# Toward Crops Prediction in Indonesi

*by* Arbi Haza Nasution

# Toward Crops Prediction in Indonesia

**Prima Wahyu Titisari, Arbi Haza Nasution, Elfis, and Winda Monika**

**Abstract** Agriculture's role and contributions in the modern era of globalization are crucially important. In order to meet the ever-increasing demands of the world's population, agriculture has faced a variety of obstacles over the years. On the other hand, climate change will impact crop growth, yields, and agricultural production. Recently, numerous researchers have introduced numerous machine learning models to address problems in diverse fields, including agriculture. Advancements in machine learning and crop simulation modeling have created new opportunities to enhance agricultural prediction. These technologies have each provided distinctive capabilities and substantial improvements in prediction performance; however, they have been primarily evaluated separately, and there may be advantages to integrating them to further improve prediction accuracy. The purpose of this study is to forecast future crops in Indonesia based on temperature, precipitation, soil condition, and humidity dataset. This study utilizes Indian dataset, as Indonesian dataset is not yet available. The dataset consists of 2200 records from 22 plant species. We compare four machine learning algorithms which are decision tree, support vector machine, random forest, and K-nearest neighbor with accuracy as evaluation metric. The result shows that random forest model can predict the suitable crop to be planted on specific

P. W. Titisari (✉) · Elfis
Department of Agrotechnology, Faculty of Agriculture, Universitas Islam Riau, Pekanbaru 28284, Indonesia
e-mail: pw.titisari@edu.uir.ac.id

Elfis
e-mail: elfisuir@edu.uir.ac.id

A. H. Nasution
Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Riau, Pekanbaru 28284, Indonesia
e-mail: arbi@eng.uir.ac.id

W. Monika
Department of Library Science, Faculty of Humanities, Universitas Lancang Kuning, Pekanbaru 28266, Indonesia
e-mail: windamonika@unilak.ac.id



P. K. Pattnaik et al. (eds.), *Proceedings of 3rd International Conference on Smart Computing and Cyber Security*, Lecture Notes in Networks and Systems 914,
https://doi.org/10.1007/978-981-97-0573-3_17

soil and weather conditions with the accuracy of 0.989697. The next step is to implement the model with Indonesian dataset. A plan optimization to choose the best set of plant species to produce or obtain will be our next challenge.

**Keywords** Agriculture · Crops · Machine learning

## 1 Introduction

The effects of global climate change, such as varying rainfall intensity, duration, and frequency; extreme weather; rising temperatures; significant variations in solar radiation; and rising greenhouse gas emissions, can have an impact on agricultural, forest, and other natural resources [1, 2]. Climate change will affect crop growth, yields, and production in the agricultural sector as a result of an increase in drought and flood events, which will have indirect effects on economic stability, although the effects will vary by region and crop type [3, 4]. Moreover, under climate change scenarios, developing nations such as Indonesia will suffer more than developed nations due to agricultural production strategies driven by economic plans. In 2030–2050, the global temperature will increase by 2–3 °C, and temperature increases of 2 °C or more are expected to reduce the yields of global staple crops such as rice, maize, and wheat [5–7]. In addition, climate change has caused significant shifts in planting and harvesting dates, which has altered the growing season due to variations and uncertainties in precipitation and temperature, thereby affecting food demand.

Food security and sustainable development continue to be the primary objectives of the agricultural sector in Indonesia. Appropriate technologies and institutional innovations can help Indonesia achieve food security. Consequently, the transformation of technological innovations will continue to be a key driver of future agricultural growth, including the increased use of crop varieties, machinery, and land/institutional reforms. Currently, mechanical technology is undergoing a rapid evolution, and a vast array of technologies are available. Existing technologies, such as remote sensing [8], robotic platforms [9], and the Internet of Things (IoT) [10], have recently become pervasive in industry, especially in the agricultural sector, resulting in smart and efficient farming [11, 12].

Precision agriculture is a trend nowadays. Precision agriculture is a modern farming technique that uses the data of soil characteristics, soil types, crop yield data, and weather conditions and suggests the farmers with the most optimal crop to grow in their farms for maximum yield and profit. This technique can reduce crop failures and will help the farmers to take informed decisions about their farming strategy.

In order to mitigate the agrarian crisis in the current status quo, there is a need for better recommendation systems to alleviate the crisis by helping the farmers to make an informed decision before starting the cultivation of crops. The technique of machine learning is the foundation of an automated crop mapping system. The system also calculates the daily national aggregation of village areas. On average, the system can upload updated information from regulatory agencies to cloud storage for user-friendly access every day. According to Schmidhuber [13], deep learning (DL) is a contemporary approach that is being utilized effectively in various machine learning techniques. It is comparable to artificial neutral networks (ANNs), but with enhanced learning capabilities; consequently, its accuracy is higher [14]. In recent years, DL technologies such as generative adversarial networks (GANs), recurrent neutral networks (RNNs), and convolutional neutral networks (CNNs) have been widely implemented and studied in various fields of research, including agriculture and farming. Agriculturalists and researchers frequently utilize a variety of software systems without evaluating the mechanisms and concepts of the techniques, such as GANs, RNNs, and CNNs, that are typically implemented in DL algorithms [15]. Morales and Villalobos [16] simulated sunflower and wheat data for 2001–2020 in five regions of Spain using OilcropSun and CERES-Wheat from DSSAT, and [17] attempts to demonstrate the feasibility of applying deep learning (DL) algorithms to the agricultural sector. Shahhosseni et al. [18] investigated whether combining crop modeling and machine learning (ML) improves predictions of corn yield in the US Corn Belt. The result indicates that weather data alone is insufficient, and that ML models require additional hydrological inputs to generate more accurate yield predictions. We plan to predict future agricultural crops in Indonesia based on datasets of temperature, precipitation, soil conditions, and humidity, using a decision tree, support vector machine, random forest, and KNN.

## 2 Method

Using machine learning can aid in making accurate predictions. Consequently, the analysis was conducted utilizing one of the machine learning techniques, supervised learning, in which labeled data from the past was utilized. The employed methodology consists of five stages. The first step describes the data collection process used to recommend plants for suitable soils. In the second phase, a description of the library to be used alongside exploratory data analysis will be provided. The third stage examines the interrelationships between the variables in the dataset. In the fourth stage, the relationship between numerical variables and one or more categorical variables is demonstrated. The fifth stage discusses machine learning techniques used for making predictions, such as the decision tree, support vector mechanism, random forest, and KNN.

### 2.1 Dataset

The data used in this paper is made by augmenting and combining various publicly available datasets of India which are obtained from the Kaggle website.[1] This data is relatively simple with very few yet useful features unlike the complicated features affecting the yield of the crop. The data has nitrogen (N), phosphorous (P), potassium (K), and pH values of the soil, humidity, temperature, and rainfall required for a particular crop as shown in Table 1. There are 2200 records containing 22 plant species which are rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

This dataset can be utilized to recommend the crop for suitable soil. This will be extremely beneficial for crop production (agriculture) without losses due to soil pH, rainfall, humidity, and other chemical components present in the soil. As indicators, the essential plant nutrients nitrogen, phosphate, and potassium are utilized. Nitrogen is so essential because it is a major component of chlorophyll, the compound that plants use to produce sugars from water and carbon dioxide (i.e., photosynthesis) using sunlight energy. In addition, it is a significant component of amino acids, the building blocks of proteins. Without proteins, plants wilt and eventually die. Phosphorus is therefore essential for cell division and tissue development. Phosphorus is also associated with the plant's complex energy transformations. Adding phosphorus to soil deficient in available phosphorus encourages root development and winter hardiness, stimulates tillering, and frequently accelerates maturation. Potassium is an essential nutrient plant absorbed from the soil and fertilizer. It increases disease resistance, promotes upright and sturdy stalk growth, enhances drought tolerance, and helps plants survive the winter.

The average bioactivity soil temperatures range from 50 to 75 °F. These values are favorable for the normal life functions of earth's biota, which ensure proper decomposition of organic matter, increased nitrogen mineralization, absorption of soluble substances, and metabolism. The pH range between 5.5 and 6.5 is optimal for plant growth due to the optimal availability of nutrients. Rainfall, in addition to disease, can affect the rate at which a crop grows from seed, including when it is ready for harvest. A favorable balance of precipitation and irrigation can result in faster-growing plants, which can reduce the time between seeding and harvest.

Figure 1 shows the data distribution of Nitrogen, Phosphorus, Potassium, Temperature, Humidity and Rainfall in the dataset. It is shown that most of the data is between the temperatures of 20–30 °C, which is a similar temperature range for agriculture in Indonesia. Figure 2 shows the categorial plot of crop distribution on temperature and rainfall.

[1] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset.

**Table 1** Dataset

| N | P | K | Temperature | Humidity | pH | Rainfall | Label |
|---|---|---|---|---|---|---|---|
| 90 | 42 | 43 | 20.8 | 82.0 | 6.5 | 202.9 | Rice |
| 85 | 58 | 41 | 21.7 | 80.3 | 7.0 | 226.6 | Rice |
| 60 | 55 | 44 | 23.0 | 82.3 | 7.8 | 263.9 | Rice |
| 74 | 35 | 40 | 26.4 | 80.1 | 6.9 | 242.8 | Rice |
| 78 | 42 | 42 | 20.1 | 81.6 | 7.6 | 262.7 | Rice |

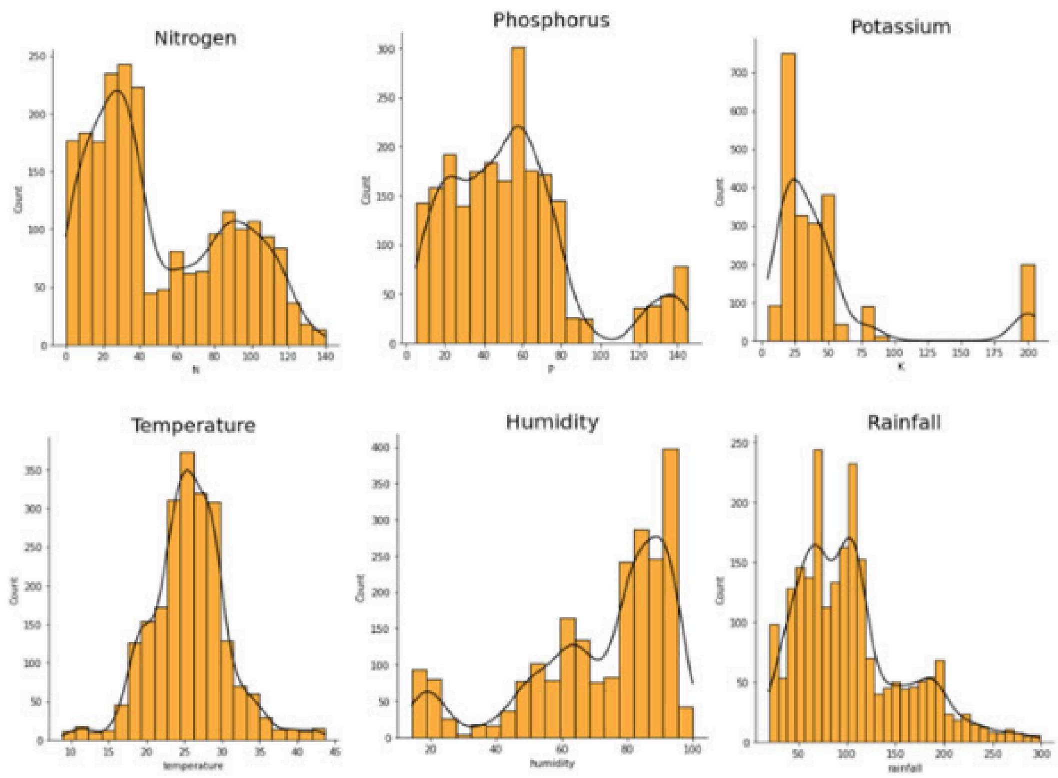


**Fig. 1** Distribution of the dataset

## 2.2 *Machine Learning Model*

The dataset is split into 1650 train sets and 550 test sets with a simple 75:25 train-test split. We compare four machine learning algorithms which are decision tree, support vector machine, random forest, and K-nearest neighbor. We evaluate the performance of those four algorithms with the standard classification evaluation metrics, which is accuracy.

The experiments are conducted several times with different parameters for each algorithm to find the optimal parameters. The decision tree classifier parameters experimented are best splitter and random splitter. For the support vector machine classifier, the regularization parameters C experimented are 1, 10, 100, and 1000, and

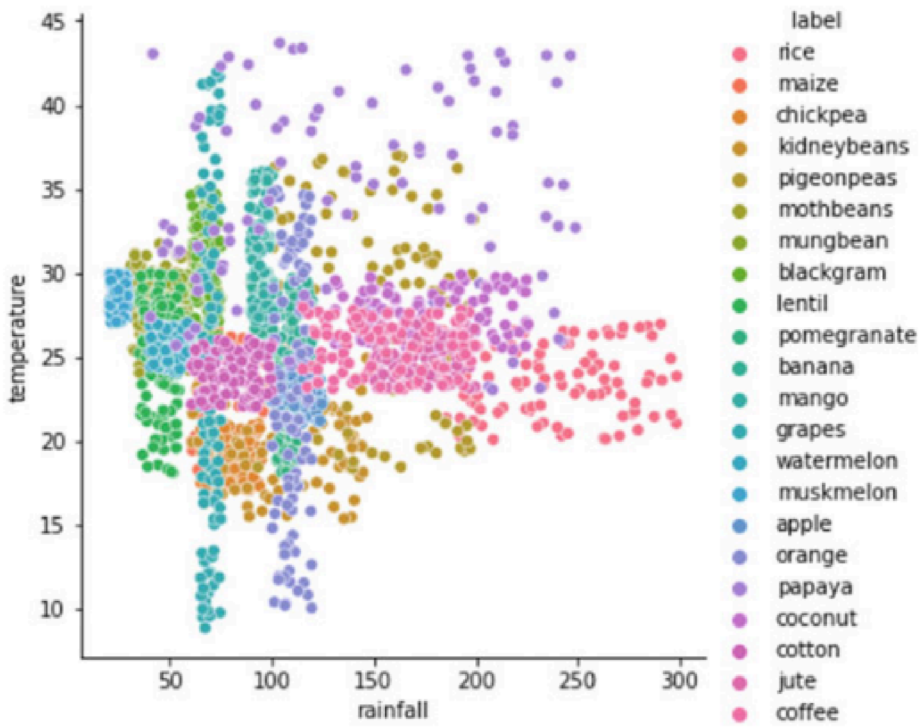


**Fig. 2** Categorical plot of crop distribution on temperature and rainfall

the kernels are rbf and linear. For the random forest classifier, the n estimators are 1, 5, and 10. Lastly, for the K-nearest neighbor classifier, the number of neighbors is 5, 10, 20, and 25.

## 3 Result and Discussion

The experiment results are listed in Table 2. The four machine learning models are compared with accuracy as an evaluation metric. The optimal parameter of each algorithm is also shown in Table 2. The best performing algorithm is random forest with the accuracy of 0.989697.

The result shows that random forest model can predict the suitable crop to be planted on specific soil and weather conditions using the Indian dataset. The next

**Table 2** Performance comparison

| Model | Accuracy | Best parameters |
|---|---|---|
| Decision tree | 0.984848 | Splitter: best |
| Support vector machine | 0.986667 | svc__C: 100, svc__kernel: rbf |
| Random forest | 0.989697 | n_estimators: 10 |
| K-nearest neighbor | 0.967273 | n_neighbors: 5 |

step is to implement the model with Indonesian dataset. However, producing or obtaining big dataset with 22 plant species will be time consuming and costly. A plan optimization to choose the best set of plant species to produce or obtain will be our next challenge. The plan optimization for producing or obtaining plant species dataset can be formalized by utilizing Markov decision process (MDP) with the goal to get a more accurate estimation of the most feasible optimal plan with the least total cost before fully executing the plan [19–21]. Furthermore, the utilization of the Large Language Model for classifying crops to be planted on specific soil and weather conditions can be explored [24].

Figure 3 depicts the predicted potential crop in Indonesia based on temperature, in the temperature range of 20–30 °C, in which temperature is optimal for the growth of some of these plants. Typical Indonesian agroforestry plants that thrive at temperatures between 20 and 30 °C are jeulotung, coconut, rubber, clove, coffee, cocoa, jackfruit, melinjo (*Gnetum gnemon)*, petai (*Parkia speciosa*), teak, tengkawang (*Shorea stenoptera*), resin, and mahogany, as well as those with low economic value, including dadap *(Erytherina variegata*), lamtoro (*Leucaena leucocephala*), and calliandra [22, 23]. Durian (*Durio zibetines*), mangosteen, avocado, banana, guava, sapodilla, salak (*Salacca jalacca*), matoa (*Pometia pinnata)*, persimmon, breadfruit, orange, and pineapple are fruiting plants that can grow under these temperature conditions. Cloves, nutmeg, cinnamon, pepper, turmeric, vanilla, cardamom, and andaliman (*Zanthoxilum piperitum*) are among the spices and medicinal plants that do well under these temperature conditions. In addition, rice, sago, and corn, tubers such as cassava, sweet potatoes, and taro potatoes, and legumes such as soybeans, peanuts, and green beans thrive in these temperatures.

Figure 4 depicts the potential crop predicted for Indonesia based on prevailing humidity. Other than the plant species depicted in the image, all of the aforementioned plant types thrive in a humidity range between 5.5 and 7.2. According to Smith and Ferguson, the season in Indonesia falls within the moderately humid category in terms of humidity. This is supported by Indonesia's geographical location, which is within the equatorial zone, and its favorable climatic conditions for photosynthesis.

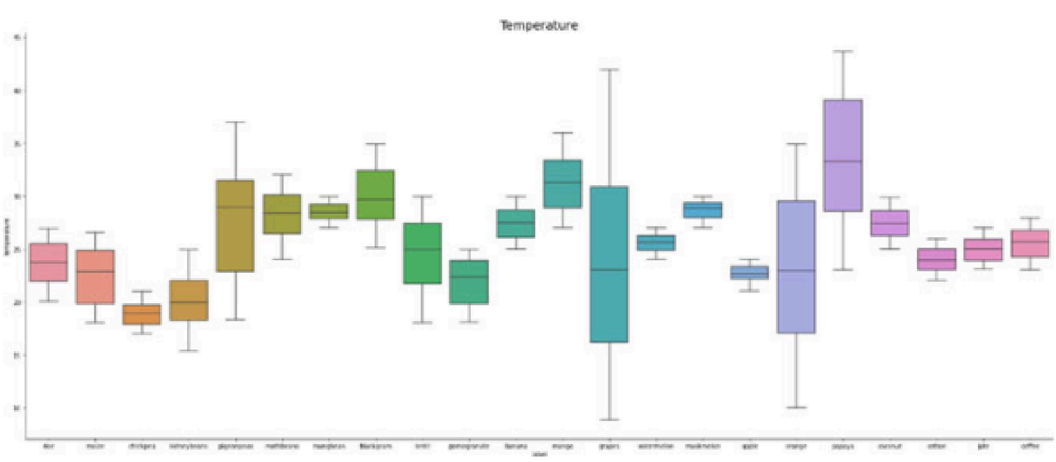


**Fig. 3** Potential crop to be predicted in Indonesia based on temperature

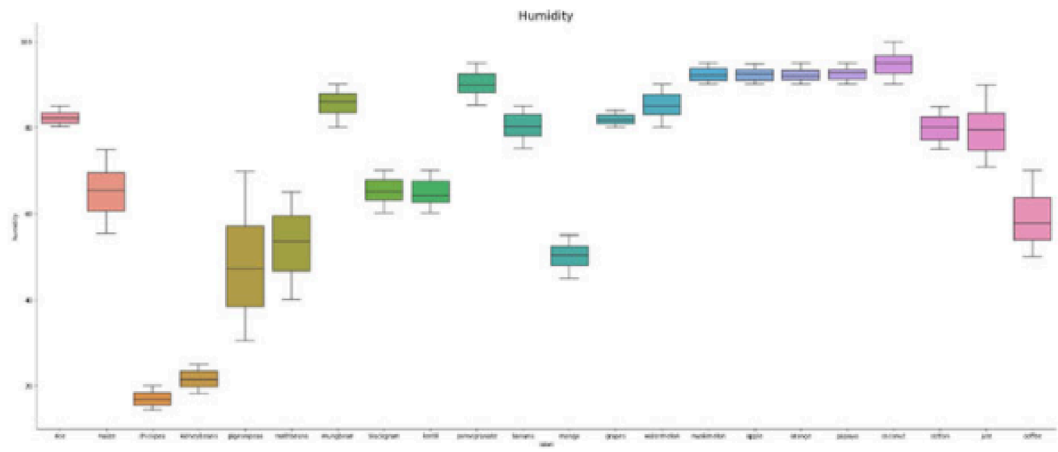


**Fig. 4** Potential crop to be predicted in Indonesia based on humidity

Therefore, climatological engineering can increase the productivity of these diverse plant species.

The average annual precipitation in Indonesia is 2000 m with spatial variations ranging from 500 to 4000 mm. This amount of precipitation has a significant impact on temperature and humidity.

In Indonesia, agricultural and plantation soils are divided into two major groups, mineral soils and organic soils, with a total of 18 types. Suitable soil characteristics for agriculture and plantations include a high concentration of organic elements, pH $< 7$, a high level of humidity even during the dry season, no hardening after planting, and the absence of a rock layer. Of the 18 types of soil, red-yellow solitary pod soil predominates. This soil is suitable for agriculture.

Complete nutrients can promote plant growth and produce high-quality produce. The absolute requirement for plant nutrients must be met because nutrients in the soil are scarce and their availability decreases. Macronutrients and micronutrients make up the plant nutrients. The macro-elements nitrogen, phosphorus, and potassium are essential for plant growth. Based on the characteristics of Indonesian soil, nitrogen, phosphorus, and potassium are classified as moderate to moderate.

## 4 Conclusion

Machine learning can be used to predict the types of Indonesian agricultural crops that can be grown under conditions of 20–30 °C temperature, humidity of 5.5–7.2%, and soil nutrient content. The use of machine learning can help farmers increase their crop yields. This is because the data utilized by machine learning can also assist farmers in determining the optimal planting and harvesting times. The result shows that random forest model can predict the suitable crop to be planted on specific soil and weather conditions using the Indian dataset. The next step is to implement the model with Indonesian dataset. A plan optimization to choose the best set of plant species to produce or obtain will be our next challenge. The plan optimization for producing or obtaining plant species dataset can be formalized by utilizing Markov decision process (MDP) with the goal to get a more accurate estimation of the most feasible optimal plan with the least total cost before fully executing the plan.

**Acknowledgements** This research is supported by Universitas Islam Riau.

## References


1. Ansari A, Lin Y-P, Lur H-S (2021) Evaluating and adapting climate change impacts on rice production in Indonesia: a case study of the Keduang Subwatershed, Central Java. Environments 8:117
2. Zhai P, Zhou B, Chen Y (2018) A review of climate change attribution studies. J Meteorol Res 32:671–692
3. Marques ÉT, Gunkel G, Sobral MC (2019) Management of tropical river basins and reservoirs under water stress: experiences from Northeast Brazil. Environments 6:62
4. Wild M (2012) Solar radiation surface solar radiation versus climate change solar radiation versus climate change. In: Meyers RA (ed) Encyclopedia of sustainability science and technology. Springer, New York, NY, USA, pp 9731–9740
5. Lehtonen H, Liu X, Purola T (2015) Balancing climate change mitigation and adaptation with socio-economic goals at farms in northern Europe. In: Paloviita A, Järvelä M (eds) Climate change adaptation and food supply chain management, vol 11. Routledge, Taylor & Francis Group, London, UK, pp 132–146
6. Nelson GC, Valin H, Sands RD, Havlík P, Ahammad H, Deryng D, Elliott J, Fujimori S, Hasegawa T, Heyhoe E (2014) Climate change effects on agriculture: economic responses to biophysical shocks. Proc Natl Acad Sci USA 111:3274–3279
7. Wassmann R, Jagadish SVK, Heuer S, Ismail A, Redona E, Serraj R, Singh RK, Howell G, Pathak H, Sumfleth K (2009) Climate change affecting rice production: the physiological and agronomic basis for possible adaptation strategies. Adv Agron 101:59–122
8. Atzberger C (2013) Advances in remote sensing of agriculture: context description, existing operational monitoring systems and major information needs. Remote Sens 5:949–981
9. Santos L, Ferraz N, dos Santos FN, Mendes J, Morais R, Costa P, Reis R (2018) Path planning aware of soil compaction for steep slope vineyards. In: Proceedings of the 2018 IEEE international conference on autonomous robot systems and competitions (ICARSC), Torres Vedras, Portugal, 25–27 April 2018
10. Patil KA, Kale NR (2016) A model for smart agriculture using IoT. In: Proceedings of the 2016 international conference on global trends in signal processing, information computing and communication (ICGTSPICC), Jalgaon, India, 22–24 Dec 2016, pp 543–545
11. Dhanaraju M, Chenniappan P, Ramalingam K, Pazhanivelan S, Kaliaperumal R (2022) Smart farming: internet of things (IoT)-based sustainable agriculture. Agriculture 12:1745
12. Walter A, Finger R, Huber R, Buchmann N (2017) Opinion: smart farming is key to developing sustainable agriculture. Proc Natl Acad Sci USA 114:6148–6150
13. Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117
14. Kamilaris A, Prenafeta-Boldú FX (2018) A review of the use of convolutional neural networks in agriculture. J Agric Sci 156:312–322
15. Albahar M (2023) A survey on deep learning and its impact on agriculture: challenges and opportunities. Agriculture 13:540
16. Morales A, Villalobos FJ (2023) Using machine learning for crop yield prediction in the past or the future. Front Plant Sci

17. Lomma LN, Jian S, Amshi AT (2020) Application of machine learning in agriculture: future scope. IRJET 07(10)
18. Shahhosseni M, Hu G, Huber I, Archontoulis SV (2021) Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Sci Rep 11:1606
19. Nasution AH, Murakami Y, Ishida T (2021) Plan optimization to bilingual dictionary induction for low-resource language families. Trans Asian Low-Resour Lang Inf Process 20:1–28
20. Nasution AH, Kadir EA, Murakami Y, Ishida T (2020) Toward formalization of comprehensive bilingual dictionaries creation planning as constraint optimization problem. In: Optimization based model using fuzzy and other statistical techniques towards environmental sustainability. Springer, pp 41–54
21. Nasution AH, Murakami Y, Ishida T (2017) Plan optimization for creating bilingual dictionaries of low-resource languages. In: 2017 international conference on culture and computing (culture and computing). IEEE, pp 35–41
22. Noldeke B, Winter E, Laumonier Y, Simamora T (2021) Simulating agroforestry adoption in rural Indonesia: the potential of trees on farmers for livelihoods and environment. Land 10:385
23. Budiastuti MTS, Purnomo D, Setyaningrum D (2022) Agroforestry system as the best vegetation management to face forest degradation in Indonesia. Rev Agric Sci 10:4–23
24. Nasution, AH, Onan A (2024) ChatGPT label: comparing the quality of human-generated and LLM-generated annotations in low-resource language NLP Tasks. IEEE Access.

Toward Crops Prediction in Indonesi

ORIGINALITY REPORT

| 33% | 23% | 31% | 20% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Marwan Albahar. "A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities", Agriculture, 2023<br>Publication | 4% |
| 2 | mdpi-res.com<br>Internet Source | 4% |
| 3 | github.com<br>Internet Source | 4% |
| 4 | Dedek Andrian, Khairul Amri, Restu K. Nugraheni, Arbi Haza Nasution, Azmansyah. "Chapter 14 Analysis of Creative Economic Situation in Pekanbaru: Statistical Analysis and Word Cloud", Springer Science and Business Media LLC, 2024<br>Publication | 3% |
| 5 | www2.mdpi.com<br>Internet Source | 2% |
| 6 | www.researchgate.net<br>Internet Source | 2% |
| 7 | Submitted to Liverpool John Moores University<br>Student Paper | 2% |
| 8 | S. NagaMallik Raj, Pyla Lohit, Doddala Jyo-theendra, Kannuru Chandana, P. Nikhil, N. Thirupathi Rao, Debnath Bhattacharyya. "Chapter 14 Prediction and Identification of Diseases to the Crops Using Machine | 2% |

Learning", Springer Science and Business Media LLC, 2023
Publication

| | | |
|---|---|---|
| 9 | dokumen.pub<br>Internet Source | 2% |
| 10 | doctorpenguin.com<br>Internet Source | 1% |
| 11 | Mohsen Shahhosseini, Guiping Hu, Isaiah Huber, Sotirios V. Archontoulis. "Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt", Scientific Reports, 2021<br>Publication | 1% |
| 12 | Submitted to Montverde Academy<br>Student Paper | 1% |
| 13 | "Charting a Sustainable Future of ASEAN in Business and Social Sciences", Springer Science and Business Media LLC, 2020<br>Publication | 1% |
| 14 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 1% |
| 15 | engrxiv.org<br>Internet Source | 1% |
| 16 | Submitted to University of the Philippines Los Banos<br>Student Paper | 1% |
| 17 | www.coursehero.com<br>Internet Source | 1% |
| 18 | "Proceedings of International Conference on Smart Computing and Cyber Security", Springer Science and Business Media LLC, 2021<br>Publication | 1% |

| | | |
|---|---|---|
| 19 | Submitted to University of Newcastle upon Tyne<br>Student Paper | 1% |
| 20 | Submitted to University of Hertfordshire<br>Student Paper | 1% |

| | | | |
|---|---|---|---|
| Exclude quotes | On | Exclude matches | < 1% |
| Exclude bibliography | On | | |