JIEET (Journal of Information Engineering and Educational Technology)

Volume 9, Number 2, 2025 pp. 89-100
E-ISSN : 2549-869X
Open Access: https://journal.unesa.ac.id/index.php/jieet

A Hybrid Clustering–Classification Framework for SMEs Success Level Prediction

Andika Dermawan Saputra[1], Wiyli Yustanti[2*]
[1,2] Program Study of Information Systems, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia

Corresponding author: wiyliyustanti@unesa.ac.id

ARTICLE INFO

Article history:
Submitted 22-11-2025
Accepted 5-12-2025
Available online 6-12-2025

Keywords:
SME;
Hybrid Method;
Prediction;
Classification;
K-Means Clustering;

DOI:
https://doi.org/10.26740/jieet.v9n2.p89-100

ABSTRACT

Micro, Small, and Medium Enterprises (SMEs) are vital to economic growth, yet their complex success determinants necessitate advanced predictive modelling. This study proposes a hybrid clustering-classification framework to classify and predict SME success levels based on 22 multidimensional indicators, including financial literacy, FinTech adoption, and entrepreneurial resilience. K-Means clustering was first applied to the survey data, yielding three optimal success personas, validated by the highest Silhouette Score (0.5238). These clusters were labelled as Beginner, Conventional, Stable Digital Adopter, and Digital Innovator SMEs. These empirically derived clusters served as pseudo-labels for the classification stage. Classification algorithms were tested with and without the Synthetic Minority Oversampling Technique (SMOTE). While ensemble methods (Random Forest, LightGBM) and SVM performed well, the K-Nearest Neighbours (KNN) algorithm consistently outperformed all others, achieving the highest F1-Score (0.9324) under SMOTE implementation. The findings validate the effectiveness of the hybrid clustering-classification approach in accurately mapping and predicting SME success levels. The resulting model serves as a robust, data-driven tool for policymakers to guide targeted interventions and digital training programs, fostering sustainable SME development.

INTRODUCTION

Small and Medium Enterprises (SMEs) play a pivotal role as the backbone of global economic development. Their success directly correlates with national growth, employment creation, and long-term financial stability. In many emerging economies, including Indonesia, SMEs account for a significant share of GDP and drive innovation and social inclusion. However, the pathway to SME success is often challenged by limited financial capital, managerial literacy gaps, technological adaptation barriers, and supply chain vulnerabilities, which collectively complicate the measurement and prediction of success in a structured and reliable manner. (Putri et al., 2025).

The digital transformation era has introduced opportunities for SMEs to overcome these challenges by adopting financial technologies (FinTech), digital payment systems, and ecommerce platforms. Empirical evidence suggests that FinTech innovation can significantly accelerate enterprise transformation by providing greater access to financial services and enhancing operational efficiency. (Luo et al., 2022). At the same time, sustainable practices, such as adopting circular economy principles and implementing green economy initiatives, have been shown to strengthen SMEs' social, economic, and environmental performance. (Astadi et al., 2022; Rodríguez-Espíndola et al., 2022). Moreover, knowledge management and organizational learning are critical mediators of sustainable innovation within SMEs, enabling firms to remain competitive in dynamic markets. (Al Koliby et al., 2025) . Beyond internal capabilities, mechanisms of appropriation, such as informal knowledge protection and supply chain resilience, also influence the long-term sustainability of SMEs (Abbas et al., 2020; Morales et al., 2022).

Despite these advances, defining the level of SME success remains challenging. Success is often reduced to narrow financial metrics, leaving multidimensional factors, such as financial literacy, entrepreneurial resilience, innovation adoption, and customer satisfaction. Previous studies have highlighted the need for holistic frameworks that incorporate both financial and non-financial dimensions to evaluate SME performance. (Ramdan et al., 2022; Lisi et al., 2024). Consequently, there is a growing demand for analytical models that capture these multidimensional features and provide robust predictive insights.

Recent advances in machine learning have demonstrated that hybrid clustering–classification frameworks, which combine unsupervised clustering with supervised classification, can uncover hidden structures in multidimensional data and improve predictive accuracy. Traditional supervised classification models often focus solely on the mapping between features and labels, ignoring underlying sample structures. (Xiao et al., 2020). Hybrid approaches overcome this limitation by first applying clustering to discover latent groups, which then serve as enriched labels for downstream classification tasks. (Dakhil et al., 2024; Xiao et al., 2020). For instance, hybrid clustering–classification models have been successfully implemented in domains such as customer churn prediction. (Bilal et al., 2022; Liu et al., 2022), traffic state classification (Shang et al., 2022), intrusion detection (Samunnisa et al., 2023), medical diagnosis (Bhavna et al., 2021; Kumar et al., 2019) and network classification (Chenghu & Thammano, 2024; Fahad et al., 2014), all of which report significant improvements over single-method models. Furthermore, novel approaches such as entropy-weighted clustering (Du, 2023; Qiao & Zhang, 2019) and end-to-end clustering-classification networks (Li et al., 2024; Pei et al., 2019) Illustrate the flexibility and generalizability of hybrid frameworks across industries.

Building on these insights, this study proposes a hybrid clustering–classification framework for predicting SME success levels. We conducted a literature review to identify 22 critical indicators of SME performance, including financial literacy, digital adoption, business growth, and entrepreneurial resilience. Using survey data collected from 150 SMEs, K-Means clustering was applied to group enterprises into three distinct clusters representing low, medium, and high success levels. These clusters were subsequently used as class labels to train supervised classification models, enabling predictive assessment of SME success based on multidimensional features. The main contributions of this study are threefold: (1) proposing a hybrid clustering–classification framework tailored for SME success prediction, (2) empirically validating the framework on primary survey data encompassing financial, managerial, technological, and entrepreneurial dimensions, and (3) providing actionable insights for policymakers and practitioners to support SME growth strategies.

METHOD

This study employs the Knowledge Discovery in Databases (KDD) framework to construct a predictive model for classifying SME success levels. The KDD process is selected because it enables the systematic transformation of raw survey data into actionable knowledge. The methodology is organised into three main components: instrument design and data collection, KDD stages, and knowledge representation. In general, the research framework can be described as Figure 1.

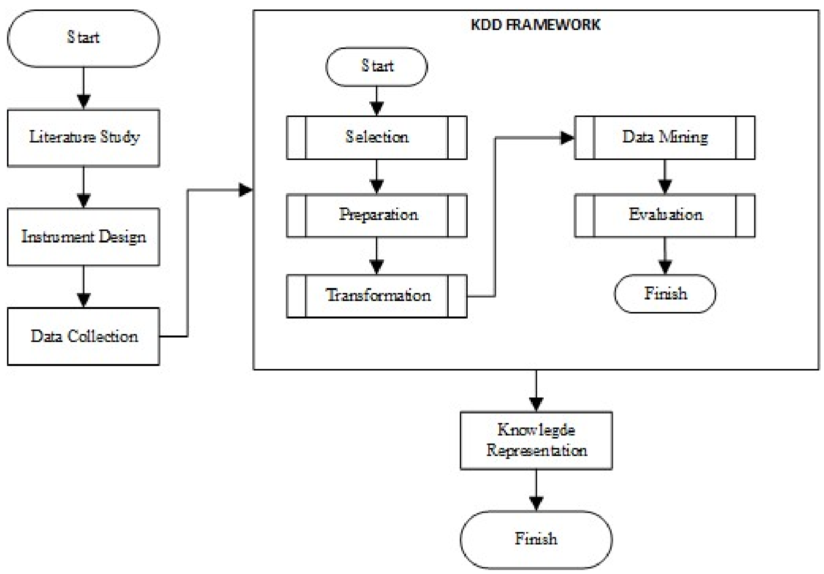

Figure 1. Research Framework

The results of the literature study are described in the previous section on prior research. This section explains the method for designing survey instruments for data collection and the stages of the KDD process.

A. Instrument Design and Data Collection

To capture the multidimensional nature of SME success, this study developed a structured questionnaire comprising 22 items, each measured on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). The instrument was designed based on prior literature that emphasises the role of financial literacy, digital adoption, business performance, and entrepreneurial resilience as critical determinants of SME sustainability and growth. Accordingly, the items were grouped into four dimensions: (i) Financial Literacy and Management, which measures the ability to manage financial resources effectively; (ii) Digital Adoption and Fintech, which evaluates the integration of digital tools and financial technology in business processes; (iii) Business Performance and Growth, which reflects operational outcomes such as revenue growth, customer satisfaction, and promotion intensity; and (iv) Entrepreneurial Capability and Resilience, which captures innovation, motivation, social networks, and adaptability to risk. The variables derived from the survey are summarised in Table I, where each item is treated as an input feature for clustering and classification. The questionnaire was distributed online using a purposive sampling technique, based on data from the trade office in Pacitan Regency, and yielded 150 respondents.

Table 1. Variables Dimension and Types

Dimension	Variable		Type (Likert Scale)
Financial Literacy and Management	X_1	Financial statement comprehension	Ordinal
	X_2	Access to formal financial services	Ordinal
	X_3	Knowledge of the break-even point	Ordinal
	X_4	Knowledge of sales safety margin	Ordinal
	X_5	Regular financial record-keeping	Ordinal
	X_6	Availability of capital	Ordinal
	X_7	Financial decision-making ability	Ordinal
Digital Adoption and Fintech	X_8	Use of digital media for marketing	Ordinal
	X_9	Provision of digital payment methods	Ordinal
	X_{10}	Use of fintech for transactions	Ordinal
	X_{11}	Knowledge of payment gateway systems	Ordinal
	X_{12}	Business process digitalisation	Ordinal
Business Performance and Growth	X_{13}	Short-term business growth	Ordinal
	X_{14}	Promotion intensity	Ordinal
	X_{15}	Long-term business growth	Ordinal
	X_{16}	Customer satisfaction	Ordinal
	X_{17}	Rational pricing strategy	Ordinal
Entrepreneurial Capability and Resilience	X_{18}	Product innovation	Ordinal
	X_{19}	Entrepreneurial motivation	Ordinal
	X_{20}	Risk-taking attitude	Ordinal
	X_{21}	Social network/ community support	Ordinal
	X_{22}	Governmental support received	Ordinal

B. KDD Stages

1) Data Selection: The first stage in the Knowledge Discovery in Databases (KDD) process is Data Selection, which focuses on defining and extracting the most relevant data for subsequent modelling. In this study, the dataset was derived from SME responses to a 22-item multidimensional questionnaire, as presented in Table 1. These items were carefully designed to capture essential aspects such as financial literacy, FinTech adoption, business growth, entrepreneurial resilience, and digital utilisation. The selection process ensured that all 22 features included in the analysis were theoretically grounded and directly aligned with the research objective, namely to model SME success clusters and classify their success levels. By systematically narrowing the dataset to only relevant and justifiable variables, the Data Selection stage establishes a solid foundation for the subsequent steps of the KDD framework.
2) Preprocessing: The second stage of the KDD framework is Preprocessing (Data Cleaning), which is essential to improve the quality and reliability of the dataset before further analysis. At this stage, several procedures were conducted to handle inconsistencies and potential noise within the survey responses. First, missing values were identified and imputed using statistical

methods, such as mode or mean substitution, to minimise information loss. Second, outliers that could distort the clustering process were examined and either adjusted or removed based on their statistical deviation from the data distribution. Third, logical and consistency checks were applied to ensure that responses across items remained coherent with the expected measurement scale. These preprocessing steps not only enhanced the dataset's validity but also ensured that the 22 selected features were clean, consistent, and ready for transformation into a modelling-ready format.

3) Transformation: The third stage of the KDD framework is Data Transformation, which prepares the cleaned dataset into a suitable representation for clustering and classification. In this study, two main transformation techniques were applied. First, all variables measured on a five-point Likert scale were normalised using Min-Max Scaling, converting the original values into a uniform range of [0,1]. This normalisation process ensured that no single variable dominated the clustering process due to differences in scale magnitude. Second, a dimensionality reduction procedure was considered using Principal Component Analysis (PCA) to summarise the dataset while retaining most of the variance. Although all 22 features were theoretically significant, PCA was tested as an optional step to evaluate whether reducing feature dimensionality could enhance computational efficiency and reveal latent structures in the data. Through these transformations, the dataset became standardised and analytically robust, providing a solid foundation for the subsequent data mining stage.
4) Data Mining: The fourth stage in the KDD process is Data Mining, which constitutes the core of knowledge discovery. In this study, a hybrid two-stage modelling scheme was implemented by integrating unsupervised and supervised learning techniques. In the first stage, K-Means clustering was used to uncover the dataset's natural structure and generate pseudo-labels for different SME success levels. The optimal number of clusters was guided by both theoretical considerations and validation using the Silhouette Score and Elbow method. The resulting clusters corresponded to three distinct success profiles. Each cluster was then assigned a categorical label based on the mean values of key performance-related features. In the second stage, the pseudo-labels obtained from clustering were used as ground-truth for supervised classification. The labelled dataset was split into training and test sets at a 80:20 ratio, ensuring balanced representation. Several classification algorithms were tested, including Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), and K-Nearest Neighbours (KNN). To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the class distribution in the training data. This hybrid clustering–classification framework allowed the model to capture hidden patterns in the dataset while ensuring reliable predictive performance, thereby strengthening its ability to classify SMEs into appropriate success levels. The method proposed in this study can be illustrated in Fig. 2.

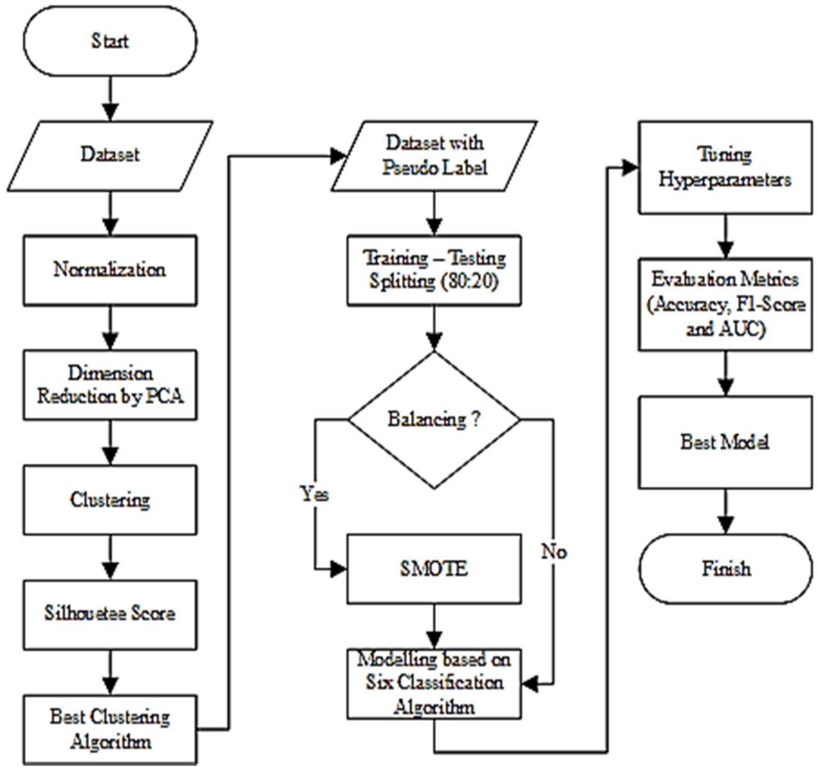

Figure 2. Two-Stage Hybrid Modelling in Data Mining

C. Knowledge Representation

This phase focuses on transforming analytical results into practical tools that can be easily accessed and utilised by stakeholders. In this study, the best-performing clustering–classification model was operationalised into a mobile-based application designed to evaluate SME success levels. This application serves a dual purpose. For government agencies and policymakers, it provides a decision-support tool that enables monitoring of SME conditions at scale, thereby assisting in the design of targeted interventions, training programs, and financial support policies. For SME owners, the application offers a self-assessment platform where entrepreneurs can enter relevant business information and receive immediate feedback on their success level. In addition, the system highlights the most influential factors driving the prediction, providing actionable insights for improvement. By embedding the predictive model into a mobile application, the knowledge gained from this research is represented in an intuitive, user-friendly format that bridges the gap between advanced data analytics and real-world business decision-making. This ensures that the research outcomes are not only theoretically valid but also practically impactful.

RESULTS AND DISCUSSION

This section presents the results of the clustering and classification experiments, along with their interpretations. The analysis focuses on determining the optimal number of clusters, evaluating the clustering quality, and assessing the classification performance. The optimal number of clusters was first determined using the Elbow Method, which analyses the relationship between the number of clusters and the within-cluster sum of squares (WCSS). The curve revealed a clear inflexion point at (k=3) as seen in Fig. 3, indicating that three clusters provide the most efficient balance between model simplicity and explanatory power.

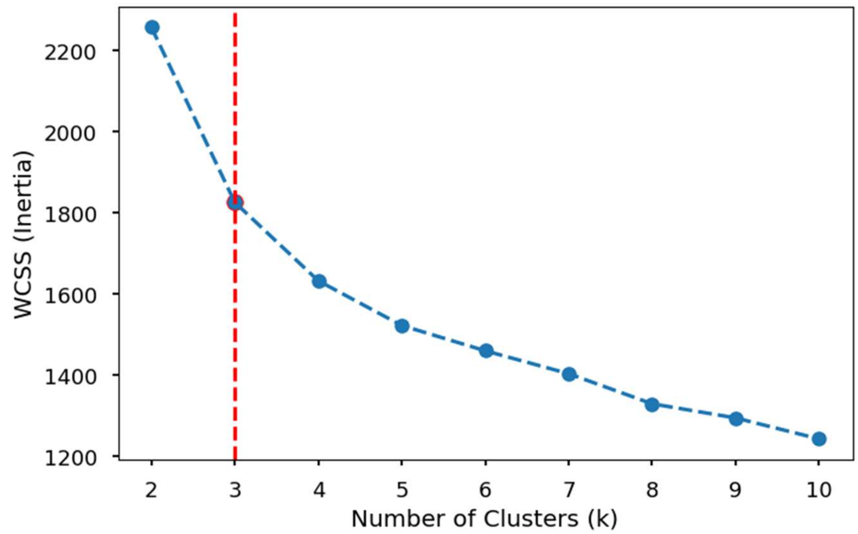

Figure 3. Optimal Cluster based on the Elbow Method

Building on this result, the clustering evaluation was further performed using five algorithms, such as K-Means, Gaussian Mixture Model (GMM), Agglomerative Clustering, BIRCH, and MiniBatch K-Means, each applied with three clusters. The comparative performance was assessed through the Silhouette Score as a validity index to measure the cohesion and separation of the clusters as described in Table II.

Table 2. Silhouette Score Of Clustering

Clustering Algorithm	Silhouette Score
K-Means	0.5238
GMM	0.3574
Agglomerative	0.4481
BIRCH	0.5221
MiniBatch K-Mean	0.4588

The results demonstrate that K-Means achieved the highest Silhouette Score (0.5238), slightly outperforming BIRCH (0.5221), while MiniBatch K-Means (0.4588), Agglomerative Clustering (0.4481), and GMM (0.3574) showed weaker clustering performance. Although the absolute Silhouette values are modest, the consistency of both the Elbow Method and Silhouette analysis confirms that K-Means provides the most effective partitioning of the SME dataset into three distinct clusters. Consequently, K-Means was selected as the primary clustering algorithm for the subsequent hybrid clustering–classification framework, as it captures the underlying data structure more reliably and generates robust pseudo-labels for classification. Visually, the clustering results can be presented as a scatter diagram using the PCA dimensions, as shown in Fig. 4.

.

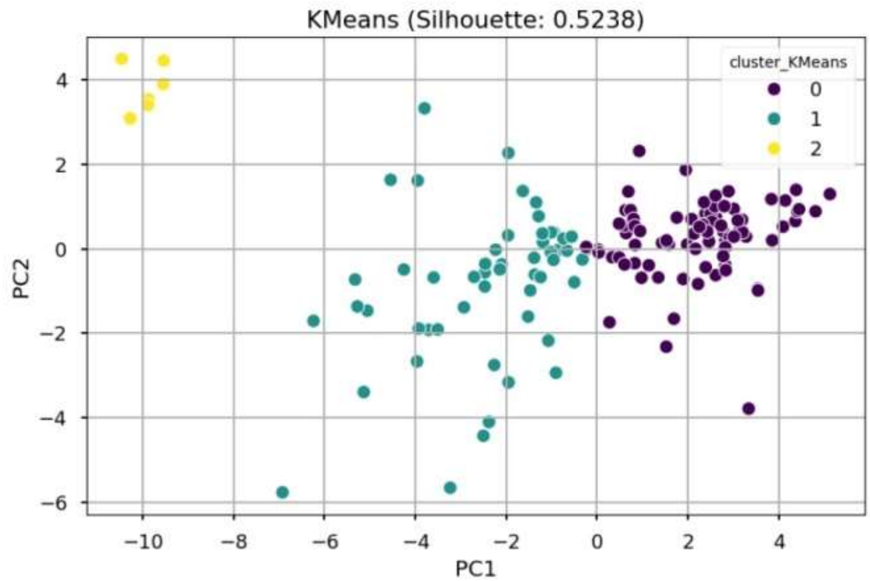

Figure 4. Scatter Diagram SMEs based on K-Means Clustering

The clustering results indicate that the dataset was divided into three distinct groups with varying proportions. Cluster 0 emerged as the dominant group, comprising 61.1% of the total samples, while Cluster 1 accounted for 34.9%. In contrast, Cluster 2 represented only 4% of the dataset, making it the smallest segment. These proportions suggest that the majority of SMEs share characteristics represented in Cluster 0, whereas a smaller portion exhibits the distinct features captured in Cluster 2.

To construct meaningful pseudo-labels for each cluster, it was essential first to examine its distinguishing characteristics. This process was carried out through descriptive statistical analysis of the dominant features derived from the principal components. Since none of the eigenvalues exceeded 1, a threshold of 0.25 was applied to the component loadings to identify the most influential variables. From this analysis, PC1 was characterised by two variables with significant loadings: X_1 (0.253364) and X_4 (0.251857). In contrast, PC2 showed more substantial loadings, with six variables exceeding the threshold: X_{13} (0.456389), X_{22} (0.435596), X_{19} (0.423412), X_{20} (0.348936), X_{17} (0.313519), and X_{15} (0.279292). These loading values reflect the degree of influence each variable has on its respective component, thereby serving as the foundation for interpreting cluster profiles. By mapping these dominant features, pseudo-labels can be generated to provide theoretical meaning to the numerical clusters, ensuring their relevance to the study's objectives.

The descriptive statistical analysis of the dominant features for each cluster characterises distinct SME profiles. Cluster 0, which represents 61.1% of the SMEs, is labelled 'Beginners' and 'Conventional'. This group is characterised by traditional business practices, limited digital adoption, constrained financial capital, and relatively simple marketing strategies. Cluster 1, accounting for 34.9% of the SMEs, is identified as Digital Adopters and Stable. Members of this cluster show early adoption of digital technologies such as online marketing, demonstrate better financial management, and exhibit greater stability in sustaining their businesses. Finally, Cluster 2, which represents 4.0% of the SMEs, is classified as Digital

Innovators. SMEs in this cluster are distinguished by strong product innovation, extensive use of digital technologies, greater access to financial capital, and a broader market orientation. These results highlight the progressive spectrum of SME development, ranging from conventional practices to advanced digital innovation.

Table 3. F1-Score Of Classification Performance

Model	With SMOTE (Testing)	Without SMOTE (Testing)	Observation
Decision Tree	0.7999	0.8358	Slightly better without SMOTE, but risk of overfitting (Train=1.0).
KNN	0.9324	0.9324	Most stable and consistent across both setups.
LightGBM	0.8687	0.8694	Very similar performance in both cases.
Logistic Regression	0.7104	0.8306	Performed much better without SMOTE, though less effective overall.
Random Forest	0.7999	0.9009	Higher score without SMOTE, but a perfect training score suggests overfitting.
SVM	0.8067	0.8622	Improved without SMOTE, but still below KNN and Random Forest.

Subsequently, the clustering labels, comprising the three SME categories (Beginners & Conventional, Digital Adopters & Stable, and Digital Innovators), were utilised as the target variable in the classification stage. In this way, the classification process aimed to learn relationships between the survey variables and pseudo-labels derived from clustering, thereby enabling predictive modelling of the success rates of new SMEs. To further evaluate the robustness of the models, classification was conducted under two conditions: with data balancing using the SMOTE technique and without balancing. The results demonstrate notable differences (Table III). When balancing was applied, the K-Nearest Neighbours (KNN) model achieved the highest F1-Score on the test dataset (0.9324), demonstrating strong generalisation across the training and validation phases. Ensemble-based methods such as Random Forest and LightGBM also performed effectively, although slightly below KNN. Meanwhile, Logistic Regression produced lower scores, indicating limited capacity to handle the dataset's nonlinear relationships. In the non-balancing scenario, KNN's performance remained consistently high, with an F1-score of 0.9324, reaffirming its stability across both experimental setups. Random Forest showed improved testing performance (0.9009) compared to the balanced case, but this was accompanied by a tendency toward overfitting, as reflected in perfect training scores. Similarly, other models such as LightGBM, SVM, and Decision Tree maintained competitive results, though slightly less balanced between training and validation compared to the SMOTE-applied case.

Overall, the findings confirm that KNN is the most stable and reliable model for predicting SME success levels, both under balanced and imbalanced conditions. This stability underscores its ability to capture and generalise the complex distributional patterns of SME survey data, making it particularly suitable as the final predictive model within the proposed hybrid clustering–classification framework. The best model is then implemented in a mobile application that can be used by related parties and MSMEs themselves to evaluate the condition of their business, including its level of success. Fig. 5 illustrates the application interface to be developed.

Figure 5. Mobile Application for SMEs Success Level Prediction

CONCLUSION

This study developed a hybrid clustering–classification framework to predict SME success levels using multidimensional survey data. K-Means was selected as the optimal clustering method, producing three distinct groups: Beginner and Conventional, Digital Adopters and Stable, and Digital Innovators. These clusters served as pseudo-labels for the classification stage. Among the tested algorithms, K-Nearest Neighbours (KNN) consistently achieved the highest and most stable performance (F1-Score of 0.9324), confirming its robustness in modelling SME patterns. The proposed framework provides a practical data-driven tool for policymakers and SMEs to evaluate success levels, design targeted interventions, and support digital growth.

REFERENCES

Abbas, J., Zhang, Q., Hussain, I., Akram, S., Afaq, A., & Shad, M. A. (2020). Sustainable innovation in small and medium enterprises: The impact of knowledge management on organisational innovation through a mediation analysis by using the SEM approach. *Sustainability (Switzerland)*, *12*(6). https://doi.org/10.3390/su12062407

Al Koliby, I. S., Noor, N. H. M., Al-Swidi, A. K., Al-Hakimi, M. A., & Mehat, N. A. B. (2025). Enhancing sustainable performance among manufacturing SMEs: the interplay of knowledge management and organisational structure. *Discover Sustainability*, *6*(1). https://doi.org/10.1007/s43621-025-01351-1

Astadi, P., Kristina, S., Retno, S., Yahya, P., & Agni Alam, A. (2022). The long path to achieving green economy performance for micro, small, and medium enterprises. In *Journal of Innovation and Entrepreneurship* (Vol. 11, Issue 1). https://doi.org/10.1186/s13731-022-00209-4

Bhavna, Verma, R., Handa, R., & Puri, V. (2021). A Hybrid Approach for Diabetes Prediction and Risk Analysis Using Data Mining. *Lecture Notes in Electrical Engineering*, *668*. https://doi.org/10.1007/978-981-15-5341-7_92

Bilal, S. F., Almazroi, A. A., Bashir, S., Khan, F. H., & Almazroi, A. A. (2022). An ensemble-based approach using a combination of clustering and classification algorithms to enhance customer churn prediction in the telecom industry. *PeerJ Computer Science*, *8*.

https://doi.org/10.7717/PEERJ-CS.854

Chenghu, C., & Thammano, A. (2024). A Novel Classification Model Based on Hybrid K-Means and Neural Network for Classification Problems. *HighTech and Innovation Journal*, *5*(3), 716–729. https://doi.org/10.28991/HIJ-2024-05-03-012

Dakhil, A. F., Ali, W. M., & Hasan, M. A. (2024). Applying Hybrid Clustering with Evaluation by AUC Classification Metrics. *International Journal of Computing and Digital Systems*, *15*(1). https://doi.org/10.12785/ijcds/150177

Du, X. (2023). A Robust and High-Dimensional Clustering Algorithm Based on Feature Weight and Entropy. *Entropy*, *25*(3). https://doi.org/10.3390/e25030510

Fahad, A., Alharthi, K., Tari, Z., Almalawi, A., & Khalil, I. (2014). CluClas: Hybrid clustering-classification approach for accurate and efficient network classification. *Proceedings - Conference on Local Computer Networks, LCN*. https://doi.org/10.1109/LCN.2014.6925769

Kumar, V. B., Vijayalakshmi, K., & Padmavathamma, M. (2019). A hybrid data mining approach for diabetes prediction and classification. *Lecture Notes in Engineering and Computer Science*, *2019-October*.

Li, L., Lu, Y., Yang, G., & Yan, X. (2024). End-to-End Network Intrusion Detection Based on Contrastive Learning. *Sensors*, *24*(7), 1–21. https://doi.org/10.3390/s24072122

Lisi, S., Mignacca, B., & Grimaldi, M. (2024). *Non-financial reporting and SMEs: A systematic review, research agenda, and novel conceptualisation. Journal of Management & Organisation, 30*(3), 600–622. https://doi.org/10.1017/jmo.2023.43

Liu, R., Ali, S., Bilal, S. F., Sakhawat, Z., Imran, A., Almuhaimeed, A., Alzahrani, A., & Sun, G. (2022). An Intelligent Hybrid Scheme for Customer Churn Prediction Integrating Clustering and Classification Algorithms. *Applied Sciences (Switzerland)*, *12*(18). https://doi.org/10.3390/app12189355

Luo, S., Sun, Y., Yang, F., & Zhou, G. (2022). Does fintech innovation promote enterprise transformation? Evidence from China. *Technology in Society*, *68*. https://doi.org/10.1016/j.techsoc.2021.101821

Morales, P., Flikkema, M., Castaldi, C., & Man, A. P. de. (2022). The effectiveness of appropriation mechanisms for sustainable innovations from small and medium-sized enterprises. *Journal of Cleaner Production*, *374*. https://doi.org/10.1016/j.jclepro.2022.133921

Pei, Y., Liu, J., Zhang, Y., & Qu, Y. (2019). An End-to-End Clustering and Classification Learning Network. *Proceedings - 2nd China Symposium on Cognitive Computing and Hybrid Intelligence, CCHI 2019*. https://doi.org/10.1109/CCHI.2019.8901950

Putri, A. N. A., Hermawan, P., Mirzanti, I. R., Meadows, M., & Sadraei, R. (2025). Unpacking green growth in SMEs: A framework for dynamic capabilities, value co-creation, and sustainable performance. *Sustainable Futures*, *10*(October 2024), 100840. https://doi.org/10.1016/j.sftr.2025.100840

Qiao, C., & Zhang, W. (2019). Research on classification method of crh maintenance parts based on entropy weight-clustering analysis. *Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019*. https://doi.org/10.1109/ITAIC.2019.8785429

Ramdan, M. R., Aziz, N. A. A., Abdullah, N. L., Samsudin, N., Singh, G. S. V., Zakaria, T., Fuzi, N. M., & Ong, S. Y. Y. (2022). SMEs Performance in Malaysia: The Role of Contextual Ambidexterity in Innovation Culture and Performance. *Sustainability (Switzerland)*, *14*(3), 1–18. https://doi.org/10.3390/su14031679

Rodríguez-Espíndola, O., Cuevas-Romo, A., Chowdhury, S., Díaz-Acevedo, N., Albores, P., Despoudi, S., Malesios, C., & Dey, P. (2022). The role of circular economy principles and sustainable-oriented innovation to enhance social, economic and environmental performance: Evidence from Mexican SMEs. *International Journal of Production Economics*, *248*. https://doi.org/10.1016/j.ijpe.2022.108495

Samunnisa, K., Kumar, G. S. V., & Madhavi, K. (2023). Intrusion detection system in distributed cloud computing: Hybrid clustering and classification methods. *Measurement: Sensors*, *25*. https://doi.org/10.1016/j.measen.2022.100612

Shang, Q., Yu, Y., & Xie, T. (2022). A Hybrid Method for Traffic State Classification Using K-Medoids Clustering and Self-Tuning Spectral Clustering. *Sustainability (Switzerland)*, *14*(17). https://doi.org/10.3390/su141711068

Xiao, J., Tian, Y., Xie, L., Jiang, X., & Huang, J. (2020). A Hybrid Classification Framework Based on Clustering. *IEEE Transactions on Industrial Informatics*, *16*(4). https://doi.org/10.1109/TII.2019.2933675