Vol. 11 No.2 Tahun 2025, pp 143-152
Jurnal Teknologi dan Manajemen Informatika
http:// http://jurnal.unmer.ac.id/index.php/jtmi
P-ISSN: 1693-6604   E-ISSN: 2580-8044

Optimized LightGBM Model for Predicting Total Cup Points of Arabica Coffee using Sensory Cupping Data

Arya Rezagama Sudrajat[1], Ricardus Anggi Pramunendar[2], Mohammad Syaifur Rohman[3]

Universitas Dian Nuswantoro, Indonesia

Article Info

Article History
Received : 30-10-2025
Revised : 18-11-2025
Accepted : 26-11-2025

Keywords
Arabica Coffee;
Sensory Cupping;
LightGBM;
Coffee Quality Prediction;
Machine Learning

Corresponding Author
Arya Rezagama Sudrajat,
Universitas Dian Nuswantoro
111202214706@mhs.dinus.ac.id

ABSTRACT

Evaluating coffee quality through sensory cupping is essential but inherently subjective, as scoring depends on the consistency and expertise of professional panelists. To improve objectivity, this study applies the Light Gradient Boosting Machine (LightGBM) algorithm to predict the Total Cup Points of Arabica coffee using sensory evaluation data. The dataset, obtained from the Coffee Quality Institute Arabica Reviews (May 2023), contains 1,509 cupping records assessed according to the Specialty Coffee Association (SCA) protocol. Nine sensory attributes aroma, flavor, aftertaste, acidity, body, balance, uniformity, clean cup, and sweetness were used as predictors. The modeling process included data preprocessing, feature selection, hyperparameter tuning using RandomizedSearchCV, and performance evaluation through 5-Fold and 10 Fold Cross-Validation. The tuned LightGBM model achieved an R^2 of 0.9634 and an RMSE of 0.4673 under the 10-Fold scheme. Comparative analysis showed that LightGBM produced lower prediction error than XGBoost, Random Forest, and Support Vector Regression (SVR) when evaluated under identical default parameter settings. Feature importance indicated that flavor, balance, clean cup, and aftertaste were the most influential contributors to total cup points. The findings provide a reliable computational framework to support more objective, consistent, and efficient coffee cupping assessments.

INTRODUCTION

Coffee is one of Indonesia’s most strategic plantation commodities, with high economic value and a significant role in international trade. According to data from Statistics Indonesia (BPS) [1], the volume and value of Indonesian coffee exports continue to show an increasing trend despite annual fluctuations. This indicates that global demand for Indonesian coffee, particularly Arabica, remains strong, and therefore maintaining product quality is crucial for sustaining export competitiveness. Arabica coffee has a higher market value and a more complex sensory profile than Robusta [2], [3], making quality consistency essential. Several factors influence Arabica quality, including altitude, which correlates with flavor complexity [4], as well as physicochemical characteristics that differentiate Arabica, Robusta, and Liberica beans [5], [6].

Coffee quality assessment is commonly conducted through the cupping test standardized by the Specialty Coffee Association (SCA) [7]. This procedure evaluates attributes such as aroma, flavor, acidity, body, aftertaste, and overall impression [8], [9]. However, because cupping relies on human perception, the results are vulnerable to subjectivity and interevaluator variability [10]. Consequently, a computational model capable of learning patterns from existing cupping data and generating more consistent, objective, and reproducible Total Cup Points estimations is needed. Alternative analytical techniques such as GC/MS-based metabolomics [11] and UV/Vis/NIR spectroscopy [12] have been explored, but these approaches require specialized equipment and are less practical for routine evaluations.

Previous studies have applied various machine learning approaches in the coffee and food domains; however, most of them have focused on classification tasks or defect detection rather than quantitative prediction of sensory scores such as Total Cup Points (TCP). Beyond coffee-focused studies, related research in Indonesia has also demonstrated advancements in predictive analysis and image classification using machine learning. For example, a study on coffee bean defect classification using transfer learning showed that deep learning techniques can effectively extract visual features and improve recognition performance [13]. Another study developed a classification model for herbal leaf images using Convolutional Neural Networks (CNN) combined with K-Nearest Neighbor, demonstrating that CNN-based feature extraction enhances predictive accuracy in multicategory datasets [14]. From a methodological perspective, cross-validation has also been applied to strengthen model reliability, as shown in a study on prostate cancer classification using MRI images with a hierarchical validation scheme [15]. However, despite these advancements, research focusing specifically on numerical TCP prediction and systematic optimization of LightGBM for quantitative sensory modeling remains limited.

Among various machine learning algorithms, LightGBM is known for its computational efficiency and strong predictive performance on structured data [16]. Its accuracy is highly influenced by hyperparameter configurations such as num leaves, max depth, learning_rate, and regularization parameters, which collectively control the model's generalization capability [17], [18]. With these advantages, LightGBM presents strong potential for predicting Total Cup Points based on standardized cupping data. This study aims to develop a predictive model for Total Cup Points using the LightGBM algorithm based on nine primary sensory attributes: aroma, flavor, aftertaste, acidity, body, balance, uniformity, clean cup, and sweetness.

The study further compares LightGBM with three regression algorithms, XGBoost, Random Forest, and Support Vector Regression (SVR) and applies hyperparameter tuning to obtain the optimal configuration. The findings are expected to support the development of a more objective, consistent, and efficient digital system for evaluating Arabica coffee quality. This paper is organized into four main sections. Section 1 presents the research background. Section 2 describes the research methodology, including data collection, preprocessing, and model development. Section 3 discusses the experimental results and analysis. Finally, Section 4 concludes the study and provides recommendations for future research.

METHODS

Research Flow

This research began with the collection of data from the Coffee Quality Institute (CQI) Arabica Reviews dataset published in May 2023 [19] as the primary data source. The collected data underwent several preprocessing steps, including the removal of outliers, selection of relevant sensory attributes, and data formatting adjustments. LightGBM model was then developed using default parameters to establish baseline performance, followed by hyperparameter tuning to obtain the optimal configuration. Model performance was evaluated using 5-fold and 10Fold Cross-Validation schemes with R^2 and RMSE metrics to assess prediction accuracy and stability in estimating the total cup points of Arabica coffee. Figure 1 shows the overall illustration of the workflow.

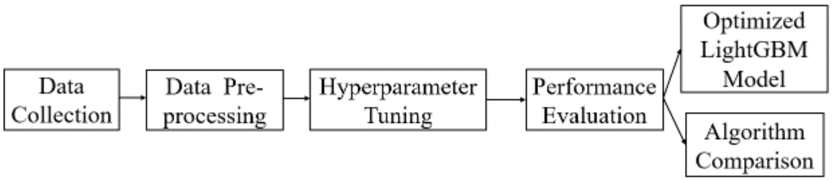

Figure 1. Research Design Flowchart

Dataset Description

The dataset used in this study was obtained from the CQI Arabica Reviews published in May 2023 [19], which is publicly available on the Kaggle platform. It consists of 1,509 Arabica coffee records evaluated by professional panelists following the SCA cupping protocol. Each record contains nine primary sensory attributes used in quality assessment, namely Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, and Sweetness. In addition to sensory attributes, the dataset also includes supplementary information such as the country of origin, producer name, altitude, coffee variety, and processing method. This diversity makes the dataset representative of Arabica coffee quality variations across major producing regions worldwide. In this research, the nine sensory attributes were used as predictor variables (input features), while the TCP served as the target variable predicted by the LightGBM model.

Data Preprocessing

The preprocessing stage was conducted to ensure the quality and consistency of the data before model development. The Coffee Quality Institute Arabica Reviews May 2023 dataset [19] was imported from Kaggle into the Python environment. An initial examination was performed to identify missing values and duplicate records. No missing values were found in the nine sensory attributes or the Total Cup Points variable, and a full-row duplication check confirmed that the dataset contained no duplicate entries; therefore, no imputation or duplicate removal procedures were required.

Records that did not comply with the SCA standard, particularly Total Cup Points values outside the 70–100 range, were identified as outliers and removed to maintain a representative distribution. Feature selection was then carried out using a domain knowledge–based approach following the evaluation guidelines of the CQI and SCA. Nine sensory attributes Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, and Sweetness were used as predictor variables, while total cup points served as the target variable.

Since all sensory attributes used as predictor variables share the same scoring scale (0–10) according to the SCA cupping protocol, no normalization or standardization was applied. The uniform scale ensures that no feature disproportionately influences the learning process, allowing the LightGBM model to process all variables without requiring additional feature scaling. The cleaned dataset was then utilized for model training and evaluation using the k-Fold Cross-Validation technique with two configurations, 5-Fold and 10-Fold. To maintain result consistency and ensure experimental reproducibility, a fixed random_state value of 42 was applied throughout all data splitting, training, and validation procedures.

Model and Evaluation

LightGBM Model

The LightGBM is a gradient boosting decision tree algorithm designed for efficient predictive modeling on large and imbalanced datasets [20]. In this study, the model was first trained using default parameters, followed by hyperparameter tuning to obtain the optimal configuration.

1. Number of Terminal Leaves defines the maximum number of leaf nodes in each decision tree [21]. The general range for this parameter is 10–256 [22], while this study used a range of 10–300 to explore a broader level of model complexity.
2. Maximum Tree Depth limits the hierarchical depth of each tree to control overfitting and underfitting [21]. The typical range for this parameter is 2–15 [22], whereas this

study employed a range of −1 to 20 to allow greater flexibility in determining the optimal tree depth

3. Learning Rate determines the contribution of each new tree to the overall model update [21]. The common range for this parameter is 0.01–0.3 [22] and in this study a range of 0.01–0.2 was used to balance training speed and learning stability.
4. Number of Boosting Iterations represents the total number of trees or boosting rounds during model training [21]. The general range is 50–1000 [22], and in this study the range of 1–2000 was used to examine the effect of iteration count on prediction accuracy.
5. Sampling Ratio per Iteration controls the proportion of training data used in each boosting round to enhance generalization and reduce overfitting [21]. The recommended range is 0.5–1.0 [22], and this study used 0.1–1.0 to analyze the effect of sample size variation on prediction results.
6. Feature Sampling Ratio per Tree specifies the proportion of features utilized when building each tree [21]. The general range is 0.5–1.0 [22], while this study used 0.1–1.0 to observe how feature sampling affects model performance.
7. L1 Regularization Strength adds a penalty based on the absolute value of leaf weights to improve sparsity and model simplicity [21]. The common range is 0–5 [22], and this study 0–10 was used to assess its influence on model stability.
8. L2 Regularization Strength introduces a penalty on squared weights to balance bias and variance [21]. The general range is 0–10 [22], and this study 0–20 was used to evaluate its impact on the overall model performance. Model Evaluation and Validation

Model Evaluation and Validation

The model performance was evaluated using two primary metrics, namely the coefficient of determination (R^2) and the Root mean Square Error (RMSE). The R^2 value measures how much variance in the data can be explained by the model, while RMSE indicates the average magnitude of prediction errors. The mathematical formulations of R^2 and RMSE used in this study are presented in Equation (1) and Equation (2), respectively. To obtain stable and unbiased results, the k-Fold Cross-Validation technique was implemented with two configurations, 5-Fold and 10-Fold. This technique allows all data to be alternately used as training and testing sets, providing a more representative estimation of the model's generalization ability.

The use of both 5-Fold and 10-Fold configurations was chosen because they can balance bias and variance during the evaluation process [23]. The mean and standard deviation values of R^2 and RMSE from each validation scheme were used to examine model consistency and to identify the most reliable configuration for subsequent analysis.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} \left(y_i - \underline{y}\right)^2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Research Design

The research began with the collection of data from the CQI Arabica Reviews published in May 2023, which served as the primary data source. The obtained dataset then underwent a series of preprocessing procedures to ensure data quality and consistency. These procedures included identifying and removing outliers based on the SCA standard, selecting relevant sensory attributes used in professional cupping evaluations, and adjusting data formats to ensure compatibility with the model training pipeline.

In this study, model evaluation did not rely on a fixed train–test split. Instead, the performance and generalization capability of the model were assessed using the k-Fold Cross-Validation technique. This approach divides the dataset into *k* equally sized subsets, where one fold is used as the testing set while the remaining $k-1$ folds serve as the training set. This iterative process was applied using both 5 fold and 10 fold configurations to obtain a more comprehensive assessment of the model's stability and predictive reliability under different data partitioning settings.

The baseline LightGBM model was initially constructed using default parameters to establish reference performance. After the baseline evaluation, a systematic hyperparameter tuning process was conducted to determine the most optimal configuration for the LightGBM model. The tuning procedure employed the RandomizedSearchCV method, which explores random combinations of parameters from predefined search spaces. A total of 150 candidate parameter configurations were evaluated using 10-Fold Cross-Validation, allowing the tuning process to capture potential interactions among hyperparameters that may influence overall model performance.

The optimized LightGBM model was subsequently evaluated using both 5 Fold and 10 Fold Cross-Validation schemes. Two primary evaluation metrics were employed: the coefficient of determination (R^2) to quantify the proportion of variance explained by the model, and the RMSE to measure the average magnitude of prediction errors. The use of these two metrics provided a comprehensive understanding of the model's predictive accuracy and stability across different validation scenarios. To further assess the effectiveness of the optimized LightGBM model, a comparative analysis was performed against three commonly used regression algorithms, namely XGBoost, Random Forest, and SVR. This comparison aimed to determine the relative performance of LightGBM in predicting Total Cup Points based on the nine sensory attributes evaluated by professional cuppers.

All implementation procedures were conducted using Python 3.12.3 within the Visual Studio Code (VS Code) environment, supported by essential libraries such as LightGBM 4.1.0, scikit-learn 1.4.2, NumPy 1.26.4, Pandas 2.2.1, Seaborn 0.13.2, and Matplotlib 3.8.4. The experiments were executed on an ASUS TUF Gaming F15 (FX506HCB) laptop equipped with an Intel Core i7-11600H processor (12 CPUs, 2.9 GHz), 16 GB RAM, and the windows 11 Home 64-bit operating system. This hardware configuration provided sufficient computational resources to ensure efficient data processing, model training, hyperparameter tuning, and evaluation.

RESULTS AND DISCUSSION

Hyperparameter Tuning

The hyperparameter tuning process was carried out using the RandomizedSearchCV method, which evaluates randomly selected combinations of parameter values from predefined search ranges. A total of 150 candidate configurations were tested using 10 Fold Cross-Validation. This approach enabled the evaluation of diverse combinations of tree structure, learning dynamics, sampling proportions, and regularization strength, while also capturing interactions among parameters that may influence predictive performance.

The search space included eight key parameters that govern the behavior of the LightGBM model: number of terminal leaves, maximum tree depth, learning rate, number of boosting iterations, sampling ratio per iteration, feature sampling ratio per tree, l1 regularization strength, and l2 regularization strength. These ranges were defined based on theoretical recommendations and prior work and were aligned with the parameter descriptions provided in the Methods section. Evaluation results indicated that model performance varied across different combinations, with some parameter ranges showing broad stability and others displaying more pronounced sensitivity. Because each configuration was evaluated using 10-

Fold Cross-Validation, the resulting performance metrics reflected consistent behavior across different data partitions.

From the 150 evaluated configurations, RandomizedSearchCV identified the parameter combination that achieved the highest mean R^2 score. This configuration was subsequently re-evaluated using 10 Fold Cross-Validation to ensure consistent performance and confirm that the selected parameters generalized well across all folds. Table 1 shows the summarized of optimal configuration obtained from the tuning process.

Table 1. Optimal LightGBM Hyperparameter Configuration

Parameter	Optimal Value
Number of Terminal Leaves	40
Maximum Tree Depth	20
Learning Rate	0.03
Number of Boosting Iterations	2000
Sampling Ratio per Iteration	0.9
Feature Sampling Ratio per Tree	0.2
L1 Regularization Strength	0.5
L2 Regularization Strength	0

Performance Comparison Between Baseline and Optimized LightGBM Models

The comparison between the baseline LightGBM model and the optimized version was carried out to examine changes in predictive performance after hyperparameter tuning. The evaluation employed a 10-Fold Cross-Validation scheme, where each subset took turns as the testing fold while the remaining folds served as the training data. Two metrics were used: mean R^2 to indicate the proportion of variance explained by the model and mean RMSE to measure the average prediction deviation.

Under the baseline configuration, LightGBM produced a mean R^2 of 0.9547 and a mean RMSE of 0.5153. After the optimization procedure using RandomizedSearchCV, the tuned model achieved a mean R^2 of 0.9634 and a mean RMSE of 0.4673. These numerical differences describe the changes in model behavior following adjustments to tree structure, learning rate, sampling ratios, and regularization strengths. Table 2 shows the summarized performance comparison between the baseline and optimized LightGBM models.

Table 2. Performance Comparison Between Baseline and Optimized LightGBM

Model	R^2 mean	RMSE mean
LightGBM Default	0.9547	0.5153
LightGBM Tuned	0.9634	0.4673

Comparison Between 5-Fold and 10-Fold Validation

The comparison between the 5-Fold and 10-Fold Cross-Validation schemes was performed to describe variations in LightGBM performance across different partitioning strategies. Both schemes were used to calculate mean R^2 and mean RMSE values for the baseline and tuned configurations. In the 5 fold setting, the baseline model produced a mean R^2 of 0.9471 and a mean RMSE of 0.5571. After tuning, the mean R^2 increased to 0.9555, and the mean RMSE was 0.5138. In the 10 fold setting, the baseline configuration resulted in a mean R^2 of 0.9547 and a mean RMSE of 0.5153. With the optimized parameters, the model obtained a mean R^2 of 0.9634 and a mean RMSE of 0.4673. Table 3 shows the presented of comparison of performance values across both validation schemes.

Table 3. Comparison of LightGBM model performance 5-Fold and 10-Fold Cross-Validation

Model	R^2 mean	RMSE mean
LightGBM Default (5-Fold)	0.9471	0.5571
LightGBM Tuned (5-Fold)	0.9555	0.5138
LightGBM Default (10-Fold)	0.9547	0.5153
LightGBM Tuned (10-Fold)	0.9634	0.4673

Comparison with Other Regression Algorithms

The comparison with other regression algorithms was conducted to provide an overview of how the tuned LightGBM model performs relative to several commonly used regression methods in sensory-based prediction tasks. The evaluation was performed using a 10-Fold Cross-Validation scheme, employing two metrics: the mean R^2 value to measure the proportion of explained variance and the mean RMSE value to quantify the average prediction deviation.

Four algorithms were included in this comparison: the tuned LightGBM model, XGBoost, Random Forest, and SVR). To maintain methodological consistency and keep the primary focus on LightGBM, only LightGBM was optimized using RandomizedSearchCV. The remaining algorithms were evaluated using their default configurations and serve as baseline references rather than fully optimized competitors. This approach aligns with the objective of the study, which is to analyze the behavior and effectiveness of LightGBM after hyperparameter tuning.

The tuned LightGBM model achieved a mean R^2 of 0.9634 and a mean RMSE of 0.4673. XGBoost produced a mean R^2 of 0.9492 and a mean RMSE of 0.5355, while Random Forest obtained a mean R^2 of 0.9333 with an RMSE of 0.6112. SVR achieved the lowest values with a mean R^2 of 0.8798 and a mean RMSE of 0.8212. These numerical differences illustrate variations in predictive behavior across different regression methods evaluated under identical validation settings. Table 4 shows a summary of the comparison results.

Table 4. Comparison of Regression Algorithms

Model	R2_Mean	RMSE_Mean
LightGBM (Tuned)	0.9634	0.4673
XGBoost	0.9492	0.5355
RandomForest	0.9333	0.6112
SVR	0.8798	0.8212

Comparison with Previous Studies

Previous research has applied LightGBM for coffee quality prediction using sensory evaluation data [24]. The reported results demonstrated acceptable predictive performance; however, the study did not incorporate extensive hyperparameter optimization and did not employ a rigorous k-fold cross-validation procedure during model evaluation.

In this work, LightGBM is developed using a broader hyperparameter search space and evaluated through RandomizedSearchCV across 150 candidate configurations. A 10-fold cross-validation scheme is applied to ensure reliable performance estimation, yielding a mean R^2 of 0.9634 and a mean RMSE of 0.4673. These results illustrate that systematic hyperparameter tuning combined with k-fold validation can enhance the robustness and predictive accuracy of LightGBM for estimating Total Cup Points.

Discussion

Cross-validation results demonstrated that the 10 fold configuration produced slightly higher mean R^2 values and lower RMSE compared to the 5 fold scheme. This pattern reflects how increasing the number of partitions can provide a more stable estimation of generalization performance for the cupping dataset due to more homogeneous sample representation across folds. The improvement observed after hyperparameter tuning in both validation settings further indicates the sensitivity of the model to structural and regularization adjustments.

The evaluation of different regression algorithms under uniform experimental settings also showed distinct behavioral differences. XGBoost produced a mean R^2 of 0.9492 with an RMSE of 0.5355, while Random Forest and SVR yielded lower values on both metrics. LightGBM, after tuning, achieved the highest performance with a mean R^2 of 0.9634 and RMSE of 0.4673. These values illustrate how each algorithm responds differently to sensory-based input spaces. Since only LightGBM was tuned, the comparison functions as a reference point rather than a claim of superiority, and the close proximity between LightGBM and XGBoost reflects the capability of both boosting algorithms to model nonlinear sensory–quality relationships effectively.

Feature importance analysis identified flavor, balance, clean cup, and aftertaste as the most influential predictors of Total Cup Points. These attributes align closely with the core sensory components emphasized in the SCA evaluation protocol, indicating that the model captures patterns relevant to professional cupping practices. The prominence of flavor-related descriptors is also consistent with findings from previous sensory modeling studies, where flavor complexity and harmony frequently contribute substantially to overall quality scoring.

Beyond methodological evaluation, several practical implications arise for Indonesian coffee stakeholders. The tuned LightGBM model—demonstrating stable performance under 10-Fold validation—can serve as a tool for early-stage screening of large sample batches during post-harvest processing. Such preliminary estimations may help cooperatives, exporters, and local processors identify promising lots more efficiently prior to formal sensory evaluation. This can be particularly beneficial in regions with limited access to certified cuppers or during peak harvest periods when sample volumes are high.

The concentration of importance in attributes such as Flavor, Balance, and Clean Cup further provides actionable insights for producers. These attributes are often influenced by post-harvest practices, including fermentation control, drying consistency, and defect management. Understanding which sensory variables carry the highest predictive weight may guide producers in prioritizing quality improvement interventions that most directly affect cupping outcomes.

Lastly, the computational efficiency of LightGBM supports its integration into lightweight digital assessment tools, enabling rapid, data-driven estimations that complement human cuppers rather than replacing them. For Indonesia's growing specialty coffee sector—where consistency, traceability, and efficiency are increasingly important—predictive modeling can serve as an additional decision-support layer to enhance competitiveness in domestic and international markets.

CONCLUSIONS AND RECOMMENDATIONS

This study demonstrates that LightGBM can provide an objective prediction of Arabica coffee Total Cup Points using standardized cupping data. After hyperparameter tuning, the model achieved stable performance under the 10-fold validation scheme, with an R^2 of 0.9634 and an RMSE of 0.4673. In the algorithm comparison, the optimized LightGBM model outperformed XGBoost evaluated under default settings and also produced higher accuracy than Random Forest and SVR. Feature importance analysis identified flavor, balance, clean cup, and aftertaste as the most influential sensory attributes. Future research is recommended to incorporate chemical, environmental, or image-based features to improve model generalization, and to validate the approach in real cupping environments across different Arabica varieties and production regions so the model can be more broadly applied within the coffee industry.

REFERENCES

[1] Badan Pusat Statistik, “Ekspor Kopi Menurut Negara Tujuan Utama Tahun 2000-2024,” Badan Pusat Statistik. [Online]. Available: Https://Www.Bps.Go.Id/Id/Statistics-Table/1/Mtaxncmx/Ekspor-Kopi-Menurut-Negara-Tujuan-Utama--2000-2024

[2] A. Ghoshray And E. Mise, “Whither The Price Of Coffee? An Analysis Of Trends And Shocks Since The Coffee Crisis,” *Empir. Econ.*, Vol. 68, No. 6, Pp. 2655–2672, 2025, Doi: 10.1007/S00181-025-02716-1.

[3] M. Al-Mahish And R. Alfayadh, “The Impact Of Coffee Quality Attributes And Ratings On Specialty Coffee Bean Prices,” *Beverage Plant Res.*, Vol. 4, No. 1, Pp. 0–0, 2024, Doi: 10.48130/Bpr-0024-0029.

[4] K. Muiz, N. Nilahayati, N. Nasruddin, J. Jamidi, And H. Hafifah, “Analisis Kualitas Fisik Dan Cita Rasa Kopi Arabika (Coffea Arabica L.) Pada Dua Rentang Ketinggian,” *J. Agrium*, Vol. 21, No. 3, P. 258, 2024, Doi: 10.29103/Agrium.V21i3.18994.

[5] M. F. Kurniawan, D. Qisthina, And T. Fitrilia, “Karakteristik Sensori Dan Fisikokimia Seduhan Kopi Sangrai Liberika, Arabika, Dan Robusta Dari Indonesia,” *Food Sci. J. Food Sci. Technol.*, Vol. 4, No. 2, Pp. 100–113, 2024, Doi: 10.33830/Fsj.V4i2.9063.2024.

[6] D. Qisthina, Muhammad Fakih Kurniawan, And Tiana Fitrilia, “Deskripsi Atribut Sensori Tiga Jenis Kopi (Arabika, Robusta, Dan Liberika) Asal Indonesia Dan Hasil Cupping Score,” *Karimah Tauhid*, Vol. 3, No. 8, Pp. 9031–9042, 2024, Doi: 10.30997/Karimahtauhid.V3i8.14611.

[7] Specialty Coffee Association, “Cupping Protocol,” 2020. [Online]. Available: Https://Sca.Coffee/Research/Protocols-Best-Practices

[8] F. Adam, R. Agustina, And R. Fadhil, “Pengujian Cita Rasa Kopi Arabika Dengan Metode Cupping Test,” *J. Ilm. Mhs. Pertan.*, Vol. 7, No. 1, Pp. 517–521, Mar. 2022, Doi: 10.17969/Jimfp.V7i1.19021.

[9] D. Hamdhana, M. Yusra, N. A. Yunita, R. G. Putri, A. Arliansyah, And M. Mardiaton, “Pembekalan Keterampilan Cupping Kopi Pada Generasi Milenial Di Aceh Utara,” *J. Pengabdi. Sos.*, Vol. 2, No. 2, Pp. 2708–2716, 2024, Doi: 10.59837/Aaamvz25.

[10] D. Giacalone, I. Steen, J. Alstrup, And M. Münchow, “Inter-Rater Reliability Of ‘Clean Cup’ Scores By Coffee Experts,” *J. Sens. Stud.*, Vol. 35, No. 5, 2020, Doi: 10.1111/Joss.12596.

[11] F. Amalia *Et Al.*, “Identification Of Potential Quality Markers In Indonesia’s Arabica Specialty Coffee Using GC/MS-Based Metabolomics Approach,” *Metabolomics*, Vol. 19, No. 11, Pp. 1–11, 2023, Doi: 10.1007/S11306-023-02051-5.

[12] Z. Iqbal *Et Al.*, “Studi Miniatur Uv/Vis/Nir Spektrometer Untuk Proses Kuantifikasi Mutu Biji Kopi Dengan Protokol Cupping Test,” *J. Teknotan*, Vol. 18, No. 1, P. 1, 2024, Doi: 10.24198/Jt.Vol18n1.1.

[13] A. Michael And J. Rusman, “Jurnal Teknologi Dan Manajemen Informatika Klasifikasi Cacat Biji Kopi Menggunakan Metode Transfer Learning Dengan Hyperparameter Tuning Gridsearch,” Vol. 9, No. 1, Pp. 37–45, 2023.

[14] H. Basri, H. Darwis, And F. Umar, “Jurnal Teknologi Dan Manajemen Informatika Klasifikasi Daun Herbal Menggunakan K-Nearest Neighbor Dan Convolutional Neural Network Dengan Ekstraksi Fourier Descriptor,” Vol. 9, No. 2, Pp. 79–90, 2023.

[15] R. P. Putra And F. Marisa, “Jurnal Teknologi Dan Manajemen Informatika Klasifikasi Jenis Kanker Prostat Melalui Citra MRI Menggunakan Pengolahan Citra Digital,” Vol. 10, No. 2, Pp. 143–153, 2024.

[16] X. Sun, “Application Of An Improved Lightgbm Hybrid Integration Model Combining Gradient Harmonization And Jacobian Regularization For Breast Cancer Diagnosis,” *Sci. Rep.*, Vol. 15, No. 1, Pp. 1–18, 2025, Doi: 10.1038/S41598-025-86014-X.

[17] P. Septiana Rizky, R. Haiban Hirzi, And U. Hidayaturrohman, “Perbandingan Metode Lightgbm Dan Xgboost Dalam Menangani Data Dengan Kelas Tidak Seimbang,” *J Stat.*

J. Ilm. Teor. Dan Apl. Stat., Vol. 15, No. 2, Pp. 228–236, 2022, Doi: 10.36456/Jstat.Vol15.No2.A5548.

[18] B. R. Prasetyo, E. D. Wahyuni, And P. M. Kusumantara, "Komparasi Performa Model Berbasis Algoritma Random Forest Dan Lightgbm Dalam Melakukan Klasifikasi Diabetes Melitus Gestasional," *J. Inform. Dan Tek. Elektro Terap.*, Vol. 12, No. 3, Pp. 481–490, 2024, Doi: 10.23960/Jitet.V12i3.4817.

[19] E. H. M. Tang, "Coffee Quality Institute - Arabica Reviews (May 2023)," 2023. [Online]. Available: Https://Www.Kaggle.Com/Datasets/Erwinhmtang/Coffee-Quality-Institute-Reviews-May2023

[20] Y. Cai, J. Feng, Y. Wang, Y. Ding, Y. Hu, And H. Fang, "The Optuna–Lightgbm–Xgboost Model: A Novel Approach For Estimating Carbon Emissions Based On The Electricity–Carbon Nexus," *Appl. Sci.*, Vol. 14, No. 11, 2024, Doi: 10.3390/App14114632.

[21] Microsoft, "Lightgbm Documentation," 2024. [Online]. Available: Https://Lightgbm.Readthedocs.Io/

[22] P. Florek And A. Zagdański, "Benchmarking State-Of-The-Art Gradient Boosting Algorithms For Classification," 2023, [Online]. Available: Http://Arxiv.Org/Abs/2305.17094

[23] I. K. Nti, O. Nyarko-Boateng, And J. Aning, "Performance Of Machine Learning Algorithms With Different K Values In K-Fold Crossvalidation," *Int. J. Inf. Technol. Comput. Sci.*, Vol. 13, No. 6, Pp. 61–71, 2021, Doi: 10.5815/Ijitcs.2021.06.05.

[24] M. F. Kurniawan *Et Al.*, "Coffee Quality Prediction With Light Gradient Boosting Machine Algorithm Through Data Science Approach," *Appl. Sci.*, Vol. 4, No. 1, Pp. 563–573, Feb. 2024, Doi: 10.33395/Sinkron.V8i1.12169.