Formosa Journal of Science and Technology (FJST)
Vol. 4, No.1, 2025 477-488

Predictive Analysis of The National Defense Index (IBN) on National Resilience Using Classification and Regression Trees (CART) Method

Farah Mufida Qotrunnada[1]*, Setiyo Budiyanto[2], Achmad Farid Wadjdi[3]
Universitas Pertahanan
Corresponding Author: Farah Mufida Qotrunnada,
farahmufidaqotrunnada@gmail.com

A R T I C L E I N F O

Keywords: Technology, National Defense Index, CART, Python, Machine Learning

Received : 31, December
Revised : 14, January
Accepted: 26, January

A B S T R A C T

IBN (National Defense Index) is an important indicator in measuring public awareness and participation in maintaining the sovereignty and stability of the country. With the increasing complexity of challenges to national resilience, a predictive approach is needed to understand the relationship between IBN and other variables that influence national resilience. The CART method is used to identify patterns and determine significant variables that play a role in strengthening national resilience. Through this model, the study aims to provide deeper insight for the formulation of more effective policies in building strong national resilience. The results of the study were obtained by identifying factors that influence the National Defense Index (IBN). The results of this study are expected to be a consideration for the basis for strategic policies in achieving the great goal of Indonesia Emas 2045 as a strong, sovereign, and sustainable country.

ISSN-E: 2964-6804
DOI: https://doi.org/10.55927/fjst.v4i1.13271
https://journal.formosapublisher.org/index.php/fjst

INTRODUCTION

The National Defense Index (IBN) is a measuring tool designed to assess the level of preparedness and commitment of citizens to national defense. In Indonesia, IBN is an important indicator to measure how far society is involved in efforts to defend national sovereignty, which includes aspects of knowledge, attitude, and participation in national defense programs (Ministry of Defense of the Republic of Indonesia, 2021). According to Rahayu (2021), IBN not only measures the level of understanding of the basic values of national defense such as love of the homeland, awareness of nation and state, and loyalty to Pancasila, but also the level of real participation in daily life, both individually and collectively. A high IBN indicates better awareness and readiness in facing potential threats from both within and outside the country. Savitri (2022) said that IBN itself is an important measuring tool designed to assess the preparedness and commitment of citizens to national defense. This index includes aspects of knowledge, attitude, and participation in national defense programs, so that it becomes a crucial indicator in measuring community involvement in efforts to defend national sovereignty (Sumartini & Purnami, 2015).

In processing IBN data and other binding factors, this study uses the Classification and Regression Trees (CART) Model. CART is a data analysis technique that is useful in exploring the relationship between variables in a dataset. CART works by dividing data into several groups based on predictor variables to classify or predict the value of the target variable. This model offers easy-to-understand visualization and can help identify patterns in data that may not be visible with other analysis techniques (Breiman et al., 1986). According to Rokach & Maimon (2005), the application of the CART model in this study is in line with the trend of using more sophisticated data analysis methods in formulating public policy.

This allows decision makers to understand the complex dynamics that affect national defense awareness in various levels of society. The use of the CART model to analyze the relationship between IBN and other variables offers several advantages. First, CART can handle complex data and has non-linear interactions between variables. Second, this model can provide easily interpreted results in the form of a decision tree, thus helping to formulate more focused and effective national defense policies or programs. According to Farikhi & Pramono (2023) the Classification and Regression Trees (CART) model is a widely used method in data analysis due to its several key advantages. One of its most prominent strengths is its ability to handle both classification and regression tasks effectively by creating a decision tree. These trees are easy to understand and interpret, which is particularly helpful when the results need to be communicated to non-technical stakeholders. CART can process large datasets with many variables, and it is efficient even when dealing with mixed data types (numerical and categorical variables) (Pratiwi & Zain, 2014).

Through this analysis, it is hoped that the main factors that influence the National Defense Index (IBN) can be found and segments of society that require special attention can be identified. A deeper understanding of these factors will enable the development of more effective and data-based policies to increase public awareness and participation in national defense. Thus, efforts to improve IBN can be more focused and target groups that need strengthening, such as the younger generation and community groups with low levels of awareness.

In addition, targeted IBN enhancement is very important to achieve the vision of Indonesia Emas 2045, where Indonesia is expected to become an advanced, sovereign, and sustainable country. With a society that is more concerned and involved in national defense efforts, national resilience will be stronger in facing various challenges, both from within and outside the country. Therefore, strengthening IBN is not only an effort to strengthen defense, but also an integral part of achieving Indonesia's great goal of becoming a strong and united country in the future.

THEORETICAL REVIEW

In this section, we will examine the factors that can influence the National Defense Index (IBN), focusing on variables such as poverty, crime, conflict, and criminality. The purpose of this study is to predict or classify the IBN based on these variables, to identify factors that can worsen or improve the level of community readiness and commitment in defending the country. With a better understanding of the relationship between the IBN and these social factors, more appropriate and data-driven policies can be designed to increase citizen awareness and participation in strengthening national resilience. The study uses several literature reviews in the form of the CART model, python, google collaborative and machine learning:

Classification and Regression Trees (CART)

Classification and Regression Trees (CART) is a data analysis method used for classification and regression by building a decision tree model. In classification, CART divides the dataset into several subsets based on certain attributes to predict target categories. In regression, this method identifies patterns in the data that allow for predicting continuous values of the target variable. The resulting decision tree consists of nodes that represent decisions or rules based on input variables, and branches that describe the results of those decisions. One of the strengths of CART is its ability to handle data with non-linear interactions between variables, as well as the ease of interpreting the results through decision tree visualization (Breiman et al., 1986).

Python

Python is a very useful programming language in implementing the Classification and Regression Trees (CART) method because it offers various libraries and tools that facilitate the creation, training, and evaluation of decision tree models. One of the main libraries in Python for implementing CART is scikit-learn, which provides the Decision Tree Classifier and Decision Tree Regressor functions for classification and regression tasks, respectively. scikit-learn makes it easy for users to build decision tree models, perform tree pruning, and evaluate model performance using various metrics. In addition, Python also provides visualization libraries such as matplotlib and graphviz that help in drawing decision trees graphically, making it easier to interpret and communicate the results of the analysis (Pedregosa et al., 2011; Van Rossum & Drake, 2009).

Google Colaboratory

Google Colaboratory, or better known as Google Colab, is a cloud-based platform that provides an interactive environment for running Python code, which is very useful in implementing the Classification and Regression Trees (CART) method. With Google Colab, users can access and run Python notebooks in the cloud without having to install the software locally. This allows for flexible and collaborative programming, as well as the utilization of powerful computing resources, including GPUs and TPUs, to accelerate model training. Google Colab also supports various Python libraries needed for CART implementation, such as scikit-learn for building and evaluating decision tree models, and matplotlib for visualizing results. In addition, Google Colab makes it easy to share notebooks with colleagues and other stakeholders, facilitating effective collaboration and documentation of analysis results (Google, 2023; Abadi et al., 2016).

Machine Learning

Machine learning plays an important role in the implementation of Classification and Regression Trees (CART) by improving the model's ability to understand and process complex data. In the context of CART, machine learning provides tools and techniques for automating and optimizing the decision tree model building process. The CART algorithm, which falls under the category of machine learning, utilizes learning techniques to divide data into more homogeneous subsets based on certain features, resulting in a model that can classify data or predict continuous values with high accuracy. With machine learning, the process of tuning and validating the model becomes more systematic, thanks to methods such as cross-validation and grid search that can be implemented to find optimal model parameters. Machine learning also allows the integration of CART with other algorithms, such as ensemble methods, which can improve model performance by combining multiple decision trees for more robust results (Breiman et al., 1986; Murphy, 2012).

METHODOLOGY

Data processing begins with understanding the structure and data type of each variable in the dataset to be analyzed. The first step is to prepare the programming environment by installing Python libraries such as Pandas, Matplotlib, and Seaborn. Next, the relationship between the independent variables and the National Defense Index (IBN) is visualized to understand possible patterns.

In the analysis, the CART (Classification and Regression Trees) model is applied to explore the relationship between IBN and other variables. The steps include selecting the appropriate dependent (target) variable and independent (predictor) variables, followed by visualization of the resulting decision tree along with the interpretation of each branch. The model is evaluated with relevant metrics and cross-validated to measure its robustness if possible. From the resulting decision tree, the variables that have the most influence on IBN are identified. The results of this analysis are then discussed to show how these findings can be interpreted in the context of improving IBN. Furthermore, a model shift is performed by setting hyperparameters, such as tree depth and minimum number of samples for branch fragmentation, to improve model performance. The results of the model before and after are then compared to assess the effectiveness of the improvements applied.

The treatment in this station goes through several phases shown schematically Fig. 1.

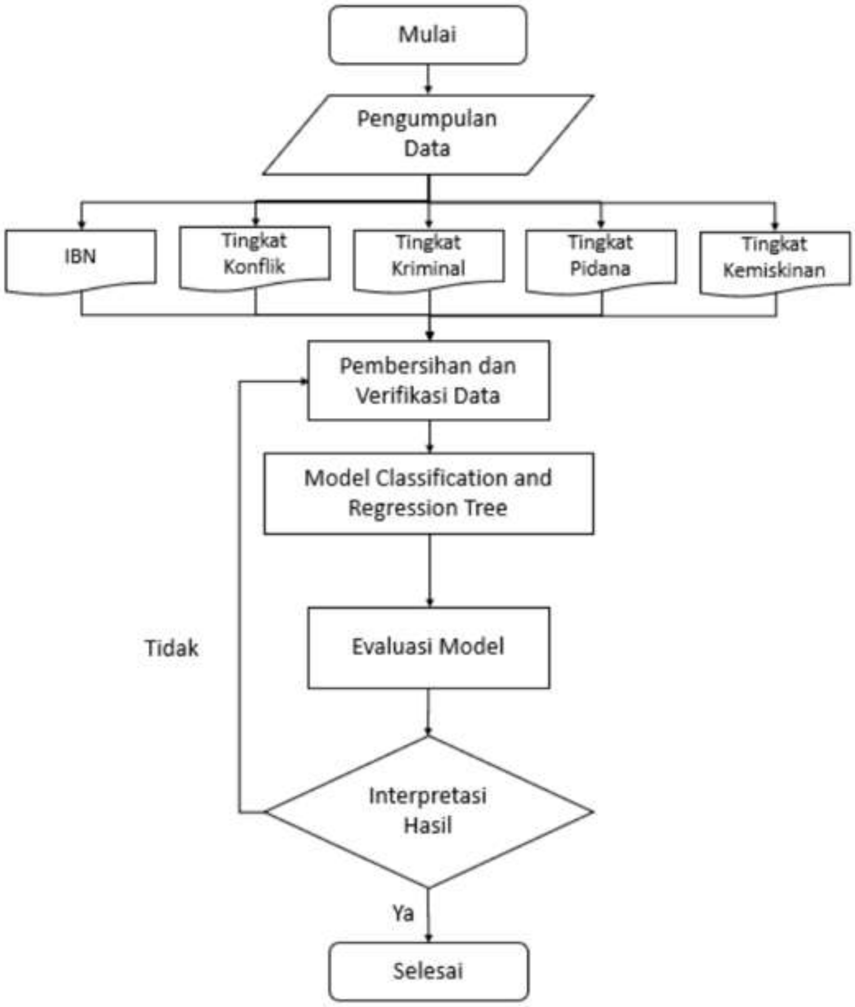


Figure 1. Schematic diagram of the physicochemical treatment process in the station of the textile industrial unit.

RESULTS AND DISCUSSION

This section discusses the technical specifications for the research on predicting the National Defense Index (IBN) based on social variables, which includes the design and functionality of various data analysis components. This includes the configuration of the Classification and Regression Trees (CART) model as the main algorithm, supporting infrastructure for managing and processing social data, and a data management system that allows for real-time monitoring and evaluation of results.

Machine Learning

Data processing is done using Google Collab and the Python programming language, as follows:

a. Data Exploration

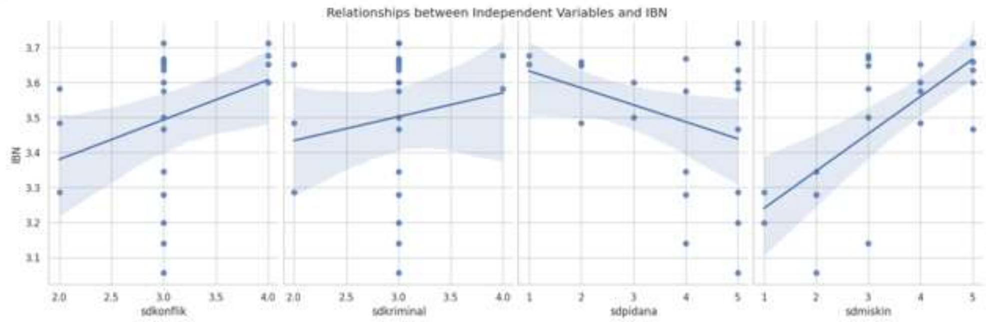
```
df.info()
print('\
')
print(df.describe())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21 entries, 0 to 20
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Provinsi    21 non-null     object
 1   IBN         21 non-null     float64
 2   sdkonflik   21 non-null     int64
 3   sdkriminal  21 non-null     int64
 4   sdpidana    21 non-null     int64
 5   sdmiskin    21 non-null     int64
dtypes: float64(1), int64(4), object(1)
memory usage: 1.1+ KB

             IBN  sdkonflik  sdkriminal   sdpidana   sdmiskin
count  21.000000  21.000000   21.000000  21.000000  21.000000
mean    3.499049   3.047619    2.952381   3.761905   3.428571
std     0.201055   0.589592    0.497613   1.410842   1.325573
min     3.056325   2.000000    2.000000   1.000000   1.000000
25%     3.344798   3.000000    3.000000   3.000000   3.000000
50%     3.583063   3.000000    3.000000   4.000000   3.000000
75%     3.651798   3.000000    3.000000   5.000000   5.000000
max     3.712173   4.000000    4.000000   5.000000   5.000000
```

With a range index of 21 entries ranging from 0 to 20, each column contains 21 values with each Province data type in the form of objects, IBN in the form of float64 and other variables in the form of int64. At the bottom are descriptive statistics, information is displayed on the number of data (count), average (mean), standard deviation (std), lowest value (min), and percentiles at 25%, 50%, and 75%, and the highest value (max) for each column.

b. Visualization of the relationship between independent variables and IBN

```
# Set the style of seaborn
sns.set(style='whitegrid')

# Create pairplot to visualize relationships between variables
sns.pairplot(df, x_vars=['sdkonflik', 'sdkriminal', 'sdpidana', 'sdmiskin'], y_vars='IBN', height=5, aspect=0.8, kind='reg')
plt.suptitle('Relationships between Independent Variables and IBN', y=1.02)
plt.show()
```



The results of the Scatter Plot visualization depicting the relationship between the independent variables and the National Defense Index (IBN) do not show any significant linear correlation between IBN and the independent variables. The data points appear to be spread randomly without a clear pattern, iIBNcating that the linear relationship between IBN and the variables is not

```
# Splitting the data into features and target variable
X = df[['sdkonflik', 'sdkriminal', 'sdpidana', 'sdmiskin']]
y = df['IBN']

# Splitting the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initializing and training the CART model
cart_model = DecisionTreeRegressor(random_state=42)
cart_model.fit(X_train, y_train)

# Making predictions on the test set
predictions = cart_model.predict(X_test)

# Calculating the mean squared error
mse = mean_squared_error(y_test, predictions)
r_squared = r2_score(y_test, predictions)

print('Mean Squared Error:', mse)
print('R-squared:', r_squared)
```

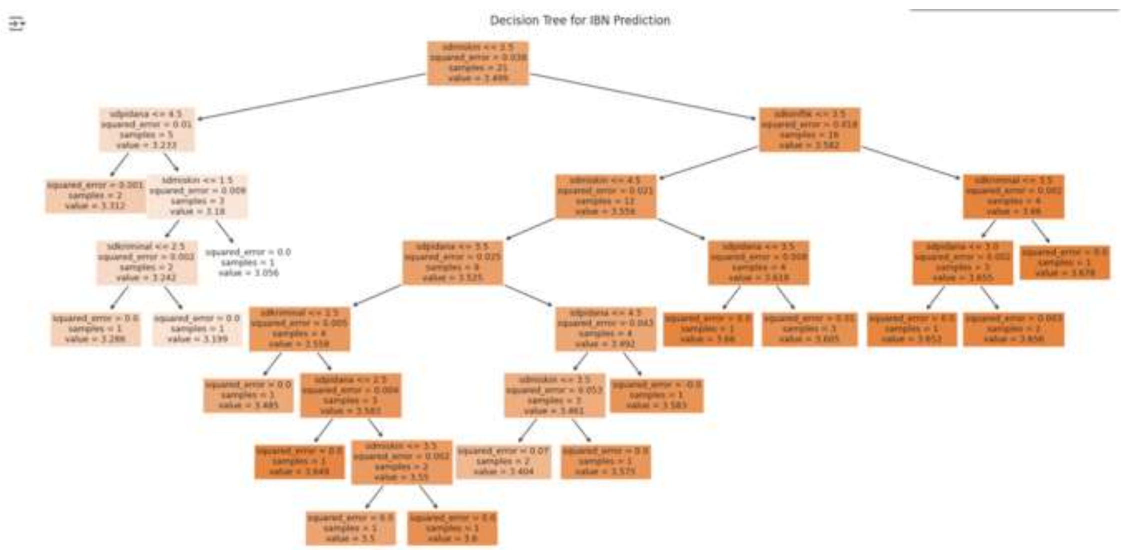
```
Mean Squared Error: 0.011559956225796229
R-squared: 0.22433795002329926
```

strong. In addition, the IBN values tend to be centered around the average, iIBNcating that the level of national defense awareness in various provinces is relatively uniform.

Modeling with CART

The Target Result of Mean Squared Error (MSE) of 0.011559956225796229 iIBNcates that the model has a fairly good performance in predicting the target with a relatively small error. While a higher R-squared value iIBNcates a better fit of the model to the data. In this case, the R-squared value of 0.22433795002329926 iIBNcates that the model explains about 22.43% of the variance in the IBN variable based on other variables.

Visualize Decision Treee



The results of the decision tree analysis can be presented as follows:

1. Root Node: sdmiskin <= 2.5 with a Squared error of 0.038 from 21 samples and has an average prediction value of 3,499. If the poverty rate (sdmiskin) is less than or equal to 2.5, then we move to the left branch. Otherwise, we move to the right branch.
2. Left Branch of Root: sdpenal <= 1.5 with a Squared error of 0.01 from 5 samples and has an average prediction value of 3,233. If the criminal rate (sdpenal) is less than or equal to 1.5, we move to the next left branch. Otherwise, we move to the right branch.
3. Right Branch of Root: sdkriminal <= 3.5 with a Squared error of 0.018 from 16 samples and has an average prediction value of 3,582. If the crime rate (sdkriminal) is less than or equal to 3.5, we move to the next left branch. Otherwise, we move to the right branch.
4. Leaf Nodes: Leaf nodes are the end points of the decision tree, where no further decisions are made due to the Squared error of 0.0 out of 1 sample. Then the node value is obtained from the predictions representing each branch. The value at this node is the final prediction for the condition that meets all the criteria along the branch.

This section addresses the overall structure of the smart dispenser system, including hardware components and network infrastructure. The system architecture refers to the layout of the dispensers, sensor arrangements, and integration of IoT components such as microcontrollers and communication modules. A well-designed system architecture

Model Evaluation

```
# Calculate metrics
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')

print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
```

```
Accuracy: 0.6
Precision: 0.8
Recall: 0.6
```

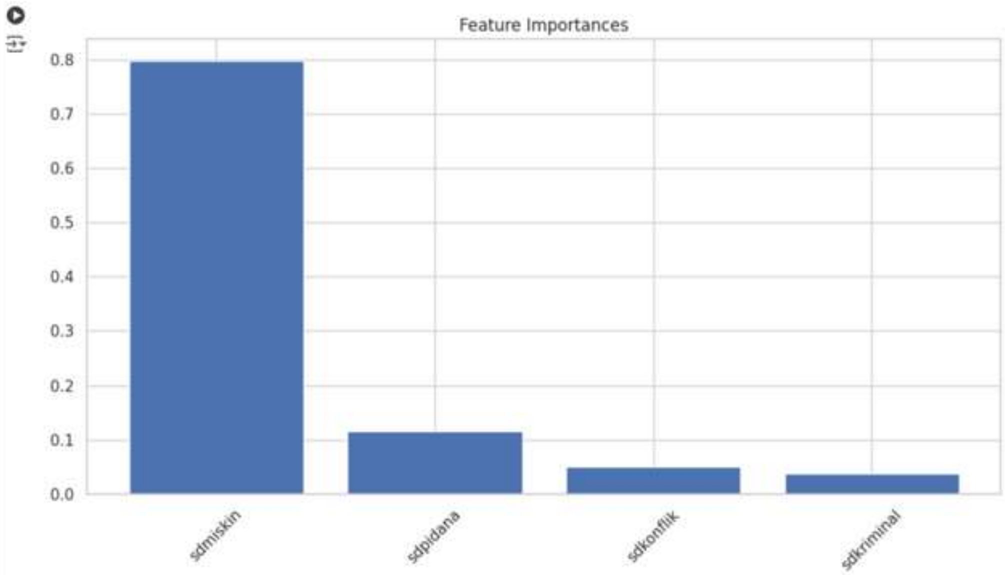
The model shows moderate performance with 60% accuracy. A precision of 80% with a Recall of 0.6 iIBNcates good positive predictive value. The class distribution shows that the 'Moderate' category is the most common.

```
# Perform 5-fold cross-validation
model = DecisionTreeClassifier(random_state=42)
cv_scores = cross_val_score(model, X, y, cv=5)

print("Cross-validation scores:", cv_scores)
print("Mean CV score:", np.mean(cv_scores))
print("Standard deviation of CV scores:", np.std(cv_scores))
```

```
Cross-validation scores: [0.6  0.75 0.5  0.25 0.25]
Mean CV score: 0.47000000000000003
Standard deviation of CV scores: 0.19646882704388502
```

The average cross-validation score of 0.47 iIBNcates that the model performance is not very robust across different subsets of the data. The high standard deviation (0.196) iIBNcates significant variability in performance across layers.

Analysis of Result




```
sdmiskin: 0.7973
sdpidana: 0.1156
sdkonflik: 0.0500
sdkriminal: 0.0371
Model training and feature importance analysis completed.
```

Based on the results of the decision tree analysis, it can be seen the influence of variables on the IBN value, namely: sdmmiskin of 0.7973, sdcriminal of 0.1156, sdconflict of 0.0500, and sdcriminal of 0.0371. So it can be concluded that the variable 'sdmiskin' has the greatest influence on IBN followed by 'sdcriminal', 'sdconflict', and 'sdcriminal'.

The results of the analysis using Google Collaboratory and the python programming language found that the variable 'sdmiskin' which measures the level of poverty or the availability of human resources in poor areas has the greatest influence on IBN. This shows that poverty or economic inability can affect the level of awareness and participation in national defense. Areas with high poverty rates may have difficulty in providing the resources or training needed to increase the National Defense Index.

To increase IBN, focusing on poverty reduction and improving social welfare can be an important step. Programs that improve economic conditions and access to education and training can contribute positively to increasing awareness and participation in national defense.

Model Optimization

```
[ ] # Optimize the Decision Tree model using GridSearchCV to find the best hyperparameters
    from sklearn.model_selection import GridSearchCV

    # Define the parameter grid
    param_grid = {
        'max_depth': [3, 5, 7, 10],
        'min_samples_split': [2, 5, 10],
        'min_samples_leaf': [1, 2, 4],
        'criterion': ['gini', 'entropy']
    }

    # Initialize the Decision Tree Classifier
    dt = DecisionTreeClassifier(random_state=42)

    # Initialize GridSearchCV
    grid_search = GridSearchCV(estimator=dt, param_grid=param_grid, cv=5, n_jobs=-1, verbose=1)

    # Fit GridSearchCV
    grid_search.fit(X_train, y_train)

    # Get the best parameters and best score
    best_params = grid_search.best_params_
    best_score = grid_search.best_score_

    print('Best Parameters:', best_params)
    print('Best Score:', best_score)

    Fitting 5 folds for each of 72 candidates, totalling 360 fits
    /usr/local/lib/python3.10/dist-packages/sklearn/model_selection/_split.py:737: UserWarning: The least
      warnings.warn(
    Best Parameters: {'criterion': 'gini', 'max_depth': 5, 'min_samples_leaf': 2, 'min_samples_split': 2}
    Best Score: 0.5499999999999999
```

After optimizing the model by providing max depth, min sample, and criteria, the best measurement was obtained on the Gini criteria with a max depth of 5, with a min number of leaves of 2 and a min number of stems of 2. From these measurements, the best result was obtained at 0.549999.

Before optimization, the model showed moderate performance with an accuracy of 0.6, precision of 0.8, and recall of 0.6. This indicates that the model has good positive predictive value, with sufficient ability to identify positive categories and avoid misclassification. However, despite having relatively high precision, lower accuracy and recall indicate that there is room for improvement in terms of model accuracy and consistency. After optimization, the model showed improvement in terms of performance measurements based on the Gini criterion with parameters max depth of 5, min number of leaves of 2, and min number of stems of 2. The best result obtained after optimization was 0.55. This indicates significant improvement in the model, with parameter adjustments that are more appropriate for the data, resulting in more optimal and representative performance in terms of classification.

CONCLUSION

From the performance of the model, the following analysis conclusions are obtained:

- The variable 'sdmiskin' (sdmiskin = 0.7973) has the greatest influence on IBN. This shows that this variable is very dominant in influencing the IBN value.
- The variable 'sdcriminal' (sdcriminal = 0.1156) also has a significant influence but is much smaller than 'sdmiskin'.
- The variable 'sdkonflik' (sdkonflik = 0.0500) has an even smaller influence on IBN.
- The variable 'sdkriminal' (sdkriminal = 0.0371) has the lowest influence on the IBN value Chapter V is a summary of the overall research analysis. It can be expressed in either a sentence or a data value of parameters.All hypertext links and bookmark sections will be deleted. If the paper needs to refer to an email address or URL in the article, the complete address or URL must be typed in plain font.

RECOMMENDATION

Then the following interpretations can be given:

1. The Influence of 'sdmiskin' on IBN:

   The variable 'sdmiskin', which measures the level of poverty or the availability of human resources in poor areas, has the greatest influence on IBN. This shows that poverty or economic inability can affect the level of awareness and participation in defeIBNng the country. Areas with high poverty rates may struggle to provide the resources or training needed to improve the National Defense Index. To improve IBN, focusing on poverty reduction and improving social welfare can be important steps. Programs that improve economic coIBNtions and access to education and training can contribute positively to increasing awareness and participation in national defense.
2. Effect of 'criminal crime' on IBN:

   The variable 'criminal crime', which relates to the level of crime or criminality in an area, has a smaller effect compared to poverty but is still significant. This suggests that the presence or level of crime can affect IBN, perhaps by creating a less safe or stable environment. Reducing crime rates and improving law enforcement and public security systems can help create coIBNtions that are more conducive to increasing IBN. Programs that improve public security can also contribute to building a spirit of national defense.

3. Effect of 'sdkonflik' and 'sdkriminal' on IBN:
   The variables 'sdkonflik' (conflict) and 'sdkriminal' (crime) have a smaller effect on IBN compared to poverty and crime. This may iIBNcate that although conflict and crime affect the social environment, their impact on IBN values is not as great as poverty. Although the effect is smaller, addressing local conflict and crime issues is still important. Providing conflict prevention programs and improving law enforcement and social reforms can make additional contributions to supporting the improvement of IBN.

FURTHER STUDY

To provide deeper insights and wider applications, more study is invited to explore this area further, addressing any potential limits and broadening the scope of analysis.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Berezovskaya, F., Bilionis, I., Bordes, A., ... & Zinn, M. (2016). TensorFlow: Large-scale machine learning on heterogeneous systems. arXiv preprint arXiv:1603.04467.

Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. Belmont, CA: Wadsworth Publishing Company.

Kementerian Pertahanan Republik Indonesia. (2021). Indeks Bela Negara: Mengukur Kesiapsiagaan Warga Negara. Jakarta: Kemhan RI.

Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press.

Google. (2023). Google Colaboratory. Retrieved from https://colab.research.google.com

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Pratiwi, F.E., & Zain, I. (2014). Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara.

Rahayu, Siti. (2021). Penguatan Kesadaran Bela Negara Pada Remaja Milenial Menuju Indonesia Emas. PEDAGOGIKA. 12. 134-151. 10.37411/pedagogika.v12i2.711.

Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers-a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 35(4), 476-487.

Suryanto, A. (2018). Implementasi bela negara untuk mewujudkan nasionalisme. Jurnal Pertahanan & Bela Negara, 8(1), 1-18.

Timofeev, R. (2004). Classification and regression trees (CART) theory and applications. Humboldt University, Berlin.

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.