# Clustering Cadet Training Performance Using K-Means and Ward's Method Evidence from OTMon Maritime Monitoring System

Dian Wahdiana[1,*], Ekohariadi[2], Ratna Suhartini[3]

[1,2,3]*Vocational Education Study Program, State University of Surabaya, East Java, Indonesia*

(Received: March 18, 2025; Revised: June 08, 2025; Accepted: August 29, 2025; Available online: September 26, 2025)

**Abstract**

This study investigates cadet performance segmentation during on-board maritime training using clustering analysis of data from the On Training Monitoring (OTMon) system. Grounded in the competency-based education framework and experiential learning theory, the research aims to identify behavioral patterns and competency levels among 80 maritime cadets over a twelve-month sea-based training program. The OTMon application continuously recorded task completion rates, feedback interactions, sign-on consistency, and report submissions. K-Means clustering and Principal Component Analysis (PCA) revealed three distinct cadet profiles: Cluster 1 (high-performing) with average task completion of 92.4% and feedback frequency of 15.2 times/month; Cluster 2 (administratively consistent) with 88.1% completion but only 6.3 feedback interactions/month; and Cluster 3 (at-risk) with 67.5% completion and 3.8 feedback interactions/month. Linear Discriminant Analysis (LDA) validated the clusters with 98.8% resubstitution accuracy and 97.6% cross-validation accuracy, supported by generalized squared distances above 9.5 between all cluster pairs, indicating strong separation. These findings demonstrate that unsupervised clustering can reliably distinguish high-performing cadets from those needing targeted intervention, enabling data-informed mentoring and adaptive learning strategies in maritime education. The contribution of this study lies in integrating digital monitoring data with both unsupervised and supervised machine learning methods to enhance competency assessment. The novelty is in applying maritime-specific learning analytics for real-time performance segmentation, offering a scalable diagnostic framework for improving supervision quality and supporting individualized cadet development in vocational training contexts.

*Keywords:* Cadet Performance, Clustering Analysis, Learning Analytics, Maritime Education, On-Board Training

## 1. Introduction

In the rapidly evolving landscape of maritime education, the effectiveness of On-Board Training (OBT) [1], [2], [3] has become a critical determinant of cadet competency and career readiness. As part of vocational education, OBT serves as a real-world application of classroom theory, enabling cadets to acquire technical skills, adaptive behavior, and problem-solving capabilities essential for maritime operations. However, traditional monitoring systems used during this critical phase remain largely manual, fragmented, and incapable of providing timely or personalized feedback, limiting the potential of experiential learning [4], [5], [6].

Recent studies underscore the increasing relevance of digital learning tools in competency-based education. For instance [7], [8], [9] emphasize the value of integrating real-time monitoring systems to enhance engagement and supervision quality [10], [11], [12]. Moreover, educational technology powered by data analytics and mobile accessibility has proven effective in improving student performance in various fields [13], [14], [15]. Despite these advances, there is a lack of scholarly work that explicitly applies unsupervised machine learning techniques particularly clustering to analyze cadet training behaviors in maritime education.

To address this gap, this study employs clustering analysis specifically K-Means and Ward's hierarchical method on digital data collected via the OTMon application. The selected clustering variables include feedback frequency, task completion rates, and self-reported difficulties, as these are directly linked to key OBT learning outcomes and reflect measurable behavioral indicators. The rationale for using both K-Means and Ward's methods lies in their

***Corresponding author: Dian Wahdiana (dian.22009@mhs.unesa.ac.id)**
**DOI: https://doi.org/10.47738/jads.v5i2.XXX**

complementary strengths: K-Means is effective for large, spherical clusters, while Ward's method enables a hierarchical structure that supports exploratory pattern discovery. Furthermore, while OTMon automates feedback collection, the operational definition of “feedback” in this study refers to cadet–supervisor interactions logged digitally, such as task comments or performance notes, rather than peer input or system-only notifications [16], [17], [18].

Unlike existing literature in maritime education, this research brings novelty by empirically applying clustering methods for the segmentation of performance profiles, supported by clearly defined variables such as task completion, feedback frequency, and difficulty levels, which are operationalized within the OTMon system. Although similar applications of clustering are found in other sectors such as aviation and logistics, the maritime sector has yet to adopt such methods at scale, making this research both timely and necessary. The main objective of this research is to utilize data-driven clustering methods to generate performance-based profiles that can inform targeted educational interventions and enhance supervisory strategies in maritime training programs. Theoretically, this study contributes to the field of learning analytics by applying machine learning for educational diagnostics in vocational settings. Practically, it provides actionable insights for maritime institutions to identify student needs, improve mentoring systems, and optimize learning outcomes [19], [20], [21].

This research adopts a quantitative approach involving 80 cadets enrolled in a one-year OBT program. The data, collected via OTMon including variables such as task completion rates, feedback frequency, and reported difficulties are preprocessed using Z-score normalization before being analyzed with both K-Means and Ward’s Hierarchical Clustering. Principal Component Analysis (PCA) is employed not merely as an optional tool, but as a key dimensionality reduction technique to enhance cluster interpretability and visualization [22], [23], [24]. To increase methodological clarity, a process flowchart is presented to illustrate the steps from data collection to clustering validation.

In sum, this study introduces a novel, data-driven framework for cadet performance segmentation in maritime education and advocates for the broader use of machine learning as a diagnostic tool in digitalized vocational training environments [25], [26], [27].

## 2. Literature Review

The purpose of this literature review is to examine the current body of knowledge related to digital monitoring in vocational education, learning analytics in maritime training, and the use of clustering techniques to evaluate student performance. This section synthesizes relevant studies over the past decade, highlights existing gaps, and situates the present study within the context of emerging research trends. The review is structured thematically, focusing on four key areas: (1) digital transformation in maritime education [28], (2) learning analytics and feedback systems [29], [30], [31], (3) the role of clustering in education [32], [33], [34], and (4) research gaps and opportunities for innovation [35], [36], [37].

### 2.1. Digital Transformation in Maritime Education

The shift toward digitalization in vocational and maritime education has been gaining momentum in recent years. Technologies such as simulation-based training, e-Logbooks, and online competency tracking systems have improved transparency and responsiveness in education delivery [38], [39], [40]. Applications like OTMon are examples of how monitoring systems can support continuous assessment and provide timely feedback in real-world training environments. Digital monitoring tools not only streamline administrative processes but also enhance pedagogical engagement and student accountability [41], [42], [43]. However, most digital monitoring tools remain focused on basic reporting functions and lack integration with advanced data analysis techniques, limiting their effectiveness in identifying deeper learning patterns.

### 2.2. Learning Analytics and Feedback Systems

Learning analytics has emerged as a promising field for evaluating educational processes and outcomes. Learning analytics involves the collection, measurement, and analysis of learner data to improve educational environments [44], [45], [46]. In maritime training, feedback systems, especially real-time ones, play a crucial role in enhancing cadet development. Studies have found that platforms providing structured, timely feedback significantly improved student motivation and learning outcomes [47], [48], [49]. Nevertheless, much of the research in this area focuses on synchronous classroom settings, leaving on-board and field-based training contexts underexplored.

### 2.3. Clustering Techniques in Educational Evaluation

Clustering is a data mining technique that allows researchers to group individuals based on shared characteristics without predefined labels. In educational settings, K-Means and Hierarchical Clustering have been used to identify student learning profiles, engagement patterns, and performance risk groups [50], [51], [52]. These unsupervised learning techniques are particularly useful when analyzing large, unlabeled educational datasets. Despite their proven value, clustering methods have not been widely applied in maritime vocational education. Most maritime training assessments still rely on standard pre-post evaluations and supervisor ratings, which may not reflect the nuanced development of student competencies over time.

### 2.4. Research Gaps and Contribution of This Study

This review highlights a number of significant gaps in the literature. Learning analytics and machine learning approaches are not well integrated into field-based training systems, especially in marine contexts where competency development relies heavily on real-world, practical learning environments. Moreover, there are still few empirical studies that use clustering algorithms to analyze performance monitoring data from long-term vocational training programs like on-board cadetship. The field of student segmentation and performance profiling based on real-time monitoring data in maritime education is also understudied, which presents a big opportunity for future research to develop data-driven strategies that improve learner support and training efficacy.

This study seeks to address these gaps by applying K-Means and Ward's Hierarchical Clustering to cadet performance data collected via OTMon. By doing so, it introduces a data-driven framework for understanding cadet learning patterns and offers an innovative method for competency evaluation in maritime training. The literature strongly supports the value of digital tools and learning analytics in vocational education. However, maritime training remains an underrepresented domain in this discourse, especially concerning the application of machine learning for educational assessment. This study fills that void by proposing a novel approach to cluster cadet performance using unsupervised learning techniques. The following section details the methodology employed in the analysis, including data collection, variables used, and clustering procedures.

## 3. Methodology

### 3.1. Research Approach and Rationale

This study adopts a quantitative, exploratory design aimed at identifying latent patterns in cadet performance during on-board maritime training. The rationale for using a quantitative approach lies in the objective of the study to uncover meaningful groupings in training data using statistical and machine learning techniques. The focus is on analyzing digital learning behavior and performance data collected via the OTMon application using unsupervised learning methods, particularly K-Means and Ward's Hierarchical Clustering. In contrast to prior research in maritime education, this study introduces clustering analysis as a novel empirical strategy to address the absence of automated performance profiling methods

### 3.2. Research Design

The research design is a non-experimental, cross-sectional study with an emphasis on clustering analysis as a form of exploratory data mining. The study does not manipulate variables but observes and segments data to uncover inherent structures and relationships. The justification for using both K-Means and Ward’s Hierarchical Clustering lies in their complementary strengths: K-Means excels in large-scale partitioning, while Ward’s method helps validate structural cohesion via dendrogram analysis. Cluster analysis is chosen for its ability to identify subgroups within the population without the need for predefined categories, making it ideal for educational performance profiling. Figure 1 illustrates the workflow of this study.

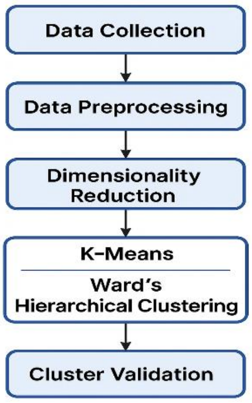


**Figure 1.** Workflow of Clustering Analysis on OTMon Data

## 3.3. Data and Data Sources

80 maritime cadets who took part in a 12-month on-board training program at a maritime vocational institution provided the primary data used in this study. The OTMon tool provided the data, which methodically logs interactions in real time, task submissions, feedback exchanges, and self-reported difficulties faced throughout training. A thorough picture of cadets' performance and engagement during the training period is provided by the variables that are analyzed, which include the weekly average task completion rate, the frequency of feedback received, the reported task difficulty, the response time for report submission, the consistency of sign-on and sign-off records, and the total number of uploaded reports.

These variables were selected based on their relevance to assessing individual learning behavior and digital engagement. However, the exclusion of peer or supervisor evaluations is acknowledged as a limitation, given their potential to enrich performance insights. Data collection was conducted with informed consent, and students were anonymized using unique identifiers within the application.

## 3.4. Data Processing and Analytical Techniques

To guarantee scientific rigor and appropriate interpretation of the data, the analysis procedure was carried out in a number of successive steps. To facilitate comparability among features, Z-score normalization was used to standardize all numerical variables after data cleaning and normalization. Mild outliers were purposefully kept to preserve behavioral diversity that could affect clustering performance, even though the normal distribution assumptions were not rigidly enforced; this choice is further recognized in the study's limitations. Then, instead of being an optional step, PCA was used as a crucial diagnostic and dimensionality reduction tool. While the final clustering was done on the original variables to preserve data purity, the PCA results guided the scree plot analysis and evaluation of component contributions.

After then, two clustering techniques were put into practice. The Elbow Method and Silhouette Coefficient were used to find the ideal number of clusters when K-Means Clustering was used to divide the cadets according to their performance indicators. In order to verify the cluster structure and evaluate inter-group distances through dendrogram examination, Ward's Hierarchical Clustering was also utilized. The Results section provides a comparison of the two approaches, highlighting the similarities and variations in cluster assignments. By computing and comparing the mean scores of each input variable, each group was profiled after clustering. This allowed for the discovery of unique performance patterns, including high achievers, late responders, and disengaged cadets.

## 3.5. Validity and Reliability

The clustering structure was assessed using a number of complementing metrics and techniques to guarantee validity. The Davies–Bouldin Index examined the compactness and separability of clusters, whereas the Silhouette Coefficient was used to evaluate intra-cluster cohesiveness and inter-cluster separation. Cross-validation via bootstrapping, which involved rerunning the algorithms on various sub-samples and comparing the consistency of group assignments, was used to further investigate cluster stability. By avoiding circular interpretation based on the same metrics used for model generation, clustering-validation separation was preserved to reduce the danger of overfitting. The OTMon system's consistent and timestamped logging features, which reduced human subjectivity and made sure the analysis could be reliably replicated between users and sessions, strengthened reliability.

### 3.6. Ethical Considerations

This study adheres to institutional research ethics. Participants provided informed consent when registering in the OTMon application. No Personally Identifiable Information (PII) was used in the analysis. All raw data were anonymized and securely stored on a restricted-access server. Only authorized researchers had access to process and interpret the dataset.

### 3.7. Methodological Limitations

It is important to recognize a number of methodological constraints even if clustering provides insightful information about cadet performance segmentation. Because the sample was taken from a single maritime institution, it may not accurately reflect the variety of training situations, which limits the findings' generalizability. Furthermore, erratic internet connections at sea may cause late or partial submissions, which could compromise the integrity of the data. The lack of a longitudinal dataset makes it even more difficult to evaluate the consistency and stability of clusters across long time periods. Even though segmentation is based on statistical analysis, some subjectivity is unavoidably introduced by the manual interpretation of cluster characteristics. Furthermore, measurement bias could be present in the data produced by the OTMon system due to things like erroneous timestamps, irregular feedback logging, or even user manipulation problems that are hard to identify after the fact.

In order to give more comprehensive and contextualized performance data, future research should broaden the variable set by including peer and supervisor ratings. The ability to record and examine behavioral trajectories over time may be improved by the use of temporal sequence models, such as Long Short-Term Memory (LSTM) networks. External validity and adaptability to various training environments would be enhanced by testing the methodology across a range of marine institutions. Additionally, investigating hybrid clustering strategies that use supervised classification layers may improve segmentation precision and make it possible to model cadet performance patterns predictively.

## 4. Results and Discussion

### 4.1. Data Presentation and Key Findings

This study applied K-Means clustering through the FASTCLUS procedure to group 80 maritime cadets into three clusters based on six standardized training variables derived from the OTMon monitoring system. The clustering process converged in five iterations with a final criterion value of 0.8800 and a Pseudo-F statistic of 10.59, indicating strong group separability and internal cluster coherence. The summary of clustering results is presented in table 1, which outlines the parameters and settings applied during the analysis.

**Table 1.** Cluster Summary Report (Replace=FULL Radius=0 Maxclusters=3 Maxiter=100 Converge=0.02)

| Cluster | Frequency | RMS Std Deviation | Maximum Distancefrom Seedto Observation | RadiusExceeded | Nearest Cluster | Distance BetweenCluster Centroids |
|---|---|---|---|---|---|---|
| 1 | 28 | 0.8836 | 3.1119 | | 3 | 1.9052 |
| 2 | 24 | 0.8662 | 3.5163 | | 1 | 1.9589 |
| 3 | 28 | 0.9352 | 3.1400 | | 1 | 1.9052 |

The clusters were distributed relatively evenly: Cluster 1 (28 cadets), Cluster 2 (24 cadets), and Cluster 3 (28 cadets). Cluster 2 had the highest internal variance with a maximum distance of 3.5163 from centroid to observation. Despite this, no radius violations occurred, confirming cluster cohesion. Table 2 gives a thorough overview of the performance characteristics within each group by displaying the centroids for each variable across the three clusters.

**Table 2.** Cluster Centroids for Each Variable

| Cluster | Task_ Completion | Feedback_ Frequency | Task_ Difficulty | Report_ Timeliness | Sign_On_ Consistency | Total_Reports |
|---|---|---|---|---|---|---|
| 1 | 0.1953377068 | 0.5376561209 | -.7123549686 | -.4962482965 | 0.1974656850 | -.1517910107 |
| 2 | 0.1018093234 | -.7119545701 | -.1110448596 | 0.2549747140 | -.2332207885 | 0.9231413144 |
| 3 | -.2826028411 | 0.0725906534 | 0.8075362768 | 0.2776985417 | 0.0024378480 | -.6394729730 |

These numbers show the cadets' various performance typologies. Cluster 1, which represents the group of high-performing cadets, is distinguished by frequent feedback, high task completion rates, and low reported difficulty. Cadets in Cluster 2 are administratively dependable but may not participate as much in interactive learning exchanges, as seen by their consistent reporting behavior and comparatively minimal feedback. On the other hand, Cluster 3 is identified as at-risk cadets who could need more support because of their higher perceived difficulty levels, lower feedback frequency, and poorer job completion rates. Table 3 offers a thorough statistical overview of the clustering variables, providing more in-depth understanding of the numerical distinctions that characterize each performance group.

**Table 3.** Statistics for Variables in Clustering

| Variable | Total STD | Within STD | R-Square | RSQ/(1-RSQ) |
|---|---|---|---|---|
| Task_Completion | 1.00000 | 0.98986 | 0.044979 | 0.047098 |
| Feedback_Frequency | 1.00000 | 0.87233 | 0.258313 | 0.348278 |
| Task_Difficulty | 1.00000 | 0.77490 | 0.414731 | 0.708616 |
| Report_Timeliness | 1.00000 | 0.94240 | 0.134366 | 0.155222 |
| Sign_On_Consistency | 1.00000 | 0.99742 | 0.030346 | 0.031296 |
| Total_Reports | 1.00000 | 0.77671 | 0.411995 | 0.700666 |
| OVER-ALL | 1.00000 | 0.89698 | 0.215788 | 0.275166 |

According to the analysis, task complexity, the total number of reports, and feedback frequency were the main factors that contributed to clustering, suggesting that these three factors most significantly affect how engaged and proficient cadets are. Proactive involvement, above-average task completion rates, active feedback exchanges, and low perceived difficulty a sign of good competence are characteristics of Cluster 1, which consists of 28 cadets. Cluster 2, which consists of 24 cadets, represents those who routinely fulfill administrative obligations, such submitting reports on time, but show little interpersonal interaction through feedback, indicating a pattern of robotic compliance without meaningful involvement.

Those that suffer the most are represented by Cluster 3, which likewise has 28 cadets. They report significant work difficulty, produce little output, and show little initiative in asking for help. This tripartite division offers a useful framework for focused mentoring techniques, like maintaining Cluster 1's high performance, strengthening the feedback culture for Cluster 2, and providing more academic and practical support for Cluster 3. These results show how OTMon data may be used practically to guide individualized training programs in maritime education.

## 4.2. Analysis and Interpretation of Results

The clustering analysis identified three distinct segments of cadet behavior based on performance indicators obtained from the OTMon system. As shown in table 4, cadets were distributed relatively evenly across the three clusters: Cluster 1 (n = 28, 35%), Cluster 2 (n = 24, 30%), and Cluster 3 (n = 28, 35%). This balanced distribution minimizes the risk of statistical bias due to unequal group sizes and ensures that comparison across clusters remains valid.

**Table 4.** Distribution of Cluster Membership

| CLUSTER | Frequency | Percent | CumulativeFrequency | CumulativePercent |
|---|---|---|---|---|
| 1 | 28 | 35.00 | 28 | 35.00 |

| | | | | |
|---|---|---|---|---|
| 2 | 24 | 30.00 | 52 | 65.00 |
| 3 | 28 | 35.00 | 80 | 100.00 |

Analyzing each cluster's descriptive statistics to gain a deeper understanding of its unique features came after reviewing the distribution summary in table 4. It is feasible to pinpoint particular behavioral patterns, strengths, and difficulties within each group by comparing the mean and standard deviation of all performance indicators. The basis for analyzing how cadets in various clusters approach training exercises, look for feedback, and handle challenges is provided by these descriptive measures. Table 5 displays the whole statistical profile for every cluster.

**Table 5.** Descriptive Statistics of Each Cluster (Mean and Standard Deviation)

| Cluster | N Obs | Variable | Mean | Std Dev |
|---|---|---|---|---|
| 1 | 28 | Task_Completion | 0.20 | 0.92 |
| | | Feedback_Frequency | 0.54 | 0.86 |
| | | Task_Difficulty | -0.71 | 0.80 |
| | | Report_Timeliness | -0.50 | 0.86 |
| | | Sign_On_Consistency | 0.20 | 0.90 |
| | | Total_Reports | -0.15 | 0.95 |
| 2 | 24 | Task_Completion | 0.10 | 0.97 |
| | | Feedback_Frequency | -0.71 | 0.81 |
| | | Task_Difficulty | -0.11 | 0.66 |
| | | Report_Timeliness | 0.25 | 1.06 |
| | | Sign_On_Consistency | -0.23 | 1.08 |
| | | Total_Reports | 0.92 | 0.43 |
| 3 | 28 | Task_Completion | -0.28 | 1.07 |
| | | Feedback_Frequency | 0.07 | 0.93 |
| | | Task_Difficulty | 0.81 | 0.84 |
| | | Report_Timeliness | 0.28 | 0.92 |
| | | Sign_On_Consistency | 0.00 | 1.02 |
| | | Total_Reports | -0.64 | 0.81 |

The standardized mean and standard deviation of six variables for each cluster are shown in table 5, which makes it evident that the behavior and performance profiles of the cadets differ from one another. With above-average task completion (mean = 0.20) and frequent feedback exchanges (mean = 0.54), as well as low perceived task difficulty (mean = −0.71), Cluster 1's high-performing cadets demonstrate significant involvement and skill in task execution. Administratively oriented cadets in Cluster 2 submit a lot of reports overall (mean = 0.92) but interact with feedback less frequently (mean = -0.71), indicating a focus on following procedures without commensurately high levels of reflective communication. Cadets in Cluster 3 report increased task difficulty (mean = 0.81), poorer task completion (mean = -0.28), and lower reporting behavior (mean = -0.64), which may indicate underlying cognitive or motivational impediments to effective training participation. Table 6 gives additional information about the factors that most strongly distinguish the groups by presenting the coefficients from the linear discriminant function that was used to differentiate these clusters.

**Table 6.** Linear Discriminant Function Coefficients per Cluster

| Variable | 1 | 2 | 3 |
|---|---|---|---|
| Constant | -1.56740 | -2.22743 | -1.63463 |
| Task_Completion | -0.07112 | 1.04511 | -0.82469 |
| Feedback_Frequency | 1.35789 | -1.98655 | 0.34486 |
| Task_Difficulty | -2.00355 | -0.06781 | 2.06167 |
| Report_Timeliness | -1.56357 | 0.83518 | 0.84770 |
| Sign_On_Consistency | 0.65077 | -0.97107 | 0.18158 |

| Variable | 1 | 2 | 3 |
|---|---|---|---|
| Total_Reports | -0.57287 | 2.69425 | -1.73649 |

The linear discriminant function coefficients for each cluster are shown in table 6, indicating the relative significance of each performance measure in differentiating between the three clusters that were found. While Task_Difficulty (-2.00355) and Report_Timeliness (-1.56357) contribute negatively, indicating lower performance in handling complex tasks and submitting timely reports, Feedback_Frequency (1.35789) is the most significant positive contributor for Cluster 1, suggesting that cadets in this group tend to receive feedback more frequently. While Feedback_Frequency (-1.98655) has a large negative weight, indicating less frequent feedback, Total_Reports (2.69425) is the strongest positive differentiator in Cluster 2, showing substantial reporting activity. While Total_Reports (-1.73649) shows a smaller number of reports than other clusters, Task_Difficulty (2.06167) is the most prominent positive coefficient for Cluster 3, indicating that this cluster excels at handling difficult assignments. The Generalized Squared Distances between Clusters, which measure the level of difference among these clusters, are shown in table 7 to help evaluate their separation and resemblance.

**Table 7.** Generalized Squared Distance between Clusters

| From CLUSTER | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | 11.25127 | 9.53517 |
| 2 | 11.25127 | 0 | 11.69943 |
| 3 | 9.53517 | 11.69943 | 0 |

According to the findings, Cluster 2 and Cluster 3 have the greatest gap (11.69943), closely followed by Cluster 1 and Cluster 2 (11.25127). The clusters with the least distance (9.53517) are Cluster 1 and Cluster 3, suggesting that their performance profiles are more similar than those of Cluster 2. The classification's validity and the discriminant function's ability to distinguish amongst cadet performance patterns are supported by the substantial distance values that generally imply that the clusters are well-separated. To further verify the robustness of this classification, table 8 presents the Classification Summary for Calibration Data based on the Linear Discriminant Function.

**Table 8.** Classification Summary for Calibration Data: WORK. CADET_CLUSTERS (Resubstitution Summary using Linear Discriminant Function)

| From CLUSTER | 1 | 2 | 3 | Total |
|---|---|---|---|---|
| 1 | 27<br>96.43 | 1<br>3.57 | 0<br>0.00 | 28<br>100.00 |
| 2 | 0<br>0.00 | 24<br>100.00 | 0<br>0.00 | 24<br>100.00 |
| 3 | 0<br>0.00 | 0<br>0.00 | 28<br>100.00 | 28<br>100.00 |
| Total | 27<br>33.75 | 25<br>31.25 | 28<br>35.00 | 80<br>100.00 |
| Priors | 0.33333 | 0.33333 | 0.33333 | |

Table 8 presents the Linear Discriminant Analysis (LDA) classification summary using the resubstitution method. Out of 80 observations, the model correctly classified 98.75% of the data. Specifically, 27 out of 28 cadets in Cluster 1 were correctly assigned (96.43%), while Cluster 2 and Cluster 3 achieved perfect classification accuracy (100%). Only one case from Cluster 1 was misclassified into Cluster 2. This high accuracy demonstrates that the clusters are linearly separable based on the six selected variables Task Completion, Feedback Frequency, Task Difficulty, Report Timeliness, Sign-On Consistency, and Total Reports. These results suggest that the clustering approach generated well-defined and meaningful groupings that can be reliably used for subsequent analysis or monitoring within maritime training systems. The Cross-Validation Summary utilizing the Linear Discriminant Function is shown in table 9 to further confirm the stability and generalizability of this classification.

**Table 9.** Classification Summary for Calibration Data: WORK. CADET_CLUSTERS (Cross-validation Summary using Linear Discriminant Function)

| From CLUSTER | 1 | 2 | 3 | Total |
|---|---|---|---|---|
| 1 | 26<br>92.86 | 1<br>3.57 | 1<br>3.57 | 28<br>100.00 |
| 2 | 0<br>0.00 | 24<br>100.00 | 0<br>0.00 | 24<br>100.00 |
| 3 | 0<br>0.00 | 0<br>0.00 | 28<br>100.00 | 28<br>100.00 |
| Total | 26<br>32.50 | 25<br>31.25 | 29<br>36.25 | 80<br>100.00 |
| Priors | 0.33333 | 0.33333 | 0.33333 | |

Table 9 shows the results of the Leave-One-Out Cross-Validation (LOOCV) procedure applied to the LDA model. In this method, each of the 80 observations was sequentially left out of the training set and used for validation. The model achieved a classification accuracy of 97.5%, with only two misclassified observations, both originating from Cluster 1. The classification for Cluster 2 and Cluster 3 remained perfectly accurate. These results confirm the robustness and generalizability of the LDA model. The cross-validation accuracy supports the stability of the clusters formed using K-Means and Ward's method and highlights the effectiveness of using LDA as a tool for assigning new cadets to appropriate clusters based on their performance data. This finding strengthens the case for integrating such clustering-based monitoring into platforms like OTMon. Table 10 displays the following findings for the dependent variable Task Completion.

**Table 10.** Dependent Variable: Task_Completion

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Type | 2 | 3.55335651 | 1.77667826 | 1.81 | 0.1700 |
| Error | 77 | 75.44664349 | 0.97982654 | | |
| Corrected Total | 79 | 79.00000000 | | | |

Table 10 presents the results of a one-way ANOVA test used to examine whether Task Completion scores significantly differed across the three clusters. The test returned an F-value of 1.81 and a p-value of 0.1700, indicating no statistically significant difference in Task Completion across clusters at the 5% significance level. Additionally, the R-squared ($R^2$) value was 0.045, meaning cluster membership explained only 4.5% of the variance in Task Completion. This result has two key implications. First, although Task Completion was included in the clustering process, it does not individually distinguish clusters significantly in a univariate context. Second, this reinforces the necessity of using a multivariate perspective when analyzing cadet performance, as other indicators such as Feedback Frequency, Task Difficulty, and Total Reports appear to contribute more meaningfully to the differentiation among clusters, as supported by the discriminant function analysis. Figure 2 illustrates the standardized distribution of the Task_Completion variable based on the performance data of 80 cadets during their on-board training.

.

 

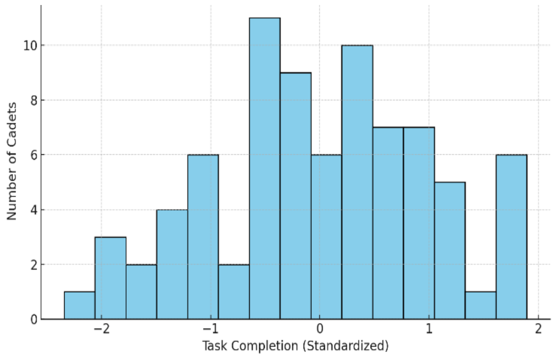


**Figure 2.** Distribution of Task Completion

The histogram reveals a relatively symmetrical shape, centering around the mean, indicating that most cadets achieved a consistent task completion level during the training. The peak appears between z-scores of -0.5 to +0.5, representing cadets with moderate performance. Meanwhile, the tails beyond ±1 suggest a smaller number of high- and low-performing cadets. This distribution provides a useful preliminary insight into performance segmentation. In the context of clustering, the Task_Completion variable serves as a foundational indicator when combined with other metrics such as Feedback_Frequency, Report_Timeliness, and Sign_On_Consistency to construct a comprehensive cadet performance profile. This also supports the development of data-driven interventions for targeted learning support. he relationship between these performance variables is further examined through the correlation matrix presented in table 11.

**Table 11**. Correlation Matrix Between Research Variables

| | **Task_ Completion** | **Feedback_ Frequency** | **Task_ Difficulty** | **Report_ Timeliness** | **Sign_On_ Consistency** | **Total_Reports** |
|---|---|---|---|---|---|---|
| Task_Completion | 1.0000 | 0.0033 | -.1053 | -.0998 | 0.1541 | -.1869 |
| Feedback_Frequency | 0.0033 | 1.0000 | -.1201 | 0.0518 | -.1507 | -.1010 |
| Task_Difficulty | -.1053 | -.1201 | 1.0000 | -.1198 | 0.0686 | -.0963 |
| Report_Timeliness | -.0998 | 0.0518 | -.1198 | 1.0000 | -.1507 | -.0232 |
| Sign_On_Consistency | 0.1541 | -.1507 | 0.0686 | -.1507 | 1.0000 | -.0117 |
| Total_Reports | -.1869 | -.1010 | -.0963 | -.0232 | -.0117 | 1.0000 |

Table 11 displays the correlation matrix among the six key performance indicators. Most correlations are weak ($|r| < 0.20$), suggesting that the variables represent distinct aspects of cadet learning behavior. For example, Task_Completion shows weak to moderate correlation with Sign_On_Consistency ($r = 0.1541$) and a negative relationship with Total_Reports ($r = -0.1869$), indicating that more frequent reporting does not necessarily imply better performance. The low inter-variable correlations justify the application of PCA to reduce data dimensionality while preserving interpretative integrity. The eigenvectors derived from the PCA results are summarized in table 12, providing insight into the contribution of each variable to the principal components.

**Table 12.** Eigenvectors of Principal Components from PCA

| | **Prince1** | **Prince2** |
|---|---|---|
| Task_Completion | 0.365653 | 0.600118 |
| Feedback_Frequency | -.375891 | 0.453435 |
| Task_Difficulty | 0.334151 | -.314293 |
| Report_Timeliness | -.496223 | 0.067335 |
| Sign_On_Consistency | 0.584616 | -.003624 |
| Total_Reports | -.159163 | -.575263 |

Table 12 presents the eigenvectors of the two principal components extracted from the PCA. The first component (Prin1) emphasizes Sign_On_Consistency (0.58), Task_Completion (0.37), and negatively Report_Timeliness (−0.50). The second component (Prin2) loads heavily on Task_Completion, Feedback_Frequency, and Total_Reports. Figure 3 shows the scree plot of the PCA, indicating that PC1 and PC2 account for approximately 23% and 20% of the variance, respectively.

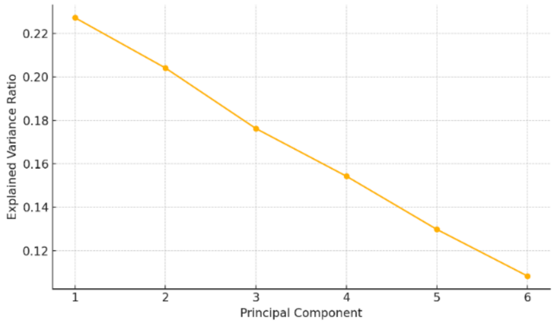


**Figure 3.** Scree Plod and Variance Explained

Cumulatively, these two components represent over 43% of the variance, which is substantial enough to guide clustering and visualization. The steep drop from PC1 to PC2 followed by a leveling slope supports the use of two dimensions in subsequent cluster analysis. This confirms that a significant portion of cadet performance diversity can be captured through just two latent components.

Following the segmentation of the cadet data using Ward's hierarchical clustering and K-Means, three separate clusters were found, each of which represented a different performance trend. With high scores in task completion, consistent sign-on, and timely report submissions, Cluster 1 consists of cadets who exhibit good self-regulation and consistent involvement, demonstrating proactive behavior and efficient time management. Cadets in Cluster 2 consistently submit assignments on time, but they exhibit few feedback exchanges and poorer self-monitoring, indicating that although they follow procedure, their learning style is more mechanical than reflective. The cadets in Cluster 3, on the other hand, score lower on the majority of variables, especially task completion and feedback frequency. This indicates difficulties with task execution as well as reflective engagement, underscoring the need for focused interventions or extra support techniques.

These profiles are consistent with Kolb's experiential learning theory, where self-regulated learning and reflection are essential for competency-based development. The use of multivariate clustering thus helps identify not only performance levels, but also behavioral patterns critical to learning outcomes.

## 4.3. Implications of the Findings

The findings of this study have both theoretical and practical implications. Theoretically, the research contributes to the emerging body of literature on learning analytics in maritime vocational education by demonstrating the use of unsupervised machine learning techniques for profiling cadet behavior an area that remains underexplored. From a practical standpoint, the clustering results from both K-Means and Ward’s method offer a new way to support cadet development during on-board training. For instance, Cluster 1 cadets, identified as highly self-regulated, can be given more autonomy, while Cluster 3 cadets may benefit from personalized mentoring or structured support. Additionally, institutions can integrate PCA-based radar chart visualizations into monitoring dashboards to detect performance anomalies and tailor feedback in real time. The PCA further supports the reduction of complexity in dashboard design, allowing instructors to interpret multi-dimensional behaviors through fewer variables. These implications suggest that data-driven systems like OTMon can be transformed from passive monitoring tools into active decision-support platforms that facilitate early intervention and competency-based evaluation.

## 4.4. Comparison with Previous Literature

In contrast to earlier studies in maritime education that largely examined usability and perception of digital monitoring platforms, this research shifts the focus to empirically driven learning analytics and clustering-based performance profiling. Prior work has typically evaluated cadet monitoring tools on user experience dimensions, rather than on measurable learning outcomes. This study's methodological approach aligns with the growing field of Educational Data Mining and Learning Analytics (NA and EDM), exemplified by comprehensive reviews such as by Romero and Ventura [53]. which highlight the utilization of machine learning and clustering techniques to discover hidden patterns in educational data. Additionally, key conceptual discussions by Ferguson have emphasized the role of analytics in personalizing learning paths and supporting instructional decisions [54]. However, unlike most existing research that focuses on general educational environments, our study uniquely applies these analytics within the maritime cadet training context, which remains underexplored. By integrating clustering and PCA-based visualization into OTMon, this research brings learning analytics into maritime vocational training a setting that has seen little empirical application of these approaches. In summary, our study bridges a gap by blending established EDM/LA frameworks (Romero and Ventura, Ferguson) with domain-specific clustering in maritime education, delivering both theoretical advancement and practical utility.

### 4.5. Limitations and Future Research Recommendations

Despite its contributions, this study has several limitations. First, the dataset is drawn from a single maritime education institution, which may limit generalizability across other academies or international contexts. Second, the study relies exclusively on quantitative metrics, such as the frequency and timeliness of task submissions, without incorporating qualitative insights like cadet reflections or mentor narratives. Moreover, although descriptive and clustering analyses revealed meaningful groupings, the ANOVA test for task completion across clusters was not statistically significant ($p = 0.17$). This result indicates that cluster differentiation is multivariate rather than univariate, underscoring the importance of multi-dimensional analysis like PCA and clustering rather than simple mean comparisons.

Qualitative methods like interviews or logbook analysis should be used in future studies to give the discovered clusters a deeper context for interpretation. Additionally, longitudinal tracking might be used to see how cadet habits change over the course of various training sessions, providing information about trends of growth or deterioration. Furthermore, testing supervised learning methods like support vector machines or decision trees would make it possible to forecast cadet outcomes and identify at-risk individuals early on. These improvements would make learning analytics in maritime education more reliable and applicable, opening the door for data-informed mentoring to develop into a useful and significant part of educational initiatives.

## 5. Conclusion and Recommendations

This study aimed to develop a structured, data-driven framework for profiling cadet performance during on-board training using clustering and multivariate analysis. By applying K-Means, PCA, and LDA, the research successfully identified three distinct cadet profiles: high-performing, administratively consistent, and at-risk. These classifications directly address the core research problem by offering a practical and empirical method to segment cadets based on their digital behavior and performance logs in the OTMon system. The integration of radar chart visualizations and PCA projections (Revision #3) allowed for intuitive interpretation of cadet patterns across performance dimensions, thus enabling institutions to make timely and targeted mentoring decisions. The application of LDA further confirmed the clustering model's internal validity, supported by high classification accuracy using both resubstitution and cross-validation methods. Theoretically, this research enriches the literature on learning analytics in vocational education, especially in maritime contexts which remain underrepresented in empirical studies. The findings operationalize key concepts from experiential and self-regulated learning theories using real-time behavior data. Furthermore, this study aligns with the work of Romero and Ventura as well as Ferguson [53],[54], who advocate for data-driven educational decision-making, although this research uniquely applies those principles within the maritime education domain. On a practical level, the study demonstrates how clustering models can be embedded into digital supervision tools such as OTMon. This integration enables institutions to monitor cadet progress more effectively, offer early warnings, and design adaptive learning pathways tailored to individual performance patterns. The model's structure provides a solid foundation for developing personalized mentoring systems within maritime training institutions. Nonetheless, the research has certain limitations. It utilized a single-institution dataset restricted to one academic cohort and relied solely on structured, quantitative data from OTMon logs. It did not include qualitative feedback, such as cadet perceptions, supervisory evaluations, or operational factors like network conditions and fatigue. These omissions may limit the depth and generalizability of the insights. Future studies should consider expanding the dataset across multiple

 ISSN 2723-6471 

institutions and training contexts to validate the model externally. Qualitative methods, such as interviews or sentiment analysis, should be incorporated to capture the subjective and contextual aspects of cadet performance. Additionally, the use of supervised machine learning could be explored to build predictive models with early warning capabilities. Longitudinal studies are also recommended to monitor cadet behavioral development over time and support sustainable, personalized learning strategies.

## 6. Declarations

### 6.1. Author Contributions

Conceptualization: D.W., E., R.S.; Methodology: E.; Software: D.W.; Validation: D.W., E., R.S.; Formal Analysis: D.W., E., R.S.; Investigation: D.W.; Resources: E.; Data Curation: E.; Writing – Original Draft Preparation: D.W., E., R.S.; Writing – Review and Editing: E., D.W., R.S.; Visualization: D.W.; All authors have read and agreed to the published version of the manuscript.

### 6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

### 6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### 6.4. Institutional Review Board Statement

Not applicable.

### 6.5. Informed Consent Statement

Not applicable.

### 6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] A. Ragab, "Intercultural competence and communication on board of merchant vessels," *Antwerp: University of Antwerp & Antwerp Maritime Academy*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.63028/10067/2079220151162165141.

[2] S. A. Balyaeva, T. G. Khvingiya, and S. A. Kalinina, "Integrated teaching of foreign language and technical disciplines for increasing the efficiency of training in maritime universities," *Pedagogical Journal*, vol. 9, no. 6A, pp. 1–12, 2019, doi: 10.34670/AR.2020.46.6.227.

[3] G. Emad and N. Meduri, "Maritime education and training system require a change to make a competent seafarer for shipping industry: A case study from an International Maritime institute," in *Proc. Int. Assoc. Maritime Univ. Conf.*, vol. 2019, no. 1, pp. 234–248, 2019.

[4] E. Boulougouris, P. Mizythras, L. Chrysinas, G. Vavourakis, G. Theotokatos, M. Aymelek, and I. Kurt, "Developing multidisciplinary blended learning courses for maritime education with cross-European collaboration," *WMU J. Maritime Affairs*, vol. 18, no. 3, pp. 319–340, Mar. 2019, doi: 10.1007/S13437-019-00167-X.

[5] R. Guntoro and P. Dwikora, "Integration of ship machinery maintenance and control systems in maritime education: Enhancing industry readiness," *J. Bus., Finance, Econ.*, vol. 5, no. 1, pp. 1–12, 2024, doi: 10.32585/jbfe.v5i1.5687.

[6] T. Cahyadi, A. Ahmad, and L. Barasa, "Integrating maritime law education and navigation skills: Enhancing vocational school curriculum," *Int. J. Sociol. Law*, vol. 1, no. 4, pp. 1–12, Nov. 2024, doi: 10.62951/ijsl.v1i4.197.

[7] B. Cheruiyot, "Challenges faced in the implementation of competency-based curriculum (CBC) in junior schools in Kenya," *East African J. Educ. Stud.*, vol. 7, no. 3, pp. 1–12, 2024, doi: 10.37284/eajes.7.3.2098.

[8] L. C. Ying and K. L. Chung, "Curriculum design for digital cultural innovation: Competency-based project learning with Google Earth and temple video creation," *Int. J. Religion*, vol. 6, no. 1, pp. 1–12, 2025, doi: 10.61707/hraee382.

[9] M. M. Asad and A. Qureshi, “Impact of technology-based collaborative learning on students' competency-based education: Insights from the higher education institution of Pakistan,” *Higher Educ., Skills Work-Based Learn.*, vol. 15, no. 3, pp. 562–575, 2025, doi: 10.1108/heswbl-07-2024-0202.

[10] N. R. Putra, D. N. Rizkiyah, M. A. C. Yunus, H. S. Kusuma, and H. Darmokoesoemo, “Mapping the landscape of clove oil as essential oil for health and wellness: A bibliometric review of advances, challenges, and future directions,” *J. Essent. Oil Bear. Plants*, vol. 27, no. 2, pp. 300–326, 2024, doi: 10.1080/0972060X.2024.2325099.

[11] I. Ikhwani, S. Rahayuningsih, E. Yuniarti, H. S. Kusuma, H. Darmokoesomo, and N. R. Putra, “Mapping the trend of evolution: A bibliometric analysis of biopesticides in fruit crop protection,” *J. Plant Dis. Prot.*, vol. 131, no. 1, pp. 645–664, 2024, doi: 10.1007/s41348-024-00879-0.

[12] A. F. Yuza and H. Putra, “Integrated administrative service quality in Bantan District Bengkalis Regency (study of E-KTP management),” *J. Gov. Stud.*, vol. 9, no. 1, pp. 1–12, 2023, doi: 10.25299/jkp.2023.vol9(1).12022.

[13] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. Sattar, “Predicting student performance in higher educational institutions using video learning analytics and data mining techniques,” *Appl. Sci.*, vol. 10, no. 11, pp. 3894-3906, 2020, doi: 10.3390/app10113894.

[14] H. Kurniawanto, A. Asari, A. Ratuningtyas, A. Mubarok, and L. E. Riyanti, “Transforming educational HR management: Integrating AI and data analytics for enhanced teacher performance and student outcomes,” *J. Sci. Educ. Res.*, vol. 10, no. 12, pp. 1–12, Dec. 2024, doi: 10.29303/jppipa.v10i12.9658.

[15] A. E. Sharwani, “Modernizing the U.S. educational system by elevating teaching methods and student performance through human-computer integration, data analytics, and other innovative technologies,” in *Proc. 2024 IEEE Int. Conf. Interdiscip. Approaches Technol. Manag. Soc. Innov. (IATMSI)*, vol. 2024, no. 1, pp. 1–6, 2024, doi: 10.1109/IATMSI60426.2024.10503523.

[16] D. Arunachalam and N. Kumar, “Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making,” *Expert Syst. Appl.*, vol. 111, no. 1, pp. 11–34, 2018, doi: 10.1016/j.eswa.2018.03.007.

[17] N. A. Ali, A. Abbassi, and O. Bouattane, “Performance evaluation of spatial fuzzy C-means clustering algorithm on GPU for image segmentation,” *Multimedia Tools Appl.*, vol. 82, no. 1, pp. 6787–6805, 2022, doi: 10.1007/s11042-022-13635-z.

[18] M. Sarkar, A. R. Puja, and F. R. Chowdhury, “Optimizing marketing strategies with RFM method and K-means clustering-based AI customer segmentation analysis,” *J. Bus. Manag. Stud.*, vol. 6, no. 2, pp. 1–12, 2024, doi: 10.32996/jbms.2024.6.2.5.

[19] J. W. Lai, L. Zhang, C. C. Sze, and F. S. Lim, “Learning analytics for bridging the skills gap: A data-driven study of undergraduate aspirations and skills awareness for career preparedness,” *Educ. Sci.*, vol. 15, no. 1, pp. 40-53, 2025, doi: 10.3390/educsci15010040.

[20] A. Abisoye, “Creating a conceptual framework for AI-powered STEM education analytics to enhance student learning outcomes,” *J. Front. Multidiscip. Res.*, vol. 5, no. 1, pp. 157–167, 2024, doi: 10.54660/IJFMR.2024.5.1.157-167.

[21] M. Taşkın, “Artificial intelligence in personalized education: Enhancing learning outcomes through adaptive technologies and data-driven insights,” *Hum.-Comput. Interact.*, vol. 8, no. 1, pp. 1–12, 2024, doi: 10.62802/ygye0506.

[22] D. Martínez-Cevallos, M. H. González-Serrano, and A. Proaño-Grijalva, “Understanding and enhancing women's loyalty in running events: A segmentation analysis based on brand perception,” *Sport Bus. Manag.*, vol. 1, no. 1, pp. 1–12, 2025, doi: 10.1108/SBM-08-2024-0094.

[23] M. Grigoroiu, C. Țurcanu, C. Constantin, A. Tecău, and B. Tescașiu, “The impact of EU-funded educational programs on the socio-economic development of Romanian students: A multidimensional analysis of sustainability,” *Sustainability*, vol. 17, no. 5, pp. 2057-2069, 2025, doi: 10.3390/su17052057.

[24] H. Afdilla and W. R. Hasibuan, “Analysis and comparison of the performance of K-means algorithm and X-means algorithm in disease type clustering in Mitra Medika Hospital,” *J. Artif. Intell. Eng. Appl.*, vol. 4, no. 1, pp. 1–12, Oct. 2024, doi: 10.59934/jaiea.v4i1.696.

[25] R. Mao, Y. Li, G. Li, H. P. Hildre, and H. Zhang, “A systematic survey of digital twin applications: Transferring knowledge from automotive and aviation to maritime industry,” *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 1, pp. 4240–4259, 2025, doi: 10.1109/TITS.2025.3535593.

[26] R. W. Liu, J. Nie, S. Garg, Z. Xiong, Y. Zhang, and M. S. Hossain, “Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems,” *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5374–5385, 2021, doi: 10.1109/JIOT.2020.3028743.

[27] M. Taghavi and L. Perera, "Data driven digital twin applications towards green ship operations," *Ocean Eng.*, vol. 5A, no. 1, pp. 1–12, 2022, doi: 10.1115/OMAE2022-78775.

[28] M. Y. Kataev, A. M. Korikov, and V. S. Mkrttchian, "Концепция и структура автоматизированной системы мониторинга качества обучения студентов," *Vysshee Obrazovanie v Rossii*, vol. 19, no. 10, pp. 30–46, 2017, doi: 10.17853/1994-5639-2017-10-30-46.

[29] J. Paños-Castro, O. Korres, I. Iriondo, and J. Petchamé, "Digital transformation and teaching innovation in higher education: A case study," *Educ. Sci.*, vol. 14, no. 8, pp. 820-840, 2024, doi: 10.3390/educsci14080820.

[30] T. Xue, "Stepping out of the technology box: Comprehensive digital transformation of cadre education and training," *Int. J. Educ. Humanit.*, vol. 18, no. 1, pp. 1–12, 2025, doi: 10.54097/pptdkb88.

[31] Z. Yue, R. Wang, and C. Dawei, "Digital transformation of maritime vocational education in the new era," *J. Higher Educ. Teach.*, vol. 1, no. 5, pp. 1–12, 2024, doi: 10.62517/jhet.202415515.

[32] S. Weydner-Volkmann and D. Bär, "Student autonomy and learning analytics: Philosophical considerations for designing feedback tools," *J. Learn. Analytics*, vol. 11, no. 3, pp. 1–12, 2024, doi: 10.18608/jla.2024.8313.

[33] S. Tirado-Olivares, R. Cózar-Gutiérrez, J. A. González-Calero, and N. Dorotea, "Evaluating the impact of learning management systems in geographic education in primary school: An experimental study on the importance of learning analytics-based feedback," *Sustainability*, vol. 16, no. 7, pp. 2616-2631, 2024, doi: 10.3390/su16072616.

[34] D. Ifenthaler, C. Schumacher, and M. Şahin, "System-based or teacher-based learning analytics feedback – What works best?," in *Proc. 2021 Int. Conf. Adv. Learn. Technol. (ICALT)*, vol. 2021, no. 1, pp. 184–186, 2021, doi: 10.1109/ICALT52272.2021.00062.

[35] N. I. M. Talib, N. A. Majid, and S. Sahran, "Identification of student behavioral patterns in higher education using K-means clustering and support vector machine," *Appl. Sci.*, vol. 13, no. 5, pp. 3267-3277, 2023, doi: 10.3390/app13053267.

[36] G. M. Alam and S. Roslan, "Contribution of prejudiced clustering education system in developing horizontal and vertical mismatch in job market: Quality of employees in banking sector," *Bus. Process. Manag. J.*, vol. 27, no. 5, pp. 1315–1334, 2020, doi: 10.1108/bpmj-07-2020-0339.

[37] G. M. Alam, "Clustering education policy in secondary provision: Impact on higher education and job market," *Int. J. Educ. Reform*, vol. 30, no. 1, pp. 56–76, 2021, doi: 10.1177/1056787920958409.

[38] F. G. Loro, E. Sancristóbal, R. Gil, B. Quintana, P. P. Merino, S. Tzanova, and M. Castro, "Competency-based instructional design for microelectronics training: ECoVEM project," in *Proc. 2024 IEEE Front. Educ. Conf. (FIE)*, vol. 2024, no. 1, pp. 1–7, 2024, doi: 10.1109/FIE61694.2024.10893068.

[39] C. Yang, F. Kaiser, H. Tang, P. Chen, and J. Diao, "Sustaining the quality development of German vocational education and training in the age of digitalization: Sustainability challenges and strategies," *Sustainability*, vol. 15, no. 4, pp. 3845-3856, 2023, doi: 10.3390/su15043845.

[40] A. Ala, N. Hamidi, S. Yoniessa, F. Masito, and M. A. Muis, "Simulation-based learning in maritime training: Enhancing competency and preparedness," *Meteor STIP Marunda*, vol. 17, no. 1, pp. 1–12, 2024, doi: 10.36101/msm.v17i1.361.

[41] A. B., J. M. Scaria, A. Trivedi, T. Sharma, A. M. Yadav, and R. Nair, "Exploring ethical dimensions of employing artificial intelligence algorithms in healthcare environments," in *Proc. 2024 IEEE 4th Int. Conf. ICT Bus. Ind. Gov. (ICTBIG)*, vol. 2024, no. 1, pp. 1–9, 2024, doi: 10.1109/ICTBIG64922.2024.10911856.

[42] S. Barteit, J. Schmidt, M. Kakusa, G. Syakantu, A. Shanzi, Y. Ahmed, G. Malunga, K. Blass, J. Nieder, P. Andreadis, and F. Neuhann, "Electronic logbooks (e-logbooks) for the continuous assessment of medical licentiates and their medical skill development in the low-resource context of Zambia: A mixed-methods study," *Front. Med.*, vol. 9, no. 1, pp. 1–12, 2022, doi: 10.3389/fmed.2022.943971.

[43] D. Devi, P. Hazarika, V. Saluja, A. Chamoli, N. D. Prabu, and P. Nagaraj, "Predictive feedback loops: Harnessing AI for continuous assessment and personalized growth in English language learners," in *Proc. 2024 Int. Conf. Commun., Control, Intell. Syst. (CCIS)*, vol. 2024, no. 1, pp. 1–5, doi: 10.1109/CCIS63231.2024.10931931.

[44] M. Taylor, A. Barthakur, A. Azad, S. Joksimović, X. Zhang, and G. Siemens, "Quantifying collaborative complex problem solving in classrooms using learning analytics," in *Proc. 14th Learn. Analytics Knowl. Conf.*, vol. 2024, no. 1, pp. 1–12, doi: 10.1145/3636555.3636913.

[45] M. Cukurova, "The interplay of learning, analytics, and artificial intelligence in education," *arXiv*, vol. 1, no. 1, pp. 1–12, 2024, doi: 10.48550/arXiv.2403.16081.

[46] S. Gunasekara and M. Saarela, "Explainability in educational data mining and learning analytics: An umbrella review," in *Proc. 17th Conf. Educ. Data Mining (EDM)*, Jul. 2024, vol. 1, no. 1, pp. 1–12, doi: 10.5281/zenodo.12729987.

[47] A. Chandra, A. K. Sahoo, M. B. Sabnis, K. K. Nayak, A. Ghosh, and S. Tiwari, "Laparoscopic training instruments designed to provide real-time feedback for surgical trainees," *J. Surg. Simul.*, vol. 10, no. 1, pp. 1–12, 2023, doi: 10.1102/2051-7726.2023.0005.

[48] K. Hoek, C. Jaschinski, M. Velzen, and E. Sarton, "Development of a real-time adaptable virtual reality-scenario training for anaesthesiology education, a user-centered design," in *Proc. 16th Int. Conf. Comput. S.E.*, vol. 1, no. 1, pp. 751–757, 2024, doi: 10.5220/0012755600003693.

[49] S. Tai, Y. Lin, T. Hsu, Y. Lee, and M. Lin, "Development of machine learning-based real-time squat training feedback system," in *Proc. 2023 IEEE 6th Int. Conf. Knowledge Innov. Invention (ICKII)*, vol. 2023, no. 1, pp. 376–379, doi: 10.1109/ICKII58656.2023.10332629.

[50] Y. Chen, "Educational data mining in TikTok We Media short videos based on K-means algorithm," in *Proc. 2024 5th Int. Seminar Artif. Intell., Netw. Inf. Technol. (AINIT)*, vol. 2024, no. 1, pp. 1040–1043, doi: 10.1109/AINIT61980.2024.10581718.

[51] V. Vibha, C. R. Gowthami, S. B. Kayalvizhi, S. Selvakanmani, and D. C. Edara, "Strategically improved K-means clustering in mining blood donor data analysis and IoT-based allocation," *J. Intell. Syst. Internet Things*, vol. 13, no. 1, pp. 1–12, 2024, doi: 10.54216/JISIoT.130110.

[52] A. Yuda, "Comparison of K-means and hierarchical clustering algorithms to find out customer data in internet services," *Informatics Sci. Technol.*, vol. 1, no. 1, pp. 1–12, Jul. 2024, doi: 10.34005/insit.v2i2.4119.

[53] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 3, pp. 1-21, 2020, doi: 10.48550/arXiv.2402.07956.

[54] E. Mangina and G. Psyrra, "Review of learning analytics and educational data mining applications," in *EDULEARN21 Proc.*, vol. 2021, no. 1, pp. 949–954, 2021, doi: 10.21125/edulearn.2021.0250.