DECODE: JURNAL PENDIDIKAN TEKNOLOGI INFORMASI
P-ISSN: 2775-2984 | E-ISSN: 2775-1813

Vol. 5, No 3, November 2025, Hal: 1044-1055
Doi: http://dx.doi.org/10.51454/decode.v5i3.1342
https://journal.umkendari.ac.id/index.php/decode

Development of a Hybrid Voting Model with SMOTE and Random Search for Classification of Religious Facility Grant Recipients

Muhamad Jamaris[1], Rita Novita[2]

[1]Informatics Engineering, Universitas Sains dan Teknologi Indonesia, Indonesia.
[2]Health Informatics, Institut Kesehatan Payung Negeri Pekanbaru, Indonesia.

Article Info

Keywords:

Hybrid Voting;
Machine Learning;
Random Search;
Religious Facility Grand;
SMOTE.

Article History:
Submitted: August 11, 2025
Accepted: October 27, 2025
Published: November 05, 2025

Corresponding Author:
Muhamad Jamaris
Email: jamaris@usti.ac.id

Abstract: The process of determining recipients of religious facility grants requires high accuracy to ensure that aid is distributed fairly and supports equitable community services. Manual selection methods often face challenges such as data imbalance, diverse assessment criteria, and subjective decision-making, which can reduce accuracy and efficiency. This study proposes a hybrid machine learning model using Voting Ensemble (Hard and Soft), combining Logistic Regression (LR), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), optimized with Random Search and supported by SMOTE to handle class imbalance. The dataset consists of religious facility grant applications in Riau Province, with preprocessing, SMOTE balancing, and Stratified K-Fold Cross Validation applied for robust evaluation. The experimental results show that the Hybrid Voting model outperforms single models, achieving an average accuracy of 99.46%, with precision, recall, and F1-score consistently above 96%, and some folds achieving 100% accuracy. These findings demonstrate that the hybrid approach enhances prediction stability, reduces misclassification of minority classes, and provides a decision-support system that is objective, accurate, and efficient for grant recipient selection.

INTRODUCTION

The process of determining grant recipients for worship facility assistance is a crucial activity that requires high accuracy to ensure that the aid is distributed to the right targets and supports equitable services for the community (Raisyah et al., 2024). Beyond its technical importance, this process also carries strong implications for policy making and social fairness, as transparent and data driven grant allocation can help reduce public distrust and promote inclusivity in government assistance programs. However, the selection process, which is still conducted manually, often faces various challenges, such as data imbalance, diverse assessment criteria, and decision-maker subjectivity. These conditions can reduce the accuracy, efficiency, and transparency of grant distribution.

Along with technological advancements, the application of machine learning (ML) offers a solution to support the classification process of grant recipients in an automated, objective, and efficient manner (Anam et al., 2025). Previous studies have demonstrated the effectiveness of ML algorithms in similar classification cases. The study in (Qadrini et al., 2022) applied Random Forest for the selection of Bidikmisi scholarship recipients in East Java and achieved good accuracy, but its performance decreased when handling imbalanced data. Meanwhile, (Aprihartha, 2024) used Support Vector Machine (SVM) for classifying non-cash food assistance recipients, showing high prediction accuracy, although the model required complex parameter optimization. Another study (Ramadani et al., 2024) compared K-Nearest Neighbor (KNN), C4.5, and Naive Bayes in classifying the eligibility of the Hope Family Program, where C4.5 outperformed the others, but its performance on large-scale datasets remained unknown. Furthermore, (Agustina & Ihsan, 2023) implemented Voting Ensemble for sentiment analysis prediction, which improved accuracy stability; however, the study did not apply SMOTE to effectively address class imbalance.

Based on these previous studies, several research gaps can be identified. First, most studies still use single models, which tend to have unstable performance on heterogeneous data. Second, the class imbalance issue has not been addressed optimally, despite the fact that the distribution of grant recipients is typically uneven among accepted, rejected, and in-process categories. Third, hyperparameter optimization in ensemble methods is rarely performed, although it is crucial to improving model performance. In addition, hybrid voting approaches combining multiple base algorithms with SMOTE and K-Fold Validation have rarely been applied to the classification of worship facility grant recipients.

This study proposes a novel approach to address these gaps through the development of a hybrid machine learning model using a voting technique that combines Logistic Regression (LR), SVM, and KNN in both soft voting and hard voting schemes, optimized using Random Search. Logistic Regression was selected for its simplicity, speed, and effectiveness in multiclass classification of tabular data (VanFC et al., 2025). SVM was chosen for its ability to generate optimal hyperplanes to accurately separate classes in non-linear data (Anam et al., 2025), while KNN leverages local data proximity to detect patterns in minority classes (Dang & Le, 2024). The application of SMOTE (Synthetic Minority Over-sampling Technique) helps balance class distribution, preventing the model from being biased toward the majority class (Putra et al., 2025). Meanwhile, Random Search is employed to efficiently find the best hyperparameter combinations, enhancing the model’s accuracy and stability (Abubakar et al., 2023). With this combination, the classification system is expected to produce more accurate, stable, and fair predictions, thereby supporting the determination of worship facility grant recipients in a transparent, efficient, and data-driven manner, ultimately contributing to evidence based policy decisions and equitable resource distribution.

METHOD

This research holds significant importance in supporting the process of determining worship facility grant recipients in a more objective, fast, and accurate manner. To date, the selection process has often faced challenges such as data imbalance, subjectivity in decision-making, and the limitations of manual methods, which are prone to misclassification. Through the application of machine learning with a hybrid voting ensemble approach, this study is expected to improve accuracy, stability, and efficiency in the classification process of grant recipients. In addition, the use of SMOTE helps balance the data so that minority classes such as PROCESS and REJECT are not neglected, while Random Search ensures the model operates optimally.

The methodological design of this study also explicitly includes evaluation using five key performance metrics, namely accuracy, precision, recall, F1 score, and confusion matrix, to comprehensively assess model performance from various perspectives. These metrics are integrated into the evaluation process to ensure that the proposed model not only achieves high predictive accuracy but also maintains balance and fairness across all classification categories.

The following figure presents the research methodology flow, covering the stages from literature review, dataset collection and preprocessing, model validation, the development of base models (Logistic regression, SVM, and KNN), hybrid voting model construction (Soft and Hard), hyperparameter optimization, to comprehensive model evaluation aimed at producing a superior classification system.

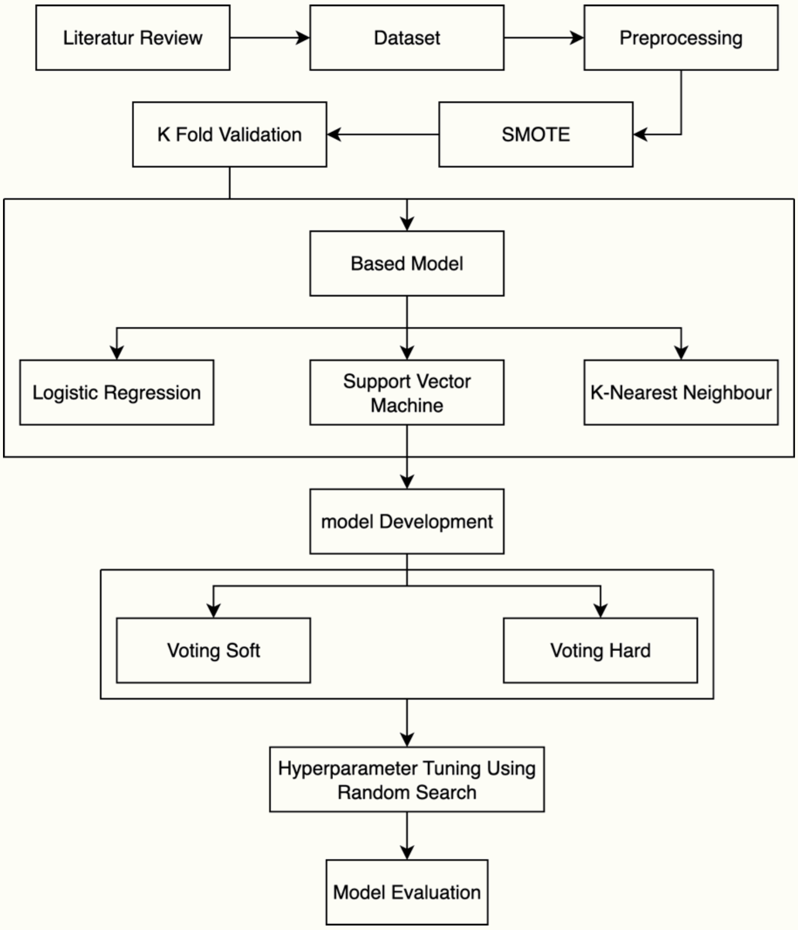

Figure 1. Research Methodology Flow

The first stage of this research is the literature review, aimed at identifying the most relevant methods, techniques, and strategies in classifying grant recipients using machine learning. This process begins by examining various previous studies related to the application of Logistic regression (LR), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and ensemble learning methods in social assistance or grant classification cases. References include (Qadrini et al., 2022), which used Random Forest for social assistance selection with high accuracy but unstable performance on imbalanced data; (Aprihartha, 2024), which implemented SVM for classifying educational assistance recipients, though requiring complex hyperparameter optimization; and (Ramadani et al., 2024), which compared KNN, C4.5, and Naive Bayes for classifying the eligibility of the Hope Family Program, finding C4.5 superior. Study (Agustina & Ihsan, 2023) improved accuracy stability through Voting Ensemble, although it did not adequately address data imbalance.

In addition, (Susrifalah et al., 2025) applied Gradient Boosting in Hope Family Program selection but faced overfitting risks without cross-validation, while (Handayani & Qutub, 2025) used bagging on Random Forest for poverty prediction but struggled with skewed class distribution. Study (Syamsiah & Purwandani, 2023) employed stacking with NN, RF, SVM, and GLM for electricity consumption prediction, yielding better accuracy than single models, while (Ridwan et al., 2024) enhanced model performance by applying SMOTE on imbalanced hate speech data. Furthermore, (Shah et al., 2023) developed a hybrid ensemble for location-based social assistance prediction with more stable results, and (Reza & Rohman, 2024) emphasized the importance of hyperparameter tuning using Random Search to increase accuracy consistency in ensemble models.

From these studies, it can be concluded that although machine learning methods have improved classification accuracy for social assistance or grant recipients, challenges remain regarding data imbalance, the limitations of single models, and the lack of hyperparameter optimization in ensemble models. Therefore, this study employs a hybrid voting approach combining LR, SVM, and KNN, supported by SMOTE and K-Fold Validation to produce a more stable, accurate, and efficient classification model.

The second stage is dataset collection, sourced from worship facility grant applications in Riau Province. The dataset contains variables related to the type of worship facility (e.g., mosque, church, vihara, pura, and klenteng), location, grant application status, and eligibility indicators such as the physical condition of the building and administrative legality. This data forms the basis for the grant recipient classification process.

The third stage is preprocessing, which includes data cleaning to remove missing or duplicate values, normalization of numerical data, and encoding of categorical variables for processing by machine learning algorithms. This stage also includes handling data imbalance using SMOTE (Synthetic Minority Over-sampling Technique) to ensure balanced class distribution. Figure 2 shows the initial label distribution.

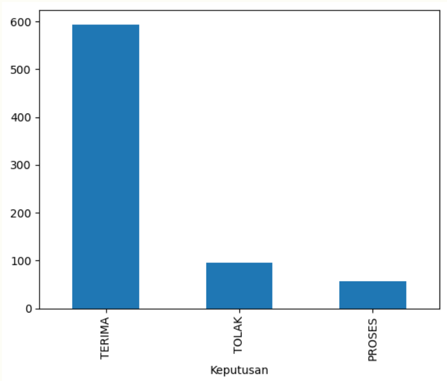

Figure 2. Label Distribution Before Data Balancing

Figure 2 shows the initial label distribution of the total 750 worship facility grant records collected from the regional grant management system. The label distribution before applying SMOTE reveals significant data imbalance. From this dataset, the ACCEPT category dominates with approximately 600 entries (80%), while the REJECT category contains around 100 entries (13%), and the PROCESS category is the smallest with about 50 to 60 entries (7%). This condition indicates that most worship facility grant applications are approved, with only a small portion rejected and an even smaller portion still in the verification process. This imbalance risks biasing the machine learning model toward the majority class (ACCEPT). Therefore, before building the classification model, balancing techniques such as SMOTE are necessary so the model can fairly learn from all classes and improve prediction accuracy for minority classes. Figure 3 presents the label distribution after balancing.

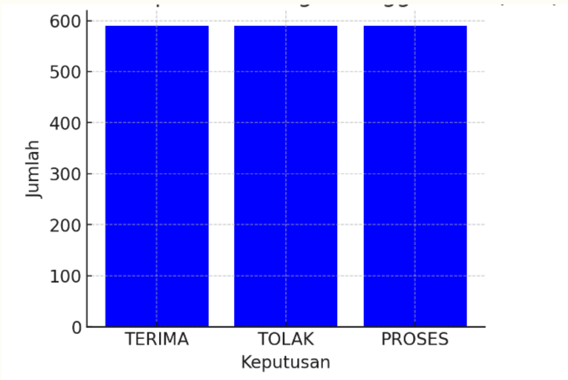

Figure 3. Label Distribution After Data Balancing

Figure 3 is a bar chart showing the label distribution after applying SMOTE. The total dataset after balancing increases to approximately 1800 records, with each of the three decision categories, ACCEPT, REJECT, and PROCESS, now having about 600 entries each. This change occurs because SMOTE generates synthetic samples for the minority classes (REJECT and PROCESS) until they match the majority class (ACCEPT). The goal is to address the previous imbalance so that the model is not biased toward the majority class and performs better in recognizing minority classes. Balanced distribution is crucial in this case to ensure fair learning and improve metrics such as accuracy, recall, and F1 score, especially for classes that were previously underrepresented.

The fourth stage is building base models consisting of three algorithms: Logistic Regression (LR), Support Vector Machine (SVM), and K Nearest Neighbour (KNN). These algorithms were selected because their characteristics complement each other in handling tabular data with varying levels of complexity and data distribution. Logistic Regression is effective for binary and multiclass classification where relationships between predictors and class labels are relatively linear, providing both interpretability and computational efficiency (Herianto et al., 2024). SVM performs well in identifying optimal decision boundaries, even when the data pattern is complex or non-linear, ensuring high generalization ability (Anam et al., 2025). KNN, as an instance based learning algorithm, captures local data structures by considering proximity among instances, which helps detect subtle variations within minority or less represented data (Anam et al., 2021).

This combination was chosen to balance interpretability, robustness, and adaptability. Logistic Regression contributes a strong analytical foundation and explainability, SVM enhances precision through margin optimization, and KNN increases flexibility in recognizing local patterns. Together, they form a complementary hybrid configuration expected to produce a stable, accurate, and efficient classification system capable of handling diverse data characteristics in worship facility grant distribution.

The fifth stage is developing a hybrid model with voting ensemble techniques. Two voting approaches are used: Soft Voting, which considers the predicted probabilities of each model, and Hard Voting, which relies on the majority vote (Rahayu et al., 2024). This combination is expected to enhance classification stability and accuracy compared to single models. The sixth stage is hyperparameter tuning using Random Search. This method enables efficient random exploration of parameter combinations to improve model accuracy and reduce overfitting risks (Aribowo et al., 2024).

The final stage is model evaluation using accuracy, precision, recall, F1-score, and confusion matrix. Evaluation is conducted with K-Fold Cross Validation (5 Fold) to ensure stable performance across all data subsets. The results will highlight the superiority of the hybrid voting method over single models, demonstrating the effectiveness of this approach in supporting an objective, transparent, and efficient grant recipient selection process in Riau Province.

RESULTS AND DISCUSSIONS

This section presents the test results of the machine learning models developed for classifying recipients of worship facility grants in Riau Province, along with their performance analysis. All models were evaluated using Stratified K-Fold Cross Validation to ensure balanced class distribution in each fold, with SMOTE applied to address data imbalance. The single models tested included Logistic Regression, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM), while the Hybrid Voting models consisted of Hard Voting and Soft Voting, combining the three base models with Random Search optimization. The results for each fold are presented in tables containing evaluation metrics such as accuracy, precision, recall, and F1-score, along with confusion matrices to show the prediction distribution for each class. The performances of these models were then compared to assess the improvement from single models to hybrid models and to identify the best model most suitable for supporting the selection process of grant recipients quickly, accurately, and transparently. Figure 4 presents the test results using Logistic Regression.

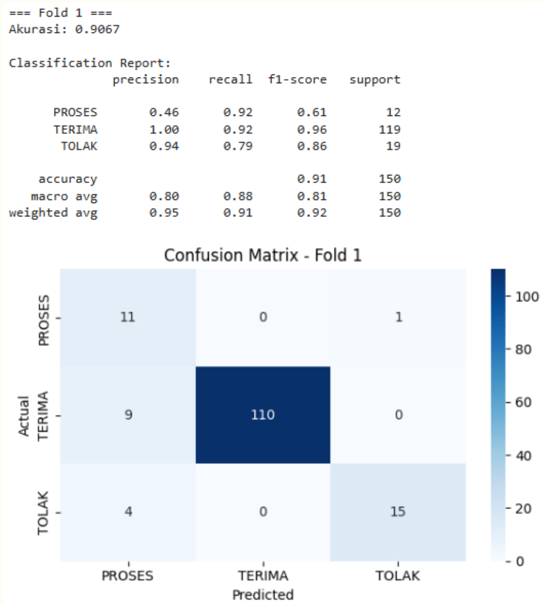

Figure 4. Test results of the Logistic Regression model

The test results of the Logistic Regression model combined with SMOTE and validated using Stratified K-Fold showed fairly good classification performance. In Fold 1, the model achieved an accuracy of 90.67%, indicating that the majority of test data was classified correctly. Based on the classification report, the PROSES class had a precision of 0.46 and recall of 0.92, indicating that most PROSES data was successfully detected, although some misclassifications into other classes occurred. The TERIMA class showed the best performance with a precision of 1.00 and recall of 0.92, meaning the model was able to recognize the majority class very well. Meanwhile, the TOLAK class had a precision of 0.94 and recall of 0.79, showing relatively stable predictions despite some data being misclassified. This misclassification pattern is visible in the confusion matrix, where some PROSES and TOLAK data were predicted as TERIMA. Table 1 presents the overall results using K-Fold (K=5).

Table 1. Logistic regression model test results using all folds

FOLD	ACCURACY	PRECISION	RECALL	F1-SCORE
FOLD 1	90.67%	80%	88%	81%
FOLD 2	90.60%	81%	90%	82%
FOLD 3	93.29%	84%	96%	87%
FOLD 4	87.25%	77%	90%	79%
FOLD 5	93.29%	81%	90%	85%

Overall, the evaluation of the model over five folds shows consistent performance, with accuracy ranging from 87.25% to 93.29%. The model's average precision was between 80%–84%, recall between 88%–96%, and F1-score between 81%–87%. The high recall values, especially for the TERIMA class, indicate that the model is sensitive in detecting approved grant recipients, which is the majority class. Meanwhile, variations in precision and recall for minority classes (PROSES and TOLAK) show that using SMOTE helped balance performance, although some misclassifications still occurred in classes with fewer data. With this performance, the Logistic Regression model proved effective for classifying worship facility grant recipients with imbalanced class distributions after SMOTE and cross-validation. Table 2 presents the results using KNN and SVM.

Table 2. KNN and SVM Model Test Results using All Folds

ALGORITHM	FOLD	ACCURACY	PRECISION	RECALL	F1-SCORE
KNN	FOLD 1	96.00%	87%	90%	89%
	FOLD 2	98.66%	95%	98%	96%
	FOLD 3	96.64%	89%	96%	92%
	FOLD 4	92.62%	81%	90%	83%
	FOLD 5	97.99%	93%	94%	93%
SVM	FOLD 1	94.67%	85%	92%	88%
	FOLD 2	96.64%	90%	97%	92%
	FOLD 3	95.30%	87%	95%	89%
	FOLD 4	91.28%	82%	93%	84%
	FOLD 5	97.32%	91%	95%	93%

Table 2 presents the test results of the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) models using Stratified K-Fold Cross Validation with five folds. Each fold shows evaluation metrics of accuracy, precision, recall, and F1-score. Briefly, KNN demonstrated excellent performance with accuracy ranging from 92.62% to 98.66%, with average precision and recall between 87%–95%, and F1-score between 83%–96%. SVM also showed stable performance with accuracy from 91.28% to 97.32%, precision from 82%–91%, recall from 92%–97%, and F1-score from 84%–93%. From these results, it can be concluded that both algorithms performed highly, with KNN slightly outperforming in some folds, while SVM offered consistent performance, especially in recall, indicating strong class recognition capability. Figure 5 presents the results using Hard Voting with Random Search.

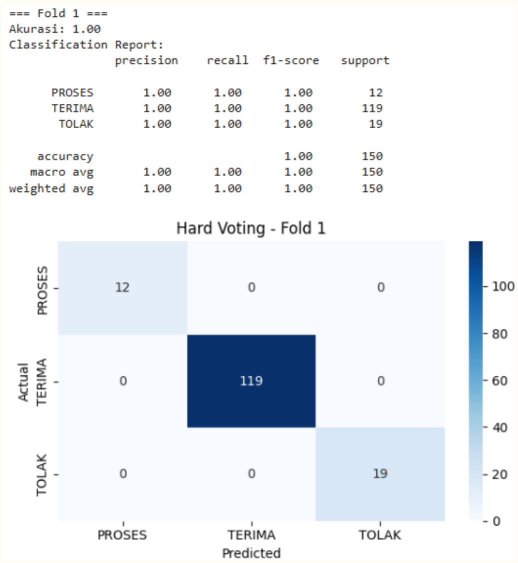

Figure 5. Test results of the Hard Voting model

The test results of the Hybrid Hard Voting model, which combines LR, SVM, and KNN with SMOTE support and hyperparameter optimization via Random Search, showed outstanding classification performance. In Fold 1, this model achieved 100% accuracy, with precision, recall, and F1-score all reaching 1.00 for all classes. The confusion matrix in this fold shows that all test data—PROSES, TERIMA, and TOLAK—were correctly classified with no errors. This confirms that the hybrid model can combine the strengths of the three base algorithms to produce highly precise predictions, especially after class distribution was balanced with SMOTE. Table 3 presents the overall results using K-Fold (K=5).

Table 3. Hard Voting Model Test Results Using All Folds

FOLD	ACCURACY	PRECISION	RECALL	F1-SCORE
FOLD 1	100%	100%	100%	100%
FOLD 2	99.33%	97%	98%	98%
FOLD 3	100%	100%	100%	100%
FOLD 4	98.66%	95%	98%	96%
FOLD 5	99.33%	98%	97%	98%

Evaluation across the five folds shows highly consistent performance, with accuracy ranging from 98.66% to 100%. Average precision ranged between 95%–100%, recall between 97%–100%, and F1-score consistently above 96%. The application of Random Search for hyperparameter tuning effectively optimized the parameter combinations of the three base models, enhancing model stability across folds. Compared to single models such as Logistic Regression, SVM, and KNN, these results confirm that Hybrid Hard Voting delivers superior performance, with high sensitivity across all classes and minimal misclassification. With this performance, the hybrid voting model is highly suitable as a decision support system for classifying worship facility grant recipients in Riau Province objectively, accurately, and efficiently. Figure 6 presents the results using Soft Voting with Random Search.

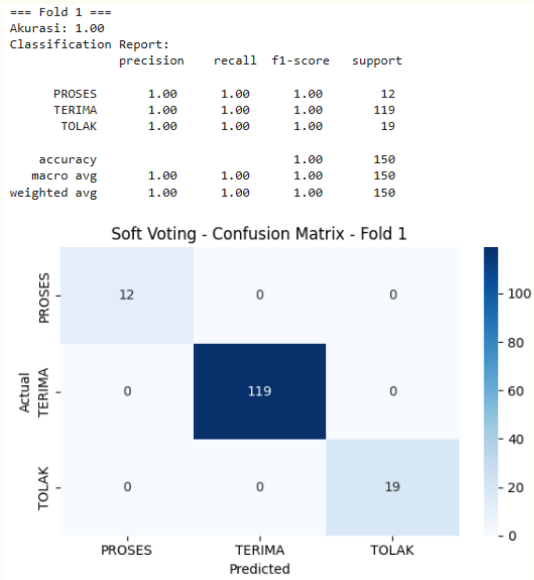

Figure 6. Test results of the Soft Voting model

The test results of the Hybrid Soft Voting model, which combines LR, SVM, and KNN with SMOTE support and hyperparameter tuning via Random Search, showed highly optimal classification performance. In Fold 1, the model achieved 100% accuracy, with precision, recall, and F1-score all at 1.00 for all classes. The confusion matrix indicated no misclassifications; all PROSES, TERIMA, and TOLAK class data were correctly classified. This confirms that the soft voting approach can leverage the strength of probability aggregation from the base models, producing accurate and stable predictions across all classes, including minority classes that were previously prone to misclassification. Table 4 presents the overall results using K-Fold (K=5).

Table 4. Soft Voting Model Test Results using All Folds

FOLD	ACCURACY	PRECISION	RECALL	F1-SCORE
FOLD 1	100%	100%	100%	100%
FOLD 2	99.33%	97%	98%	98%
FOLD 3	100%	100%	100%	100%
FOLD 4	98.66%	95%	98%	96%
FOLD 5	99.33%	98%	97%	98%

Across all five folds, the model's performance remained highly consistent, with accuracy ranging from 98.66%–100%. Average precision ranged between 95%–100%, recall between 97%–100%, and F1-score consistently above 96%. The use of SMOTE effectively balanced the data distribution, improving model sensitivity to minority classes, while Random Search helped find the best hyperparameter combinations to maximize ensemble voting performance. Overall, Hybrid Soft Voting delivered performance comparable to Hybrid Hard Voting, and both showed significant improvements over single models such as Logistic Regression, SVM, and KNN. With this performance, Hybrid Soft Voting is highly suitable as a decision support system for classifying worship facility grant recipients in Riau Province, as it can provide accurate, stable, and minimally erroneous results.

The test results reveal significant differences between single and hybrid models in classifying worship facility grant recipients in Riau Province. For single models, Logistic Regression achieved an average accuracy of 91.62%, with the main weakness being in the precision of minority classes (PROSES and TOLAK), although sensitivity to the majority TERIMA class was quite high. Performance improvements were observed in K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), which achieved average accuracies of 96.30% and 95.44%, respectively. KNN showed better sensitivity to minority classes due to its distance-based nature, while SVM demonstrated high stability with non-linear data through an optimal separating margin. SMOTE proved effective in improving the F1-score of single models as data distribution became more balanced, making predictions for minority classes more accurate.

Model performance increased significantly when using the Hybrid Voting approach. Both Hard Voting and Soft Voting, which combine LR, SVM, and KNN with SMOTE support and hyperparameter optimization via Random Search, achieved an average accuracy of 99.46%, with some folds reaching 100% with no misclassifications. Hard Voting relies on majority votes from the base models, while Soft Voting uses probability aggregation, but both showed almost identical results. The confusion matrix in the first fold showed all data from all three classes were correctly classified, indicating that the model combination could leverage the strengths of each base algorithm to produce stable, accurate, and consistent predictions.

Overall, these results confirm that the hybrid voting model is the best approach in this study. The combination of SMOTE for data balancing and Random Search for hyperparameter optimization played an important role in enhancing model performance. With accuracy nearing 100% and virtually no misclassification, this model is highly suitable for use as a decision support system for selecting worship facility grant recipients in Riau Province, which demands speed, transparency, and minimal errors.

The near perfect performance achieved by the hybrid voting models can be explained by the complementary characteristics of the three base algorithms used. Logistic Regression provides strong generalization ability in handling linear relationships and produces a stable decision boundary across well separated data. Support Vector Machine improves the model's capability to manage complex and nonlinear class boundaries through kernel based optimization, enhancing separation between overlapping data points. Meanwhile, K Nearest Neighbor adds local adaptability by identifying subtle variations in neighboring data, which increases prediction accuracy for minority classes.

By combining these three algorithms in both Hard and Soft Voting schemes, the ensemble effectively integrates their respective strengths, including global generalization from Logistic Regression, margin based precision from SVM, and local sensitivity from KNN. This balanced interaction minimizes the weaknesses of individual models, reduces classification bias caused by imbalanced data, and enhances robustness through collective decision making. The use of SMOTE and Random Search further improves class balance and parameter optimization, allowing the ensemble to achieve consistently high accuracy and minimal misclassification across all folds.

In practical terms, the findings of this study can be directly applied by local governments to enhance the transparency and efficiency of grant allocation systems. The hybrid voting model can be integrated into existing electronic grant management systems to automate the evaluation process of proposals, reducing human bias and subjectivity in decision making. By using this model, local government agencies can prioritize data driven policy decisions, ensure fair distribution of assistance across regions, and improve accountability through traceable and consistent evaluation outcomes. This integration will also support digital governance initiatives in public service delivery by promoting objectivity, efficiency, and fairness in aid distribution.

However, despite its excellent performance, this study has several limitations that need to be acknowledged. First, the very high accuracy values, which reach almost 100%, may indicate a potential risk of overfitting, especially since the dataset used is limited to worship facility grant data from Riau Province. This condition suggests that the model may have learned specific patterns unique to this dataset, which could affect its generalizability when applied to data from other provinces or regions with different characteristics. Second, the model's predictive strength relies heavily on data quality and feature representation, meaning that variations in data collection or differences in assessment criteria may reduce performance consistency. Therefore, future research is recommended to validate and refine this model using more diverse datasets from multiple regions, employ additional regularization techniques, or incorporate external validation to ensure broader applicability and robustness in real-world grant selection contexts.

CONCLUSIONS

This study successfully developed a hybrid machine learning model based on the Voting Ensemble method (Hard and Soft) that combines LR, Support Vector Machine (SVM), and K Nearest Neighbor (KNN) with the support of SMOTE and Random Search optimization to classify recipients of worship house grants in Riau Province. The results show that the hybrid voting models achieved the highest performance, with an average accuracy of 99.46 percent and consistent precision, recall, and F1 score values above 96 percent. Compared to single models, the hybrid model proved more stable and accurate, making it suitable as a decision support system for a fast, transparent, and efficient grant selection process. Nevertheless, the dataset used in this study was limited to Riau Province, and the model was developed using only three base algorithms, without comparison to more advanced ensemble or deep learning methods. These factors may limit the model's generalizability to other regions. Future studies are encouraged to expand the dataset to a cross-provincial scale, explore deep learning or advanced ensemble techniques, and integrate the model into a web-based decision support system to enable real-time, data-driven decision-making for grant management.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Education, Culture, Research, and Technology through the Directorate General of Higher Education, Research, and Technology for the funding support in carrying out this research. This study is part of an activity funded under a derivative research contract from LLDIKTI Region 17 with the University of Science and Technology of Indonesia, Number: 015/LL17/DT.05.00/PL/2025.

REFERENCES

Abubakar, M. A., Muliadi, M., Farmadi, A., Herteno, R., & Ramadhani, R. (2023). Random Forest Dengan Random Search Terhadap Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung. *Jurnal Informatika*, *10*(1), 13–18. https://doi.org/10.31294/inf.v10i1.14531

Agustina, N., & Ihsan, C. N. (2023). Pendekatan Ensemble Untuk Analisis Sentimen Covid19 Menggunakan Pengklasifikasi Soft Voting. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *10*(2), 263–270. https://doi.org/10.25126/jtiik.20231026215

Anam, M. K., Lestari, T. P., Yenni, H., Nasution, T., & Firdaus, M. B. (2025). Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *19*(2), 159–167. https://doi.org/10.37936/ecti-cit.2025192.257815

Anam, M. K., Putra, P. P., Malik, R. A., Karfindo, K., Putra, T. A., Elva, Y., Mahessya, R. A., Firdaus, M. B., Ikhsan, & Gunawan, C. R. (2025). Enhancing the Performance of Machine Learning Algorithm for Intent Sentiment Analysis on Village Fund Topic. *Journal of Applied Data Sciences*, *6*(2), 1102–1115. https://doi.org/10.47738/jads.v6i2.637

Aprihartha, A. (2024). Implementasi Metode Support Vector Machine (SVM) pada Klasifikasi Status Penerima Bantuan Pangan Non Tunai. *JSI : Jurnal Sistem Informasi*, *16*(2), 313–324. https://doi.org/10.18495/jsi.v16i2.123

Aribowo, A. S., Cahyana, N. H., & Fauziah, Y. (2024). Enhancing Semi-Supervised Sentiment Analysis Through Hyperparameter Tuning Within Iterations: A Comparative Study Using Grid Search and Random Search. *International Conference on Advanced Informatics and Intelligent Information Systems*, 248–260. https://doi.org/10.2991/978-94-6463-366-5_23

Dang, X. T., & Le, T. T. (2024). KNN-SMOTE: An Innovative Resampling Technique Enhancing the Efficacy of Imbalanced Biomedical Classification. *Springer Nature Switzerland*, 111–121. https://doi.org/10.1007/978-3-031-63929-6_11

Handayani, D. N., & Qutub, S. (2025). Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(2), 405–412. https://doi.org/10.31004/riggs.v4i2.512

Putra, P. P., Anam, M. K., Chan, A. S., Hadi, A., Hendri, N., & Masnur, A. (2025). Optimizing Sentiment Analysis on Imbalanced Hotel Review Data Using SMOTE and Ensemble Machine Learning Techniques. *Journal of Applied Data Sciences*, *6*(2), 936–951. https://doi.org/10.47738/jads.v6i2.618

Qadrini, L., Hikmah, H., & Megasari, M. (2022). Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejawa Timur Tahun 2017. *Journal of Computer System and Informatics (JoSYC)*, *3*(4), 386–391. https://doi.org/10.47065/josyc.v3i4.2154

Rahayu, K., Anam, M. K., Efrizoni, L., Nurjayadi, & Fitri, T. A. (2024). Optimasi Teknik Voting Pada Sentimen Analisis Pemilihan Presiden 2024 Menggunakan Machine Learning. *The Indonesian Journal of Computer Science*, *13*(4), 6289–6302. https://doi.org/10.33022/ijcs.v13i4.4119

Raisyah, S. I., Zufria, I., & Triase, T. (2024). Sistem Pendukung Keputusan Dalam Menentukan Kelayakan Bantuan Dana Hibah Fasilitas Rumah Ibadah Menggunakan Metode AHP. *TIN: Terapan Informatika Nusantara*, *5*(3), 196–208. https://doi.org/10.47065/tin.v5i3.5558

Ramadani, P., Fadillah, R., Adawiyah, Q., & Al Ghazali, B. R. (2024). Perbandingan Algoritma Naïve Bayes, C4.5, dan K-Nearest Neighbor untuk Klasifikasi Kelayakan Program Keluarga Harapan. *Jurnal Media Informatika*, *6*(2), 775–782.

Reza, A. A. R., & Rohman, M. S. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *Journal Of Informatics And Telecommunication Engineering*, *7*(2), 534–544. https://doi.org/10.31289/jite.v7i2.10628

Ridwan, R., Hermaliani, E. H., & Ernawati, M. (2024). Penerapan Metode SMOTE Untuk Mengatasi Imbalanced Data Pada. *Computer Science (CO-SCIENCE)*, *4*(1), 80–88. https://doi.org/10.31294/coscience.v4i1.2990

Shah, S., Ghomeshi, H., Vakaj, E., Cooper, E., & Mohammad, R. (2023). An Ensemble-Learning-Based Technique for Bimodal Sentiment Analysis. *Big Data and Cognitive Computing*, *7*(2), 1–20. https://doi.org/10.3390/bdcc7020085

Susrifalah, A., Vionanda, D., Kurniawati, Y., & Sulistiowati, D. (2025). Penerapan Algoritma Extreme Gradient Boosting dengan ADASYN untuk Klasifikasi Rumah Tangga Penerima Program Keluarga Harapan di Provinsi Sumatera Barat. *UNP Journal of Statistics and Data Science*, *3*(2), 232–239. https://doi.org/10.24036/ujsds/vol3-iss2/369

Syamsiah, N. O., & Purwandani, I. (2023). Penerapan Metode Stacking Untuk Meningkatkan Akurasi Hasil Peramalan Konsumsi Listrik. *Journal of System and Computer Engineering (JSCE) ISSN*, *4*(1), 15–25. https://doi.org/10.47650/jsce.v4i1.665

VanFC, L. L., Anam, M. K., Bukhori, S., Mahamad, A. K., Saon, S., & Nyoto, R. L. V. (2025). The Development of Stacking Techniques in Machine Learning for Breast Cancer Detection. *Journal of Applied Data Sciences*, *6*(1), 71–85. https://doi.org/10.47738/jads.v6i1.416