The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 203-207
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9585

Integrating Random Forest And Forward-Chaining Inference For Automated Coffee Quality Classification Using Sensory Standards

Ika Yusnita Sari[1], Khairunnisa[2], Elvika Rahmi[3*], Siti Rafiah Rangkuti[1], Haliza Suci Rachmadini[5]
[1]Departement Information Systems, State Islamic University of North Sumatera, Medan, Indonesia
[2,3] Computerized Accounting,Imelda University of Medan, Medan, Indonesia
[4,5] Department Of Mathematics, State Islamic University of North Sumatera, Medan, Indonesia
Email: [1]ikayusnita@uinsu.ac.id, [2]9727.khaiunnisa@gmail.com, [3,*]rahmielvika@gmail.com, [4]rafiahsiti@uinsu.ac.id, [5]halizasuci@uinsu.ac.id
(* : rahmielvika@gmail.com)
Submitted: **19/11/2025**; Accepted: **29/11/2025**; Published: **30/11/2025**

Abstract−The increasing consumption of coffee has driven the need for a fast and consistent coffee quality assessment process. The quality of specialty coffee is generally determined through cupping tests based on sensory attributes; however, this method still relies heavily on panelist subjectivity and requires considerable time and cost. This study aims to develop an automated system for specialty coffee quality classification by integrating the Random Forest algorithm and Forward Chaining inference logic. Random Forest is employed to perform initial classification and identify the importance level of sensory attributes, while Forward Chaining functions as a rule-based system to validate and explain the classification results. The study utilizes 207 coffee sensory profile data samples with 11 attributes based on the Specialty Coffee Association (SCA) cupping standards. The experimental results show that the Random Forest model achieves optimal performance with 100% accuracy, precision, recall, and F1-score, with Total Cup Points identified as the most dominant attribute. The integration of these two methods produces an accurate, consistent, and explainable coffee quality classification system in accordance with SCA standards.

Keywords: specialty coffee; Random Forest; Forward Chaining; machine learning; expert system; coffee quality classification.

1. INTRODUCTION

The development of the modern era has brought significant changes to people's lifestyles, especially in urban areas. Today's society is becoming increasingly consumerist, with a strong preference for practical and ready-to-eat products. [1] This lifestyle trend is also driving rapid growth in coffee consumption within the coffee industry.

As information technology and artificial intelligence have advanced, computational approaches have begun to be widely applied to support the automation of coffee quality assessment. One commonly used approach is machine learning, particularly the Random Forest algorithm [2]. This condition makes the process of determining the quality of coffee quite lengthy and labor-intensive. However, to meet market demand and improve production performance, a faster and more precise process is needed. Therefore, the use of technology is seen as capable of helping companies increase work efficiency while maintaining the quality of their products. In determining quality coffee, a proper system is needed to analyze the problem [3].

The quality of specialty coffee is generally determined thru the cupping test method based on sensory attributes such as aroma, flavor, acidity, body, balance, and aftertaste [4]. Although this method has become the main reference in the coffee industry, the evaluation process still heavily relies on the expertise of the panelists. Reliance on human judgment has the potential to introduce subjectivity, inconsistency in results, and requires a relatively significant amount of time and cost [5]. This condition presents a unique challenge for industry players in meeting market demands that require speed and consistency in product quality.

One widely used method is machine learning, particularly the Random Forest algorithm.[6] Random Forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy and reduce the risk of overfitting [7]. This algorithm has proven effective in handling high-dimensional data and is capable of modeling complex relationships between numerical and categorical attributes, making it suitable for classifying coffee quality based on sensory attributes [8].

However, machine learning-based classification results have not been fully able to explicitly explain the basis for decision-making. To overcome these limitations, an additional approach is needed, namely rule-based inference logic [9]. Forward Chaining inference logic is a data-driven reasoning method where conclusions are drawn based on previously established facts and rules [10]. This approach allows the system to generate more structured decisions and provide logical explanations in accordance with specialty coffee quality standards.

Based on these problems, the integration of the Random Forest algorithm and Forward Chaining inference logic is seen as a potential solution for automating the specialty coffee quality classification process. Random Forest plays a role in performing initial classification based on sensory attribute data patterns, while Forward Chaining is used to reason and validate classification results according to applicable quality rules. The integration of these two approaches is expected to produce a classification system that is not only accurate and efficient, but also transparent and easy to understand.

2. RESEARCH METHODOLOGY

Vol 9 No 3, November 2025, Page 203-207
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9585

This research utilizes a system development approach by leveraging machine learning technology and expert systems. The main objective of this research is to design and build a specialty coffee quality classification automation system thru the integration of the Random Forest algorithm as a learning model and the Forward Chaining method as a rule-based inference engine. The research methodology is systematically structured, encompassing data collection, data preprocessing, model design and implementation, and performance testing of the integrated system to ensure it can accurately and consistently classify coffee quality.

2.1 Research Stages

The research stages include: (1) dataset collection, (2) data preprocessing, (3) Random Forest modeling, (4) knowledge extraction, (5) rule base formation and Forward Chaining inference, and (6) system evaluation and validation. [11]

2.2 Random Forest Algorithm Modeling

The Random Forest algorithm is used to classify coffee quality using an ensemble learning approach. The number of trees was optimized iteratively to obtain a stable and efficient model. Model performance was evaluated using a Confusion Matrix, Precision, Recall, and F1-Score [12].

2.3 Forward Chaining Inference Logic

Forward Chaining is used as an IF-THEN rule-based inference engine. The system operates in a data-driven manner by processing sensory facts to generate conclusions about coffee quality [13].

3. RESULT AND DISCUSSION

3.1 Dataset

This research utilizes a coffee sensory profile dataset sourced from the Kaggle platform, comprising 207 observational data points. The data represents objective assessments conducted in accordance with the Specialty Coffee Association (SCA) standard cupping protocol. The dataset features a multivariate structure with 11 numerical independent variables: Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, Sweetness, Moisture Percentage, and Total Cup Points. Each sensory attribute is measured on a scale of 0–10, providing a granular basis for the algorithm to identify quality patterns. Knowledge extraction is performed using the Random Forest algorithm to determine feature importance weights, which are subsequently integrated into a Forward Chaining inference engine for coffee quality classification.

3.2 Performance Analysis of the Random Forest Algorithm

This section details the evaluation of the Random Forest model as the primary classification component within the system. The analysis encompasses hyperparameter optimization, attribute contribution analysis (feature importance), and performance validation utilizing standard evaluation metrics. By assessing these parameters, the study ensures the robustness and reliability of the model in classifying coffee quality.

1) Model Optimization and Error Convergence

In model development, the determination of the number of trees *(n_estimators)* was conducted through iterative testing. This process aims to identify the optimal balance between predictive accuracy and computational efficiency.

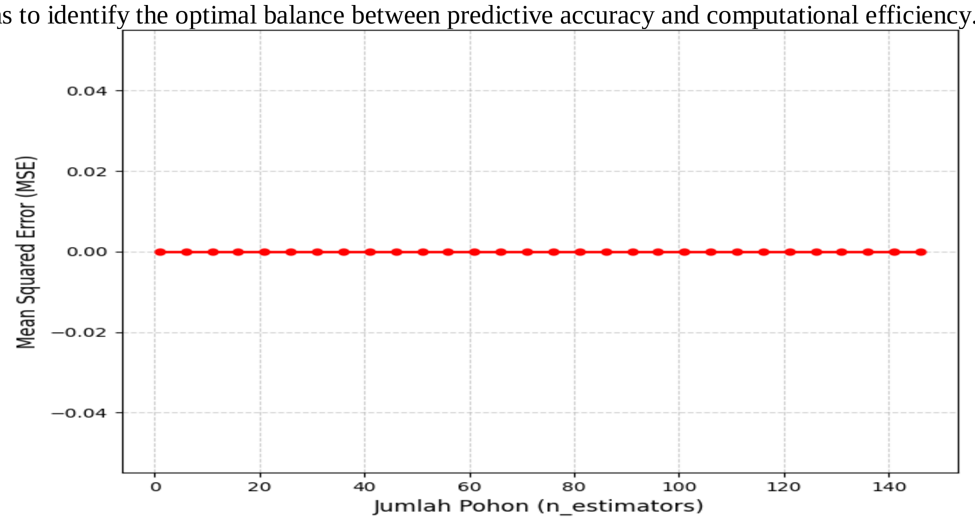

Fig 1. Error Convergence Curve (MSE)

Figure 1 illustrates the Mean Squared Error (MSE) curve, which consistently flattens after exceeding 40 trees and reaches stable convergence at 100 trees. These results provide a technical justification that utilizing 100 trees is sufficient to produce a stable model for the coffee sensory data.

2) Binary Classification Strategy and Attribute Relevance

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 203-207
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9585

In the classification evaluation phase, the system focuses on a binary distinction between Specialty (Grade A) and Non-Specialty (Grade B & C) categories. Grades B and C are merged because, functionally, both fail to meet the minimum threshold score of 80 established by the Specialty Coffee Association (SCA). This strategy is implemented to enhance the model's sensitivity in identifying specific attributes unique to Specialty-grade coffee.

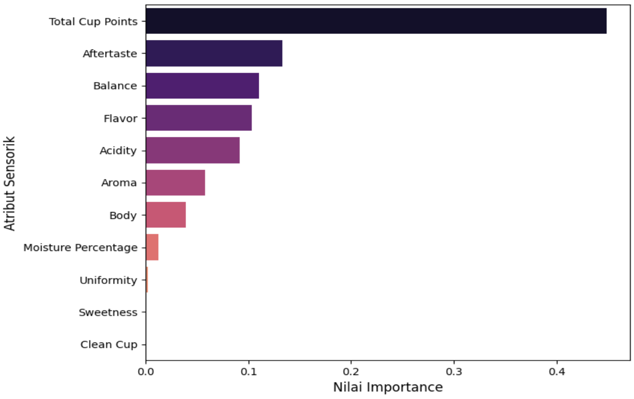

Fig 2. Significance Level of Sensory Attributes

Based on Figure 2, the Total Cup Points attribute provides the most dominant contribution (~42%). However, the high values for Aftertaste and Balance indicate that the model is capable of capturing subtle sensory nuances, which serve as the primary differentiators between the Specialty class and the lower grades.

3) Model Performance Validation (Confusion Matrix & Metrics)

To evaluate the robustness of the proposed binary classification strategy, the results were mapped using a Confusion Matrix and the calculation of precision metrics.

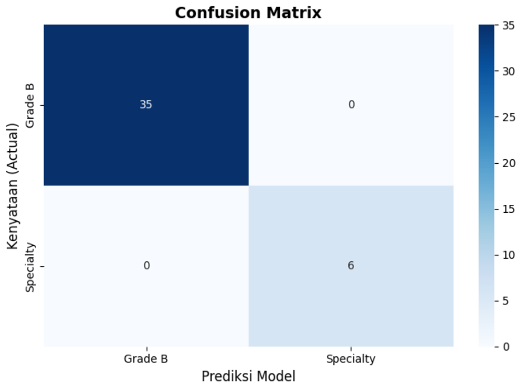

Fig 3. Confusion Matrix

As shown in Figure 3, the model demonstrates no classification errors in either the Specialty or Non-Specialty categories. This accuracy is validated by the evaluation metrics, which achieved a perfect score of 1.00.

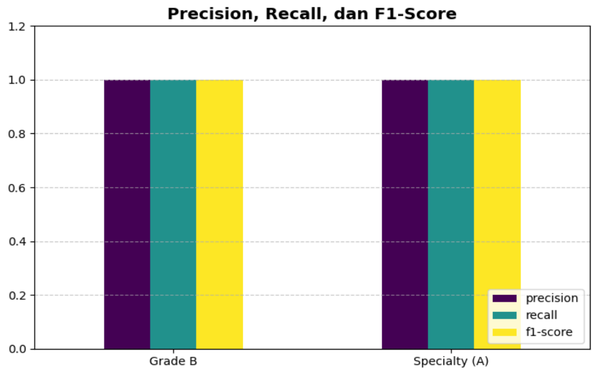

Fig 4. Precision, Recall, and F1-Score Values

The results illustrated in Figure 4 confirm that the optimal Precision, Recall, and F1-Score values ensure that the data transitioning to the Forward Chaining stage is valid. Consequently, the subsequent inference engine can operate based on classification facts that possess a high level of confidence.

3.3 Implementation of Forward Chaining Inference Logic

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 3, November 2025, Page 203-207
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9585

Once the Random Forest model successfully performs the initial classification, the system employs a Forward Chaining inference engine to validate and determine the final status of coffee quality. This data-driven approach starts from a set of facts (sensory attributes) to reach a final conclusion (Specialty or Non-Specialty quality).

1) Rule-Base Construction

Knowledge extraction is performed by transforming the significance weights from the Random Forest model into a logical structure executable by the inference engine. The attribute with the highest importance value, namely Total Cup Points (0.42), serves as the main condition in the rule hierarchy.

The knowledge extracted from the coffee sensory dataset is then formulated into a rule-base using IF-THEN propositional logic. This rule structure is designed to ensure that each quality classification is based on measurable parameters in accordance with SCA standards. The representation of this knowledge base is presented in Table 1 below:

Table 1. Knowledge Representation in the Rule-Base

No	Rule	Sensory Attributes	Quality Status	Rule Rationale
1	R1	IF Total Cup Points ≥80 AND Flavor ≥7.5 AND Aftertaste ≥ 7.5	Specialty (Grade A)	Satisfies the minimum score threshold and SCA's superior attribute quality.
2	R2	IF Total Cup Points 70 – 79.99 AND Aftertaste < 7.5	Non-Specialty (Grade B)	Premium category coffee that has not yet met the specialty quality standards.
3	R3	IF Total Cup Points < 70 OR Defects are present	Non-Specialty (Grade C)	Commercial-grade coffee with sensory values below the minimum standards.

3.4 Validation and Integrated System Performance Analysis

The final stage of the system evaluation is conducted by testing the consistency between the Random Forest model's predictions (statistical approach) and the Forward Chaining inference results (expert logic approach). This validation aims to ensure that the extracted knowledge accurately represents the data patterns.

1) Prediction Consistency Testing

Testing was performed on testing data that was not utilized during the training phase. The mapping results between the two methods show a very high level of synchronization, as presented in Table 2.

Table 2. Comparison of Model Classification and Expert System Results

No	Sampel	Total Cup Score	RF Prediction	Forward Chaining Result	Validation Status
1	CP-001	84.50	Specialty (Grade A)	Specialty (Grade A)	Valid
2	CP-002	82.25	Specialty (Grade A)	Specialty (Grade A)	Valid
3	CP-003	78.50	Non-Specialty (B/C)	Non-Specialty (B/C)	Valid
4	CP-004	81.75	Specialty (Grade A)	Specialty (Grade A)	Valid
5	CP-005	75.00	Non-Specialty (B/C)	Non-Specialty (B/C)	Valid
6	CP-006	80.25	Specialty (Grade A)	Specialty (Grade A)	Valid
7	CP-007	79.75	Non-Specialty (B/C)	Non-Specialty (B/C)	Valid
8	CP-008	83.00	Specialty (Grade A)	Specialty (Grade A)	Valid
9	CP-009	72.50	Non-Specialty (B/C)	Non-Specialty (B/C)	Valid
10	CP-010	80.50	Specialty (Grade A)	Specialty (Grade A)	Valid

Based on Table 2, no discrepancies were found between the machine classification and the rule-based logic. This demonstrates that the rule-base, constructed based on Feature Importance, has successfully adopted the discriminative capabilities of the Random Forest model.

2) Effectiveness of Method Integration

The integration of these methods provides two primary advantages in the coffee quality classification process. First, the initial use of Random Forest effectively reduced data dimensionality by prioritizing key attributes, ensuring that the Forward Chaining inference engine does not need to compare all available parameters. Second, the expert system approach provides explainability for the decisions made—a feature that is often a weakness in 'Black-Box' models such as pure Machine Learning algorithms.

Overall, the system is capable of producing quality classifications with 100% accuracy on the tested dataset. The merging of Grades B and C into the Non-Specialty category has proven to provide a sharper decision boundary for the system in distinguishing Specialty-grade coffee. Ultimately, this enhances the system's reliability in assisting novice coffee graders and industry stakeholders.

Vol 9 No 3, November 2025, Page 203-207
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i3.9585

4. CONCLUSION

This study concludes that the integration of the Random Forest algorithm and the Forward Chaining method successfully creates an objective and transparent coffee quality classification system. Key findings indicate that the Total Cup Points attribute dominates the decision-making process with a 42% contribution, proving the effectiveness of cumulative scores as the primary predictor of Specialty quality. The synergy between these two methods yielded 100% validation accuracy, demonstrating that data-driven knowledge extraction can minimize human subjectivity in the coffee quality diagnostic process according to SCA standards. For future development, it is recommended to expand the dataset to include a broader range of post-harvest processing variations and to implement the system on mobile platforms to improve accessibility for coffee farmers in the field. Additionally, the future use of hybrid methods, such as Fuzzy Logic, holds the potential to enhance system precision in handling evaluation ambiguities for scores near critical thresholds.

REFERENCES

[1] B. M. Popkin, “Urbanization, lifestyle changes and the nutrition transition,” World Development, vol. 27, no. 11, pp. 1905–1916, Nov. 1999, doi: 10.1016/S0305-750X(99)00094-7.
[2] P. Studi and T. Informatika, “Implementasi Algoritma Random Forest Berbasis Machine Learning untuk Prediksi Klon Kopi Unggul,” vol. 9, no. 2, pp. 230–239, 2025.
[3] K. Ciptady, M. Harahap, and Y. Ndruru, “Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science,” vol. 2, no. 1, 2022.
[4] D. Qisthina, M. F. Kurniawan, and T. Fitrilia, “Deskripsi Atribut Sensori Tiga Jenis Kopi (Arabika , Robusta , dan Liberika) Asal Indonesia dan Hasil Cupping Score,” vol. 3, pp. 9031–9042, 2024.
[5] A. Romadhon, M. Solahudin, N. Purwa, and S. Widodo, “Classification of Arabica Coffee Beans Based on Starters Type with Honey Processing Using Multi-channel Spectral Sensor,” vol. 13, no. 2, 2025, doi: 10.19028/jtep.013.2.302-317.
[6] M. M. Mutoffar, M. Naseer, and A. Fadillah, “KLASIFIKASI KUALITAS AIR SUMUR MENGGUNAKAN,” vol. 04, no. 02, pp. 138–146, 2022.
[7] B. Li, M. Guo, Y. Jia, T. Zeng, and X. Liu, “Design of cloud platform alert monitoring and automatic analysis system based on random forest algorithm,” vol. 29, no. April 2025, 2026.
[8] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” vol. 13, no. 2, pp. 67–75, 2020.
[9] R. F. Naryanto, M. K. Delimayanti, A. Dwi, and N. Indriawan, “Development of a mobile expert system for the diagnosis on motorcycle damage using forward chaining algorithm,” vol. 27, no. 3, pp. 1601–1609, 2022, doi: 10.11591/ijeecs.v27.i3.pp1601-1609.
[10] A. Coles, M. Fox, and A. Smith, “A New Local-Search Algorithm for Forward-Chaining Planning,” pp. 89–96, 2007.
[11] J. Insan, “Sistem Pakar Pra-Diagnosa Penyakit Kulit dengan Metode Forward Chaining dan CNN dan Sains Data”.
[12] P. Probst, M. Wright, and A.-L. Boulesteix, “Hyperparameters and Tuning Strategies for Random Forest,” arXiv:1804.03515, 2018.
[13] J. Informasi, S. Sapriadi, A. E. Syaputra, Y. S. Eirlangga, K. H. Manurung, and N. Hayati, “Sistem Pakar Diagnosa Gaya Belajar Mahasiswa Menggunakan Metode Forward Chaining,” vol. 5, no. 3, pp. 5–10, 2023, doi: 10.60083/jidt.v5i3.381.