JSCE (Journal of System and Computer Engineering)
Vol.6, No.3, July 2025, pp. 211-219
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i3.1955

Crop Recommendation Based on Soil and Weather Conditions Using the K-Nearest Neighbors Algorithm

Yuliyanto [1,a,*]; Supriadi Sahibu [2,b]; Taufik Imran [3,c]; Andriansyah Oktafiandi Arisha [4,d]; Munawirah [5,e];

[1,2,3] Universitas Handayani Makassar, Jl. Adiyaksa Baru, Makassar and 90231, Indonesia
[4] Teknik Informatika, AMIK Tridharma Palu, Jl. Undata No. 10, Palu and 94111, Indonesia
[5] Sistem Informasi, Universitas Tomakaka, Jl. Ir Juanda, Mamuju and 91515, Indonesia
[a] uhmyuliyanto@gmail.com; [b] supriadi@handayani.ac.id; [c] imran_taufik@handayani.ac.id; [d] ecpand@gmail.com; [e] munawirahkadir@gmail.com;*
** Corresponding author*

Abstract

The national food self-sufficiency program demands innovation in optimizing the selection of agricultural commodities based on environmental and weather conditions. This challenge is rooted in a fundamental problem faced by farmers—achieving harmony among soil characteristics, weather patterns, and suitable crops. In support of this initiative, it is necessary to develop a crop recommendation system based on machine learning that utilizes key soil and weather condition parameters. This study employs the K-Nearest Neighbors (KNN) algorithm, which functions by identifying the optimal value of 'K' to maximize classification accuracy. The KNN algorithm is implemented in a crop recommendation system to classify 1,100 datasets representing ideal growing conditions for 11 crop types. These datasets were generated using a normal distribution approach with a 5% variation from the mean values, and were validated using a clipping function to ensure the data remained within ideal ranges. The results of this study demonstrate that the KNN algorithm achieves high accuracy 96,67% in utilizing soil and weather parameters to generate crop recommendations. The average probability score for the recommended crops was 83.33%. Based on experimental testing, rice was recommended during the rainy and extreme rainy seasons, soybeans were recommended during the dry season, and mung beans were most suitable during extreme dry conditions

Keywords *: Crop Recommendation, K-Nearest Neighbors, Environmental Parameters*

1. Introduction

Agriculture is currently facing major challenges in improving productivity while maintaining environmental sustainability (Jat et al., 2020). Understanding soil and crop requirements can increase efficiency by up to 40%, through precise fertilization and maintaining optimal crop productivity (Bhargava et al., 2021). One way to achieve this goal is by selecting suitable crops to be planted based on soil conditions, climate, and other environmental factors. Therefore, a system capable of providing efficient crop recommendations by considering various dynamic environmental factors is needed (Sharma et al., 2022).

Machine learning-based systems have great potential in processing large-scale data, such as soil and weather sensor data, and using algorithms to generate more accurate crop selection recommendations (Ali et al., 2023). These machine learning-based recommendation systems have the potential to improve prediction accuracy in selecting appropriate crops. Previous studies have shown that machine learning algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest have been used to predict agricultural yields and determine crops suitable for local environmental conditions (Kumar & Singh, 2022).

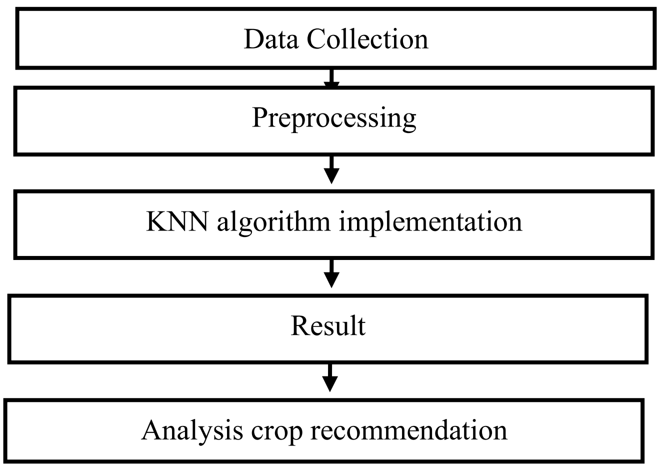
Research on agricultural recommendation systems has been conducted by Gosai and Raval, who discussed a crop recommendation system using machine learning. Their study compared six algorithms: Decision Tree, Naïve Bayes, Support Vector Machine, Logistic Regression, Random Forest, and XGBoost (Gosai & Raval, 2021). Another relevant study was conducted by Suresh and Kumar, which focused on crop prediction and recommendation using the Support Vector Machine (SVM) algorithm, based on soil nutrient content as the recommendation basis (Suresh & Kumar, 2022).

Several variables are used in this study. Nitrogen is essential for protein and chlorophyll formation. Soil with nitrogen content below 10 ppm is considered too low for crops and leads to growth failure (Tiwari et al., 2019). Nitrogen levels above 150 ppm can cause nutrient imbalances and toxicity in plants, resulting in excessive vegetative growth that reduces yields (Meena et al., 2021). Phosphorus supports root development and flowering. Phosphorus content below 5 ppm prevents proper plant growth (Barman et al., 2020). Phosphorus levels above 60 ppm may hinder the absorption of other micronutrients and inhibit growth (Verma et al., 2021). Potassium is vital for stomatal regulation and plant resistance. Soil with potassium below 50 ppm is unsuitable for crops as it causes weak stems and damaged leaves (Pandey & Tripathi, 2022). Potassium above 400 ppm can cause nutrient imbalances (Reddy et al., 2020). Soil pH affects nutrient availability. Soil with a pH below 4.0 contains toxic levels of Al and Fe that inhibit plant growth (Dwivedi et al., 2019). Soil with a pH above 8.0 reduces the availability of micronutrients and hinders plant growth (Singh & Jha, 2021). The optimal temperature for tropical crops is between 22–32°C. Temperatures below 10°C and above 40°C can cause stress and plant death (Patel et al., 2022). The ideal humidity for tropical crops is between 60–85%. Humidity below 40% causes leaf wilting, while humidity above 90% increases the risk of fungal diseases (Malik et al., 2023). An annual rainfall of at least 500 mm is required for plant growth; below this level, plants will suffer from drought (Deshmukh et al., 2022). Rainfall above 3000 mm can cause flooding and root damage (Joshi et al., 2021).

To ensure novelty in this research, the KNN algorithm is used to provide efficient crop recommendations based on a machine learning approach by leveraging soil and weather condition data. The K-Nearest Neighbors (KNN) algorithm works by identifying proximity between data points, has the ability to handle multidimensional data, and offers flexibility in matching existing data patterns (Kaur & Kaur, 2023). Therefore, KNN is selected in this study to provide crop recommendations for farmers, helping them determine the most suitable crops according to prevailing soil and weather conditions.

2. Method

This study employs a quantitative approach, and the type of research is experimental in nature, where the scope of the problem is addressed through library research, field data collection methods, as well as system design and testing. Review Process.



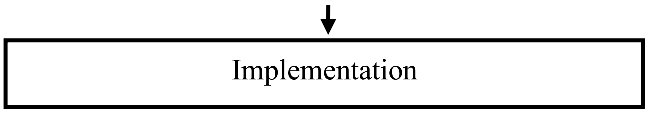


Figure 1. *Research phases.*

The model in this study was developed using the K-Nearest Neighbors (KNN) algorithm implemented in the Python 3.10 programming language. Key libraries utilized include NumPy for numerical operations, Pandas for data manipulation, and Matplotlib for visualization. The KNN algorithm was trained and evaluated using the Scikit-learn library. To facilitate real-time field testing and user interaction, a graphical user interface was built using Streamlit. All experiments were conducted on a standard personal computer without GPU acceleration.

3. Results And Discussion

The data collection method was carried out by obtaining data from various sources. All datasets were compiled based on validated thresholds or the ideal requirements for crops, derived from official and accessible sources to ensure data accuracy.

3.1 Data Set Description

A total of 1,100 datasets were used in this study, which were categorized into two parts: 80% as training data and 20% as testing data. The dataset consists of seven variables: Nitrogen (ppm), Phosphorus (ppm), Potassium (ppm), Soil pH, Temperature (°C), Humidity (%), and Rainfall (mm/month).

Table 1. Ideal Range of NPK Requirements

Crop	Nitrogen-Min	Nitrogen-Max	Phosphorus -Min	Phosphorus -Max	Potassium -Min	Potassium -Max	Soil pH-Min	Soil pH -Max
Rice	80	150	15	30	60	120	5	6,5
Maize	100	150	20	35	80	130	5,5	6,8

Table 2. Ideal Range of Weather Requirements

Crop	Temperature -Min	Temperature -Max	Humidity - Min	Humidity - Max	Rainfall - Min	Rainfall -Max
Rice	20	30	70	90	150	300
Maize	24	32	65	85	120	200

3.2 Pre-processing

The performance of classification techniques is highly influenced by the training dataset. The entire dataset was constructed based on the ideal range of each crop's requirements. These values were then subjected to small random variations following a normal distribution to simulate realistic field conditions (e.g., variations in nitrogen levels or daily temperature). This variation was constrained within reasonable limits using a parameter of variation = 0.05 (i.e., ±5%). To ensure compliance with the ideal range during dataset generation, the np.clip function in Python was employed to explicitly restrict minimum and maximum values within safe bounds.

Table 3. Sample Dataset for Crop Data

Nitrogen	Phosphorus	Potassium	Temperature	Humidity	pH	Rainfall	Label
103,4	19,8	78,7	24,8	81,9	5,4	210,1	Rice
127,6	24,7	105,3	27,6	77,0	5,7	155,1	Maize
77,7	20,3	71,5	27,3	68,2	5,9	158,2	Soybean
95,9	30,9	98,2	30,5	62,9	6,2	125,6	Melon

3.3 KNN Implementation

Following the data preprocessing stage, the next step involved implementing the K-Nearest Neighbors (KNN) algorithm as the primary method in the crop recommendation system. The KNN algorithm operates on the principle that data points with similar characteristics are located close to each other in the feature space. After performing cross-validation for each value of K, the average accuracy was calculated and visualized using a graph. This graph illustrates the relationship between the K values and the average accuracy obtained from the cross-validation results.

Once the model was trained using the optimal K value determined from the previous step, the next phase was to evaluate the model's performance. One commonly used evaluation method in classification tasks is the confusion matrix. The confusion matrix provides a comprehensive overview of the model's prediction results by comparing the predicted labels with the actual ground truth labels from the test dataset.

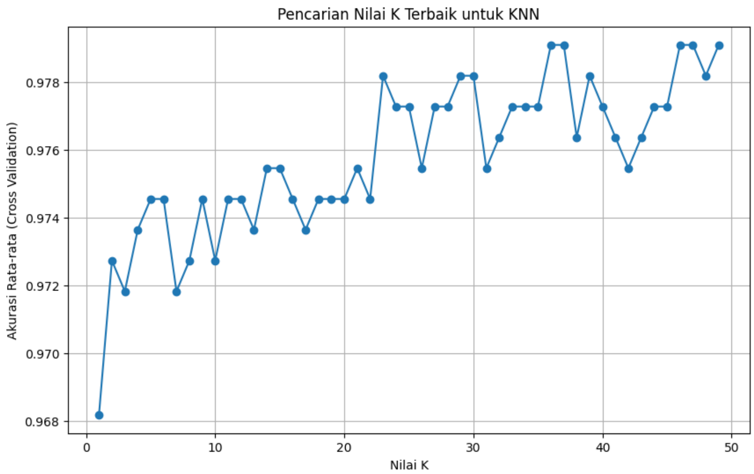


Figure 2. *Best* K.

The optimal selection of the *K* parameter in the *K-Nearest Neighbors* (KNN) algorithm was performed through a cross-validation process. This cross-validation technique is essential to determine the most suitable *K* value, which in this study employed the 10-fold cross-validation method. Cross-validation is widely recommended in machine learning model evaluation to enhance the reliability of results (Kuhn & Johnson, 2013). The *K* values were tested in the range of 1 to 15, with each value evaluated based on the average accuracy on the validation data. The experimental results revealed that *K* = 36 yielded the best performance, achieving the highest accuracy of 96.67%. Therefore, this value was selected as the final configuration for the model. The use of cross-validation in this context aims to improve model generalization and minimize the risk of overfitting to the training data.

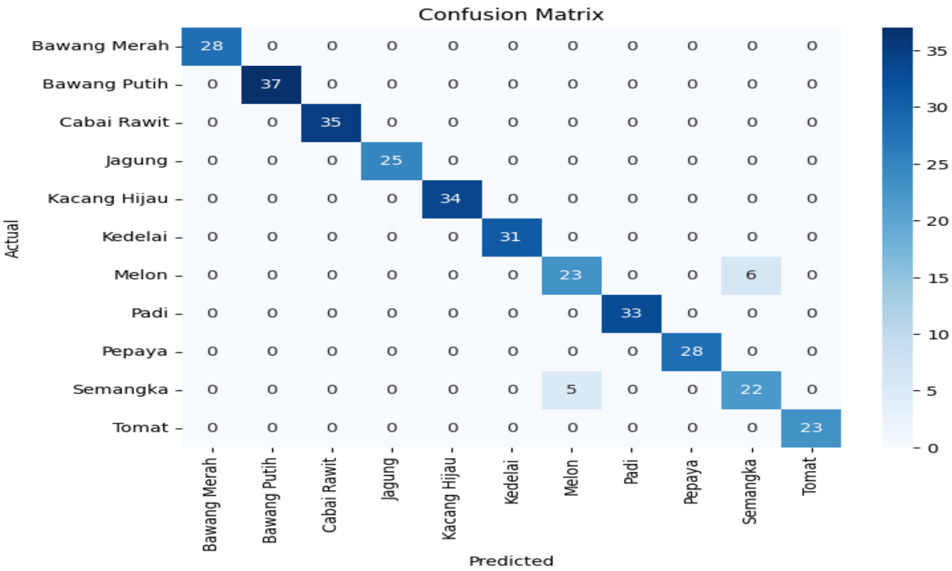


Figure 3. *Confusion Matrix*

```
=== Classification Report Gabungan ===
              precision    recall  f1-score  support  accuracy_per_class
Bawang Merah   1.000000  1.000000  1.000000     28.0            1.000000
Bawang Putih   1.000000  1.000000  1.000000     37.0            1.000000
Cabai Rawit    1.000000  1.000000  1.000000     35.0            1.000000
Jagung         1.000000  1.000000  1.000000     25.0            1.000000
Kacang Hijau   1.000000  1.000000  1.000000     34.0            1.000000
Kedelai        1.000000  1.000000  1.000000     31.0            1.000000
Melon          0.821429  0.793103  0.807018     29.0            0.793103
Padi           1.000000  1.000000  1.000000     33.0            1.000000
Pepaya         1.000000  1.000000  1.000000     28.0            1.000000
Semangka       0.785714  0.814815  0.800000     27.0            0.814815
Tomat          1.000000  1.000000  1.000000     23.0            1.000000

Akurasi Model Keseluruhan: 96.67%
```

Figure 4. *Classification report*

The confusion matrix and classification report of the KNN model illustrate classification performance across 11 crop classes. Most crops, including shallot, garlic, chili, maize, mung bean, soybean, rice, papaya, and tomato, were classified with perfect precision, recall, and F1-score (1.000). However, bidirectional misclassification occurred between melon and watermelon, as evidenced in the confusion matrix and their lower F1-scores (0.87 and 0.85, respectively). These misclassifications are likely due to overlapping agronomic features. Overall, the model achieved a high accuracy of 96.67%, demonstrating strong predictive performance with minor errors in closely related crop classes.

3.4 Result

This testing phase represents one of the final stages for evaluating the performance of the classification model used in the crop recommendation system based on weather data and soil conditions. The testing was conducted using a set of input parameters: Nitrogen (N), Phosphorus (P), Potassium (K), Temperature (S), Humidity (KL), Soil pH, and Rainfall (CH), to determine the resulting crop recommendations. The tests were carried out across three soil types Latosol, Alluvial, and Regosol under four seasonal conditions: Rainy Season, Dry Season, Extreme Rainy Season, and Extreme Dry Season. These soil types are commonly found in tropical regions such as Indonesia

Table 4. Results

No	Soil Types	Parameter Input							Crop Recommendation					
		N	P	K	S	KL	pH	CH	Crop 1	%	Crop 2	%	Crop 3	%
1	Latosol	101,55	23,46	62,22	24	75,8	5,66	318,8	Rice	100	-	-	-	-
2	Latosol	68,34	17,46	75,93	30,6	65,8	5,65	147,3	Soybean	61.11	Mung Bean	38.89	-	-
3	Latosol	84,51	16,72	83,37	25,9	91,8	5,04	505,7	Rice	91.67	Papaya	8.33	-	-
4	Latosol	65,76	10,76	54,77	31,2	55,2	6,06	26,1	Mung Bean	100	-	-	-	-
5	Aluvial	96,97	20,15	62,59	24,5	83,1	5,95	318,8	Rice	83.33	Papaya	16.67	-	-
6	Aluvial	83,35	16,63	58,6	30	61,9	6,41	147,3	Soybean	61.11	Mung Bean	38.89	-	-
7	Aluvial	94,19	14,24	79,26	24,8	85,7	5,3	505,7	Rice	91.67	Papaya	8.33	-	-
8	Aluvial	71,16	14,77	67,7	32,8	50	6,3	26,1	Mung Bean	100	-	-	-	-
9	Regasol	86,98	20,09	73,69	23	77,9	6,08	318,8	Rice	66.67	Shallot	22.22	Papaya	11.11
10	Regasol	63,25	17,12	50,8	30	68,8	5,55	147,3	Soybean	77.78	Mung Bean	22.22	-	-
11	Regasol	74,3	17,79	70,41	24	91,7	5,81	505,7	Rice	66.67	Papaya	25	Shallot	8.33
12	Regasol	63,89	12,77	62,62	31,7	49,8	5,88	26,1	Mung Bean	100	-	-	-	-
									Average	83.33	Average	15.05	Average	1.62

3.5 Analysis Recommendation

This analysis aims to evaluate the model's ability to classify the most suitable crop types based on the given input parameters. In addition, the recommended crops can be assessed in terms of their suitability for cultivation across different soil types and seasonal conditions (Latosol (L), Aluvial (A), Regosol (R) anda Rainy Season, Dry season, Extreme Rainy, Extreme Dry) .

Table 5. Analysis Recommendation

No	Crop	Soil Types			Season				J	%	Description
		L	A	R	Rainy	Dry	Extreme Rainy	Extreme Dry			
1	Rice	✓	✓	✓	✓	-	✓	-	5	71.4	Rice is suitable for cultivation in all soil types and during the rainy and extreme rainy seasons.
2	Soybean	✓	✓	✓	-	✓	-	-	4	57.1	Soybean is suitable for cultivation in all soil types during the dry season.
3	Mung bean	✓	✓	✓	-	✓	-	✓	5	71.4	Mung bean is suitable for cultivation in all soil types during the dry and extreme dry seasons.
4	Shallot	-	-	✓	✓	-	✓	-	3	42.9	Shallot is suitable for cultivation in Regosol soil during the rainy and extreme rainy seasons.
5	Papaya	✓	✓	✓	✓	-	✓	-	5	71.4	Papaya is suitable for cultivation in all soil types during the rainy and extreme rainy seasons.

3.6 Discussion

The K-Nearest Neighbors (KNN) algorithm employed in this study achieved a classification accuracy of 96.67%. The optimal value of K was determined to be 36 based on evaluation using 10-fold cross-validation. This result highlights the novelty of the research by positioning the model competitively compared to previous studies. The study by Gosai and Raval (2021) compared six different machine learning algorithms, where XGBoost emerged as the best-performing model with an accuracy of only 91.34%. Although the study offered a comprehensive comparison, it did not apply cross-validation techniques, which may have limited the reliability of the reported performance metrics. Meanwhile, the research by Suresh and Kumar (2022) implemented the Support Vector Machine (SVM) algorithm and achieved an accuracy of 91.15%, relying solely on soil nutrient data. However, that study did not consider environmental variables such as temperature and rainfall, which are critical factors influencing crop productivity in tropical regions. In contrast, the present study incorporates both soil nutrient parameters (N, P, K, and pH) and climatic variables such as temperature, humidity, and rainfall. With a simple yet efficient KNN approach and a rigorous cross-validation procedure, the model demonstrates higher predictive accuracy and is well-suited for application in crop recommendation systems tailored to tropical agro-environments.

Despite the strong performance of the model, several limitations must be acknowledged. One of the primary constraints is the model's sensitivity to the selection of the K value. A small K may lead to overfitting, while a large K could result in underfitting. In this study, the optimal $K = 36$ was determined through systematic evaluation to achieve the best classification accuracy. It is also important to note that the dataset used to train the model primarily reflects tropical environmental conditions. This may limit the model's generalizability when applied to regions with temperate or arid climates. Future studies should consider expanding the dataset to enhance the model's flexibility across diverse agro-ecological zones.

The confusion matrix analysis revealed bidirectional misclassification between **melon** and **watermelon**, where six instances of melon were misclassified as watermelon, and five watermelon samples were classified as melon. This error likely stems from the high degree of agronomic similarity between the two crops, including comparable nitrogen requirements, soil pH, temperature, and humidity preferences. Since KNN relies on distance-based similarity, the algorithm struggles to differentiate between classes with overlapping feature distributions. Moreover, the current system does not account for dynamic factors such as soil type variation, seasonal crop rotation, or economic considerations like market price and profit margins. These factors play a crucial role in real-world agricultural decision-making. Future work is recommended to incorporate temporal and economic dimensions to improve the model's contextual relevance and adaptability. In conclusion, while the KNN-based crop recommendation model developed in this study demonstrates high predictive accuracy and computational efficiency, it also presents several limitations that offer valuable directions for future improvement and the development of a more context-aware, comprehensive recommendation system.

4. Conclusions

The conclusions experimental results of this study demonstrate that the KNN algorithm achieves high accuracy 96,67% in utilizing soil and weather parameters to generate crop recommendations. The average probability score for the recommended crops was 83.33%. Based on experimental testing, rice was recommended during the rainy and extreme rainy seasons, soybeans were recommended during the dry season, and mung beans were most suitable during extreme dry conditions

References

Ali, R., Khan, M., & Zafar, N. (2023). Application of machine learning in precision agriculture. Computers and Electronics in Agriculture, 198, 107059.

Barman, A., Roy, D., & Paul, S. (2020). Effect of phosphorus on growth and development of crop plants. Agricultural Sciences, 11(2), 123–129.

Bhargava, M., Sharma, P., & Choudhary, R. (2021). Enhancing fertilizer use efficiency in crops: A review. Agronomy Research, 19(4), 1234–1246.

Deshmukh, S., Kale, R., & Patil, R. (2022). Impact of rainfall variability on crop production in India. Journal of Climatology Studies, 12(3), 78–85.

Dwivedi, R., Srivastava, A., & Pathak, H. (2019). Soil pH and nutrient availability in acidic soils. Indian Journal of Soil Science, 64(1), 45–50.

Gosai, S., & Raval, D. (2021). Crop Recommendation System Using Machine Learning. International Journal of Computer Applications, 183(17), 1–6.

Jat, R., Yadav, A., & Rathore, R. (2020). Sustainable agricultural practices to increase productivity. International Journal of Environmental Science, 8(2), 75–82.

Joshi, M., Sharma, A., & Negi, S. (2021). Effect of excessive rainfall on soil and plant health. Soil & Water Research, 16(2), 89–97.

Kaur, G., & Kaur, P. (2023). KNN algorithm and its application in agricultural recommendation. Journal of Intelligent Systems, 32(4), 212–218.

Kumar, A., & Singh, R. (2022). Comparative analysis of ML models in crop prediction. Journal of Agricultural Informatics, 13(1), 34–40.

Malik, S., Khan, F., & Qureshi, A. (2023). Impact of humidity on fungal disease incidence in crops. Plant Protection Journal, 35(2), 110–117.

Meena, R., Lal, M., & Dey, A. (2021). Nitrogen management for sustainable crop productivity. Advances in Agronomy, 170, 95–112.

Pandey, S., & Tripathi, R. (2022). Role of potassium in improving plant health. Indian Journal of Agronomy, 67(4), 336–344.

Patel, K., Raval, J., & Desai, V. (2022). Temperature stress in tropical crops: Physiology and adaptation. Journal of Plant Research, 38(1), 56–66.

Reddy, B., Rao, G., & Prasad, Y. (2020). Soil potassium and its impact on crop yields. Journal of Soil Fertility, 6(3), 89–96.

Sanchez, P. A. (1976). Properties and Management of Soils in the Tropics.

Wiley. Sharma, V., Mehta, N., & Chandel, A. (2022). Smart farming using AI and IoT: A review. Artificial Intelligence in Agriculture, 5, 34–45.

Singh, N., & Jha, S. (2021). Alkaline soils and micronutrient deficiency in crops. Journal of Environmental Agronomy, 10(3), 140–146.

Suresh, A., & Kumar, P. (2022). SVM-based crop recommendation using soil nutrients. Journal of Machine Learning in Agriculture, 4(2), 98–106.

Tiwari, V., Singh, K., & Yadav, L. (2019). Low nitrogen effects on crop productivity. International Journal of Plant & Soil Science, 27(4), 229–235.

Verma, S., Bharti, R., & Chauhan, N. (2021). Phosphorus toxicity and crop growth inhibition. Agronomy Updates, 9(1), 62–70.

Gosai, H., & Raval, P. (2021). Crop recommendation system using machine learning algorithms. International Journal of Advanced Research in Computer and Communication Engineering, 10(2), 54–58.

Suresh, K., & Kumar, M. (2022). Prediction and recommendation of crops using support vector machine. Journal of Artificial Intelligence and Data Mining, 10(1), 17–24.

Havlin, J. L., Tisdale, S. L., Nelson, W. L., & Beaton, J. D. (2013). Soil fertility and fertilizers: An introduction to nutrient management (8th ed.). Pearson Education.

Brady, N. C., & Weil, R. R. (2008). The nature and properties of soils (14th ed.).

Pearson Prentice Hall. Sanchez, P. A. (1976). Properties and management of soils in the tropics. Wiley-Interscience.

FAO. (2006). Plant nutrition for food security: A guide for integrated nutrient management. Food and Agriculture Organization of the United Nations.

Huang, B., & Fry, J. D. (1998). Soil temperature effects on plant growth and physiology. In M. Pessarakli (Ed.), Handbook of Plant and Crop Physiology (pp. 215–230). Marcel Dekker.

Nugroho, H. A., & Purwandari, B. (2020). Sistem rekomendasi tanaman berdasarkan klasifikasi tanah menggunakan metode Naïve Bayes. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 4(3), 456–462.

Prasetyo, E., & Handayani, R. (2021). Implementasi algoritma KNN untuk sistem rekomendasi tanaman hortikultura berbasis kondisi tanah dan cuaca lokal. Jurnal Teknik Pertanian Lampung, 10(2), 73–82.

Wibowo, M., & Ramadhani, R. (2021). Perbandingan metode klasifikasi data mining untuk rekomendasi tanaman pangan. Jurnal Media Informatika Budidarma, 5(3), 913–921.