Proceedings Academic Universitas Jambi
Volume 2, Issue 3, September, 2026

Prediction of Cropping Season Onset Using Machine Learning as a Project-Based Learning (PjBL) Module in a Computational Physics Course

Tugiyo Aminoto[1], Nova Susanti[2], Rahma Dani[3], Suharli AJ[4]

[1,2,3,4] Department of Physics Education, Jambi University, Jambi, Indonesia

[*]Coressponding Authors: tugiyo@unja.ac.id

Abstract

Background: Climate change has led to increasing air temperatures and changing rainfall patterns, which directly affect the agricultural sector, particularly in determining cropping calendars. Conventional cropping calendars have become less accurate, potentially reducing crop productivity and increasing the risk of crop failure. The challenge of predicting cropping calendars is closely related to computer programming concepts taught in the Computational Physics course, and integrating these two areas is strategically important to support Indonesia's national food security program, particularly in Jambi Province. Objective: This study aims to (1) develop a climate change-based cropping onset calendar using a machine learning approach, and (2) integrate the development of this cropping calendar as a project in the Computational Physics course. Methods: The study used rainfall, air temperature, and large-scale climate indices, including the El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and Madden–Julian Oscillation (MJO) as predictors. The research methodology consists of data preprocessing, exploratory data analysis, and predictive modeling using a machine learning algorithm (Random Forest). Model performance is evaluated using Mean Absolute Error (MAE). Results: The model predicted MT1 with the MAE improving from 1.0 to 0.67 dasarian after feature selection. However, MT2 prediction remained less accurate, with the MAE decreasing only slightly from 3.3 to 3.0 dasarians. The modeling process can be integrated into the Computational Physics course through a project-based learning approach. Students participate in data processing, model development, and result interpretation, thereby enhancing their computational and data analysis skills through real-world problem-solving. Conclusion: The outputs of this research include an adaptive planting time prediction model that accounts for climate change, a data-driven cropping calendar prototype, and a contextual learning innovation for Computational Physics education. This study is expected to contribute both to the advancement of data-driven agricultural science and to the improvement of teaching quality in higher education.

Keywords: cropping season onset; climate change; machine learning; computational physics; project-based learning

Cite This Article

Aminoto, T., Susanti, N., Dani, R., & Suharli, A. J. (2026). Prediction of Cropping Season Onset Using Machine Learning as a Project-Based Learning (PjBL) Module in a Computational Physics Course. *Proceedings Academic Universitas Jambi*, 1(1): e59986. DOI: 10.22437/proca.v2i3.59986

Editor

Dr. Robi Hendra,M.Pd

Article info

Received: August 4, 2026. Revised: August 26, 2026. Accepted: September 1, 2026

© 2026 The Author(s). Published with license by LPPM Universitas Jambi. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0 International). This license enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator.

INTRODUCTION

Climate change is one of the greatest challenges facing the global agricultural sector because it directly affects crop productivity. Rising temperatures, changing rainfall patterns, and more frequent extreme weather events have created uncertainty in agricultural production systems, including rice cultivation, the primary staple food crop in Asia (1, 2). Numerous studies have shown that climate change can reduce crop productivity by altering growing environmental conditions. Globally, increasing temperatures have been proven to negatively affect the yields of major crops, including rice, although the magnitude of these impacts varies spatially (3, 4). This indicates that conventional agricultural systems relying on historical seasonal patterns are becoming increasingly inadequate under dynamic climate conditions.

In Indonesia, particularly in the Sumatra region, climate variability has been shown to significantly influence rice production. The integration of climate and agricultural data demonstrates that factors such as temperature, rainfall, and humidity play dominant roles in determining rice yields (5, 6). Therefore, a more adaptive and data-driven approach to agricultural planning is required, one of which is the development of a planting calendar that incorporates climate change factors. Traditional planting calendars, which farmers have long relied upon, are generally based on experience and historical seasonal patterns and are therefore less capable of accommodating current climate dynamics (7, 8). However, determining the optimal planting time is crucial for increasing productivity and reducing the risk of crop failure. Recent studies indicate that optimizing planting schedules using climate data and predictive models can significantly improve crop yields (25–27). A summary of the literature review is presented in Table 1, which includes five of the most relevant related studies.

Table 1. Summary of selected literature review

No.	Author	Method	Main findings
1	Benos et al.	ML Review	Machine learning effectively improves prediction accuracy and supports data-driven decision-making (9).
2	Sharma et al.	Review	Machine learning can model complex relationships between climate factors and agricultural production (10).
3	Wolfert et al.	Big Data Review	Integration of big data and IoT enhances the efficiency and adaptability of agricultural systems (11).
4	Lipper et al.	Conceptual	Climate-adaptive approaches are essential for maintaining food security (12).
5	Javaid et al.	AI & IoT Review	AI and IoT improve agricultural efficiency, productivity, and resource management (13).

Along with technological advances, machine learning (ML)-based approaches have been widely applied in agriculture to model the complex relationships between climate variables and crop yields. ML models are capable of processing large volumes of data and producing more accurate predictions than conventional statistical methods (14, 15). Several studies have demonstrated that algorithms such as Random Forest and Neural Networks achieve high performance in predicting rice production based on

climatic factors (16, 17, 19). Furthermore, the application of artificial intelligence in agriculture supports the development of smart agriculture systems that are adaptive to climate change. This technology enables the integration of climate, soil, and crop data to generate more effective and sustainable agricultural management recommendations (18, 24). On the other hand, in higher education, particularly in Computational Physics courses, innovative teaching approaches are needed to integrate theoretical concepts with real-world applications. Project-based learning has proven to be an effective approach for improving students' analytical, programming, and data-driven problem-solving skills (19). Implementing real-world projects, such as developing a machine learning-based rice planting calendar, not only enhances students' understanding of computational concepts but also contributes to solving practical agricultural problems.

The problem-solving framework of this study adopts an integrative approach that combines climate change analysis with machine learning techniques to develop an adaptive and data-driven rice planting calendar. The initial stage involves collecting and processing climate data (such as air temperature, rainfall, and climate variability) together with rice production data to identify patterns and relationships among these variables. This approach is consistent with the concept of climate-smart agriculture, which emphasizes adapting agricultural systems to climate dynamics through the use of data and technology (12).

Subsequently, machine learning models are developed to learn the complex relationships between climatic factors and optimal planting times. Algorithms such as Random Forest, Support Vector Machine, and Neural Network can be employed to improve prediction accuracy because of their capability to handle multivariate and nonlinear data (9). The resulting models are then used to construct a planting calendar that is adaptive to current climatic conditions and future projections. The next stage involves evaluating model performance by comparing its predictions with conventional methods based on historical seasonal patterns. The evaluation is conducted using Mean Absolute Error (MAE) to ensure that the developed model achieves satisfactory accuracy and reliability (10). The research findings are then implemented through project-based learning in the Computational Physics course, where students actively participate in data processing, model development, and result interpretation. This approach not only provides practical solutions to agricultural problems but also enhances students' computational skills, data analysis capabilities, and science-based contextual problem-solving abilities.

Recent research trends indicate that integrating artificial intelligence into agriculture has become a major focus in addressing climate change challenges and global food security. Smart farming approaches based on machine learning and deep learning have been extensively applied to various tasks, including crop yield prediction, land suitability classification, and precision resource management. Recent studies demonstrate that machine learning models can process large-scale and complex datasets to produce more accurate predictions than conventional methods while supporting data-driven decision-making in modern agricultural systems (9). In addition, integrating technologies such as the Internet of Things (IoT) and real-time data analytics further strengthens the concept of precision agriculture that adapts to environmental dynamics (20).

Meanwhile, climate-smart agriculture has emerged as a strategic approach for improving agricultural productivity while enhancing resilience to climate change. Recent studies show that the application of AI and machine learning within this framework can improve production efficiency, optimize resource utilization, and significantly increase crop yields (12, 13). However, most existing studies still focus primarily on crop yield prediction or general recommendations and have not specifically integrated climate change factors into practical planting calendar systems, particularly within local and educational contexts.

The novelty of this research lies in its multidisciplinary integration of climate change, machine learning, and project-based learning in higher education. Unlike previous studies, which generally focused only on crop yield prediction or agricultural classification (9, 10), this study develops an adaptive rice planting calendar based on climate data by utilizing machine learning models that consider both temporal and spatial climate variability. Furthermore, this research not only develops a predictive model but also implements it as a project-based learning activity in a Computational Physics course, thereby contributing to both scientific advancement and educational innovation. This integrative approach remains relatively uncommon, particularly in linking climate data analysis with project-based learning implementation in computational science education (19). Consequently, this study has the potential to produce a model that is not only scientifically robust but also practical and relevant for enhancing educational quality while providing tangible solutions to agricultural challenges in the era of climate change.

Based on the above research background, this study aims to develop a rice planting calendar based on climate change factors and machine learning while implementing it as a project in Computational Physics education. This research is expected not only to produce a predictive model that is adaptive to climate change but also to improve learning quality through the integration of physics, computation, and real-world problems, particularly in the agricultural sector, which currently plays a strategic role in strengthening national food security. This study addresses the following research questions:

1. How do climate change factors (rainfall and climate variability) influence the determination of the optimal planting times?
2. How can a machine learning model be developed to generate an adaptive planting calendar based on climate data?
3. How accurate is the proposed machine learning model compared with conventional methods for determining the planting calendars?

METHODS

Study design and setting

This study was conducted in Jambi Province, Indonesia, using meteorological observations from the Sultan Thaha Meteorological Station (BMKG). Jambi is located on the eastern coast of Sumatra and is characterized by a humid tropical rainforest climate. Rainfall variability in the region is strongly influenced by both local atmospheric processes (21,22,23) and large-scale climate variability (20), including the El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), the Madden–Julian Oscillation (MJO). These climate phenomena modulate seasonal rainfall patterns and consequently affect the onset of agricultural planting seasons (25,26). Daily meteorological observations from 1991–2019 were used to develop the prediction

model. The target variables were the onset of the first planting season (MT1) and the second planting season (MT2), expressed as the corresponding dasarian (10-day period) of each year.

Procedure and data collection

Daily meteorological data were obtained from the Indonesian Agency for Meteorology, Climatology, and Geophysics (BMKG). The variables included: Daily rainfall (mm), Maximum air temperature (°C), Minimum air temperature (°C), Mean air temperature (°C), Large-scale climate indices were incorporated as external predictors: Niño 3.4 Index (ENSO), Dipole Mode Index (IOD), Madden–Julian Oscillation (MJO), ENSO and IOD were obtained as monthly indices, whereas MJO was available at daily temporal resolution. These indices have been widely recognized as dominant drivers of rainfall variability over the Maritime Continent (27,28,29).

Unlike the other climate indices, the MJO is an intraseasonal phenomenon with a typical period of 30–90 days (30). Daily MJO observations were therefore summarized into monthly predictors, including the monthly mean MJO amplitude and the monthly proportions of active MJO days occurring in phases 2–3 and 4–5. Only days with an MJO amplitude greater than one were considered active. Phases 2–3 and 4–5 were selected because they are commonly associated with enhanced convective activity and increased rainfall over the Maritime Continent, including Sumatra (31), making them particularly relevant for representing atmospheric variability affecting Jambi.

To represent delayed climate responses, lagged variables were generated for the large-scale climate indices. ENSO and IOD were represented by one-, two-, three-, and four-month lags. MJO predictors (mean amplitude and phase proportions) were represented by one- and two-month lags to account for their shorter persistence. These lagged predictors were included to capture the delayed influence of large-scale climate variability on rainfall and, consequently, on the timing of cropping season onset.

The response variables, namely the onset of the first (MT1) and second (MT2) cropping seasons, were determined at the dasarian (10-day) scale. Daily rainfall was aggregated into dasarian totals, and the onset of each cropping season was identified based on three consecutive wet dasarians. Thus, while the predictor variables were maintained at the monthly scale, the Random Forest model was developed to predict cropping season onset at the finer dasarian resolution, enabling monthly climate information to be translated into operational 10-day planting recommendations.

The daily meteorological data were first quality-controlled by converting all meteorological variables into numeric format and replacing missing-value codes with null values. Records outside the study period were excluded.
Reference evapotranspiration (ET_0) was estimated using the Hargreaves equation:

$$ET_0 = 0.0023(T_{mean} + 17.8)\sqrt{T_{max} - T_{min}}$$

where T_mean, T_max, and T_min denote daily mean, maximum, and minimum air temperatures, respectively (32). Water balance (WB) was then calculated as

$$WB = P - ET_0$$

where (P) is daily rainfall.

All daily variables were aggregated into dasarian (10-day) periods, producing 36 observations per year. For each dasarian, total rainfall, average temperature, average ET_0, total water balance, and rainy days were calculated. Missing predictor values were imputed using median imputation prior to model development, while target variables were retained without imputation.

The onset of cropping seasons was identified using rainfall-based criteria adapted from the operational definition employed by BMKG. For each dasarian, a wet dasarian was defined as a 10-day period with cumulative rainfall of at least 50 mm and a minimum of three rainy days. The onset of the first planting season (MT1) was defined as the first occurrence of three consecutive wet dasarians. The onset of the second planting season (MT2) was determined after the occurrence of three consecutive dry dasarians, followed by another sequence of three consecutive wet dasarians occurring at least ten dasarians after MT1. The resulting onset dates were expressed as annual dasarian numbers (1–36).

Meteorological variables were aggregated into dasarian (10-day) intervals to match the temporal resolution of the target variables. The aggregated predictors included cumulative rainfall, mean air temperature, mean reference evapotranspiration (ET_0), and cumulative water balance for each dasarian. Large-scale climate indices, including the El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and Madden–Julian Oscillation (MJO), were integrated with the dasarian dataset. Monthly climate indices (ENSO and IOD) were assigned to the three corresponding dasarians within each month, whereas daily MJO observations were aggregated into dasarian averages. Specifically, the mean MJO amplitude and the proportions of active MJO days in phases 2–3 and 4–5 were calculated for each dasarian. To account for delayed climate influences on rainfall and cropping season onset, lagged climate variables were generated. ENSO and IOD by lags of three, six, nine, and twelve dasarians, and MJO by lags of one and two dasarians. These lagged predictors were included to capture the temporal persistence of large-scale climate variability and its delayed effects on regional hydroclimatic conditions.

Random Forest regression was used to independently predict the onset of the first (MT1) and second (MT2) cropping seasons. The predictor set comprised cumulative rainfall, mean air temperature, reference evapotranspiration (ET_0), water balance, ENSO, IOD, MJO, and all corresponding lagged climate variables. The Random Forest models were implemented using 500 decision trees (*ntree* = 500) with a minimum terminal node size of three observations (*nodesize* = 3). Missing values in the predictor variables were imputed using median values prior to model training. A walk-forward validation approach was employed to simulate operational forecasting conditions. In each iteration, the model was trained using all available historical data up to a specified year and then used to predict the cropping season onset for the subsequent year. Three forecasting experiments were conducted: (1) training with data from 1991–2015 to predict 2016, (2) training with data from 1991–2016 to predict 2017, and (3) training with data from 1991–2017 to predict 2018. Random Forest was selected because it can effectively model complex nonlinear relationships, is robust to multicollinearity, noise, and missing data, and provides reliable measures of predictor importance through permutation-based variable importance (%IncMSE) (33,34). This capability makes

Random Forest particularly suitable for modeling the interactions between climate variability and cropping season onset.

Statistical analysis

Prediction performance was evaluated using the Mean Absolute Error (MAE) (35):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|]$$

where y_i denotes the observed cropping season onset and hat y_i represents the predicted onset. MAE is expressed in dasarian units and measures the average prediction error.

The relative contribution of each predictor was assessed using the permutation-based Variable Importance (%IncMSE) produced by the Random Forest algorithm. The average importance across the walk-forward experiments was calculated, and only the highest-ranked variables within each climate-index group (ENSO, IOD, and MJO) were retained to construct the final prediction model. This feature selection procedure reduced redundancy while preserving the most informative predictors.

RESULTS and DISCUSSION

The influence of large-scale climate variability on rainfall in the study area was first investigated. Pearson correlation analysis was performed to quantify the relationships between rainfall and the major climate indices, namely the El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and Madden–Julian Oscillation (MJO). The analysis was performed at both 10-day (dasarian or dekad) and monthly temporal scales. Comparing these two temporal resolutions allows the assessment of how temporal aggregation affects the strength of the relationship between rainfall and climate indices. In addition to contemporaneous climate indices, lagged climate variables were also included in the monthly analysis to evaluate the delayed response of rainfall to large-scale climate variability. The correlation coefficients obtained at both temporal scales are presented in Table 2.

Variable importance

Figure 1 presents the variable importance (%IncMSE) obtained from the Random Forest models for predicting the onset of the first (MT1) and second (MT2) cropping seasons using monthly aggregated climate predictors. The results indicate that large-scale climate variability plays a more important role than local meteorological variables in determining cropping season onset in Jambi.

Table 2. Pearson correlation coefficients between rainfall and climate indices

Climate Index	Dasarian (r)	Monthly (r)
ENSO	-0.142	-0.214
IOD	-0.155	-0.219
MJO Amplitude	0.034	0.081
MJO Phase 2–3	0.116	0.041
MJO Phase 4–5	-0.005	0.077

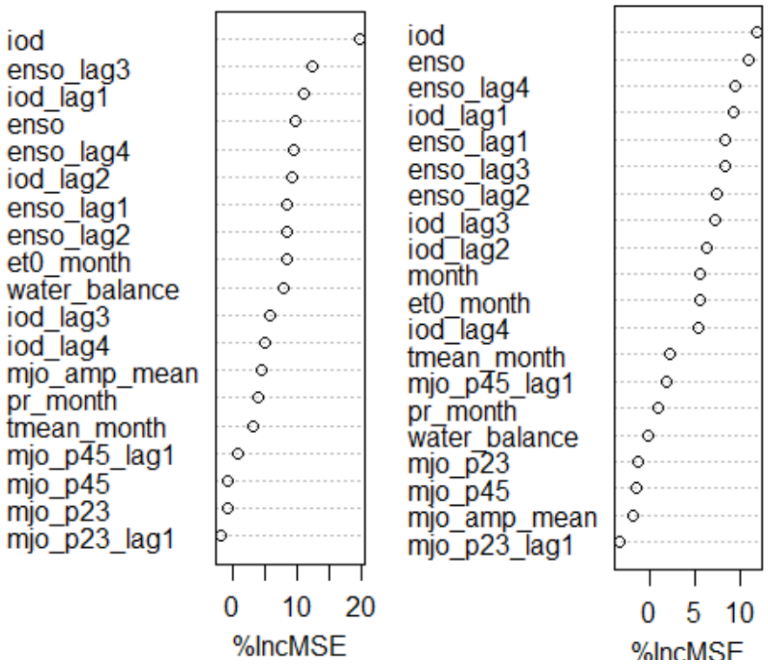

Figure 1. Variable importance (%IncMSE) of monthly climate predictors for Random Forest models predicting (a) MT1 and (b) MT2 onset.

Prediction performance of the random forest model

Table 3 presents the observed and predicted cropping season onset for MT1 and MT2 during the validation period (2016–2018). Although the predictor variables were aggregated at the monthly scale, the Random Forest model successfully predicted cropping season onset at the finer dasarian (10-day) resolution. The Random Forest model successfully predicted the onset of MT1 with relatively high accuracy, yielding a mean absolute error (MAE) of approximately one dasarian. In contrast, MT2 predictions were less accurate, with an MAE of approximately 3.3 dasarians.

For MT1, the predicted onset closely matched the observed values, with errors ranging from 0 to 2 dasarians and a mean absolute error (MAE) of approximately 1.0 dasarian. The model accurately predicted the onset in 2017, while the predictions for 2016 and 2018 differed by only one and two dasarians, respectively. These results indicate that monthly climate variability contains sufficient information to predict the onset of the first planting season with relatively high accuracy. In contrast, MT2 prediction exhibited larger discrepancies, with a mean absolute error of approximately 3.3 dasarians. The prediction errors were two dasarians in 2016 and four dasarians in both

2017 and 2018. The lower accuracy for MT2 suggests that the second planting season is influenced by more complex hydroclimatic conditions than MT1. After the onset of the rainy season, soil moisture, antecedent rainfall, irrigation practices, and short-term atmospheric variability become increasingly important, introducing additional uncertainty that may not be fully represented by monthly aggregated climate predictors.

Table 3. Observed and predicted of MT for 3 year validation period

Year	MT1 Observed	MT1 Predicted	MT2 Observed	MT2 Predicted
2016	4	3	29	27
2017	6	6	25	29
2018	3	5	24	28

Note: MT1 = 4 corresponds to the first dasarian of February, whereas MT2 = 24 corresponds to the third dasarian of August

Despite using monthly climate variables as predictors, the model produced predictions at the dasarian scale, demonstrating that Random Forest effectively learned the nonlinear relationships between seasonal climate variability and the timing of cropping season onset. This finding suggests that monthly climate indices preserve sufficient seasonal information for predicting agricultural events at a finer temporal resolution, thereby reducing data dimensionality without substantially compromising prediction accuracy. The better performance of MT1 compared with MT2 is also consistent with the variable importance analysis, where large-scale climate indices, particularly IOD and ENSO, were identified as the dominant predictors. These climate modes primarily influence the onset of the rainy season over western Indonesia, making MT1 more directly linked to large-scale climate variability. In contrast, MT2 occurs later in the agricultural calendar, when local hydrological conditions and intraseasonal atmospheric processes play a greater role, reducing the predictive skill of models based primarily on large-scale climate indices. Overall, the results demonstrate that monthly aggregated climate predictors can effectively forecast cropping season onset at the dasarian scale, particularly for the first planting season. This approach provides a practical framework for operational agricultural forecasting because monthly climate indices are routinely available several weeks to months in advance, allowing timely prediction of planting schedules while maintaining a relatively fine temporal resolution.

Effect of feature selection on prediction performance

To reduce predictor redundancy and improve model efficiency, a feature selection procedure was applied based on the permutation importance (%IncMSE) obtained from the Random Forest model. Rather than retaining all lagged climate variables, only the most important predictor from each climate-index group (ENSO, IOD, and MJO) was selected, while the remaining variables with lower importance scores were excluded. Consequently, several lagged variables, including ENSO, ENSO lag1, lag2, lag4, IOD lag1–lag4, MJO phase 2–3, MJO phase 4–5, and their corresponding lagged variables, were removed from the final model. This strategy reduced predictor dimensionality while preserving the dominant climate signals identified by the Random Forest algorithm.

Table 4. Prediction performance after feature selection based on variable importance.

Year	Observed MT1	Predicted MT1	Observed MT2	Predicted MT2
2016	4	4	29	28
2017	6	6	25	29
2018	3	5	24	28

Model performance

Metric	MT1	MT2
Mean Absolute Error (MAE)	0.67 dasarian	3.00 dasarians

The simplified model improved the prediction of the first cropping season (MT1). As shown in Table 4, the prediction error for 2016 decreased from one dasarian to zero, resulting in a reduction of the MAE from approximately 1.0 to 0.67 dasarian. The predictions for 2017 remained unchanged, while the prediction for 2018 still differed by two dasarians. These results indicate that eliminating redundant predictors reduced model complexity and slightly enhanced the model's generalization ability for MT1. In contrast, the prediction accuracy for the second cropping season (MT2) showed only marginal improvement. The prediction error for 2016 decreased from two to one dasarian; however, the errors for 2017 and 2018 remained relatively large, resulting in an overall MAE of approximately 3.0–3.3 dasarians. This finding suggests that MT2 is governed by more complex environmental processes than MT1, and its prediction cannot be substantially improved through feature selection alone.

The improved performance of the reduced-feature model is consistent with previous studies showing that removing irrelevant or highly correlated predictors can enhance Random Forest performance by reducing noise and improving model interpretability. Although Random Forest is relatively robust to multicollinearity, redundant predictors may dilute the contribution of informative variables and increase prediction variance, particularly when the sample size is limited (33,36). Variable selection therefore enables the model to focus on the most influential predictors while maintaining predictive accuracy.

The retained predictors were primarily associated with IOD and ENSO, reinforcing the conclusion that large-scale ocean–atmosphere variability is the dominant driver of cropping season onset in Jambi. Similar findings have been reported in previous studies, where ENSO and IOD were identified as the primary sources of rainfall variability over the Maritime Continent and were among the most influential predictors in machine-learning models for seasonal climate forecasting (Aldrian & Susanto, 2003; Hendon, 2003). In contrast, most MJO-derived variables contributed little additional predictive information after accounting for ENSO and IOD, likely because the intraseasonal variability represented by the MJO is partially averaged out when predictors are aggregated to the monthly scale.

Overall, these results demonstrate that feature selection based on Random Forest variable importance can produce a more parsimonious prediction model without sacrificing predictive performance. The reduction in predictor dimensionality simplifies model interpretation and decreases computational complexity while preserving the key climate signals controlling cropping season onset. Such an approach

is particularly beneficial for operational forecasting systems, where simpler models are generally preferred because they require fewer input variables while maintaining reliable prediction accuracy.

Beyond its application in agricultural planning, the proposed machine learning-based cropping calendar provides an interdisciplinary learning resource for Computational Physics courses. Its development process aligns well with the principles of Project-Based Learning (PjBL), emphasizing inquiry, collaboration, and problem-solving through real-world applications. The development of a machine learning-based cropping calendar provides an authentic interdisciplinary project that can be effectively implemented using the Project-Based Learning (PjBL) approach in Computational Physics courses. Computational Physics emphasizes the application of computational methods to solve real-world scientific problems through mathematical modeling, data analysis, and computer programming (38). In this context, predicting cropping season onset using climate variability and machine learning enables students to integrate concepts from physics, climatology, agriculture, statistics, and artificial intelligence into a single computational project. Such authentic problems are fundamental to PjBL because they encourage students to apply theoretical knowledge to meaningful societal challenges (37).

The learning project can be organized into several stages following the PjBL framework. Students first identify the agricultural problem associated with climate variability and review the physical mechanisms linking large-scale climate phenomena, such as ENSO, IOD, MJO, and monsoon variability, to regional rainfall. They then collect and preprocess meteorological and climate-index datasets, calculate derived variables such as evapotranspiration and water balance, construct machine learning datasets, and develop prediction models using Random Forest or other machine learning algorithms. Finally, students evaluate model performance using appropriate statistical metrics, interpret feature importance, and communicate the results through technical reports and presentations. This workflow closely resembles the complete scientific research process, thereby promoting computational thinking, data literacy, programming skills, and scientific reasoning.

Using a cropping calendar prediction as a PjBL project also supports the development of higher-order thinking skills. Students are required to formulate hypotheses, evaluate data quality, optimize model parameters, compare alternative prediction models, and interpret the physical significance of climate predictors rather than simply generating prediction results. Such learning activities foster critical thinking, problem-solving, collaboration, and self-directed learning, which are recognized as essential competencies for science and engineering graduates in the twenty-first century (38,39). Furthermore, because agriculture and climate change are highly relevant issues in Indonesia, the project enhances contextual learning by connecting computational techniques with local environmental and societal challenges.

From an instructional perspective, this project can be completed over one academic semester. Students work in small collaborative groups to develop an operational prototype of a machine learning-based cropping calendar using programming languages such as R or Python. The final outputs may include data visualization, feature importance analysis, prediction accuracy evaluation, and a decision-support dashboard for determining planting schedules. Through this process, students acquire

not only computational competencies but also experience in applying artificial intelligence to environmental and agricultural decision-making. Consequently, integrating a machine learning-based cropping calendar into a Computational Physics course represents an effective strategy for implementing Project-Based Learning while strengthening interdisciplinary skills aligned with Education 4.0 and sustainable agricultural development.

CONCLUSIONS

This study demonstrated the potential of Random Forest for predicting cropping season onset in Jambi, Indonesia, using monthly aggregated climate variables while producing predictions at the finer dasarian (10-day) scale. The results showed that large-scale climate variability, particularly the Indian Ocean Dipole (IOD) and the El Niño–Southern Oscillation (ENSO), was more influential than local meteorological variables in determining the onset of both the first (MT1) and second (MT2) cropping seasons. Variable importance analysis consistently identified IOD and ENSO, together with their lagged effects, as the dominant predictors, highlighting the significant role of ocean–atmosphere interactions in controlling seasonal rainfall variability over the study area. Feature selection based on Random Forest variable importance successfully reduced predictor redundancy while maintaining the dominant climate signals. The simplified model improved the prediction accuracy of MT1, reducing the mean absolute error (MAE) to 0.67 dasarian, whereas the prediction accuracy of MT2 remained lower, with an MAE of approximately 3.0 dasarians. These findings suggest that the onset of the first cropping season is more directly controlled by large-scale climate variability, whereas the second cropping season is likely influenced by additional local hydrological and intraseasonal atmospheric processes that were not fully captured by the monthly predictors. Overall, the study demonstrates that monthly climate aggregation provides sufficient predictive information for forecasting cropping season onset at the operational dasarian scale. This approach offers a practical balance between forecast lead time and temporal resolution, making it suitable for supporting agricultural planning and climate-informed cropping calendar development in Indonesia. Future studies should incorporate additional predictors, such as soil moisture, satellite-derived vegetation indices, and seasonal climate forecasts, and evaluate other machine learning algorithms, including Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM) networks, to further improve the prediction of cropping season onset, particularly for the second planting season. Moreover, validating the proposed framework across different agroclimatic regions in Indonesia would help assess its robustness and broader applicability. For future work, additional predictors representing local hydroclimatic conditions, such as soil moisture, irrigation, and satellite-derived vegetation indices, should be incorporated to improve the prediction accuracy of the second cropping season (MT2). In addition, the proposed framework should be evaluated using other advanced machine learning algorithms, such as Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Long Short-Term Memory (LSTM) networks, and validated across different agroclimatic regions in Indonesia to assess its robustness and generalizability.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING
This research was funded by Universitas Jambi under the Applied Research Scheme (Skema Penelitian Terapan) for the Fiscal Year 2026, Faculty of Teacher Training and Education (Fakultas Keguruan dan Ilmu Pendidikan), under Contract Agreement Number: 445/DST/UN21.11/PT.01.05/SPK/2026.

DECLARATION OF ARTIFICIAL INTELLIGENCE USE
This study utilized artificial intelligence (ChatGPT) across two primary capacities: data analysis and manuscript writing support. For data analysis and modeling, machine learning algorithms, including random forest techniques, were executed using R coding to analyze the dataset and predict relevant outcomes. For manuscript writing support, ChatGPT was employed to perform language refinement—improving grammar, sentence structure, and general readability—as well as content summarization to assist in concisely condensing the findings and conclusions. The authors confirm that all AI-assisted processes were critically reviewed to maintain the integrity and reliability of the results, and all final decisions or interpretations presented in this article were made solely by the authors.

REFERENCES
1. Tasneem, K. T., Shahzad, M. U., Rashid, J., Othman, K. M., Zafar, T., & Faheem, M. (2025). Predicting rice yield and impact of climate change on rice production using machine learning models. Theoretical and Applied Climatology, 156, 665.
2. Erlin E, Yunianta A, Wulandhari LA, Desnelita Y, Nasution N, Junadhi. Enhancing rice production prediction in Indonesia using advanced machine learning models. IEEE Access. 2024.
3. Hu, T., Zhang, X., Khanal, S., Wilson, R., Leng, G., Toman, E. M., Wang, X., Li, Y., & Zhao, K. (2024). Climate change impacts on crop yields: A review of empirical findings, statistical crop models, and machine learning methods. Environmental Modelling & Software, 179, 106119.
4. Mulyaqin T, Nurmalina R, Kusnadi N, Trisasongko BH. Effect of climate and economic factors on rice production: Machine learning approach. J Innov Sustain. 2025;9(2).
5. Kurniawan, Z., Tiaharyadini, R., Jayadi, P., & Widhyanty, W. (2025). Analyzing climate impacts on rice production in Sumatra through spatiotemporal machine learning models. Jurnal Ilmu Pengetahuan dan Teknologi Komputer, 11(2), 550.
6. Rizqi B, Manessa MDM. Balancing urban growth and food autonomy: An integrated machine learning and agricultural statistics framework for local rice self-sufficiency in Indonesia. Forum Geogr. 2025.
7. Zhang, X., et al. (2024). Game analysis of future rice yield changes in China based on explainable machine-learning and planting date optimization. Field Crops Research, 317, 109557.
8. Aditya N, Munthe IR, Sihombing V. A comparative analysis of machine learning algorithms for predicting paddy production. Sinkron. 2024;8(2):1200–1207.
9. Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2025). Machine learning in agriculture: A comprehensive updated review. Sensors, 25(3), 1021.
10. Sharma, A., Jain, A., Gupta, P., & Chowdary, V. M. (2024). Machine learning applications for precision agriculture: A review. *Computers and Electronics in Agriculture, 219*, 108189.
11. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2024). Big data in smart farming: A review. *Agricultural Systems, 212*, 103782.
12. Lipper, L., McCarthy, N., Zilberman, D., Asfaw, S., & Branca, G. (2024). Climate-smart agriculture for food security. Nature Reviews Earth & Environment, 5(2), 88–102.
13. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2025). Enhancing smart agriculture using artificial intelligence and Internet of Things. *Sustainable Production and Consumption, 46*, 1–15.

14. Yasin F, Firmansyah MR, Aldo D, Amrustian MA. Multivariate forecasting of paddy production: A comparative study of machine learning models. J Tek Inform. 2025.
15. Mulyaqin, T., Nurmalina, R., Kusnadi, N., & Trisasongko, B. H. (2025). Effect of climate and economic factors on rice production: Machine learning approach. Journal of Innovations and Sustainability, 9(2).
16. Adiba JIQ, Slamet J, Teguh AM, Septiyanto AF, Sarno R. Advanced deep learning and hybrid models for forecasting rice production in Indonesia. In: Proc Int Conf Inf Technol Comput Electr Eng (ICITACEE). 2024.
17. Dhanaraj, R. K., Maragatharajan, M., Sureshkumar, A., & Balakannan, S. P. (2025). On-device AI for climate-resilient farming with intelligent crop yield prediction using lightweight models. Scientific Reports, 15, 31195.
18. Prince, M., & Felder, R. (2024). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education, 113*(1), 1–30.
19. Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2024). Big data in smart farming: A review. *Agricultural Systems, 212*, 103782.
20. Aminoto T, Faqih A, Koesmaryono Y, Dasanto BD. Rainfall anomaly response to ENSO and IOD teleconnections in the CORDEX-SEA simulations. In: IOP Conference Series: Earth and Environmental Science. 2024;1359(1):012008.
21. Aminoto T, Faqih A. Tren curah hujan di Asia Tenggara berdasarkan model CORDEX-SEA dan data ERA5. Journal Online of Physics. 2024;10(1):7–13.
22. Aminoto T, Susanti N, Dani R, Faqih A. Extreme rainfall indices trends in Indonesia during 1971–2020. Jurnal Penelitian Pendidikan IPA. 2026;12(3):456–467.
23. Aminoto T. Performa model hujan CORDEX-SEA pada aspek spatial-temporal dan implementasi faktor bobot pada rata-rata ansambel [dissertation]. Bogor: IPB University; 2024. Available from: https://repository.ipb.ac.id/handle/123456789/158280
24. Islam, M. M., Alharthi, M., Alkadi, R. S., Islam, R., & Masum, A. K. M. (2024). Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience. AIMS Agriculture and Food, 9(4), 980–1003.
25. Aldrian, E., & Susanto, R. D. (2003). Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature. *International Journal of Climatology, 23*(12), 1435–1452. https://doi.org/10.1002/joc.950
26. Chang, C.-P., Wang, Z., McBride, J., & Liu, C.-H. (2005). Annual cycle of Southeast Asia–Maritime Continent rainfall and the asymmetric monsoon transition. *Journal of Climate, 18*(2), 287–301.
27. Hendon, H. H. (2003). Indonesian rainfall variability: Impacts of ENSO and local air–sea interaction. *Journal of Climate, 16*(11), 1775–1790.
28. Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature, 401*(6751), 360–363.
29. Wang, B., & Fan, Z. (1999). Choice of South Asian summer monsoon indices. *Bulletin of the American Meteorological Society, 80*(4), 629–638).
30. Wheeler, M. C., & Hendon, H. H. (2004). An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review, 132*(8), 1917–1932. https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2
31. Zhang, C. (2005). Madden–Julian Oscillation. *Reviews of Geophysics, 43*(2), RG2003, doi:10.1029/2004RG000158.
32. Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture, 1*(2), 96–99.
33. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
34. Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*(3), 18–22.

35. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, *30*, 79-82.
36. Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics, 7*, Article 3. https://doi.org/10.1186/1471-2105-7-3
37. Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools, 19*(3), 267–277. https://doi.org/10.1177/1365480216659733
38. Landau, R. H., Páez, M. J., & Bordeianu, C. C. (2020). *Computational physics: Problem solving with Python* (3rd ed.). Wiley.
39. Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(2), 39–43. https://doi.org/10.1080/00098650903505415