



Thermal Comfort Projection on Northern Coast of Central Java Using Machine Learning

Afandi¹, Aji Supriyanto^{2*}

^{1,2}Department of Information Technology, Faculty of Information Information and Industry, Universitas Stikubank, Semarang, Indonesia

¹afandi0022@mhs.unisbank.ac.id, ²ajisup@edu.unisbank.ac.id

Abstract

The Thermal Humidity Index (THI) serves as a critical measure of environmental thermal comfort, particularly vital for living beings in densely populated regions. This study projects and classifies THI in the western northern coastal areas of Central Java using Machine Learning (ML) techniques. Utilizing temperature and humidity data from 2018 to 2024, THI projections were conducted using the XGBoost algorithm, whereas comfort level classifications were performed using the Random Forest algorithm. The results indicate that Semarang City, eastern Kendal, Pemalang, and Tegal frequently experienced slightly uncomfortable conditions (THI 27–30), particularly during the rainy and transitional seasons, whereas other regions maintained comfortable levels (THI < 27). The THI projection model for 2025–2029 achieved an accuracy of 73%, while the classification model attained a remarkably high accuracy of 99.94%. These findings highlight the need for enhanced regional management strategies in areas with reduced thermal comfort.

Keywords: coastal north central java; random forest (RF); temperature humidity index (THI); thermal comfort; XGBoost

How to Cite: A. Afandi and Aji Supriyanto, “Thermal Comfort Projection on Northern Coast of Central Java Using Machine Learning”, *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 9, no. 4, pp. 865 - 878, Aug. 2025.
Permalink/DOI: <https://doi.org/10.29207/resti.v9i4.6537>

Received: April 13, 2025

Accepted: June 16, 2025

Available Online: August 19, 2025

*This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia*

1. Introduction

The rapid increase in population, particularly in urban areas, has emerged as a major concern in many countries, including Indonesia. Its impacts have become increasingly evident, especially in coastal regions. Significant land-use changes have occurred due to various factors, along with urban expansion that contributes to urban sprawl, resulting in the loss of green spaces and the alteration of microclimatic characteristics [1] - [3]. The assessment of thermal comfort based on the Temperature-Humidity Index (THI) is crucial for sustainable urban planning, particularly in the northern coastal region of western Central Java, encompassing areas from Semarang to Tegal.

Thermal comfort directly affects public health and lifestyle, making the THI a useful proxy for assessing the impact of environmental changes. THI is a composite indicator that evaluates human comfort based on a combination of air temperature and relative humidity[4]. Surface Air Temperature (SAT) in Java Island increased by approximately 0.11°C to 1.24°C

during the period 1990–2019, with the highest temperatures recorded in the northern region of the island and the lowest in the western highlands [5]. This temperature increase has been influenced in part by the reduction of green open spaces due to large-scale infrastructure development[6]. These changes have contributed to elevated THI values, which reduce thermal comfort, increase the risk of heat-related health issues, and negatively affect productivity and quality of life for communities in affected areas.

The THI is widely used to assess thermal comfort by utilizing the variability of temperature and humidity, which are analyzed based on observational data. In Indonesia, such data can be obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG). Previous studies on the Temperature-Humidity Index (THI) in the northern coastal region of Central Java have primarily employed conventional methods. While these approaches provide an overview of average thermal conditions, they fall short in capturing the spatial-temporal dynamics that are critical in this environment, where elevated THI values indicate

significant thermal stress[7]. Therefore, these limitations underscore the need for a more holistic approach, one that can effectively ensure optimal levels of thermal comfort in practice.

On the other side of current research, emerging artificial intelligence (AI)-based methods offer more effective and efficient alternatives for predicting the Temperature-Humidity Index (THI). Several AI studies with Machine Learning (ML) have demonstrated that Random Forest (RF) excels in effectively processing complex spatial-temporal patterns due to its parallel processing capabilities[8], [9]. Additionally, RF is known for its ability to reduce overfitting and enhance prediction stability. Extreme Gradient Boosting (XGBoost) leverages strong regularization techniques to improve predictive accuracy[10]. As an alternative to conventional THI calculations, this study proposes a more adaptive approach by integrating RF and XGBoost to project THI in the western part of the northern coast of Central Java, aiming to improve the accuracy of thermal comfort analysis.

Research on thermal comfort continues to advance by utilizing THI as a key indicator for assessing the impact of temperature and humidity on human comfort. A study by Diyah et. Al.,(2024), indicated that urbanization in Bandung City has led to an increase in surface temperature, with a reduction in vegetated land contributing to a decline in thermal comfort. THI data revealed a dominance of uncomfortable zones ($>26^{\circ}\text{C}$) in 2023, despite a slight increase in moderately comfortable zones ($24\text{--}26^{\circ}\text{C}$). These findings suggest that land cover changes play a significant role in reducing thermal comfort and should be considered in urban spatial planning [11].

On the other side of current research, investigated ML algorithms for thermal comfort prediction and found that XGBoost achieved the highest accuracy in modeling the THI compared to other algorithms such as Random Forest (RF) and Neural Networks (NN). The findings indicate that the application of machine learning algorithms can enhance the accuracy of thermal comfort predictions and provide a robust foundation for more adaptive environmental planning[12].

This study aims to analyze the changing patterns of the THI in urban coastal areas of western Central Java, from Semarang City to Tegal Regency, using temperature and humidity data from BMKG observation stations. Additionally, this research develops a THI prediction model utilizing Random Forest (RF) for thermal comfort classification and Extreme Gradient Boosting (XGBoost) to enhance projection accuracy. Furthermore, the study evaluates the model's performance in capturing the spatiotemporal variations of THI and formulates mitigation and adaptation strategies to improve thermal comfort and the quality of life for local communities.

The key recommendations of this study encompass three aspects: health, sustainable urban planning, and outdoor workers. In the health sector, the recommendations focus on mitigating the impact of thermal stress through improved healthcare facilities and the provision of green open spaces. In sustainable urban planning, the findings support the development of green spaces, environmentally friendly building designs, and adaptation strategies for changing climate patterns. Meanwhile, for outdoor workers, this study advocates for more flexible work scheduling and the provision of protective facilities to minimize heat exposure and safeguard health.

2. Methods

This study utilizes temperature and humidity datasets from BMKG observation stations in the western part of the northern coastal region of Central Java, covering the period from 2018 to 2024 over a span of seven years. The THI variables used are Temperature and Humidity which affect human comfort. The analysis process consists of four main stages, as illustrated in Figure 1.

2.1 Data Preprocessing

Data preprocessing is essential for eliminating noise, handling missing values, and refining data to ensure its readiness for further analysis[13],[14]. The data processing consists of five key stages in handling temperature (T) and relative humidity (RH) data, including Exploratory Data Analysis (EDA), which serves as an initial step in preprocessing to comprehend the dataset structure and detect potential anomalies before conducting further analysis[15].

In this study, the dataset is read using the `read_csv()` function from the Pandas library, enabling efficient processing of data in CSV format. Subsequently, an initial exploration is performed using the `df.info()` function to examine the data count, variable types, and potential missing values. This step is crucial for identifying dataset characteristics and preparing for subsequent preprocessing stages. Moreover, EDA is utilized to analyze patterns in large-scale data[16], playing a pivotal role in uncovering hidden correlations and patterns, thereby supporting optimal model selection. The next step involves displaying descriptive statistics to understand data distribution and variability. The `df.describe()` function is employed to compute statistical measures such as mean, median, standard deviation, minimum and maximum values, and quartiles for each variable in the dataset. Descriptive statistics are critical in identifying outliers or non-normal data distribution patterns, which may influence the analysis outcomes.

Subsequently, handling missing values is performed by detecting and imputing or removing NaN values[17]. This process is conducted using the `.isnull().any(axis=1)` method to identify whether any variables contain missing values and to assess their proportion [18]. After identifying missing values, the

next step is to handle them through data removal using `df.dropna(inplace=True)`[19]. Additionally, duplicate rows are removed to ensure data cleanliness and improve the reliability of subsequent analysis. The method `df1 = df.drop_duplicates()`, is employed to

eliminate duplicate entries from the dataset. This preprocessing procedure ensures that the dataset is properly prepared for predictive modeling with high accuracy and reliability [20].

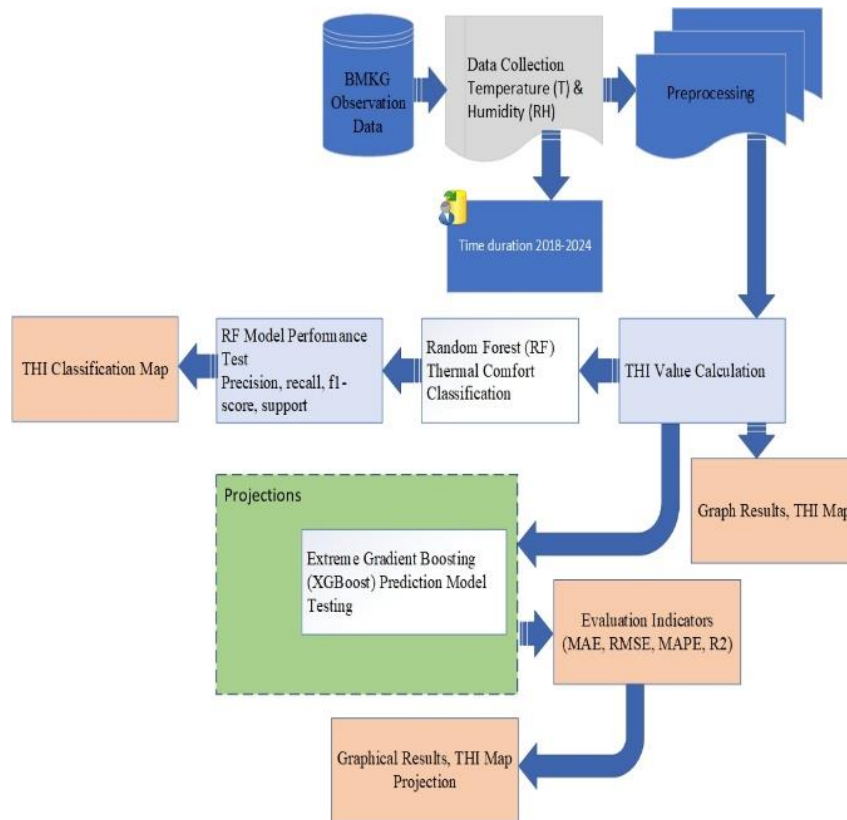


Figure 1. Data Processing Method for THI Projection in Thermal Comfort Using (RF, XGBoost)

2.2 Calculating the THI for the Period 2018–2024

The THI is utilized as an indicator of thermal comfort by considering air temperature and relative humidity[21]. The THI calculation in this study the formula developed using Equation 1[22]:

$$THI = 0.8 * T + \frac{RH * T}{500} \quad (1)$$

The Temperature-Humidity Index (THI) is calculated based on air temperature (T) in degrees Celsius (°C) and relative humidity (RH) in percentage (%). This formula includes a conversion factor of 500 to adjust the units of temperature and humidity. The constant 0.8 in $0.8 \times T$ represents the direct influence of temperature on thermal comfort, while the $(RH \times T)/500$ component reflects the combined effect of temperature and humidity. Higher relative humidity results in an increased THI value, indicating a higher level of thermal discomfort.

2.3 Thermal Comfort Classification Using Random Forest (RF)

Random Forest (RF) is utilized in this study to classify thermal comfort levels based on THI categories, where the THI classification determines comfort levels concerning heat stress. A THI value of less than 27 is

considered comfortable, 27–30 is categorized as slightly uncomfortable, 30–40 represents moderate heat, 40–45 indicates severe heat, and 45–54 is classified as dangerous. A THI value above 54 signifies extreme heat stress[23].

The classification process is carried out by dividing the dataset into thermal comfort categories based on the THI threshold values. Model performance evaluation is conducted using precision, recall, and F1-score metrics to assess the accuracy of the classification.

2.4 THI Projection Using XGBoost

Extreme Gradient Boosting (XGBoost) was implemented in this study to predict THI values based on temperature and humidity data from 2018 to 2024. XGBoost was selected due to its ability to handle large datasets, optimize training processes through boosting techniques, and apply strong regularization to prevent overfitting [24].

The model was evaluated using several metrics, including R-squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to assess prediction errors. The predicted THI values generated by XGBoost were compared with actual values to evaluate model

accuracy and performance in projecting thermal comfort in the northern coastal region of Central Java.

3. Results and Discussions

The analysis of the Temperature Humidity Index (THI) in this study not only involves calculating the index itself but also integrates machine learning techniques to enhance the accuracy of THI classification and projection. The Random Forest (RF) model is employed to classify thermal comfort levels, while the XGBoost model is applied to project THI values with greater precision [25], [26]. The models are evaluated using accuracy metrics, including precision, recall, and F1-score, along with R-squared (R^2), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) to ensure the reliability of THI predictions.

3.1 Research Location Map

The research area covers the western north coast of Central Java, including the cities of Semarang, Kendal, Batang, Pekalongan, Pemalang, and Tegal. This region was chosen because of its unique climate characteristics, significant urbanization influence, and the impact of climate change on thermal comfort patterns.

Urbanization occurs because the northern coastal areas of Central Java tend to have faster industrial growth. The data used in this study were collected from several BMKG stations spread across the study area, with a time span of 2018 to 2024. Figure 2 shows a spatial representation of the research location, which is the basis for analyzing the distribution of the THI.

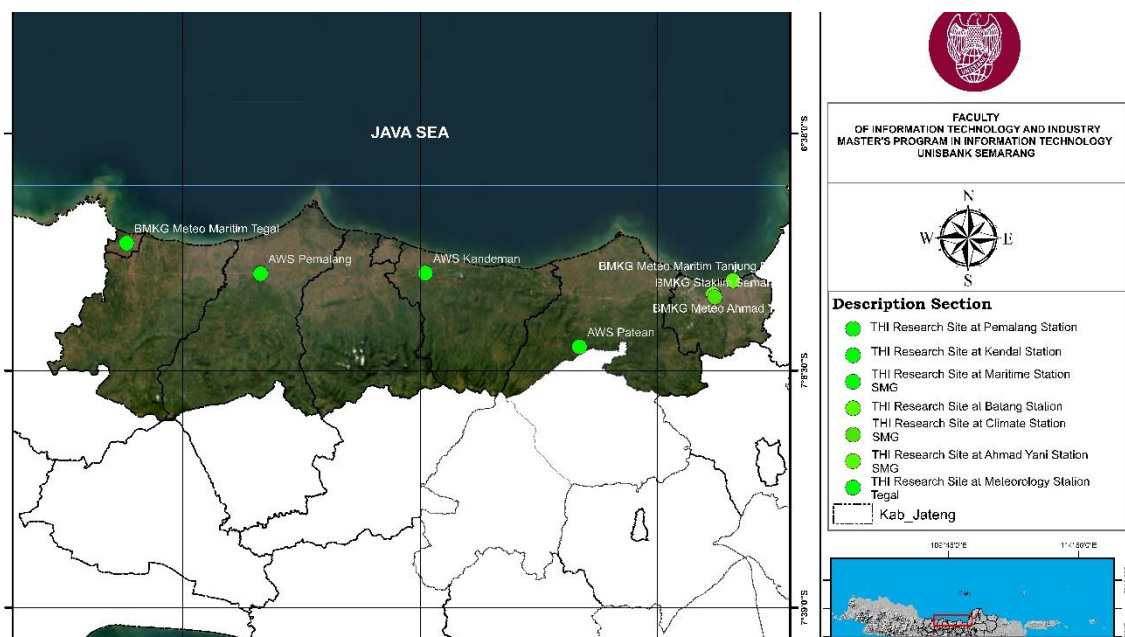


Figure 2. Map of Weather Observation Stations and THI Research Sites Northwestern Coast of Central Java

3.2 Analysis of Temperature and Humidity Trends for 2018-2024

In this study, the average temperature from seven research locations along the northern coast of Central Java was analyzed to identify temperature change trends over the period from 2018 to 2024. The visualization of average temperature, as presented in the graph, illustrates daily temperature fluctuations at various measurement points, which were then averaged to provide a comprehensive overview of temperature conditions in the region.

Figure 3, depicts the changes in average temperature at each research location throughout the 2018–2024 period. The graph highlights the patterns of daily temperature fluctuations observed at each site, which were subsequently compared to identify general trends in the northern coastal region of Central Java. The

analysis results indicate temperature variations influenced by seasonal factors, shifts in climate patterns, and environmental characteristics at each measurement location.

Descriptive statistics are utilized to examine the characteristics of data distribution, including central tendency (mean), variability (standard deviation), and data range (min-max values). This is seen in Table 1.

The descriptive statistical analysis of daily average temperature and annual temperature variations in the Northern Coastal Region of Central Java from 2018 to 2024 reveals a fluctuation pattern indicating temperature stability, with an average daily temperature of 28.00°C. The relatively small standard deviation (0.90°C) suggests minimal daily temperature variability, despite normal fluctuations. The recorded minimum temperature was 24.71°C, while the maximum reached 30.54°C. The first quartile (25%)

was at 27.40°C, and the third quartile (75%) at 28.63°C, indicating a non-extreme temperature range. Annually, the average temperature experienced a slight decline of 0.09°C in 2019 but showed an upward trend in

subsequent years, with the highest increase recorded in 2023 (0.41°C). In 2024, the temperature rose by 0.13°C, which, although relatively small, still reflects a long-term warming trend.

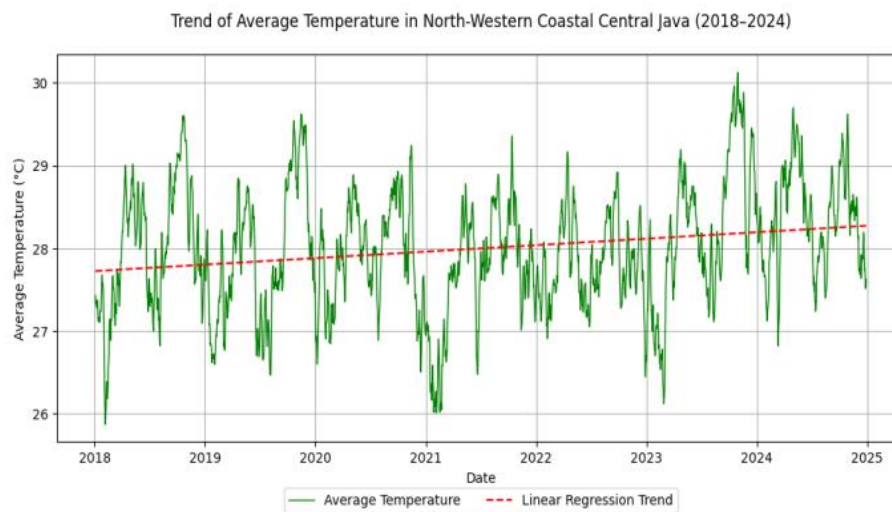


Figure 3. Trend of Average Temperature in the Northern Coast of Central Java (2018–2024)

Table 1. Descriptive Statistics and Changes in Annual Average Temperature

Descriptive Statistics	Mean Temp (°C)	Year	Mean Temp (°C)	Temp Change (°C)
mean	28.00	2018	28.00	0.00
std	0.90	2019	27.91	-0.09
min	24.71	2020	27.93	0.01
25%	27.40	2021	27.69	-0.23
50%	28.00	2022	27.83	0.14
75%	28.63	2023	28.24	0.41
max	30.54	2024	28.37	0.13

3.3 Analysis of Average Relative Humidity (RH) Trends for 2018–2024

This analysis evaluates the trends in average relative humidity in the Northern Coast of Central Java from 2018 to 2024. The graph illustrates daily humidity fluctuations with a linear regression line to identify the overall trend. Variations in humidity levels significantly impact thermal comfort and weather conditions.

Descriptive statistics and an annual humidity trend analysis are presented to better understand these patterns. Figure 4, depicts the changes in average humidity across the study locations over the 2018–2024 period. Descriptive statistics are employed to analyze the characteristics of data distribution, including central tendency (mean), variability (standard deviation), and data range (minimum to maximum values). This can be seen in Table 2.

Table 2. Descriptive Statistics and Annual Average Humidity Changes

Descriptive Statistics	Mean RH (%)	Year	Mean Humidity (%)	RH Change (%)
mean	78.30	2018	75.21	0.00
std	7.25	2019	75.68	0.47
min	51.29	2020	80.79	5.11
25%	72.71	2021	80.68	-0.12
50%	79.57	2022	80.73	0.05
75%	83.86	2023	76.82	-3.90
max	95.57	2024	78.17	1.34

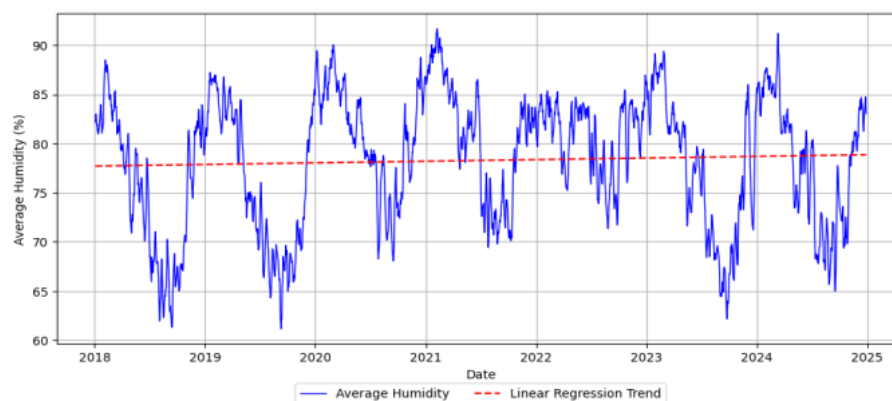


Figure 4. Trend of Average Humidity in the Northern Coast of Central Java (2018–2024)

Based on the analysis of descriptive statistical data, the average daily humidity in the northern coastal region of

Central Java from 2018 to 2024 was recorded at 78.30%, with a standard deviation of 7.25%. This

indicates significant daily humidity variations while remaining generally stable. The minimum humidity was recorded at 51.29%, while the maximum reached 95.57%, reflecting substantial fluctuations influenced by seasonal weather factors and extreme weather phenomena. Additionally, the first quartile (25%) was at 72.71%, and the third quartile (75%) was recorded at 83.86%, indicating that humidity mostly ranged between 72% and 84%. In 2019, a slight increase in humidity of 0.47% was observed, while 2020 experienced a significant surge of 5.11%, potentially caused by extreme weather events or distinct climatic patterns compared to previous years. However, in 2021, humidity decreased by 0.12%, indicating smaller fluctuations. A significant decline of 3.90% occurred in 2023, which may have been influenced by weather phenomena such as prolonged droughts or El Niño events, leading to reduced humidity in the region. In 2024, humidity increased again by 1.34%, suggesting a relatively small rise that could be a response to climate change or broader global weather phenomena. Overall, these humidity fluctuations have a significant impact on thermal comfort.

3.4 Calculation of Temperature Humidity Index (THI) for 2018–2024

In this step, data processing was conducted to calculate the Temperature Humidity Index (THI) for the period 2018–2024 based on average temperature (T_{avg}) and relative humidity (RH_{avg}) data. The dataset consists of 17,897 rows and six columns, including station names, geographic coordinates, observation dates, average temperature, average humidity, and calculated THI values.

In general, the THI calculations indicate that an increase in humidity, even with the same or lower temperatures, can reduce thermal comfort. Therefore, THI serves as a valuable indicator for understanding changes in thermal conditions in coastal areas, which significantly impact public health and comfort. The analysis also reveals substantial THI variations throughout the observation period, highlighting seasonal fluctuations that influence thermal comfort. Table 3 presents the THI calculations based on average temperature and humidity data from various observation stations during the 2018–2024 period. The THI was computed using Equation 2[22]:

$$THI = 0.8 * T + \frac{RH * T}{500} \quad (2)$$

Table 3. THI Calculation Results Based on Temperature and Humidity (2018–2024)

Station_Name	latitude	longitude	Date	Mean_Temperature(°C)	Mean_Humidity(%)	THI
BMKG_Tegal	-6.86817	109.121	01/01/2018	28.9	73	27.3
BMKG_Klimatologi Jateng	-6.984	110.381	01/01/2018	27.5	81	26.5
BMKG_Maritim Semarang	-6.9486	110.4199	01/01/2018	28.0	78	26.8
BMKG_Ahmad Yani	-6.97683	110.3778	01/01/2018	27.5	78	26.3
AWS Pemalang	-6.93359	109.4080	01/01/2018	27.8	80	26.7
AWS_Batang	-6.93231	109.7611	01/01/2018	29.0	75	27.6
AWS Patean Kendal	-7.0905	110.091	01/01/2018	27.8	78	26.6
.....						
BMKG_Tegal	-6.86817	109.121	31/12/2024	27.1	82	26.1
BMKG_Klimatologi Jateng	-6.984	110.381	31/12/2024	27.4	82	26.4
AWS Patean Kendal	-7.0905	110.091	31/12/2024	26.8	78	25.6

3.5 Analysis of Temperature and THI Trends (2018–2024)

The THI is a widely utilized indicator for assessing thermal comfort levels based on air temperature and relative humidity. This approach aims to generate more precise THI projections, providing insights into potential changes in thermal comfort along the northern coast of Central Java and supporting evidence-based decision-making for mitigating heat-related impacts in the region. Figure 5 illustrates temperature and THI trends over the 2018–2024 period. Annual variations are analyzed to evaluate potential warming trends or humidity changes that contribute to THI variability. The results of this analysis are presented in Table 4 and serve to provide a clearer understanding of the dynamics of thermal conditions over time. During the period from 2018 to 2024, fluctuations were observed in the Temperature Humidity Index (THI) and average

temperature values, as presented in Figure 4. Overall, the trend indicates an increase in both THI and temperature, although certain years exhibited a decline.

The THI increased from 26.13 in 2018 to 27.38 in 2024, with annual variations ranging from -0.74 to 0.83. A significant decrease occurred in 2021, with a change of -0.74, while the highest increase was recorded in 2024 at 0.83. Similarly, the average temperature followed the same pattern, rising from 27.02°C in 2018 to 28.44°C in 2024. The highest increase was observed in 2024, at 1.01°C, whereas the most substantial decline occurred in 2021, at -0.99°C. The graph illustrates a general upward trend in temperature and THI over time. However, the considerable annual fluctuations suggest that external factors, such as regional weather patterns and climate variability, contribute to these changes.

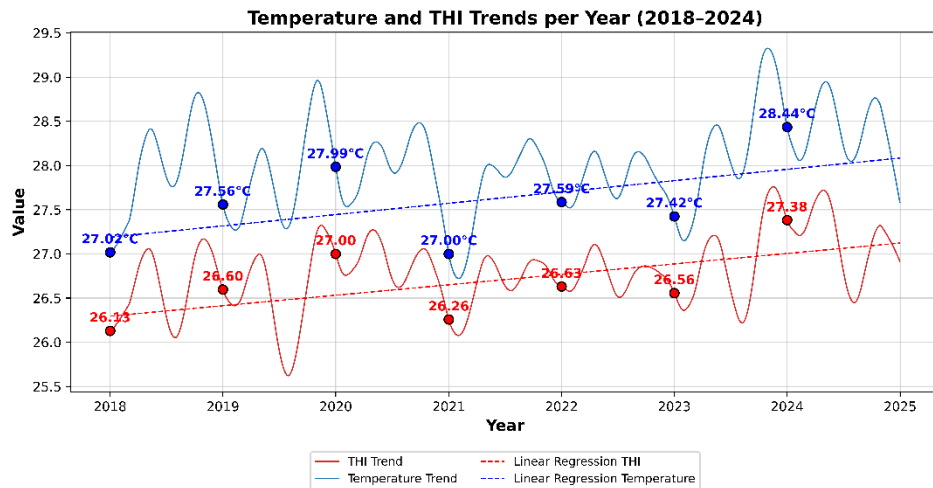


Figure 5. Analysis of Temperature and THI Trends (2018-2024)

Table 4. Average THI and Temperature and Their Annual Changes

Year	THI	THI Change	Temperature (°C)	Temp. Change (°C)
2018	26.13	0.00	27.02	0.00
2019	26.60	0.47	27.56	0.54
2020	27.00	0.40	27.99	0.43
2021	26.26	-0.74	27.00	-0.99
2022	26.63	0.37	27.59	0.59
2023	26.56	-0.07	27.42	-0.16
2024	27.38	0.83	28.44	1.01

3.6 Seasonal Average of THI (2018-2024)

The seasonal average of the THI for the period between 2018 and 2024 is calculated in this section. The objective of this calculation is to understand the fluctuations in thermal comfort based on variations in temperature and humidity across different seasons and to assess their impact on the quality of life of

communities in the northern coastal region of Central Java. This analysis provides a clearer understanding of how seasonal changes influence thermal comfort in coastal areas. This change can be seen in Figure 6.

The analysis of the Temperature-Humidity Index (THI) from 2018 to 2024 indicates significant seasonal fluctuations, reflecting the dynamics of temperature and humidity influenced by both global and local climatic factors. During the rainy season, THI generally decreased from the beginning to the end of the season, except in 2023, which exhibited an increase of +1.65, suggesting warmer and more humid conditions compared to previous years. Conversely, in 2024, the most substantial THI decline of -2.55 was observed, likely due to increased rainfall or shifts in weather patterns that resulted in lower temperatures at the end of the season.

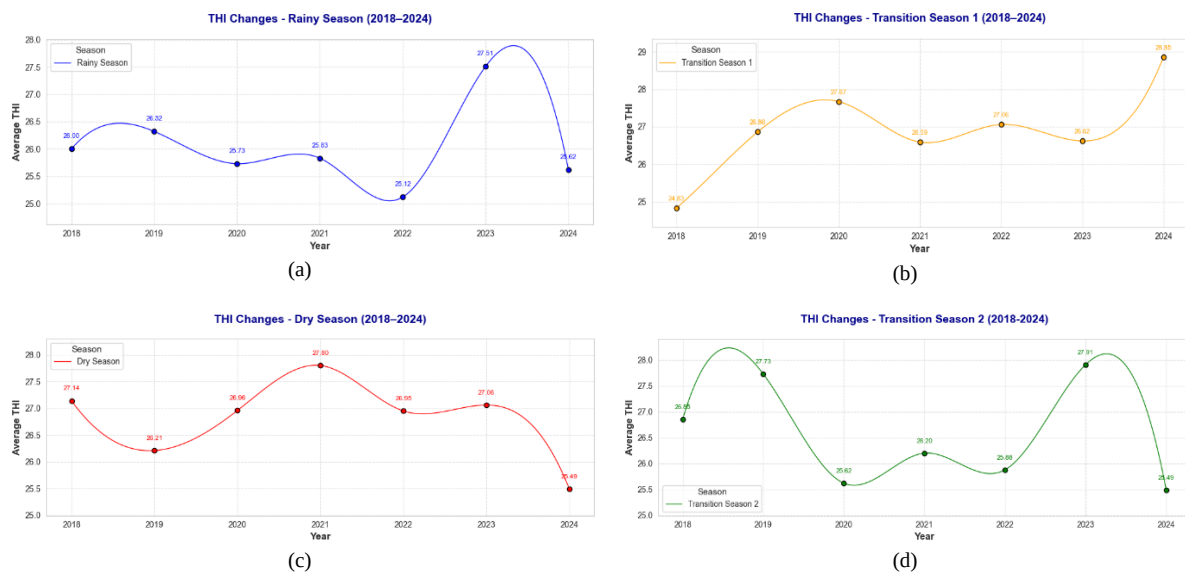


Figure 6. Seasonal Average of THI (2018-2024): (a) Rainy Season, (b) Dry Season, (c) Transition Season 1, (d) Transition Season 2

During the first transition season, a significant increasing trend in the Temperature-Humidity Index (THI) has been observed in recent years, with the

highest value recorded in 2024 (28.85). This increase indicates progressively warmer and more humid environmental conditions during the first transition

season. Meanwhile, in most years, THI exhibited a decline from the beginning to the end of the season, as observed in 2018 (-2.48) and 2021 (-1.27). However, certain years, such as 2023 (+1.14) and 2024 (+0.76), showed an increase in THI, suggesting the influence of external factors contributing to higher temperature and humidity levels towards the end of the season.

During the dry season, the Temperature-Humidity Index (THI) exhibited a varied trend. In certain years, such as 2018 (+1.05), 2019 (+0.17), and 2023 (+0.69), an increase in THI was observed, indicating hotter and more humid dry seasons. Conversely, a significant decline in THI occurred in 2024 (-2.57). The El Niño phenomenon in that year likely contributed to this decrease, as higher temperatures were accompanied by lower humidity levels, resulting in a reduced perception of heat despite elevated air temperatures.

Transitional Season 2 also exhibits significant annual fluctuations. In 2024, a substantial decrease in the THI (-2.36) was observed, contrasting with the increase recorded in 2023 (+1.14). This decline may be associated with El Niño, which resulted in higher temperatures but lower humidity, leading to a significant reduction in the THI.

Overall, the findings indicate that THI is influenced not only by air temperature but also by humidity, precipitation patterns, wind systems, and global climate phenomena such as El Niño. The seasonal fluctuations emphasize the importance of continuous monitoring of temperature and humidity to better understand the impact of changing climate patterns on thermal comfort and environmental equilibrium.

3.7 THI Mapping in the Western Part of the Northern Coast of Central Java (2018–2024)

This map aims to illustrate the distribution of thermal comfort based on temperature and humidity in the coastal region, highlighting the variations in the Temperature-Humidity Index (THI) throughout the year. By visualizing THI data in a map format, this study seeks to provide a clearer understanding of how thermal conditions vary across different locations and their potential impact on coastal communities, particularly in response to shifting climate patterns and extreme weather events.

The analysis results indicate that the Temperature-Humidity Index (THI) in the northwestern coastal region of Central Java during the 2018–2024 period ranged from 22.36 to 30.96, with an average of 26.77. The spatial distribution reveals variations in THI values across different areas, where lower THI values (22.3–26.1) were observed in Batang, Kendal, Pekalongan, and Pemalang Regencies, while higher THI values

(27.2–29.2) were predominantly found in parts of Tegal Regency. Semarang City exhibited moderate THI values, ranging from 26.5 to 27.5. The THI map employs a color scale from blue to red, where blue represents lower THI values, indicating better thermal comfort levels, while orange to red signifies higher THI values, reflecting hotter thermal conditions. Batang, Kendal, Pekalongan, and Pemalang appear in shades of blue, suggesting better thermal comfort, whereas certain areas in Tegal and Semarang are depicted in bright orange, indicating warmer thermal conditions.

During the rainy season (November–April) [27], the north coast of Central Java exhibited significant spatial variation in THI values for the 2018–2024 period. Based on the visualization map, THI values during this period ranged from 24.3 to 28.6, with an average falling between 26.0 and 27.0. Lower THI values (approximately 24.3–25.5) were observed in Kendal Regency and the border of Batang Regency, represented by blue color gradients. Conversely, higher THI values (reaching 27.7–28.6) were identified in Tegal, parts of Semarang City, and the Kendal-Semarang border, depicted in red color gradients. The central region of Batang to the midsection of Pekalongan during the 2018–2024 rainy seasons appeared in bright orange, indicating slightly higher THI values ranging from 26.7 to 27.2. The increase in THI values in this area suggests that despite temperatures remaining within a moderate range, relatively high humidity contributed to slightly uncomfortable thermal conditions. The Pemalang region, extending to the northeastern border of Tegal, which appears in light blue-gray shades, exhibited lower THI values, ranging from 26.0 to 26.5. This region falls within the category of relatively stable and more comfortable thermal conditions compared to other areas. The THI mapping visualization for the western part of the northern coast of Central Java (2018–2024) is presented in Figure 7.

Conversely, during the dry season (June–September) [28], THI values were higher compared to the rainy season, ranging from 23.8 to 28.3, with an average of 25.5–26.5. Areas with lower THI values (23.8–25.0), such as Batang, Kendal, and Pekalongan, tend to have more green spaces that contribute to reducing temperatures. In contrast, regions including Semarang, southeastern Kendal, Tegal, Pemalang, and the western border of Pekalongan exhibited higher THI values (27.4–28.3), influenced by rising air temperatures due to urbanization, low precipitation, and geographical conditions that are susceptible to heat accumulation. Overall, the dry season leads to an increase in temperature and a decrease in humidity, which subsequently exacerbates thermal discomfort along the northern coast of Central Java.

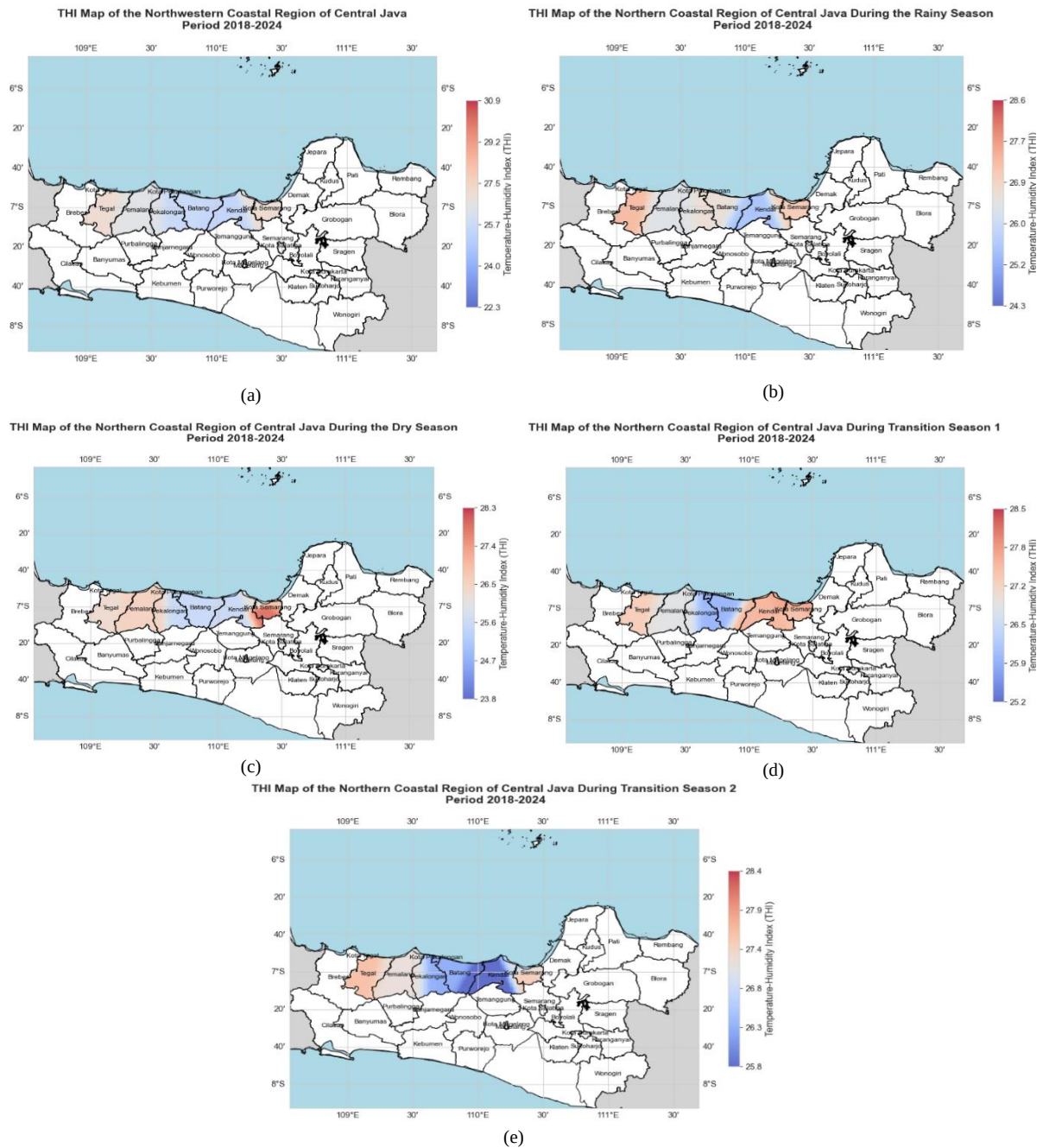


Figure 7. THI Mapping in the Western Part of the Northern Coast of Central Java (2018–2024): (a) Average THI Map, (b) THI Map for the Rainy Season, (c) THI Map for the Dry Season, (d) THI Map for Transition Season 1, (e) THI Map for Transition Season 2

The first transition season (May) [29], in the northern coastal region of Central Java during the 2018–2024 period exhibits significant spatial variation, with THI values ranging from 25.2 to 28.5 and an average between 26.5 and 27.5. Lower THI values (25.2–27.0) are observed in Pekalongan Regency, Pemalang, and parts of Batang, likely influenced by higher humidity levels and stronger sea breezes. Conversely, higher THI values (27.5–28.5) are concentrated in Tegal, Kendal, Semarang City, and eastern Batang, which are presumably affected by increased air temperature and lower humidity. Semarang City tends to have higher THI values compared to surrounding areas, which may be attributed to human activities, the use of heat-absorbing materials, and limited vegetation cover.

Overall, the first transition season exhibits higher thermal conditions compared to the rainy season, reflecting the shift toward the dry season with increased solar radiation.

During the second transition season (October) [30], the northern coast of Central Java exhibits THI variations ranging from 25.8 to 28.4, reflecting the transition from the dry season to the rainy season. Regions with lower THI values (25.8–26.5), such as Batang, Kendal, and Pekalongan, experience higher humidity, which helps mitigate heat effects. In contrast, Tegal and Semarang City exhibit higher THI values (27.5–28.4) due to elevated air temperatures and the influence of urbanization. The western coastal area of Semarang

shows a lower THI (26.5–26.8), indicating relatively more comfortable thermal conditions compared to surrounding areas. Pemalang and the Pekalongan-Pemalang border (26.5–27.0) exhibit moderate conditions with a potential for thermal discomfort. Overall, the second transition season presents a more complex THI distribution pattern compared to both the dry and rainy seasons. The analysis indicates that no region experiences extreme thermal discomfort or heat stress.

3.8 Implementation of the Random Forest (RF) Algorithm for Thermal Comfort Classification Based on THI

The Random Forest (RF) method is employed in this study to classify thermal comfort levels based on the Temperature Humidity Index (THI). The RF model has been demonstrated to effectively handle complex non-linear datasets, making it superior in predicting device output characteristics with minimal error [31]. By utilizing a historical data-based analysis of the Temperature Humidity Index (THI), this study provides insights into thermal comfort trends, identifies patterns of change, and offers recommendations based on the classification results. Figure 8 presents a visualization of the RF classification results for thermal comfort levels based on THI.

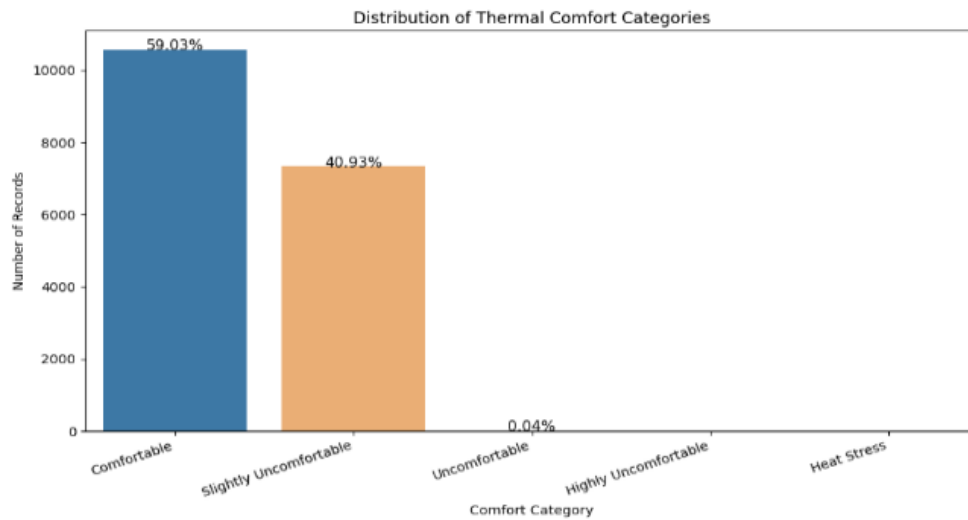


Figure 8. Random Forest (RF) Classification of THI for Thermal Comfort Analysis

Performance evaluation of the model was conducted using a confusion matrix and a classification report, which provide insights into the accuracy, precision, recall, and F1-score for each thermal comfort category. The model's accuracy results are presented in Table 5, illustrating the extent to which the model can accurately classify thermal comfort levels. In the THI classification (Temperature Humidity Index), the Gini Index is employed to measure impurity or the clarity of separation between classes in the decision tree, where p_i represents the proportion of data in class i , and c denotes the number of classes. A lower Gini Index value indicates a purer separation between classes, meaning the decision tree can more effectively distinguish data. The Gini Index is calculated in Equation 3[32] :

$$g = 1 - \sum_{i=1}^c p_i^2 \quad (3)$$

In prediction using the Random Forest (RF) model, the ensemble model operates such that $\hat{Y}$ represents the final prediction of the RF model, N denotes the number of decision trees, and $f_i(x)$ refers to the prediction generated by the i -th tree for input x . In other words, Random Forest aggregates the predictions from individual trees and computes an averaged prediction to minimize overfitting and enhance accuracy. The mathematical formulation is given in Equation 4:

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (4)$$

Table 5. Classification Report for Temperature-Humidity Index (THI) Using (RF)

Classification Report	Precision	Recall	F1-Score	Support
Discomfort	1.00	0.50	0.67	2
Comfortable	1.00	1.00	1.00	2059
Slightly Uncomfortable	1.00	1.00	1.00	1519
Accuracy			1.00	3580
Macro avg	1.00	0.83	0.89	3580
Weighted avg	1.00	1.00	1.00	3580
MSE Regresi				
0.000441				

Based on the classification results, the distribution of thermal comfort categories indicates that the majority of conditions fall within the "Comfortable" category, accounting for 59.03%, followed by the "Slightly Uncomfortable" category at 40.93%, and the "Uncomfortable" category, which comprises only 0.04% of the total data. No data were found in the "Highly Uncomfortable" or "Heat Stress" categories, suggesting that most observed weather conditions remain within thermal comfort limits. This distribution indicates that although a portion of the data falls into the "Slightly Uncomfortable" category, overall weather conditions are still considered comfortable for humans.

The classification results obtained using the Random Forest (RF) algorithm demonstrate a notably high accuracy of 99.94%. According to the classification report, the precision, recall, and F1-score values for the "Comfortable" and "Slightly Uncomfortable" categories reach 1.00, indicating that the model effectively distinguishes between these two categories. However, for the "Uncomfortable" category, the recall value is only 0.50, which may be due to the limited number of data points in this category, preventing the model from sufficiently learning the distinguishing patterns. In addition to classification, the regression model was evaluated using the Mean Squared Error (MSE), which yielded a value of 0.00044.

3.9 Confusion Matrix for Evaluating the Performance of Random Forest (RF) in Thermal Comfort Classification Based on THI

The confusion matrix as in Figure 9 is used to evaluate the performance of the Random Forest (RF) model in classifying thermal comfort levels based on the Temperature Humidity Index (THI).

As a result of the analysis, the model demonstrated exceptionally high accuracy in classifying the "Comfortable" and "Slightly Uncomfortable" categories, with 10,565 and 7,325 correctly predicted samples, respectively. However, the "Uncomfortable" category had a very limited number of samples, with only 7 instances.

3.10 Evaluation of Feature Importance in Determining THI

Feature importance in the Random Forest (RF) model is utilized to identify the most influential variables in classifying thermal comfort levels based on the THI as in Figure 10. Feature importance analysis indicates that air temperature is the primary factor influencing thermal comfort, with various studies demonstrating that temperature variations significantly impact the Temperature Humidity Index (THI) [33]. The average temperature contributes 79.4% to the classification process of thermal comfort levels, whereas the average humidity accounts for only 20.6%.



Figure 9. Confusion Matrix for Random Forest (RF) Performance in THI Thermal Comfort Classification

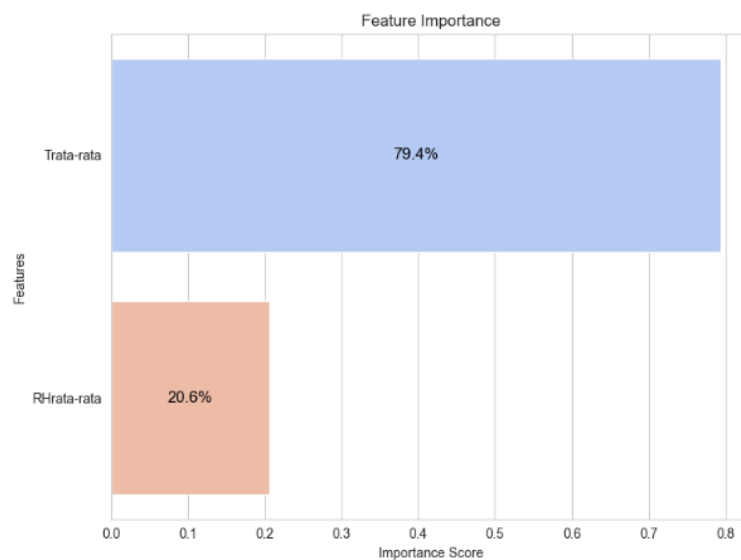


Figure 10. Evaluation of Feature Importance in Determining THI

3.11 Thermal Comfort Classification Map of THI - RF for the Northern Coast of Central Java (2018-2024)

In this study, THI classification is based on the following value ranges: Comfortable ($THI < 27$), Slightly Uncomfortable ($27 \leq THI < 30$), Discomfort

($30 \leq THI < 40$), High Discomfort ($40 \leq THI < 45$), and Heat Stress ($THI \geq 45$).

The results of the THI analysis indicate that the majority of the northern coastal regions of Central Java fall within the slightly uncomfortable category ($THI 27-30$), particularly in Semarang, the eastern part of Kendal, Pemalang, and Tegal, as seen in Figure 11.

The primary factors contributing to this condition include urbanization, industrial activities, and the urban heat island (UHI) effect. In contrast, Batang and the western part of Kendal exhibit more comfortable conditions (THI < 27), likely attributed to the extensive

green areas and the cooling effect of sea breezes. The Pekalongan region represents a transition zone between comfortable and slightly uncomfortable conditions, where coastal areas tend to be cooler than inland areas.



Figure 11. Thermal Comfort Classification Map of THI – RF (2018-2024)

3.12 THI Projection Using XGBoost

By employing machine learning-based predictive models, particularly XGBoost, this study forecasts the development of THI values in the northern coastal region of Central Java for the period 2025–2029 to identify future trends in thermal comfort variations. The

analysis of THI predictions and projections for the 2025–2029 period using the XGBoost model indicates a relatively stable upward trend with minor fluctuations as Figure 12. The historical trend and projection graphs demonstrate a general increase in THI over time, supported by a regression line that indicates an upward trajectory.

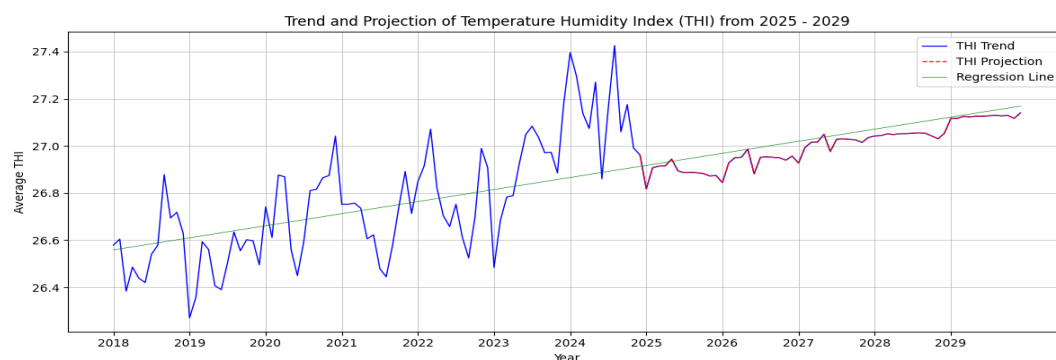


Figure 12. Projected Temperature-Humidity Index (THI) Using XGBoost for the Period 2025-2029

The predictive model exhibits strong performance, with a Mean Absolute Error (MAE) of 0.32, Mean Squared Error (MSE) of 0.18, and Root Mean Squared Error (RMSE) of 0.42. Furthermore, an R-squared (R^2) value of 0.73 suggests that the model explains approximately 73% of the data variability, indicating a high level of accuracy in predicting THI changes. The Mean Absolute Percentage Error (MAPE) of only 1.21% further confirms that the prediction error is relatively low, enhancing confidence in the model's reliability for

future THI trend estimations. Annual trend analysis reveals that THI in 2025 starts at 26.81, experiencing fluctuations throughout the year. In 2026, a gradual increase is observed, accompanied by minor temporary declines. A significant upward trend emerges from 2027 to 2029, with THI reaching its highest value of approximately 27.14 by the end of 2029. This trend suggests increasing temperature and humidity levels, which may impact thermal comfort in the study area.

The results of the validation model show that the 5-year prediction of most of the heavily populated Pantura areas tends to be less comfortable. This is the basis for policy makers to take steps in controlling THI.

4. Conclusions

In conclusion, this study projects thermal comfort levels in the northern coastal region of Central Java for the 2025–2029 period using the Temperature-Humidity Index (THI) approach. The XGBoost model was employed to predict THI values, and the Random Forest (RF) algorithm was utilized to classify comfort levels. The results highlight significant variations in thermal conditions across regions and seasons, influenced by geographical factors, urbanization, and local microclimates. In general, regions such as Batang and Pekalongan tend to exhibit more comfortable thermal conditions ($THI < 27$) throughout the year, while areas like Semarang, eastern Kendal, Pemalang, and Tegal often experience slightly uncomfortable conditions ($THI 27\text{--}30$), especially during the rainy and transitional seasons. This is likely due to the urban heat island effect and higher humidity levels.

During the dry season, most areas show improved thermal conditions with lower THI values compared to other seasons, but certain urban areas, such as Semarang and eastern Kendal, continue to exhibit higher THI values than surrounding regions. This highlights the urgent need for heat mitigation strategies, particularly in urban areas, to reduce the impact of rising temperatures and humidity on public health and productivity. The model evaluation results show that the Random Forest method achieves a high classification accuracy (99.94%) with strong precision and recall for the Comfortable and Slightly Uncomfortable categories. However, limitations exist in identifying the Uncomfortable category due to imbalanced data distribution. The XGBoost model demonstrated relatively good predictive performance, with an R^2 of 0.73, indicating that it explains approximately 73% of the variability in THI data.

For future strategies, this study suggests that heat mitigation approaches are essential for improving the quality of life, particularly for public health, outdoor workers, and urban planning in coastal areas. Urban areas, where the urban heat island effect is prominent, must incorporate green spaces and enhance cooling strategies to manage rising thermal discomfort. Additionally, better urban planning and coastal area management, considering the implications on workers and local communities, are crucial for reducing thermal risks and improving overall living conditions. Future research should integrate additional factors such as wind speed, land cover, and climate change scenarios to further enhance predictive accuracy and provide more comprehensive recommendations for urban thermal environment management.

References

- [1] H. Rahmadyani and M. Fahri, "Analisis Performa Kenyamanan Termal Pada Ruang Terbuka (Studi Kasus: Alun-Alun Taman Merdeka Pangkalpinang)," 2024. [Online]. Available: <http://journals.ums.ac.id/index.php/sinektika>
- [2] D. Maheng, B. Bhattacharya, C. Zevenbergen, and A. Pathirana, "Changing Urban Temperature and Rainfall Patterns in Jakarta: A Comprehensive Historical Analysis," *Sustainability (Switzerland)*, vol. 16, no. 1, Jan. 2024, doi: 10.3390/su16010350.
- [3] D. Maheng et al., "Impact of land use land cover changes on urban temperature in Jakarta: insights from an urban boundary layer climate model," *Front Environ Sci*, vol. 12, 2024, doi: 10.3389/fenvs.2024.1399041.
- [4] R. Aghamolaei, M. M. Azizi, B. Aminzadeh, and J. O'Donnell, "A comprehensive review of outdoor thermal comfort in urban areas: Effective parameters and approaches," Sep. 01, 2023, SAGE Publications Inc. doi: 10.1177/0958305X221116176.
- [5] S. Prasetyo, U. Hidayat, Y. D. Haryanto, and N. F. Riama, "Variasi dan Trend Suhu Udara Permukaan di Pulau Jawa Tahun 1990-2019," *Jurnal Geografi: Media Informasi Pengembangan dan Profesi Kegeografian*, vol. 18, no. 1, pp. 60–68, Jun. 2021, doi: 10.15294/jg.v18i1.27622.
- [6] H. N. Saputri, B. D. Dasanto, and R. Hidayati, "Tingkat Kenyamanan Termal Lingkungan Kampus IPB Dramaga Menggunakan Pendekatan Physiological Equivalent Temperature (PET)," *Jurnal Ilmu Lingkungan*, vol. 21, no. 2, pp. 299–307, 2023, doi: 10.14710/jil.21.2.299-307.
- [7] B. T. Putri and Q. A. Besila, "Implementasi Kenyamanan Termal: Hasil Pemetaan Area Dan Rekomendasi Desain Lanskap Rekreasi Pantai Pulau Kodok Tegal," *JURNAL BHUWANA*, pp. 1–9, Jul. 2023, doi: 10.25105/bhuwana.v3i1.17526.
- [8] H. Talebi, L. J. M. Peeters, A. Otto, and R. Tolosana-Delgado, "A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling," *Math Geosci*, vol. 54, no. 1, Jan. 2022, doi: 10.1007/s11004-021-09946-w.
- [9] D. Wu et al., "Downscaling the Midsummer Temperature-Humidity Index Based on Multiple Machine Learning Methods," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 16, pp. 7124–7134, 2023, doi: 10.1109/JSTARS.2023.3299459.
- [10] N. Effendy, M. Z. A. Fadhilah, D. W. Kraton, and H. A. Abrar, "The prediction of thermal sensation in building using support vector machine and extreme gradient boosting," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 2963–2970, Sep. 2024, doi: 10.11591/ijai.v13i3.pp2963-2970.
- [11] A. Dhiya, ulhaq Nasihien, A. Wahid Hasyim, and F. Usman, "The Spirit of Society Journal Thermal Comfort Level Based on the Temperature Humidity Index Value in Bandung City," Online, 2024. [Online]. Available: <https://jurnal.narotama.ac.id/index.php/scj/index>
- [12] T. Xi, M. Wang, E. Cao, J. Li, Y. Wang, and S. U. Sa'ad, "Preliminary Research on Outdoor Thermal Comfort Evaluation in Severe Cold Regions by Machine Learning," *Buildings*, vol. 14, no. 1, Jan. 2024, doi: 10.3390/buildings14010284.
- [13] A. Tuppad and S. D. Patil, "Journal of Advanced Zoology Data Pre-processing Issues in Medical Data Classification", [Online]. Available: <https://jazindia.com>
- [14] N. Hafidz and D. Yanti Liliana, "Klasifikasi Sentimen pada Twitter Terhadap WHO Terkait Covid-19 Menggunakan SVM, N-Gram, PSO," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 213–219, Apr. 2021, doi: 10.29207/resti.v5i2.2960.
- [15] D. Varma, A. Nehansh, and P. Swathy, "Data Preprocessing Toolkit: An Approach to Automate Data Preprocessing," *Interantional Journal Of Scientific Research In Engineering And Management*, vol. 07, no. 03, Mar. 2023, doi: 10.55041/ijrsrem18270.
- [16] M. A. N D, M. S. M, and Ms. S. Y. G TILAK, "Exploratory Data Analysis (EDA) and data visualization," *IJIREEEICE*, vol. 12, no. 6, May 2024, doi: 10.17148/ijireeice.2024.12608.

- [17] P. R. Sihombing, S. Suryadiningrat, D. A. Sunarjo, and Y. P. A. C. Yuda, "Identifikasi Data Outlier (Pencilan) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya," *Jurnal Ekonomi Dan Statistik Indonesia*, vol. 2, no. 3, pp. 307–316, Jan. 2023, doi: 10.11594/jesi.02.03.07.
- [18] W. McKinney, "pandas: a Foundational Python Library for Data Analysis and Statistics." [Online]. Available: <https://www.researchgate.net/publication/265194455>
- [19] N. Kovač, K. Ratković, H. Farahani, and P. Watson, "A practical applications guide to machine learning regression models in psychology with Python," *Methods in Psychology*, vol. 11, Dec. 2024, doi: 10.1016/j.metip.2024.100156.
- [20] P. Sinthong, D. Patel, N. Zhou, S. Shrivastava, A. Iyengar, and A. Bhamidipaty, "DQDF: Data-Quality-Aware Dataframes," in *Proceedings of the VLDB Endowment*, VLDB Endowment, 2021, pp. 949–957. doi: 10.14778/3503585.3503602.
- [21] A. Taufiq and C. Wulandari, "Comfort Level Of Green Open Space In Bandar Lampung Based On Climate And Humidity," *Jurnal Belantara*, vol. 5, no. 1, pp. 01–13, Mar. 2022, doi: 10.29303/jbl.v5i1.847.
- [22] M. M. Fadhlurrahman and N. Nasrullah, "Study of Thermal Comfort under the Shade of Varied Tree Canopy Form and Distance from the Stem," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Jun. 2020. doi: 10.1088/1755-1315/501/1/012048.
- [23] F. Masitoh and A. N. Rusydi, "Climatological Human Comfort Using Heat and Humidity Index (Humidex) in Gadingkulon, Malang," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics Publishing, Jan. 2020. doi: 10.1088/1755-1315/412/1/012026.
- [24] K. Cai and M. R. Rodavia, "XGBoost Analysis based on Consumer Behavior," *Frontiers in Computing and Intelligent Systems*, vol. 5, no. 2, pp. 85–89, 2023, doi: 10.54097/fcis.v5i2.12974
- [25] N. B. Assymkhan and A. Z. Kartbaev, "thermal comfort prediction using SVM and random forest model," *Вестник КазУТБ*, vol. 4, no. 25, Dec. 2024, doi: 10.58805/kazutb.v.4.25-400.
- [26] Y. Boutahri and A. Tilioua, "Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings," *Results in Engineering*, vol. 22, Jun. 2024, doi: 10.1016/j.rineng.2024.102148.
- [27] A. Z. A. Zaini, M. Vonnisa, M. Marzuki, and R. Ramadhan, "Seasonal Variation of Rainfall in Indonesia under Normal Conditions without ENSO and IOD Events from 1981-2021," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 11, pp. 9899–9909, Nov. 2023, doi: 10.29303/jppipa.v9i11.4569.
- [28] H. Akhsan, M. Irfan, and I. Iskandar, "El Niño Southern Oscillation (ENSO), Indian Ocean Dipole (IOD), and the Rise of Extreme Temperatures in Eastern Sumatra: Exploring Climate Change Dynamics," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 2, pp. 600–608, Feb. 2023, doi: 10.29303/jppipa.v9i2.3084.
- [29] A. N. R. Irawan and D. Komori, "Beyond Fixed Dates and Coarse Resolution: Developing a Dynamic Dry Season Crop Calendar for Paddy in Indonesia from 2001 to 2021," *Agronomy*, vol. 14, no. 3, Mar. 2024, doi: 10.3390/agronomy14030564.
- [30] R. Hayati, M. Irman, and L. N. Agia, "Sell in May and Go Away or Just Another January Effect? Studied of Anomaly in Indonesia Stock Exchange."
- [31] A. Salah and D. Yevick, "A Random Forest Model for Predicting and Analyzing the Performance of CNT TFET with Highly Doped Pockets," *Adv Theory Simul*, Jan. 2024, doi: 10.1002/adts.202400607.
- [32] B. Lala et al., "The Challenge of Multiple Thermal Comfort Prediction Models: Is TSV Enough?," *Buildings*, vol. 13, no. 4, Apr. 2023, doi: 10.3390/buildings13040890.
- [33] Hanny Nirwani, R. Hidayati, and Perdinan, "Thermal Comfort for various Altitudes and Land Covers in North Sumatra," *Agromet*, vol. 37, no. 2, pp. 91–98, Oct. 2023, doi: 10.29244/j.agromet.37.2.91-98.