COMPILER
Vol. 13, No. 1, May 2024, pp. 45~54
ISSN: 2252-2403 (Print), ISSN: 2549-2403 (Online)
Accredited Fourth Grade by Kemenristekdikti, Decree No:200/M/KPT/2020
DOI: 10.28989/compiler.v13i1.2158

A Comparative Analysis between K-Means and Agglomerative Clustering Techniques in Maritime Skill Certification

Deny Adi Setyawan[1,*], Agustina Purwatiningsih[2]
[1]Department of Nautical Studies, Politeknik Pelayaran Banten, Indonesia
[2]Department of Informatic Engineering, Universitas Duta Bangsa Surakarta, Indonesia

Article Info

Article history:

Received March 9, 2024
Accepted May 9, 2024
Published May 31, 2024

Keywords:

Maritime Skill Certification
Clustering
K-Means
Agglomerative
Data Mining

ABSTRACT

The maritime industry must constantly adjust seafarer training to meet evolving operational demands and ensure compliance with new regulations. This study addresses the challenge of assessing the relevance of Certificate of Proficiency (COP) services by categorizing them to determine which qualifications are essential for marine professionals. The goal is to identify obsolete or misaligned training programs that need updates or enhancements to better serve industry needs. To this end, the study employed two clustering algorithms, K-Means and Agglomerative Clustering, on data from 2021 to 2023. K-Means was chosen for its efficiency in processing large datasets and creating clear, non-overlapping groups. Agglomerative Clustering was selected for its ability to offer a detailed, hierarchical view of data, which helps in understanding the complex structure of certification demands more comprehensively. The analysis identified three main clusters; notably, Cluster 2 indicated a high demand for critical certifications, while Cluster 1, containing the majority of certifications, received little interest, suggesting they may be less relevant. This insight encourages training providers to consider refining their offerings. Although comprehensive, the study's three-year timeframe suggests extending this period in future research for a more detailed trend analysis and forecasting in maritime training adaptations.

Corresponding Author:

Deny Adi Setyawan,
Department of Nautical Studies,
Politeknik Pelayaran Banten,
Jl. Raya No.1, Karang Serang, Kec. Sukadiri, Kabupaten Tangerang, Banten, Indonesia
Email : denyadi333@gmail.com

1. INTRODUCTION

The maritime industry is critical to global trade, acting as the foundation of international trade by facilitating the movement of commodities and people across oceans. The industry's complexity and inherent dynamic necessitate a steadfast commitment to navigation safety and operational efficiencies. Within this framework, sailors' proficiency and preparedness are critical, as they are responsible for the safe and efficient administration of ship operations. The Certificate of Proficiency (COP) is a testimonial to a seafarer's talents, demonstrating they have the skills and knowledge required to function efficiently in difficult conditions. These certifications are more than just a formality; they are critical to sustaining the high safety standards that the industry values [1].

The COP consists of a series of training modules, including Basic Safety Training (BST) and Survival Craft and Rescue Boat (SCRB), that are methodically intended to prepare crew members for a wide range of maritime scenarios, from daily operations to emergency response. The strict criteria established by the International Convention on Criteria of Training, Certification, and Watchkeeping for Seafarers (STCW) regulate this complete training approach. The convention is critical to ensuring maritime safety and environmental stewardship by establishing minimum certification criteria for seafarers [2], [3]. By meeting and exceeding these regulatory benchmarks, the COP not only complies with legislative mandates but also instils trust in the competency of maritime professionals among shipping companies, insurers, and regulatory bodies, resulting in smoother operations and fewer maritime mishaps [4].

In recent years, the proliferation of big data and the advancement of advanced analytics have created new opportunities for the maritime industry to monitor and enhance seafarer training and deployment. Cluster analysis, using approaches like K-Means and Agglomerative Clustering, has developed as an important analytical tool. It enables the analysis of large datasets related to seafarer certificates, finding existing trends and diagnosing skill shortages. This type of study helps maritime training schools connect their curricula with the industry's changing expectations, ensuring that the training provided is both relevant and effective [5].

A survey of the existing literature demonstrates the several applications of K-Means Clustering in maritime management. For example, [6] demonstrated how Using K-means to determine functional needs. This technology uses automatic identification system (AIS) data from a reference ship to generate a representative operational profile, which can help decision-makers identify the functional requirements of ships performing similar missions to the reference ship. Another study by [7] This study uses McQueen k-means to discover the best placement for container storage and distribution centres (CSDC) in order to improve the efficiency of the container transportation network. In addition, to improve clustering results, employ K-Means and Agglomerative clustering. To determine the number of quality clusters suited for a data set with an optimal k value, the agglomerative hierarchical clustering (AHC) algorithm and the K-means method were utilized [8]. Another study compared the agglomerative, k-means, and advanced k-means methods for an RFM (Recency, Frequency, and Monetary)--based market segmentation methodology [9].

In this context, the current study seeks to undertake a comparative analysis of the efficacy of K-Means and Agglomerative Clustering in assessing seafarer certification patterns. K-Means and agglomerative clustering were used in this study because they offer complementary benefits for data processing. While agglomerative clustering gives flexibility in distance measures and is sensitive to more complicated data structures with its hierarchical representation through dendrograms, K-Means is good for handling huge datasets with fast convergence and well-defined clusters. Combining these two techniques improves the accuracy and dependability of clustering results by enabling cross-validation and more diverse viewpoints, especially when working with heterogeneous and unevenly dispersed data. This study aims to determine popular training preferences, identify potential skill gaps, and assess the alignment of present training programs with market expectations. The study hopes to gain valuable insights into the development of more robust training strategies and human resource management practices in the maritime sector, with the ultimate goal of improving maritime safety and operational efficiency as we face the challenges of the twenty-first century [10].

2. RESEARCH METHOD

In this research, the Maritime Skills Certification clustering method is divided into several stages, namely data collection, preprocessing, clustering, cluster validation, cluster analysis by comparing the results of K-Means clusters and Agglomerative clusters, and cluster visualization.. Figure 1 shows these steps.

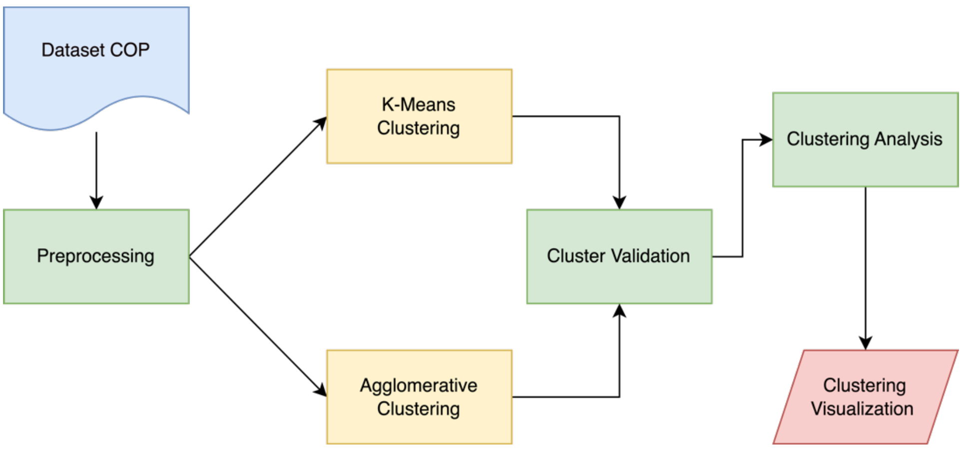

Figure 1. Research stages

2.1. Dataset

The information used in this study is an overview of seafarers' COP services from 2021 to 2023, including 36 monthly data points from recapitulations during that time. There are 41 distinct certification services available in total, encompassing a range of training specialities that correspond to the competencies needed by seafarers to guarantee efficiency, safety, and adherence to maritime laws. Every service includes specialist knowledge for managing particular kinds of vessels and commodities, as well as operational and fundamental safety abilities. The number of COP registrants each month and the amount of the amount of the monthly registration fee are other variables included in this research. Table 1 offers a thorough and well-structured summary of the many kinds of COP services.

Table 1. Certificate of Proficiency Service Data

No.	Certificate of Proficiency Service	Jan 20211	Feb 2021	…	Des 2023
1	Basic Safety Training	399	143		529
2	Arpa Simulator	12	10		139
3	Radar Simulator	29	5		286
4	Proficiency In Survival Craft And Rescue Boats	263	130		434
5	Proficiency In Fast Rescue Boats	0	0		0
6	Advanced Fire Fighting	256	73		617
7	Medical First Aid	38	24		467
8	Medical Care On Board Ship	26	39		587
9	Tanker Familiarization	0	0		0
10	Oil Tanker Specialized Training Programme	0	0		0
11	Liquified Gas Tanker Specialized Training Programme	0	0		0
12	Crowd Management	0	19		57
13	Basic Safety Training for Domestic and Fishing Vessels	0	79		285
14	Crisis Management And Human Behaviour	0	11		81
15	Bridge Resource Management	48	20		273
16	Ship Security Officer	60	54		323
17	Engine Room Simulator	0	0		0
18	Engine Room Resource Management	65	39		136
19	Operational Use Of Ecdis Training Programme	7	10		159
20	Dangerous, Hazardous Harmful Cargoes (Imdg Code) Training Programme	3	4		21
21	Security Awareness Training	24	51		419
22	Seafarers With Designated Security Duties	6	25		168
23	Ratings Forming Part Of A Navigational Watch	1	1		18
24	Ratings As Able Seafarer Deck	0	1		25
25	Ratings Forming Part Of A Watch In Engine Room	2	0		9
26	Basic Training For Liquefied Gas Tanker Cargo Operations	14	1		46
27	Advanced Training For Liquefied Gas Tanker Cargo Operations	0	3		12
28	Basic Training For Oil And Chemical Tanker Cargo Operations	30	28		223
29	Advanced Training For Chemical Tanker Cargo Operations	0	25		28
30	Advanced Training For Oil Tanker Cargo Operations	0	0		92
31	Ratings As Able Seafarer Engine	0	4		10
32	Marine High Voltage	0	0		0
33	Electro Technical Ratings	0	0		0
34	Basic Code of Safety for Ship Using Gas or Other Low Flashpoint Fuels	0	0		0
35	Advance Code of Safety for Ship Using Gas or Other Low Flashpoint Fuels	0	0		0
36	Leadership Training	0	0		0
37	Marine Environment Awareness Training	0	0		0
38	The Electro Technical Rating Program	0	0		0
39	Maritime English	0	0		0
40	Reprint of COP	0	0		34
41	Proficiency in Goc for the GMDSS	0	0		0

2.2. Preprocessing

Preprocessing is an important stage in data analysis, particularly when working with algorithms that require numerical input, such as cluster analysis. In this study, encoding techniques are employed to convert categorical data into numeric format. This transformation allows categorical data, which is initially incompatible with many machine learning algorithms, to be effectively processed. Methods such as One-Hot Encoding and Label Encoding ensure that each category in the dataset is analysable, facilitating more accurate and comprehensive analysis [11]. This step is critical for algorithms that cannot parse text data, such as K-Means or Agglomerative Clustering, which are widely employed to analyse clusters within data. Each service category, such as Basic Safety Training, Radar Simulator, and Survival Craft Proficiency, is allocated a unique number identity that the algorithms can process.

Furthermore, the presence of missing values in a dataset might distort or prejudice the analytic results. As a result, it is normal practice to manage missing values before proceeding with the analysis. One typical strategy is to eliminate records with missing values, particularly if the missing data accounts for a tiny proportion of the dataset and its removal has no substantial influence on the dataset's integrity or analysis findings [12].

2.3. Clustering

The next step after establishing the ideal number of clusters using the Elbow Method and Dendrogram analysis is to carry out clustering. This grouping is carried out utilizing two main algorithms: K-means grouping and Agglomerative Clustering [13]. After clustering using both techniques, the results are compared to determine which algorithm produces the most coherent and interpretable cluster formations.

2.3.1. K-Means Clustering

K-means Clustering is a partitioning strategy that splits data into k non-overlapping subsets (or clusters) that lack internal structure [14], [15]. K-means clustering involves the following steps:

- Randomly generate k centroids from data points.
- Form k clusters by assigning each data point to the nearest centroid using Euclidean distance.
- Recalculate the centroid of each cluster.
- Repeat steps 2 and 3 until the centroids stabilize (no substantial movement in location or predefined number of rounds).

The primary formula (1) for K-means is the calculation of the Euclidean distance, which is used to assign data points to clusters [16]:

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^{n} (x_j + c_j)^2} \quad (1)$$

d = distance
n = number of objects
j = (starting from 1 to n)
x_j = feature of the j object regarding x
c_j = centroid of the j feature

2.3.2. Agglomerative Clustering

Agglomerative Clustering is a hierarchical clustering technique that creates a hierarchy of clusters through a series of iterative mergers [17], [18]. The steps include the following:

1. Consider each data point as a cluster.
2. Combine the closest (most similar) clusters to create a single cluster.
3. Recompute the distances (similarities) between the new and old clusters.
4. Repeat steps 2 and 3 to group all data points into a single cluster.

The algorithm utilized in this study is average linkage. Average linkage is an agglomerative grouping of things based on their average. Using a formula (2) that can be expressed as follows [19]:

$$d_{(AB)C} = \frac{d_{(AC)} + d_{(BC)}}{n_{(AB)} n_C} \quad (2)$$

Where $n_{(AB)}$ is the number of members in clusters A and B, and n_C is the number of members in cluster C.

2.4. Elbow Method

Clustering is the process of grouping a set of objects so that those in the same group, or cluster, are more similar to one another than to those in other groups [20]. One of the most important tasks in clustering analysis is finding the best number of clusters to capture the data's intrinsic structure [21]. The Elbow Method is especially effective when dealing with K-Means clustering. This method entails graphing the explained variance against the number of clusters and selecting the elbow of the curve as the number of clusters to employ [13], [22]. The essential premise is to select a number of clusters such that adding another cluster does not result in significantly better data modelling. More formally, the method examines the percentage of variance explained as a function of the number of clusters: one should select a number of clusters such that adding another cluster does not significantly improve data modelling. This method is frequently represented graphically by the number of clusters on the x-axis and the within-cluster sum of squares (WCSS) on the y-axis [22].

2.5. Dendrograms

Dendrograms are an important technique for hierarchical grouping. They serve to depict the arrangement of the clusters generated throughout the analysis. Researchers can use the dendrogram to determine the number of clusters by searching for a substantial jump in the distance between successive merging or splitting [13]. This distance measures the dissimilarity of clusters, and a big rise indicates that the next merging or split may combine clusters that are less similar than those within the present ones [23].

3. RESULTS AND ANALYSIS

3.1. Cluster Validation

The experimental results from the Elbow Method show that three clusters are the most optimal number for the dataset. When using the Elbow Method, the experiment involved examining a variety of possible clusters ranging from 1 to 15. This method calculates the within-cluster sum of squares (WCSS) for each conceivable number of clusters and searches for the point at which the reduction in WCSS begins to decelerate, showing decreasing returns on the explanatory power of more clusters. In this situation, the WCSS plot showed the 'elbow' at k=2.

Although the Elbow method suggested that two (2) clusters be used in this study, we decided to use three (3) clusters based on the analysis of the Distortion Scores, which demonstrated a significant reduction upon the addition of a third cluster, indicating improved homogeneity and proximity among the data within that cluster. The analytical requirement to pinpoint more precise subtypes or dataset segments that are pertinent to the goals of the investigation also supports this choice [24]. To provide deeper and more useful insights through more exact data segmentation, three clusters were chosen. This ensures that the analysis more closely represents the underlying structure and dynamics of the data under study. Figure 2 shows the results of the elbow approach.

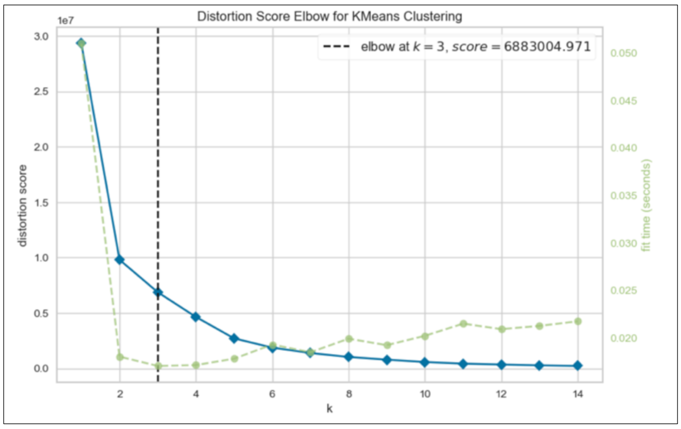

Figure 2. The distortion Score elbow for K-Means Clustering

The Dendrogram, a technique used in hierarchical clustering, visualized the data merging process. Initially, each data point was treated as a separate cluster at the diagram's base. As the algorithm progressed, these clusters were merged based on similarity, which is frequently assessed using distance metrics like Euclidean distance. The Dendrogram's structure revealed that the data naturally formed three separate clusters. This was visible as a point on the diagram where merging any farther would result in a considerable increase in the distance between freshly generated clusters, indicating a natural divide in the data. Figure 3 shows the dendrogram method's results.

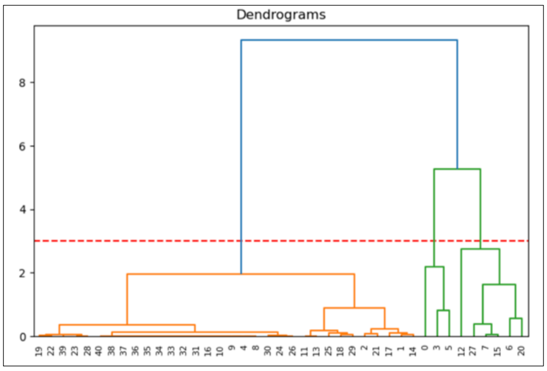

Figure 3. The result of the dendrograms method

Both the Elbow Method and the Dendrogram have thus reached the conclusion that three clusters provide the most relevant and interpretable categorization of the data, striking a balance between too-fine-grained and too-coarse grouping. This conclusion lays the groundwork for the following use of clustering techniques such as K-means and Agglomerative Clustering to divide the data into three different clusters for further investigation.

3.2. Clustering Analysis

The clustering findings from the K-Means and Agglomerative Clustering algorithms show parallels in the overall distribution of data into three major clusters. However, following closer analysis, the two approaches differ slightly in the attribution of specific COP services to these clusters. Figure 4 shows the clustering result.

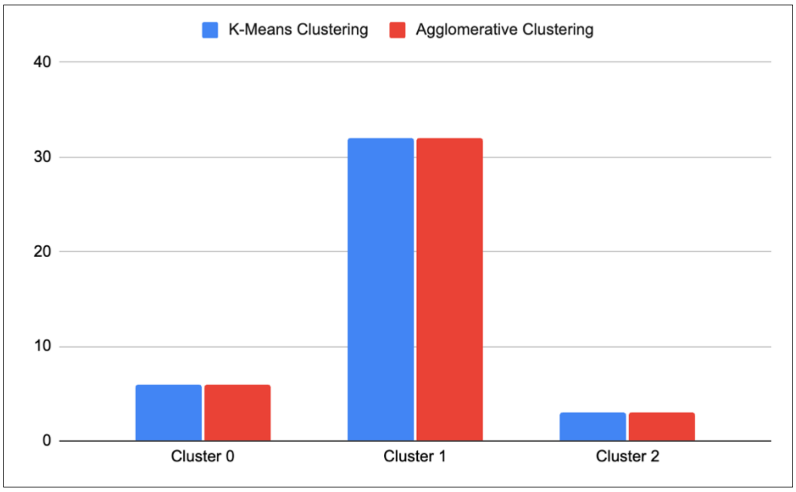

Figure 4. The clustering result

The 'Basic Safety Training for Domestic and Fishing Vessels' certification is placed in Cluster 0 of the K-Means clustering output. 'Proficiency in survival craft and rescue boats', on the other hand, is classified as Cluster 1. This arrangement implies that the K-Means algorithm associates 'Basic Safety Training for Domestic and Fishing Vessels' with a group of certifications with a certain level of enrolee interest, which could indicate certifications with a broader or more general appeal within the maritime sector. Table 2 shows the cluster data of COP.

Table 2. Cluster data of COP

No	Certificate of Proficiency Service	K-Means Cluster	Aglomerative Cluster
1	Basic Safety Training	2	2
2	Advanced Fire Fighting	2	2
3	Basic Safety Training for Domestic and Fishing Vessels	2	0
4	Proficiency In Survival Craft And Rescue Boats	0	2
5	Security Awareness Training	0	0
6	Medical First Aid	0	0
7	Medical Care On Board Ship	0	0
8	Ship Security Officer	0	0
9	Basic Training For Oil And Chemical Tanker Cargo Operations	0	0
10	Seafarers With Designated Security Duties	1	1
11	Radar Simulator	1	1
12	Bridge Resource Management	1	1
13	Engine Room Resource Management	1	1
14	Arpa Simulator	1	1
15	Crisis Management And Human Behaviour	1	1
16	Crowd Management	1	1
17	Basic Training For Liquefied Gas Tanker Cargo Operations	1	1
18	Operational Use Of Ecdis Training Programme	1	1
19	Advanced Training For Oil Tanker Cargo Operations	1	1
20	Ratings As Able Seafarer Deck	1	1
21	Reprint of COP	1	1
22	Dangerous, Hazardous Harmful Cargoes (Imdg Code) Training Programme	1	1
23	Advanced Training For Chemical Tanker Cargo Operations	1	1
24	Ratings Forming Part Of A Navigational Watch	1	1
25	Ratings As Able Seafarer Engine	1	1
26	Advanced Training For Liquefied Gas Tanker Cargo Operations	1	1
27	Ratings Forming Part Of A Watch In Engine Room	1	1
28	Proficiency In Fast Rescue Boats	1	1
29	Tanker Familiarization	1	1
30	Oil Tanker Specialized Training Programme	1	1
31	Liquified Gas Tanker Specialized Training Programme	1	1
32	Engine Room Simulator	1	1
33	Marine High Voltage	1	1

No	Certificate of Proficiency Service	K-Means Cluster	Aglomerative Cluster
34	Electro Technical Ratings	1	1
35	Basic Code of Safety for Ship Using Gas or Other Low Flashpoint Fuels	1	1
36	Advance Code of Safety for Ship Using Gas or Other Low Flashpoint Fuels	1	1
37	Leadership Training	1	1
38	Marine Environment Awareness Training	1	1
39	The Electro Technical Rating Program	1	1
40	Maritime English	1	1
41	Proficiency in Goc for the GMDSS	1	1

In contrast, the Agglomerative Clustering technique allocates 'Basic Safety Training for Domestic and Fishing Vessels' to Cluster 1 and 'Proficiency in Survival Craft and Rescue Boats' to Cluster 0. This inversion indicates that the hierarchical nature of Agglomerative Clustering sees a stronger relationship or similarity between 'Proficiency In Survival Craft And Rescue Boats' and the other certifications in Cluster 0, which could reflect a different aspect of enrolee interest or the nature of the certifications themselves. Figure 5 depicts the data distribution for the K-Means and Agglomerative clusters.

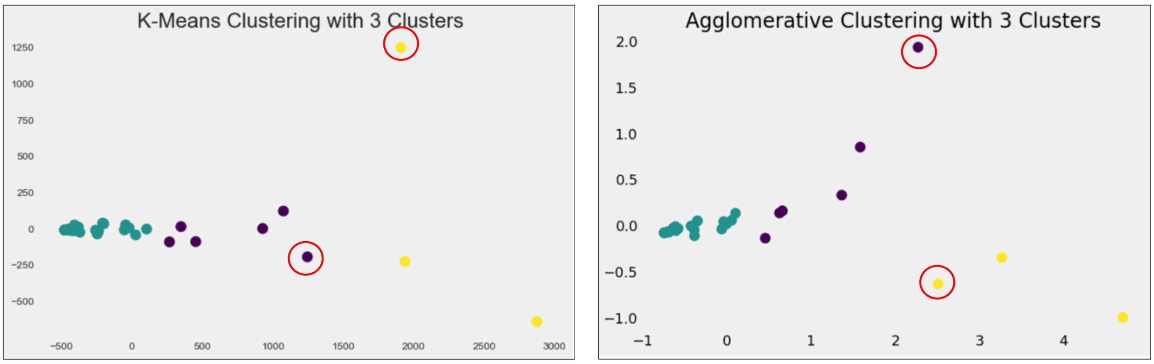

Figure 5. Cluster Visualization

These differences emphasize the distinct traits and factors that each algorithm takes into account during the clustering process. K-Means focuses on centroid-based similarity, which can lead to a different grouping than Agglomerative Clustering, which creates clusters based on a hierarchy of links between data points. Such differences in clustering outcomes highlight the need of exploring several clustering algorithms in order to acquire a thorough knowledge of data structures and linkages within the dataset [25]. These findings highlight that, whereas clustering algorithms frequently arrive at identical conclusions, the specifics of their assignments might differ due to the various approaches they use. This variance is critical for data analysts and decision-makers because it affects the interpretation of data clusters as well as subsequent actions or suggestions based on clustering results [26].

The clustering study has neatly classified the seafarer certification data into three separate clusters, providing substantial insights regarding applicant enrolment trends for various COP services during the three-year period from 2021 to 2023. Cluster 2 has the greatest enrolment, with over 9,000 candidates in the provided timeframe. The qualifications in this cluster are clearly in great demand, indicating their crucial importance to a substantial segment of the maritime profession. This cluster is likely to include critical qualifications required for a variety of maritime positions, as well as those highly prized for career advancement within the industry. Given their popularity, it is critical to focus on improving the delivery and quality of these certifications in order to maximize service efficiency and handle the large number of candidates.

Cluster 0 is a moderately popular set of certificates, with each seeing upwards of 4,000 applicants. While the demand for these certificates is high, there is still space for expansion. Improvements in the quality and delivery of certifications in this cluster may boost their desirability and lead to increased enrolment numbers. This could include changing the curriculum to meet current industry standards, including more hands-on training, or improving the entire learning experience. Cluster 1 contains certificates with little applicant interest, containing 14 categories of certifications that have drawn few or no applicants. The low enrolment figures necessitate a detailed examination of these certificates. The question is whether they continue to play an important role in the maritime business or have become obsolete. Decisions must be taken on whether or not to continue offering these services. To evaluate the lack of interest and assess the future of these certifications, a market analysis and feedback from industry stakeholders may be required. In some circumstances, renaming, repurposing, or combining them with more widely used certificates may help to better match them with industry demands.

The results of the cluster analysis show that the clusters provide a clear indicator of the levels of demand for different seafarer certifications. The insights acquired from this analysis will help training providers manage resources more effectively, improve program offerings, and eventually match with the marine industry's strategic objectives. It is critical that the proper authorities and institutions interpret and act on these findings in order to retain a qualified, competent, and suitably certified maritime workforce.

4. CONCLUSION

The researchers successfully divided the preferences for seafarer certification into three distinct clusters using the K-Means and Agglomerative clustering algorithms. During a three-year period, Cluster 2, comprising three qualifications, had the highest demand with over nine thousand registrants, indicating the importance of maritime certificates. There was a possibility for growth with better quality and service in Cluster 0, which had six certificates and a moderate demand. With 32 certificates, Cluster 1 had the most, yet there was little interest in them, suggesting that their value should be re-evaluated. Combining approaches is necessary for comprehensive data analysis, as evidenced by the similarities and differences in grouping assignments between the K-Means and Agglomerative methods. However, because this study only examined data from the last three years, it has limitations. A larger dataset spanning more than five years would be advantageous, allowing future studies to conduct forecasting analysis to decide which certificates should be expanded further.

ACKNOWLEDGEMENTS

The author would like to thank the Research and Community Service (P3M) Banten Merchant Marine Polytechnic of Agency of Human Resource Development of the Ministry of Transportation for the support provided to the author in the form of research funding assistance in 2024 of Penelitian Dosen Pemula (PDP) scheme.

REFERENCES

[1] D. Vassalos, “A Pioneer of Maritime Safety,” *Ships and Offshore Structures*, pp. 462–467, 2020, doi: 10.1080/17445302.2020.1787598.

[2] C. D. Mantoju, “Analysis of impact of the maritime labour convention, 2006: A seafarer’s perspective,” *Journal of International Maritime Safety, Environmental Affairs, and Shipping*, vol. 5, no. 3, pp. 107–119, Jul. 2021, doi: 10.1080/25725084.2021.1955475.

[3] M. L. Fotteler, D. Andrioti Bygvraa, and O. C. Jensen, “The impact of the Maritime Labor Convention on seafarers’ working and living conditions: an analysis of port state control statistics,” *BMC Public Health*, vol. 20, no. 1, pp. 1–9, Dec. 2020, doi: 10.1186/S12889-020-09682-6/FIGURES/4.

[4] T. Myint, “AN ANALYSIS OF PROCESS AND DEVELOPMENT OF MYANMAR MARITIME EDUCATION & TRAINING SYSTEM WHICH IMPACTS ON MYANMAR SEAFARER’S EMPLOYMENT OPPORTUNITIES,” *International Journal on Recent Trends in Business and Tourism*, vol. 3, no. 2, pp. 26–30, 2019.

[5] E. Smith Johnson and N. Immanuel Paul, “Exploring the sustainability of specialized higher education curricula as exemplified by Jamaica’s maritime education and training curriculum,” *Journal of Applied Research in Higher Education*, vol. ahead-of-print, no. ahead-of-print, 2024, doi: 10.1108/JARHE-08-2023-0375/FULL/PDF.

[6] J. Park and M. Choi, “A K-Means Clustering Algorithm to Determine Representative Operational Profiles of a Ship Using AIS Data,” *J Mar Sci Eng*, vol. 10, no. 9, Sep. 2022, doi: 10.3390/jmse10091245.

[7] O. Moskvichev, E. Moskvicheva, and A. Bulatov, “Clustering Methods for Determination of Optimal Locations of Container Storage and Distribution Centers,” in *Transportation Research Procedia*, Elsevier B.V., 2021, pp. 461–469. doi: 10.1016/j.trpro.2021.02.096.

[8] P. Patel, B. Sivaiah, and R. Patel, “Approaches for finding Optimal Number of Clusters using K-Means and Agglomerative Hierarchical Clustering Techniques,” in *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, IEEE, Jul. 2022, pp. 1–6. doi: 10.1109/ICICCSP53532.2022.9862439.

[9] S. H. Shihab, S. Afroge, and S. Z. Mishu, “RFM Based Market Segmentation Approach Using Advanced K-means and Agglomerative Clustering: A Comparative Study,” *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, Apr. 2019, doi: 10.1109/ECACE.2019.8679376.

[10] E. Impact, “Global Maritime Trends 2050,” London, 2023.

[11] N. Ashokkumar, B. Redddiprasanna, B. Varshitha, S. Farishma, P. H. Shankar, and K. Thandapani, “Data Encoding Techniques for Reducing Energy Consumption in Network-On-Chip,” *ViTECoN 2023 - 2nd IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies, Proceedings*, 2023, doi: 10.1109/VITECON58111.2023.10157433.

[12] M. R. Maarif, A. R. Saleh, M. Habibi, N. L. Fitriyani, and M. Syafrudin, “Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI),” *MPDI*, vol. 14, no. 265, 2023.

[13] D. A. Jethva, S. M. Khamkar, A. A. Pachchhapur, and S. Kulkarni, “Customer Segmentation Analysis using K-means Algorithm with Elbow method and Dendrogram,” in *2023 6th International Conference on Advances in Science and Technology (ICAST)*, IEEE, Dec. 2023, pp. 296–301. doi: 10.1109/ICAST59062.2023.10454952.

[14] C. Hayat, "Enhanced K-Means Clustering Approach for Diagnosis Types of Acne," in *2021 2nd International Conference on Innovative and Creative Information Technology, ICITech 2021*, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 39–43. doi: 10.1109/ICITech50181.2021.9590124.

[15] U. A. Nasron and M. Habibi, "Analysis of Marketplace Conversation Trends on Twitter Platform Using K-Means," *Compiler*, vol. 9, no. 1, pp. 51–61, 2020, doi: 10.28989/compiler.v9i1.579.

[16] K. Kusumaningtyas *et al.*, "Tweet Analysis of Mental Illness Using K-Means Clustering and Support Vector Machine," *Telematika : Jurnal Informatika dan Teknologi Informasi*, vol. 20, no. 3, pp. 295–308, Feb. 2024, doi: 10.31315/TELEMATIKA.V20I3.9820.

[17] M. Wati, D. Adela, and Muh. Jamil, "Implementation of Hierarchical Agglomerative Clustering Method to East Kalimantan Unemployment Analysis," Institute of Electrical and Electronics Engineers (IEEE), Jan. 2024, pp. 395–399. doi: 10.1109/icitisee58992.2023.10405078.

[18] Z. Cai, R. Li, and H. Wu, "Truncated representation graph with adaptive weighted and manifold regularization for agglomerative clustering," in *2022 5th International Conference on Data Science and Information Technology, DSIT 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/DSIT55514.2022.9943817.

[19] M. Bibi, W. Aziz, M. Almaraashi, I. H. Khan, M. S. A. Nadeem, and N. Habib, "A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis," *IEEE Access*, vol. 8, pp. 68580–68592, 2020, doi: 10.1109/ACCESS.2020.2983859.

[20] S. Wang and H. Li, "Adaptive K-valued K-means clustering algorithm," in *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 1442–1445. doi: 10.1109/ICMCCE51767.2020.00316.

[21] J. Chen, L. Zhao, M. Zhou, Y. Liu, and S. Qin, "An Approach to Determine the Optimal k-Value of K-means Clustering in Adaptive Random Testing," in *Proceedings - 2020 IEEE 20th International Conference on Software Quality, Reliability, and Security, QRS 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 160–167. doi: 10.1109/QRS51102.2020.00032.

[22] D. Marutho, S. Hendra Handaka, and E. Wijaya, "The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News," in *2018 International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia: IEEE, Oct. 2018. doi: 10.1109/ISEMANTIC.2018.8549751.

[23] F. Bajaber, "Enhancing Subcluster Identification in IoT Sensor Networks with Hierarchical Clustering Algorithms and Dendrograms," in *8th International Conference on Software Engineering and Computer Systems, ICSECS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 459–464. doi: 10.1109/ICSECS58457.2023.10256413.

[24] N. Jalali, N. K. Tran, A. Sen, and P. P. Morita, "Identifying the Socioeconomic, Demographic, and Political Determinants of Social Mobility and Their Effects on COVID-19 Cases and Deaths: Evidence From US Counties," *JMIR Infodemiology*, vol. 2, no. 1, Jun. 2022, doi: 10.2196/31813.

[25] H. I. Abdalla, "A Brief Comparison of K-means and Agglomerative Hierarchical Clustering Algorithms on Small Datasets," *Lecture Notes in Electrical Engineering*, vol. 942 LNEE, pp. 623–632, 2022, doi: 10.1007/978-981-19-2456-9_64/TABLES/12.

[26] K. B, "A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 1600–1604, May 2020, doi: 10.30534/ijeter/2020/20852020.