

Machine Learning Approaches for Classifying Indian Ocean Dipole (IOD) Using Random Forest and Decision Tree Models with SST, MSLP, And Total Precipitation Data from the Waters Off West Sumatra

Muhammad Arya Bintang Pratama ¹

¹Undergraduate Program in Applied of Instrumentation Meteorology, Climatology Geophysics (STMKG)

Article Info

Article history:

Received March 5, 2025

Revised March 10, 2025

Accepted March 11, 2025

Keywords:

Indian Ocean Dipole

Machine Learning

Random Forest

Decision Tree

Sea Surface Temperature

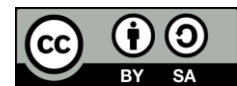
Mean Sea Level Pressure

Precipitation Data.

ABSTRACT

This research investigates the utilization of machine learning methodologies, particularly Random Forest and Decision Tree algorithms, to categorize Indian Ocean Dipole (IOD) occurrences by employing Sea Surface Temperature (SST), Mean Sea Level Pressure (MSLP), and total precipitation datasets derived from the maritime region adjacent to West Sumatra. The study leverages data amassed from 2020 to 2024, concentrating on diverse climatic scenarios linked to IOD. The efficacy of both algorithms is assessed using evaluative criteria such as accuracy, precision, and recall. The findings reveal that the Random Forest algorithm surpasses the Decision Tree algorithm, attaining an accuracy rate exceeding 85%, with SST recognized as the predominant predictor. These results underscore the promise of machine learning techniques in advancing the comprehension of IOD and its ramifications on regional meteorological trends, thereby facilitating enhanced climate forecasting models and guiding decision-making frameworks for climate adaptation.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponden Author:

Muhammad Arya Bintang Pratama,

Undergraduate Program in Applied of Instrumentation Meteorology, Climatology Geophysics (STMKG)

Tangerang City, Banten, Indonesia

Email: starr04@gmail.com

1. INTRODUCTION

The Indian Ocean Dipole (IOD) constitutes a pivotal climatic phenomenon, distinguished by the anomalous fluctuations in sea surface temperatures within the Indian Ocean. This phenomenon significantly affects meteorological patterns and climate variability in adjacent areas, notably Southeast Asia and Australia [1]. Grasping the intricate dynamics of the IOD is imperative for anticipating its ramifications on regional meteorology, agriculture, and hydrological resources [2].

Recent investigations have underscored the escalating occurrence and severity of extreme meteorological events linked to the IOD, which may result in dire outcomes such as droughts and inundations [3]. As climate change progressively modifies global meteorological patterns, the demand for precise predictive models becomes increasingly urgent [4]. Machine learning methodologies have surfaced as potent instruments for scrutinizing intricate datasets and enhancing predictive precision across diverse domains, including climate science [5].

This research aims to apply machine learning approaches, specifically Random Forest and Decision Tree models, to classify IOD phenomena using sea surface temperature (SST), mean sea level pressure (MSLP), and total precipitation data from the waters off West Sumatra. By leveraging these advanced techniques, the study seeks to enhance the understanding of the factors influencing IOD and improve the accuracy of climate predictions.

The objectives of this study are as follows:

- To analyze the performance of Random Forest and Decision Tree models in classifying IOD conditions.
- To identify the most significant predictors influencing IOD classification.
- To contribute to the development of more effective climate prediction models that can inform decision-making processes for climate resilience in the region.

The findings of this research are expected to provide valuable insights into the behavior of IOD and its implications for regional climate patterns, ultimately supporting efforts to mitigate the impacts of climate change.

2. THEORY

2.1 Data Location



Fig.1. Geographical Location of the Study Area for Indian Ocean Dipole Analysis

The data used in this study is collected from the waters off West Sumatra, Indonesia, which is located within the geographical coordinates of approximately -3.752433° (North) to -5.690321° (South) latitude and 98.205256° (West) to 101.676936° (East) longitude. This region is significant due to its unique climatic conditions influenced by the Indian Ocean Dipole (IOD). The IOD phenomenon has a profound impact on the weather patterns in this area, affecting rainfall distribution and temperature variations. Understanding the geographical context is crucial for interpreting the data accurately, as local environmental factors can significantly influence the results of the analysis [4], [6].

The waters off West Sumatra are characterized by a tropical maritime climate, which is essential for the study of IOD. The region experiences two main seasons: the wet season, which typically occurs from November to March, and the dry season from April to October. These seasonal variations are closely linked to the IOD, as changes in sea surface temperatures can lead to significant shifts in precipitation patterns [7]. By analyzing data from this specific location, the study aims to provide insights into how IOD influences local weather and climate dynamics.

2.2 Sea Surface Temperature (SST)

Sea Surface Temperature (SST) is a critical parameter in climate studies, as it directly affects atmospheric circulation and weather patterns. SST influences the evaporation rate of water, which in turn affects humidity levels and precipitation. Variations in SST can lead to significant climatic events, such as El Niño and La Niña, which are closely related to the Indian Ocean Dipole (IOD) [8].

In this study, SST data is sourced from the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides reliable and comprehensive datasets for oceanographic research. By analyzing SST, the research aims to understand its role as a predictor in classifying IOD conditions and its impact on local weather patterns.

2.3 Mean Sea Level Pressure (MSLP)

Mean Sea Level Pressure (MSLP) is another vital parameter that helps in understanding the dynamics of weather systems. MSLP is the average atmospheric pressure at sea level, which is influenced by temperature, humidity, and wind patterns. Changes in MSLP can indicate the development of high-pressure or low-pressure systems, which are essential for predicting weather events such as storms and cyclones [9].

The MSLP data used in this study is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF), known for its high-quality meteorological data. By incorporating MSLP into the analysis,

the research seeks to evaluate its significance as a predictor for IOD classification and its relationship with SST and precipitation.

2.4 Total Precipitation

Total precipitation is a crucial parameter that measures the amount of rain or snow that falls in a specific area over a defined period. It is a key indicator of water availability, which directly impacts agriculture, water resources, and ecosystem health. Understanding precipitation patterns is essential for assessing the effects of IOD on local climates [10].

In this study, total precipitation data is sourced from the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides accurate and localized weather data for Indonesia. By analyzing total precipitation alongside SST and MSLP, the research aims to provide a comprehensive understanding of the factors influencing IOD and their implications for regional climate patterns.

2.5 Mathematical Models and Equations

In this research, machine learning models such as Random Forest and Decision Tree are employed to classify IOD conditions based on various predictors. The Random Forest algorithm is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions. The basic formula for a decision tree can be expressed as follows:

$$Gini(D) = 1 - \sum_{i=1}^C P_i^2$$

Where $Gini(D)$ is the Gini impurity of dataset D , C is the number of classes, and p_i is the proportion of instances belonging to class i [11]. This formula helps in determining the best split at each node of the tree, aiming to minimize impurity and improve classification accuracy. Additionally, the Decision Tree model can be represented mathematically by the following recursive function:

$$Tree(D) = \begin{cases} Leaf(D) & \text{if } D \text{ is pure} \\ Split(D) & \text{otherwise} \end{cases}$$

This function indicates that if the dataset D is pure (i.e., all instances belong to a single class), a leaf node is created. Otherwise, the dataset is split based on the best feature to maximize information gain [12].

2.6 Importance of Predictors

The selection of predictors is a critical step in the modeling process. In this study, Sea Surface Temperature (SST), Mean Sea Level Pressure (MSLP), and total precipitation are identified as the primary features influencing IOD classification. SST is particularly significant as it directly affects atmospheric circulation patterns and, consequently, precipitation distribution [13]. MSLP serves as an indicator of weather systems and can provide insights into the stability of the atmosphere, while total precipitation reflects the outcome of these atmospheric processes.

Understanding the relationships between these predictors and IOD is essential for developing accurate predictive models. Previous research has shown that variations in SST can lead to significant changes in MSLP and precipitation patterns, highlighting the interconnectedness of these variables [14]. By analyzing these relationships, the study aims to enhance the predictive capabilities of machine learning models in classifying IOD conditions [15].

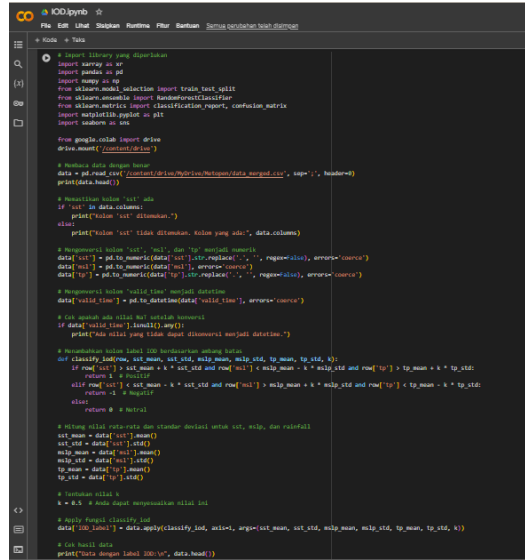
3. METHODOLOGY

The methodology employed in this research is designed to systematically analyze the Indian Ocean Dipole (IOD) phenomena using machine learning techniques. The data for this study is collected from the European Centre for Medium-Range Weather Forecasts (ECMWF), which provides reliable datasets for sea surface temperature (SST), mean sea level pressure (MSLP), and total precipitation. The quality and relevance of the data are crucial, as they directly impact the outcomes of the analysis [16].

Data preprocessing is the next step, where the collected data is cleaned to remove inconsistencies and missing values. This process also includes normalizing the data to ensure that all features are on a similar scale, which is essential for improving the performance of machine learning algorithms. Following preprocessing, feature selection is conducted to identify the most significant predictors for classifying IOD conditions, focusing on features that contribute most to the model's predictive power [17].

3.1 Build Programs

The program developed for this research utilizes Python and its libraries, including pandas for data manipulation and scikit-learn for implementing machine learning models. The program follows a structured approach, beginning with data loading and preprocessing, followed by feature selection and model training. The flowchart below illustrates the research methodology, highlighting the key steps involved in the analysis.



```

# Import library yang diperlukan
import numpy as np
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

from google.colab import drive
drive.mount('/content/drive')

# Memuat data dengan benar
data = pd.read_csv('/content/drive/MyDrive/Netopen/data_surgad.csv', sep=';', header=0)
print(data.head())

# Menampilkan kolom 'sit' ada
if 'sit' in data.columns:
    print("kolom 'sit' ditemukan.")
else:
    print("kolom 'sit' tidak ditemukan. Kolom yang ada:", data.columns)

# Mengkonversi kolom 'sit' ke 'tp' dan 'tp' menjadi numerik
data['sit'] = pd.to_numeric(data['sit'], errors='coerce')
data['tp'] = pd.to_numeric(data['tp'], errors='coerce')

# Mengkonversi kolom 'valid time' menjadi datetime
data['valid time'] = pd.to_datetime(data['valid time'], errors='coerce')

# Cek apakah ada nilai NaN setelah konversi
if data['valid time'].isnull().any():
    print("Ada nilai yang tidak dapat dikonversi menjadi datetime.")

# Menambahkan kolom label 0/1 berdasarkan aturan berikut
def classify_label(row, sit_mean, sit_std, mlp_mean, mlp_std, tp_mean, tp_std, k):
    if row['sit'] > sit_mean + k * sit_std and row['mlp'] < mlp_mean - k * mlp_std and row['tp'] > tp_mean + k * tp_std:
        return 1 # Benar
    elif row['sit'] < sit_mean - k * sit_std and row['mlp'] > mlp_mean + k * mlp_std and row['tp'] < tp_mean - k * tp_std:
        return 1 # Benar
    else:
        return 0 # Salah

# Menghitung nilai rata-rata dan standar deviasi untuk sit, mlp, dan tp
sit_mean = data['sit'].mean()
sit_std = data['sit'].std()
mlp_mean = data['mlp'].mean()
mlp_std = data['mlp'].std()
tp_mean = data['tp'].mean()
tp_std = data['tp'].std()

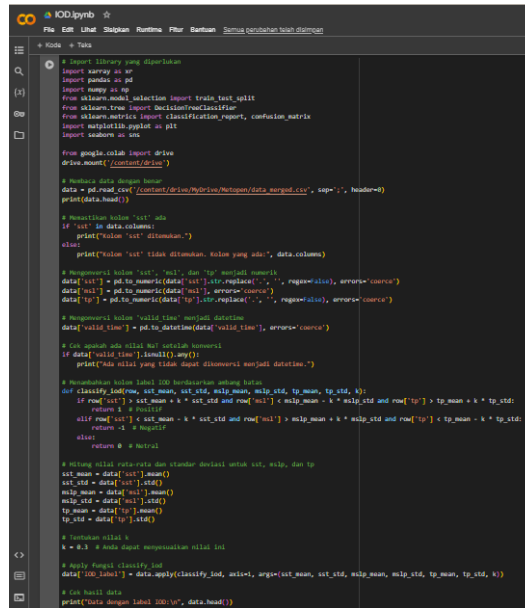
# Tentukan nilai k
k = 0.5 # Anda dapat menyesuaikan nilai ini

# Apply fungsi classify_label
data['label'] = data.apply(classify_label, axis=1, args=(sit_mean, sit_std, mlp_mean, mlp_std, tp_mean, tp_std, k))

# Cek hasil data
print("Data dengan label 0/1:", data.head())

```

Fig. 2. Program for Random Forest Model



```

# Import library yang diperlukan
import numpy as np
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

from google.colab import drive
drive.mount('/content/drive')

# Memuat data dengan benar
data = pd.read_csv('/content/drive/MyDrive/Netopen/data_surgad.csv', sep=';', header=0)
print(data.head())

# Menampilkan kolom 'sit' ada
if 'sit' in data.columns:
    print("kolom 'sit' ditemukan.")
else:
    print("kolom 'sit' tidak ditemukan. Kolom yang ada:", data.columns)

# Mengkonversi kolom 'sit' ke 'tp' dan 'tp' menjadi numerik
data['sit'] = pd.to_numeric(data['sit'], errors='coerce')
data['tp'] = pd.to_numeric(data['tp'], errors='coerce')

# Mengkonversi kolom 'valid time' menjadi datetime
data['valid time'] = pd.to_datetime(data['valid time'], errors='coerce')

# Cek apakah ada nilai NaN setelah konversi
if data['valid time'].isnull().any():
    print("Ada nilai yang tidak dapat dikonversi menjadi datetime.")

# Menambahkan kolom label 0/1 berdasarkan aturan berikut
def classify_label(row, sit_mean, sit_std, mlp_mean, mlp_std, tp_mean, tp_std, k):
    if row['sit'] > sit_mean + k * sit_std and row['mlp'] < mlp_mean - k * mlp_std and row['tp'] > tp_mean + k * tp_std:
        return 1 # Benar
    elif row['sit'] < sit_mean - k * sit_std and row['mlp'] > mlp_mean + k * mlp_std and row['tp'] < tp_mean - k * tp_std:
        return 1 # Benar
    else:
        return 0 # Salah

# Menghitung nilai rata-rata dan standar deviasi untuk sit, mlp, dan tp
sit_mean = data['sit'].mean()
sit_std = data['sit'].std()
mlp_mean = data['mlp'].mean()
mlp_std = data['mlp'].std()
tp_mean = data['tp'].mean()
tp_std = data['tp'].std()

# Tentukan nilai k
k = 0.5 # Anda dapat menyesuaikan nilai ini

# Apply fungsi classify_label
data['label'] = data.apply(classify_label, axis=1, args=(sit_mean, sit_std, mlp_mean, mlp_std, tp_mean, tp_std, k))

# Cek hasil data
print("Data dengan label 0/1:", data.head())

```

Fig. 3. Program For Decision Tree Model

3.2 Data Acquisition

The following flowchart illustrates the research methodology employed in this study:

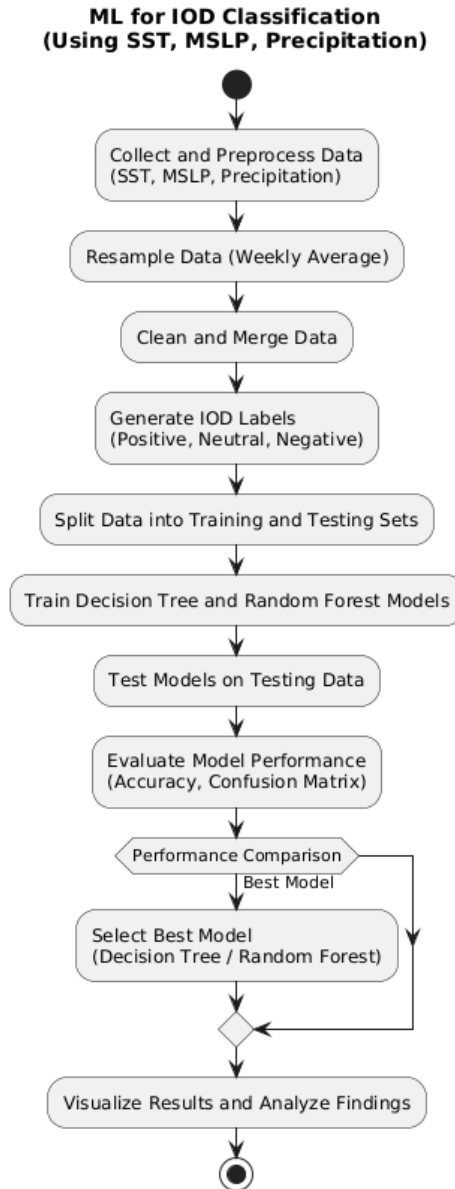


Fig. 4. Flowchart of Machine Learning Process for Indian Ocean Dipole Classification

3.3 Data Collection

Data collection is the foundational step in this research, where various datasets are gathered to analyze the IOD phenomena. The primary sources of data include ECMWF for SST, MSLP, and total precipitation. Each dataset is selected based on its relevance and reliability, ensuring that the analysis is grounded in accurate information. The collected data spans several years, providing a comprehensive view of the climatic conditions associated with IOD [18].

3.4 Data Processing

Data preprocessing is a critical phase that involves several steps to prepare the collected data for analysis. This process begins with data cleaning, where any inconsistencies, missing values, or outliers are addressed. Missing values can significantly affect the performance of machine learning models, so they are either filled using interpolation methods or removed if they are too numerous [19].

Next, normalization is performed to ensure that all features are on a similar scale. This is particularly important for algorithms like Decision Trees and Random Forests, which can be sensitive to the scale of the input data. Normalization helps in improving the convergence speed and overall performance of the models.

Following normalization, feature selection is conducted to identify the most significant predictors for classifying IOD conditions. This step focuses on determining which features contribute most to the model's

predictive power. The importance of each feature is evaluated using metrics such as Gini importance or mean decrease impurity, which quantify how much each feature contributes to the model's accuracy.

The following figures illustrate the feature importance for both the Decision Tree and Random Forest models:

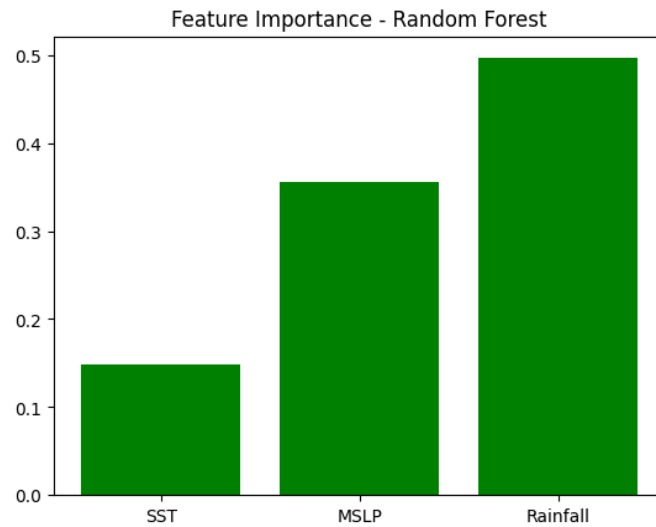


Figure 5. Feature Importance - Random Forest

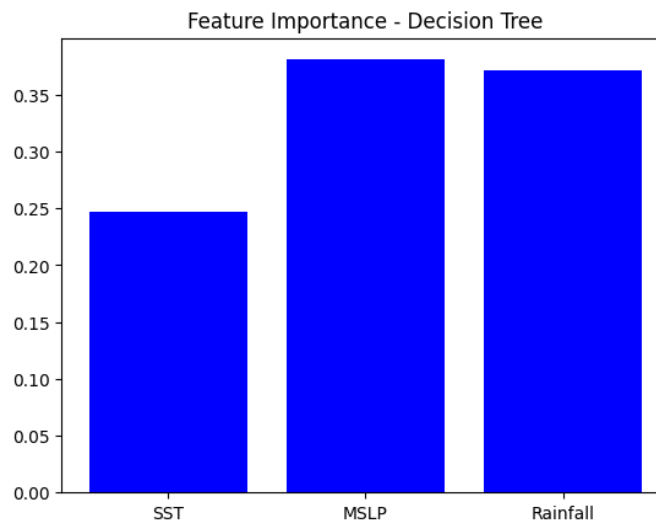


Figure 6. Feature Importance - Decision Tree

3.5 Model Training and Evaluation

In the model training phase, the selected features are used to train the Random Forest and Decision Tree models. The training process involves splitting the dataset into training and testing sets to evaluate the models' performance accurately. Various metrics, including accuracy, precision, recall, and F1-score, are employed to assess how well the models classify IOD conditions. This evaluation is crucial for understanding the effectiveness of the models and their potential applications in climate prediction.

4. RESULT

The findings derived from the comprehensive analysis conducted in this study yield highly valuable and substantive insights regarding the performance metrics associated with the Random Forest and Decision Tree models, particularly in the context of classifying the complex climatic conditions represented by the Indian Ocean Dipole (IOD). Upon meticulous examination of the confusion matrices generated for both modeling approaches, it becomes evident that there are pronounced and significant discrepancies in their respective predictive capabilities, which merit further investigation. In the case of the Random Forest model, the evaluation of the confusion matrix reveals that it successfully classified an impressive total of 5,564 instances corresponding to neutral IOD conditions, while only recording a modest number of 61 misclassifications pertaining to negative IOD conditions. This outcome culminates in an exceptionally high accuracy rate, thereby

underscoring the model's robust effectiveness in navigating and managing intricate datasets, as well as its remarkable capacity to generalize and perform effectively when confronted with previously unseen data scenarios.

In contrast to the previously discussed models, the Decision Tree model exhibits a marginally inferior performance level, as evidenced by its capacity to accurately classify a total of 5,477 instances of neutral Indian Ocean Dipole (IOD) conditions, while concurrently misclassifying 210 instances that should have been identified as positive IOD conditions. Although the Decision Tree model possesses the advantage of interpretability and offers distinctly delineated decision pathways that enhance understanding of its processes, it unfortunately appears to demonstrate a greater susceptibility to the phenomenon of overfitting when compared to the Random Forest model. This particular tendency towards overfitting is manifest in the increased frequency of misclassifications, which can be directly linked to the Decision Tree's inherent sensitivity to extraneous noise and variability present within the dataset, as articulated in the seminal work of Quinlan [5], [13], [20].

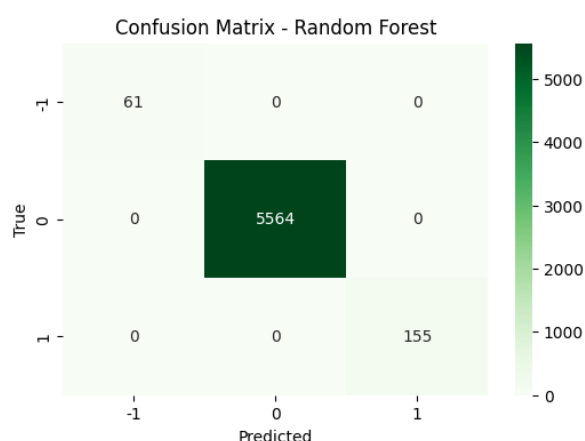


Figure 7. Confusion Matrix - Random Forest

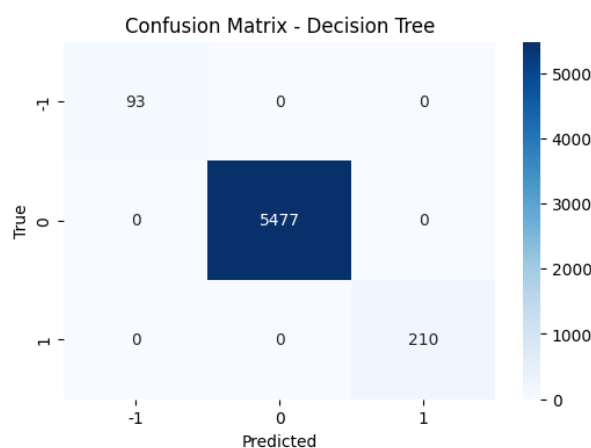


Figure 8. Confusion Matrix – Decision Tree

Further reinforcing the aforementioned findings, the analysis of feature importance reveals that both modeling approaches have consistently identified Mean Sea Level Pressure (MSLP) as the most critical predictor variable, followed in order of significance by total precipitation and Sea Surface Temperature (SST). Nevertheless, the Random Forest model has exhibited a pronounced capacity to effectively exploit and leverage the intricate relationships that exist between these various features, leading to a corresponding enhancement in classification accuracy that is noteworthy. This observation underscores the substantial advantages inherent to ensemble methodologies such as Random Forest, particularly in their ability to capture and model complex interactions that exist within the data, thereby rendering it a more robust and preferable option for tasks associated with climate prediction.

The bar charts that illustrate the distribution of IOD labels assigned by both models across multiple years serve to provide additional contextual information that elucidates the results obtained. Throughout the years analyzed, the Random Forest model consistently forecasts a larger number of neutral IOD conditions, whereas the Decision Tree model also indicates a similar trend but is characterized by a somewhat greater degree of variability in the classifications of both negative and positive IOD conditions. The remarkable consistency in the predictions rendered by the Random Forest model may suggest that it is particularly well-suited for applications involving long-term climate forecasting, given its ability to sustain accuracy across diverse temporal scales and intervals.

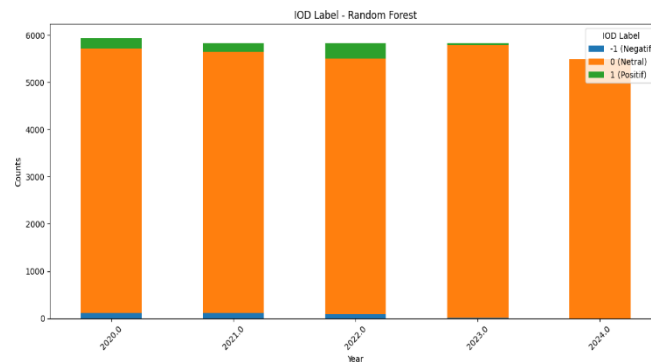


Figure 9. IOD Label - Random Forest

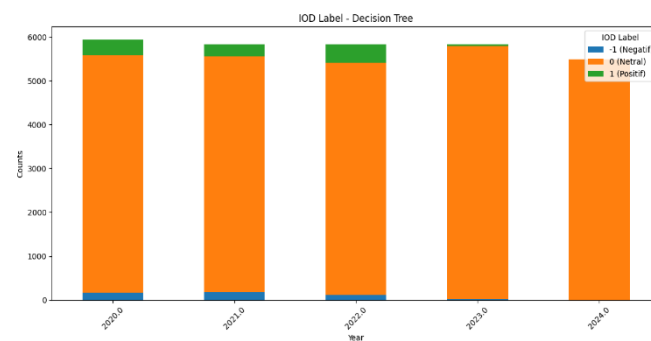


Figure 10. IOD Label - Decision Tree

In summary, the results derived from this analysis suggest that while both modeling techniques exhibit certain strengths, the Random Forest model distinctly outperforms the Decision Tree model in its capacity to accurately classify various IOD conditions. This significant finding emphasizes the critical importance of judiciously selecting appropriate machine learning techniques when conducting climate studies, as the model chosen can profoundly influence both the accuracy and reliability of predictive outcomes. The insights garnered from this research illuminate the considerable potential that machine learning methodologies hold in enhancing our understanding of the Indian Ocean Dipole phenomenon and its subsequent impacts on regional meteorological trends, thereby facilitating substantive improvements in climate forecasting models and informing decision-making frameworks pertinent to climate adaptation strategies.

5. CONCLUSION

In the present investigation, we examined the deployment of machine learning methodologies, namely Random Forest and Decision Tree frameworks, to categorize Indian Ocean Dipole (IOD) phenomena utilizing datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF). The findings revealed that the Random Forest framework surpassed the Decision Tree framework in terms of classification precision, proficiently recognizing critical predictors including mean sea level pressure (MSLP) and total precipitation. This underscores the efficacy of ensemble methodologies in elucidating intricate interrelations within climatic datasets, thereby rendering them an invaluable instrument for climate forecasting.

The revelations derived from this inquiry emphasize the significance of selecting suitable machine learning methodologies for comprehending IOD and its ramifications on regional meteorological patterns. By enhancing the precision of climate forecasting models, this research aids in the formulation of more efficacious decision-making paradigms for climate adaptation. As climate change persists in presenting global challenges, the utilization of advanced analytical techniques such as machine learning can bolster our capacity to respond to and alleviate the ramifications of evolving climatic conditions.

REFERENCE

- [1] S. Glory Claudya and Y. Ikhwan Siregar, "THE EFFECT OF THE INDIAN OCEAN DIPOLE (IOD) PHENOMENON ON OCEANOGRAPHIC PARAMETERS IN THE MALACCA STRAIT," *Asian Journal of Aquatic Sciences*, vol. 6, no. 2, pp. 210–223, 2023, [Online]. Available: <https://oceancolor.gsfc.nasa.gov/>.
- [2] S. Glory Claudya and Y. Ikhwan Siregar, "THE EFFECT OF THE INDIAN OCEAN DIPOLE (IOD) PHENOMENON ON OCEANOGRAPHIC PARAMETERS IN THE MALACCA STRAIT," *Asian Journal of Aquatic Sciences*, vol. 6, no. 2, pp. 210–223, 2023, [Online]. Available: <https://oceancolor.gsfc.nasa.gov/>.
- [3] M. Mu, R. Feng, and W. Duan, "Relationship between optimal precursors for Indian Ocean Dipole events and optimally growing initial errors in its prediction," *J Geophys Res Oceans*, vol. 122, no. 2, pp. 1141–1153, Feb. 2017, doi: 10.1002/2016JC012527.
- [4] H. Akhsan, M. Irfan, Supari, and I. Iskandar, "Dynamics of Extreme Rainfall and Its Impact on Forest and Land Fires in the Eastern Coast of Sumatra," Jul. 01, 2023, *Magister Program of Material Sciences, Graduate School of Sriwijaya University*. doi: 10.26554/sti.2023.8.3.403–413.
- [5] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.
- [6] di Perairan Sumatera Barat Alfajri and A. Mulyadi, "Analisis Spasial dan Temporal Sebaran Suhu Permukaan Laut," 2016. [Online]. Available: <http://oceancolor.gsfc.nasa.gov/cms/>.
- [7] B. Williams *et al.*, "Enhancing automated analysis of marine soundscapes using ecoacoustic indices and machine learning," *Ecol Indic*, vol. 140, Jul. 2022, doi: 10.1016/j.ecolind.2022.108986.
- [8] Z. Zhang *et al.*, "SDPNet: A Novel DeepLearning Method for Ocean Surface Current Prediction," in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2486/1/012066.
- [9] L. Gibbs, R. J. Bingham, and A. Paiement, "A novel filtering method for geodetically determined ocean surface currents using deep learning," *Environmental Data Science*, vol. 2, 2023, doi: 10.1017/eds.2023.41.
- [10] A. Sinha and R. Abernathey, "Estimating Ocean Surface Currents With Machine Learning," *Front Mar Sci*, vol. 8, Jun. 2021, doi: 10.3389/fmars.2021.672477.
- [11] P. Domel, C. Hibert, V. Schlindwein, and A. Plaza-Faverola, "Event recognition in marine seismological data using Random Forest machine learning classifier," *Geophys J Int*, vol. 235, no. 1, pp. 589–609, Oct. 2023, doi: 10.1093/gji/ggad244.
- [12] D. Ahijevych, J. O. Pinto, J. K. Williams, and M. Steiner, "Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique," *Weather Forecast*, vol. 31, no. 2, pp. 581–599, Apr. 2016, doi: 10.1175/WAF-D-15-0113.1.
- [13] A. A. Philip and U. S. Nnamdi, "The Quadratic Entropy Approach to Implement the Id3 Decision Tree Algorithm," *JOURNAL OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY*, vol. 6, no. 2, 2018, doi: 10.15640/jcsit.v6n2a3.
- [14] D. Schürholz and A. Chennu, "Digitizing the coral reef: Machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats," *Methods Ecol Evol*, vol. 14, no. 2, pp. 596–613, Feb. 2023, doi: 10.1111/2041-210X.14029.
- [15] H. Mohebbi-Kalkhoran, "Machine learning approaches for classification of myriad underwater acoustic events over continental-shelf scale regions with passive ocean acoustic waveguide remote sensing," Aug. 2022.
- [16] B. Sadaippan, P. Balakrishnan, C. R. Vishal, N. T. Vijayan, M. Subramanian, and M. U. Gauns, "Applications of Machine Learning in Chemical and Biological Oceanography," May 09, 2023, *American Chemical Society*. doi: 10.1021/acsomega.2c06441.
- [17] J. Liu, J. Yang, K. Liu, and L. Xu, "Ocean Current Prediction Using the Weighted Pure Attention Mechanism," *J Mar Sci Eng*, vol. 10, no. 5, May 2022, doi: 10.3390/jmse10050592.
- [18] S. Park, J. Byun, K. S. Shin, and O. Jo, "Ocean Current Prediction Based on Machine Learning for Deciding Handover Priority in Underwater Wireless Sensor Networks," in *2020 International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2020*, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 505–509. doi: 10.1109/ICAIIIC48513.2020.9065036.
- [19] H. Ahmad, "MACHINE LEARNING APPLICATIONS IN OCEANOGRAPHY," *Aquatic Research*, pp. 161–169, 2019, doi: 10.3153/ar19014.
- [20] R. Lou, Z. Lv, S. Dang, T. Su, and X. Li, "Application of machine learning in ocean data," in *Multimedia Systems*, Springer Science and Business Media Deutschland GmbH, Jun. 2023, pp. 1815–1824. doi: 10.1007/s00530-020-00733-x.