lecturer so that the lecturer is ready to make changes to
learning methods? Lecturer profile analysis is carried out using a Machine Learning approach using the K-means algorithm. Lecturers will be grouped into similar clusters based on the data from the answers to the questionnaires provided. The contribution to be made from the results of this research can be used to guide various stakeholders, such as higher education institutions and policymakers, to be effective and efficient in implementing e-learning.

**Method**

To analyze and view lecturer profiles based on the lecturer's answers to the questionnaires distributed, there are several methodological steps carried out in this study. In general, these steps can be seen in the following diagram:

![Figure 1. Research methodology](image)

Data is first pre-processed through data transformation, after which it is collected. The Elbow Method and K-means Algorithm are used in the subsequent step of cluster analysis to find the ideal number of clusters. The next step is to identify each cluster's pattern characteristics to collect crucial factors that establish the cluster's uniqueness. The next stage is to evaluate and extrapolate the cluster pattern's characteristics to identify the profile of a lecturer's readiness for the online teaching and learning process.

**Data**

Data was taken from answers to questionnaires from Sanata Dharma University lecturers that were circulated in April–June 2022. A total of 274 respondents answered a questionnaire about lecturers' readiness to deal with changes in learning methods from classroom learning to online learning due to the COVID-19 pandemic. Questionnaire answers were made on a Likert scale of 1–4, with a value of 1 strongly disagreeing and 4 strongly agreeing. The number of questions is 27, with a focus on student independence, learning materials, learning methods, and exploration of new knowledge. An example of a questionnaire question can be seen in Table 1.

<table>
<thead>
<tr>
<th>No</th>
<th>Fill in the questionnaire questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I use different methods in online learning compared to offline.</td>
</tr>
<tr>
<td>6</td>
<td>I don't need to make many changes in the way I teach for the success of teaching online learning.</td>
</tr>
<tr>
<td>11</td>
<td>I provide various lecture materials (text, ppt, video) in the LMS that I use.</td>
</tr>
<tr>
<td>17</td>
<td>The quizzes, tests, and assignments that I give are in accordance with the learning objectives and in accordance with the circumstances of students studying online.</td>
</tr>
<tr>
<td>22</td>
<td>Online lectures provide me with many possibilities that enrich the way I teach.</td>
</tr>
<tr>
<td>24</td>
<td>I don't have any significant problems in operating the LMS to manage lectures.</td>
</tr>
<tr>
<td>26</td>
<td>Online learning makes it easier for me to involve students in class discussions.</td>
</tr>
</tbody>
</table>
**Data preprocessing**

Data taken from real life, such as questionnaire answer data and primary data, where the data is obtained directly from the source, is often inconsistent, noisy, incomplete, and/or missing. So the first step after the data is obtained is preprocessing the data. This data is collected from various sources using data mining or data warehousing techniques.

Data pre-processing entails the procedures we must carry out to alter or encode data so that machines can read it readily (García et al., 2016). In the Knowledge Data Discovery process, data pre-processing is the second step after data collection (see Figure 2). The main idea is that for the model to make predictions accurately and precisely, the algorithm must be able to easily interpret data features.

![Figure 2. Data Pre-processing in Knowledge Data Discovery](image)

Data mining algorithms cannot effectively find patterns in noisy data; hence, they will not produce high-quality results. As a result, processing data is crucial to raising the level of data quality overall. The data will display erroneous statistical data as a whole if there are gaps or duplicate entries. Similar to outlier data, inconsistent data tends to throw off the learning model as a whole and result in incorrect predictions. In other words, sound judgments must be supported by solid evidence. To obtain data of this caliber, data pre-processing is crucial.

**Cluster analysis**

Clustering is an unsupervised learning approach used in machine learning and is a widely used method for statistical data analysis in various domains. The process of clustering involves putting data points into groups. We can classify each piece of data into a specific group within a single dataset using a clustering technique (Sarker, 2021).

Theoretically, data points belonging to the same group ought to share similar characteristics, but those belonging to other groups ought to have highly different characteristics. By examining which group is at the center of the data when we use the clustering method, clustering may be utilized in data science to extract certain valuable pattern characteristics from the data. The K-means algorithm is one of the more well-known algorithms.
**K-Means algorithm**

Probably the most well-known clustering algorithm is K-Means. It is quite simple to comprehend and use this algorithm. To begin, we first decide the classes or groups to employ and then randomly initialize the centers of each group. It is useful to take a cursory look at the data and make an effort to recognize the various categories to determine the appropriate number of classes. The center point, which represents the "X" on the graph we are making, is a vector with the same length as each vector representing a data point (Ghazal, 2021).

Each data point is assigned to the group whose center is closest to it once the distance between it and each group's center has been calculated. The group center is recalculated using these identified locations by averaging all of the vectors inside the group. This process should be repeated several times, or until there is little change in the group's center between iterations. The group center can be randomly selected a few times, and you can then pick the method that produces the greatest outcomes.

Implementing the K-Means algorithm has the benefit of being incredibly quick. Because there are so few computations involved—just calculating the distance between the point and the group's center—the process is quick. K-Means, on the other hand, has several shortcomings. We must first decide how many groups or courses there will be. The K-means algorithm starts with a random selection of cluster centers, which can result in different clustering when running multiple algorithms, making the choice of this class not always simple. As a result, the outcomes could be unpredictable and inconsistent.

**Elbow method**

The Elbow approach is the most traditional way to determine how many possible clusters are ideal for the dataset under analysis. The fundamental strategy is to choose $K = 2$ as the initial ideal cluster number $K$, increase $K$ constantly until it reaches a maximum for the predicted prospective optimal cluster number, and then identify the associated potential optimal cluster number $K$ (Humaira & Rasyidah, 2020).

The Elbow technique examines the relationship between the total WSS and the number of clusters. Additionally, a sufficient number of clusters must be chosen to prevent the addition of another cluster from raising the overall WSS. The following definition of the ideal cluster size applies:

1. Vary $k$ from 1 to 10 clusters to compute a clustering algorithm, such as k-means, for different $k$ values.
2. Determine the total number of squares in the cluster (WSS) for each $k$.
3. Plot the WSS curve based on the $k$-cluster count.
4. The number of clusters is often estimated from the position of the elbows in the plot.

**Variable importance**

In this study, we used the Boruta Technique to conduct a Variable Importance analysis to determine the elements that affect lecturer preparation for online learning. A feature selection technique called Boruta is based on the Random Forest Classifier algorithm. Making copies of features from the original dataset is how the algorithm operates. To create randomness in this copy, the values in each column
are mixed up. Shadow characteristics are the name for these jumbled characteristics. The original features and the shadow features are then combined to create a new feature space with dimensions twice as large as the original dataset.

The algorithm then determines whether the maximum importance of the shadow feature is greater than that of the original feature. The following stage is to build a classification Random Forest on this new feature space to compute a Z-score, a statistical test, to assess its significance. If a characteristic is deemed relevant, it is retained; otherwise, it is eliminated from the dataset. The dataset utilized in the second iteration is created from the features that, in the first iteration, satisfied the requirements. These traits are used again to construct shadow features, and the algorithm assesses their importance as it did in the initial iteration. While certain aspects were eliminated, others were preserved. This continues until a predetermined number of iterations are completed, all features are accepted, or all features are dropped (Naik & Mohan, 2019).

**Findings and Discussion**

We applied the Elbow method to count the number of clusters. This technique is employed to establish the ideal number of clusters. The Elbow method's findings demonstrate that k = 2, or a cluster of 2, is the best cluster.

![Figure 3. Number of Clusters](image)

To be sure whether it is true that k=2 is the optimal cluster, here is a description of k=2, 3, 4, and 5
It can be seen that for \( k = 3, 4, \) and \( 5 \), there is a mix between cluster members. In other words, the cluster does not or cannot differentiate more clearly between members. If \( k = 2 \) enlarges the image (Fig. 5), a more definite distinction will be seen between members in cluster 1 and cluster 2.

To find out what variables have an influence (Variable Importance), the Boruta technique was used in this study. The results of variable importance are as follows:
The green boxplot shows the variables that influence the purpose of the questionnaire. The yellow boxplot is a variable that can be considered an influential variable or not. While the red boxplot is a variable that does not support or has no effect on the intent and purpose of holding the questionnaire.

It can be seen that the order of the top 10 variables of importance can be explained as follows:

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Lecturer Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(J13) I believe the learning activities that I present in the LMS will make it easier for students to study independently.</td>
</tr>
<tr>
<td>2</td>
<td>(J12) The learning materials that I present in my LMS are designed to make it easier for students to study independently.</td>
</tr>
<tr>
<td>3</td>
<td>(J10) I give lecture material that I believe is easy for students to learn and master.</td>
</tr>
<tr>
<td>4</td>
<td>(J3) I can still teach well and with quality using this online mode.</td>
</tr>
<tr>
<td>5</td>
<td>(J7) I explore new knowledge according to my interests during online teaching from home.</td>
</tr>
<tr>
<td>6</td>
<td>(J25) In online learning, I can get reference books easily.</td>
</tr>
<tr>
<td>7</td>
<td>(J11) I provide various lecture materials (text, ppt, video) in the LMS that I use.</td>
</tr>
<tr>
<td>8</td>
<td>(J18) The learning evaluation tool that I gave to students objectively assessed students' abilities according to the learning objectives that had been agreed upon at the beginning of the lecture.</td>
</tr>
<tr>
<td>9</td>
<td>(J9) The lectures that I present in the LMS are equipped with clear learning objectives, learning materials, and instructions along with learning activities.</td>
</tr>
<tr>
<td>10</td>
<td>(J21) I am more enthusiastic about carrying out this online teaching assignment.</td>
</tr>
</tbody>
</table>
Of the 10 Variable Importance, it can be seen that three factors make lecturers ready to carry out online learning. The first factor is that lecturers can present quality learning activities and lecture materials and make it easier for students to study independently (J13, J12, J10, J3). In this case, the instructor modifies face-to-face classroom learning activities into online-learning-appropriate ones. Similarly, lecture content must be adapted for the online learning model. In other words, instructors can adapt learning methods to be more applicable to online education. Additionally, instructors are expected to provide prompt feedback on online assignments so that students know what needs improvement (Suparwito et al., 2021).

Aside from that, activities and materials must be designed to be simple to understand so that students have no trouble with online learning. In its implementation, instructors must foster student well-being. Educational psychology research emphasizes the significance of happiness as a crucial element of educational success (Kislyakov et al., 2014). It is believed that lecturers who foster an atmosphere of openness provide greater opportunities for communication and problem-solving. This type of emotional closeness is the foundation for the development of mutual trust, so students and professors do not feel too far apart. Meaningful feedback will further strengthen a student's closeness and familiarity with the teacher during the implementation phase of learning (Thurlings et al., 2013).

The second factor is that the lecturer explores new knowledge, obtains references, and provides various lecture materials (text, ppt, video) in LMS (J7, J25, J11). The ability and skills of lecturers to teach online are very important to the quality of successful online education (Kim, 2006). This also includes knowledge of online learning tools and features, as well as skills in using them in online learning. Furthermore, lecturers must be able to plan and develop diverse and high-quality lecture materials for online learning (Farmer & Ramsdale, 2016).

The third factor is the ability of lecturers to determine learning evaluation tools to objectively assess students' abilities according to learning objectives (J18, J9). The selection of the right evaluation tool is believed to be able to measure student learning success. There are several ways of evaluating and assessing appropriate online learning in tertiary institutions (Baldwin, 2017).

The study’s findings are consistent with the characteristics of successful online learning supported by various factors, including (Mulyatiningisih et al. 2020): (1) students who can learn independently and have high motivation to learn; (2) lecturers who master online learning technology; (3) learning strategies that provide opportunities for interaction between students, lecturers, and learning content; (4) learning content that is simple and has clear learning instructions; (5) short duration video media, and (6) educational institutions providing learning facilities such as online libraries, LMS, and lecturer training.

On the other hand, there are four variables (J1, J4, J8, and J19) that do not provide significant value for lecturer clustering in adapting online learning. Table 3 below presents four variables that do not contribute to the clustering of lecturers' adaptation to online teaching. In other words, the presence or absence of these variables will not have much effect on the lecturer's ability to teach online.
Table 3. Variable Importance that least influences the readiness of lecturers in online learning

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Lecturer Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(J8) I feel more enthusiastic about teaching when I can interact physically with students.</td>
</tr>
<tr>
<td>2</td>
<td>(J19) I return work, tests, or feedback within a reasonable time.</td>
</tr>
<tr>
<td>3</td>
<td>(J4) I don’t need to make many changes in the way I teach for the success of teaching.</td>
</tr>
<tr>
<td>4</td>
<td>(J1) I use different methods in online learning compared to offline.</td>
</tr>
</tbody>
</table>

**Lecturer profile**

Based on the important variable that has been analyzed in this study, the results of a cluster analysis are obtained, which show the profiles of USD lecturers in online learning. There are two groups of first and second-ranking lecturer clusters related to the readiness of lecturers for online learning (see Figure 5). Cluster 1 (red) = 106, with 4 data intersecting with cluster 2 (blue). Cluster 2 (blue) = 168, with 3 data intersecting with cluster 1 (red).

Four main factors influence whether lecturers are included in the ready category in online learning, as follows:

1. Lecturers can provide students with high-quality learning activities and lecture materials, and make it simpler for students to study independently. The lecturer modifies the learning model, activities, and lecture materials to accommodate online learning. In other words, lecturers can adapt offline learning methods so that they are more applicable to online learning, including the ability to provide prompt feedback on online assignments so that students know where they can improve.

2. Lecturers can design material and it is easy to learn so that students have no difficulty in online learning.

3. Lecturers are ready to explore new knowledge, obtain references, and provide various lecture materials (text, ppt, video) in the LMS. In this case, the abilities and skills of the lecturer are needed to learn about online learning tools and features, and how to use them in online learning. Furthermore, lecturers must be able to plan and develop diverse and high-quality lecture materials for online learning.

4. Lecturers can determine learning evaluation tools to objectively assess student abilities according to learning objectives. The selection of the right evaluation tool is believed to be able to measure student learning success.

On the other hand, three main factors influence whether lecturers are included in the category of being less prepared for online learning, namely:

1. Lecturers are used to teaching face-to-face, so they feel more enthusiastic about teaching when they can interact physically with students. This means that lecturers need to find new methods of interaction with students because offline methods cannot be fully used in online methods.

2. Lecturers feel that they do not need to make many changes in the way they teach face-to-face in class compared to online learning. This can mean that lecturers are reluctant to change, in other words, they are reluctant to change methods and create new learning materials. There should be changes in the
way of teaching, activities, and material delivery in online classes compared to offline classes.

3. The lecturer returns the results of work, tests, or feedback in an unreasonable timeframe. It is better that even though the system is online, feedback on online assignments is still given within a reasonable time because students need to know which understanding is correct and what still needs to be improved.

Conclusion
The lecturer profiles were effectively analyzed using a machine learning strategy and the K-means algorithm. By employing the Elbow Method, two clusters of professors are obtained. In addition, a variable importance analysis was conducted to determine the significant factors that determined the lecturers' preparedness for online learning. The first benefit is that instructors can provide students with high-quality learning activities and lecture materials, making it simpler for students to study independently. The second factor is lecturers investigating new information, obtaining references, and providing diverse lecture materials through the LMS. The third factor is the instructor's capacity to determine learning assessment instruments. These three factors can serve as a starting point for stakeholders preparing instructors to provide effective online learning.

References


