

Interpretable Decision Tree Model for Bank Health Classification

Andriansyah Latif
Gunadarma University
Margonda Raya Street
Pondok Cina Depok

Sari Noorlima Yanti
Gunadarma University
Margonda Raya Street
Pondok Cina Depok

Dina Kusuma Astuti
Gunadarma University
Raya Street
Pondok Cina Depok

ABSTRACT

The health of the banking sector is a key determinant of financial stability and economic resilience, making reliable assessment methods essential for regulators and stakeholders. Conventional statistical models, however, often lack interpretability and alignment with supervisory frameworks. This study introduces a transparent classification approach for bank health evaluation by integrating the CAMEL framework covering capital, asset quality, management, earnings, and liquidity with the Decision Tree C4.5 algorithm. Using secondary data from the Financial Services Authority of Indonesia, the dataset comprises 170 observations from 34 banks during 2019–2023. The research process involves data cleaning, preprocessing, label encoding, and partitioning into training and testing subsets with an 80:20 ratio. The C4.5 algorithm constructs the decision tree by iteratively selecting attributes with the highest information gain to minimize entropy, producing a model that classifies banks into four categories: healthy, fairly healthy, less healthy, and unhealthy. Results show that the composite CAMEL score offers the strongest discriminative power, serving as the root node and the most significant predictor of bank health. Model evaluation through precision, recall, and F1-score confirms consistent predictive performance while preserving interpretability. Unlike prior studies focusing mainly on credit risk or bankruptcy prediction, this research delivers a regulatory-oriented framework that is comprehensive, practical, and relevant for strengthening supervisory practices.

General Terms

Banking, Financial Stability, Bank Soundness Evaluation, Regulatory-Based Classification, Interpretable Machine Learning.

Keywords

CAMEL, Classification, Credit Risk, Decision Tree C4.5, Financial Ratios.

1. INTRODUCTION

Banks play a strategic role in maintaining financial stability while simultaneously fostering national economic growth [1]. As financial intermediaries, banks mobilize and allocate public funds to enhance living standards and strengthen a country's economic resilience [2]. The regulatory framework in Indonesia, through various provisions issued by the national banking authorities [3], emphasizes the importance of an objective, transparent, and internationally aligned bank soundness assessment system. The evaluation of bank soundness serves as a crucial instrument for management, regulators, investors, and the public to assess performance and mitigate potential financial risks [4].

One of the most widely adopted methods for assessing bank soundness is the CAMEL approach (Capital, Asset Quality,

Management, Earnings, Liquidity) [5]. The CAMEL ratios enable the classification of banks into categories of sound, fairly sound, less sound, or unsound [6]. This classification assists banks in improving policies and financial performance, enhancing operational efficiency and effectiveness, and ultimately strengthening customer and investor confidence [7].

Numerous prior studies have examined bank soundness assessment. Research on stability and risk has explored the role of societal trust and organizational trust in mitigating market risk within the banking sector. Using a large global sample comprising 10,616 bank-year observations from 45 countries, study [8] demonstrates that higher levels of trust significantly reduce overall bank risk. This mitigating effect is more pronounced in banks operating in countries with weak investor protection, dissatisfaction with government economic policies, and political instability. The findings underscore that trust can function as an alternative governance mechanism that complements or even substitutes weak formal institutions, with important implications for financial regulators worldwide.

Study [9] analyzed the efficiency of conventional and Islamic banks using a machine learning approach. The dataset consisted of annual financial reports from 11 banks (seven conventional and four Islamic) over the period 2014–2018. Research variables were categorized into five primary financial ratio groups: profitability, managerial practices, assets and loans, capital adequacy (CAR), and liquidity. Linear regression was employed to analyze relationships among variables, while predictive modeling was conducted by comparing four algorithms: Support Vector Machine (SVM), Chi-squared Automatic Interaction Detector (CHAID), Neural Network, and linear regression. The results indicate that profitability, liquidity, and managerial practices significantly influence bank efficiency, with the CHAID algorithm demonstrating the best predictive performance.

Study [10] examined the application of artificial intelligence in the financial sector, highlighting how machine learning has revolutionized credit risk analysis and loan management. The study notes that traditional methods, such as expert judgment and statistical regression, face limitations in addressing the complexity of modern financial data. In contrast, machine learning techniques—including Decision Trees, Random Forests, and Deep Learning—offer more accurate predictions of creditworthiness by leveraging large and diverse datasets. However, the study also emphasizes ethical concerns, such as algorithmic bias and the lack of transparency in “black box” models, thereby underscoring the need for more interpretable, transparent, and fair decision-making models in banking.

Study [11] conducted a systematic review of artificial intelligence models used to predict financial institution bankruptcy, employing the PRISMA and PICOC frameworks. The review evaluated the effectiveness of traditional models—

such as logistic regression, K-Nearest Neighbors (KNN), Random Forest, and neural networks—as well as more advanced approaches, including XGBoost, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and ensemble methods. The findings indicate that while traditional models excel in simplicity and computational efficiency, advanced AI-based models achieve substantially higher predictive accuracy.

Study [12] discussed the application of machine learning (ML) for credit risk prediction, with a particular focus on the financial industry, especially microfinance institutions. The study aimed to assess the extent to which ML improves default prediction accuracy and to identify its challenges and limitations. The findings reveal that boosted model families (such as XGBoost and Gradient Boosting) are the most frequently used algorithms for credit risk prediction, followed by Random Forest, Logistic Regression, KNN, and Neural Networks.

Despite these contributions, several limitations remain. Study [9] is constrained by its limited sample size—only 11 banks over a five-year period—thus restricting generalizability. Moreover, its focus is confined to cost efficiency without considering broader dimensions of bank soundness, and model evaluation relies solely on R^2 across four basic algorithms, without comparison to more robust ML methods. Study [10] primarily concentrates on individual credit risk analysis and does not extend its perspective to the institutional level within a regulatory framework such as CAMEL, even though institutional bank soundness is critical to financial system stability. Study [11] emphasizes AI-based bankruptcy prediction but does not integrate fundamental financial indicators commonly used by regulators, limiting its direct relevance to supervisory practices. Meanwhile, study [12], although mapping the state of the art in ML algorithms for credit risk prediction, continues to face challenges related to interpretability (black box issues) and data imbalance.

This study implements the Decision Tree C4.5 model to classify bank soundness levels based on CAMEL ratios (CAR, NPL, BOPO, NIM, ROA, and LDR). Unlike previous studies, this research utilizes official data from the Financial Services Authority (OJK) for the period 2019–2023, covering 34 banks with a total of 170 observations, thereby offering broader and more representative institutional-level coverage. The research process includes data cleaning, preprocessing, an 80:20 data-splitting scheme, model training, and evaluation using a confusion matrix with precision, recall, and F1-score metrics to ensure classification quality.

The specific contribution of this study lies in integrating the regulatory CAMEL framework with the Decision Tree C4.5 algorithm, which is inherently transparent and interpretable. Consequently, the classification results not only hold academic value but also offer practical relevance for regulators, bank management, and stakeholders. This study complements the literature on credit risk analysis and bankruptcy prediction by providing a more applicable, regulation-oriented, and interpretable predictive model to support banking stability supervision.

2. RESEARCH METHOD

This study classifies bank soundness levels through several systematic stages based on the CAMEL ratios. The use of financial ratios such as CAR, NPL, ROA, and LDR as predictive indicators of bank stability is consistent with recent empirical studies on banking risk modeling and institutional financial health assessment [13,14]. The first stage involves collecting CAMEL indicator data—comprising CAR, NPL,

BOPO, NIM, ROA, and LDR from the official website of Otoritas Jasa Keuangan (OJK). Subsequently, the credit score for each component is calculated, followed by the determination of the final CAMEL composite rating according to the predetermined weighting scheme. The designed model is then trained using the training dataset. The output of this process is a trained Decision Tree model that can be utilized to classify bank soundness levels, in line with interpretable machine learning approaches increasingly adopted in financial supervision research [15].

2.1 Bank Soundness Data

This study employs CAMEL ratios obtained from OJK for the 2019–2023 period, covering 34 banks with a total of 170 observations. The use of panel banking data across multiple years enhances the robustness of institutional risk modeling and aligns with recent financial distress prediction frameworks [13,16]. Table 1 presents a sample of the bank soundness rating data used in this research.

Table 1. Sample of Bank Soundness Rating Classification

Bank Name	Year	CAMEL Percentage	Ranking
Bank Central Asia	2019	93,24%	1
Bank Central Asia	2020	92,52%	1
Bank QNB Indonesia	2019	49,99%	4
Bank QNB Indonesia	2020	83,16%	1
Bank INA Perdana	2019	74,66%	2
Bank INA Perdana	2020	84,49%	1
Bank Sinarmas	2019	49,35%	4
Bank Sinarmas	2020	63,55%	3
Bank CIMB Niaga	2019	86,01%	1
Bank CIMB Niaga	2020	84,21%	1

2.2 Data Cleaning

In this study, the data cleaning procedure was conducted using built-in Python functions from the pandas library. The function `df.isnull().sum()` was employed to detect and identify missing or null values in each column, enabling an assessment of data completeness. Furthermore, the function `df.duplicated().sum()` was utilized to calculate the number of duplicate rows corresponding to previous entries, thereby allowing redundant data to be identified and removed. Systematic data validation and preprocessing are essential steps in financial machine learning pipelines to ensure model reliability and to reduce bias caused by incomplete or redundant observations [15,17].

Suppose there is a dataset D of size $n \times m$, where n represents the number of rows (records) and m represents the number of columns (attributes). Each data element is denoted as x_{ij} , where $i=1,2,\dots,n$ and $j=1,2,\dots,m$. The number of missing values in the j -th column is calculated using Equation (1):

$$MV_j = \sum_{i=1}^n \delta(x_{ij}) \quad (1)$$

Using the indicator function $\delta(x_{ij})$ which is defined as shown in Equation (2):

$$\delta(x_{ij}) = \begin{cases} 1, & \text{if } x_{ij} = \emptyset \text{ (null)} \\ 0, & \text{if } x_{ij} \neq \emptyset \text{ (null)} \end{cases} \quad (2)$$

Thus, the total number of missing values in the dataset is calculated using Equation (3):

$$MV_{total} = \sum_{j=1}^m MV_j \quad (3)$$

2.2 Data Preprocessing

After the data cleaning process was completed, the next step involved data preprocessing, which aims to transform categorical attributes into numerical form [18-19]. This transformation is necessary because most machine learning algorithms can only process numerical data [20]. In this study, the *LabelEncoder* from the *scikit-learn* library was employed to convert textual labels into integer representations. Formally, this process is defined as a mapping function $f: X \rightarrow Z$, where each unique category c_j in the set of categorical attributes is mapped to an integer value such that $f(c_j) = j-1$. The preprocessed dataset is not only more structured and simplified but also compatible with various analytical algorithms applied in the modeling stage [21]. The integration of data cleaning and preprocessing results in a dataset that has undergone both validation and transformation. For example, if $X = \{Bank A, Bank B, Bank C\}$, then the transformation yields: $f(Bank A) = 0, f(Bank B) = 1, f(Bank C) = 2$.

2.3 Development of the Bank Soundness Classification Model

The modeling stage was conducted by developing a classification model based on the Decision Tree C4.5 algorithm to learn patterns from the dataset that had undergone preprocessing and data splitting. The C4.5 algorithm was selected due to its ability to generate clear and interpretable classification rules, as well as its effectiveness in handling both numerical and categorical attributes. Unlike previous studies that primarily focused on individual credit risk analysis or bankruptcy prediction using “black-box” models such as neural networks and ensemble methods, this study emphasizes institutional-level bank soundness classification within the regulatory CAMEL framework recognized by banking authorities.

In this study, the training dataset was used to construct the decision tree, while the testing dataset was employed to evaluate the performance of the resulting model. The decision tree construction process was carried out iteratively by selecting the attribute that contributed the most to reducing data uncertainty (entropy). In this case, the CAMEL composite score was selected as the root node because it produced the highest information gain compared to other attributes, making it the most significant variable in determining bank soundness levels.

This finding highlights the novelty of the study, namely the integration of the regulatory CAMEL indicators with the interpretable C4.5 model, resulting in a bank soundness classification that is not only accurate but also transparent and practically relevant for regulators [22-23]. The importance of interpretable machine learning models in financial supervision has been increasingly emphasized in recent literature, particularly in contrast to black-box ensemble and deep learning approaches [24-25]. Subsequently, the dataset was partitioned into branches based on the values of the selected attribute. Each sample was allocated to a branch according to its attribute value, and the splitting process continued until all samples within a branch were classified into the same class, following the standard recursive partitioning framework widely adopted in decision tree learning [26-27].

In general, the stages of constructing the C4.5 decision tree in this study consist of the following steps:

1. Calculating the entropy and information gain values for each attribute using Equations (4) and (5). Entropy is used to measure class impurity, while information gain quantifies the reduction in uncertainty after splitting the dataset based on a given attribute [28-29]. These measures are fundamental criteria in decision tree optimization and have been extensively validated in classification problems across financial and risk modeling domains [30].

$$Entropy(S) = - \sum_{i=1}^k p_i \log_2(p_i) \quad (4)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (5)$$

where p_i denotes the probability of class i within the data subset S .

2. Selecting the attribute with the highest information gain as the root node. The attribute with the maximum information gain is chosen as the root node of the decision tree using Equation (6). This greedy selection strategy ensures that each split maximally improves class separation at every stage of the tree construction process [26,31].

$$Root = \arg \max_{A \in Attributes} Gain(S, A) \quad (6)$$

3. Partitioning the dataset into branches based on the selected attribute. After attribute A is chosen as the root node (based on the highest information gain), the dataset S is divided into subsets S_j according to the possible values of attribute A :

$$S = \bigcup_{j=1}^v S_j, \quad S_i \cap S_j = \emptyset \text{ for } i \neq j \quad (7)$$

Where v denotes the number of possible values of attribute A , and $S_j = \{(x, y) \in S | A(x) = v_j\}$, where each branch of the tree represents a data subset S_j that contains a specific value of attribute A .

4. The process is repeated for each branch until all samples within a branch have homogeneous class labels or the remaining attributes are no longer significant. For each subset S_j , the algorithm repeats the selection of the optimal attribute using the information gain criterion. The process terminates when one of the following conditions is satisfied:

- 1) Class Homogeneity. If all samples in S_j belong to the same class, $Entropy(S_j) = 0$, then S_j becomes a leaf node labeled with the dominant class [28].
- 2) No Remaining Attributes/Zero Gain. If no attribute is able to produce positive information gain using Equation (8), the node is converted into a leaf node and assigned the dominant class label :

$$\max_{A \in Attributes} Gain(S_j, A) = 0 \quad (8)$$

In this case, the node is also converted into a leaf node, and the class label is determined based on the majority of samples using Equation (9), where C denotes the set of all classes :

$$Class(S_j) = \arg \max_{c \in C} |\{y \in S_j | y = c\}| \quad (9)$$

2.4 Evaluation of the Bank Health Classification Model

The performance of the classification model was evaluated by comparing the predictions generated by the C4.5 Decision Tree algorithm with the actual labels in the test dataset. The evaluation process began with the construction of a confusion matrix, which represents the distribution of predictions across four bank health categories, namely healthy, fairly healthy, less

healthy, and unhealthy. Based on this confusion matrix, the primary evaluation metrics “ precision, recall, and F1-score were calculated.

Precision measures the accuracy of the model in producing correct positive predictions according to the actual class, while recall assesses the model’s ability to identify all samples that truly belong to a specific class. The F1-score provides a more balanced assessment, as it represents the harmonic mean of precision and recall, thereby enabling a more comprehensive evaluation of model performance. In addition, overall accuracy was computed to determine the proportion of correct predictions relative to the total test data.

By combining these metrics, the evaluation not only emphasizes general accuracy but also considers the consistency of predictions across each bank health category. Consequently, the results offer a more in-depth and objective overview of the quality and reliability of the developed model.

3. RESULT AND DISCUSSION

This section presents the empirical findings obtained from the application of the C4.5 Decision Tree algorithm in classifying bank health levels in Indonesia based on the CAMEL framework. The presentation of results focuses on two main aspects: the technical performance of the classification model and its implications for banking supervision practices.

First, the structure of the decision tree is analyzed to identify the most influential attributes in determining bank health categories. Subsequently, the predictive performance of the model is evaluated using a confusion matrix and performance metrics, including accuracy, precision, recall, and F1-score, to ensure the reliability of the classification results. In addition, the discussion emphasizes the interpretability of the derived decision rules, enabling the model to produce not only accurate classifications but also results that are relevant from a regulatory perspective and supportive of transparent supervisory practices. These findings are then compared with previous studies to highlight the contribution and novelty of this research in the context of risk assessment and financial health evaluation of banks.

3.1 Percentage of Bank Health Levels

Based on Circular Letter No. 9/24/DPbS concerning the standard CAMEL final assessment categorization, which consists of four categories: healthy, fairly healthy, very healthy, and unhealthy—the bank health level is determined through the calculation of Capital (CAR), Asset (NPL), Management (NIM), Earnings (ROA and BOPO), and Liquidity (LDR) ratios. These values are weighted according to Bank Indonesia Regulation No. 6/10/PBI/2004.

1. Capital. Capital adequacy is a key indicator in evaluating the financial strength of a banking institution. It is measured using the Capital Adequacy Ratio (CAR), which compares capital to risk-weighted assets. Table 2 presents ten CAR samples along with their corresponding credit values. Table 2 shows that five banks during the 2019–2020 period recorded CAR values exceeding 9%, indicating that all samples were classified as very healthy and obtained maximum credit scores. This result demonstrates that the sampled banks possessed strong capital structures. The CAR credit value contributes 25% to the final CAMEL score.

Table 2. Sample of CAR Calculation

Bank Name	Year	CAR Ratio	Credit Value
BCA	2019	23,40%	100%
BCA	2020	26,90%	100%
QNB Indonesia	2019	21,08%	100%
QNB Indonesia	2020	24,53%	100%
INA Perdana	2019	37,41%	100%
INA Perdana	2020	40,08%	100%
Sinarmas	2019	17,32%	100%
Sinarmas	2020	17,29%	100%
CIMB Niaga	2019	21,47%	100%
CIMB Niaga	2020	21,92%	100%

2. Asset Quality. Asset quality is assessed using several indicators, including the Non-Performing Loan (NPL) ratio, which reflects the proportion of problematic loans relative to total loans. Table 3 indicates that Bank Central Asia (BCA) achieved a very healthy status due to its NPL ratio being below 2%. In 2020, Bank INA Perdana also attained a very healthy classification, reflecting improved asset quality. The NPL credit value contributes 30% to the final CAMEL score.

Table 3. Sample of NPL Calculation

Bank Name	Year	NPL Ratio	Credit Value
BCA	2019	1,30%	94,67%
BCA	2020	1,80%	91,33%
QNB Indonesia	2019	9,83%	37,80%
QNB Indonesia	2020	4,66%	72,27%
INA Perdana	2019	4,76%	71,60%
INA Perdana	2020	1,43%	93,80%
Sinarmas	2019	8,83%	44,47%
Sinarmas	2020	4,75%	71,67%
CIMB Niaga	2019	2,79%	84,73%
CIMB Niaga	2020	3,62%	79,20%

3. Management. Management performance is evaluated based on capital management, asset quality, profitability, liquidity, and general management practices. The Net Interest Margin (NIM) is used as an indicator, reflecting the comparison between net income and operating income. Table 4 shows that Bank Central Asia and Bank Sinarmas were classified as healthy during 2019–2020, as their NIM values ranged between 5% and 9.5%. The NIM credit value contributes 25% to the final CAMEL score.

Table 4. Sample of NIM Calculation

Bank Name	Year	OJK NIM	NIM Ratio	Credit Value
BCA	2019	6,20%	79,35%	79,35%
BCA	2020	5,70%	80,48%	80,48%
QNB Indonesia	2019			
QNB Indonesia	2020	2,56%	12,44%	12,44%
INA Perdana	2019	1,61%	156,45%	100%
INA Perdana	2020	3,78%	59,45%	59,45%
INA Perdana	2020	3,40%	62,18%	62,18%
Sinarmas	2019	7,31%	0,77%	0,77%
Sinarmas	2020	6,25%	24,00%	24,00%
CIMB Niaga	2019	5,31%	74,18%	74,18%
CIMB Niaga	2020	4,88%	67,46%	67,46%

4. Earnings. Return on Assets (ROA) reflects the efficiency of banks in utilizing their assets to generate pre-tax profits and serves as a key profitability indicator. Table 4 presents ten ROA samples. Table 4 shows that Bank Central Asia achieved a very healthy classification due to ROA values exceeding 1.5%. In 2019, Bank Sinarmas also attained a very healthy status, although a decline was observed in 2020. The ROA credit value contributes 5% to the final CAMEL score. Operational efficiency is further assessed using the BOPO ratio, which measures operating expenses relative to operating income. Table 5 presents ten BOPO samples. Table 5 indicates that Bank Central Asia and Bank CIMB Niaga were classified as very healthy due to BOPO values below 94%. In 2020, Bank INA Perdana also achieved a very healthy classification. The BOPO credit value contributes 5% to the final CAMEL score

Table 5. Sample of NIM Calculation

Bank Name	Year	BOPO Ratio	Credit Value
BCA	2019	59,10%	100%
BCA	2020	63,50%	100%
QNB	2019		
Indonesia		99,40%	8,50%
QNB	2020		
Indonesia		116,14%	0,00%
INA Perdana	2019	96,08%	50,00%
INA Perdana	2020	93,60%	81,00%
Sinarmas	2019	119,43%	0,00%
Sinarmas	2020	111,70%	0,00%
CIMB Niaga	2019	82,44%	100%
CIMB Niaga	2020	89,38%	100%

5. Liquidity. Liquidity is evaluated using the Loan to Deposit Ratio (LDR), which compares total loans to total deposits. Table 6 presents ten LDR samples. Table 6 shows that Bank INA Perdana achieved a very healthy classification due to LDR values below 75%. In 2020, Bank Central Asia and Bank Sinarmas also attained very healthy status. The LDR credit value contributes 10% to the final CAMEL score.

Table 6. Sample of NIM Calculation

Bank Name	Year	LDR Ratio	Credit Value
BCA	2019	80,50%	100%
BCA	2020	65,80%	100%
QNB	2019		
Indonesia		84,70%	100%
QNB	2020		
Indonesia		97,02%	72,92%
INA Perdana	2019	62,94%	100%
INA Perdana	2020	41,26%	100%
Sinarmas	2019	81,95%	100%
Sinarmas	2020	56,97%	100%
CIMB Niaga	2019	97,64%	70,44%
CIMB Niaga	2020	82,91%	100%

3.2 Results of C4.5 Decision Tree Modeling

The modeling results are illustrated in Figure 1, which depicts the structure of the decision tree with the CAMEL score as the root node. The tree demonstrates how the dataset is hierarchically partitioned based on CAMEL values to determine bank health classifications. Several important findings from the decision tree are summarized as follows:

- Root Node (CAMEL ≤ 0.798)

This node represents the starting point of the decision tree, where the most significant attribute, CAMEL score, is selected. A Gini impurity value of 0.414 indicates moderate class heterogeneity. Among 136 samples, the class distribution [99, 32, 3, 2] shows that most banks were classified as healthy. Consequently, this node was provisionally labeled as Class 1 (healthy)

- Left Branch (CAMEL ≤ 0.656)
The Gini impurity decreased to 0.243, indicating increased homogeneity. Among 37 samples, most banks were classified as fairly healthy (Class 2), suggesting that lower CAMEL scores tend to correspond to fairly healthy banks.
- Right Sub-Branch (CAMEL ≤ 0.564)
This node exhibited a higher Gini impurity (0.480), reflecting mixed class distribution. With samples distributed between Classes 3 and 4, this node represents a critical boundary between less healthy and unhealthy banks.
- Left Sub-Sub-Branch
A Gini impurity of 0.0 indicates perfect homogeneity. All 32 samples were classified as Class 2 (fairly healthy), demonstrating strong and consistent patterns.
- Right Sub-Sub-Branch
This node also achieved perfect homogeneity (Gini = 0.0), with all samples classified as Class 4 (unhealthy), highlighting the model's capability to detect critical financial conditions.
- Right Sub-Sub-Branch of the Right Branch
With a Gini impurity of 0.0, all three samples were classified as Class 3 (less healthy), indicating distinct financial patterns associated with this category.

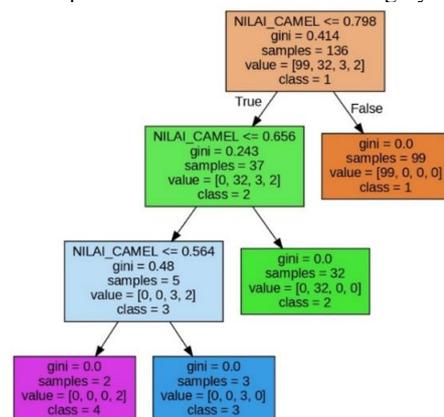


Figure 1 : Results of C4.5 Decision Tree Modeling for Bank Health Classification

3.3 Evaluation Results of the C4.5 Decision Tree Model

This evaluation stage was conducted to compute the confusion matrix based on the testing dataset and to calculate three performance metrics: precision, recall, and F1-score. The overall accuracy of the model was also determined. Based on Figure 2, the confusion matrix indicates 24 true positive (TP) instances for Class 1, 0 false positives (FP), 1 false negative (FN), and 9 true negatives (TN).

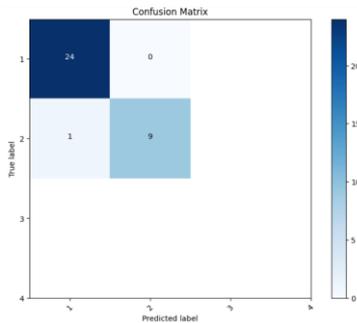


Figure 2: Confusion Matrix Result

The overall results demonstrate that the C4.5 algorithm performed exceptionally well in predicting class labels, achieving an overall accuracy of 97%. The high precision, recall, and F1-score values indicate that the model is highly capable of identifying specific classes accurately. Therefore, this evaluation confirms that the proposed model can be effectively applied for classifying new banking data.

4. CONCLUSION

This study developed an interpretable bank soundness classification model by integrating the CAMEL regulatory framework with the Decision Tree C4.5 algorithm using official data from Otoritas Jasa Keuangan for the 2019–2023 period. Based on 170 observations from 34 banks, the results indicate that the CAMEL composite score is the most influential attribute in determining bank health categories, forming the root of the decision tree and generating clear classification thresholds. The model achieved an overall accuracy of 97%, supported by high precision, recall, and F1-score values, demonstrating strong reliability across health categories. These findings confirm that integrating CAMEL indicators with an interpretable C4.5 model provides an accurate, transparent, and regulation-oriented approach to bank soundness classification, offering practical value for supervisory monitoring and decision-making.

5. ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to Universitas Gunadarma for the institutional support and academic environment that facilitated the completion of this research.

6. REFERENCES

[1] S. A. Athari, F. Irani, and A. AlAl Hadood, “Country risk factors and banking sector stability: Do countries’ income and risk-level matter? Evidence from global study,” *Heliyon*, vol. 9, no. 10, 2023, Art. no. e20398. DOI: 10.1016/j.heliyon.2023.e20398.

[2] A. Demirgüç-Kunt, L. Klapper, D. Singer, and S. Ansar, *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. Washington, DC, USA: World Bank, 2022. DOI: 10.1596/978-1-4648-1897-4.

[3] European Central Bank, *Financial Stability Review*. Frankfurt, Germany: European Central Bank, Nov. 2023.

[4] A. Citterio, “Bank failure prediction models: Review and outlook,” *Socio-Economic Planning Sciences*, vol. 92, 2024, Art. no. 101818. DOI: 10.1016/j.seps.2024.101818.

[5] E. Baffour Gyau, M. Appiah, B. A. Gyamfi, T. Achie, and M. A. Naeem, “Transforming banking: Examining the role of AI technology innovation in boosting banks

financial performance,” *International Review of Financial Analysis*, vol. 96, 2024, Art. no. 103700. DOI: 10.1016/j.irfa.2024.103700.

[6] O. Abdelsalam, A. Chantziaras, N. L. Joseph, and N. Tsileponis, “Trust matters: A global perspective on the influence of trust on bank market risk,” *Journal of International Financial Markets, Institutions & Money*, vol. 92, 2024, Art. no. 101959. DOI: 10.1016/j.intfin.2024.101959.

[7] H. F. Assous, “Prediction of banks efficiency using feature selection method: Comparison between selected machine learning models,” *Complexity*, vol. 2022, 2022, Art. no. 3374489. DOI: 10.1155/2022/3374489.

[8] L. J. Vásquez-Serpa, C. Rodríguez, J. R. Pérez-Núñez, and C. Navarro, “Challenges of artificial intelligence for the prevention and identification of bankruptcy risk in financial institutions: A systematic review,” *Journal of Risk and Financial Management*, vol. 18, no. 1, 2025, Art. no. 26. DOI: 10.3390/jrfm18010026.

[9] J. P. Noriega, L. A. Rivera, and J. A. Herrera, “Machine learning for credit risk prediction: A systematic literature review,” *Data*, vol. 8, no. 11, 2023, Art. no. 169. DOI: 10.3390/data8110169.

[10] A. Citterio and F. S. Fontana, “Bank distress prediction: A machine learning approach,” *Expert Systems with Applications*, vol. 177, 2021, Art. no. 114960. DOI: 10.1016/j.eswa.2021.114960.

[11] T. M. Pham and H. T. Nguyen, “Determinants of bank stability: Evidence from emerging markets,” *Research in International Business and Finance*, vol. 60, 2022, Art. no. 101602. DOI: 10.1016/j.ribaf.2021.101602.

[12] S. Barboza, H. Kimura, and E. Altman, “Machine learning models and bankruptcy prediction,” *Journal of Banking & Finance*, vol. 140, 2022, Art. no. 106531. DOI: 10.1016/j.jbankfin.2022.106531.

[13] M. S. Khan, S. Scheule, and E. Wu, “Predicting bank failure using machine learning techniques,” *International Review of Financial Analysis*, vol. 74, 2021, Art. no. 101662. DOI: 10.1016/j.irfa.2021.101662.

[14] A. Alaka *et al.*, “Systematic review of supervised machine learning in financial risk prediction,” *Expert Systems with Applications*, vol. 168, 2021, Art. no. 114280. DOI: 10.1016/j.eswa.2020.114280.

[15] S. García, J. Luengo, and F. Herrera, “Data preprocessing in data mining,” *Knowledge-Based Systems*, vol. 215, 2021, Art. no. 106740. DOI: 10.1016/j.knosys.2020.106740.

[16] A. K. Jain and B. B. Gupta, “Machine learning based financial risk prediction: A systematic review,” *IEEE Access*, vol. 9, pp. 131825–131845, 2021. DOI: 10.1109/ACCESS.2021.3114306.

[17] M. Zięba, S. K. Tomczak, J. M. Tomczak, and J. Świątek, “Ensemble boosted trees with synthetic features generation in financial prediction,” *Expert Systems with Applications*, vol. 157, 2020, Art. no. 113477. DOI: 10.1016/j.eswa.2020.113477.

[18] P. Biecek and T. Burzykowski, “Explanatory model analysis in machine learning for financial applications,”

- Information Sciences*, vol. 571, pp. 158–176, 2021. DOI: 10.1016/j.ins.2021.04.043.
- [19] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324.
- [20] J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986. DOI: 10.1023/A:1022643204877.
- [21] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.
- [22] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [23] F. Petropoulos, A. Apiletti, V. Assimakopoulos, and M. Papadopoulos, “Explainable artificial intelligence in finance: A systematic review,” *Expert Systems with Applications*, vol. 206, 2022, Art. no. 117749. DOI: 10.1016/j.eswa.2022.117749.
- [24] A. Bussmann, N. Giudici, D. Marinelli, and J. Papenbrock, “Explainable machine learning in credit risk management,” *Computational Economics*, vol. 57, no. 1, pp. 203–216, 2021. DOI: 10.1007/s10614-020-10042-0.
- [25] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.
- [26] J. Friedman, T. Hastie, and R. Tibshirani, “Additive logistic regression: A statistical view of boosting,” *Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000. DOI: 10.1214/aos/1016218223.
- [27] G. Louppe, L. Wehenkel, A. Suter, and P. Geurts, “Understanding variable importances in forests of randomized trees,” in *Advances in Neural Information Processing Systems*, vol. 26, 2013.
- [28] H. Ishwaran and U. B. Kogalur, “Random survival forests for R,” *R News*, vol. 7, no. 2, pp. 25–31, 2007.
- [29] R. B. Kearns and Y. Mansour, “On the boosting ability of top-down decision tree learning algorithms,” *Journal of Computer and System Sciences*, vol. 58, no. 1, pp. 109–128, 1999. DOI: 10.1006/jcss.1998.1605.
- [30] P. Domingos, “A few useful things to know about machine learning,” *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012. DOI: 10.1145/2347736.2347755.
- [31] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.