Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:23
https://doi.org/10.28919/cmbn/9662
ISSN: 2052-2541

DATA-DRIVEN APPROACH FOR TEMPERATURE AND RELATIVE HUMIDITY FORECASTING IN SOLAR DRYER DOME FACILITY

ADVENDIO DESANDROS[1,*], GREGORIUS NATANAEL ELWIREHARDJA[1,3], ENDANG DJUANA[4,5], ARIEF SURIADI BUDIMAN[6,7,8], BENS PARDAMEAN[1,2]

[1]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta 11480, Indonesia

[2]Computer Science Department, BINUS Graduate Program - Master of Computer Science Bina Nusantara University, Jakarta 11480, Indonesia

[3]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

[4]Electrical Engineering Department, Faculty of Industrial Technology, Universitas Trisakti, Jakarta 11440, Indonesia

[5]Center for Artificial Intelligence and Advanced Technology (CAPTIVATE), Institution for Research and Community Services, Universitas Trisakti, Jakarta 11440, Indonesia

[6]Industrial Engineering Department, BINUS Graduate Program - Master of Industrial Engineering, Bina Nusantara University, Jakarta 11480, Indonesia

[7]Department of Manufacturing and Mechanical Engineering and Technology, Oregon Institute of Technology, USA

[8]Oregon Renewable Energy Center (OREC), Oregon Institute of Technology, USA

Abstract: Accurate prediction of temperature and relative humidity is essential in Solar Dryer Dome (SDD) facilities to ensure strict quality of the agricultural products. Compared to traditional control models, ML-based approaches provide benefits in terms of adaptability, as they are capable of capturing complex and non-linear patterns in time series data. In this project, primary time series datasets collected from SDD facility in North Jakarta were used to

*Corresponding author
E-mail address: advendiodesandros1@gmail.com
Received October 25, 2025

forecast the next hour temperature and relative humidity values within the 5-minutes resolution. There were three sensors placed in three different locations: outside the facility near the front door, inside the facility near the front door, and near the fan inlet. The collected data showed daily seasonality patterns with temperature and relative humidity possess an inversely correlated relationship due to their inherent physical interaction. On the modeling part, this research compared three different predictive models such as XGBoost, Linear Regression, and Facebook's Prophet based on the RMSE and R^2 scores on the test set. XGBoost and Linear Regression provided on-par performance, with XGBoost showing its best at temperature prediction (RMSE: 1.47°C, R^2: 0.96) and Linear Regression showing its best at relative humidity prediction (RMSE: 3.20%, R^2: 0.96). Meanwhile, Prophet exhibits the lowest performance in both variables with RMSE: 3.56°C and R^2: 0.72 for temperature, and RMSE: 8.42% and R^2: 0.76 for relative humidity forecasting. Despite the low performance, the Prophet model was advantageous in terms of consistency across horizons, while the two best models show decreasing performance in higher horizon settings. These findings suggest the promising application of ML techniques in SDD facilities to support sustainability in agro-industrial business.
Keywords: time series; temperature forecasting; humidity forecasting; climate data analysis; solar dryer dome.
2020 AMS Subject Classification: 62M10, 68T09, 86A08, 62P12.

1. INTRODUCTION

To reduce post-harvest losses of agricultural crops, drying has been a popular method to be implemented [1, 2]. The technique is known to be eco-friendly, effective in maintaining product quality, and cost-efficient [3], hence the wide utilizations of Solar Dryer Dome (SDD) buildings in various countries [2, 4]. The building itself has become a preferable infrastructure for optimal harvest product drying [5–7]. However, enhancements on its control systems should continuously be explored to further amplify its impact on agricultural products and optimize energy usages. Examples include smarter HVAC control using Machine Learning (ML) [8].

Accurate indoor microclimate forecasting is very important to be implemented in SDDs to implement smart control systems. This is because such buildings must be able to maintain its ideal indoor environmental conditions, as the availability of solar energy is limited only at daytimes and the temperature may drop significantly during night, especially in remote areas with extreme weather conditions [9]. The microclimate variables that play an important role in maintaining ideal environmental conditions include temperature τ_i and relative humidity η_j [5, 6]. Mathematically, the general first-order derivative forms of the climate state variables in the SDD can be expressed as:

$$\frac{\partial Y_C}{\partial_t} = f(Y_C, Y_P, X_L, X_E) \tag{1}$$

where $y_c \in Y_C$ refers to climate state variables including temperature τ_i and humidity η_j, $y_p \in Y_P$ refers to plant state variables, $x_l \in X_L$ refers to control inputs such as ventilation rate and light intensity, and $x_e \in X_E$ refers to external inputs such as solar radiation. To control such variables, actuators are implemented in smart SDDs which usages can be optimized to improve energy consumption [5].

In this study, we collected data from an SDD that is aimed to assist the drying of coffee beans, which are required to be dried up to limit the moisture content at only below 12.5% according to the Indonesian standard [10]. As coffee beans are stored inside, it is crucial to continuously control the ideal τ_i and η_j. Past simple approach include the Albright's Complete Linear Model for greenhouse dynamic control system [11], which is formulated as follows:

$$\frac{\partial \tau_{in}(t)}{\partial t} = \frac{1}{C_p V}\rho^{-1}[\theta(t) + \sigma(t) - \lambda f(t) - \mu \Delta T] - \omega \Delta T \tag{2}$$

$$\frac{\partial \eta_{in}(t)}{\partial t} = \frac{1}{V}\rho^{-1}[f(t) + \varepsilon(t)] - \omega\rho^{-1}\Delta H \tag{3}$$

$$\omega = \frac{v}{V} \tag{4}$$

$$\Delta T = \tau_{in} - \tau_{out} \tag{5}$$

$$\Delta H = \eta_{in} - \eta_{out} \tag{6}$$

where τ_{in} and τ_{out} denote indoor and outdoor temperatures, η_{in} and η_{out} are indoor humidity and outdoor relative humidity, C_p is the air specific heat, V is the volume of the building, ρ is the air density, θ is the heat provided by local SDD heater, σ is the solar radiant energy, λ is the latent vaporization heat, f is the water capacity of the fog system, μ is the heat transfer coefficient, and ε is the evapotranspiration rate of the plants. However, this linear model was still unable to accurately accommodate the complex relationship between τ_i and η_j. This research is motivated by the idea that accurate predictions of changes in temperature and humidity can support greater energy savings by deactivating actuators for certain period of time according to the predicted environmental conditions [12]. An example to this is turning off heaters when the temperature is predicted to rise [5].

In Machine Learning (ML) realm, predicting the future values based on the historical data is known as time series forecasting techniques. In the past, several studies has been conducted to forecast temperature and relative humidity using ML techniques such as Multilayer Perceptron (MLP) and Radial Basis Function (RBF) networks [13], Least Square Support Vector Machine (LSSVM) [14], as well as variant of time series Deep Learning (DL) models including Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU) and Transformer-based techniques [15–17]. This research has produced promising result in terms of prediction accuracy, highlighting the superiority of ML to make an accurate forecasting model compared to a more traditional approach like Albright's Complete Linear Model.

However, despite the DL-based time series forecasting models' capability to learn complex and non-linear patter in time series sequence, they would require large and diverse datasets to generalize well, since they have many trainable parameters [18, 19]. Thus, a simple and traditional ML methods are more advantageous in small dataset availability, especially if the time-series data a relatively exhibit consistent seasonal patterns. Moreover, traditional models offer easier interpretability [20] as simple as extracting the regression coefficient for linear models or extracting the feature importances for the tree-based models.

This study was conducted using limited temperature and relative humidity data collected from several sensors in an SDD facility, with the goal of developing a forecasting model useful for sustainable agricultural practice. Given this data context, traditional ML techniques were chosen due to their advantages under such constraints. The paper is written in a well-structured manner which begin with motivation, methodology (covering dataset collection, modelling approach, and evaluation metrics), and discussion. A major emphasis is placed on data and model analysis, to provide actionable insights to farmers, rather than focusing only on gaining the best model performance on paper.

2. RESEARCH METHODOLOGY

2.1. Dataset Collection. The temperature and humidity time series data were acquired from a pilot SDD facility located in Sunter, North Jakarta using the Elitech GSP-6 temperature and humidity data logger (accessible at https://github.com/dioz95/SmartDomeDataset). To capture temperature and humidity variations within different location, three sensors were strategically positioned at different points within and around the dome: (1) an external sensor placed outside the room near the front door, (2) an internal sensor placed inside the room at the front near the air inlet, and (3) another internal sensor placed inside the room at the back near the ventilation fan.

The sensors were configured to perform measurements at a constant time interval of five minutes, which provided sufficient temporal resolution for time series modeling while maintaining manageable data volume. Data collection was conducted continuously over a 12-day period, spanning from 16 October 2021 to 28 October 2021. Due to minor variations in logging duration, the total recording time slightly differed among the three devices: 11 days 21 hour 20 minutes for the outdoor sensor and 11 days 21 hour 15 minutes for the front indoor sensor and back indoor sensor. Despite these minor discrepancies, the dataset remained temporally aligned across all sensors, offering a comprehensive view of the temperature and humidity dynamics within the dome environment. The whole collected dataset was saved in comma separated value (.csv) format to ensure data transferability and analysis flexibility. Overall, this data collection procedure was considered sufficient to capture short-term temporal pattern of the temperature and humidity values, making the dataset suitable for developing and evaluating forecasting models over multiple prediction horizons.

2.2. Time Series Forecasting. Time series forecasting is a statistical use case created to predict the future values based on the historical values of a time series dataset [21]. In our case, the recorded temperature and humidity data on the facility is considered as time series data, as the data were collected in chronological order. In this research, three distinct time series forecasting algorithms were employed to predict the future temperature and humidity values: (1) Extreme Gradient Boosting (XGBoost), (2) Linear Regression, and (3) Prophet.

2.2.1 *Extreme Gradient Boosting (XGBoost).* XGBoost is an ensemble of tree-based method that known for its superior performance in regression and classification tasks [22, 23]. Instead of fitting one large tree, the model adds trees sequentially, where each new tree is trained to correct the errors made by the previous ones. The final prediction for a given input is the sum of contributions from all trees in the ensemble. As an optimized version of the standard gradient boosting algorithm, XGBoost includes both L1 and L2 regularization terms that penalize overly complex trees [24], to support better generalization. Formally, XGBoost minimizes the prediction loss L by:

$$L = \sum_{n} (y_n - \hat{y}_n)^2 + \sum_{k} \Omega(f_k) \tag{7}$$

where f_k are regression trees, and $\Omega(f)$ is a regularization term (consist of L1 and L2) that penalizes tree complexity to prevent overfitting [25].

However, this model does not inherent sequential capability, emerging the needs for feature engineering, to generate relevant features for time series forecasting purpose. In this experiment, several feature engineering techniques were applied to generate features such as:

- Lags: Lags features are needed to capture short-term temporal patterns that are useful to estimate the future values. Specifically for this research, we used several lag values L where $L \in \{1, 2, 3, 6, 12, 24, 48, 288\}$. The new features are the shifted original values $y_{t-l}^{(s)}$ where s is the unique identifier of each sensor, t is the current time index, and l is the scalar element of lag window values L. The lag value of 288 (as well as rolling window length) was specifically selected to capture daily seasonality, given that the sampling frequency of the data was 5 minutes. This corresponds to one full day of observations, as calculated by (24 hours × 60 minutes) / 5 minutes = 288.
- Rolling statistics: Rolling statistics features are the features generated by computing descriptive statistics of the y values over certain time windows. In this experiment, rolling windows set $W \in \{6, 12, 24, 288\}$ are used to generate new rolling statistics features with statistical measures such as mean and standard deviation.
- Calendar features: Calendar features were extracted from the timestamp index of the time series dataset to describe single time-related features such as hour, day of week, and whether the day is weekend.

The XGBoost model was trained using multiple horizon values to investigate the effect of horizon length to the model's overall performance. The forecasting objective is to predict the values of temperature and humidity for the next horizon h time steps, where for example $h = 12$ correspond to the 5 minutes × 12 = 60 minutes ahead (given 5 minutes is the value of sampling frequency). For each horizon $H \in \{1, .., h\}$, a separated supervised learning problem was defined as,

$$\hat{y}_{t+h} = f_h(x_t) \tag{8}$$

where f_h is an XGBoost model trained exclusively to predict values within horizon h. The XGBoost models were trained using hyperparameters configuration as described in Table 1.

TABLE 1. Hyperparameters configuration for training XGBoost model.

Hyperparameters	**Value**
Objective	Squared error regression (l_2 loss)
Evaluation metric	Root Mean Squared Error (RMSE)
Learning rate	0.05
Maximum tree depth	6
Subsample and column sampling	0.8
Early stopping	50

2.2.2 *Linear Regression*. Linear regression was employed as a model to forecast temperature and humidity using the collected dataset. This model is a simple yet widely used statistical method that assumes a linear relationship between input features and the target variable [26]. Linear regression predicts the unknown value by searching for regression coefficients β_i that explain the linear relationship between the target *y* and features *x* by,

$$\hat{y}_i = \beta_0 + \sum_{i=1}^{p} \beta_i x_i \tag{9}$$

where β_0 is the intercept and *p* is the value of the feature dimension [27]. Due to its simplicity, this provides a more interpretable results compared to the more complex XGBoost models.

Like XGBoost, linear regression was originally designed for standard regression tasks without sequential capabilities. To adapt the model for time series forecasting, feature engineering had to be performed. To promote fairness in the evaluation, the experiment conducted using similar feature engineering techniques as explained previously in the XGBoost section, such as generating lag features, rolling statistics, and calendar features. Specifically, for each timestamp, we created lag features $(y_{\{t-1\}}, y_{\{t-2\}}, y_{\{t-3\}}, y_{\{t-6\}}, y_{\{t-12\}}, \ y_{\{t-24\}}, y_{\{t-48\}}, y_{\{t-288\}})$, rolling means and standard deviations (of window sizes 6, 12, 24, and 288 steps), and time-based categorical features (hour, day of the week, and weekend indicator). This design allows the linear model to capture short- and long-term dependencies, as well as daily seasonality patterns captured in the dataset. The linear regression was also trained with multiple horizon strategy to observe the horizon effect to the model performance, as performed on the XGBoost experiment.

2.2.3 *Facebook Prophet.* Facebook Prophet forecasting model was applied as a comparison model for our proposed XGBoost-based temperature and humidity time series forecasting. Different with other two tested models, Prophet is based on an additive model that decomposes the series into three interpretable components: (1) trend, (2) seasonality, and (3) holiday/events [28]. This model forecast the future value *y(t)* following this equation model,

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (10)$$

where *g(t)* is the trend that modelled using long-term growth using piecewise linear or logistic functions, *s(t)* is the seasonality that modelled recurring patterns with specified periodicity using Fourier series, *h(t)* is the optional component for user-specified special events or anomalies, and ε_t is the error term [29].

In this research, Prophet was configured with daily seasonality enabled to reflect the strong diurnal cycles present in temperature and humidity data without considering holiday or any other anomalous events. The model was trained on the historical data up to the last two days and forecasts were generated for the corresponding horizon in the validation period. This approach allows Prophet to flexibly capture both non-linear trends and recurring daily changes without the need for additional feature engineering as performed in XGBoost and Linear Regression model. Like the other tested models, Prophet was evaluated across multiple forecast horizons to assess its short- and long-term predictive performance.

2.3. Evaluation. To ensure fair evaluation performance of the three models, the primary dataset was split into train and validation set, where train set was used for the models to learn the pattern of the temperature and humidity time series data while the validation set used to test evaluate the trained model. The last three days of the dataset collection period (2021-09-23 to 2021-09-25) were determined as the validation set. During the evaluation, the predicted values $\hat{y}$ of each model were compared with the actual values *y* from the sensor measurement based on the root mean square error (RMSE) and coefficient of determination (R^2) which have this following formulation,

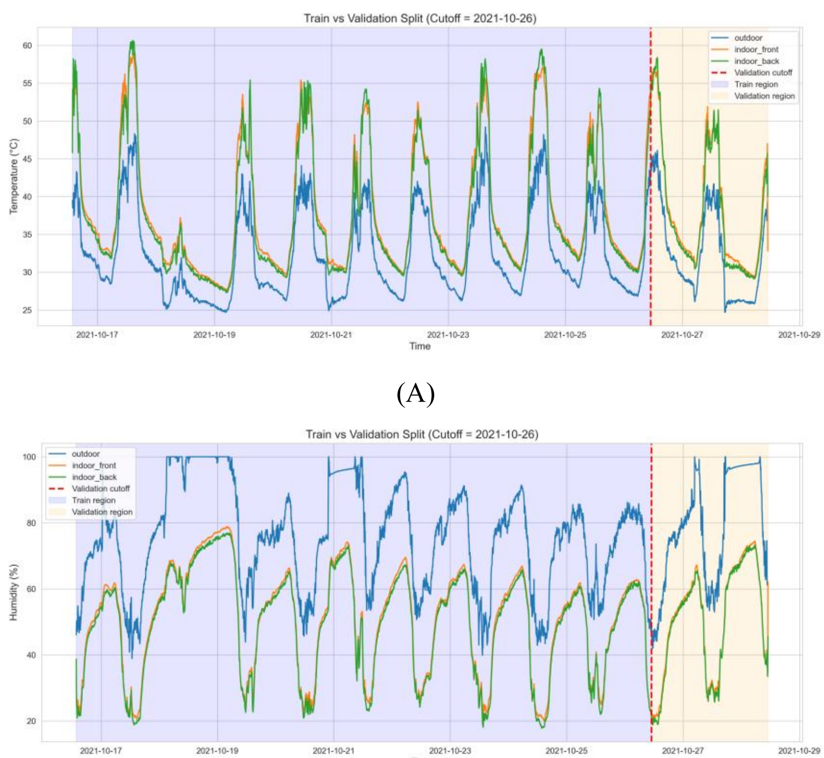
$$RMSE(h) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{t_i+h}^{(s)} - \hat{y}_{t_i+h}^{(s)})\verb|^|2} \quad (11)$$

$$R^2(h) = 1 - \frac{\sum_{i=1}^{N}(y_{t_i+h}^{(s)} - \hat{y}_{t_i+h}^{(s)})\verb|^|2}{\sum_{i=1}^{N}(y_{t_i+h}^{(s)} - \bar{y}_{t_i+h}^{(s)})\verb|^|2} \quad (12)$$

3. Results And Discussion

3.1. Data Analysis. Figure 1 shows the time series data collected from the SDD facility at Sunter, North Jakarta. The measurements were obtained from three sensors which named according to their placement: *outdoor*, *indoor_front*, and *indoor_back*. The figure also highlights the cutoff point to split the training and validation set. The graph indicates that both temperature and humidity exhibit clear cyclical patterns, with noticeable differences between daytime and nighttime averages. These cyclical behaviors are important for forecasting models, as they suggest the presence of strong seasonal components even at the daily scale, which must be captured to achieve accurate predictions. Table 2 provides a detailed summary of these measurements, grouped by each sensor's unique identifier and separated by day (6am to 6pm) and night (6pm to 6am) periods.

The observed variation between day and night highlights the influence of environmental and operational conditions on the facility's microclimate, as shown in the Table 2. External sensors capture fluctuations driven by ambient weather, while internal sensors reflect a more controlled environmental condition regulated by the local air circulation system. The average internal temperatures within the facility are recorded higher compared to external measurement, while internal humidity levels remain lower than the outside, as the dome captures solar radiation through its cover, traps heat inside and raises the internal temperature of the drying chamber [30].



(A)

(B)

FIGURE 1. Collected (A) temperature and (B) humidity time series data from SDD facility in North Jakarta.

Meanwhile, the lower relative humidity inside the dome is a natural consequence of warmer temperature. Warm air has a greater capacity to hold moisture, thus decreasing the relative humidity while sharing the same level of absolute moisture level [31]. Additionally, the controlled air circulation provided by inlets and fans further removes wet air and replaces it with drier air from outside continuously. The combination of higher temperature and lower humidity are desirable for drying facilities as it will accelerate moisture removal from agricultural products, thereby reducing overall drying time [32]. A faster drying process is also helpful to preserve nutritional quality and prevents microbial growth, making the overall product quality better compared to the traditional open-air drying methods [33].

TABLE 2. Average daytime and nighttime temperature and humidity for each sensor ID.

Sensor ID	Period	Average Temperature (°C)	Average Humidity (%)
outdoor	Day	36.24	67.46
	Night	28.46	84.45
indoor_back	Day	43.00	39.27
	Night	32.24	59.44
indoor_front	Day	43.49	40.51
	Night	32.76	60.80

3.2. Models Performance and Analysis. Table 3 highlights the average performance of the three tested models over 12 horizon lengths and 3 sensor placements. One noticeable remark is that XGBoost and Linear Regression have quite similar performance where XGBoost was slightly better at temperature forecasting, while Linear Regression was slightly better at humidity forecasting according to the validation RMSE and R^2 score recorded during the experiment. This highlights the capability of XGBoost to capture nonlinearities and complex interactions in the data makes it highly suitable for modeling the temporal dynamics of temperature and humidity within the facility. Linear Regression, despite using simpler modelling approach, succeed to deliver competitive results, suggesting that a simple linear assumption can approximate short-term dependencies reasonably well in small-sized dataset.

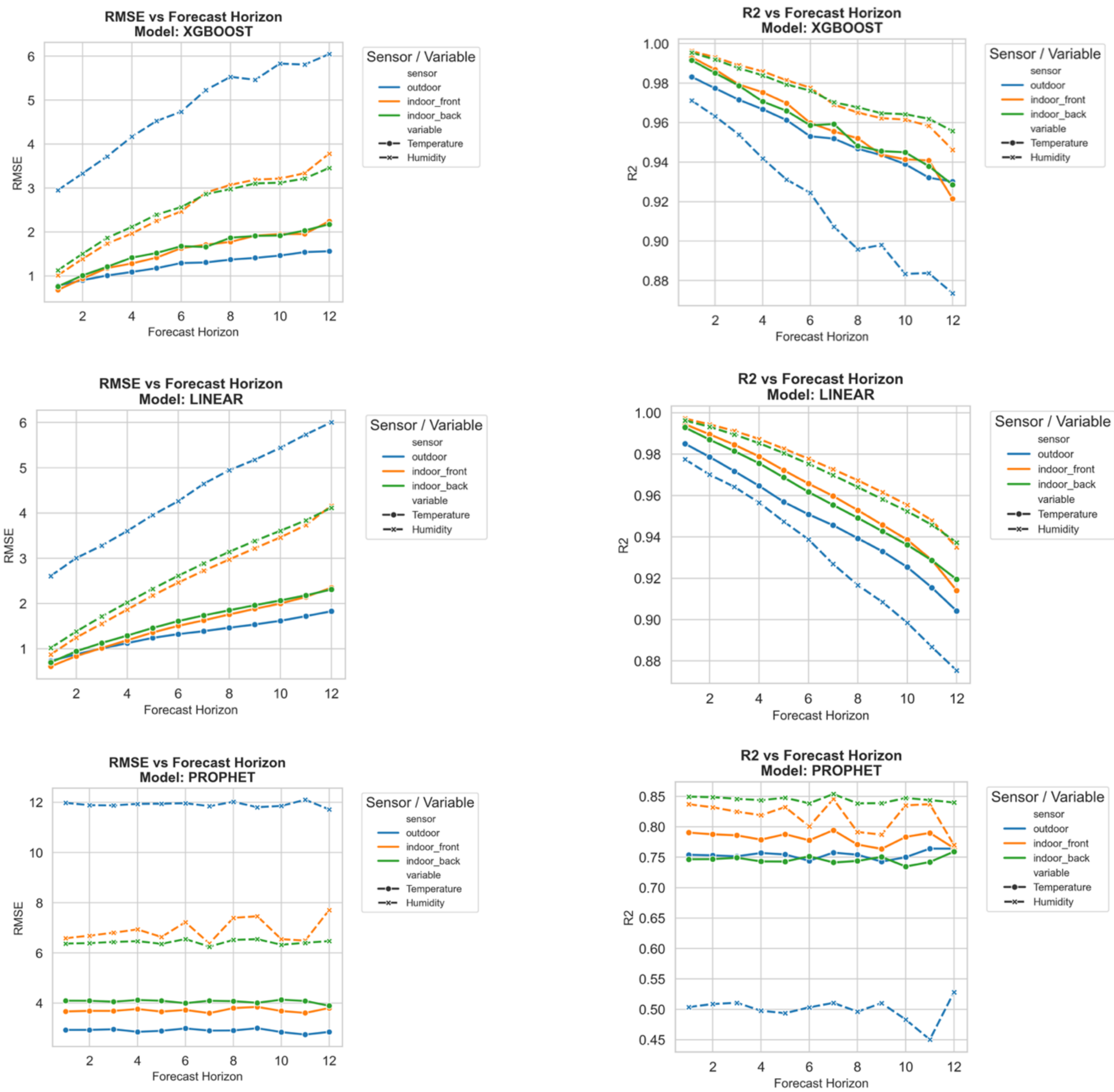
On the other hand, Prophet model scored the lowest overall performance among the tested models. This happens because Prophet relies only on additive decompositions of trend, seasonality, and other optional events. Such an approach is effective for stable and regularly repeating series but struggles with irregularities that might occurs due to weather anomaly or sensor-specific dynamics within the facility. Consequently, Prophet forecasts tend to reflect the baseline cyclical pattern without adequately adapting to nonlinear deviations, resulting in a less accurate prediction compared to the other two models.

TABLE 3. Average daytime and nighttime temperature and humidity for each sensor ID.

Measured Variable	**Average RMSE**			**Average R^2**		
	XGBoost	**LR**	**Prophet**	**XGBoost**	**LR**	**Prophet**
Humidity	3.28	3.20	8.42	0.96	0.96	0.72
Temperature	1.47	1.49	3.56	0.96	0.96	0.76

However, although the tested Prophet model yielded relatively low RMSE and R^2 scores, its performance trend remained stable as the forecasting horizon increased. Prophet predicts future values by fitting one model to trend, seasonality, and any other specific events, then extends these components forward. So, forecast error is relatively uniform across horizons because the same seasonal cycle is being projected, making a plateaued performance that appears consistent across horizons. Meanwhile for the XGBoost and Linear Regression which originally made for standard classification/regression tasks, as the horizon length increases, the target (y_{t+h}) becomes further away from the reference point. Thus, the strength of engineered features (lags, rolling mean and standard deviation, and time features) become less informative within high horizon settings. The complete models' performance trends over multiple horizons and sensor placement can be seen in Figure 2.

Comparing these results to other works, we found that our modelling approach (although simple and fast) could provide a comparable result to the others that utilize more complex models and/or larger size of dataset. For example, [13] had compare forecasting performance on temperature and humidity time series dataset for the next 30-minute using Bayesian optimized RBF model and multi-layer perceptron (MLP) and retrieve the best result of humidity forecasting with RMSE of 4.29% and an R^2 of 0.95 and with RMSE of 1.58°C and R^2 of 0.96 for temperature forecasting. Meanwhile, the best Linear Regression model in our experiment can exceed these numbers by scoring RMSE of 2.46% and an R^2 of 0.98 for humidity forecasting, while scoring RMSE of 1.51°C and an R^2 of 0.97 for temperature forecasting within the same forecasting horizon. Moreover the models constructed by [13] consider other variables besides historical data such as indoor soil temperature and light intensity, highlighting the efficiency of our proposed approach.



(A) (B)

FIGURE 2. Model performance trend across multiple horizons, measured by (A) RMSE and (B) R^2.

Another research also describes an effort to predict temperature and relative humidity values on a greenhouse facility in the South China region [14]. The best model was least-squares support-vector machines (LSSVM) with RMSE of 0.87°C and an R^2 of 0.92 for temperature and RMSE of 3.91% and an R^2 of 0.95 for relative humidity forecasting for the next 15 minutes. Our best Linear Regression model can exceed all these numbers, except the RMSE of temperature prediction where our model scores a slightly higher value of 1.02°C within the same forecasting horizon. However, this can be understood since [14] collected the data in a longer period of time of 44 days. These comparisons highlight the effectiveness of gradient boosting algorithms in capturing complex patterns in forecasting temperature and relative humidity in a short-term period, despite using relatively small-sized dataset on the experiment.

3.2.1. *XGBoost Model Analysis*. Table 4 presents the five most important features of the XGBoost model for forecasting temperature and humidity values across the three sensors with a six-step-ahead horizon, which is equivalent to 30 minutes. The results indicate that short-term lag variables (lags 1 and 2) and hour of the day provide the highest predictive contribution for temperature forecasting. These are followed by rolling statistics features such as the six-point rolling mean and the 24-point rolling standard deviation, along with the day of week and lag 48 which also plays a notable role in capturing daily patterns.

TABLE 4. Feature importances of XGBoost model to forecast temperature and humidity.

Variable	**Sensor ID**	**I**	**II**	**III**	**IV**	**V**
Humidity	*outdor*	*lag_1*	*lag_2*	*hour*	*roll_mean_6*	*dayofweek*
	indoor_back	*lag_1*	*lag_2*	*hour*	*roll_std_24*	*lag_48*
	indoor_front	*lag_1*	*lag_2*	*hour*	*lag_48*	*roll_std_24*
Temperature	*outdor*	*lag_1*	*lag_2*	*hour*	*roll_mean_6*	*dayofweek*
	indoor_back	*lag_1*	*lag_2*	*hour*	*roll_mean_6*	*is_weekend*
	indoor_front	*lag_1*	*lag_2*	*hour*	*roll_mean_6*	*dayofweek*

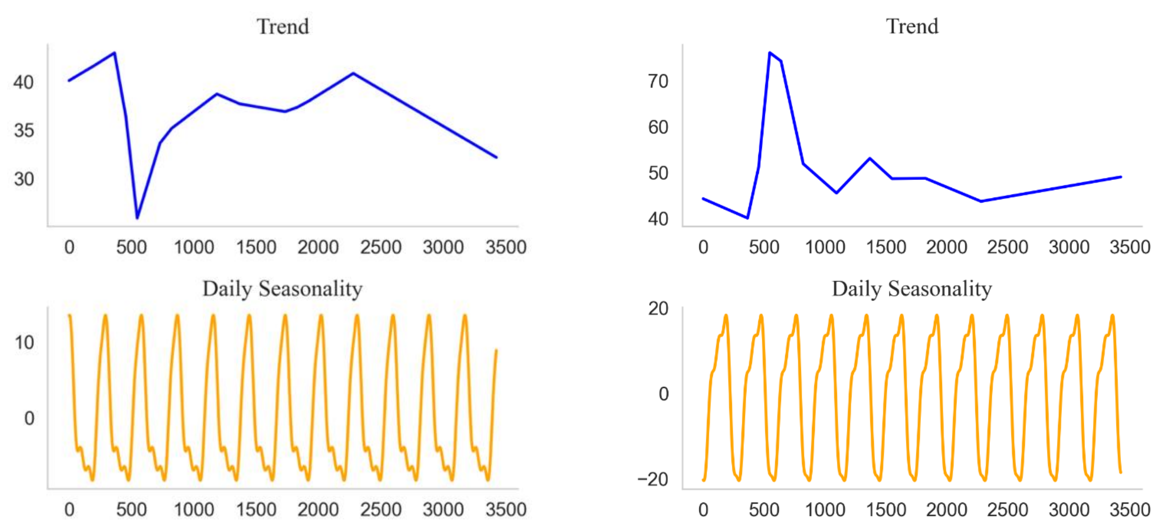
The humidity forecasting model shares quite similar feature importance, except the 24-point rolling standard deviation and lag 48 features do not appear on the list. These results demonstrate that the temperature and humidity changes around the facility are influenced by both short-term dependencies and repeated daily cycles, which the XGBoost model effectively captures for improved forecasting accuracy.

3.2.2. *Linear Regression Model Analysis***.** Table 5 presents the ranking of regression coefficients for temperature and humidity forecasting across six horizons. Slightly different from the XGBoost model, our linear regression model forecast future humidity and temperature value by considering wider range of features, including short-term lags (e.g., lag 1 and lag 2), temporal indicators (weekend and day-of-week), as well as rolling statistics such as rolling means (6, 12, and 24 windows) and rolling standard deviations with (12 and 24 windows). One advantage of ranking feature importances based on the regression coefficients is that it provides better interpretability since regression coefficients directly indicate both the magnitude and the direction of influence, although it relies on simple linear assumptions between the independent and dependent variables.

TABLE 5. Feature importances of linear regression model to forecast temperature and humidity.

Variable	Sensor ID	I	II	III	IV	V
Humidity	*outdor*	*lag_1*	*roll_std_24*	*lag_2*	*roll_mean_24*	*dayofweek*
	indoor_back	*lag_1*	*lag_2*	*roll_mean_6*	*roll_mean_12*	*roll_mean_24*
	indoor_front	*lag_1*	*lag_2*	*roll_mean_6*	*roll_mean_12*	*roll_mean_24*
Temperature	*outdor*	*lag_1*	*roll_std_24*	*roll_std_12*	*lag_2*	*lag_12*
	indoor_back	*lag_1*	*lag_2*	*roll_std_24*	*roll_std_12*	*roll_mean_24*
	indoor_front	*lag_1*	*lag_2*	*roll_mean_6*	*roll_mean_12*	*roll_std_24*

3.2.3. *Facebook Prophet Model Analysis***.** Analysis on the trained Prophet model performed by extracting the decomposed time series features such as trend and daily seasonality as illustrated on Figure 3 (A) and (B) for both temperature and humidity forecasting for *indoor_back* sensor. The level shows a smoothed estimate of the current value of the series, the trend shows the estimated rate of change in the level, and seasonality reveals the cyclical pattern of the time series value. For temperature, the level shows values ranging from +- 20 °C to +- 40 °C with sudden drops around the time index of 1000. The seasonality plot displays a strong cyclical pattern with a clear daily periodicity where peaks and valleys occur at regular intervals. Interestingly, the extracted humidity level and seasonality components exhibit the similar but inverted pattern compared to the temperature ones. This behavior occurs due to the physical relationship between temperature and relative humidity: as the temperature increase, the air capacity to hold moisture also increases, making the relative humidity measurement decrease.



(A) (B)

FIGURE 3. Visualization of level, trend, and seasonality components of the (A) temperature and (B) relative humidity time series of the *indoor_back* sensor.

3.3. Further Research Direction. Future further research investigation could provide additional, deeper insights by using larger data sets crossing multiple years and involving multiple sensor locations, so the model may capture temperature and humidity patterns from multiple seasons (dry and wet), different years, and additional sensor locations. This would allow the utilization of data-hungry architecture such as LSTM and self-attention-based models. Moreover, a longer dataset also allows safer use of train/validation/test splits that preserve temporal order and still leave sufficient held-out data for robust evaluation.

With recent smart agricultural trends involving more and more technologies being integrated with greenhouses such as Smart Dryer Dome (SDD) in agricultural complexes in Indonesia [4, 15], United States [34–37] and other agricultural-based economies all around the world, the potential of the use of data-driven approach might be increasing too and so will its potential impacts for agricultural productivity be for the said economies. The current study paves the way and opens possibilities for further, richer, more exciting research investigations enabling smarter and more productive agriculture for the economies and for the betterment of our societies. For instance, when photovoltaics (PV) is integrated with an agricultural greenhouse [37], the same plot of land that previously is just used for agricultural productivity, has now become productive too for renewable energy production [34] which is critical for energy transition and averting the climate crisis in the United States. The use of data-driven approach here would be enabled for not only increasing agricultural productivity (by more precise temperature and humidity prediction/control inside the greenhouse, etc.), but also for saving the energy use for the agricultural operations (by allowing more optimum battery charging cycles [38] and hence increase its lifetime) and thus allows for more sustainable agriculture and energy transition in our modern world [39, 40].

Another research use case such as probabilistic forecasting with prediction intervals and anomaly detection for sensor fault identification could be leveraged to enrich the impact of the research. Probabilistic forecasting is especially useful in the case where uncertainty factors are present. For example, when weather/climate anomaly occurs during a certain period or when the drying chamber is less occupied due to operational break. Anomaly detection on the other hand is practically helpful to know whether the recorded values are subject to sensor fault. However, this approach would require an additional effort in labelling the time series data to distinguish the period where the sensor operates normally from those affected by faults. Together, these strategies would lead to more resilient decision-support tools that ensure consistent drying performance and operational efficiency even under uncertain environmental conditions.

4. CONCLUSION

This report presents original research in applying a time-series machine learning approach for forecasting temperature and relative humidity values from the primary data collected by three sensors installed in a SDD facility in the North Jakarta region. Data-wise, recorded temperature and humidity data exhibit distinct seasonal patterns, fluctuating upwards and downwards within a relatively consistent interval on a daily basis. Moreover, the recorded temperature values were inversely proportional to the recorded relative humidity values, suggesting an explainable physical interaction among these physical measures.

Multiple algorithms were evaluated to forecast such variables within the interval of 5 minutes, up to the duration of 1 hour such as XGBoost, Linear Regression, and Facebook Prophet. Since XGBoost and Linear Regression were not designed for time series forecasting, several features need to be engineered to produce independent variables, while the Prophet model automatically decomposes the time series into trend and seasonality values. From the experiment, XGBoost and Linear Regression were the top performing models with the highest average RMSE and R^2 score across 12 horizons, while Prophet was the lowest performing model. However, despite inferior performance of Prophet among the tested algorithms, this model recorded a steady performance across increasing prediction horizon, while XGBoost and Linear Regression recorded a decreased performance trend in higher horizon values. These findings highlight the superiority of XGBoost and Linear Regression in short-term forecasting with seasonal time-series dataset, as well as highlight the consistency of Prophet forecasting model in long-horizon settings.

AUTHOR CONTRIBUTIONS

Advendio Desandros: Methodology, Writing – Original draft, Formal analysis.

Gregorius Natanael Elwirehardja: Methodology, Writing – Review & Editing.

Endang Djuana: Conceptualization, Investigation, Writing – Review & Editing, Funding acquisition.

Arief Suriadi Budiman: Conceptualization, Writing – Review & Editing, Funding acquisition.

Bens Pardamean: Conceptualization, Writing – Review & Editing, Funding acquisition.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] R. Osae, G. Essilfie, R.N. Alolga, S. Akaba, X. Song, et al., Application of Non-thermal Pretreatment Techniques on Agricultural Products Prior to Drying: A Review, J. Sci. Food Agric. 100 (2020), 2585-2599. https://doi.org/10.1002/jsfa.10284.

[2] P. Udomkun, S. Romuli, S. Schock, B. Mahayothee, M. Sartas, et al., Review of Solar Dryers for Agricultural Products in Asia and Africa: An Innovation Landscape Approach, J. Environ. Manag. 268 (2020), 110730. https://doi.org/10.1016/j.jenvman.2020.110730.

[3] P. Morya, M. Agarwal, G.D. Agarwal, A Comprehensive Insight into Solar Drying Systems for Agricultural and Industrial Applications, Sol. Energy 293 (2025), 113494. https://doi.org/10.1016/j.solener.2025.113494.

[4] A.S. Budiman, F. Gunawan, E. Djuana, B. Pardamean, S. Romeli, et al., Smart Dome 4.0: Low-Cost, Independent, Automated Energy System for Agricultural Purposes Enabled by Machine Learning, J. Phys.: Conf. Ser. 2224 (2022), 012118. https://doi.org/10.1088/1742-6596/2224/1/012118.

[5] N. Hananda, A. Kamul, C. Harito, E. Djuana, G.N. Elwirehardja, et al., Solar Drying in Indonesia and Its Development: A Review and Implementation, IOP Conf. Ser.: Earth Environ. Sci. 1169 (2023), 012084. https://doi.org/10.1088/1755-1315/1169/1/012084.

[6] N.R. Pochont, M.N. Mohammad, B.T. Pradeep, P.V. Kumar, A Comparative Study of Drying Kinetics and Quality of Indian Red Chilli in Solar Hybrid Greenhouse Drying and Open Sun Drying, Mater. Today: Proceedings 21 (2020), 286-290. https://doi.org/10.1016/j.matpr.2019.05.433.

[7] J.W. Baurley, A. Budiarto, M.F. Kacamarga, B. Pardamean, A Web Portal for Rice Crop Improvements, Int. J. Web Portals 10 (2018), 15-31. https://doi.org/10.4018/IJWP.2018070102.

[8] S. Zhou, A. Shah, P. Leung, X. Zhu, Q. Liao, A Comprehensive Review of the Applications of Machine Learning for HVAC, DeCarbon 2 (2023), 100023. https://doi.org/10.1016/j.decarb.2023.100023.

[9] F. Sevgin, Machine Learning-Based Temperature Forecasting for Sustainable Climate Change Adaptation and Mitigation, Sustainability 17 (2025), 1812. https://doi.org/10.3390/su17051812.

[10] S. Suherman, H. Hadiyanto, M.A. Asy-Syaqiq, G.‘. Ghassani, M. Ajundasari, Energy and Exergy Performance Evaluation of a Drying Coffee Beans System Using a Photovoltaic–Direct Solar Dryer at Different Drying Temperature Conditions, Int. J. Ambient. Energy 45 (2024), 2344548. https://doi.org/10.1080/01430750.2024.2344548.

[11] L.D. Albright, K.G. Arvanitis, A.E. Drysdale, Environmental Control for Plants on Earth and in Space, IEEE Control. Syst. 21 (2001), 28-47. https://doi.org/10.1109/37.954518.

[12] R.E. Caraka, M. Tahmid, R.M. Putra, A. Iskandar, M.A. Mauludin, et al., Analysis of Plant Pattern Using Water Balance and Cimogram Based on Oldeman Climate Type, IOP Conf. Ser.: Earth Environ. Sci. 195 (2018), 012001. https://doi.org/10.1088/1755-1315/195/1/012001.

[13] C. Yan, T. Na, Q. Zhen, Y. Sun, K. Liu, Prediction of Air Temperature and Humidity in Greenhouses via Artificial Neural Network, PLOS One 20 (2025), e0325650. https://doi.org/10.1371/journal.pone.0325650.

[14] X. Wei, Y. Luo, X. Zhou, J. Zhao, H. Lu, et al., Temperature and Relative Humidity Prediction in South China Greenhouse Based on Machine Learning, Sci. Rep. 15 (2025), 24855. https://doi.org/10.1038/s41598-025-08964-6.

[15] F.E. Gunawan, A.S. Budiman, B. Pardamean, E. Juana, S. Romeli, et al., Multivariate Time-Series Deep Learning for Joint Prediction of Temperature and Relative Humidity in a Closed Space, Procedia Comput. Sci. 227 (2023), 1046-1053. https://doi.org/10.1016/j.procs.2023.10.614.

[16] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Sequence to Sequence Deep Learning Architecture for Forecasting Temperature and Humidity Inside Closed Space, in: 2022 10th International Conference on Cyber and IT Service Management (CITSM), IEEE, 2022: pp. 1-7. https://doi.org/10.1109/CITSM56380.2022.9936008.

[17] K.E. Setiawan, G.N. Elwirehardja, B. Pardamean, Comparison of Deep Learning Sequence-to-Sequence Models in Predicting Indoor Temperature and Humidity in Solar Dryer Dome, Commun. Math. Biol. Neurosci. 2022 (2022), 98. https://doi.org/10.28919/cmbn/7655.

[18] X. Song, L. Deng, H. Wang, Y. Zhang, Y. He, et al., Deep Learning-Based Time Series Forecasting, Artif. Intell. Rev. 58 (2024), 23. https://doi.org/10.1007/s10462-024-10989-8.

[19] J. Baurley, A. Perbangsa, A. Subagyo, B. Pardamean, A Web Application and Database for Agriculture Genetic Diversity and Association Studies, Int. J. Biosci. Biotechnol. 5 (2013), 33-42. https://doi.org/10.14257/ijbsbt.2013.5.6.04.

[20] J.W. Choi, M.S. Hidayat, S.B. Cho, W. Hwang, H. Lee, et al., Recent Trends in Machine Learning, Deep Learning, Ensemble Learning, and Explainable Artificial Intelligence Techniques for Evaluating Crop Yields Under Abnormal Climate Conditions, Plants 14 (2025), 2841. https://doi.org/10.3390/plants14182841.

[21] M.A.B. Syed, M.R. Hasan, N.I. Chowdhury, M.H. Rahman, I. Ahmed, A Systematic Review of Time Series Algorithms and Analytics in Predictive Maintenance, Decis. Anal. J. 15 (2025), 100573. https://doi.org/10.1016/j.dajour.2025.100573.

[22] B.N. Rumahorbo, K. Nanggala, G.N. Elwirehardja, B. Pardamean, Analyzing Important Statistical Features from Facial Behavior in Human Depression Using XGBoost, Commun. Math. Biol. Neurosci. 2023 (2023), 35. https://doi.org/10.28919/cmbn/7916.

[23] J.S. Sulistyawan, K. Setyo Nugroho, B. Pardamean, Explainable Tree-Based Boosting Algorithms for Smoker Detection Using Bio-Signals, in: 2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), IEEE, 2024: pp. 114-119. https://doi.org/10.1109/ISRITI64779.2024.10963483.

[24] R.E. Caraka, S.A. Bakar, B. Pardamean, A. Budiarto, Hybrid Support Vector Regression in Electric Load during National Holiday Season, in: 2017 International Conference on Innovative and Creative Information Technology (ICITech), IEEE, 2017: pp. 1-6. https://doi.org/10.1109/INNOCIT.2017.8319127.

[25] D. Tarwidi, S.R. Pudjaprasetya, D. Adytia, M. Apri, An Optimized Xgboost-Based Machine Learning Method for Predicting Wave Run-Up on a Sloping Beach, MethodsX 10 (2023), 102119. https://doi.org/10.1016/j.mex.2023.102119.

[26] Z. Patonai, R. Kicsiny, G. Géczi, Multiple Linear Regression Based Model for the Indoor Temperature of Mobile Containers, Heliyon 8 (2022), e12098. https://doi.org/10.1016/j.heliyon.2022.e12098.

[27] Y. Li, K. Wei, K. Chen, J. He, Y. Zhao, et al., Forecasting Monthly Water Deficit Based on Multi-Variable Linear Regression and Random Forest Models, Water 15 (2023), 1075. https://doi.org/10.3390/w15061075.

[28] S.J. Taylor, B. Letham, Forecasting at Scale, Am. Stat. 72 (2018), 37-45. https://doi.org/10.1080/00031305.2017.1380080.

[29] M. Navratil, A. Kolkova, Decomposition and Forecasting Time Series in the Business Economy Using Prophet Forecasting Model, Cent. Eur. Bus. Rev. 8 (2019), 26-39. https://doi.org/10.18267/j.cebr.221.

[30] Y.D. Kokate, P.R. Baviskar, S.D. Suryawanshi, Performance Investigation of Newly Developed Novel Hemispherical Solar Dryer for Sustainable Food Preservation: Comparative Analysis with Traditional Methods, Sol. Energy 283 (2024), 113036. https://doi.org/10.1016/j.solener.2024.113036.

[31] E.L. Rulazi, J. Marwa, B. Kichonge, T. Kivevele, Development and Performance Evaluation of a Novel Solar Dryer Integrated with Thermal Energy Storage System for Drying of Agricultural Products, ACS Omega 8 (2023), 43304-43317. https://doi.org/10.1021/acsomega.3c07314.

[32] Z. Wang, L. Xu, D. Liu, Q. Zhang, A. Hu, et al., Effects of Air Temperature and Humidity on the Kinetics of Sludge Drying at Low Temperatures, Energies 14 (2021), 7722. https://doi.org/10.3390/en14227722.

[33] B.N.D. Pagukuman, M.K. Wan Ibrahim, A Review of the Significance Effect of External Factors of the Solar Dyer Design to Dried Foods Product Quality, J. Eng. Des. Technol. 20 (2021), 1765-1786. https://doi.org/10.1108/JEDT-01-2021-0033.

[34] I. Bermudez Rivera, F.S. Mozar, C. Harito, D.N.N. Putri, A. Zago, et al., AGRO-PV Dome - Developing Agrivoltaics Solution Suitable for OR Through Multidisciplinary Hands-On Student Design Project, in: 2025 ASEE Annual Conference & Exposition Proceedings, ASEE Conferences. https://doi.org/10.18260/1-2--57583.

[35] S.K. Tippabhotla, N.G. Diesta, X. Zhang, S. Sridhara, C. Stan, et al., Thermomechanical Residual Stress Evaluation in Multi-Crystalline Silicon Solar Cells of Photovoltaic Modules with Different Encapsulation Polymers Using Synchrotron X-Ray Microdiffraction, Sol. Energy Mater. Sol. Cells 193 (2019), 387-402. https://doi.org/10.1016/j.solmat.2019.01.016.

[36] V.A. Handara, G. Illya, S.K. Tippabhotla, R. Shivakumar, A.S. Budiman, Center for Solar Photovoltaics (CPV) at Surya University: Novel and Innovative Solar Photovoltaics System Designs for Tropical and Near-Ocean Regions (an Overview and Research Directions), Procedia Eng. 139 (2016), 22-31. https://doi.org/10.1016/j.proeng.2015.09.211.

[37] A. Budiman, G. Illya, S. Anbazhagan, S. Tippabhotla, W. Song, et al., Enabling Lightweight Polycarbonate-Polycarbonate (pc-Pc) Photovoltaics Module Technology – Enhancing Integration of Silicon Solar Cells into Aesthetic Design for Greener Building and Urban Structures, Sol. Energy 235 (2022), 129-139. https://doi.org/10.1016/j.solener.2022.02.018.

[38] A.S. Budiman, R. Fajarna, M. Asrol, F.S. Mozar, C. Harito, et al., Estimation of Lead Acid Battery Degradation—A Model for the Optimization of Battery Energy Storage System Using Machine Learning, Electrochem 6 (2025), 33. https://doi.org/10.3390/electrochem6030033.

[39] H. Soeparno, A.S. Perbangsa, B. Pardamean, Best Practices of Agricultural Information System in the Context of Knowledge and Innovation, in: 2018 International Conference on Information Management and Technology (ICIMTech), IEEE, 2018: pp. 489-494. https://doi.org/10.1109/ICIMTech.2018.8528187.

[40] M.F. Kacamarga, B. Pardamean, H. Wijaya, Lightweight Virtualization in Cloud Computing for Research, in: R. Intan, C.-H. Chi, H.N. Palit, L.W. Santoso (Eds.), Intelligence in the Era of Big Data, Springer, Berlin, Heidelberg, 2015: pp. 439–445. https://doi.org/10.1007/978-3-662-46742-8_40.