Research Article

Machine Learning Approach to Analyze the Relationship Between State Defense Index and Human Development to Strengthen National Defense

[1,*]Farah Mufida Qotrunnada, [2]Setiyo Budiyanto, [3]Achmad Farid Wajdi

[1,2,3] Department of Remote Sensing, Faculty of Defense Science and Technology, Defense University, Bogor City, Indonesia

* Corresponding Author: farahmufidaqotrunnada@gmail.com

Abstract: In efforts to strengthen national defense, it is important to understand how factors in human development, such as education, health, and economic welfare, can influence public awareness of national defense. This study aims to analyze the relationship between the National Defense Index (IBN) and the Human Development Index (IPM) in Indonesia using a Machine learning approach. To strengthen national defense, it is essential to understand how factors in human development, such as education, health, and economic welfare, can affect public awareness of national defense. Machine learning methods are applied to analyze the significant relationship between IBN and IPM, which is expected to provide insights for the development of more data-driven national defense policies. The results show that the Machine learning model can predict IBN values with high accuracy, supported by a Mean Squared Error (MSE) of 0.000638 and an R-squared value of 0.9026. This indicates that 90.26% of the variability in IBN values can be explained by the model, suggesting accurate predictions that are relevant for data-driven policies. Collaboration with various stakeholders is expected to enhance the application of these findings in further studies and the formulation of national defense policies.

Keywords: Machine learning, Python, Human Development Index, National Defense Index, Collaboration

Citation: Qotrunnada, F. M., Budiyanto, S., & Wajdi, A. F. (2025). Machine learning approach to analyze the relationship between state defense index and human development in an effort to strengthen national defense. *Iota*, 5(1). https://doi.org/10.31763/iota.v5i1.869

Academic Editor: Adi, P.D.P

Received: January 04, 2025

Accepted: January 23, 2025

Published: February 08, 2025

1. Introduction

Indonesia as an archipelago that is diverse in social, cultural, and economic aspects faces complex challenges in maintaining national stability and resilience. Sustainable human development is an important factor in strengthening national defense, where community involvement and state defense awareness are indispensable to maintaining national stability. Therefore, understanding the relationship between the Index of State Defense (IBN) and the Human Development Index (HDI) is important so that the government can formulate targeted policies to improve national security and resilience.

Along with rapid digital development, mastery of data processing has become a much-needed capability in various sectors, such as business, industry, science, and technology. This ability plays an important role in increasing efficiency and supporting developments in various fields with information-based decision-making (Mahmud et al., 2022). To realize this potential, a deep understanding of data analysis is required to extract useful insights from the available information. These skills are becoming increasingly valuable as data plays a crucial role in strategic decision-making (Hastie et al., 2009).

This research uses exploratory data analysis or EDA as an important step in the data analysis process that allows users to understand the structure, patterns, and basic characteristics of the available datasets. Using statistical and visualization techniques, EDA enables the identification of anomalies, hidden patterns, and relationships between variables that can support further analysis (Tukey, 1977).

EDA also helps identify and address data hilling and outliers, which is important to ensure the data is in optimal condition. The use of machine learning and platforms such as Google Collaboratory to process data with Python is used in this study, as it is a strategic step in developing a competent workforce that is ready to face future challenges (Van, 2009). In this context, machine learning analysis is a useful approach to explore the relationship between the Index of State Defense (IBN) and the Human Development Index (IPM) in Indonesia.

This research aims to apply a machine learning approach in analyzing the data from both indices to understand how factors in HDI, such as education, health, and economic well-being, contribute to people's national defense awareness. This is expected to provide deeper insights for policymakers to devise a more responsive and data-driven national defense strategy. To support the research objectives, reference is needed from several previous studies.

Research by Atmojo et al. compared the accuracy of several machine learning methods in predicting the Human Development Index (HDI) in Indonesia. This research used four methods: K-Nearest Neighbors (KNN), Random Forest, AdaBoost, and Support Vector Machine (SVM), focusing on three main indicators: Life Expectancy, Average Years of Schooling, and Expenditure per Capita. The results showed that while SVM had the highest Mean Squared Error (MSE) on the training data, it produced the lowest MSE on the testing data, making it the method with the most accurate predictions. Followed by AdaBoost, Random Forest, and KNN, these findings highlight the effectiveness of SVM in HDI analysis and provide important insights for the application of machine learning in human development policy (Atmojo et al., 2024).

Research by Pamungkas and Widiyanto discusses the use of the Support Vector Machine (SVM) method to classify the Human Development Index (HDI) in Indonesia in 2022. This research aims to measure the accuracy of HDI classification and uses the Confusion Matrix as a performance measurement tool. The results showed that SVM, with a gamma parameter setting of 1 and C=1, 10, and 100, achieved a classification accuracy rate of 95.36%. The study also found that increasing the amount of training data is not always directly proportional to classification accuracy, so it is important to consider the ratio between training data and test data scientifically (Pamungkas, 2023).

Furthermore, research by Fahrurrozi et al. discussed the relationship between the Human Development Index (HDI) and economic resilience in East Lombok Regency. The results showed that four variables had a positive effect on HDI, namely Life Expectancy, Expected Years of Schooling, Average Years of Schooling, and Expenditure. However, the Expected Years of Schooling and Expenditure indicators show a low contribution to HDI progressivity (Fahrurrozi et al., 2023).

Different from previous studies that focus on social welfare or stability, this research links human development directly to state defense awareness as an element of national security. Machine learning approaches enable the identification of hidden patterns and correlations in data, resulting in a predictive model that supports evidence-based defense policy. In addition to understanding current conditions, the model also provides preventive insights that contribute to a more strategic national defense policy. Machine learning approaches can reveal a significant relationship between the Index of State Defense (IBN) and the Human Development Index (HDI), where a high HDI will have a positive impact on people's state defense awareness. Thus, the results of this analysis are expected to support the development of a more strategic and evidence-based national defense policy in Indonesia.

2. Theory

2.1 *National Defense Index (IBN)*

The National Defense Index is a measure or assessment used to measure the level of participation and commitment of citizens in supporting national defense and security. This index covers various aspects, such as national awareness, the spirit of state defense, involvement in activities that support defense, and active contributions to strengthening national security. The purpose of the State Defense Index is to measure the extent of public awareness and involvement in national defense efforts, as well as to encourage greater participation from various levels of society in supporting defense and security efforts. The State Defense Index may include indicators such as participation in state defense training, the number of civil defense volunteers, support for defense policies, levels of patriotism and nationalism, and positive attitudes towards national values and symbols.

2.2 *Python*

Python is one of the most popular programming languages in data processing due to its advantages in code readability and a variety of libraries that support efficient data analysis. By using Python, users can perform various data processing processes such as processing, analysis, visualization, and data modeling easily. Libraries such as NumPy, Pandas, and Matplotlib provide powerful functions for array and data frame manipulation, tabular data processing, and the creation of attractive graphs and visualizations. In addition, Python also supports the development of artificial intelligence and machine learning through libraries such as Scikit-Learn, TensorFlow, and PyTorch, which allow users to build complex predictive models and perform deeper data analysis. With these features, Python has become the first choice for data scientists, data analysts, and software developers to process data efficiently and effectively in a variety of application contexts.

2.3 *Machine Learning*

Machine learning (ML) is a branch of artificial intelligence (AI) that teaches computers to learn from data and experience without the need to be explicitly programmed (S. Subramanian, et.al, 2021), (T. Jadhav et al, 2023), (T. R. N and R. Gupta, 2020), (B. Wang and W. Zhang, 2023), (R. Farhat, et.al, 2020), (R. Yadav and A. Bhat, 2024), (S. -F. Zhang, et.al, 2019). The main goal of machine learning is to develop algorithms that allow computers to perform certain tasks automatically based on patterns and information contained in data. In general, there are several approaches to machine learning:

- Supervised Learning: Models are trained using data that is already labeled, to predict or classify new data. Example applications include image recognition, spam email classification, and stock price prediction.
- Unsupervised Learning: Models are trained on unlabeled data, to discover hidden structures or patterns in the data. Example applications include clustering and dimensionality reduction.
- Reinforcement Learning: Models learn through interaction with the environment, receiving feedback in the form of rewards or punishments, to maximize a particular outcome. Example applications include robotics control and video games.

Machine learning utilizes various techniques and algorithms such as neural networks, decision trees, support vector machines, and others (W. F. C. Rocha, et.al. 2022), (C. Richardson, et.al, 2022), (N. M. Kailash Varma, et.al, 2024).

2.4 *Google Colab*

Google Colab (Collaboratory) is a cloud-based platform provided by Google to conduct research and development in data science for free. Colab allows users to write and execute Python code in a Jupyter Notebook environment integrated with Google Drive. One of the main advantages of Colab is its ability to use powerful computing resources without requiring additional setup or costs, as Colab uses Google's cloud infrastructure.

3. Method

This research uses a quantitative approach to analyze data on the Index of State Defense (IBN) and the Human Development Index (HDI), as well as data related to the factors that influence both. This research focuses on analyzing the relationship between previously measured variables.

The research phase began with the collection of IBN and HDI prediction data, as well as factors that might affect both indices. This data was then verified and cleaned to ensure quality and consistency. Next, descriptive analysis was conducted to summarize the characteristics of the data, such as mean, standard deviation, and frequency distribution, thus providing an overview of the variables under study. After that, correlation analysis was conducted using the Pearson correlation coefficient to measure the strength and direction of the linear relationship between IBN and HDI predictions, to determine whether or not there was a significant relationship between the two variables.

The next step was a multiple linear regression analysis used to identify factors that might influence the relationship between IBN and HDI. This regression model was used to measure the relative influence of each factor on HDI, by including the IBN prediction as one of the independent variables.

The result of this research is a predictive model that shows a significant relationship between the Index of State Defense (IBN) and the Human Development Index (HDI) in Indonesia. The results can be used to support more strategic and evidence-based policymaking to strengthen national resilience. The government can use these results to design more targeted policies to improve people's quality of life and build awareness of state defense. The framework of this study is organized as in Figure 1.

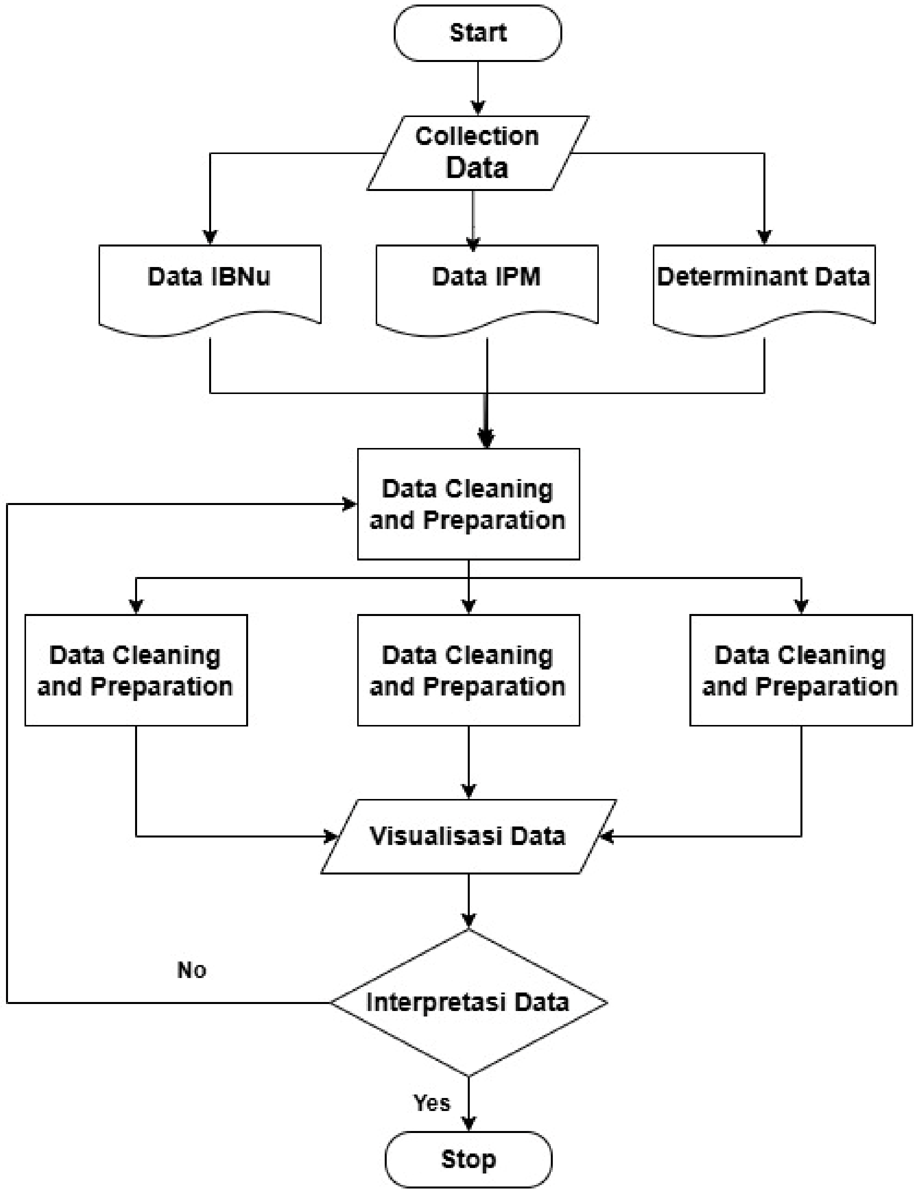

Figure 1. Flowchart System

Figure 1 describes the structured data analysis process. The process starts with a "Start" point, followed by a "Data Collection" step, where two types of data ("IBNu Data" and "HDI Data") are collected along with "Determinant Data". Once the data is collected, there is a "Data Cleaning and Preparation" step to ensure the data is ready for analysis. Once the data was prepared, three types of analysis were conducted: "Descriptive Analysis", "Correlation Analysis", and "Regression Analysis". These analyses are complemented by "Data Visualization" to help understand the results. After visualization, there is an "Interpretation of Results" step, where the results of the analysis are reviewed. If the interpretation is deemed inadequate or incomplete, the process will return to conduct further analysis. If the interpretation is adequate ("Yes"), the process reaches the "Finish" stage, which signifies the end of the analysis workflow. This iterative process ensures that the analysis is thorough and of high quality, as per the research being conducted.

4. Result and Discussion

4.1 Data Processing

The data that has been processed is processed using Google Colab with the following steps:

4.1.1 *Pre-Processing*: prepare the program by importing the libraries that will be used.

```python
# Impor pustaka yang dibutuhkan
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
```

Figure 2. Library required

4.1.2 *Dataset processing*: displaying available datasets, filtering data according to the provinces in the data, and merging data based on provinces.

```python
# membaca dataset
data_thn = pd.read_csv('content/penjualan_thn_per_provinsi_19-20.csv')
data_jan = pd.read_csv('content/jan2021_2023_30_01.csv')
data_feb = pd.read_csv('content/jan2023_20.02.csv')
data_prev = pd.read_csv('content/predTargetData_20.02.csv')

# Filter data sesuai provinsi yang ada di data_thn
provinsi_list = data_thn['Provinsi'].unique()
data_prev_filtered = data_prev[data_prev['Provinsi'].isin(provinsi_list)]
data_jan_filtered = data_jan[data_jan['Provinsi'].isin(provinsi_list)]
data_feb_filtered = data_feb[(data_feb['Provinsi'].isin(provinsi_list)) & (data_feb['2023'

# Menggabungkan data berdasarkan provinsi
data_merge = data_thn.merge(data_jan_filtered, on='Provinsi', how='left').merge(data_jan_f
```

Figure 3. Datasets, data filters, and data merging

4.1.3 *Exploratory Data Analysis:* The process of exploring the characteristics of the dataset before conducting further modeling.

	Provinsi	Y_target	X_jan_prev	pred_periode	target_pred
0	Riau	1.00000	0.85947	0.89871	0.89871
1	Jawa Timur	1.00000	0.88492	0.88492	0.88492
2	Jawa Barat	0.92604	0.90000	0.90000	0.90000
3	DKI Jakarta	1.00000	0.89871	0.89871	0.89871
4	Lampung	1.00000	0.89871	0.89871	0.89871
5	Bengkulu	0.82510	0.85947	0.85947	0.85947
6	Sumatera Selatan	0.93156	0.88095	0.88095	0.88095
7	Maluku	0.91234	0.88492	0.88492	0.88492
8	Kalimantan Barat	1.00000	0.88492	0.88492	0.88492
9	Maluku Utara	0.94290	0.89093	0.89093	0.89093

	target_periode	error_margin	kriteria	realisasi	prediksi	persentase
0	0.89871	0.27416	0.24563	0.82510	0.82510	0.42970
1	0.88492	0.24563	0.27416	0.85947	0.85947	0.44793
2	0.90000	0.24563	0.27416	0.85947	0.85947	0.44793
3	0.89871	0.27416	0.24563	0.82510	0.82510	0.42970
4	0.89871	0.27416	0.24563	0.82510	0.82510	0.42970
5	0.85947	0.27416	0.24563	0.82510	0.82510	0.42970
6	0.88095	0.24563	0.27416	0.85947	0.85947	0.44793
7	0.88492	0.24563	0.27416	0.85947	0.85947	0.44793
8	0.88492	0.24563	0.27416	0.85947	0.85947	0.44793
9	0.89093	0.24563	0.27416	0.85947	0.85947	0.44793

Figure 4. Data Analysis

4.1.4 *Machine Learning Process:* Processing data by testing data, validation data, and training data.

```python
# Import required libraries
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt
import seaborn as sns

# Generate predictions
y_pred = model.predict(X_test)

# Evaluate model using regression metrics
print("Mean Squared Error:", mean_squared_error(y_test, y_pred))
print("R-squared Score:", r2_score(y_test, y_pred))

# Visualize actual vs predicted values
plt.figure(figsize=(10, 6))
plt.scatter(y_test, y_pred, color='green')
plt.xlabel('Actual Values')
plt.ylabel('Predicted Values')
plt.title('Actual vs Predicted Values')
plt.show()
```

Figure 5. Machine Learning Process

Based on the analysis, the Mean Squared Error (MSE) value of 0.000638 indicates that the average squared error between the IBN value predicted by the model and the actual IBN value is very small. A low MSE indicates that the model has good predictive ability, with relatively minimal prediction error. The R-squared of 0.9026 indicates that 90.26% of the variation in IBN values can be explained by the model. A high R-squared indicates that the model can explain most of the variation in the data and produce accurate predictions. Overall, the low MSE and high R-squared indicate that the machine learning model used in this study performs very well in predicting IBN values.

4.1.5 *Model Evaluation:* Testing the performance of the constructed model by displaying the relationship between parameters in the model.

```python
# Import required libraries
import matplotlib.pyplot as plt
import seaborn as sns

# Visualize distribution plot for target variable
plt.figure(figsize=(10, 6))
sns.histplot(data=data, x='Target TBN mgr Provinsi', color='blue')
plt.xlabel('Distribution')
plt.title('Target TBN mgr Provinsi')
plt.show()

# Create boxplot
plt.figure(figsize=(8, 5))
sns.boxplot(x='TBN', y='Provinsi', data=data)
plt.xlabel('TBN')
plt.ylabel('Provinsi')
plt.show()

# Create heatmap
heatmap_data = data.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(data=heatmap_data, cmap='RdBu', annot=True, fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```

Figure 6. Model Evaluation

4.1.6 Prediction Result and Visualization

Furthermore, from data processing using phyton (Google Collaboratory) based on CSV data, the results in Figure 7 are as follows. Figure 7 describes the graph where the blue dots are mostly around the green line with little dispersion, indicating that the model performs well and the predictions are close to the actual values. Conversely, if the dots are scattered far from the green line, it indicates a larger prediction error by the model. The R-squared value of 0.902 indicates that there is a strong relationship between the actual and predicted values. In addition, the small Mean Squared Error (MSE) value of 0.0006 indicates that the average squared error between the predicted and actual values is very small. This indicates that the model has good performance and minimal prediction error.

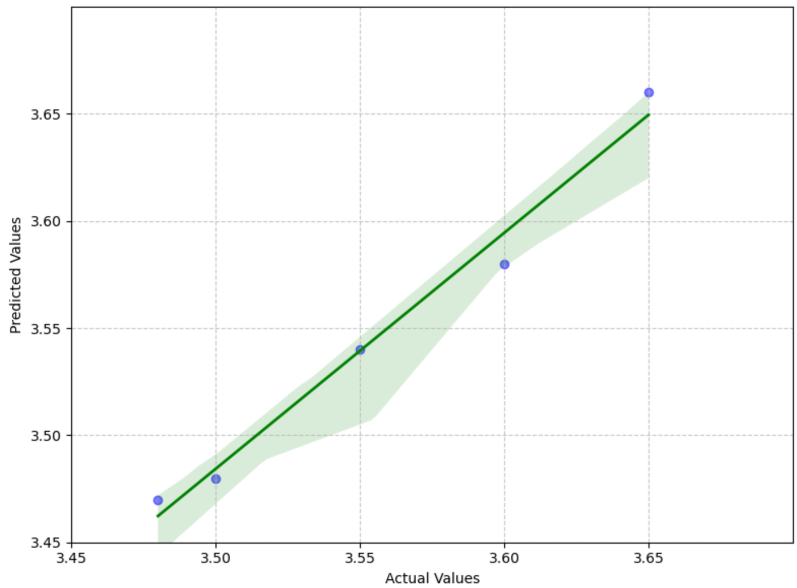

Figure 7. Graph between Actual Value and Predicted Value

Furthermore, a graphical visualization of the relationship between IBN and all provinces was conducted as shown in Figure 8.

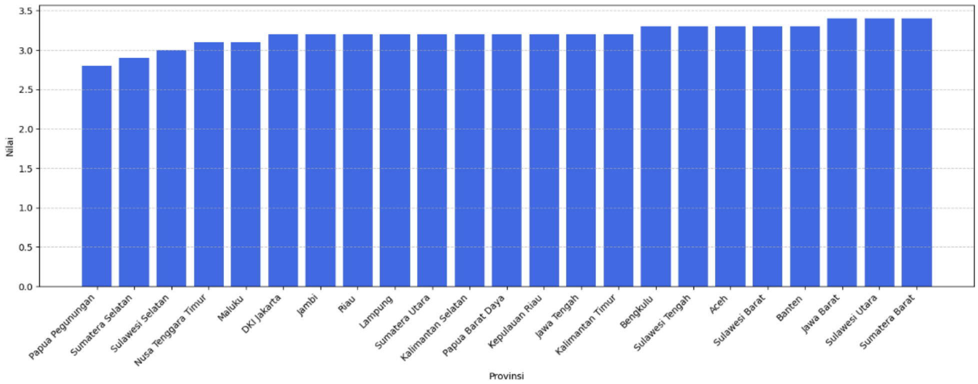

Figure 8. Relationship between IBN and Province

Figure 8 displays a graph where each bar represents the IBN value for each province. Provinces with higher IBN values are shown with taller bars on the graph. Conversely, provinces with lower IBN values are shown with shorter bars. This visualization shows the variation in IBN values between provinces. Some provinces such as Papua Mountains and South Sumatra show lower IBN values, while other provinces such as Central Sulawesi and West Nusa Tenggara show higher IBN values. Figure 9 shows the distribution graph of the relationship between IBN and provinces.

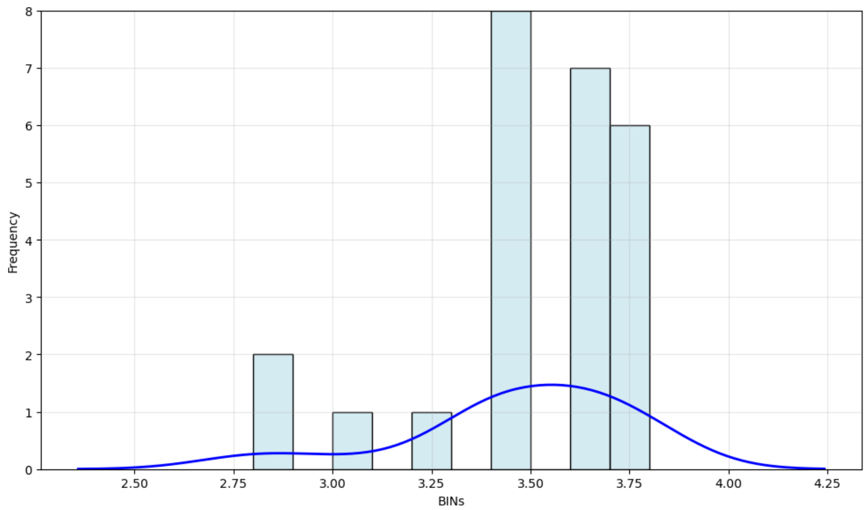

Figure 9. Distribution graph of the relationship between IBN and Province

Moreover, from Figure 9, which shows the distribution graph of the relationship between IBN and provinces, it can be seen that the IBN values fall into several main groups spread across provinces. The highest peaks on the histogram are around IBN values of 3.4 and 3.6, indicating that many provinces have IBN values ranging between these values. On the other hand, some provinces have lower IBN values, around 3.0.

5. Conclusions

The research shows that the machine learning model used to predict the value of the State Defense Index (IBN) has good performance. This is indicated by the analysis of the actual vs predicted values graph, where most of the points are close to the green line which is the model prediction line. Conversely, points scattered far from the green line indicate a greater potential for prediction error by the model. The model evaluation results are also reinforced by the Mean Squared Error (MSE) value of 0.000638, which indicates that the average squared error between the IBN value predicted by the model and the actual IBN value is very small. In addition, the R-squared value of 0.9026 indicates that 90.26% of the variability in IBN values can be explained by the model. This indicates that the model can explain most of the variation in the data and produce accurate predictions.

Based on the research results that show the good performance of the machine learning model in predicting the value of the Index of State Defense (IBN), it is recommended to continue optimizing the model by considering the integration of additional relevant features or the exploration of new techniques in machine learning that can improve prediction accuracy. In addition, it is important to develop a deeper interpretation of the analysis results to understand what factors influence the IBN values in each province. Collaboration with various stakeholders and related institutions can also improve the applicability of the results of this research in further studies.

Acknowledgments: Thanks to the work team that has completed this research. Hopefully, this research can continue to be developed, and advance machine learning in Indonesia.

Author contributions: The authors were responsible for building Conceptualization, Methodology, analysis, investigation, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision of project administration, funding acquisition, and have read and agreed to the published version of the manuscript.

Funding: The study was conducted without any financial support from external sources.

Availability of data and Materials: All data are available from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Additional Information: No Additional Information from the authors.

References

1. S. Subramanian, B. Tseng, R. Barbieri and E. N. Brown, "Unsupervised Machine Learning Methods for Artifact Removal in Electrodermal Activity," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 399-402, doi: 10.1109/EMBC46164.2021.9630535.
2. T. Jadhav et al., "Predicting Urban Land Cover Using Classification: A Machine Learning Approach," 2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC), Rajkot, India, 2023, pp. 450-454, doi: 10.1109/R10-HTC57504.2023.10461930.
3. T. R. N and R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques," 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.190.
4. B. Wang and W. Zhang, "Research on Edge Network Topology Optimization Based on Machine Learning," 2023 5th International Conference on Applied Machine Learning (ICAML), Dalian, China, 2023, pp. 41-46, doi: 10.1109/ICAML60083.2023.00018.
5. R. Farhat, Y. Mourali, M. Jemni and H. Ezzedine, "An overview of Machine Learning Technologies and their use in E-learning," 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA), Tunis, Tunisia, 2020, pp. 1-4, doi: 10.1109/OCTA49274.2020.9151758.
6. R. Yadav and A. Bhat, "A Survey on Skin Lesion Detection and Classification using Machine Learning," 2024 2nd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), Namakkal, India, 2024, pp. 1-5, doi: 10.1109/AIMLA59606.2024.10531571.
7. S. -F. Zhang, J. -H. Zhai, B. -J. Xie, Y. Zhan and X. Wang, "Multimodal Representation Learning: Advances, Trends and Challenges," 2019 International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, 2019, pp. 1-6, doi: 10.1109/ICMLC48188.2019.8949228.
8. W. F. C. Rocha, A. U. Lucena, G. F. Sarmanho, R. C. Félix, S. A. Miqueleti and T. C. Dourado, "Machine learning protocol from ultrasound data for monitoring, predicting, and supporting the analysis of dam slopes," 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 2022, pp. 1619-1623, doi: 10.1109/ICMLA55696.2022.00084.

9. C. Richardson, S. Shah, and X. Yuan, "Semi-Supervised Machine Learning for Analyzing COVID-19 Related Twitter Data for Asian Hate Speech," 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 2022, pp. 1643-1648, doi: 10.1109/ICMLA55696.2022.00252.
10. N. M. Kailash Varma, S. H. Mattaparty, S. Ismail, J. Thaduri, G. Deep Arora and A. B, "Sentiment Analysis: A Machine Learning Perspective," 2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP), New Delhi, India, 2024, pp. 1-6, doi: 10.1109/ICECSP61809.2024.10698402.
11. Kemhan.go.id. (2020). https://www.kemhan.go.id/pothan/2020/01/16/kegiatanindikator kinerja-utama-iku-direktorat-bela-negara-ditjen-pothankemhan.html [Accessed: Jul. 20, 2024]
12. Atmojo, Fernando & Nurlita, Catarina & Nurchim, Nurchim. (2024). Analisis Pemanfaatan Machine learning Guna Prediksi Indeks Pembangunan Manusia. Simtek : jurnal sistem informasi dan teknik komputer. 9. 89-96. 10.51876/simtek.v9i2.390.
13. Fahrurrozi, Muh & Mohzana, Mohzana & Haritani, Hartini & Yunitasari, Dukha & Basri, Hasan. (2023). Peningkatan Indeks Pembangunan Manusia Regional Dan Implikasinya Terhadap Ketahanan Ekonomi Wilayah. Jurnal Ketahanan Nasional. 29. 70. 10.22146/jkn.83425.
14. Mahmud, Malissa & Wong, Shiau Foong. (2022). Digital age: The importance of 21st century skills among the undergraduates. Frontiers in Education. 7. 10.3389/feduc.2022.950553.
15. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Stanford, CA: Stanford University. https://web.stanford.edu/~hastie/Papers/ESLII.pdf https://doi.org/10.1007/978-0-387-84858-7
16. Pamungkas, C. A., & Widiyanto, W. W. (2023). Klasifikasi Indeks Pembangunan Manusia Di Indonesia Tahun 2022 Dengan Support Vector Machine. Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer, 2(3), 139–145. https://doi.org/10.55606/juisik.v3i1.407.
17. Prasetyo E. 2012. Data Mining Konserp dan Aplikasi Menggunakan MATLAB. Yogyakarta: Andi.
18. Rahman, S., Sembiring, A., Siregar, D., Prahmana, I. G., Puspadini, R., & Zen, M. (2023). Python: Dasar dan Pemrograman Berorientasi Objek. Penerbit Tahta Media.
19. Tukey, J.W. (1977) Exploratory Data Analysis. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.
20. Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.