

Peer Review

Review of: "Data-Driven Innovation in Workforce Selection: A Clustering-Based Workflow for Technology Adoption in Indonesian Construction SMEs"

Dalila Boughaci¹

1. Université des Sciences et de la Technologie Houari Boumediène, Algeria

In this paper, the authors investigate the potential of K-Means clustering as a tool to improve recruitment practices in Indonesian construction, specifically for small and medium-sized enterprises (SMEs). The authors use a data-driven approach to classify candidates according to three main skills: AutoCAD drafting, report writing, and adaptability.

K-Means clustering is a widely used unsupervised learning algorithm, known for its simplicity and scalability. One of the reasons for selecting K-Means is its effectiveness with small datasets, making it particularly suited to the SME context, where large volumes of data may not be available. Additionally, K-Means can efficiently handle the three competencies evaluated in the study: technical skills (AutoCAD drafting), communication skills (report writing), and soft skills (adaptability).

In the study, the authors apply K-Means clustering to assign candidates into three groups: Rejected, Under Consideration, and Accepted. They argue that this method offers a more transparent, equitable, and efficient alternative to traditional, threshold-based recruitment processes, which are often influenced by subjective judgments and biases. By using clustering, the method provides a clearer, data-driven framework for decision-making.

Although the clustering procedure is clearly explained, the paper has several key limitations that weaken the depth of its findings.

Lack of Numerical Results: While the use of 2D and 3D scatter plots effectively visualizes how the algorithm groups applicants based on their competencies, the authors fail to provide any numerical details, such as mean scores, standard deviations, cluster sizes, or the distribution of candidates within

each cluster. This lack of quantitative information limits the ability to assess the significance of the clustering results and hinders a deeper understanding of the specific characteristics of each group (e.g., how candidates in the "Under Consideration" cluster compare to those in the "Accepted" group).

Absence of Statistical Analysis: The paper lacks any discussion on the statistical significance of the clustering results, such as how effectively the clusters differentiate based on competencies or whether the method holds any predictive value. A more comprehensive analysis, like comparing the clustering outcomes with real-world hiring results, would offer a clearer understanding of the method's practical relevance.

Interpretation of Results: The paper lacks a detailed interpretation of the clustering outcomes. While the clusters are identified, there is no explanation of why candidates are grouped in specific ways or how soft skills and technical skills influence the decision-making process. This lack of analysis reduces the practical usefulness of the findings for employers.

Comparison with Other Methods: The study does not compare K-Means clustering with other clustering methods, such as Hierarchical Clustering, DBSCAN, or Gaussian Mixture Models. Including such a comparison would offer a stronger justification for choosing K-Means and could help determine if other algorithms might perform better in terms of cluster cohesion or interpretability.

Limited Competency Assessment: The paper focuses on just three competencies—AutoCAD drafting, report writing, and adaptability—which may not fully capture the wide range of skills needed in modern construction roles. Important competencies like project management, leadership, or digital skills (e.g., Building Information Modeling -BIM- knowledge) are excluded, limiting the applicability of the results to a broader spectrum of roles within the sector.

Small Sample Size: The study uses a relatively small sample size of just 30 shortlisted candidates. While this may be sufficient for a proof of concept, the limited sample size reduces the generalizability of the findings and raises questions about the reliability of the results. A larger dataset would enhance the robustness of the clustering and offer more meaningful insights.

Distance Metric Not Discussed: The authors do not specify which distance metric was used in the K-Means algorithm (e.g., Euclidean distance, Manhattan distance, or Cosine similarity). Since the choice of distance metric can greatly influence the clustering results, discussing this would improve the methodological transparency of the paper.

Limited Discussion on Clustering Validity: Although the Davies–Bouldin Index (DBI) is mentioned as a measure of cluster validity, the paper lacks a more comprehensive discussion on how the clustering process was validated. It would be helpful to include cross-validation techniques or external validation through comparisons with expert judgment or real-world outcomes to ensure the clusters are meaningful and accurately reflect the competencies needed in construction projects.

Scalability and Generalizability: The study is based on data from a single SME in the Indonesian construction sector, so the results may not be directly applicable to other industries or regions. While the methodology is valuable, future research could explore how well it generalizes to different contexts or sectors. Additionally, there is limited discussion on how the proposed method would scale with larger datasets or across other industries. While the study focuses on a small sample of construction candidates, it would be helpful to understand whether this clustering approach could be applied to other SMEs or sectors, or if adjustments are needed to account for industry-specific requirements.

In conclusion, while the paper introduces an interesting and innovative approach to recruitment using K-Means clustering, it lacks essential numerical analysis, interpretation, and comparative assessments that would provide a deeper understanding of the method's effectiveness. Addressing these gaps would make the study more valuable, offering actionable insights for SMEs looking to implement data-driven decision-making tools in their workforce selection process.

Declarations

Potential competing interests: No potential competing interests to declare.