Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 305-316
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Comparison of the Random Forest and LightGBM Algorithms in MSME Performance Classification

Juwita Permata Sari*, Bunga Sabila

Information System Study Program, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Jl. Jenderal Sudirman Blok A No. 1/2/3, Siantar Barat, Kota Pematang Siantar, Sumatera Utara, Indonesia
Email: [1,*]permatasarij84@gmail.com, [2]sabilabunga89@gmail.com
Correspondence Author Email: permatasarij84@gmail.com
Submitted: **01/06/2026**; Accepted: **12/06/2026**; Published: **30/06/2026**

Abstract−The performance assessment of Micro, Small, and Medium Enterprises (MSMEs) requires a data-driven approach, as business performance is influenced by multiple dimensions, including financial, customer, digital, operational, and business environment indicators. This study aims to compare the performance of Random Forest (RF) and Light Gradient Boosting Machine (LightGBM) models in classifying MSME performance into four categories: Critical, Struggling, Growth, and Elite. The study utilized a synthetic secondary dataset obtained from Kaggle, consisting of 150,000 records and 15 attributes, which were processed using Google Colab. The research methodology comprised data preprocessing, categorical feature encoding, feature-target separation, training and testing data partitioning using an 80:20 ratio with stratified sampling, model development, model testing, performance evaluation, and comparative analysis. Model performance was assessed using accuracy, macro-precision, macro-recall, macro F1-score, and confusion matrix analysis. The experimental results demonstrate that RF achieved superior performance, with an accuracy of 0.998267, a macro-precision of 0.997563, a macro-recall of 0.996832, and a macro F1-score of 0.997197. In comparison, LightGBM achieved an accuracy of 0.997567, a macro-precision of 0.996465, a macro-recall of 0.996517, and a macro F1-score of 0.996491. Furthermore, the confusion matrix analysis revealed that RF generated 52 misclassifications out of 30,000 testing instances, whereas LightGBM generated 73 misclassifications. Feature importance analysis of the RF model identified Burn_Rate_Ratio, Net_Profit_Margin (%), Avg_Historical_Rating, Repeat_Order_Rate (%), and Peak_Hour_Latency as the most influential variables in distinguishing MSME performance categories. These findings indicate that RF is a highly effective model for MSME performance classification based on multidimensional business indicators. Nevertheless, further validation using empirical MSME datasets is recommended to enhance the generalizability and practical applicability of the proposed approach.

Keywords: MSMEs; Random Forest; LightGBM; Machine Learning; Classification

1. INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs) play a crucial role in supporting economic growth, creating employment opportunities, and enhancing community welfare (Isabella & Sanjaya, 2022). MSME performance is an important aspect to consider, as it reflects a firm's ability to sustain operations, improve competitiveness, and adapt to changes in the business environment (Nuraini & Darmawan, 2024). The improvement of MSME performance is influenced by both internal and external factors, including managerial capabilities, environmental uncertainty, technology adoption, information systems, and business control mechanisms (Alansori & Listyaningsih, 2022; Hadi, 2022). In the digital era, MSME performance can no longer be assessed solely based on revenue or profit, as business activities are increasingly associated with e-commerce, accounting information systems, financial literacy, financial technology, digital payment systems, innovation, and customer behavior (Ernis, 2024; Jannah, 2025; Masriah & Ispriyahadi, 2025; Ulyasari et al., 2023). Therefore, the classification of MSME performance is necessary to map business conditions into more distinct categories, enabling business owners and stakeholders to understand the position of an enterprise more quickly, objectively, and in a data-driven manner.

In the field of machine learning, various algorithms have been employed to address classification tasks, including K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) (Erkamim et al., 2023; Shinzi, 2025). Methods such as KNN, DT, SVM, and ANN offer advantages in terms of implementation simplicity, interpretability, or the ability to capture specific classification patterns. However, their performance may be affected by feature scaling, parameter selection, data complexity, and the risk of overfitting when dealing with complex data structures (Erkamim et al., 2023; Shinzi, 2025).

Consequently, ensemble models based on decision trees have gained considerable attention due to their ability to improve classification stability and predictive performance through the aggregation of multiple base learners (Djaka & Winarsih, 2025; Firnanda et al., 2025). RF has emerged as a widely adopted approach because it employs a bagging strategy to combine multiple DTs, reduces reliance on a single decision tree, and provides feature importance measures that facilitate model interpretation (Alfianti & Supriyanto, 2024; Firnanda et al., 2025). LightGBM is also a relevant benchmark model, as it utilizes an efficient gradient boosting framework that offers high computational speed and is capable of handling large-scale datasets in classification tasks (Pratama et al., 2025; Zhao et al., 2024). Therefore, a comparison between RF and LightGBM is particularly important because both methods are based on decision trees yet employ fundamentally different learning mechanisms, namely bagging and boosting.

Several previous studies have shown that machine learning can be used to support business performance prediction and classification. First, research conducted by (Saputra & Yustanti, 2025) developed a hybrid clustering-

classification framework to predict the success rate of Small and Medium Enterprises (SME) based on multidimensional indicators, such as financial literacy, FinTech adoption, and entrepreneurial resilience. The results of this study indicate that the machine learning approach is capable of mapping business success levels based on various indicators, with KNN achieving the highest F1 score of 0.9324 after applying the Synthetic Minority Oversampling Technique (SMOTE) (Saputra & Yustanti, 2025). An important contribution of this study lies in the use of multidimensional indicators closely related to the context of SME success. However, this study has not specifically compared RF and LightGBM for classifying MSME performance into operational categories such as 'Critical', 'Struggling', 'Growth', and 'Elite'. Second, research conducted by (Erkamim et al., 2023) compared RF and XGBoost in MSME performance classification. The results showed that RF achieved an accuracy and F1-score of 0.944, while XGBoost achieved an accuracy of 0.944 and an F1-score of 0.950 (Erkamim et al., 2023).

The strength of this study is its closeness to the MSME context and the use of an ensemble model to classify business performance. However, the research did not include LightGBM as a comparison model, and the indicators used were predominantly financial. Furthermore, (Ramadhan et al., 2025) Developing a predictive model for the success of MSMEs using DT, RF, and SVM.. The results of this study indicate that RF is able to provide good classification performance compared to other models (Ramadhan et al., 2025). This study demonstrates the potential of using decision tree-based algorithms to predict MSME success. However, it does not compare RF with LightGBM, which represents a bagging and boosting approach. In addition, (Bimawan et al., 2024) and (Shinzi, 2025) showed that different classification algorithms may produce different results depending on data characteristics; however, these studies were not specifically directed toward MSME performance classification based on multidimensional indicators.

Based on a review of previous research, an unresolved issue lies in how MSME performance classification models are constructed, evaluated, and interpreted. Several studies have used classification algorithms such as KNN, DT, SVM, RF, and XGBoost, but not many have specifically compared RF and LightGBM as two different tree-based ensemble approaches, namely bagging and boosting, in MSME performance classification (Erkamim et al., 2023; Saputra & Yustanti, 2025; Ulyasari et al., 2023). Furthermore, some studies still use limited indicators, especially on the financial aspect, so they do not fully represent MSME performance as a condition influenced by a combination of financial, customer, digital, operational, and business environment indicators (Masriah & Ispriyahadi, 2025). From an evaluation perspective, a more comprehensive model assessment needs to be conducted through accuracy, macro precision, macro recall, macro F1-score, confusion matrix, and feature importance. This ensures that the research results not only show the best-performing model but also explain the dominant indicators that differentiate each MSME performance category (Rainio, 2024; Zeng, 2025). Therefore, this study aims to develop a classification model that is not only predictive but also interpretable in mapping MSME performance levels.

To address these issues, this study proposes a comparison of the RF and LightGBM algorithms to classify MSME performance levels into four categories, namely `Critical`, `Struggling`, `Growth`, and `Elite`. RF is used because it has the ability to build stable predictions through DT sets and supports feature interpretation through feature importance (Alfianti & Supriyanto, 2024; Firnanda et al., 2025). LightGBM is used as a comparison model because of its training efficiency and its ability to handle tabular data-based classification tasks (Pratama et al., 2025; Zhao et al., 2024). Model evaluation is carried out using accuracy, precision macro, recall macro, F1-score macro, and confusion matrix, while feature importance analysis is used to identify indicators that contribute most to differentiating MSME performance categories (Rainio, 2024; Zeng, 2025). The objectives of this study are to determine the model with the best performance between RF and LightGBM and to identify the dominant features that influence MSME performance classification. The contribution of this research lies in providing an empirical comparison of two ensemble learning algorithms, the use of more comprehensive multiclass evaluation, and the presentation of feature interpretation as an initial basis for data-driven business decision making.

2. RESEARCH METHODS

2.1 Dataset

The dataset used in this study is a synthetic secondary dataset on the performance of Micro, Small, and Medium Enterprises (MSMEs) obtained from Kaggle under the name UMKM Dataset (Zkyfauzi, n.d.). This dataset is used to support the process of classifying MSME performance levels based on indicators representing financial, customer, digital, operational, and business environment aspects. The dataset file used is synthetic_umkm_data.csv with 150,000 data and 15 attributes. The target variable in this dataset is Class, which indicates the MSME performance categories, namely Critical, Struggling, Growth, and Elite. To support the repeatability of the experiment, a copy of the dataset and data processing notebook are also stored in the researcher's Google Drive repository. Some sample data used in this study are shown in Table 1 to provide an initial overview of the attribute structure in the dataset.

Table 1. Sample of MSME Dataset

No	Monthly_ Revenue	Net_Profit _Margin	Burn_Rat e_Ratio	Avg_Histori cal_Rating	Repeat_Or der_Rate	Digital_Adop tion_Score	Peak_Hour _Latency	Class
1	6,680,716	22.72	0.811	4.75	19.40	4.24	Low	Grow th

No	Monthly_ Revenue	Net_Profit _Margin	Burn_Rat e_Ratio	Avg_Histori cal_Rating	Repeat_Or der_Rate	Digital_Adop tion_Score	Peak_Hour _Latency	Class
2	5,819,101	4.46	0.968	4.21	14.87	1.27	Med	Grow th
3	5,236,404	-10.12	1.047	3.51	21.00	3.37	Med	Strug gling
…	…	…	…	…	…	…	…	…
150 000	6,071,256	4.22	0.954	4.34	20.87	2.67	Low	Grow th

Source: Repositori Google Drive peneliti: [Dataset]; diadaptasi dari Kaggle UMKM Dataset (Zkyfauzi, n.d.)

Based on Table 1, the dataset contains numeric and categorical attributes that describe the condition of MSMEs from several aspects. The Monthly_Revenue, Net_Profit_Margin (%), and Burn_Rate_Ratio attributes represent the financial condition of the business. The Avg_Historical_Rating and Repeat_Order_Rate (%) attributes describe customer response and loyalty. The Digital_Adoption_Score attribute indicates the level of digital adoption, while Peak_Hour_Latency describes operational conditions during peak hours. The Class variable is used as a classification target that indicates the MSME performance categories, namely Critical, Struggling, Growth, and Elite.

2.2 Research Framework

The research framework was developed to illustrate the experimental flow in comparing the RF and LightGBM algorithms in classifying MSME performance levels. The research flow begins with problem identification and literature review, dataset utilization, data preprocessing, data division using a train-test split, model training, model testing, model evaluation, model comparison, results and discussion, and conclusion drawing. The research framework is shown in Figure 1.

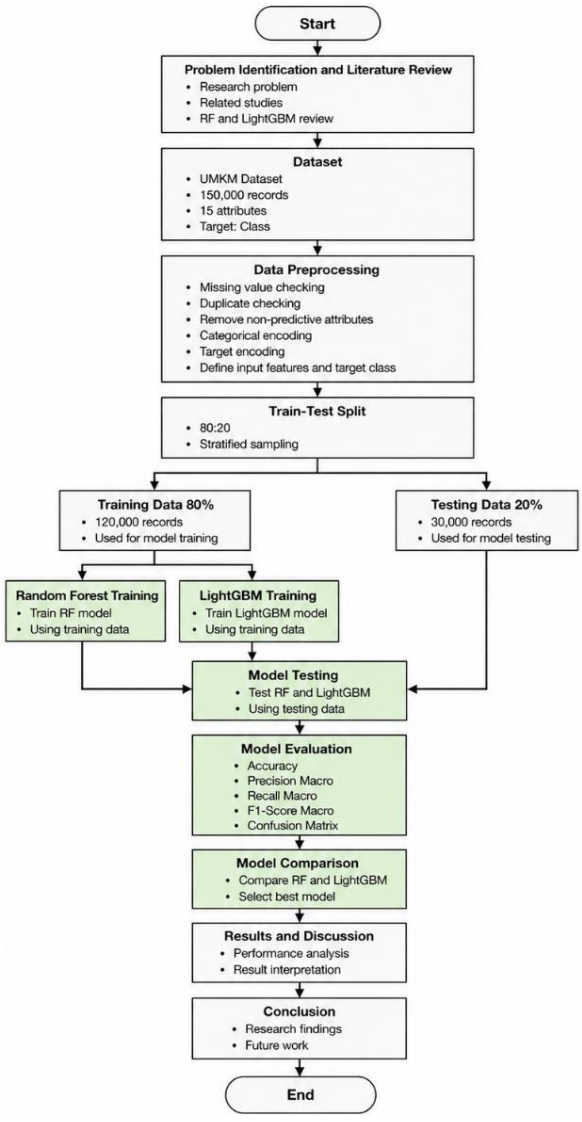

Figure 1. Research Framework of RF and LightGBM Comparison for MSME Performance Classification

Based on Figure 1, the research began with problem identification and literature review to strengthen the research foundation regarding MSME performance classification, the RF model, the LightGBM model, and multiclass classification evaluation. After the dataset was determined, it underwent a data preprocessing stage, which included checking for missing values, checking for duplicate data, removing non-predictive attributes, categorical encoding, target encoding, and determining input features and target classes. This stage was carried out to ensure the dataset was in a format suitable for use in the modeling process. After the preprocessing stage was completed, the dataset was divided using a train-test split scenario with a ratio of 80:20 and stratified sampling. This ratio was used to provide the model with sufficient training data to learn classification patterns, while still providing sufficient test data to assess the model's ability on data not yet used during training. Of the total 150,000 data sets, 120,000 data sets were used as training data and 30,000 data sets were used as test data.

The training data were used to train the RF and LightGBM models separately. After the training process was completed, both models were tested using the same test data to ensure consistent comparison of the evaluation results. The evaluation was conducted using accuracy, precision macro, recall macro, F1-score macro, and confusion matrix. The first four metrics were used to assess the numerical performance of the model, while the confusion matrix was used to observe the distribution of correct and incorrect predictions for each target class. The evaluation results of the RF and LightGBM models were then compared to determine the best-performing model. The results are then discussed in the results and discussion section before drawing conclusions. The evaluation formula used in this study was formulated according to the multiclass classification scenario and the metrics used in the model evaluation stage. The formula is shown below.

$$\text{Accuracy} = \frac{\sum_{i=1}^{K} \text{TP}_i}{N} \tag{1}$$

$$\text{Precision}_{\text{Macro}} = \frac{1}{K}\sum_{i=1}^{K} \frac{\text{TP}_i}{\text{TP}_i+\text{FP}_i} \tag{2}$$

$$\text{Recall}_{\text{Macro}} = \frac{1}{K}\sum_{i=1}^{K} \frac{\text{TP}_i}{\text{TP}_i+\text{FN}_i} \tag{3}$$

$$\text{F1-Score}_{\text{Macro}} = \frac{1}{K}\sum_{i=1}^{K} \frac{2\times\text{Precision}_i\times\text{Recall}_i}{\text{Precision}_i+\text{Recall}_i} \tag{4}$$

$$CM = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1K} \\ C_{21} & C_{22} & \cdots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \cdots & C_{KK} \end{bmatrix} \tag{5}$$

In these equations, K denotes the number of target classes, namely Critical, Struggling, Growth, and Elite, while N denotes the number of test data. TP_i denotes the number of class i data correctly predicted as that class, FP_i denotes the number of data from other classes incorrectly predicted as class i, and FN_i denotes the number of class i data incorrectly predicted as another class. Precision_i and Recall_i denote the precision and recall values for class i. CM denotes the confusion matrix, while C_{ij} denotes the number of data with class i that were actually predicted as class j .

2.3 Random Forest

RF is a decision tree-based ensemble learning algorithm that works by building multiple decision trees and combining the predictions from each tree to produce a final decision (Bimawan et al., 2024). In classification tasks, RF uses a bagging or bootstrap aggregating approach, which forms several data subsets from the training data, trains a decision tree on each subset, and then determines the final class based on a majority voting mechanism (Ramadhan et al., 2025). This approach makes RF more stable than using a single DT because the model's decisions do not rely solely on a single tree structure (Bimawan et al., 2024). This stability makes RF widely used in various classification tasks based on tabular data and business data (Erkamim et al., 2023; Iwung & Rahman, 2026).

In this study, RF was used to classify the performance levels of MSMEs into four categories: Critical, Struggling, Growth, and Elite. RF was chosen because of its ability to handle tabular data, process complex feature relationship patterns, and produce stable classification performance across various classification cases (Erkamim et al., 2023; V et al., n.d.). Several previous studies have shown that RF can provide competitive performance when compared to other algorithms such as DT, SVM, KNN, and XGBoost (Bimawan et al., 2024; Ramadhan et al., 2025). Furthermore, RF also has the advantage of providing feature importance information that can be used to understand the contribution of each attribute to the classification results (Adawiyah & Kartikasari, 2025; Alfianti & Supriyanto, 2024). This interpretative capability is important in tabular data-based research because model results are assessed not only by numerical performance but also by indicators that play a role in distinguishing target classes (Alfianti & Supriyanto, 2024; Erkamim et al., 2023).

The application of RF in this study was carried out using training data amounting to 80% of the entire dataset. The RF model was trained to learn the relationship between input features and target classes, then tested using 20% of the test data. The model configuration used included n_estimators=100, random_state=42, class_weight='balanced', and n_jobs=-1. The n_estimators=100 parameter was used to build 100 decision trees, random_state=42 was used to

ensure reproducibility of the experimental results, class_weight='balanced' was used to help the model pay more attention to class distributions, and n_jobs=-1 was used to allow the training process to utilize available computing resources. The results of the RF model testing were then evaluated using metrics established in the research framework, namely accuracy, precision macro, recall macro, F1-score macro, and confusion matrix. These evaluation results were used as a basis for assessing RF's ability to classify MSME performance and as a comparison against the LightGBM model in the model comparison stage.

2.4 LightGBM

LightGBM is a gradient boosting-based ensemble learning algorithm designed to improve training efficiency and classification performance on large datasets (Oueslati et al., 2025; Zhao et al., 2024). Unlike RF, which uses a bagging approach, LightGBM builds models incrementally through a boosting process, where each new tree is formed to correct errors from the previous model (Anggraini et al., 2024; Pratama et al., 2025). This mechanism enables LightGBM to produce competitive models in various classification tasks because the learning process is gradual and targeted towards prediction errors (Prajesh & Jain, 2025; Zhao et al., 2024). In this study, LightGBM was used as a comparison model against RF in classifying MSME performance levels. LightGBM was chosen because it has good computational efficiency and is able to handle large amounts of tabular data (Pratama et al., 2025). Furthermore, LightGBM is also relevant for use in multiclass classification tasks because it can process multiple target classes in a single modeling scenario (Oikonomou & Damigos, 2024; Zhao et al., 2024). The use of LightGBM as a benchmark allows this study to evaluate the performance differences between the bagging approach on RF and the boosting approach on LightGBM.

The implementation of LightGBM in this study was carried out using training data amounting to 80% of the entire dataset. The LightGBM model was trained using the same training data as RF to ensure consistent performance comparisons. The model configurations used included objective='multiclass', n_estimators=200, learning_rate=0.05, random_state=42, and class_weight='balanced'. The objective='multiclass' parameter was used because the classification target consists of four categories: Critical, Struggling, Growth, and Elite. The n_estimators=200 parameter was used to determine the number of tree formation iterations, while learning_rate=0.05 was used to adjust the model's learning rate. The random_state=42 parameter was used to maintain the consistency of the experimental results, while class_weight='balanced' was used to help the model pay more attention to class distributions. The results of the LightGBM model testing were then evaluated using the same metrics as RF, namely accuracy, precision macro, recall macro, F1-score macro, and confusion matrix. The use of the same metrics aims to ensure a fair and consistent comparison between RF and LightGBM. The results of the LightGBM evaluation are then used in the model comparison stage to determine the best model in classifying MSME performance.

3. RESULTS AND DISCUSSION

3.1 Data Preprocessing and Train-Test Split

The initial stage in the results section was carried out by reviewing the output from the data pre-processing process to ensure the dataset could be used in the modeling process. Based on the inspection results, the dataset had no missing values or duplicate data, so the data size remained at 150,000. The main changes at this stage were made by adjusting the data format to suit the needs of the classification model. Non-predictive attributes such as ID and Review_Text were not used as input features because they do not represent key indicators in the classification process. Next, the categorical feature Peak_Hour_Latency was converted to numeric form, while the Target Class was converted to numeric labels so that it could be processed by the RF and LightGBM models. A summary of the data pre-processing results is shown in Table 2.

Table 2. Data Preprocessing Results

Preprocessing Step	Result
Missing value checking	No missing values were found
Duplicate checking	No duplicate records were found
Non-predictive attribute handling	ID and Review_Text were excluded from input features
Categorical feature encoding	Peak_Hour_Latency was converted into numeric format
Target encoding	Class was converted into numeric labels
Feature-target separation	Input features and target class were separated
Final data status	Dataset is ready for train-test split

Based on Table 2, the data preprocessing stage shows that the process focused on preparing the data format for modeling, rather than reducing the amount of data. Non-predictive attributes were removed to ensure the model only used features relevant to MSME performance indicators. The encoding process is necessary because machine learning models require data in numerical form, both for categorical features and classification targets. Once the features and targets are separated, the dataset is ready to enter the training and test data division stage. After the preprocessing

stage is complete, the dataset is divided into training and test data using an 80:20 ratio using stratified sampling. This division resulted in 120,000 training data sets and 30,000 test data sets out of a total of 150,000 data sets.

The training data sets are used to build the RF and LightGBM models, while the test data sets are used to measure the model's performance on data not used during the training process. The use of stratified sampling aims to maintain a balanced proportion of each target class in the training and test data sets. The class distribution across all data, training data, and testing data is shown in Figure 2. This visualization is used to ensure that each target class remains proportionally represented after the data splitting process.

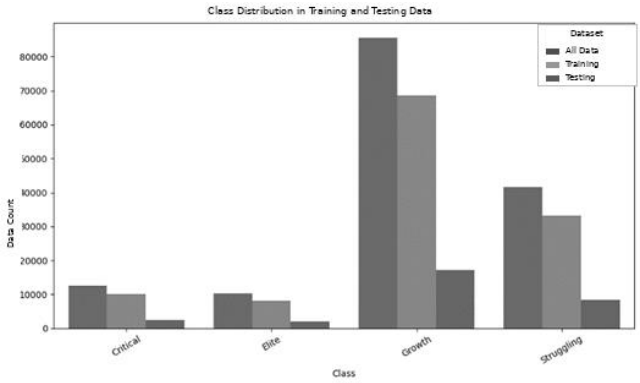

Figure 2. Class Distribution in Training and Testing Data

Based on Figure 2, the Growth class has the largest data set compared to other classes, across all data sets, training data, and testing data. The similar distribution patterns across the training and testing data indicate that the data splitting process successfully maintained the proportions of the target classes. This is important because the model needs to be trained and tested on a representative class distribution to avoid bias in favor of a particular class. To clarify the composition of the data split, a summary of the training and testing data sets is presented in Table 3.

Table 3. Train-Test Split Summary

Data Subset	Percentage	Number of Records	Function
Training data	80%	120,000	Used to train RF and LightGBM models
Testing data	20%	30,000	Used to test and evaluate trained models
Total data	100%	150,000	Complete dataset used in the experiment

Based on Table 3, the data distribution composition shows that the majority of the data is used for model training, while some data is set aside for testing. The 80% training data composition provides a sufficient sample size for the RF and LightGBM models to learn MSME performance classification patterns. Meanwhile, the 20% testing data is used to evaluate the model's performance on data that has not been used during the training process. This data composition provides a consistent basis for training and testing the RF and LightGBM models in the next stage.

3.2 Model Training of RF and LightGBM

The model training phase was conducted after the training data and testing data were clearly separated. At this stage, 120,000 training data sets were used to train the RF and LightGBM models. The RF model was trained using a bagging ensemble approach, while the LightGBM model was trained using a gradient boosting approach. Both models used the same training data to allow for consistent performance comparisons. The training output from Google Colab shows that both models were successfully trained and produced very high initial performance on the training data. To demonstrate the efficiency of the training process, a comparison of the training times of the two models is shown in Figure 3.

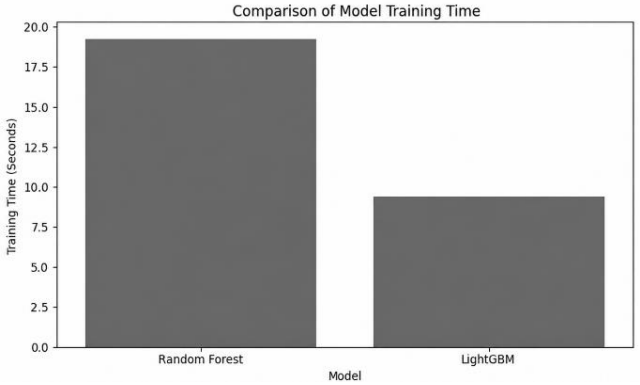

Figure 3. Model Training Time Comparison

Based on Figure 3, the RF model requires a training time of 19.3226 seconds, while the LightGBM model requires a training time of 9.3585 seconds. These results indicate that LightGBM has a faster training time than RF on the used dataset. The difference in training time is understandable because RF builds multiple decision trees in parallel, while LightGBM uses a boosting approach designed for computational efficiency on tabular data. However, training time is not the only basis for determining the best model, as final performance must still be assessed through testing results on the testing data. A summary of the configuration and training results of both models is shown in Table 4.

Table 4. Model Training Configuration and Results

Model	Algorithm	Estimator	Class Weight	Random State	Learning Rate	Training Time (s)	Train Accuracy	Train Precision Macro	Train Recall Macro	Train F1-Score Macro
RF	Bagging Ensemble	100	balanced	42	-	19.3226	1.0000	1.0000	1.0000	1.0000
LightGBM	Gradient Boosting	200	balanced	42	0.05	9.3585	1.0000	1.0000	1.0000	1.0000

Based on Table 4, the two models use different configurations according to their algorithmic characteristics. The RF model uses 100 estimators with class_weight='balanced', random_state=42, and a bagging ensemble approach. Meanwhile, the LightGBM model uses 200 estimators, learning_rate=0.05, class_weight='balanced', random_state=42, and a gradient boosting approach. The train accuracy, train precision macro, train recall macro, and train F1-score macro values for both models reached 1.0000, indicating that both models were able to learn patterns in the training data very well. However, performance on the training data cannot yet be used as a final basis for determining the best model. Therefore, both models were further tested using testing data to assess the model's generalization ability to data not used during the training process.

3.3 Testing Results of RF

The RF model testing phase was conducted using 30,000 test data sets not used in the training process. This testing aimed to determine the model's ability to classify MSME performance levels using new data. The trained RF model then generated predictions for four target classes: Critical, Struggling, Growth, and Elite. These predictions were analyzed using a confusion matrix to determine the distribution of correct and incorrect predictions within each class. The RF model confusion matrix visualization is shown in Figure 4. This image is used to demonstrate the classification error pattern in more detail before the numerical results are summarized in an evaluation table.

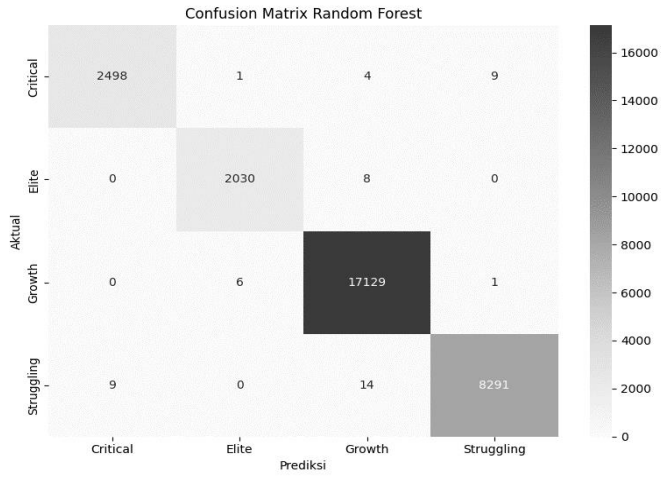

Figure 4. Confusion Matrix of RF

Based on Figure 4, the RF model shows a very high number of correct predictions for each target class. In the Critical class, the model was able to correctly classify 2,498 data points, while a small portion of the data was misclassified into other classes. In the Elite class, the model produced 2,030 correct predictions with a very small number of errors. In the Growth class, the model produced 17,129 correct predictions, making this class the class with the largest number of correct predictions due to its dominant data set. In the Struggling class, the model produced 8,291 correct predictions with several misclassifications into the Critical and Growth classes. Overall, the RF model produced 29,948 correct predictions from a total of 30,000 test data points, resulting in only 52 misclassifications. A summary of the RF model evaluation results is shown in Table 5.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 305-316
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

Table 5. Evaluation Results of RF

Model	Accuracy	Precision Macro	Recall Macro	F1-Score Macro	Correct Predictions	Misclassifications	Total Testing Data
RF	0.998267	0.997563	0.996832	0.997197	29,948	52	30,000

Based on Table 5, the RF model achieved an accuracy of 0.998267, a macro precision of 0.997563, a macro recall of 0.996832, and a macro F1-score of 0.997197. The very high accuracy value indicates that most of the testing data was correctly classified. The macro precision value indicates that the model's predictions for each class have a very good level of accuracy. The macro recall value indicates that the model is able to recognize most of the actual data in each target class. Meanwhile, the high macro F1-score value indicates that the balance between precision and recall in the RF model is at an excellent level. With these results, the RF model demonstrates strong classification capabilities and is worthy of being compared with the LightGBM model in the next testing stage.

3.4 Testing Results of LightGBM

The LightGBM model testing phase was conducted using the same testing data as the RF model, namely 30,000 data points. The use of the same testing data is necessary so that the evaluation results of both models can be compared fairly and consistently. The trained LightGBM model was then used to predict the performance class of MSMEs in four target categories: Critical, Struggling, Growth, and Elite. The prediction results were visualized using a confusion matrix to see the distribution of correct and incorrect predictions in each class. The confusion matrix visualization of the LightGBM model is shown in Figure 5. This image is used to show the classification pattern of the LightGBM model before the evaluation results are presented in the form of numerical metrics.

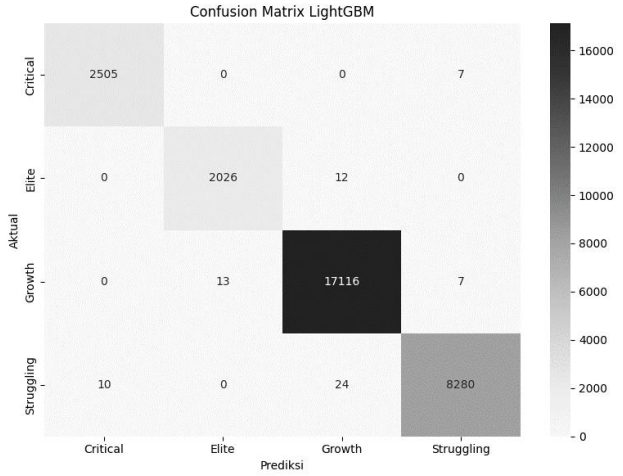

Figure 5. Confusion Matrix of LightGBM

Based on Figure 5, the LightGBM model is able to classify most of the test data into the correct class. The main diagonal values in the confusion matrix indicate the number of correct predictions for each class, while off-diagonal values indicate misclassifications. The prediction pattern of the LightGBM model shows that classes with large data sets, such as Growth and Struggling, can still be recognized well by the model. However, there are still a small number of misclassifications in some target classes, indicating the presence of some data with close characteristics between categories. Overall, the LightGBM model produced 29,927 correct predictions from a total of 30,000 test data sets, resulting in 73 misclassifications. A summary of the LightGBM model evaluation results is shown in Table 6.

Table 6. Evaluation Results of LightGBM

Model	Accuracy	Precision Macro	Recall Macro	F1-Score Macro	Correct Predictions	Misclassifications	Total Testing Data
LightGBM	0.997567	0.996465	0.996517	0.996491	29,927	73	30,000

Based on Table 6, the LightGBM model achieved an accuracy of 0.997567, a macro precision of 0.996465, a macro recall of 0.996517, and a macro F1-score of 0.996491. These values indicate that LightGBM is capable of producing very high classification performance on the testing data. An accuracy value close to 1 indicates that most of the testing data was successfully classified correctly. High macro precision and macro recall values indicate that the model has a good ability to maintain prediction accuracy and recognize actual data in each target class. Meanwhile, the macro F1-score value indicates that the balance between precision and recall in the LightGBM model is also at an excellent level. However, these results still need to be compared with the RF model to determine the model with the best overall performance.

Proceeding of International Conference on Advanced Technologies, Innovations and Science
Vol 1, No 1, June 2026, Page 305-316
ISSN XXXX-XXXX (Media Online)
https://journals.adaresearch.or.id/icatis

3.5 Model Comparison

After the test results for the RF and LightGBM models were obtained, the next step was to directly compare the performance of the two models. The comparison was conducted using predetermined evaluation metrics: accuracy, macro precision, macro recall, and macro F1-score. These four metrics were used to assess the model's ability to classify MSME performance levels across the testing data. Because the performance values of both models were within a very high range and had small differences, the comparison visualization was displayed at a larger scale to more clearly demonstrate the performance differences. The visualization of the RF and LightGBM model performance comparison is shown in Figure 6.

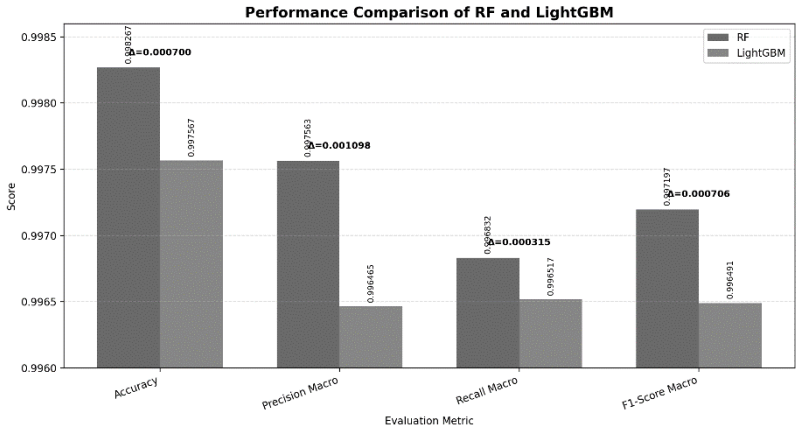

Figure 6. Performance Comparison of RF and LightGBM

Based on Figure 6, the RF model obtained higher scores than LightGBM in all evaluation metrics. In the accuracy metric, RF obtained a score of 0.998267, while LightGBM obtained a score of 0.997567, with a difference of 0.000700. In the macro precision metric, RF obtained a score of 0.997563, while LightGBM obtained a score of 0.996465, with a difference of 0.001098. In the macro recall metric, RF obtained a score of 0.996832, while LightGBM obtained a score of 0.996517, with a difference of 0.000315. In the macro F1-score metric, RF obtained a score of 0.997197, while LightGBM obtained a score of 0.996491, with a difference of 0.000706. These results indicate that RF has slightly superior performance compared to LightGBM, although the performance difference is relatively small. A summary of the evaluation metric values of both models is shown in Table 7.

Table 7. Performance Comparison of RF and LightGBM

Model	Accuracy	Precision Macro	Recall Macro	F1-Score Macro
RF	0.998267	0.997563	0.996832	0.997197
LightGBM	0.997567	0.996465	0.996517	0.996491

Based on Table 7, RF is the model with the best performance because it obtained the highest score in all evaluation metrics. The macro F1-score value is used as the main basis in determining the best model because this metric considers the balance between precision and recall in each target class. In this study, RF obtained a macro F1-score of 0.997197, while LightGBM obtained a macro F1-score of 0.996491. This difference indicates that RF is slightly more stable in maintaining a balanced classification performance in each MSME class. In addition to numerical metrics, model comparison is also strengthened by the number of correct predictions and misclassifications in the testing data, as shown in Table 8.

Table 8. Correct and Incorrect Prediction Comparison

Model	Correct Predictions	Misclassifications	Total Testing Data
RF	29.948	52	30.000
LightGBM	29.927	73	30.000

Based on Table 8, the RF model produced 29,948 correct predictions and 52 misclassifications from a total of 30,000 test data sets. Meanwhile, the LightGBM model produced 29,927 correct predictions and 73 misclassifications. Thus, RF produced 21 fewer errors than LightGBM. These results reinforce the findings in Table 7 that RF consistently outperforms, both based on evaluation metrics and the number of misclassifications. Therefore, RF was determined to be the best model in this study for classifying MSME performance levels in the dataset used.

3.6 Discussion

Based on the test results and model comparisons, RF demonstrated superior performance compared to LightGBM across all evaluation metrics used. RF's superiority is evident in its higher accuracy, macro precision, macro recall,

and macro F1-score values, as well as its lower number of misclassifications on the testing data. Although the performance difference between the two models is relatively small, the results still indicate that RF is more stable in classifying MSME performance levels on the used dataset. This stability can be attributed to RF's characteristics as a bagging-based ensemble model that builds multiple decision trees and combines the results to produce a final prediction. This finding aligns with previous studies showing that RF is capable of delivering competitive performance in tabular data-based classification tasks and MSME performance classification (Erkamim et al., 2023; Ramadhan et al., 2025). These results also support the use of ensemble models in classification because this approach can improve prediction stability compared to a single model (Bimawan et al., 2024; Firnanda et al., 2025).

In terms of the distribution of prediction results, both models are able to follow the actual class pattern very well, especially in the Growth class, which has the largest data set. However, RF produced a lower number of classification errors than LightGBM, with 52 errors compared to 73 errors out of a total of 30,000 test data sets. This difference indicates that RF is slightly better at maintaining prediction consistency across target classes, although both models achieved very high performance. These results also demonstrate that MSME performance classification can be effectively performed using a multidimensional indicator-based machine learning approach. The indicators used in this study encompass not only financial aspects but also customer, digital, operational, and business environment aspects. This approach aligns with previous research emphasizing that MSME performance is influenced by various factors, such as financial literacy, digitalization, innovation, information systems, and business environment dynamics (Alansori & Listyaningsih, 2022; Ernis, 2024; Hadi, 2022; Masriah & Ispriyahadi, 2025).

In addition to comparing numerical performance, the discussion is also strengthened by examining the dominant features of the best models. Since RF achieved the best performance based on the macro F1-score and a lower number of misclassifications, the interpretation of dominant features was performed using the feature importance of the RF model. This analysis was used to understand which indicators contributed most to differentiating MSME performance categories. Thus, the results of the study not only indicated the best predictive model but also provided an initial overview of the most influential factors in the classification process. The visualization of dominant features based on the RF model is shown in Figure 7.

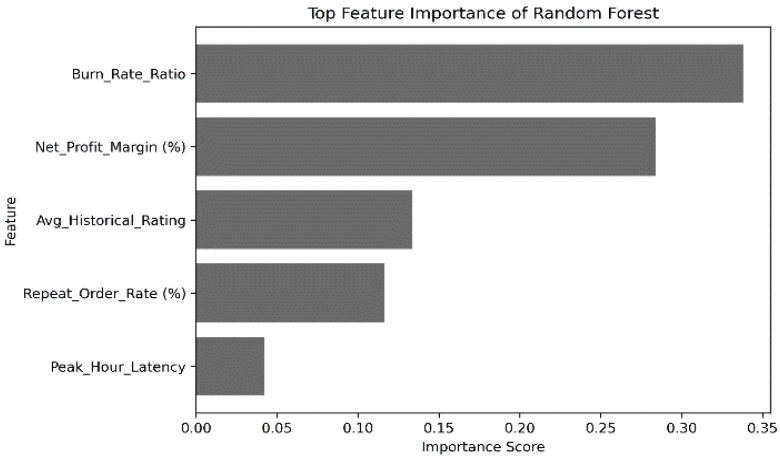

Figure 7. Dominant Features Based on RF Feature Importance

Based on Figure 7, the Burn_Rate_Ratio feature is the attribute with the largest contribution in the classification of MSME performance. This feature represents the pressure of costs or operational burdens on business conditions, so it is natural that it plays a strong role in differentiating performance categories. The next feature is Net_Profit_Margin (%), which indicates the business's ability to generate net profit from earned revenue. In addition to financial indicators, the Avg_Historical_Rating and Repeat_Order_Rate (%) features also emerged as important indicators, so that customer response and repeat purchase loyalty play a role in differentiating MSME performance levels. The Peak_Hour_Latency feature is also a dominant feature, indicating that operational aspects during peak hours can influence business performance classification. These findings strengthen the argument that MSME performance assessment is not sufficient based solely on financial indicators, but needs to consider a combination of customer and operational indicators. Compared to previous research, this study broadens the focus of the study by comparing RF and LightGBM in MSME performance classification based on multidimensional indicators. Research (Saputra & Yustanti, 2025) shows that the success of small and medium enterprises can be predicted using multidimensional indicators, but has not specifically compared RF and LightGBM in the context of MSME performance classification. Research (Erkamim et al., 2023) has compared RF and XGBoost in MSME performance classification, but has not included LightGBM as a gradient boosting-based comparison. This study fills this gap by comparing RF and LightGBM using the same multiclass evaluation metrics, namely accuracy, macro precision, macro recall, macro F1-score, and confusion matrix. In addition, this study also adds dominant feature interpretation to explain indicators that contribute to MSME performance classification. Thus, the contribution of this study lies in the

empirical comparison of two decision tree-based ensemble models and the presentation of feature interpretation to support understanding of the classification results.

Although the research results demonstrate very high model performance, there are several limitations that need to be considered. The dataset used is a synthetic dataset from Kaggle, so the resulting data patterns may not fully represent the complexity of real-world MSME conditions. Therefore, the results of this study need further validation using empirical MSME data so that the model can be tested in a wider variety of business conditions. Furthermore, the feature interpretation analysis in this study still uses the built-in feature importance of the RF model, so further research can use a more in-depth interpretability approach such as SHAP to examine feature contributions at the global and local levels. Additional testing can also be conducted by comparing other models such as XGBoost, SVM, or ANN to obtain a broader performance picture. With these developments, the MSME performance classification model can be directed into a more robust, transparent, and relevant analytical tool to support data-driven decision-making.

4. CONCLUSION

This study compares the RF and LightGBM models for classifying MSME performance levels based on financial, customer, digital, operational, and business environment indicators. The dataset used is a secondary synthetic dataset from Kaggle with 150,000 data points and 15 attributes. Data preprocessing results show that the dataset has no missing values and duplicate data, so the data remains the same after the initial inspection process. The transformation process is carried out by removing non-predictive attributes, encoding categorical features, encoding targets, and separating features and targets before data division. The dataset is then divided using a ratio of 80:20 with stratified sampling, resulting in 120,000 training data points and 30,000 testing data points. The training results show that both models are able to learn patterns in the training data very well. However, testing results on the testing data show that RF obtained the best performance compared to LightGBM in all evaluation metrics. The RF model produces an accuracy of 0.998267, a macro precision of 0.997563, a macro recall of 0.996832, and a macro F1-score of 0.997197. Meanwhile, LightGBM obtains an accuracy of 0.997567, a macro precision of 0.996465, a macro recall of 0.996517, and a macro F1-score of 0.996491. Based on the confusion matrix results, RF also produces a lower number of classification errors, namely 52 errors out of 30,000 testing data, while LightGBM produces 73 errors. Based on the highest macro F1-score value and a lower number of classification errors, RF is determined as the best model in this study. The feature importance results from the RF model show that Burn_Rate_Ratio, Net_Profit_Margin (%), Avg_Historical_Rating, Repeat_Order_Rate (%), and Peak_Hour_Latency are the dominant features in differentiating MSME performance categories. This finding indicates that MSME performance classification is not only influenced by financial aspects, but also by customer and operational indicators. Thus, the RF model has the potential to be used as an approach to MSME performance classification based on multidimensional indicators. Although the results demonstrate very high performance, this study still has limitations due to the synthetic nature of the dataset used. Therefore, further research is recommended using empirical data from MSMEs so that the model can be validated in more realistic and diverse business conditions. Furthermore, future research can add interpretability approaches such as SHAP to explain feature contributions in more depth, both at the global and local levels. Testing with other models such as XGBoost, SVM, or ANN can also be conducted to obtain broader performance comparisons. With these developments, machine learning-based MSME performance classification can be directed into a more transparent, measurable, and relevant analytical tool to support data-driven business decision-making.

ACKNOWLEDGMENT

The authors would like to express their gratitude to all parties who have supported the completion of this research. They also acknowledge the availability of the open dataset from Kaggle, which served as a source of experimental data in this study. Furthermore, they appreciate the support provided by the computing facilities and Google Colab-based software used in data processing, model training, evaluation, and visualization of the research results.

REFERENCES

Adawiyah, R., & Kartikasari, D. C. J. (2025). Comparison Of Lung Cancer Classification Using Decision Tree And Random Forest. In *Komputika : Jurnal Sistem Komputer* (Vol. 14, Issue 1, pp. 79–85). Universitas Komputer Indonesia. https://doi.org/10.34010/komputika.v14i1.15501

Alansori, A., & Listyaningsih, E. (2022). Pengaruh Kinerja Umkm Terhadap Kesejahteraan Umkm Di Bandar Lampung. In *Adbispreneur* (Vol. 7, Issue 1, P. 39). Universitas Padjadjaran. Https://Doi.Org/10.24198/Adbispreneur.V7i1.37930

Alfianti, M. L. T., & Supriyanto, R. (2024). Perbandingan Kinerja Algoritma Random Forest, Adaboost, Dan Xgboost Dalam Memprediksi Resiko Penyakit Osteoporosis. In *Jurnal Ilmu Komputer Dan Agri-Informatika* (Vol. 11, Issue 2, Pp. 172–184). Institut Pertanian Bogor. Https://Doi.Org/10.29244/Jika.11.2.172-184

Anggraini, N., Putra, S. J., Wardhani, L. K., Dhiya, F., & Arif, U. (2024). *A Comparative Analysis Of Random Forest , Xgboost , And Lightgbm Algorithms For Emotion Classification In Reddit Comments*. *17*(1), 88–97.

Bimawan, Z. I., Astuti, T., & Arsi, P. (2024). *Comparison Of Random Forest , K-Nearest Neighbor , Decision Tree , And Xgboost Algorithms For Detecting Stunting In Toddlers Komparasi Algoritma Random Forest , K-Nearest Neighbor ,*. *5*(6), 1599–1607.

Djaka, T. P. D., & Winarsih, N. A. S. (2025). Analisis Kinerja Ensemble Learning Dan Algoritma Tunggal Dalam Klasifikasi Sindrom Ovarium Polikistik Menggunakan Random Forest, Logistic Regression, Dan Xgboost. In *Infotekmesin* (Vol. 16, Issue 1, Pp. 43–51). Politeknik Negeri Cilacap. Https://Doi.Org/10.35970/Infotekmesin.V16i1.2504

Erkamim, M., Zidni, M., & Widarti, E. (2023). *Komparasi Algoritme Random Forest Dan Xgboosting Dalam Klasifikasi Performa UMKM*. *02*, 127–134.

Ernis, P. D. (2024). Pengaruh Literasi Keuangan, Inklusi Keuangan Dan Peran Inovasi Terhadap Kinerja UMKM Di Kota Pekanbaru. In *Jurnal Ilmiah Raflesia Akuntansi* (Vol. 10, Issue 2, Pp. 31–40). Politeknik Raflesia. Https://Doi.Org/10.53494/Jira.V10i2.287

Firnanda, P. A., Shofwatillah, L., Rahma, F., & Fauzi, F. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket. In *Emerging Statistics and Data Science Journal* (Vol. 3, Issue 1, pp. 445–461). Universitas Islam Indonesia (Islamic University of Indonesia). https://doi.org/10.20885/esds.vol3.iss.1.art2

Hadi, H. K. (2022). Pengaruh Dinamisme Lingkungan Dan Ketidakpastian Lingkungan Terhadap Kinerja Umkm Di Kota Mojokerto. In *Jurnal Ilmu Manajemen* (pp. 902–910). Universitas Negeri Surabaya. https://doi.org/10.26740/jim.v10n3.P902-910

Isabella, A. A., & Sanjaya, P. N. (2022). Efektivitas Pendampingan Konsultan Pendamping Umkm Terhadap Kinerja Umkm: Studi Kasus Pada Umkm Kabupaten Mesuji. In *Derivatif : Jurnal Manajemen* (Vol. 16, Issue 2, pp. 279–285). Muhammadiyah Metro University. https://doi.org/10.24127/jm.v16i2.1072

Iwung, H., & Rahman, B. (2026). Perbandingan Kinerja Algoritma Random Forest dan Convolutional Neural Network (CNN) Untuk Klasifikasi Citra Kucing. In *Jurnal Informatika: Jurnal Pengembangan IT* (Vol. 11, Issue 1, pp. 53–62). Politeknik Harapan Bersama Tegal. https://doi.org/10.30591/jpit.v11i1.10156

Jannah, Z. T. (2025). Pengaruh financial literacy, financial technology, dan entrepreneurial orientation terhadap kinerja UMKM. In *Jurnal Ilmu Manajemen* (pp. 447–459). Universitas Negeri Surabaya. https://doi.org/10.26740/jim.v13n2.p447-459

Masriah, N., & Ispriyahadi, H. (2025). Pengaruh Literasi Keuangan Dan Digital Payment. In *Jurnal Ekobis : Ekonomi Bisnis & Manajemen* (Vol. 15, Issue 1, pp. 54–76). Universitas Teknologi Muhammadiyah Jakarta (UTM Jakarta). https://doi.org/10.37932/kwahcf80

Nuraini, R., & Darmawan, D. (2024). Keberlanjutan UMKM: Dampak Orientasi Kewirausahaan untuk Peningkatan Kinerja Bisnis. In *Jurnal Ilmu Ekonomi, Manajemen dan Bisnis* (Vol. 2, Issue 2, pp. 85–91). Universitas Aisyiyah Surakarta. https://doi.org/10.30787/jiembi.v2i2.1574

Oikonomou, K., & Damigos, D. (2024). Short term forecasting of base metals prices using a LightGBM and a LightGBM - ARIMA ensemble. In *Mineral Economics* (Vol. 38, Issue 1, pp. 37–49). Springer Science and Business Media LLC. https://doi.org/10.1007/s13563-024-00437-y

Oueslati, R., Ouertani, M. W., Amdouni, A., & Manita, G. (2025). MS-FSOA-LightGBM: Multi-Strategy Starfish Optimization Algorithm with LightGBM for Software Defect Prediction. In *Procedia Computer Science* (Vol. 270, pp. 4054–4063). Elsevier BV. https://doi.org/10.1016/j.procs.2025.09.530

Prajesh, A., & Jain, P. (2025). HBA-LightGBM: Honey Badger Algorithm With LightGBM Model For Solar Irradiance Forecasting. In *IEEE Transactions on Industry Applications* (Vol. 61, Issue 3, pp. 5081–5090). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/tia.2025.3542730

Pratama, F., Ali, E., Rahmaddeni, & Agustin, W. (2025). Perbandingan Kinerja Xgboost Dan Lightgbm Dalam Klasifikasi Depresi Pada Mahasiswa Berdasarkan Faktor Demografi Dan Akademik. In *Jurnal Algoritma* (Vol. 22, Issue 2, pp. 53–64). Institut Teknologi Garut. https://doi.org/10.33364/algoritma/v.22-2.2439

Rainio, O. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 1–14. https://doi.org/10.1038/s41598-024-56706-x

Ramadhan, S., Alamin, Z., Jannah, M., Akbar, M., & Fikriyansah, R. (2025). *Data-driven MSME Success Prediction Using Decision Tree-Based Machine Learning Techniques Prediksi Keberhasilan UMKM Berbasis Data Menggunakan Teknik Machine Learning Berbasis Pohon Keputusan*. *1*(1), 1–9.

Saputra, A. D., & Yustanti, W. (2025). *A Hybrid Clustering – Classification Framework for SMEs Success Level Prediction*. *9*(2), 89–100.

Shinzi. (2025). *Perbandingan Kinerja Algoritma Knn , Decision Tree , Svm , Dan Ann Untuk Melakukan Klasifikasi Pada*. *20*(1), 24–32.

Ulyasari, O. R., Agustina, D., Wardhani, R. S., & Ilhamsyah, A. W. (2023). Pengaruh E-Commerce Dan Sistem Informasi Akuntansi Terhadap Kinerja Umkm Terhadap Kinerja Umkm Sektor Industri. In *Jurnal Ilmiah Global Education* (Vol. 4, Issue 2, pp. 799–808). LPPM Institut Pendidikan Nusantara Global. https://doi.org/10.55681/jige.v4i2.642

V, M. K. M., Almuraqab, N., Moonesar, I. A., Braendle, U. C., & Rao, A. (n.d.). *How critical is SME financial literacy and digital financial access for financial and economic development in the expanded BRICS block ?*

Zeng, G. (2025). *Invariance Properties and Evaluation Metrics Derived from the Confusion Matrix in Multiclass Classification*. *1997*.

Zhao, X., Liu, Y., & Zhao, Q. (2024). SMOTE-CHL-LightGBM: An Enhanced LightGBM for Credit Card Fraud Detection. In *2024 International Conference on Machine Learning and Cybernetics (ICMLC)* (pp. 599–606). IEEE. https://doi.org/10.1109/icmlc63072.2024.10935079

Zkyfauzi. (n.d.). *UMKM Dataset*. Kaggle. https://www.kaggle.com/datasets/zkyfauzi/umkm-dataset