Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Optimization Accuracy Value of Agricultural Land Fertility Classification Using Soft Voting Method

Khaliq Pradana[1)]*, F Budiman[2]
[1,2)]Department of Informatics, Computer Sciences, Universitas Dian Nuswantoro, Indonesia
[1)] 111202013159@mhs.dinus.ac.id , [2)]fikri.budiman@dsn.dinus.ac.id

Submitted : Nov 7, 2023 | **Accepted** : Nov 30, 2023 | **Published** : Jan 1, 2024

Abstract: Soil fertility on an agricultural land is very influential with agricultural yields, where plants can grow well and fertile if nutrient intake is met. The purpose of this research is to improve the accuracy in predicting soil fertility by utilizing machine learning by combining two classification algorithms using soft voting methods in the classification of agricultural land fertility. In this research, one of the ensemble learning methods called soft voting is employed. Soft voting is used to enhance accuracy by optimizing the combination of algorithms based on the highest probability provided by each model. The Gaussian Naive Bayes algorithm is used to predict classes in the sample data based on the Gaussian distribution of numerical data, while the decision tree is utilized to predict classes by constructing a decision tree using soil content attributes for the classification of fertile or infertile soil. The use of the Gaussian Naive Bayes algorithm in identifying fertile and infertile soil based on existing attributes achieved an accuracy rate of 87.2%. The decision tree algorithm, based on decision tree modeling, helped identify important attributes for decision-making with an accuracy rate of 88.3%. The soft voting method played a crucial role in improving accuracy by combining both algorithms, resulting in an accuracy rate of 88.8%. Based on the accuracy results obtained, the use of soft voting optimization in predicting soil fertility has the highest accuracy because it combines the Gaussian naïve bayes algorithm and the decision tree algorithm.

Keywords: Agricultural Agency, classification, decisiom tree, gaussian nave bayes, soft voting.

INTRODUCTION

Soil fertility on an agricultural land is very influential with agricultural yields, where plants can grow well and fertile if nutrient intake is met. The difficult food needs and the large demand for food encourage to conduct this research, because with agricultural land that has fertile soil will provide abundant harvests. By utilizing computer science, especially data mining with machine learning in agriculture, it can determine soil fertility using algorithms as a classification and predict fertile or infertile soil. Soil fertility research is still very rarely done, so this research was conducted to provide the results of soil fertility classification and prediction. By utilizing optimization to increase accuracy using ensemble learning, ensemble learning is an optimization technique that aims to improve prediction performance by combining predictions from several models (Dong et al., 2020; Matloob et al., 2021; Zhou et al., 2021). There are several common ensemble learning methods, including bagging, boosting, stacking, and soft voting (Athar et al., 2021; Kumari et al., 2021; Taha, 2021).

*name of corresponding author

The Soft Voting method optimizes accuracy by combining the highest probability accuracy values from the Gaussian Naïve Bayes and Decision Tree algorithms which have never been used before. The Gaussian Naïve Bayes and Decision Tree algorithms are classification algorithms that can analyze data by extracting models and describing them according to data classes, making them suitable for analyzing soil fertility levels. The Gaussian Naïve Bayes algorithm excels at processing large-scale data, provides fast classification processing speed, and works well with datasets that have many features, especially numerical data that fits a Gaussian distribution. It can provide good results for binary classification (Kamel et al., 2019; Rafique et al., 2019; Wibowo et al., 2023). On the other hand, the Decision Tree algorithm has advantages in data optimization as it is easy to understand, generates decisions through a tree structure, and can provide accurate results for classification and regression tasks. It is robust to outliers and able to handle nonlinear data relationships (Ahmim et al., 2019; Humbird et al., 2018; Patel et al., 2018).

The rationale for using the Gaussian Naïve Bayes and Decision Tree algorithms in this research is based on previous studies that have illustrated the success of these algorithms in various contexts. For example, a study highlighted the effectiveness of the Gaussian Naïve Bayes algorithm in making marketing strategy decisions by utilizing information systems technology and minimizing the existing problems by 71% (Valentinus et al., 2023). In addition, the application of data mining and Gaussian Naïve Bayes algorithm for cancer disease classification achieved 90% accuracy (Kamel et al., 2019). The Decision Tree algorithm has also been used to search surrogate statistical procedure data, to extract text, medical certified fields and also in search engines. with an accuracy of 92.36% (Sanjay et al., 2019). In the medical context, it is used in the diagnosis of liver diseases with 72.67% accuracy (Setiawati et al., 2019). Furthermore, it has been applied to the analysis of poverty levels in Indonesia with an accuracy of 88.6% (Kaunang, 2018). All of these previous studies inspired new research in the application of the Gaussian Naïve Bayes and Decision Tree algorithms to predict soil fertility on agricultural land.

Previous research has used the Gaussian Naïve Bayes algorithm (Jayachitra et al., 2021) and Decision Tree (Denny et al., 2019) separately for classification modeling regarding plant fertility. However, there has been no comprehensive research comparing the effectiveness of both algorithms by applying the Soft Voting method to optimize soil fertility classification accuracy. This research is expected to provide more comprehensive guidance on the use of Gaussian Naïve Bayes and Decision Tree algorithms for classification optimization using the Soft Voting method where there is an increase in accuracy values, as well as a better understanding of when and how to use these methods more effectively in different contexts. The main difference of this research compared to previous research is the combination of the two algorithms with the Soft Voting method resulting in higher accuracy.

This research aims to provide farmers and land managers with better information for decision-making regarding land use, resource utilization and sustainable agricultural practices through soil fertility assessment. In addition, the contribution of this research can help reduce environmental damage by avoiding excessive use of fertilizers or pesticides. Therefore, this research has important value in supporting more efficient and sustainable agriculture. Besides, in reality, Agricultural Extension Agencies and Agricultural Offices in an area still perform soil fertility classification manually. This involves physically visiting the land for inspection, using agricultural science expertise, and performing manual calculations to determine soil fertility. This research plays a role in assisting soil fertility classification through the utilization of machine learning.

LITERATURE REVIEW

Research using classification algorithms has been conducted extensively in various contexts. The Gaussian Naive Bayes algorithm is used in marketing strategy decision-making by leveraging information system technology and minimizing existing issues (Valentinus et al., 2023). The Gaussian Naive Bayes algorithm is applied in cancer disease classification (Kamel et al., 2019). The Decision Tree algorithm is used to find substitute statistical procedure data, to extract text, in certified medical fields, and also in search engines (Sanjay et al., 2019). The Decision Tree algorithm is employed in liver disease diagnosis (Handayani et al., 2019; Setiawati et al., 2019). Combining two or more

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

optimized classification algorithms using the soft voting method for soil fertility is still relatively uncommon, where soil fertility is a critical aspect in agriculture that directly affects crop productivity and harvest yields. Therefore, research to optimize soil fertility has a significant impact on modern agriculture. In this context, classification methods have become valuable tools for analyzing and classifying soil fertility. One intriguing approach is the soft voting method, which combines various classification algorithms to enhance prediction accuracy (Islam et al., 2019; Karlos et al., 2020; Salur et al., 2022; Saqlain et al., 2019). In previous studies, various classification methods have been used, including Gaussian Naive Bayes and Decision Tree.

Gaussian Naive Bayes is a classification algorithm based on Bayes' theorem and is suitable for classification problems with numeric attributes (Jayachitra et al., 2021; Rafique et al., 2019). On the other hand, Decision Tree is a classification model that understands the relationships between attributes in data and is used to make decisions based on the constructed decision tree (Pasha et al., 2023; Pratama et al., 2022). The soft voting method combines the outputs of multiple classification algorithms to improve accuracy (Kumari et al., 2021; Verma et al., 2023). Previous research has shown that combining Gaussian Naive Bayes and Decision Tree using the soft voting method can produce more accurate predictions regarding soil fertility. This study will further investigate this method and consider ways to enhance prediction quality through the improved utilization of these algorithms in the context of soil fertility. Thus, this literature review will examine the theoretical foundations and contributions of previous research in the field of soil fertility optimization classification using the soft voting method, with a particular focus on the use of Gaussian Naive Bayes and Decision Tree as its primary components.

METHOD

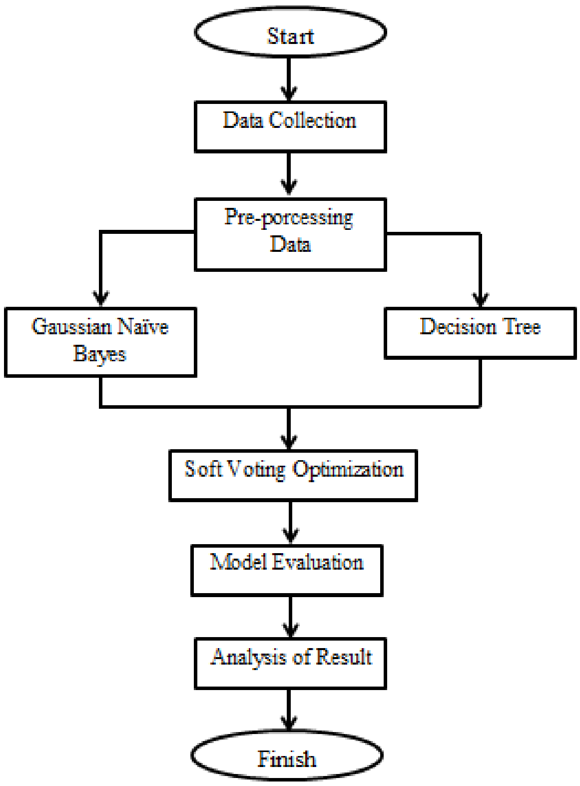

Figure 1. Research Flow Diagram

Datasets

In this study, the data used data obtained from the Grobogan District Agriculture Office regarding soil fertility, where the soil contains various nutrient elements. The research utilizes a total of 5,000

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

records, comprising 2,520 records of fertile soil and 2,480 records of infertile soil, with 16 attributes and 2 class labels. This data is related to the values of nutrient elements contained in the soil, and there are also parameters that determine whether the soil is fertile or infertile. For data specifications, please refer to the following table 1.

Table 1. Data

Data	Class	Attribute	Record
Agriculture	2	16	5.000

The attributes consist of variables with specifications as follow table 2.

Table 2. Agricultural Data Attributes

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

Pre-processing

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Data preprocessing is an important stage in analyzing data. Where in preprocessing there are three stages, namely:

Mapping is the process of converting data from binary to numeric form. Data format before the mapping, can be seen in table 3.

Table 3. Data Before Mapping

	pH	**EC**	**OC**	**OM**	**N**	**P**	**K**	**Zn**	**Fe**	**Cu**	**Mn**	**Sand**	**Silt**	**Clay**	**CaCO3**	**CEC**	**Output**
0	9.44	0.70	0.02	0.16	160	30.90	181	1.13	4.7	0.51	1.3	86.3	7.2	11.6	13.59	7.71	Fertile
1	7.24	0.02	0.05	0.17	89	22.83	112	0.67	4.9	0.34	0.5	96.2	8.4	5.5	3.42	5.08	Non Fertile

The binary nature of the class labels in the data needs to be converted into numeric form using a mapping command, similar to the one shown in Figure 2.

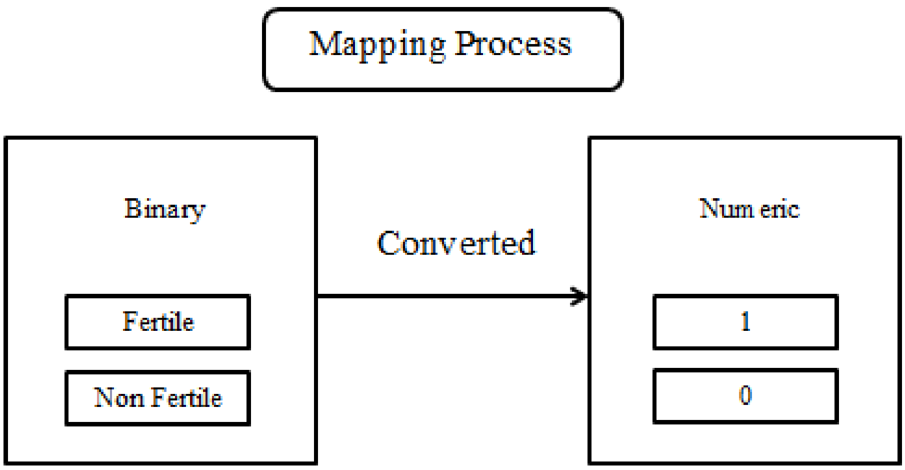

Figure 2. Mapping Process

Data format after the mapping is done, can be seen in table 4.

Table 4. Data After Mapping

	pH	**EC**	**OC**	**OM**	**N**	**P**	**K**	**Zn**	**Fe**	**Cu**	**Mn**	**Sand**	**Silt**	**Clay**	**CaCO3**	**CEC**	**Output**
0	9.44	0.70	0.02	0.16	160	30.90	181	1.13	4.7	0.51	1.3	86.3	7.2	11.6	13.59	7.71	1
1	7.24	0.02	0.05	0.17	89	22.83	112	0.67	4.9	0.34	0.5	96.2	8.4	5.5	3.42	5.08	0

Data balancing is a technique used to overcome data imbalance, so unbalanced data needs to be balanced to ensure the model is not biased between minority data and majority data.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

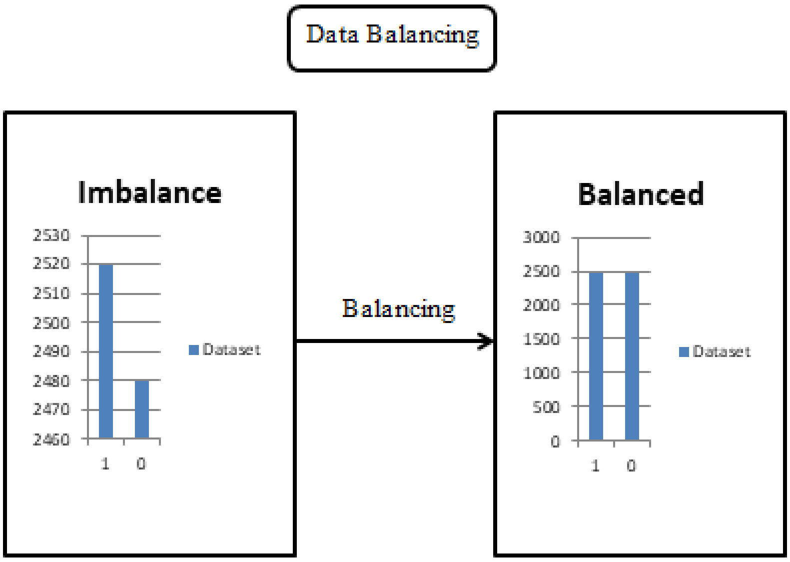

Figure 3. Data Balancing

- Data Standar Scaler is the process of rescaling and redistributing data to have a mean of zero and a standard deviation of one.

Standar Scaler

Before

Attribute	Mark
pH	9.44
EC	0.70
OC	0.02
OM	0.16
N	160
P	30.90
K	181
Zn	1.13
Fe	4.7
Cu	0.51
Mn	1.3
Sand	86.3
Silt	7.2
Clay	11.6
CaCO3	13.59
CEC	7.71

Rescaling
Redistributing

After

Attribute	Mark
pH	-0.0809732
EC	-0.75650712
OC	0.14914766
OM	-0.92171679
N	-0.58311814
P	144.858.555
K	8.125.117
Zn	-0.69348926
Fe	0.56356323
Cu	-0.35288724
Mn	0.45332742
Sand	-0.80675251
Silt	-0.73224745
Clay	-0.57234476
CaCO3	-0.98553453
CEC	-0.33758344

Figure 4. Data Standar Scaler

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Gaussian Naïve Bayes Algorithm Classification

Gaussian Naive Bayes is a classification algorithm used to predict the category or class of a sample based on the Gaussian (normal) distribution of numerical features. Equation (1) in the Gaussian Naïve Bayes Classifier.

$$P(y \mid x_1, x_2, ..., x_n) = \frac{P(y) \bullet P(x\_1 \mid y) \bullet P(x\_2 \mid y) \bullet ... \bullet P(x\_n \mid y)}{P(x\_1) \bullet P(x\_2) \bullet ... \bullet P(x\_n)} \quad (1)$$

Where is the formula for equation (1) :
P(y | x1, x2, ..., xn) = The probability of class y given features x1, x2, ..., xn.
P(y) = The prior probability of class y.
P(xi | y) = The probability of feature xi within class y.
P(x1) • P(x2) • ... • P(xn) = The joint probability of all features.

The collection of training datasets consists of features and corresponding labels or categories. This algorithm calculates the prior probability P(y) for each class, which indicates how often each class appears in the training dataset. Next, the algorithm estimates the Gaussian distribution P(xi | y) for each feature x_i within each class y. This involves calculating the mean and standard deviation for each feature within each class. When it's time to make predictions for unknown samples, the algorithm calculates the probability P(y | x1, x2, ..., xn) for each class y and then selects the class with the highest probability as the prediction. The Gaussian Naive Bayes model can be evaluated using metrics such as accuracy, precision, and recall to measure how well the model performs classification on test data.

Decision Tree Algorithm Classification

A Decision Tree is a predictive model that represents decision flow as a tree. This tree is used to make decisions based on input data characteristics. Each node on the tree represents a search on a data entity, each branch represents the outcomes of that search, and each leaf represents a label or prediction value. The goal is to divide the input data into smaller subsets to make accurate decisions or predictions. A Decision Tree begins by selecting the best feature when splitting nodes based on impurity measures. The feature with the lowest impurity measure is chosen. Data is divided into two subsets based on the selected best feature's value.

This process is carried out recursively for each subset until a stopping condition is met, such as reaching the maximum depth or the minimum number of samples in a node. The process of selecting the best feature and splitting data continues until a complete decision tree is formed, encompassing all decisions and predictions. When new input data arrives, it traverses the decision tree starting from the root node to a leaf node. At each node, a functional test is performed, and the data is passed to one of the branches based on the test result until it reaches a leaf node that yields a label or prediction value.

Soft Voting Method

The Soft Voting ensemble method combines predictions from multiple models by assigning weights to these predictions based on the probabilities provided by each model. The final result is obtained by calculating the weighted average or maximum likelihood of the prediction probabilities. The first step is the collection of training datasets, which contain features and corresponding labels or categories. Next, several different classification models, such as Gaussian Naive Bayes and Decision Tree, are created. Each model is trained using the same training dataset.

The way the soft voting algorithm works is that the gaussian naïve bayes and decision tree algorithms make predictions from the same dataset. Give weight to the prediction probability of each model according to the level of confidence in the model. After weighting, combine the prediction results obtained for the final prediction by taking the average of the prediction results. The soft voting method works by predicting the probability of a particular class from the consideration of the weights obtained from each model.

*name of corresponding author

Evaluate the Results Using a Confusion Matrix

In classification testing use an evaluation method called a confusion matrix to assess the extent to which the created classification model can be utilized. The confusion matrix allows us to see the comparison between the classification results produced by the algorithm and the actual data in the form of a matrix table, as shown in Figure 2. Confusion Matrix.

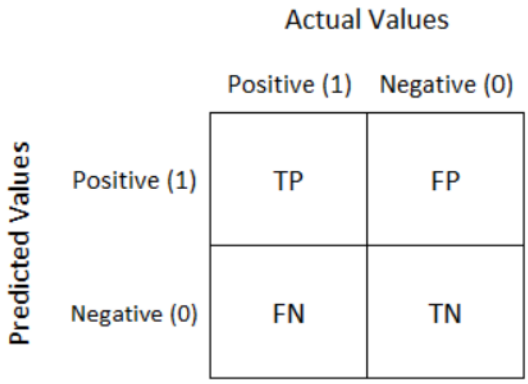

Figure 3. Confusion Matrix

By using a confusion matrix, we can measure the performance of a classification algorithm more comprehensively and gain a better understanding of how effective the model is in predicting the correct category or class for the given data. Within the confusion matrix, we can calculate the accuracy, precision, and recall values of the algorithm used. Accuracy represents the correct prediction rate of all the data used.

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

The precision value represents the true positive predictions compared to the total positive predictions.

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

The recall value represents the true positives divided by the total number of samples that should be positive.

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

RESULT

The use of machine learning in the process of classifying fertile and infertile soil uses the Gaussian naïve bayes algorithm and the decision tree algorithm, to improve the performance results of these algorithms, soft voting optimization is used by combining the two algorithms. As explained above, the research has data input in the form of 16 attributes where all of them are elements contained in the soil that affect fertility. The output of this research is to predict whether the soil is fertile or infertile with data consisting of 16 attributes. The final results of this study are described in the analysis of the best classification results using optimization, where the highest accuracy results are obtained than the use of a single model.

Gaussian Naïve Bayes Algorithm

The search for accuracy values using this algorithm should have already undergone data equalization and standardization to avoid errors when calculating accuracy. It is essential to ensure a balanced representation of fertile and infertile soil data. The algorithm's classification process involves taking the available attributes as input, and calculating the probability of membership in each soil

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

fertility class. The final step is to assign the label of the class with the highest probability as the classification decision. The Gaussian Naïve Bayes algorithm, in performing machine learning classification, takes in training data to learn from, which is then used to make predictions on previously unseen test data. The prediction results can be used to measure the model's performance and make decisions based on those predictions.

Table 5. Accuracy Gaussian Naive Bayes Algorithm

Algorithm	**Accuracy**
Gaussian Naïve Bayes	87.2%

Decision Tree Algorithm

In seeking accuracy values, this algorithm follows similar steps to Gaussian Naïve Bayes, where data equalization and standardization should have already been performed to avoid errors in accuracy calculation. This algorithm utilizes soil content attributes to construct a decision tree and make classification decisions by following a series of decisions based on features that yield the highest confidence level label classes, such as classifying soil as fertile or infertile. The decision tree algorithm in machine learning classification trains a decision tree model with training data and then uses it to make predictions on testing data. The prediction results can be used to assess the model's performance and make decisions based on those predictions.

Table 6. Accuracy Decision Tree Algorithm

Algorithm	**Accuracy**
Decision Tree	88.3%

Gaussian Naïve Bayes + Decision Tree + Soft Voting

After both algorithms have performed classification, we will combine them using the soft voting method. Soft Voting involves merging predictions from multiple machine-learning models and selecting the final predicted class based on the average probabilities of the models. By combining different algorithm models that may have different strengths and weaknesses, we aim to make overall predictions that are more robust and accurate.

Table 7.Accuracy Gaussian Naive Bayes + Decision Tree + Soft Voting

Algorithm	**Accuracy**
Gaussian Naïve Bayes + Decision Tree + Soft Voting	88.8%

It can be understood that this program creates an ensemble of models consisting of Decision Trees and Gaussian Naïve Bayes models, and then trains this ensemble using training data. Ensemble learning is frequently used to enhance model performance by combining different machine learning methods.

Comparison of algortithm usage and optimization additions

From each algorithm and the soft voting method, their performance can be evaluated using a confusion matrix, which can be observed in the figure 5.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13159 p-ISSN : 2541-044X

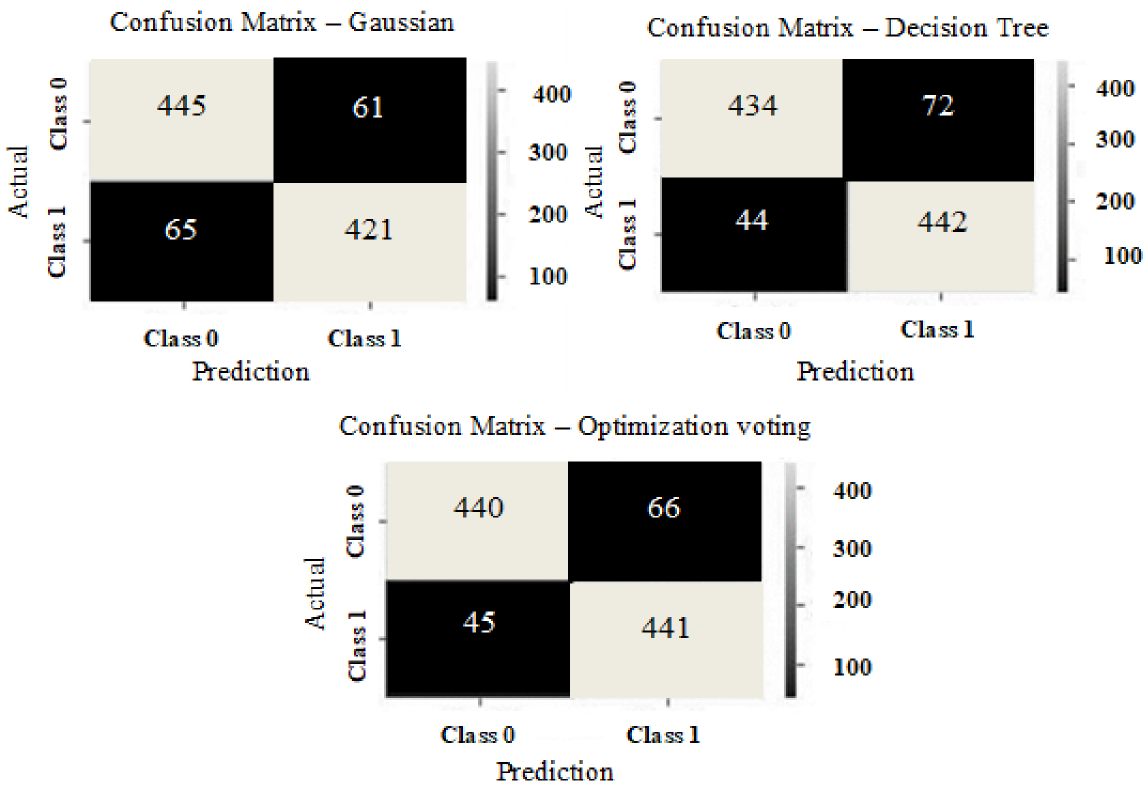

Figure 5. Confusion Matrix

The classification results using the Gaussian naïve bayes algorithm, decision tree algorithm, and also the use of soft voting optimization by combining the Gaussian naïve bayes and decision tree algorithms can be seen in table 5.

Table 8. Accuracy, Precision, and Recall

Algorithm	Gaussian Naïve Bayes	Decision Tree	Gaussian Naïve Bayes + Decision Tree + Soft Voting
Accuracy	87.2%	88,3%	88.8%
Precision	87.3%	85.9%	86.9%
Recall	86.6%	90.4%	90.7%

From the classification results in table 12, it can be seen that the classification results with the soft voting optimization process get the highest value compared to the use of a single algorithm model.

Analysis of the best classification results using optimization

The application of the soft voting method in this study provides the best results, because it obtains the highest accuracy, precision, and recall on fertile soil classification than the use of a single algorithm model without optimization. Based on the research and classification testing results, an accuracy value of 88.8% was achieved with the use of the Soft Voting optimization method. The accuracy of the Gaussian Naive Bayes and Decision Tree classification algorithms increased by 0.5%, with Gaussian Naïve Bayes achieving 87.2% accuracy and Decision Tree achieving 88.3% accuracy. Can be seen in Figure 6.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v9i1.13159 p-ISSN : 2541-044X

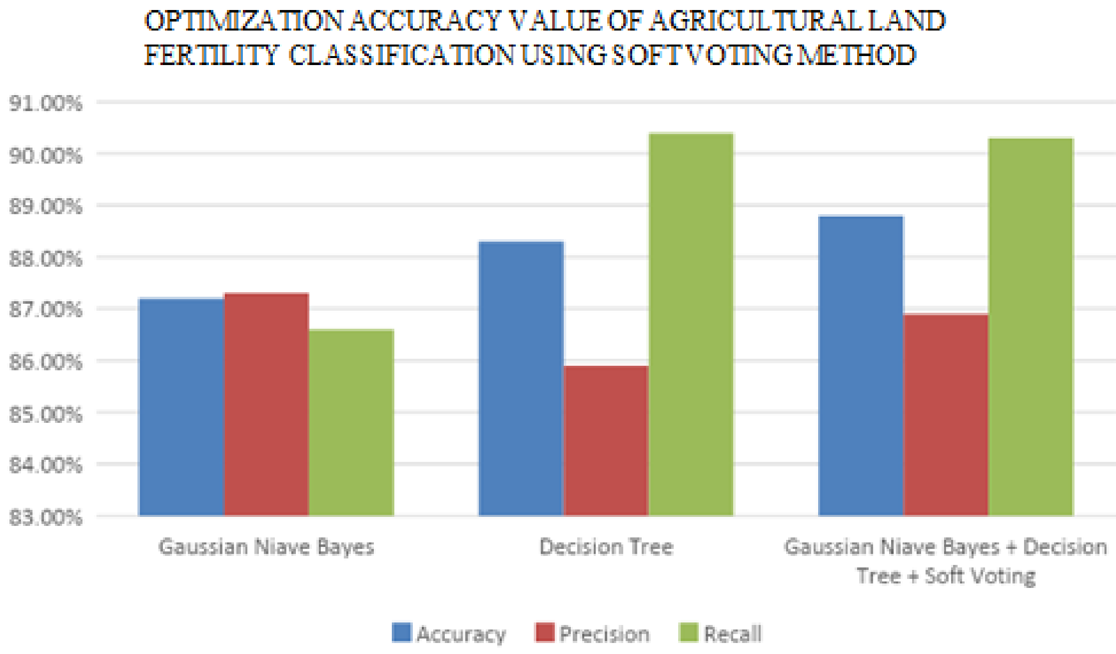

Figure 6. Comparison Accuracy, Precision, and Recall

DISCUSSIONS

Evaluation using the confusion matrix revealed different average precision and recall values for the Gaussian Naive Bayes and Decision Tree algorithms, as well as for the Soft Voting method. Based on these results, the addition of Soft Voting optimization to the Gaussian Naive Bayes and Decision Tree algorithms has an impact on accuracy, precision, and recall values. The use of the Naive Bayes algorithm is not effective when dealing with complex data, as this independence assumption can become unrealistic. In textual data, words are often interrelated and have complex contextual dependencies. Naive Bayes does not handle such dependencies well. The use of the Decision Tree algorithm is more effective when dealing with complex data because it is more flexible in handling highly complex data or closely related attributes. In this study, it is not relevant for highly complex data, and there is a performance anomaly between the Gaussian Naive Bayes and Decision Tree algorithms. This is a concern as it may affect the accuracy values in this research.

CONCLUSION

The research involves building a soil fertility classification modeling system using Gaussian Naïve Bayes and Decision Tree algorithms with the Soft Voting method. Based on the performance results, the accuracy rate of the Gaussian Naïve Bayes algorithm was found to be 87.2%, the Decision Tree algorithm had an accuracy rate of 88.3%, and the combined accuracy of both algorithms using the Soft Voting method reached 88.8%. This was achieved using a dataset comprising 5,000 data points, consisting of 2,520 fertile soil samples and 2,480 infertile soil samples. The findings of this research have a positive impact on the understanding of soil fertility and can be utilized by farmers and land managers to make more informed decisions regarding agricultural land management. Furthermore, this soft voting method can be applied in various other data analysis fields to enhance prediction accuracy. Thus, this research holds significant implications in supporting more efficient and sustainable agriculture and in advancing data analysis methods.

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

REFERENCES

Ahmim, A., Maglaras, L., Ferrag, M. A., Derdour, M., & Janicke, H. (2019). A novel hierarchical intrusion detection system based on decision tree and rules-based models. *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 228–233.

Athar, A., Ali, S., Sheeraz, M. M., Bhattachariee, S., & Kim, H.-C. (2021). Sentimental analysis of movie reviews using soft voting ensemble-based machine learning. *2021 Eighth International Conference on Social Network Analysis, Management and Security (SNAMS)*, 1–5.

Denny, A., Raj, A., Ashok, A., Ram, C. M., & George, R. (2019). i-hope: Detection and prediction system for polycystic ovary syndrome (pcos) using machine learning techniques. *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, 673–678.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241–258.

Handayani, P., Nurlelah, E., Raharjo, M., & Ramdani, P. M. (2019). Prediksi Penyakit Liver Dengan Menggunakan Metode Decision Tree dan Neural Network. *CESS (Journal of Computer Engineering, System and Science)*, *4*(1), 55–59.

Humbird, K. D., Peterson, J. L., & McClarren, R. G. (2018). Deep neural network initialization with decision trees. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(5), 1286–1295.

Islam, R., & Shahjalal, M. A. (2019). Soft voting-based ensemble approach to predict early stage DRC violations. *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, 1081–1084.

Jayachitra, S., & Prasanth, A. (2021). Multi-feature analysis for automated brain stroke classification using weighted Gaussian naïve Bayes classifier. *Journal of Circuits, Systems and Computers*, *30*(10), 2150178.

Kamel, H., Abdulah, D., & Al-Tuwaijari, J. M. (2019). Cancer classification using gaussian naive bayes algorithm. *2019 International Engineering Conference (IEC)*, 165–170.

Karlos, S., Kostopoulos, G., & Kotsiantis, S. (2020). A soft-voting ensemble based co-training scheme using static selection for binary classification problems. *Algorithms*, *13*(1), 26.

Kaunang, F. J. (2018). Penerapan algoritma J48 decision tree untuk analisis tingkat kemiskinan di Indonesia. *Cogito Smart Journal*, *4*(2), 348–357.

Kumari, S., Kumar, D., & Mittal, M. (2021). An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier. *International Journal of Cognitive Computing in Engineering*, *2*, 40–46.

Matloob, F., Ghazal, T. M., Taleb, N., Aftab, S., Ahmad, M., Khan, M. A., Abbas, S., & Soomro, T. R. (2021). Software defect prediction using ensemble learning: A systematic literature review. *IEEE Access*, *9*, 98754–98771.

Pasha, M. R., Hidayat, R. R., & Abas, M. I. (2023). Implementasi Decision Tree C4. 5 dalam Memilih Perguruan Tinggi Pendamping Program SMK Pusat Keunggulan. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *3*(6), 1129–1139.

Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74–78.

Pratama, A., Wicaksana, A. A., & Razi, A. (2022). Analisa Kesesuaian Lahan Tanah Untuk Tanaman Padi (Oryza Sativa L.) Dengan Metode Decision Tree Berbasis Web (Studi Kasus Kabupaten Aceh Utara). *Jurnal Informatika Kaputama (JIK)*, *6*(1), 1–23.

Rafique, A. A., Jalal, A., & Ahmed, A. (2019). Scene understanding and recognition: statistical segmented model using geometrical features and Gaussian naïve bayes. *IEEE Conference on International Conference on Applied and Engineering Mathematics*, *57*.

Salur, M. U., & Aydın, İ. (2022). A soft voting ensemble learning-based approach for multimodal sentiment analysis. *Neural Computing and Applications*, *34*(21), 18391–18406.

Sanjay, K. S., & Danti, A. (2019). Detection of fake opinions on online products using Decision Tree and Information Gain. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, 372–375.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 1, January 2024
DOI : https://doi.org/10.33395/sinkron.v9i1.13159
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Saqlain, M., Jargalsaikhan, B., & Lee, J. Y. (2019). A voting ensemble classifier for wafer map defect patterns identification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, *32*(2), 171–182.

Setiawati, I., Permana, A., & Hermawan, A. (2019). Implementasi Decision Tree Untuk Mendiagnosis Penyakit Liver. *Journal of Information System Management (JOISM)*, *1*(1), 13–17.

Taha, A. (2021). Intelligent ensemble learning approach for phishing website detection based on weighted soft voting. *Mathematics*, *9*(21), 2799.

Valentinus, F., Sujono, F., Ariansyah, I., & Capah, D. A. H. (2023). Implementatio Of Data Mining With Classification And Forecasting Method Use Model Gaussian Naive Bayes For Building Store (Studi Case: Tb Sinar Jaya). *Jurnal Teknik Informatika (Jutif)*, *4*(2), 413–420.

Verma, R., & Chandra, S. (2023). RepuTE: A soft voting ensemble learning framework for reputation-based attack detection in fog-IoT milieu. *Engineering Applications of Artificial Intelligence*, *118*, 105670.

Wibowo, R., Soeleman, M. A., & Affandy, A. (2023). Hybrid Top-K Feature Selection to Improve High-Dimensional Data Classification Using Naïve Bayes Algorithm. *Scientific Journal of Informatics*, *10*(2).

Zhou, K., Yang, Y., Qiao, Y., & Xiang, T. (2021). Domain adaptive ensemble learning. *IEEE Transactions on Image Processing*, *30*, 8008–8018.

*name of corresponding author