Jurnal Teknologi Informasi dan Pendidikan
Volume 16, No. 2, September 2023
https://doi.org/10.24036/jtip.v16i2.800

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id JTiP-UNP

Optimizing Classification Algorithms Using Soft Voting: A Case Study on Soil Fertility Dataset

Fatkhurridlo Pranoto Kamarudin[1*✉], Fikri Budiman[1], Sri Winarno[1], Defri Kurniawan[1]
[1]Department of Informatics Engineering, Dian Nuswantoro University, Semarang, Indonesia
Corresponding Author: 111202013093@mhs.dinus.ac.id

Article Information

Article history:

No. 800
Rec. October 04, 2023
Rev. November 26, 2023
Acc. November 27, 2023
Pub. December 22, 2023
Page. 255 – 268

Keywords:

- *Ensemble learning*
- *Soft Voting Ensemble*
- *Agricultural Analysis*
- *Algorithm Optimization*

ABSTRACT

This study aims to optimize classification algorithms in soil fertility analysis using ensemble learning techniques, specifically Ensemble Soft Voting. In the context of modern agriculture, this research uses a dataset from the Purwodadi Agricultural Department to compare the performance of various classification algorithms such as Random Forest, Gradient Boosting, and Support Vector Machine (SVM), as well as the implementation of Ensemble Soft Voting. Each algorithm was evaluated separately, with Random Forest achieving an accuracy of 90.93%, Gradient Boosting at 91.53%, and SVM at 88.91%. After applying Ensemble Soft Voting, there was an increase in accuracy to 91.63%, with an average precision of 91.21%, recall of 91.77%, and an F1-Score of 91.49%. This study used a data split of 80% for training and 20% for testing. The results indicate that Ensemble Soft Voting can enhance the effectiveness in classifying soil fertility levels, potentially improving agricultural productivity and sustainability. These findings affirm that optimizing classification algorithms through ensemble techniques is crucial in enhancing the accuracy and effectiveness of predictive models in the agricultural sector.

How to Cite:

Kamarudin, F. P., et al. (2023). Optimizing Classification Algorithms Using Soft Voting: A Case Study on Soil Fertility Dataset. Jurnal Teknologi Informasi Dan Pendidikan, 16(2), 255-268. https://doi.org/10.24036/jtip.v16i2.800

1. INTRODUCTION

Soil fertility is critical for sustainable agriculture and adequate food production, necessitating a deep understanding of soil conditions and efficient management [1]–[3]. In modern agriculture, technology and data analysis, particularly classification techniques and machine learning, play a significant role in predicting and understanding soil fertility [4]–[6]. This research focuses on optimizing classification algorithms using the Soft Voting ensemble approach, which combines multiple machine-learning models to enhance performance and robustness in soil classification [7]–[9]. The Soft Voting approach utilizes probabilities from various models for more accurate predictions, improving the capability to classify soil precisely [10]–[12]. This study concentrates on three classification algorithms: Random Forest, Gradient Boosting, and Support Vector Machine (SVM), chosen for their respective strengths in dealing with soil fertility complexities. Random Forest is selected for its robustness, Gradient Boosting for its predictive power, and SVM for its flexibility. The study integrates their individual strengths by evaluating and optimizing separate parameters for each algorithm. The Soft Voting ensemble aims to address the challenges of soil fertility classification more effectively[13]–[17].

Previous studies in agriculture, such as those by Pragathi (2021) and Mella & Pentakoti (2022), have utilized various soil attributes to train classification models with machine learning algorithms. Still, this research focuses on a soil fertility dataset [18], [19]. Karlos, Kostopoulos, & Kotsiantis (2020) also employed the Soft Voting method but focused more on semi-supervised learning and co-training [20]. This research's primary innovation lies in ensemble learning techniques, particularly Soft Voting, in soil fertility analysis. Consequently, this research aims to optimize classification algorithms in soil fertility analysis using ensemble learning techniques, specifically Ensemble Soft Voting. The expected outcome is to enhance the accuracy of soil fertility predictions, support decision-making in soil resource management, and improve agricultural productivity and farmers' income while supporting sustainable agriculture.

2. RESEARCH METHOD

Research methods are a framework for carrying out the stages of research, as depicted in Figure 1.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

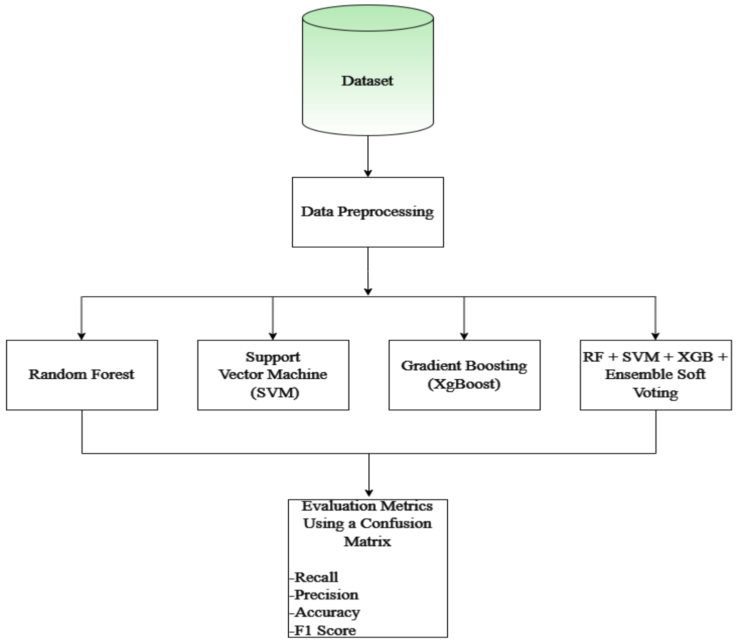

Figure 1. Research Method Flow

2.1. Dataset

This study utilizes private data obtained from the Agriculture Department of Grobogan Regency, focusing on soil fertility. The dataset comprises 5,000 data records, with 2,520 records indicating fertile land and 2,480 records indicating infertile land. There are 16 attributes included, which cover information about soil element contents and parameters to determine the soil fertility status.

Table 1. Data

Data	Class	Attribute	Records
Agriculture	2	16	5.000

The attributes consist of variables with specifications as below:

Table 2. Agricultural Data Attributes

Variable	Attribute	Data Type
Inputs	pH, EC, OC, OM, N, P, K, Zn, Fe, Cu, Mn, Sand, Silt, Clay, CaCO3, CEC	Numeric
Outputs	Fertile/Non Fertile	Binary

2.2. Pre-Processing Data

Data preprocessing is a crucial initial stage in data analysis that involves the investigation, cleaning, and preparation of raw data. There are two main steps: resampling, which addresses the imbalance in the number of data between classes in the dataset, and data standardization, which aims to normalize the range of values for all attributes to avoid the dominance of specific attributes in classification.

2.3. Classification of Random Forest Algorithms

Classification using Random Forests, a part of ensemble learning, combines multiple decision trees to enhance accuracy through majority voting. Effective against overfitting and suitable for imbalanced data and large datasets, Random Forest provides insights into feature importance. This study compares its performance with Gradient Boosting, SVM, and Soft Voting in predicting soil fertility.

2.4. Support Vector Machine (SVM) Classification Algorithm

The Support Vector Machine (SVM) algorithm is an effective classification method that separates data into categories or classes by finding the best hyperplane and maximizing the margin between two classes. The hyperplane equation is as follows:

$$\omega . x + b = 0 \tag{1}$$

Where ω is the normal vector to the hyperplane, x is the input feature vector, and b is the bias. SVM introduces the concept of "support vectors," which are the data points closest to the hyperplane and crucial in determining the position and orientation of the hyperplane. The use of SVM in this research aims to understand the accuracy of SVM in classifying soil fertility levels based on relevant attributes.

2.5. Gradient Boosting Algorithm Classification

The Gradient Boosting algorithm is an effective machine learning method for classification, building successive decision trees and enhancing prediction accuracy, particularly in predicting soil fertility with a focus on complex data. Its implementation will be further explained in this research.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

2.6. Ensemble Soft Voting Approach

In this research, the ensemble Soft Voting approach combines Random Forest, Gradient Boosting, and SVM models through "majority voting" to improve the accuracy of soil fertility predictions. This approach is expected to provide more precise and reliable predictions for modern agriculture, with its performance evaluation discussed in the results and discussion chapter.

2.7. Results Evaluation using Confusion Matrix

In classification testing, the evaluation method uses a confusion matrix that allows assessing the reliability of the classification model by comparing prediction results with actual data in a matrix form. The confusion matrix helps in measuring the performance of classification algorithms in more depth, offering an insight into the model's effectiveness in predicting the correct categories. There are four main components in the confusion matrix: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive covers data that is actually positive and predicted as positive, True Negative for negative data predicted as negative, False Positive for negative data incorrectly predicted as positive, and False Negative for positive data incorrectly predicted as negative.

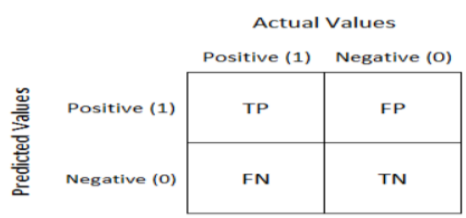

Figure 2. Confusion Matrix

Accuracy measures how correct the predictions are out of all the data used and is calculated using the formula:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision measures the extent to which positive predictions are correct compared to the total positive predictions, calculated using the formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Recall measures how many of the positive predictions are positive compared to the total number of samples that should be positive, calculated using the formula:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

F1 Score is a combined metric that integrates precision and recall and is calculated using the formula:

$$\text{F1 Score} = = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (5)$$

3. RESULTS AND DISCUSSION

3.1. Dataset

This study utilized a dataset from the Grobogan District Agricultural Department for soil fertility analysis, comprising 5,000 data records with attributes such as pH, EC, OC, OM, N, P, K, Zn, Fe, Cu, Mn, Sand, Silt, Clay, CaCO3, and CEC. The dataset categorizes soil into "Fertile" and "NonFertile," serving as the foundation for predicting soil fertility in modern agriculture.

3.2. Data Preprocessing

In the data preprocessing stage, we undertook two crucial steps, data resampling and data standardization, to ensure the reliability and performance of the classification model we were developing.

3.2.1. Data Resampling

Our preprocessing began with addressing the class imbalance in the dataset. Initially, the fertile land class (1) had 2,520 samples, while the non-fertile land class (0) had only 2,480 samples. This condition could affect the model's performance, tending to predict the majority class. To counter this, we employed a resampling technique to equalize the samples in both classes. After resampling, each class had 2,480 samples, as shown in Table 3 below:

Table 3. Number of Samples Before Class and After Resampling

Class	Before Resampling	After Resampling
0	2.480	2.480
1	2.520	2.480

3.2.2. Data Standardization

After addressing the class imbalance, the data was standardized using the StandardScaler from sklearn. This process normalized all attributes to a similar range of

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

values, with a mean of 0 and a standard deviation of 1, preparing the dataset for training and testing the classification model.

3.3. Random Forest Classification

We employed the Random Forest algorithm for data classification, dividing the dataset into 80% training and 20% testing data. This algorithm was applied to 4000 training and 1000 testing entries, constructed decision trees and combined their predictions for effective modelling. The model's evaluation was conducted using a confusion matrix.

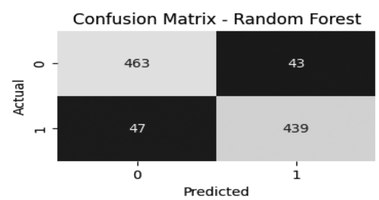

Figure 3. Confusion Matrix Random Forest

By using the confusion matrix, we obtain several performance evaluation metrics for the model, namely:

Table 4. Random Forest Classification Results

Confusion Matrix	Accuracy
Accuracy	90.9%
Precision	91.1%
Recall	90.3%
F1 Scores	90.7%

The Random Forest model on this dataset has an accuracy of 90.9%, precision of 91.1%, recall of 90.3%, and an F1-Score of 90.7%, indicating good performance and reliability in predicting soil fertility levels.

3.4. SVM Classification

After dividing the dataset, we applied SVM with a linear kernel and C=1.0 to the training data, trained the model, and then made predictions with the test data to evaluate the performance of the SVM model in soil fertility classification. Further evaluation was conducted using a confusion matrix to measure the accuracy of predictions.

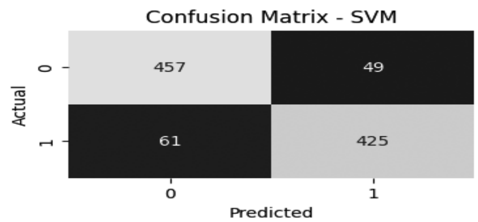

Figure 4. Confusion Matrix Support Vector Machine (SVM)

By utilizing the confusion matrix, we obtained several performance evaluation metrics for the model, namely:

Table 5. Random Forest Classification Results

Confusion Matrix	Score
Accuracy	88.9%
Precision	89.7%
Recall	87.4%
F1 Scores	88.5%

The SVM model effectively classifies soil fertility levels, demonstrating an accuracy of 88.9%, precision of 89.7%, recall of 87.4%, and an F1-Score of 88.5%, reflecting its ability in precise prediction and a balance between precision and recall.

3.5. Gradient Boosting Classification

The application of the Gradient Boosting algorithm on the training data with parameters such as the number of estimators (100), learning rate (0.1), and maximum tree depth (3), followed by model evaluation using a confusion matrix.

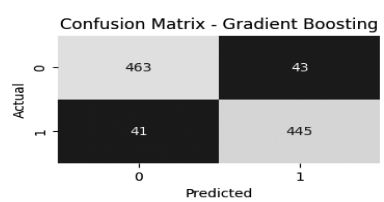

Figure 5. Confusion Matrix Gradient Boosting

By using the confusion matrix, we get several model performance evaluation metrics, namely:

Table 6. Random Forest Classification Results

Confusion Matrix	Score
Accuracy	88.9%
Precision	89.7%
Recall	87.4%
F1 Scores	88.5%

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

Confusion matrix analysis on the Random Forest model reveals high performance with an accuracy of 88.9%, precision of 89.7%, recall of 87.4%, and an F1 Score of 88.5%, demonstrating its effectiveness in soil fertility classification.

3.6. Ensemble Voting with Soft Voting

After dividing the data, we employed Ensemble Soft Voting, which combines Random Forest, Gradient Boosting, and SVM, making decisions based on the highest probability of prediction. Subsequently, the model's performance was evaluated using a confusion matrix to measure the accuracy of class predictions.

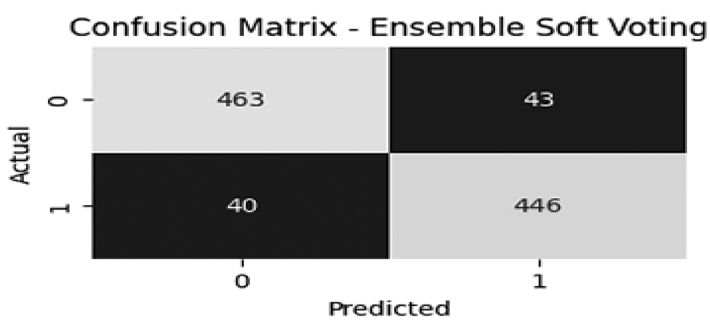

Figure 6. Confusion Matrix Ensemble Soft voting

By using the confusion matrix, we get several model performance evaluation metrics, namely:

Table 7. Ensamble Soft Voting Results

Confusion Matrix	Score
Accuracy	91.6%
Precision	91.2%
Recall	91.8%
F1 Scores	91.5%

Confusion matrix analysis on Ensemble Soft Voting indicates a high model performance with an accuracy of 91.6%, precision of 91.2%, recall of 91.8%, and F1 Score of 91.5%, suggesting accurate class predictions, precise identification of positive categories, and a balance between precision and recall.

Jurnal Teknologi Informasi dan Pendidikan
Volume 16, No. 2, September 2023
https://doi.org/10.24036/jtip.v16i2.800

3.7. Evaluation of Results Using Confusion Matrix

We will compare the performance of the Random Forest, Gradient Boosting, and SVM classification algorithms before and after applying Ensemble Voting. The focus will be on metrics such as accuracy, precision, recall, and F1-Score, accompanied by a graphical visualization of the evaluation results. Below are the graphical visualizations of the evaluation results for each algorithm and Ensemble Voting.

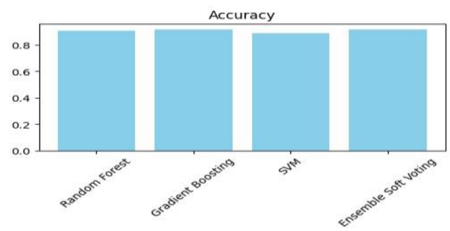

Figure 7. Accuracy graph

The first graph indicates that Ensemble Voting has the highest accuracy (91.6%), followed by Gradient Boosting (91.5%), Random Forest (90.9%), and SVM (88.9%), illustrating the performance differences among the algorithms.

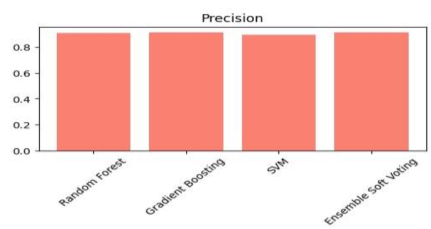

Figure 8. Precision Graph

The second graph indicates that Ensemble Voting leads in precision (91.21%), followed by Gradient Boosting (91.19%), Random Forest (91.08%), and lower SVM (89.66%), providing an overview of the accuracy of positive predictions for the models.

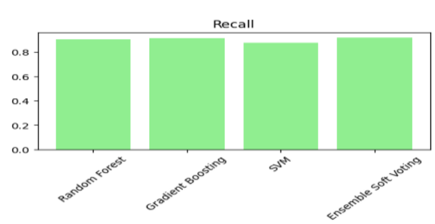

Figure 9. Recall graph

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

The third graph indicates that Ensemble Voting has the highest recall (91.8%) in identifying positive data, followed by Gradient Boosting (91.6%), Random Forest (90.3%), and SVM (87.4%), with Ensemble Voting as the best-performing model.

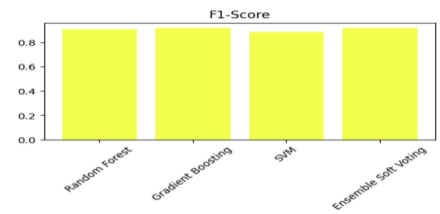

Figure 10. F1-Score graph

The F1-Score graph indicates Ensemble Voting as the best method with the highest performance, achieving a score of 91.49% and outperforming other methods such as Gradient Boosting, Random Forest, and SVM in terms of accuracy, precision, and recall. Ensemble Voting stands out with an accuracy of 91.63%, while SVM has the lowest values in all metrics. These results confirm the effectiveness of Ensemble Voting in enhancing the performance of the soil fertility classification model.

This research addresses the challenges of modern agriculture and sustainable farming by focusing on soil, a valuable agricultural asset whose fertility supports sustainable food production [21], [22]. Efficient soil management requires an in-depth understanding of its fertility, where modern technology and data analysis, including machine learning, play a crucial role. Previous studies have shown that ensemble learning techniques, particularly Ensemble Soft Voting, enhance the performance of classification models [23]–[25]. Ensemble learning combines results from various machine learning models to improve performance and resilience [26]–[28]. This approach reduces the variability and errors of a single model, supporting the use of Ensemble Soft Voting for more accurate soil fertility prediction.

This study encompasses three steps. First, data preprocessing, including resampling and standardization, ensures the model's reliability [35], [36]. Second, applying individual classification algorithms, namely Random Forest, Gradient Boosting, and SVM [29], [30]. Third, implementing Ensemble Soft Voting combines the results of these three algorithms. Evaluations show that Ensemble Soft Voting provides an accuracy of about 91.63%, precision of 91.21%, recall of 91.77%, and F1-Score of 91.49%, better than a single algorithm. This research confirms the effectiveness of ensemble learning in soil fertility analysis [31], [32]. Thus, this study contributes to developing classification models for soil fertility analysis, supporting sustainable agriculture and managing agricultural resources. These findings also support previous research, such as Pragathi (2021) and Mella & Pentakoti (2022), which also identified the benefits of ensemble learning. This opens up opportunities for further development, including communication technology in disseminating soil fertility information and fertilization recommendations.

4. CONCLUSION

This study uses ensemble learning techniques, specifically Ensemble Soft Voting, to optimize soil fertility classification, utilizing a Grobogan District Agricultural Department dataset comprising 5,000 data records and 16 attributes. The applied methods include data pre-processing, evaluation of individual classification algorithms (Random Forest, SVM, Gradient Boosting), and their amalgamation through Ensemble Soft Voting. The primary results indicate that Ensemble Soft Voting significantly enhances accuracy, precision, recall, and F1-Score, achieving the highest values compared to single algorithms. The study recommends further exploration in applying this technology in agricultural resource management and disseminating soil fertility information to support sustainable agriculture.

REFERENCES

[1] A. A. Adenle, K. Wedig, and H. Azadi, "Sustainable agriculture and food security in Africa: The role of innovative technologies and international organizations," *Technol. Soc.*, vol. 58, no. 1, pp. 1–54, 2019, doi: 10.1016/j.techsoc.2019.05.007.

[2] N. Khan, R. L. Ray, G. R. Sargani, M. Ihtisham, M. Khayyam, and S. Ismail, "Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture," *Sustain.*, vol. 13, no. 9, pp. 1–31, 2021, doi: 10.3390/su13094883.

[3] M. Bertola, A. Ferrarini, and G. Visioli, "Improvement of soil microbial diversity through sustainable agricultural practices and its evaluation by -omics approaches: A perspective for the environment, food quality and human safety," *Microorganisms*, vol. 9, no. 7, pp. 1–22, 2021, doi: 10.3390/microorganisms9071400.

[4] E. Z. Baskent, "A framework for characterizing and regulating ecosystem services in a management planning context," *Forests*, vol. 11, no. 1, pp. 1–20, 2020, doi: 10.3390/f11010102.

[5] M. Javaid, A. Haleem, I. H. Khan, and R. Suman, "Understanding the potential applications of Artificial Intelligence in Agriculture Sector," *Adv. Agrochem*, vol. 2, no. 1, pp. 15–30, 2023, doi: 10.1016/j.aac.2022.10.001.

[6] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Enhancing smart farming through the applications of Agriculture 4.0 technologies," *Int. J. Intell. Networks*, vol. 3, no. 7, pp. 150–164, 2022, doi: 10.1016/j.ijin.2022.09.004.

[7] N. Khan, M. A. Kamaruddin, U. U. Sheikh, Y. Yusup, and M. P. Bakht, "Oil palm and machine learning: Reviewing one decade of ideas, innovations, applications, and gaps," *Agric.*, vol. 11, no. 9, pp. 1–26, 2021, doi: 10.3390/agriculture11090832.

[8] M. S. Suchithra and M. L. Pai, "Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters," *Inf. Process. Agric.*, vol. 7, no. 1, pp. 72–82, 2020, doi: 10.1016/j.inpa.2019.05.003.

[9] S. K. S. Durai and M. D. Shamili, "Smart farming using Machine Learning and Deep Learning techniques," *Decis. Anal. J.*, vol. 3, no. 3, pp. 1–30, 2022, doi: 10.1016/j.dajour.2022.100041.

[10] X. Peng, X. Yu, Y. Luo, Y. Chang, C. Lu, and X. Chen, "Prediction Model of Greenhouse Tomato Yield Using Data Based on Different Soil Fertility Conditions," *Agronomy*, vol. 13, no. 7, pp. 1–

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id

19, 2023, doi: 10.3390/agronomy13071892.

[11] Y. Shahare *et al.*, "A Comprehensive Analysis of Machine Learning-Based Assessment and Prediction of Soil Enzyme Activity," *Agric.*, vol. 13, no. 7, pp. 1–18, 2023, doi: 10.3390/agriculture13071323.

[12] D. Ganesh, K. J. A. Yeshwanth, M. Satheesh, M. G. S. V. Reddy, T. Chirudeep, and S. N. K. Polisetty, "Extreme Learning Mechanism for Classification & Prediction of Soil Fertility index," *J. Pharm. Negat. Results*, vol. 13, no. 6, pp. 37–43, 2022, doi: 10.47750/pnr.2022.13.S06.006.

[13] Z. Aslam, N. Javaid, A. Ahmad, A. Ahmed, and S. M. Gulfam, "A combined deep learning and ensemble learning methodology to avoid electricity theft in smart grids," *Energies*, vol. 13, no. 21, pp. 1–24, 2020, doi: 10.3390/en13215599.

[14] M. Alipour and D. K. Harris, "Increasing the robustness of material-specific deep learning models for crack detection across different materials," *Eng. Struct.*, vol. 206, no. 2, pp. 1–14, 2020, doi: 10.1016/j.engstruct.2019.110157.

[15] C. Chen and H. Liu, "Dynamic ensemble wind speed prediction model based on hybrid deep reinforcement learning," *Adv. Eng. Informatics*, vol. 48, no. 4, pp. 1–15, 2021, doi: 10.1016/j.aei.2021.101290.

[16] K. Lavanya, A. J. Obaid, I. S. Thaseen, K. Abhishek, K. Saboo, and R. Paturkar, "Terrain mapping of landsat8 images using mnf and classifying soil properties using ensemble modelling," *Int. J. Nonlinear Anal. Appl.*, vol. 11, no. 1, pp. 527–541, 2020, doi: 10.22075/IJNAA.2020.4750.

[17] A. Gasmi, C. Gomez, A. Chehbouni, D. Dhiba, and M. El Gharous, "Using PRISMA Hyperspectral Satellite Imagery and GIS Approaches for Soil Fertility Mapping (FertiMap) in Northern Morocco," *Remote Sens.*, vol. 14, no. 16, pp. 1–22, 2022, doi: 10.3390/rs14164080.

[18] K. Pragathi, "Crop Yield Prediction, Forecasting and Fertilizer Recommendation using Voting Based Ensemble Classifier," *Int. J. Innov. Res. Technol.*, vol. 8, no. 6, pp. 510–516, 2021, doi: 10.14445/23488387/ijcse-v7i5p101.

[19] N. V. V. P. Mella and V. M. Pentakoti, "Crop yield prediction and Fertilizer Recommendation using Voting Based Ensemble Classifier," *J. Eng. Sci.*, vol. 13, no. 8, pp. 262–270, 2022, doi: 10.14445/23488387/ijcse-v7i5p101.

[20] S. Karlos, G. Kostopoulos, and S. Kotsiantis, "A Soft-Voting Ensemble Based Co-Training Scheme Using Static Selection for Binary Classification Problems," *Algorithms*, vol. 13, no. 1, pp. 1–19, 2020, [Online]. Available: https://doi.org/10.3390/a13010026

[21] K. Pawlak and M. Kołodziejczak, "The role of agriculture in ensuring food security in developing countries: Considerations in the context of the problem of sustainable food production," *Sustain.*, vol. 12, no. 13, 2020, doi: 10.3390/su12135488.

[22] X. Wang, "Managing Land Carrying Capacity: Key to Achieving Sustainable Production Systems for Food Security," *Land*, vol. 11, no. 4, pp. 1–21, 2022, doi: 10.3390/land11040484.

[23] G. T. Reddy *et al.*, "An Ensemble based Machine Learning model for Diabetic Retinopathy Classification," in *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, 2020, pp. 1–6. doi: 10.1109/ic-ETITE47903.2020.235.

[24] N. Peppes, E. Daskalakis, T. Alexakis, E. Adamopoulou, and K. Demestichas, "Performance of machine learning-based multi-model voting ensemble methods for network threat detection in agriculture 4.0," *Sensors*, vol. 21, no. 22, pp. 1–17, 2021, doi: 10.3390/s21227475.

[25] A. Taha, "Intelligent ensemble learning approach for phishing website detection based on weighted soft voting," *Mathematics*, vol. 9, no. 21, pp. 1–13, 2021, doi: 10.3390/math9212799.

[26] L. Liu *et al.*, "Deep neural network ensembles against deception: Ensemble diversity, accuracy and robustness," *Proceedings - 2019 IEEE 16th International Conference on Mobile Ad Hoc and Smart Systems, MASS 2019*. pp. 274–282, 2019. doi: 10.1109/MASS.2019.00040.

[27] A. Abbasi, A. R. Javed, C. Chakraborty, J. Nebhen, W. Zehra, and Z. Jalil, "ElStream: An Ensemble Learning Approach for Concept Drift Detection in Dynamic Social Big Data Stream Learning," *IEEE Access*, vol. 9, no. 1, pp. 1–12, 2021, doi: 10.1109/ACCESS.2021.3076264.

[28] G. Saxena, D. K. Verma, A. Paraye, A. Rajan, and A. Rawat, "Improved and robust deep learning agent for preliminary detection of diabetic retinopathy using public datasets," *Intell. Med.*, vol. 3–4, no. 1, pp. 1–11, 2020, doi: 10.1016/j.ibmed.2020.100022.

[29] P. Srivastava, A. Shukla, and A. Bansal, "A comprehensive review on soil classification using deep learning and computer vision techniques," *Multimed. Tools Appl.*, vol. 80, no. 10, pp. 14887–14914, 2021, doi: 10.1007/s11042-021-10544-5.

[30] O. Folorunso *et al.*, "Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review," *Big Data Cogn. Comput.*, vol. 7, no. 2, pp. 1–25, 2023, doi: 10.3390/bdcc7020113.

[31] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 1, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[32] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood, and G. S. Choi, "A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis," *PLoS One*, vol. 16, no. 2, pp. 1–23, 2021, doi: 10.1371/journal.pone.0245909.

P.ISSN: 2086 – 4981
E.ISSN: 2620 – 6390
tip.ppj.unp.ac.id