



Contents lists available at [Indonesian Scholar Society](https://journal.solusiriset.com/index.php/ijdr/index)  
**Indonesian Journal of Data Risk Research**  
Journal Homepage:  
<https://journal.solusiriset.com/index.php/ijdr/index>  
e-ISSN: XXXX-XXXX



# Comparative Analysis of 2024 Election Results Classification Using Machine Learning (K-Means + KNN) and Deep Learning (MLP) Approaches

Revaldo Dafa Fahmino, Christyan Tamaro Nadeak\*, Luluk Muthoharoh

Data Science Study Program, Sumatra Institute of Technology, Indonesia

Corresponding Email: [christyan.nadeak@sd.itera.ac.id](mailto:christyan.nadeak@sd.itera.ac.id)

## ABSTRACT

*General elections, as a foundation of democracy, require a thorough understanding of voting patterns influenced by socio-economic factors. This research compares two classification approaches: K-Means combined with K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP), in predicting the 2024 election results based on education and economic indicators. The dataset comprises 514 districts and municipalities in Indonesia, using six numerical features from the Central Statistics Agency (BPS) and class labels derived from the presidential candidate pair with the highest vote share. To address the mixed data structure, the KNN model employed Gower Distance, while the MLP model utilized standardized data and hyperparameter tuning via Bayesian Optimization. Evaluation results indicate that the K-Means + KNN model achieved an accuracy of 0.8932 and a weighted F1 score of 0.8709. Conversely, the MLP model attained the highest ROC-AUC score of 0.7420, reflecting its advantage in probabilistic estimation. These findings suggest that model selection should align with the primary classification objective, either predictive precision or the ability to model uncertainty in minority classes.*

## ARTICLE INFO

**Received**  
12 Jan 2026  
**Revised**  
17 March 2026  
**Accepted**  
25 Mei 2026

**Keywords:** Election Classification; Gower Distance; K-Means; K-Nearest Neighbors; Multilayer Perceptron.

## I. INTRODUCTION

General elections (Pemilu) are a key foundation in a democratic system, where the people have the right to directly determine their leaders [1]. In Indonesia, election results often reflect the varying socio-economic dynamics of each region. Data from the Central Statistics Agency (BPS)

shows significant variations in indicators such as education levels, poverty, and GRDP between districts/cities, which indirectly influence people's political preferences [2]. This condition makes classifying election results a challenge, especially when involving complex and heterogeneous social data structures.

Traditional classification methods generally face limitations in handling non-linear and multivariate data. Therefore, this research proposes two approaches: a combination of K-Means with K-Nearest Neighbors (KNN) as a machine learning representation and Multilayer Perceptron (MLP) as a deep learning approach. K-Means + KNN is used because of its ability to group numerical data with high variation through initial clustering and support mixed features using Gower Distance [3]. In contrast, MLP is able to model complex relationships between variables effectively, thanks to its layered architecture and adaptive weight optimization [4].

Several previous studies have tested the effectiveness of this method in various classification contexts. Jain et al. applied K-Means and Support Vector Machine (SVM) for multiclass classification, which resulted in increased accuracy through clustering [5]. Gupta et al. built a hybrid model of K-Means with KNN, SVM, and J48 Decision Tree, with an F-measure of 0.91, Kappa of 0.82, and ROC area of 0.91 [6]. Alarbi and Albayrak implemented a combination of K-Means and KNN in the Class Core Algorithm (CCA), which achieved a maximum accuracy of 85.4% in nonlinear data classification [7]. Meanwhile, Badrul compared the Neural Network and the C4.5 algorithm for predicting legislative election results and showed that the Neural Network produced an accuracy of 98.50% and an AUC of 0.982, higher than C4.5, which only achieved an accuracy of 97.84% and an AUC of 0.970 [8].

Although hybrid and neural network methods have proven effective individually, there is still no study explicitly comparing K-Means + KNN and MLP in the context of classifying election results based on education and economic variables. Therefore, this research aims to fill this gap by analyzing and comparing the performance of both methods on the 2024 Indonesian election data. Utilizing six socio-economic indicators from the Central Statistics Agency (BPS) and candidate pair labels based on data from the General Elections Commission (KPU), this research will evaluate model performance using accuracy metrics, weighted F1-score, and ROC-AUC to identify the most effective classification approach in handling complex socio-political data.

## II. METHODS

This research is a comparative quantitative study aimed at evaluating and comparing the performance of two classification approaches in the context of predicting vote results for the 2024 General Election based on education and economic variables. The two approaches used in this research are a combination of K-Means with K-Nearest Neighbors (KNN) as a machine learning-based model representation and Multilayer Perceptron (MLP) as a deep learning approach.

The research design is experimental, with both models applied separately but under equivalent test scenarios. The K-Means + KNN model utilizes an initial clustering process to enrich the data structure before proceeding to the distance-based classification stage using Gower Distance. Meanwhile, the MLP uses a multilayer neural network architecture to model non-linear relationships between variables. Both approaches are optimized through a hyperparameter tuning process to obtain the best model configuration. The entire process is conducted using secondary data from the Central Statistics Agency (BPS) and the General Elections Commission (KPU), with the final result being a comparison of the classification performance between the two methods.

## 1. Research Data and Variables

This research uses secondary data obtained from two main sources, namely the Central Statistics Agency (BPS) for socio-economic variables and the General Elections Commission (KPU) for the 2024 election vote results data. The dataset covers all administrative areas at the district/city level in Indonesia, totaling 514 entities.

There are six numerical predictor variables used as features in the classification process, namely, (1) Average Years of Schooling, which represents the level of education of the adult population in a region; (2) Expected Years of Schooling, as an estimate of the potential length of education for children; (3) Percentage of Poor Population (P0), as an indicator of the absolute level of poverty; (4) Poverty Depth Index (P1), which measures the average intensity of poverty; (5) Population Density, expressed in units of people per square kilometer; and (6) Gross Regional Domestic Product (GRDP) at current prices, which reflects regional economic activity.

The target variable in this research is the classification results of the most votes in the 2024 presidential election in each district/city. Class labels are categorized into three classes based on the presidential and vice-presidential candidate pairs that received the most votes: Candidate Pair 1, Candidate Pair 2, and Candidate Pair 3. This target data is categorical and is organized into labels for multiclass classification purposes.

## 2. Architecture and Model Stages

### a. K-Means + K-Nearest Neighbors

In the first approach, the classification process begins by normalizing all numeric variables with MinMaxScaler so that each feature falls within the range [0, 1], thereby eliminating scale dominance among variables. Next, an initial clustering is performed using the K-Means algorithm to explore hidden patterns in the data and generate regional representations based on socio-economic characteristics [9]. The clustering output is added as an additional categorical feature to the main data.

The data was then divided into training and test data using a stratified split with a ratio of 80:20 to maintain the distribution of class proportions. The baseline K-Nearest Neighbors (KNN)

model was applied to the training data, with distance measurements between observations using Gower Distance because the data contained both numeric and categorical features (K-Means clustering results).

The Gower Distance between two observations  $i$  and  $j$  is calculated as [10]:

$$D_{ij} = \frac{1}{p} \sum_{k=1}^p d_{ijk} \quad (1)$$

where  $p$  is the number of features, and  $d_{ijk}$  is the partial distance between observations  $i$  and  $j$  on the  $k$ -th feature, which is defined as:

$$\begin{aligned} d_{ijk} &= \frac{|x_{ik} - x_{jk}|}{\max(x_k) - \min(x_k)}, & \text{if the numeric feature} \\ d_{ijk} &= 0, & \text{if } x_{ik} = x_{jk} \text{ (categorical features)} \\ d_{ijk} &= 1, & \text{if } x_{ik} \neq x_{jk} \text{ (categorical features)} \end{aligned} \quad (2)$$

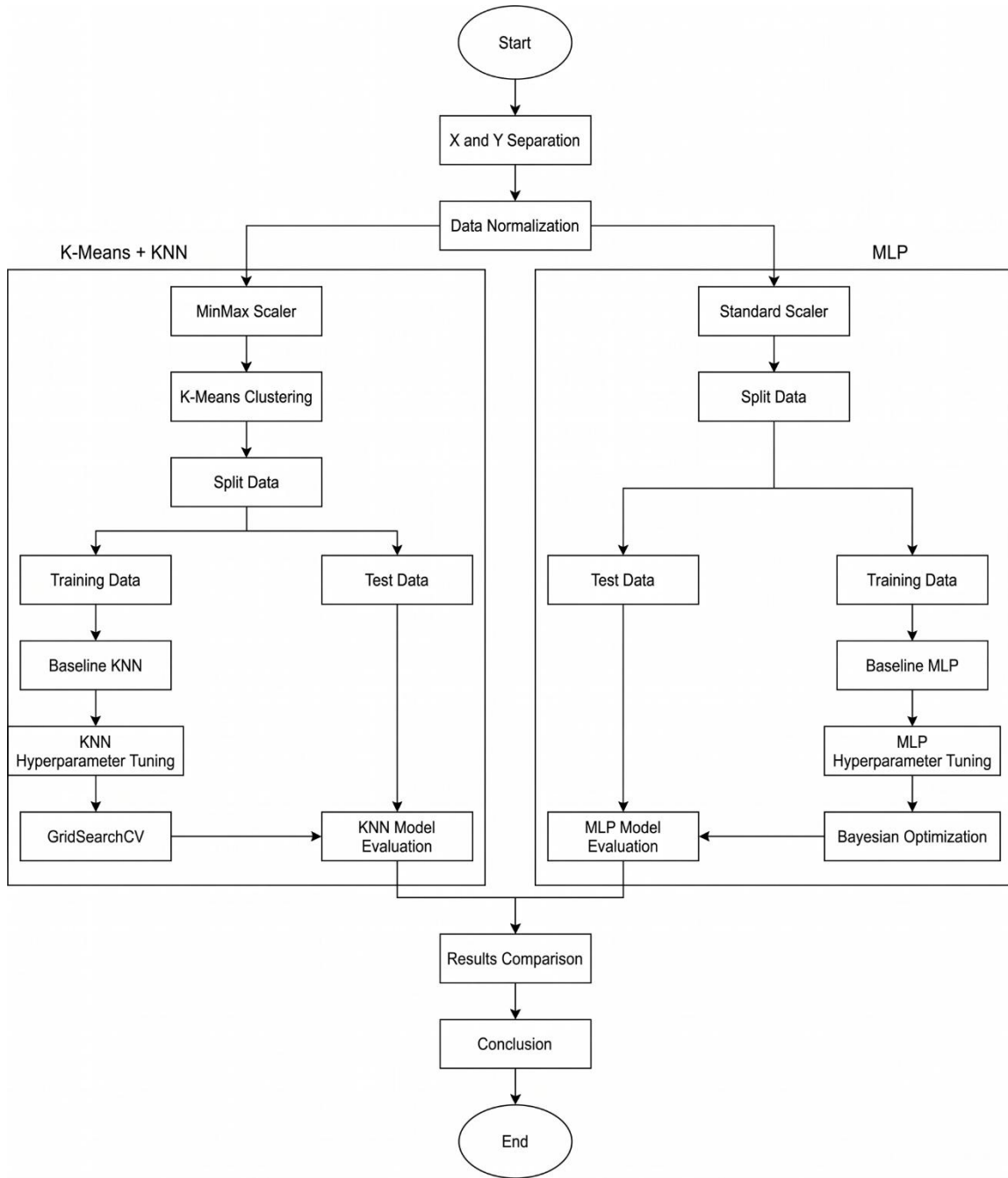
Hyperparameter tuning was performed using GridSearchCV, exploring the parameters  $n$  neighbors and weights. Model evaluation was performed on the test data using accuracy metrics, weighted F1-score, and ROC-AUC to assess classification performance both explicitly and probabilistically.

#### b. Multilayer Perceptron (MLP)

The second approach uses a Multilayer Perceptron (MLP), which begins with a standardization process for numerical data using StandardScaler. After that, the data is divided into training and test data in a stratified 80:20 ratio. The basic MLP architecture consists of six neurons in the input layer (depending on the number of features), one or more hidden layers with varying numbers of neurons, and an output layer of three neurons for each class, using the softmax activation function [11].

The baseline MLP model was built with an initial configuration and evaluated before tuning. To optimize model performance, hyperparameter tuning was performed using the Bayesian Optimization method. The parameters explored included the number of hidden layers, activation function (ReLU, Tanh, and Leaky ReLU), learning rate, batch size, dropout rate, optimizer (Adam and SGD), L2 regularization, and number of epochs. All training was performed on the training data, and performance evaluation was performed on the test data using the same metrics as the previous approach.

An overview of the overall flow of the classification process for both approaches is shown in Figure 1, which includes stages from preprocessing to model evaluation for both the K-Means + KNN and MLP paths.



**Figure 1.** Research Flowchart

### 3. Model Evaluation

The performance evaluation of the classification model was performed using several metrics relevant to multiclass classification, particularly in the context of imbalanced label distributions. The main metrics used include accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix.

Accuracy is used as a basic metric to measure the proportion of correct predictions compared to all predictions made. While accuracy provides a general overview of model performance, this metric tends to be less representative if there is dominance of a particular class. The general formula for calculating accuracy is stated as [12]:

$$Accuracy = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i + FN_i)} \quad (3)$$

where  $TP_i$  is the number of correct predictions for the  $i$ -th class, while  $FP_i$  and  $FN_i$  represent the number of false positive and false negative predictions for that class, respectively.

To describe the model's sensitivity to each class, precision and recall are calculated per class. Precision represents the model's accuracy in classifying a particular class without false positives, while recall measures how many instances of the true class the model successfully recognizes. The respective formulas are stated as [12]:

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (4)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5)$$

Next, the F1-score is calculated as the harmonic mean of the precision and recall for each class [8].

$$F1_i = \frac{2 \times Precision_i \times Recall_i}{Precision_i + Recall_i} \quad (6)$$

These three metrics are precision, recall, and F1-score, averaged using a macro average (for unweighted inter-class assessment) and a weighted average (which takes into account the distribution of data within each class). The weighted F1-score is the primary focus in evaluating explicit classification ability.

As a complement to the numerical evaluation, a confusion matrix is used to illustrate the distribution of model predictions relative to actual labels. This matrix is useful in identifying patterns of prediction errors, such as dominance of the majority class or model failure to recognize the minority class.

Finally, to evaluate the model's performance in terms of probabilistic classification, the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve was used. The approach used was One-vs-Rest, where the ROC-AUC is calculated for each class against the other classes. This metric reflects the model's ability to probabilistically distinguish classes, with

AUC values close to 1 indicating excellent discrimination ability, while values close to 0.5 indicate performance equivalent to random guessing.

### **III. RESULTS AND DISCUSSION**

#### **1. Target Class Distribution**

The distribution of target labels in the dataset shows a significant imbalance between classes. Of the 514 regencies/cities, 455 regions are classified into class 1, representing the region favored by presidential candidate pair number 2. Meanwhile, class 0 encompasses 49 regions, and class 2 comprises only 10 regions. This imbalanced distribution presents a challenge in building an effective classification model. Models tend to be biased toward the majority class, resulting in a high risk of failing to recognize the minority class. This imbalance needs to be addressed in the process of evaluating model performance, particularly through metrics that consider the weighted class distribution, such as the weighted F1-score and ROC-AUC.

#### **2. Feature Clustering Results (K-Means)**

Initial clustering was performed using the K-Means algorithm on each of the six numeric variables in the dataset individually. The goal was to extract latent structure from the socio-economic data across regions and to add categorical features from the clusters as additional information during the classification stage. The optimal number of clusters was determined using the Elbow method, which showed the most stable bending point at  $k = 3$ .

Cluster quality was evaluated using the Silhouette Score for each clustered feature. Silhouette Score values ranged from 0.58 to 0.82, with the highest score being 0.8185 for the Gross Regional Domestic Product (GRDP) variable. These values indicate good and consistent separation between clusters across each feature. The resulting clusters were then represented as new categorical features and inserted into the data before being applied to the K-Nearest Neighbors model.

#### **3. Evaluation of the K-Nearest Neighbors (KNN) Model**

The KNN model was first built as a baseline using training data enriched with features from K-Means clustering. Distances between observations were measured using Gower Distance to accommodate a combination of numerical and categorical features. At this initial stage, the model demonstrated an accuracy of 88.35% and a weighted F1-score of 0.8445. However, further evaluation using a confusion matrix showed that the model only recognized the majority class (class 1) well, while class 0 had an F1-score of 0.17, and class 2 failed to be recognized at all. The average ROC-AUC value obtained at this baseline stage was 0.6296.

To improve model performance, the hyperparameters, number of nearest neighbors, and weighting scheme were tuned using the GridSearchCV approach. The best configuration was obtained at  $k = 8$  with uniform weighting. The tuned model showed an overall performance

improvement, with accuracy reaching 89.32%, the weighted F1-score increasing to 0.8709, and the average ROC-AUC increasing to 0.6362. The F1-score for class 0 increased to 0.40, although class 2 was still not successfully classified.

#### 4. Evaluation of Multilayer Perceptron (MLP) Model

The baseline Multilayer Perceptron (MLP) model was built with a single hidden layer of 64 neurons using ReLU activation and a Softmax output layer. The input data was normalized using z-score standardization before training. The evaluation results showed an accuracy of 87.38% and a weighted F1-score of 0.8240. The model recognized the majority class (class 1) well, achieving an F1-score of 0.93, while classes 0 and 2 were not detected at all. The average ROC-AUC value obtained at this stage was 0.6534.

After hyperparameter tuning using Bayesian Optimization, an optimal configuration with five hidden layers, each containing 128 neurons and using the tanh activation function, was obtained. Model complexity regulation was applied through a dropout of 0.3796 and L2 regularization of 0.0866 to reduce the risk of overfitting. The model was optimized using the Adam algorithm with a learning rate of 0.0375. Evaluation of the tuned model showed an increase in the F1-score for class 0 to 0.27, although class 2 was still not detected. The model accuracy was recorded at 86.41%, the weighted F1-score increased to 0.8436, and the ROC-AUC reached 0.7420, the highest value among all models tested.

#### 5. Summary of Comparison Results

Table 1 presents a summary of the performance of each tested classification model, both before and after tuning. The K-Nearest Neighbors (KNN) model demonstrated the best accuracy (0.8932) and weighted F1-score (0.8709) after tuning. Meanwhile, the optimized Multilayer Perceptron (MLP) model performed best in terms of probabilistic capability, reflected by the highest ROC-AUC value of 0.7420.

The performance differences between the models highlight the strengths of each approach, both in explicit classification and probabilistic estimation. These evaluation results serve as the basis for further discussion of the relative advantages and limitations of the models in the following sections.

**Tabel 1.** Summary of Model Evaluation Results

Metric	KNN Baseline	KNN Tuning	MPL Baseline	MLP Tuning
Accuracy	0.8835	<b>0.8932</b>	0.8738	0.8641
Precision	0.8358	<b>0.8606</b>	0.7796	0.8312
Recall	0.8835	<b>0.8932</b>	0.8738	0.8641
F1-Score	0.8445	<b>0.8709</b>	0.8240	0.8436
ROC-AUC	0.6296	0.6362	0.6534	<b>0.7420</b>

The results of the study show that the K-Nearest Neighbors (KNN) model, with additional features resulting from K-Means clustering and the use of Gower Distance, excels in explicit classification of socio-economic data from the 2024 election results. After hyperparameter tuning, this model recorded the highest accuracy of 89.32% and a weighted F1-score of 0.8709. This performance reflects that distance-based methods such as KNN, when combined with categorical feature representation from clustering and adaptive distance measurement, are able to provide more stable and accurate classification on mixed data structures. The improved performance against minority classes, especially class 0, also indicates that this approach is relatively robust in handling class inequality.

In contrast, the Multilayer Perceptron (MLP) model demonstrated superiority in probabilistic estimation, with the highest ROC-AUC of 0.7420. Although the accuracy and weighted F1-score of the MLP were still below those of the KNN model, the ROC-AUC values indicated that the MLP was more sensitive in distinguishing classes at the probability level. However, the MLP's weakness remained in explicitly recognizing minority classes, where the F1 score for class 2 remained zero and only slightly improved for class 0 after tuning. This indicates that the complexity of the network architecture is not sufficient to offset the sensitivity to skewed class distributions, particularly in the context of actual election data.

These findings offer a different perspective compared to previous research by [8], which demonstrated that deep learning approaches performed better than machine learning in the context of predicting legislative election results. In his study, the deep learning neural network model achieved 98.50% accuracy and an AUC of 0.982. However, in the context of this research, which used actual 2024 election data with a mixed structure and extreme class imbalance, the K-Means and KNN-based machine learning approaches demonstrated more stable performance in explicit classification. This difference can be explained by differences in dataset characteristics, a much more imbalanced class distribution, the feature representation approach used, and the classification objective, which places greater emphasis on predictive stability between classes.

The high-class distribution imbalance in the dataset, with class 1 dominating, significantly contributes to the challenges faced by all models in recognizing the minority class. The F1-score for class 2 is zero in all models, and although there is an improvement in class 0 after tuning, it remains low, at 0.40 for KNN and 0.27 for MLP. This imbalance causes the models to be biased toward the majority class, and without special balancing techniques, performance against the minority class is difficult to significantly improve. However, it is important to note that oversampling and undersampling techniques were not applied in this research because the data used are a direct representation of actual election results, making proportional evaluation approaches such as weighted F1-score and ROC-AUC more appropriate.

#### IV. CONCLUSION

This research compares the performance of the K-Means + K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) methods in classifying vote counts for the 2024 election based on socio-economic variables. The evaluation results show that the KNN-based approach with additional clustering features and the use of Gower Distance is superior in explicit classification, as indicated by the highest accuracy and weighted F1 score. Meanwhile, the MLP performs better in probabilistic estimation, reflected by higher ROC-AUC values. However, the imbalance in class distribution remains a major challenge for both approaches. This is evident in the model's low performance on minority classes, despite hyperparameter tuning. These findings emphasize the importance of selecting appropriate classification approaches and evaluation metrics in the context of complex and imbalanced socio-political data.

For future development, similar research could consider generating labels based on more granular voter preference patterns, as well as exploring other model architectures such as 1D CNNs or attention-based approaches. Adding contextual features such as demographics, religion, and regional infrastructure also has the potential to improve the model's predictive power. Furthermore, a more adaptive classification approach to class inequality and evaluation of different election datasets could be development directions to improve the model's generalizability in the future.

#### V. AUTHOR CONTRIBUTION

The authors contributed equally to the formulation of the research concept, data collection, data analysis, manuscript writing, and approval of the final manuscript for publication.

#### VI. CONFLICT OF INTEREST

The authors declare that there are no potential conflicts of interest related to the research, writing, or publication of this article.

#### REFERENCES

- [1] D. K. Kirana, M. O. E. Setiawan, and S. Priza, "Demokrasi Indonesia dalam kapasitas pemilu yang luber jurdil," *Journal of Law and Social Society*, vol. 1, no. 1, 2024, doi: 10.70656/jolasos.v1i1.80.
- [2] S. Rahmawati and H. Halking, "Pengaruh tingkat status sosial ekonomi terhadap pemahaman pemilu pemilih pemula pada pilpres 2024 di desa suka mulia kecamatan secanggih kabupaten Langkat," *Madani Jurnal Politik dan Sosial Kemasyarakatan*, vol. 16, no. 02, 2024, doi: 10.52166/madani.v16i02.7385.
- [3] D. Zhao *et al.*, "K-means clustering and kNN classification based on negative databases," *Appl. Soft Comput.*, vol. 110, 2021, doi: 10.1016/j.asoc.2021.107732.

- [4] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5–6, 1991, doi: 10.1016/0925-2312(91)90023-5.
- [5] R. K. Jain and G. K. Singh, "Clustering-based hybrid approach for multiclass classification using SVM," in *Advances in Intelligent Systems and Computing*, 2020. doi: 10.1007/978-981-32-9515-5\_35.
- [6] S. Gupta, B. Parekh, and A. Jivani, "A hybrid model of clustering and classification to enhance the performance of a classifier," in *Communications in Computer and Information Science*, 2019. doi: 10.1007/978-981-15-0111-1\_34.
- [7] A. Alarbi and Z. Albayrak, "Core classifier algorithm: A hybrid classification algorithm based on class core and clustering," *Applied Sciences (Switzerland)*, vol. 12, no. 7, 2022, doi: 10.3390/app12073524.
- [8] M. Badrul, "Perbandingan algoritma C4.5 dan neural network untuk memprediksi hasil pemilu legislatif dki jakarta," *Jurnal Pilar Nusa Mandiri*, vol. 10, no. 2, pp. 127–138, 2014.
- [9] T. Jo, "K Means Algorithm," in *Machine Learning Foundations*, Cham: Springer International Publishing, 2021, pp. 217–240. doi: 10.1007/978-3-030-65900-4\_10.
- [10] G. Tuerhong and S. B. Kim, "Gower distance-based multivariate control charts for a mixture of continuous and categorical variables," *Expert Syst. Appl.*, vol. 41, no. 4 PART 2, 2014, doi: 10.1016/j.eswa.2013.08.068.
- [11] K. B. Prakash, "Multilayer Perceptron (in Neural Networks)," in *Machine Learning for Industrial Applications*, Wiley, 2025, pp. 51–61. doi: 10.1002/9781394268993.ch3.
- [12] S. Adiningsi, B. Pramono, A. Mashur Sajjah, and R. Adi Saputra, "Identifikasi kualitas ikan cakalang segar berbasis citra mata menggunakan metode Support Vector Machine (SVM) dengan fungsi kernel radial basis function," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 1, 2025, doi: 10.36040/jati.v9i1.12641.