Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

SOFT SKILLS SEGMENTATION AND PROFILING ANALYSIS

Nur Anis Juliana Shafidan[1]
Shamshuritawati Sharif[2]

[1]School of Quantitative Sciences, Universiti Utara Malaysia (UUM), Malaysia,
(E-mail: nuranisjuliana@gmail.com)
[2]School of Quantitative Sciences, Universiti Utara Malaysia (UUM), Malaysia,
(Email: shamshurita@uum.edu.my)

Article history
Received date : 23-6-2023
Revised date : 24-6-2023
Accepted date : 10-10-2023
Published date : 15-10-2023

To cite this document:
Shafidan, N. A. J., & Sharif, S. (2023). Soft skills segmentation and profiling analysis. *Journal of Islamic, Social, Economics and Development (JISED)*, 8 (56), 275 – 290.

Abstract: *Internship is an intervention program to improve student' soft skills. Soft skill is crucial in preparing them for the workforce and helping them gain a better understanding of the workplace. Eventually the results of how well undergraduate students have used their soft skills in a real-world work environment is reported, but which students with low score soft skills are not being identified. By having that information, higher institution can help in preparing special training for certain soft skills with low score after they have finished their internship. With clustering techniques, segmentation, and the division of a large group of students into more focused subgroups (or cluster) based on shared soft skills can be accomplished. In this study, a segmentation of the internship students according to their soft skills is investigated. In this study, before implementing the outperform algorithm into real data, 20 hierarchical clustering algorithms were compared using cophenetic correlation coefficient. The 81 real data based on 12 soft skills were analysed using MATLAB software with the combination of Euclidean, Squared Euclidean, Spearman, Correlation and Minkowski distances with four linkages methods which are single, complete, average and Ward's method. The results show that the combination of Euclidean and Minkowski distance techniques with average-linkage method are the best combination with the cophenetic correlation coefficient value of 0.8412. Overall, it shows that S1 (knowledge and understanding skill) had the highest score among the three clusters produced with the value of 9.650 while S4 (communication skill) has the lowest score between other skills with the value of 4.333 respectively. As a summary, it has shown that clustering approaches can be used to explore students' soft skills and understanding their profiling.*

Keywords: *Soft Skills, Clustering Analysis, Cophenetic Correlation Coefficient, Mean Centroid*

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

Introduction

A bachelor's degree in a particular profession or major is the end objective of undergraduate study. Students also receive a foundational education to set them up for future employment (Klafter, 2020). This involves having the capacity to use theoretical ideas to solve practical issues and acquiring practical abilities for their future employment. As a result, a well-rounded education is created that covers a range of topics and specialties. Students are required to build a strong foundation of knowledge during their undergraduate studies in subjects including history, literature, mathematics, the natural sciences, and social sciences (Behle, 2020), along with a variety of skills that will equip them for success. Undergraduate education may also entail participating in research, internships, or other chances for experiential learning to develop real-world knowledge of the subject (Radović, G. K. Hummel & Vermeulen, 2021).

The problem occurred when the graduate's employability rate falls from year to year specially after the MCO. For example, the employability of Malaysian graduates had declined at 1.8% from 2019 (86.2%) to 2020 (84.4%), according to the study by Ali et al. (2021). The previous study by Yao & Tuliao (2019) was emphasizing and developing essential soft skills which is the key solution to boost the employability rate. Therefore, it is crucial to prepare students for the future by enhance them with soft skills that can adapt business demands and rapidly expanding technology. By doing so, Malaysian talent able to fulfil future demand and their graduate's employability in the labour market.

Undergraduate education aims to develop students' soft skills, such as communication, leadership, cognitive thinking, numeracy, digital literacy, practical, entrepreneurial, autonomy, knowledge and understanding, ethics, personal, and interpersonal skills (Hashim, 2019; Md Pazil & Che Razak, 2019). Communication, leadership, cognitive, and interpersonal abilities are among the soft talents that can aid students in acquiring strong soft skills (Denney et al., 2020; Guzmán et al., 2020; Morsidi et al., 2021). In practically every context, strong oral and written communication skills are necessary (Morsidi et al., 2021). These abilities include the capacity for crystal-clear speaking and writing in formal and academic settings. Additionally, undergraduate education aims to support students' personal and intellectual growth and development as persons. This entails exposure to many viewpoints and ideas, chances for intellectual and personal development, the building of a feeling of responsibility, and involvement in the community.

Previous studies had found that internship programmes have a number of advantages for students, including improving their skills and preparing them for the workforce (Anjum, 2020), helping them gain a better understanding of the workplace and soft skills related to self-management (Lin & Anantharajah, 2019), and allowing them to practise social skills like self-assurance and communication (Abdul Hamid, Islam, & Abd Manaf, 2014). Additionally, it was shown that students who always think positively are more likely to be productive, less worried, and tend to produce a higher standard work (Lin & Anantharajah, 2019). Students who participate in the internship programmes have the chance to learn about the needs and working conditions of the professional business environment. Consequently, it makes it possible for people to understand how theory and practise are connected, which enhances their comprehension of their ability to accomplish the goals set forth for various occupations (Landon-Hays et al., 2020). The internship programme is essential because it guarantees the professional preparation of upcoming graduates in several ways, such as by assisting them in understanding their chosen sector and their prospective future work chances. When sending undergraduate students to internship programmes, most higher-level institutions aim the

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

students having a chance to get real-world experience and use the skills they had learned in the classroom. The higher-level institutions consider it a success when the students get employment right after graduation because they do not have to waste theirs' time looking for a position.

In addition, organizations which employ students for internship programs continue to emphasise soft skills as essential employability skills and hold higher education accountable for failing to equip graduates with these talents. They frequently note that students lack of soft skills required for today's rapidly changing workplaces (Al Asefer & Zainal Abidin, 2021). The ability to assess how well undergraduate students have used their soft skills in a real-world work environment is one of the main justifications for looking into their soft skills after they had finished their internship. In such, soft skills often entail character traits and behaviours that are frequently linked to one another. For instance, successful communication frequently necessitates strong listening abilities, empathy, and the capacity to comprehend various viewpoints. Additionally, effective cooperation frequently involves strong leadership, communication, and problem-solving abilities.

In statistical discipline, there are one method that can be used to segment an observation into several clusters. It is useful for data exploration where we can understand which observation fall in the same cluster than those which differ from the others. A common unsupervised learning method used to cluster together comparable observation is called clustering technique (Mohamed Nafuri et al., 2022). This technique has found widespread use in several sectors, and in recent years, several clustering algorithms have been utilised to address a range of issues and enhance clustering performance (Abualigah & Khader, 2017; Cutting et al., 2017; Wang et al., 2016). With clustering techniques, segmentation, and the division of a large group of students into more focused subgroups (or cluster) based on shared soft skills can be accomplished. Therefore, we suggest a segmentation of the internship students according to their soft skills. By understanding the soft skills that students commonly possess a low, medium, and high score, and which students are in that level, employers and educators can put students with comparable soft skills together. It helps them better understand the needs and preferences of each group. In order to better address the specific demands of each student's segment, this information may subsequently be used to personalise training tactics and professional development skill offerings.

Research Questions

This study aims to produce a clustering model thus, below are the research questions that this study is trying to solve:

1. What is the overall student's soft skills score after an internship program?
2. What are the soft skills that students commonly possess a low, medium, and high score?
3. Which students are in the low, medium, and high score of soft skills?

Research Objectives

1. To determine the overall student's soft skill score after an internship program.
2. To determine the soft skills that students commonly possess a low, medium, and high score.
3. To investigate the list of students which are in the low, medium, and high score of soft skills.

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

Literature Review

Graduate Employability

Graduates are students who have successfully completed their studies (Md Pazil & Che Razak, 2019), and graduates' employability refers to the ability of graduates to gain and maintain employment that is relevant to their field of study, and which provides them with a reasonable level of job security, career progression opportunities, and adequate compensation (Behle, 2020). Employability has emerged as a highly significant result of higher education, even though its conceptualization within the sector remains somewhat vague. Nowadays, in the context of globalization and internationalization, the educational institutions are redefined. Universities are now tasked with the responsibility of equipping their students to have a broad spectrum of skills and capabilities in its workforce (Cheng et al., 2021). Generally, employability skills are measure the qualities, aptitudes, and technical skills that are applied to problem-solving and decision-making in the workplace (Abbasi et al., 2018). Employability also indicated the ability to find, keep, and progress in graduate employment (Behle, 2020), and sometimes employability is defined as the ability of a person to find work, hold a job, and negotiate the labour market in order to realise their full potential through a secure job. In Malaysia, the government has implemented various plans and initiatives to improve the graduate employability rate, including the Malaysian Education Blueprint (2015-2025) which aims to transform the Malaysian education system in order to produce graduates who possess the necessary skills and knowledge to meet the demands of the workforce and improve their employability (Sivalingam, 2020). The objective of this initiative is to enhance the quality of education, develop necessary skills among graduates, and increase their employability. This, in turn, will help to reduce unemployment rates and contribute to the economic development of the country. Furthermore, an internship or practicum program is a unique educational program that combines learning with organized and performance-related experience (Borders et al., 2015). This program aims to bridge the gap between the skills and knowledge that students acquire in their undergraduate education and the skills and knowledge that are required by employers in the workforce.

In the context of higher education, employability refers to the process through which students will be equipped to balance graduate life and employment. Higher education needs to comprehend the nature and future of employment in order to affect this process. According to the findings by Abelha et al (2020), higher education institutions are focused with employing techniques to improve the skills needed by graduates to find employment. The growth of graduate employability and competency depends on higher education institutions fostering a strong sense of creativity and cooperation. The development of graduates who are ready to handle life's responsibilities and work outside of their specialty must be the main goal of graduate employment. Employability must place an emphasis on aptitude, be at the core of the curriculum, accept diversity, and incorporate the metacognitive skills that equip graduates of higher education not just to enter the workforce but also be ready to learn (Bennett, 2019). In order to promote graduates' level of competency, Ma'dan et al (2020) recommends for more significant institutional adjustments and reforms, such as enhancing teaching strategies and available programs. To establish a competitive and capable talent pool in the nation, university-industry partnership must also be increased. From the previous study by Okolie et al (2019), it showed that the important recommendations on how higher education teachers may better encourage and inspire students to gain the information, skills, and attributes that employers value. From all of the above findings and recommendations, it can be seen that the skills of the

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

students from the higher education plays an important role in enhancing the graduate employability.

Soft Skills

A person's capacity to adapt to their environment is measured by their soft skills, which include communication, critical thinking, problem solving, practical knowledge, and understanding (Qizi, 2020). Soft skills are also related with personal values and interpersonal abilities. Additionally, soft skills are character traits that people need in order to function in daily life (Rodzalan et al., 2022). Employers and organisations are looking for people with strong academic credentials who can improve the performance of their organisation in the employment market. But it is necessary to make sure that academic credentials are also targeted towards people with critical soft skills and not just at highly qualified applicants (Succi & Canovi, 2019). From the previous study, it showed that the employers place a higher value on professionalism, adaptability to change, creativity and innovation, customer focus, and teamwork, whereas students place a higher value on networking and conflict resolution abilities (Pardo-Garcia et al., 2020). Importantly, academia must forge closer relationships with industry and collaborate successfully to guarantee ready-for-the-job graduates (Heng et al., 2021). In addition, professional associations, independent training programmes, and university education institutions are all resources for acquiring skills. Soft skills, on the other hand, are frequently poorly defined or conceptualised in the marketplace in contrast to technical talents. Other names for soft skills are employable skills, transferrable skills, non-technical abilities, and people skills (Hirudayaraj et al., 2021). Soft skills may therefore be applied in a variety of spheres of life, not only the job. Early on in their academic careers, undergraduate students can benefit from developing their interpersonal, social, and civic engagement abilities, which can help them become more well-rounded people (Idrus, 2014) and enhance their general quality of life.

In many aspects of life, including interpersonal interactions, workplace settings, and education, effective communication is crucial. In order to communicate ideas and information, two or more individuals must interact in order to be in communication (Reith-Hall & Montgomery, 2022). Therefore, the educators need to embrace a more relevant viewpoint than the conventional belief that in-person instruction is preferable to online instruction for communication skill development (Besley et al., 2015). A person can get training to increase self-efficacy, which is the conviction that they have the capacity to communicate effectively, in order to deal with this issue of inadequate communication (Sousa Mata et al., 2021). It is crucial to have an effective communication to build a strong relationship.

In addition, Akhmetshin et al. (2019) discovered a weakness to this study of soft skills in the inability to execute a program-corrective experiment that might momentarily change how students are taught. This is due to the busy schedule final-year students have since they must complete their final-year projects and get ready for their final exams. This study also discovered that final-year students whose intended careers involved entrepreneurship tended to highly value the traits that contributed to their unique perspectives on the world of business. The university's education program and atmosphere that foster entrepreneurial thinking are highly scrutinized by working students. The earlier study by Md Pazil and Che Razak (2019) tries to close the gap in defining and categorising the soft skills of the current graduates from the viewpoint of Asian companies. It resulted in eleven (11) major domains related to graduates' soft skills which are communication skills, entrepreneurial skills, interpersonal skills, lifelong learning skills, management skills, numeracy skills, problem solving skills, professional ethics and moral skills, self-management skills, technological skills and thinking skills. The soft skills

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

of graduates must be developed or improved by higher education institutions in order to prepare them for the workplace. The quality of graduates is more important than the number in the workforce. This study is significant because it might help students and higher education institutions provide students with the necessary soft skills to fulfil the organization demands. From several research paper, a summary on the soft skills is shown in Table 1.

Table 1: List of soft skills

Soft skills	References
Communication	(Abdul Hamid, Islam, & Abd Manaf, 2014); (Hashim, 2019); (Md Pazil & Che Razak, 2019)
Entrepreneurial	(Abdul Hamid, Islam, & Abd Manaf, 2014); (Hashim, 2019); (Md Pazil & Che Razak, 2019)
Interpersonal	(Abdul Hamid, Islam, & Abd Manaf, 2014); (Hashim, 2019); (Md Pazil & Che Razak, 2019)
Lifelong learning	(Md Pazil & Che Razak, 2019)
Management	(Abdul Hamid, Islam, & Abd Manaf, 2014); (Md Pazil & Che Razak, 2019)
Numeracy	(Hashim, 2019); (Md Pazil & Che Razak, 2019)
Problem solving	(Md Pazil & Che Razak, 2019)
Professional ethics and moral	(Hashim, 2019); (Md Pazil & Che Razak, 2019)
Self-management	(Md Pazil & Che Razak, 2019)
Technological	(Md Pazil & Che Razak, 2019)
Thinking	(Abdul Hamid, Islam, & Abd Manaf, 2014); (Md Pazil & Che Razak, 2019)

Research Methodology

An overview of several clustering approaches, assessments, uses, difficulties, and potential future study is given in this review of the literature. There are several kinds of clustering approaches, including partitional, hierarchical, density-based, grid-based, and model-based clustering, according to Ghosal et al., (2020). By assessing the calibre of the clustering results, it is crucial to establish the right number of clusters and the efficacy of clustering (Nagargoje et al., 2020). The gap statistic, the elbow approach, and silhouette analysis are examples of common cluster evaluation techniques (Sagala & Gunawan, 2021). Sagala and Gunawan (2020) state that silhouette analysis evaluates how similar a data point is to its own cluster in comparison to other clusters, the elbow method plots the sum of squared distances against the number of clusters to determine the number of clusters, and the gap statistic evaluates the within-cluster dispersion in comparison to a reference distribution. In this study, we use the Silhouette index to evaluate the clustering performance. According to Shutaywi and N. Kachouie (2021), the formula are as follows;

$$S(x_i) = \frac{b(x_i) - a(x_i)}{max\{b(x_i), a(x_i)\}} \quad (1)$$

where x_i is an element in cluster π_k, $a(x_i)$ is the average distance of x_i to all other elements in the cluster π_κ (within dissimilarity), and

$$b(x_i) = min\{d_\iota(x_i)\}, \text{among all clusters } \iota \neq \kappa$$

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

where $d_\iota(x_i)$ is the average distance from x_i to all points in cluster π_ι for $\iota \neq \kappa$ (between dissimilarity). From Equation (1) the value of the Silhouette width can vary between -1 and 1.

Jarman (2020) asserts that using the HCA approach, researchers may decipher the structure and significance of data from a big unstructured dataset. However, the clustering results also depends on its algorithms (the way we choose algorithm). The algorithm are as follows given a set of m items to be clustered:

Step 1: Initially consider each instance as a disjoint cluster at level *K(m)*, *m* = 0.

Step 2: Calculate the **dissimilarity matrix** to hold all distances (y_a, y_b).

Step 3: Locate similar cluster based on minimum distance, $D\ (A, B) = \{d\ (y_a, y_b)\}$

Step 4: Move to next sequence $m = m + 1$. Merge cluster (A) and (B) into a single cluster to form the next level cluster $(m) = d[(A)(B)]$ using **linkage methods**.

Step 5: Adjust the similarity matrix *D*, by removing the row and column values corresponding to cluster (a) and (b) and adding a row and column corresponding to the newly formed cluster. The proximity between the new cluster (*A*, *B*) and old cluster in the previous level (*K*) is denoted as *[(K), (A, B)] = {d [(K), (A)], d [(K), (B)]}*.

Step 6: Repeat step 2 until single cluster formation.

Figure 1: Hierarchical Clustering Algorithms

The decision on which distance and which linkage to be combined are crucial. Researcher will choose based on the popularity however, it supposed to be based on the nature of the data such as normal distribution, exponential distribution, gamma distribution or others.

Distance Measures

Distance metrics in HCA are essential since they indicate how similar or different items or groups are. The Euclidean, Manhattan, and Mahalanobis distances are often used distance metrics in hierarchical clustering (Sulc & Rezankova, 2019). The most used distance metric for continuous data is euclidean distance, whereas cosine distance is widely employed for text data (Bu et al., 2020). For high-dimensional data, the Mahalanobis distance is a commonly used distance metric (Bulut, 2020). For particular types of data, other distance metrics including the Minkowski, correlation, and spearman distances are also utilised (Baldini & Geneiatakis, 2019). The distance between points on a straight line is referred to as the Euclidean distance. The Pythagorean theorem is employed in this distance calculation approach, which is the one that machine learning processes most frequently (Suwanda et al., 2020). According to Thant and Aye (2020), this approach has certain benefits, including the fact that the distance between any two clusters will not be affected by the addition of new objects to the study that could be outliers. The same equation as the Euclidean distance metric is then used by Squared-Euclidean, but it does not take the squared root (Khulaidah & Irsalinda, 2020). Following that, the Spearman distance between two clusters is defined as the square of their Euclidean distance (Crispino et al., 2021). The degree of all potential linkages between two clusters is then

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

measured by correlation distance (Hou et al., 2022). This distance's key benefit is that it can quantify dependency between two places of varying dimensions. In order to calculate it, one is subtracted from the sample correlation between the points, which is thought of as a series of numbers (Ruiz et al., 2020). Finally, Minkowski distance is a possible generalisation of both Manhattan distance and Euclidean distance (Thant & Aye, 2020).

Linkages Method

In hierarchical clustering, the rules or algorithms used to compute the distance between two clusters are referred to as linkage techniques. The aim of hierarchical clustering is to form clusters from comparable data points. The final clusters' formation is influenced by the linking standards since they control how the distance between two clusters is determined. Single-linkage, complete-linkage, average-linkage, and Ward's technique are some of the main linkage standards used in hierarchical clustering (Hossen, Md., & Hoque, 2015). A summary on the comparison of four linkages method mentioned above is shown in Table 2.

Table 2: Comparison of four linkages method

Method	Linkage criteria	Comment	Selected references
Single-Linkage	Shortest distance between any two points in the two clusters	Easy to use, sensitive to noise, outliers, and may not always produce optimal results	(Jarman, 2020; Ros & Guillaume, 2019)
Complete-Linkage	Maximum distance between any two points in the two clusters	Tends to produce more compact, spherical clusters that are useful for identifying clusters with similar variances	(Jarman, 2020; Yildirim & Birant, 2017)
Average-Linkage	Average distance between all pairs of points in the two clusters	Suitable for classification tasks with multiple samples and big data	(Bu et al., 2020; Jarman, 2020)
Ward's clustering method	Group similar objects into clusters based on the similarity of their attributes	Commonly used with the objective function of minimum variance	(Großwendt, Röglin, & Schmidt, 2019; Strauss & Maltitz, 2017)

Cluster Evaluation

The cophenetic correlation for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree, and the original distances (or dissimilarities) used to construct the tree (Saraçli, Dogan, & Dogan, 2013). It is a measure of how faithfully the tree represents the dissimilarities among observations. The formula is as follows,

$$c = \frac{\sum_{a<b}(x_{ab}-x)(y_{ab}-y)}{\sqrt{\sum_{a<b}(x_{ab}-x\sum_{a<b})(y_{ab}-y)^2}} \quad (2)$$

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

where x_{ab} is the distance between objects *a* and *b* in *X*, y_{ab} is the cophenetic distance between objects *a* and *b*, from *Y*, while *x* and *y* are the average of *X* and *Y*, respectively. The coefficient value represents the value between distance matrix and cophenetic matrix. As the value of the cophenetic correlation coefficient is quite close to 1 by comparing the coefficients, it can be said that the clustering is fit. Overall, based on Table 2 there were 20 algorithms generated throughout this study.

Findings

A series of analysis have been run using MATLAB software with the combination of five different distance techniques, namely Euclidean, Squared Euclidean, Spearman, Correlation and Minkowski distances with four linkages methods which are single, complete, average and Ward's method. Performance is measured based on cophenetic correlation coefficient values. The cophenetic correlation coefficient measures how well a categorization fits a collection of data ranges from -1 to 1 (Saraçli, Dogan, & Dogan, 2013). For a high-quality solution, the magnitude of this value should be very close to 1.

Table 3: Combination of Distance Techniques and Linkages Method

Distance Techniques	Linkages Method	Cophenetic Correlation Coefficient	Ranking
Euclidean	Single	0.7400	3
	Complete	0.8177	2
	Average	**0.8412**	**1**
	Ward	0.5441	4
Squared Euclidean	Single	0.7234	4
	Complete	0.7974	2
	Average	**0.8063**	**1**
	Ward	0.7827	3
Spearman	Single	0.4328	4
	Complete	0.4675	2
	Average	**0.5605**	**1**
	Ward	0.4522	3
Correlation	Single	0.4439	2
	Complete	0.4279	4
	Average	**0.6032**	**1**
	Ward	0.4432	3
Minkowski	Single	0.7400	3
	Complete	0.8177	2
	Average	**0.8412**	**1**
	Ward	0.5441	4

As overall, the combination of Euclidean and Minkowski distance techniques with average-linkage method are the best combination with the cophenetic correlation coefficient value of 0.8412 respectively which is expected because this average-linkage method produced better partitions than the single-linkage, complete-linkage, and Ward's methods.

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

Silhouette index

Table 4: Silhouette index result

Number of clusters	Silhouette index
3	0.454

There are three number of clusters produced from the clustering model with 0.454 silhouette index that indicate an average clustering result. The cophenetic correlation coefficient value is 0.845 which shows that a well clustering result matches the original similarity.

Membership and Cluster Size

Table 5: Results by Cluster

Cluster	1	2	3
Number of objects by cluster	74	4	3
Cluster Size (%)	91.36	4.94	3.70

As shown in Table 5, three clusters are derived from the clustering model. Cluster 1 constitute the largest group comprising 91.36% for 74 students while the Cluster 2 present the second highest cluster comprising 4.94% with four students respectively. Whereas, the lowest cluster is Cluster 3, making up 3.70% of the entire cluster each consist of three respondents which indicates the low soft skills cluster.

Mean Centroid

Overall, the mean score for S1 (knowledge and understanding skill) is 9.5, S2 (ethics skill) is 9.2, S3 (autonomy skill) is 9.1, S4 (communication skill) is 8.4, S5 (leadership skill) is 8.1, S6 (practical skill) is 8.6, S7 (cognitive skill) is 8.6, S8 (interpersonal skill) is 9.0, S9 (numeracy skill) is 8.4, S10 (entrepreneurial skill) is 8.9, S11 (personal skill) is 8.9 and S12 (digital skill) is 8.5 respectively. For each cluster, the heat map analysis can identify more than two crucial characteristics, if they are strongly correlated. The best method for large-scale data visualisation is heat map analysis. Light green denotes an attribute with a low mean centroid value, and darker green denotes an attribute with a high mean centroid value on the colour scale used to represent the relevance of the attributes.

Table 6. List of Mean Centroids for Each Clustering Model from Heat Map Analysis

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Cluster 1	9.649	9.459	9.270	8.676	8.284	8.797	8.851	9.230	8.608	9.095	9.081	8.703
Cluster 2	6.500	6.500	7.250	6.250	6.250	7.000	6.500	7.750	6.250	7.500	7.500	5.750
Cluster 3	9.000	6.333	7.000	4.333	5.000	5.667	5.667	6.000	5.333	6.000	6.000	7.333

Indicators:

Low	Medium	High

In Cluster 1, all soft skills were in the high level of centroid values. The top three were S1 (knowledge and understanding skill) with the value of 9.649, S2 (ethics skill) with the value of 9.459, and S3 (autonomy skill) with the value of 9.270. It represents students that can work independently, prioritize tasks, and meet deadlines without constant supervision and ethical decision-makers, making them valuable assets to any organization or community.

In Cluster 2, five soft skills were in medium level of centroid values which were S8 (interpersonal skill) with the value of 7.750, S10 (entrepreneurial skill) and S11 (personal skill) with the value

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

of 7.500, S3 (autonomy skill) with the value of 7.250 and S6 (practical skill) with the mean centroid value of 7.000 respectively meanwhile the others seven soft skills were in the low level of centroid values which were S1 (knowledge and understanding skill), S2 (ethics skill) and S7 (cognitive skill) with the value of 6.500, S4 (communication skill), S5 (leadership skill) and S9 (numeracy skill) with the value of 6.250 and finally S12 (digital skill) with the mean centroid value of 5.750 respectively. It represents the students that reasonably well collaborate in teams and effectively handle the interpersonal conflicts arise. They may also explore new opportunities with a moderate level of comfort. But with a low level of digital skills, they may struggle to leverage online platforms, resources, or e-learning opportunities to enhance their skills and knowledge.

In Cluster 3, most of the soft skills were in the low level of centroid values, two skills which were S12 (digital skill) with the value of 7.333 and S3 (autonomy skill) with the value of 7.000 were in the medium level of centroid values while only one soft skill in the high level of centroid values which was S1 (knowledge and understanding skill) with the value of 9.000 respectively. This person would possess a wealth of knowledge in their field of expertise, but less advance in their digital skills. However, with low communication skills in verbal and written, they may struggle with effectively expressing their thoughts and ideas to others. Overall, it shows that S1 (knowledge and understanding skill) had the highest score among the three clusters produced with the value of 9.650 while S4 (communication skill) has the lowest score between other skills with the value of 4.333 respectively. It is shown that clustering techniques can help in exploring students' soft skills and understanding their profiling.

Descriptive Statistics

Table 7: Descriptive Analysis for Clustering Model Analysis

	Description	Cluster 1	Cluster 2	Cluster 3
Gender	Male	17 (23%)	1 (25%)	0 (0%)
	Female	57 (77%)	3 (75%)	3 (100%)
Industry / sector categories	Government	29 (39%)	2 (50%)	0 (0%)
	Private	24 (32%)	1 (25%)	1 (33%)
	Non-Government Organization (NGO)	3 (4%)	0 (0%)	0 (0%)
	Government-Linked Companies (GLC)	7 (9%)	0 (0%)	1 (33%)
	Multinational Company	5 (7%)	1 (25%)	1 (33%)
	Statutory Body	6 (8%)	0 (0%)	0 (0%)
Working experience before joining internship	No experience	11 (15%)	2 (50%)	0 (0%)
	Part time worker	38 (51%)	1 (25%)	0 (0%)
	Online business	9 (12%)	0 (0%)	0 (0%)
	Family business	4 (5%)	1 (25%)	0 (0%)
	Others	12 (16%)	0 (0%)	3 (100%)
Received job interview	No	56 (76%)	4 (100%)	3 (100%)
	Yes	18 (24%)	0 (0%)	0 (0%)

In Cluster 1, there were a total of 74 students where 57 students were female which was represented by 77% while 17 students were male which was represented by 23% of the overall students in this cluster. Furthermore, 39% of the students in this cluster had gone theirs' internship program at government sector, represented by 29 students and 24 (32%) of the

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

students in the private sector. 9% of the students represented by 7 students in this cluster had gone theirs' internship program at government-linked companies' sector, 6 students at statutory body, 5 students at multinational company and the least sector represented by the students were non-government organization with 3 students (4%) only. Next, 51% (38) of the students in this cluster had a working experience as a part time worker before joining the internship, 16% (12) of the students had others experience, 15% (11) of the students had no working experience, 12% (9) of the students had an experience in online business while the other 5% (4) of the students work in the family business as a working experience before joining the internship program. Besides, most of the students in this cluster did not receive their job interview which was represented by 56 (76%) students while 18 (24%) students had received it.

In Cluster 2, there were a total of 4 students where 3 students are female while only 1 student is male. Furthermore, 2 students in this cluster had gone theirs' internship program at government sector, 1 student at private sector and only 1 student at multinational company. Next, 50% (2) of the students had no working experience, while only 25% (1) of the students work as a part time worker and in the family business for each experience respectively. Besides, all 4 students in this cluster did not receive their job interview.

In Cluster 3, there were a total of 3 students only where all the students are female and they were Malay races. Furthermore, 1 student in this cluster had gone theirs' internship program at private sector, 1 student at government-linked companies' sector and only 1 student at multinational company. Next, all the 3 students had others working experience before joining the internship program and they also did not receive their job interview.

Conclusion

In conclusion, the overall student's soft skills score after an internship program is in a good score because the mean centroid values are ranging from 8.1 to 9.5. Besides, this study also identifies which soft skills that students commonly possess a low, medium, and high score. The soft skills with a high score belong to S1 (knowledge and understanding skill), S2 (ethics skill), S3 (autonomy skill), S8 (interpersonal skill), while the medium score of soft skills represented by S10 (entrepreneurial skill), S11 (personal skill), S6 (practical skill), S7 (cognitive skill), S12 (digital skill), S4 (communication skill) and S9 (numeracy skill) respectively but only one soft skill is in the low score which is S5 (leadership skill) only. Besides, it shows that S1 (knowledge and understanding skill) had the highest score among the three clusters produced with the value of 9.650 while S4 (communication skill) has the lowest score between other skills with the value of 4.333 respectively. In addition, this study also produces three clusters which are represented by the internship students. Overall, the students in Cluster 1 possess a high score of soft skills while the students in Cluster 2 possess a medium score of soft skills and finally the students in Cluster 3 possess a low score of soft skills respectively.

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

References

Abbasi, F. K., Ali, A., & Bibi, N. (2018). Analysis of skill gap for business graduates: managerial perspective from banking industry. *Education+ Training*, 60(4), 354-367.

Abdul Hamid, M. S., Islam, R., & Abd Manaf, N. H. (2014). Employability Skills Development Approaches: An Application of the Analytic Network Process. *Asian Academy of Management Journal*, 93-111.

Abelha, M., Fernandes, S., Mesquita, D., Seabra, F., & Ferreira-Oliveira, A. T. (2020). Graduate employability and competence development in higher education—A systematic literature review using PRISMA. Sustainability, 12(15), 5900.

Abualigah, L. M., & Khader, A. T. (2017). Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. *The Journal of Supercomputing*, 73, 4773-4795.

Akhmetshin et al., E. M. (2019). Acquisition of entrepreneurial skills and competences: Curriculum development and evaluation for higher education. *Journal of Entrepreneurship Education*, ().

Al Asefer, M., & Zainal Abidin, N. S. (2021). Soft Skills and Graduates' Employability in the 21st Century From Employers' Perspectives: A Review of Literature. *International Journal of Infrastructure Research and Management*, 44-59. From https://iukl.edu.my/rmc/wp-content/uploads/sites/4/2022/02/5.-Al-Asefer_compressed.pdf

Ali et al., A. J. (2021). The Lesser Prominent Factors Contributing to Malaysian. *ICCD*, 287-292. doi:10.33068/iccd.Vol3.Iss1.363

Anjum, S. (2020). Impact of internship programs on professional and personal development of business students: a case study from Pakistan. *Future Business Journal*, 1-13. doi:10.1186/s43093-019-0007-3

Baldini, G., & Geneiatakis, D. (2019). A Performance Evaluation on Distance Measures in KNN for Mobile Malware Detection. *IEEE Systems, Man, and Cybernetics Society*, 193-198.

Behle, H. (2020). Students' and graduates' employability. A framework to classify and measure employability gain. *Policy Reviews In Higher Education*, 105-130. doi:10.1080/23322969.2020.1712662

Bennett, D. (2019). Graduate employability and higher education: Past, present and future. HERDSA Review of Higher Education, 5, 31-61.

Besley, J. C., Dudo, A., & Storksdieck, M. (2015). Scientists' views about communication training. Journal of research in science teaching, 52(2), 199-220.

Borders, L. D., Brown, J. B., & Purgason, L. L. (2015). Triadic supervision with practicum and internship counseling students: A peer supervision approach. The Clinical Supervisor, 34(2), 232-248.

Bu et al., J. &. (2020). Comparative Study of Hydrochemical Classification Based on Different Hierarchical Cluster Analysis Methods. *International Journal of Environmental Research and Public Health*, 9515. doi:10.3390/ijerph17249515

Bulut, H. (2020). Mahalanobis distance based on minimum regularized covariance determinant estimators for high dimensional data. *Communications in Statistics - Theory and Methods*, (). doi:10.1080/03610926.2020.1719420

Cheng, M., Adekola, O., Albia, J., & Cai, S. (2022). Employability in higher education: a review of key stakeholders' perspectives. Higher Education Evaluation and Development, 16(1), 16-31.

Crispino, M., Mollica, C., Astuti, V., & Tardella, L. (2022). Efficient and accurate inference for mixtures of Mallows models with Spearman distance. *arXiv preprint*

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

arXiv:2209.09821.

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Tukey, J. W. (2017, August). Scatter/gather: A cluster-based approach to browsing large document collections. In *ACM SIGIR Forum* (Vol. 51, No. 2, pp. 148-159). New York, NY, USA: ACM.

Denney et al., V. &. (2020). Project Management Leadership and Interpersonal Skills: The Past, Present, and Future. *Global Journal of Management and Marketing*, 135. From https://commons.erau.edu/publication/1623

Ghosal et al., A. &. (2020). A Short Review on Different Clustering Techniques and Their Applications. *Emerging Technology in Modelling and Graphics*, 69-84. doi:10.1007/978-981-13-7403-6_9

Großwendt, A., Röglin, H., & Schmidt, M. (2019). Analysis of Ward's Method. *()*, 1-9.

Guzmán et al., V. E. (2020). Characteristics and Skills of Leadership in the Context of Industry 4.0. *Procedia Manufacturing*, 543-550. doi:10.1016/j.promfg.2020.02.167

Hashim, N. H. (2019). *Taklimat Semakan Kurikulum Program Pascasiswazah.* Johor Bahru: UTM.

Heng, K., & Sol, K. (2021). COVID-19 and Cambodian higher education: Challenges and opportunities. Online learning during COVID-19 and key issues in education, 31-48.

Hirudayaraj, M., Baker, R., Baker, F., & Eastman, M. (2021). Soft skills for entry-level engineers: What employers want. *Education Sciences*, 11(10), 641.

Hossen, M. B., Md., S.-U.-D., & Hoque, A. (2015). Methods for Evaluating Agglomerative Hierarchical Clustering for Gene Expression Data: A Comparative Study. *Computational Biology and Bioinformatics*, 88-94. doi:10.11648/j.cbb.20150306.12

Hou, J., Ye, X., Feng, W., Zhang, Q., Han, Y., Liu, Y., ... & Wei, Y. (2022). Distance correlation application to gene co-expression network analysis. *BMC bioinformatics*, 23(1), 1-24.

Idrus, H. (2014). Developing well-rounded graduates through integration of soft skills in the teaching of engineering courses. *IEEE Frontiers in Education Conference (FIE) Proceedings* (pp. 1-9). IEEE.

Jarman, A. M. (2020). Hierarchical Cluster Analysis: Comparison of Single linkage, Complete linkage, Average linkage and Centroid Linkage Method. *Research Gate*, (). doi:10.13140/RG.2.2.11388.90240

Kenayathulla, H. B., Ahmad, N. A., & Idris, A. R. (2019). Gaps between competence and importance of employability skills: evidence from Malaysia. *Higher Education Evaluation and Development*, 13(2), 97-112.

Khulaidah, E. Z., & Irsalinda, N. (2020, August). FCM using squared euclidean distance for e-commerce classification in Indonesia. In *Journal of Physics: Conference Series* (Vol. 1613, No. 1, p. 012071). IOP Publishing.

Klafter, C. E. (2020). Undergraduate Education and the Maturation of Students. *Immaturity*, 34-45. doi:10.1007/s12129-019-09851-z

Lin, C. Y., & Anantharajah, S. (2019). Perceived Expectations of Internships: Case Study of a Private University in Malaysia. *Asia-Pacific Journal of Innovation in Hospitality and Tourism*, 1-16. From https://expert.taylors.edu.my/file/rems/publication/100280_5735_1.pdf#page=8

Landon-Hays et al., M. &.-A. (2020). Learning to Teach: How a Simulated Learning Environment Can Connect Theory to Practice in General and Special Education Educator Preparation Programs. *Education Sciences*, 1-17. doi:10.3390/educsci10070184

Ma'dan, M., Ismail, M. T., & Daud, S. (2020). Strategies to enhance graduate employability: insights from Malaysian public university policy-makers. Malaysian Journal of Learning and Instruction, 17(2), 137-165.

Md Pazil, A. H., & Che Razak, R. (2019). Perspectives of Asian Employers on Graduates' Soft

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

Skills: A Systematic Review. *Universal Journal of Educational Research*, 2397-2405. doi:10.13189/ujer.2019.071117

Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering Analysis for Classifying Student Academic Performance in Higher Education. *Applied Sciences*, 12(19), 9467

Morsidi et al., S. (2021). WhatsApp and Its Potential to Develop Communication Skills among University Students. *International Journal of Interactive Mobile Technologies*, 23. doi:10.3991/ijim.v15i23.27243

Nagargoje et al., A. &. (2020). Performance Evaluation of the Data Clustering Techniques and Cluster Validity Indices for Efficient Toolpath Development for Incremental Sheet Forming. *Journal of Computing and Information Science in Engineering*, (). doi:10.1115/1.4048914

Okolie, U. C., Nwosu, H. E., & Mlanga, S. (2019). Graduate employability: How the higher education institutions can meet the demand of the labour market. Higher education, skills and work-based learning, 9(4), 620-636.

Pardo-Garcia, C., & Barac, M. (2020). Promoting employability in higher education: A case study on boosting entrepreneurship skills. Sustainability, 12(10), 4004.

Qizi, K. N. U. (2020). Soft skills development in higher education. *Universal journal of educational research*, 8(5), 1916-1925.

Radović, S., G. K. Hummel, H., & Vermeulen, M. (2021). The challenge of designing 'more'experiential learning in higher education programs in the field of teacher education: A systematic review study. *International Journal of Lifelong Education*, 545-560. doi:10.1080/02601370.2021.1994664

Reith-Hall, E., & Montgomery, P. (2022). The Teaching and Learning of Communication Skills in Social Work Education. *Research on Social Work Practice*, 793-813. doi:10.1177/10497315221088285

Rodzalan et al., S. &. (2022). TVET Skills Gap Analysis in Electrical and Electronic Industry: Perspectives from Academicians and Industry Players. *Journal of Techinical Education and Training*, 158-177. doi:10.30880/jtet.2022.14.01.014

Ros, F., & Guillaume, S. (2019). A hierarchical clustering algorithm and an improvement of the single linkage criterion to deal with noise. *Expert Systems With Applications*, 96-108. doi:10.1016/j.eswa.2019.03.031

Ruiz, L. G. B., Pegalajar, M. C., Arcucci, R., & Molina-Solana, M. (2020). A time-series clustering methodology for knowledge extraction in energy consumption data. *Expert Systems with Applications*, 160, 113731.

Sagala, N. T., & Gunawan, A. A. (2021). Discovering the Optimal Number of Crime Cluster Using Elbow, Silhouette, Gap Statistics, and NbClust Methods. *Computer, Mathematics and Engineering Applications*, 1-10. doi:10.21512/comtech.v13i1.7270

Saraçli, S., Dogan, N., & Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Applications*, 1-8. From http://www.journalofinequalitiesandapplications.com/content/2013/1/203

Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6), 759.

Sivalingam, A. D., & Mansori, S. (2020). How Organizations Should View Reskilling and Upskilling The Workforce. doi:10.5281/zenodo.4302700.

Strauss, T., & Maltitz, M. J. (2017). Generalising Ward's Method for Use with Manhattan Distances. *Plos One*, 1-21. doi:10.1371/journal.pone.0168288

Volume: 8 Issues: 56 [September, 2023] pp. 275 - 290
Journal of Islamic, Social, Economics and Development (JISED)
eISSN: 0128-1755
Journal website: www.jised.com
DOI: 10.55573/JISED.085624

Sousa Mata et al., Á. N. (2021). Training in communication skills for self efficacy of health professionals: a systematic review. *Human Resources for Health*, 1-9. doi:10.1186/s12960-021-00574-3

Succi, C., & Canovi, M. (2020). Soft skills to enhance graduate employability: comparing students and employers' perceptions. *Studies in higher education*, 45(9), 1834-1847.

Sulc, Z., & Rezankova´, H. (2019). Comparison of Similarity Measures for Categorical Data in Hierarchical Clustering. *Journal of Classification*, (). doi:10.1007/s00357-019-09317-5

Suwanda, R., Syahputra, Z., & Zamzami, E. M. (2020, June). Analysis of euclidean distance and manhattan distance in the K-means algorithm for variations number of centroid K. In *Journal of Physics: Conference Series* (Vol. 1566, No. 1, p. 012058). IOP Publishing.

Thant, A. A., Aye, S. M., & Mandalay, M. (2020). Euclidean, Manhattan and Minkowski Distance Methods For Clustering Algorithms. *Int. J. Sci. Res. Sci. Eng. Technol*, 7.

Wang, Y., Chen, Q., Kang, C., & Xia, Q. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE transactions on smart grid*, 7(5), 2437-2447.

Yao, C. W., & Tuliao, M. D. (2019). Soft skill development for employability: A case study of stem graduate students at a Vietnamese transnational university. *Higher Education, Skills and Work-Based Learning.*

Yildirim, P., & Birant, D. (2017). K-Linkage: A New Agglomerative Approach for Hierarchical Clustering. *Advances in Electrical and Computer Engineering*, 77. doi:10.4316/AECE.2017.04010