Vol. 4 No. (1) (June 2026) 141-152

Journal of Data Insights

http://journalnew.unimus.ac.id/index.php/jodi

MSMEs Digital Transformation Readiness Prediction for Data-Driven Decision Support

Andi Riansyah*[1], Maya Indriastuti[2], Maulana Ahmad Widiarta[3]
[1]Department of Informatics, Universitas Islam Sultan Agung, Indonesia
[2]Department of Accounting, Universitas Islam Sultan Agung, Indonesia
[3]Department of Informatics, Universitas Islam Sultan Agung, Indonesia

DOI: https://doi.org/10.26714/jodi.v4i1.1193

Article Info

Article history:
Received January 5, 2026
Revised February 20, 2026
Accepted April 10, 2026

Keywords:
Classification; Data-driven decision support; Digital transformation; MSMEs; Pseudo-labeling; Random Forest.

Abstract

Digital transformation readiness among micro, small, and medium enterprises (MSMEs) varies across business sectors and organizational capabilities. This study develops a supervised classification model to predict MSMEs digital transformation readiness for data-driven decision support. Target labels were constructed from previous fuzzy clustering results and organized into three readiness levels: low, moderate, and high. Random Forest was selected as an interpretable ensemble baseline for small tabular MSMEs survey data because it can handle nonlinear feature interactions, reduce single-tree instability, and provide feature importance for decision support. Data preprocessing included categorical encoding, stratified train-test separation, and class imbalance handling applied only to the training data to avoid data leakage. Model evaluation on the hold-out set produced 91.40% accuracy, 91.36% macro precision, 92.16% macro recall, and 91.72% macro F1-score. The confusion matrix showed that 85 of 93 test observations were correctly classified, with most errors occurring between adjacent readiness levels. Feature importance analysis indicated that dynamic capability was the most influential predictor, followed by workforce transformation and MSMEs performance. The findings should be interpreted as evidence from the Semarang MSMEs dataset and as an operationalization of clustering-derived pseudo-labels into a practical decision-support prototype, rather than as a universal generalization beyond the study context.

✉ Correspondence Address:
E-mail: andi@unissula.ac.id

e-ISSN: 2988 - 2109

INTRODUCTION

Digital transformation has become an important driver of competitiveness for micro, small, and medium enterprises (MSMEs). Digital platforms, financial technology, online marketplaces, and data-driven business processes provide opportunities for MSMEs to expand market access, improve operational efficiency, and respond to changes in customer behavior. The importance of readiness assessment is amplified by the economic role of MSMEs: they represent the overwhelming majority of business units in many economies, contribute substantially to employment and GDP, and remain a major target of digital transformation programs in Indonesia [1]-[5]. However, MSMEs do not adopt digital technology at the same speed. There is significant variation in readiness levels because of disparities in workforce skills, organizational agility, sustainability performance, financial inclusion, and access to digital infrastructure.

Prior studies show that predictive analytics can convert heterogeneous organizational data into actionable categories, but the relevance for MSMEs readiness prediction depends on whether the model can support operational decisions rather than merely report descriptive patterns. Therefore, this study positions machine learning as a decision-support mechanism for readiness stratification, not only as a generic classification task. Studies on business intelligence, digital innovativeness, financial inclusion, and MSMEs performance indicate that organizational readiness is influenced by business capability, human resources, and technology-related constraints [6]–[13].

Previous research on MSMEs in Semarang utilized fuzzy k-means to cluster 257 MSMEs based on workforce transformation, dynamic capability, and MSMEs performance [14]. The clustering approach was useful for describing heterogeneity among MSMEs and produced three readiness-related groups. Nevertheless, the research gap remains that no prior study has operationalized fuzzy clustering-derived MSMEs readiness groups into a supervised predictive model that can classify new MSMEs and support deployment-oriented decision-making. Predictive classification is practically more valuable because local governments and support institutions can score new MSMEs profiles without rerunning clustering on the whole dataset. At the same time, the use of cluster-derived labels requires methodological caution because such labels are pseudo-labels rather than expert-validated ground truth. This interpretation is also consistent with dynamic capability theory, which emphasizes sensing, seizing, and reconfiguring resources as core organizational capabilities for adaptation.

Random forest is selected in this study not because it is assumed to be universally superior to XGBoost, LightGBM, CatBoost, Support Vector Machine, or gradient boosting, but because it is a suitable interpretable baseline for a relatively small tabular dataset with mixed predictor types and potential nonlinear interactions. Compared with a single decision tree, random forest reduces variance through ensemble aggregation; compared with more complex boosting algorithms, it has a lower tuning burden and provides a straightforward impurity-based feature importance mechanism for policy interpretation [15]–[20]. Previous applications across prediction, IoT, environmental evaluation, cost-sensitive learning, and explainable modeling demonstrate that random forest is flexible for applied decision-support tasks, although algorithmic superiority should be verified through broader benchmark experiments in future work [21]–[32].

To address this gap, this study makes three original contributions. Scientifically, it operationalizes fuzzy clustering-derived pseudo-labels as supervised classification targets for MSMEs digital readiness in an Indonesian regional context. Methodologically, it develops a random forest-based pipeline that includes stratified data splitting, training-only class imbalance handling, multiclass evaluation, and feature importance interpretation. Practically, it translates predicted readiness classes into a decision-support interface with stratified intervention recommendations for MSMEs assistance programs.

THEORETICAL BASIS AND PROPOSED CLASSIFICATION FRAMEWORK

This section discusses the theoretical basis and proposed classification framework for the MSMEs digital transformation readiness prediction model. The proposed framework is based on the assumption that digital readiness is a multidimensional organizational condition. It is not determined only by technology ownership, but also by workforce transformation, dynamic capability, MSMEs performance, and sectoral business characteristics. These dimensions provide the input structure for the predictive model and connect the technical classification task with the practical need for MSMEs development support.

The first theoretical component is cluster-derived pseudo-labeling. Previous Fuzzy K-Means clustering produced readiness-related groups by using workforce transformation, dynamic capability, and MSMEs performance as key dimensions [14]. In the present study, the cluster membership with the highest degree is converted into a discrete readiness label. This conversion may introduce circularity because the supervised classifier learns from labels generated by an unsupervised model using related predictors. Therefore, this study frames the classifier as an operationalization and generalization of previous clustering results, not as independent proof of the true readiness status of each MSMEs. The pseudo-labels are used to enable decision-support scoring of new observations, while their validity should be strengthened in future studies through expert assessment and clustering validation metrics.

The second theoretical component is the Random Forest classifier. Random Forest is an ensemble learning method that builds several decision trees trained on bootstrap samples and random subsets of features. The ensemble structure reduces the instability of a single decision tree and improves generalization by aggregating predictions from many trees [16], [17]. This characteristic is suitable for MSMEs survey data because the variables are tabular, may contain nonlinear relationships, and may include correlated indicators derived from organizational and business dimensions. The mathematical formulation below summarizes the main principles used in the proposed classification framework.

Dynamic capability theory further supports the interpretation of readiness as an organizational capability construct. According to the dynamic capability perspective, firms sustain competitiveness in turbulent environments by sensing opportunities, seizing them, and reconfiguring internal and external resources. In the MSMEs context, this means that digital readiness is not merely a function of hardware ownership or platform registration. It also depends on whether the enterprise can learn, adapt, reorganize work routines, and align digital tools with market opportunities.

The mathematical formulation of the Random Forest classifier used in this study is described as follows. Let D = {(x, y)} denote the training dataset, where x represents the MSMEs predictor vector and y represents the readiness class. A Random Forest consists of B decision trees, and each tree h_b(x) is trained using bootstrap sampling and random feature selection [16], [17]. For multiclass classification, the final prediction is obtained through majority voting across all trees:

$$\hat{y}(x) = {}_{\text{majority\_vote}}\left(h_1(x), h_2(x), \dots, h_B(x)\right) \qquad (1)$$

where ŷ(x) is the predicted readiness class, h_b(x) is the prediction of the b-th tree, and B is the total number of trees in the forest. In this study, the target class consists of three categories: low, moderate, and high readiness.

During tree construction, each split is selected by reducing class impurity. The Gini impurity at node t is defined as:

$$G(t) = 1 - \sum_{k} p_k(t)^2 \qquad (2)$$

where $p_k(t)$ is the proportion of training samples belonging to class k at node t. A candidate split divides node t into left and right child nodes. The decrease in impurity is calculated as:

$$\Delta G = G(t) - \frac{n_L}{n_t} G(t_L) - \frac{n_R}{n_t} G(t_R) \backslash hfill \qquad (3)$$

where nt is the number of samples at node t, while nL and nR are the numbers of samples in the left and right child nodes. The split with the largest impurity decrease is selected. Because the dataset contains three readiness classes, class weighting is also used to reduce bias toward the majority class. The balanced class weight for class k is calculated as:

$$w_k = \frac{N}{K \times n_k} \qquad (4)$$

where N is the total number of training samples, K is the number of classes, and n_k is the number of training samples in class k. This weighting scheme increases the penalty for misclassifying minority classes. Finally, feature importance is computed from the relative contribution of each predictor to impurity reduction across the forest:

$$FI_j = \frac{\text{Total impurity decrease by feature } j}{\text{Total impurity decrease by all features}} \tag{5}$$

where FI_j is the importance score of feature j. This formulation explains why Random Forest is not only used for prediction, but also for identifying the dominant factors that influence MSMEs digital transformation readiness.

The proposed algorithm integrates these components into a decision support pipeline. First, MSMEs data are collected and cleaned. Second, readiness labels are formed from the previous clustering output and mapped into low, moderate, and high readiness classes. Third, Random Forest is trained using the prepared predictors, with imbalance handling restricted to the training subset. Fourth, the trained model is evaluated using multiclass metrics and interpreted through feature importance analysis. Finally, the predicted classes are translated into development strategies so that the model output can support targeted MSMEs assistance programs. Because the labels are cluster-derived, the pipeline is intended as a practical scoring framework and should be validated with expert-labeled data before being generalized to other regions.

METHOD

This study follows a quantitative data mining design guided by the CRISP-DM logic, namely business understanding, data understanding, data preparation, modeling, evaluation, and deployment-oriented interpretation. Figure 1 presents the research development flowchart used to build the MSMEs digital transformation readiness classification model. The workflow explicitly separates the training and testing subsets, places SMOTE-NC only on the training path, evaluates the model using multiclass metrics, and extracts feature importance before translating the results into development strategies.

As shown in Figure 1, the process begins with MSMEs data collection and data understanding, where survey data are gathered from business respondents and examined for completeness and consistency. The data then pass through preprocessing and cleaning before cluster-derived pseudo-label formation is performed based on the previous Fuzzy K-Means analysis [14]. These pseudo-labels are mapped into low, moderate, and high readiness classes. The labeled dataset is then divided into a training set and a testing set using stratified splitting. SMOTE-NC is applied only to the training set, while the testing set is kept unchanged for model validation and evaluation.

Before model training, class imbalance handling is applied only to the training data using SMOTE-NC, which helps improve minority-class representation while preventing data leakage [18]. In this experiment, the SMOTE-NC procedure used business type as the categorical feature, k_neighbors = 5, sampling_strategy = auto, and random_state = 42. The balanced training data are then used for Random Forest model training. The model is evaluated on the hold-out testing set using accuracy, precision, recall, F1-score, confusion matrix, and a binomial confidence interval for accuracy [15]-[17].

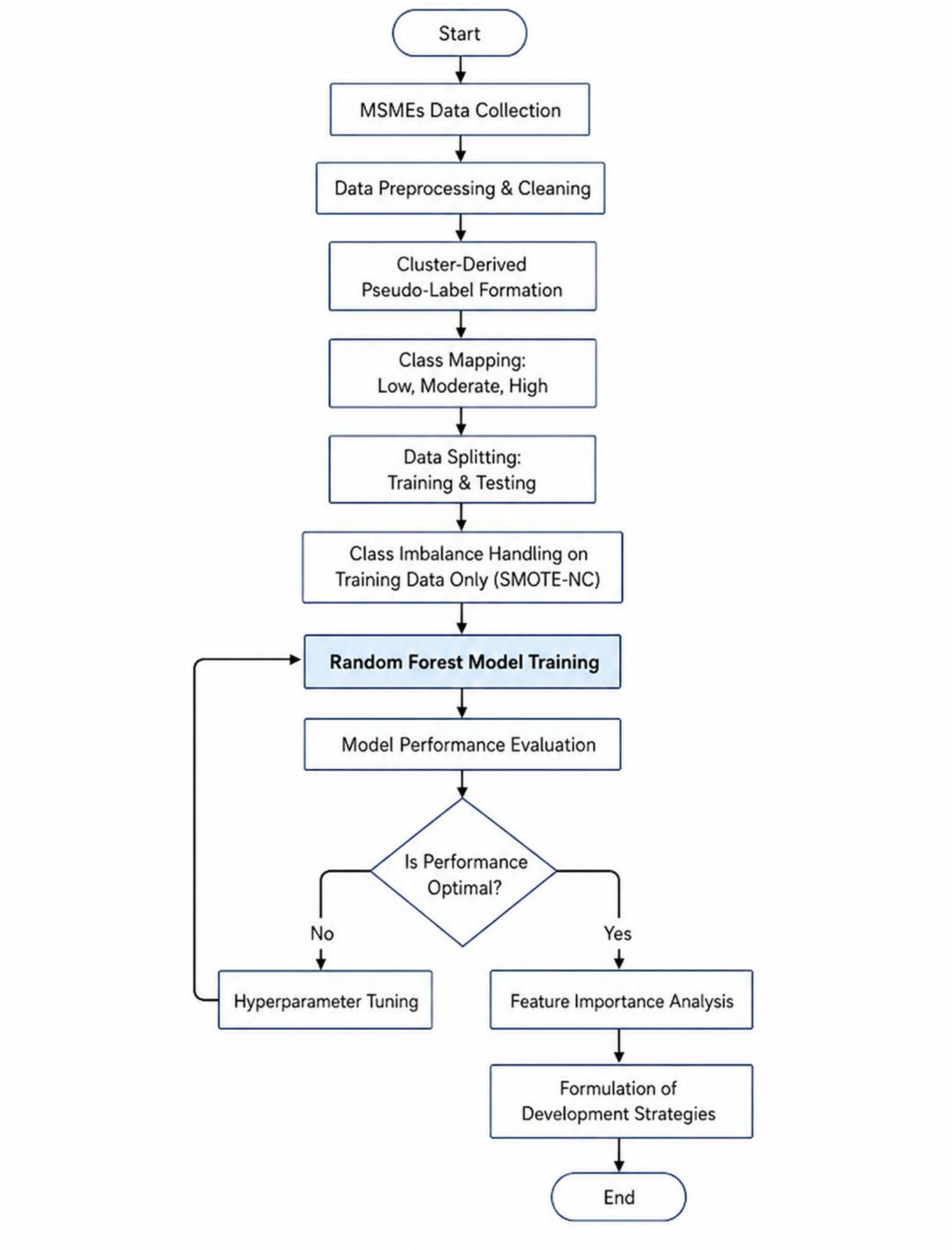


Figure 1. Research development flowchart for MSMEs readiness classification

If the performance is determined to be satisfactory, the workflow continues to feature importance extraction to identify the most significant predictors for classifying readiness levels. Hyperparameter review is included in the workflow to show how the model could be refined; however, the reported experiment uses the baseline configuration in Table 3 and does not claim exhaustive optimization. The last stage is the formulation of development strategies, in which the classification results and feature importance findings are translated into concrete recommendations for MSMEs support programs.

Table 1 summarizes the main variables used in the classification model. The predictors cover both organizational and business dimensions. Workforce transformation reflects digital skill adaptation, dynamic capability reflects organizational agility, MSMEs performance reflects business condition, and business type captures sectoral differences among MSMEs.

Table 1. Predictor variables used in the Random Forest model

Variable	Operational meaning	Role in the model
Business type	Sector of MSMEs activity, such as food and beverage, craft, fashion, or other sectors.	Captures sectoral characteristics that may affect readiness.
Workforce transformation	Digital skills, employee adaptability, and internal preparation for digital business processes.	Represents human resource readiness for digital change.
Dynamic capability	Ability to sense market changes, seize opportunities, and reconfigure business resources.	Represents organizational agility and innovation capability.
MSMEs performance	Business performance indicators related to sales, operational efficiency, and market outcomes.	Represents the business condition supporting digital investment.

The categorical variables were encoded before model training. Missing or inconsistent values were checked during preprocessing, and no missing values remained in the modeling dataset. The dataset contained 257 observations and four main predictor groups, consisting of one categorical predictor group (business type) and three numerical or ordinal predictor groups; therefore, 25% of the main predictor groups were categorical. The full pseudo-label distribution used in the classification experiment was 84 low-readiness observations (32.68%), 110 moderate-readiness observations (42.80%), and 63 high-readiness observations (24.51%). The dataset was divided using a stratified train-test split so that the class distribution was represented in both subsets; the hold-out test set contained 93 observations: 30 low-readiness, 40 moderate-readiness, and 23 high-readiness MSMEs. SMOTE-NC was applied only to the training subset after the train-test split, while the testing subset was kept in its original structure to prevent data leakage and ensure unbiased evaluation [18]. Table 3 presents Dataset characteristics and experimental setup.

Table 2. Dataset characteristics and experimental setup

Item	Description
Total observations	257 MSMEs after preprocessing
Missing values	No missing values remained after cleaning
Predictor groups	Business type, workforce transformation, dynamic capability, and MSMEs performance
Categorical predictor groups	1 of 4 predictor groups (25%): business type
Full pseudo-label distribution	Low: 84 (32.68%); Moderate: 110 (42.80%); High: 63 (24.51%)
Hold-out test support	Low: 30; Moderate: 40; High: 23; total: 93
Imbalance handling	SMOTE-NC applied only to the training subset

The algorithmic principles and mathematical formulation of Random Forest have been explained in Section 2. For the experimental implementation, Random Forest was chosen due to its robustness for tabular data, its ability to model nonlinear relationships, and its reduction of overfitting by combining decision trees through bootstrap aggregation and random feature selection [15]-[17], [21]-29]. Each tree produced a class prediction, and the final class was determined via majority voting. The model was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. Accuracy was calculated from the ratio of correct predictions to all test instances, while precision, recall, and F1-score were reported using a macro average to give equal attention to each readiness class. The parameters were manually selected as a transparent baseline after preliminary stability checks; no grid search, random search, or Bayesian optimization was conducted. Therefore, the results should be interpreted as baseline Random Forest performance rather than fully optimized performance. Table 3 presents the main Random Forest configuration used as the experimental baseline.

Table 3. Random Forest baseline configuration

Parameter	Value	Purpose
n_estimators	100	Number of decision trees in the ensemble.
max_depth	10	Controls tree complexity and overfitting.
min_samples_split	2	Minimum samples required to split a node.
min_samples_leaf	1	Minimum samples required at a leaf.
class_weight	balanced	Balances class penalties during training.
random_state	42	Ensures reproducible stochastic processes.

RESULTS AND DISCUSSION

The model was evaluated on a hold-out test set containing 93 observations. The confusion matrix produced 85 correct predictions and eight misclassifications. Therefore, the accuracy was calculated as 85 divided by 93, resulting in 91.40%. Using a Wilson binomial interval for 85 correct predictions out of 93 test observations, the 95% confidence interval for accuracy was approximately 83.93% to 95.58%. This interval indicates that, although the point estimate is high, the uncertainty associated with the limited test-set size should be considered. Table 4 presents the classification report for each readiness class.

Table 4. Classification report of the Random Forest model

Class	Precision	Recall	F1-score	Support
Low readiness	0.9032	0.9333	0.9180	30
Moderate readiness	0.9211	0.8750	0.8974	40
High readiness	0.9167	0.9565	0.9362	23
Macro average	0.9136	0.9216	0.9172	93
Weighted average	0.9142	0.9140	0.9137	93

The performance values in Table 4 show that the model was able to classify the three readiness levels with balanced performance. The high readiness class obtained the highest recall, meaning that most truly high-readiness MSMEs were detected correctly. The moderate readiness class produced a slightly lower recall, indicating that this transitional group was more difficult to separate because its characteristics overlap with both low and high readiness groups. However, because the evaluation is based on a single hold-out set rather than k-fold cross-validation, the performance values should be interpreted as preliminary evidence rather than definitive generalization beyond the dataset.

Figure 2 provides a more detailed view of the prediction distribution. Most observations are located on the main diagonal: 28 low-readiness MSMEs were correctly predicted as low, 35 moderate-readiness MSMEs were correctly predicted as moderate, and 22 high-readiness MSMEs were correctly predicted as high.

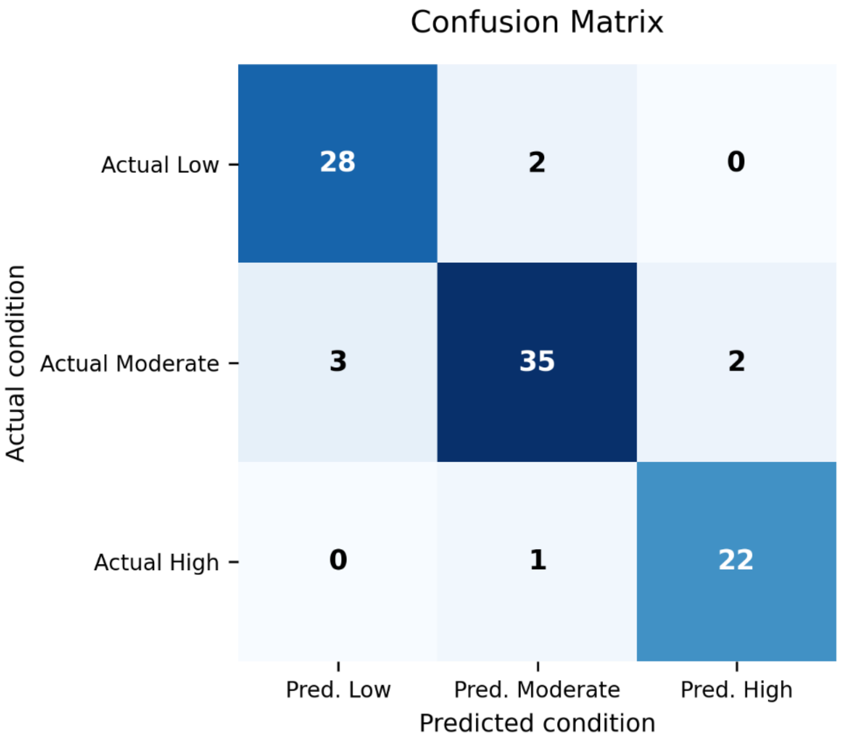


Figure 2. Confusion matrix of the Random Forest model

The misclassification pattern in Figure 2 is relatively small and mostly occurs between adjacent readiness levels. Two low-readiness MSMEs were predicted as moderate, while three moderate-readiness MSMEs were predicted as low and two were predicted as high. Only one high-readiness MSMEs was predicted as moderate. This pattern is acceptable in the context of readiness assessment because the boundary between adjacent readiness levels is naturally gradual rather than strictly binary.

Besides the predictive performance, the Random Forest model provides interpretable information through feature importance. As seen in Figure 3, dynamic capability has the highest importance score, followed by workforce transformation and MSMEs performance.

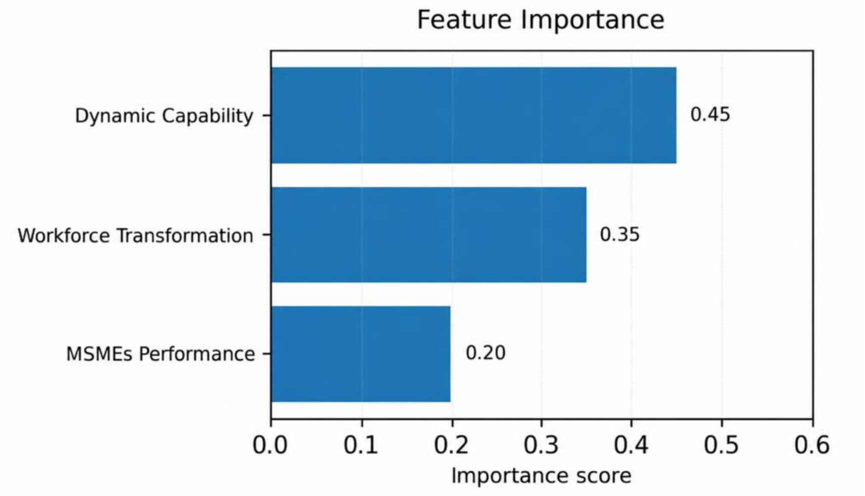
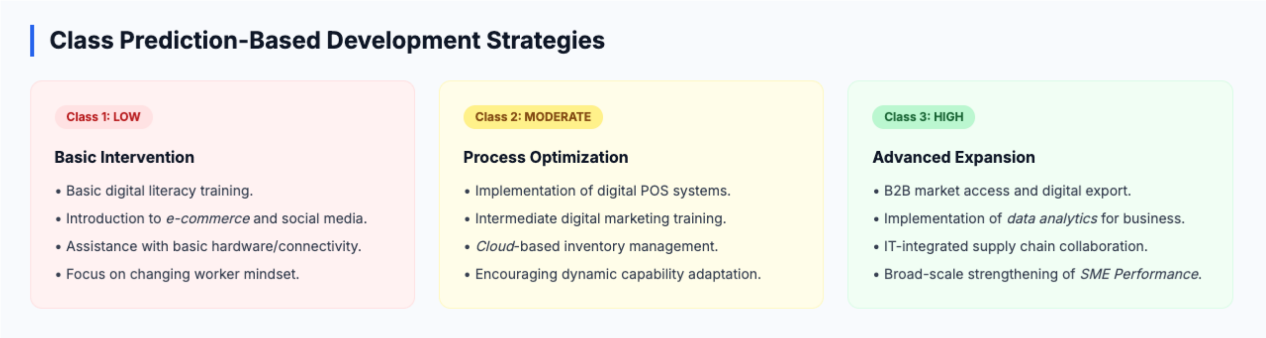


Figure 3. Feature importance of the Random Forest model

Dynamic capability obtained the highest feature importance value (0.451), indicating that the ability of MSMEs to sense market shifts, seize opportunities, and reconfigure resources is the most significant predictor of readiness classification. Workforce transformation contributed 0.351, emphasizing the role of digital skills and employee adaptability. MSMEs performance contributed 0.198, suggesting that business performance supports digital readiness but is less dominant than organizational agility and human resource transformation.

The dominance of dynamic capability is consistent with dynamic capability theory, which states that organizational agility the capacity to sense, seize, and reconfigure resources is a prerequisite for sustaining competitive advantage in changing environments. In the MSME context, this finding suggests that support programs should not focus only on hardware provision or platform onboarding. Capability-building interventions, such as managerial learning, process redesign, digital market sensing, and adaptive workforce development, are also needed to ensure that digital tools are translated into business value.

To translate the classification results into practical decision support, the predicted readiness classes were mapped into different MSME development strategies. Each class represents a different level of intervention need. Low-readiness MSMEs need basic digital literacy and infrastructure support, moderate-readiness MSMEs need process optimization and integration of digital tools and high-readiness MSMEs can be directed towards advanced expansion and business scaling. Figure 4 illustrates the prototype interface that presents class prediction-based development strategies for MSME assistance programs.



Class Prediction-Based Development Strategies

Class 1: LOW

Basic Intervention

- Basic digital literacy training.
- Introduction to *e-commerce* and social media.
- Assistance with basic hardware/connectivity.
- Focus on changing worker mindset.

Class 2: MODERATE

Process Optimization

- Implementation of digital POS systems.
- Intermediate digital marketing training.
- *Cloud*-based inventory management.
- Encouraging dynamic capability adaptation.

Class 3: HIGH

Advanced Expansion

- B2B market access and digital export.
- Implementation of *data analytics* for business.
- IT-integrated supply chain collaboration.
- Broad-scale strengthening of *SME Performance*.

Figure 4. Class prediction-based MSMEs development strategy interface

As shown in Figure 4, the proposed interface organizes the recommendation output into three readiness-based strategy groups. Class 1, with low readiness, deals with basic interventions, such as digital literacy training, introduction to e-commerce, hardware or connectivity support, and worker mindset development. Class 2, with moderate readiness, addresses process optimization, such as digital POS systems, integrated digital marketing, cloud-based inventory management, and dynamic capability adaptation. Class 3, with high readiness, enables advanced expansion, such as B2B market access, data analytics implementation, IT-integrated supply chain collaboration, and broader strengthening of MSMEs performance. This visualization demonstrates how the classification model can be connected to a practical decision support system for targeted MSME development.

The results support the use of the model as a decision support instrument. Table 5 maps the predicted classes into targeted recommendations. The purpose is not to replace policy judgment, but to provide a data-driven basis for prioritizing assistance programs. This interpretation aligns with previous business intelligence perspectives for MSMEs and with wider studies showing that predictive models can support operational and policy-related decisions [13], [22]-[32].

Table 5. Decision support recommendations based on predicted readiness

Predicted class	Profile interpretation	Recommended intervention
Low readiness	MSMEs with limited digital adoption, weak workforce readiness, and low organizational agility.	Basic digital literacy, marketplace onboarding, digital payment introduction, and intensive mentoring.
Moderate readiness	MSMEs that have started using digital channels but still have fragmented internal processes.	Integrated digital marketing, bookkeeping applications, product legality assistance, and operational standardization.
High readiness	MSMEs with strong dynamic capability, adaptive workforce, and better performance potential.	Advanced analytics training, business scaling support, export readiness programs, and access to growth financing.

From a practical perspective, the classification results can help local governments and MSMEs support institutions allocate programs more precisely. Low-readiness MSMEs require foundational support, moderate-readiness MSMEs need integration and consolidation, and high-readiness MSMEs can be prioritized for scaling and advanced digital programs. This targeted approach can reduce the risk of one-size-fits-all interventions and is consistent with policy-oriented digitalization frameworks that emphasize capability building, inclusive access, and staged support for smaller enterprises.

This study has several limitations that should be considered when interpreting the results. First, the labels were derived from clustering results, so they should be considered pseudo-labels rather than expert-validated ground truth. This means that the classifier operationalizes previous clustering patterns and may inherit any uncertainty from the clustering stage. Second, clustering validation metrics such as Silhouette Score, Davies–Bouldin Index, or Dunn Index were not available in the present dataset, so the reliability of pseudo-label formation cannot be fully assessed within this manuscript. Third, the model was evaluated using a single hold-out test set of 93 observations; therefore, future studies should apply stratified k-fold cross-validation and external regional validation. Fourth, comparison with other classifiers such as SVM, KNN, XGBoost, LightGBM, CatBoost, and Gradient Boosting should be conducted before claiming algorithmic superiority. Fifth, feature importance in this study provides global variable influence; future studies should compare it with permutation-based importance or local explanation methods such as SHAP to improve interpretability [19], [20].

CONCLUSION

This study developed a Random Forest classification model to predict MSMEs digital transformation readiness using cluster-derived labels. The model classified MSMEs into low, moderate, and high readiness levels and attained 91.40% accuracy, 91.36% macro precision, 92.16% macro recall, and 91.72% macro F1-score on the hold-out test set. The 95% confidence interval for accuracy suggests that the result is promising but still subject to sampling uncertainty. The confusion matrix revealed that most prediction errors occurred between neighboring readiness categories, which is reasonable for a gradual readiness construct.

The feature importance analysis indicated that dynamic capability was the most important predictor, followed by workforce transformation and MSMEs performance. The main scientific contribution of this study is the conversion of fuzzy clustering-based MSMEs readiness groups into a supervised predictive framework. Its methodological contribution lies in combining cluster-derived pseudo-labels, training-only imbalance handling, Random Forest classification, confidence-aware evaluation, and feature importance interpretation. Its practical contribution is the translation of predicted readiness classes into stratified intervention recommendations. Nevertheless, because the labels are pseudo-labels and the evaluation is limited to one regional dataset, future research should validate the labels with expert judgment, report clustering validation metrics, apply cross-validation, compare alternative classifiers, and test the framework in other regions before wider deployment.

ACKNOWLEDGMENTS

The authors thank the MSMEs respondents in Semarang for their participation and the institutional partners who supported the data collection process. The authors also extend their sincere gratitude to the Faculty of Industrial Technology and the Center for Industrial Intelligence Studies at Universitas Islam Sultan Agung (UNISSULA) for the computing facilities, academic support, and constructive environment provided. Furthermore, the authors appreciate the encouragement and valuable input provided throughout the development of the MSMEs digital transformation readiness study and the preparation of this manuscript.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest related to this study. The research was conducted independently, and no financial, personal, or institutional relationships influenced the design, data analysis, interpretation of results, or preparation of this manuscript.

REFERENCES

[1] M. Indriastuti and A. Riansyah, “Sustainable Performance: Evidence of SMEs in Indonesia,” Asian Journal of Business and Accounting, vol. 17, no. 1, pp. 291–313, Jun. 2024, doi: 10.22452/ajba.vol17no1.10.

[2] J. Secka and K. Szopik-Depczyńska, “Digital Innovativeness of SMEs: A Systematic Literature Review,” in Springer Proceedings in Complexity, A. Visvizi, O. Troisi, V. Corvello, and M. Grimaldi, Eds., Szczecin, Poland: Springer Science and Business Media B.V., 2026, pp. 803–820. doi: 10.1007/978-3-031-78623-5_62.

[3] N. A. Mutamimah and M. Indriastuti, “Fintech, financial literacy, and financial inclusion in Indonesian SMEs,” Int. J. Entrep. Innov. Manag., vol. 27, no. 1/2, p. 137, 2023, doi: 10.1504/ijeim.2023.129331.

[4] P. Mitropoulos, “Assessing the sustainability performance of SMEs across countries and sectors,” Environ. Impact Assess. Rev., vol. 116, 2026, doi: 10.1016/j.eiar.2025.108104.

[5] M. Alqahtani, N. Beloff, and M. White, “Decoding Cloud Adoption: A New Model Integrating Quantitative and Qualitative Insights for Saudi SMEs,” in Lecture Notes in Business Information Processing, E. W. Ziemba, W. Chmielarz, and J. Watróbski, Eds., Bisha, Saudi Arabia: Springer Science and Business Media Deutschland GmbH, 2026, pp. 197–234. doi: 10.1007/978-3-031-93580-0_9.

[6] R. Boselli, M. Cesarini, F. Mercorio, and M. Mezzanzanica, “Classifying online Job Advertisements through Machine Learning,” Future Generation Computer Systems, vol. 86, pp. 319–328, 2018, doi: 10.1016/j.future.2018.03.035.

[7] B. Kıyak, N. Biswas, F. Ertam, H. F. Öztop, and F. Selimefendigil, “Thermal performance of a fan-driven airflow domain with AI analysis using LightGBM classifier and regression,” International Communications in Heat and Mass Transfer, vol. 170, Jan. 2026, doi: 10.1016/j.icheatmasstransfer.2025.110011.

[8] N. Aggarwal, B. S. Saini, and S. Gupta, “A multi-layered stacked classifier developed for diagnosis of Parkinson’s disease and SWEDD patients using fusion of multimodal data,” Biomed. Signal Process. Control, vol. 108, no. February, p. 107924, 2025, doi: 10.1016/j.bspc.2025.107924.

[9] M. Çelik, “Antimicrobial Use at a Tertiary-Care Hospital in Türkiye According to the WHO Access, Watch and Reserve (AWaRe) Classification: A Point Prevalence Survey,” Infectious Diseases and Clinical Microbiology, vol. 7, no. 3, pp. 263–272, 2025, doi: 10.36519/idcm.2025.707.

[10] I. Abdelbaky, M. Ahmed, and M. Taha, “Machine learning classification approaches for prediction of effective diabetes drugs,” Egyptian Informatics Journal, vol. 31, no. October 2024, p. 100786, 2025, doi: 10.1016/j.eij.2025.100786.

[11] J. Sun, “Behavior Spectrum-Based Pedestrian Risk Classification via YOLOv8-ByteTrack and CRITIC-Kmeans,” Applied Sciences Switzerland, vol. 15, no. 18, 2025, doi: 10.3390/app151810008.

[12] F. Zulfiqar, U. Ijaz Bajwa, and Y. Mehmood, “Multi-class classification of brain tumor types from MR images using EfficientNets,” Biomed. Signal Process. Control, vol. 84, no. October 2022, p. 104777, 2023, doi: 10.1016/j.bspc.2023.104777.

[13] A. Fitrianingrum, M. Indriastuti, A. Riansyah, A. Basir, and D. Rusdi, “Business Intelligence: Alternative Decision-Making Solutions on SMEs in Indonesia,” in Lecture Notes on Data Engineering and Communications Technologies, vol. 161, Springer Science and Business Media Deutschland GmbH, 2023, pp. 500–507. doi: 10.1007/978-3-031-26281-4_52.

[14] A. Riansyah, M. Qomaruddin, M. Indriastuti, and M. Sagaf, “Clustering Digital Transformation of Small and Medium Enterprises (SMEs) Using Fuzzy K-means Method,” in International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2023, pp. 540–544. doi: 10.1109/eecsi59885.2023.10295664.

[15] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[16] T. K. Ho, “Random decision forests,” in Proceedings of the 3rd International Conference on Document Analysis and Recognition, 1995, pp. 278–282, doi: 10.1109/ICDAR.1995.598994.

[17] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[19] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.

[20] S. M. Lundberg et al., “From local explanations to global understanding with explainable AI for trees,” Nature Machine Intelligence, vol. 2, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[21] Q. Xia, X. Liu, Y. Li, Y. Xiong, and Q. Wu, “Shear fatigue characteristics and random forest-based life prediction model of ancient masonry under hygrothermal environment,” Construction and Building Materials, vol. 494, Oct. 2025, doi: 10.1016/j.conbuildmat.2025.143482.

[22] M. Jiang, J. Wang, L. Hu, and Z. He, “Random forest clustering for discrete sequences,” Pattern Recognition Letters, vol. 174, pp. 145–151, 2023, doi: 10.1016/j.patrec.2023.09.001.

[23] M. Domb, E. Bonchek-Dokow, and G. Leshem, “Lightweight adaptive Random-Forest for IoT rule generation and execution,” Journal of Information Security and Applications, vol. 34, pp. 218–224, 2017, doi: 10.1016/j.jisa.2017.03.001.

[24] P. D. F. Isles, “A random forest approach to improve estimates of tributary nutrient loading,” Water Research, vol. 248, p. 120876, 2024, doi: 10.1016/j.watres.2023.120876.

[25] Z. Mustaffa and M. H. Sulaiman, “Random forest based wind power prediction method for sustainable energy system,” Cleaner Energy Systems, vol. 12, Dec. 2025, doi: 10.1016/j.cles.2025.100210.

[26] F. A. Khalifa, H. M. Abdelkader, and A. H. Elsaid, “An analysis of ensemble pruning methods under the explanation of Random Forest,” Information Systems, vol. 120, p. 102310, 2024, doi: 10.1016/j.is.2023.102310.

[27] T. P. Dinh, C. Pham-Quoc, T. N. Thinh, B. K. Do Nguyen, and P. C. Kha, “A flexible and efficient FPGA-based random forest architecture for IoT applications,” Internet of Things, vol. 22, 2023, doi: 10.1016/j.iot.2023.100813.

[28] K. Pramilarani and P. Vasanthi Kumari, “Cost based Random Forest Classifier for Intrusion Detection System in Internet of Things,” Applied Soft Computing, vol. 151, p. 111125, 2024, doi: 10.1016/j.asoc.2023.111125.

[29] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, “An improved random forest based on the classification accuracy and correlation measurement of decision trees,” Expert Systems with Applications, vol. 237, p. 121549, 2024, doi: 10.1016/j.eswa.2023.121549.

[30] S. Zhu and W. Jiang, “A comprehensive evaluation of environmental quality for health ecotourism in Huangshan National Forest Park,” Mobile Information Systems, vol. 2022, 2022, doi: 10.1155/2022/8666276.

[31] A. Ismail et al., “Enhancing rock slope stability prediction using random forest machine learning: A case study,” China Geology, vol. 8, no. 4, pp. 691–706, Nov. 2025, doi: 10.31035/cg2023102.

[32] R. M. Longo, A. L. da Silva, A. I. Ribeiro, R. C. Gomes, F. C. Sperandio, and A. N. Nunes, “Evaluating the environmental quality of forest remnants using landscape metrics,” Sustainability, vol. 16, no. 4, 2024, doi: 10.3390/su16041543.