International Journal of Computer, Network Security and Information System (IJCONSIST)
Vol. 5 No. 1, September 2023

Implementation of the LightGBM Algorithm for Predicting the Nutritional Status of Infants and Toddlers in Doko Village, Kediri Regency

Muhammad Thoriqulhaq
Sains Data, Computer Science Faculty
Universitas Pembangunan Nasional Jawa Timur
Surabaya, Jawa Timur, Indonesia
21083010094@student.upnjatim.ac.id

Mohammad Idhom
Sains Data, Computer Science Faculty
Universitas Pembangunan Nasional Jawa Timur
Surabaya, Jawa Timur, Indonesia
idhom@upnjatim.ac.id

Kartika Maulida
Sains Data, Computer Science Faculty
Universitas Pembangunan Nasional Jawa Timur
Surabaya, Jawa Timur, Indonesia
kartika.maulida.ds@upnjatim.ac.id

***Abstract*— This study aims to implement a nutritional status index for infants and toddlers in Doko Village, Kediri Regency, using the LightGBM algorithm. Child health issues in Indonesia, particularly stunting, are a serious concern due to chronic malnutrition, recurring infections, and insufficient psychological stimulation during early developmental stages. Doko Village was selected as the research location due to significant challenges related to child nutrition in the area. The LightGBM algorithm was chosen for its ability to process large and imbalanced datasets while providing accurate predictions. The data used in this study comes from weight and height measurements of children at the local Posyandu. The main objective of this research is to develop a predictive model that can help healthcare workers identify children at risk of malnutrition, enabling more precise interventions. Additionally, this study developed a web-based application to monitor nutritional status in real-time, which is expected to improve the quality of life for children in Doko Village and nearby areas facing similar challenges.**

Keywords—Nutritional Status; Infants and Toddlers; LightGBM; Posyandu; Nutrition Prediction; Doko Village; Monitoring Application.

© 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution ShareAlike (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/).

I. INTRODUCTION

Early childhood malnutrition continues to be a serious public health issue worldwide, particularly in developing countries. In Indonesia, the prevalence of stunting and other forms of undernutrition among children under five years of age remains high, despite numerous government and community-level interventions. The World Health Organization (WHO) emphasizes the importance of monitoring anthropometric indicators such as weight-for-age (BB/U), height-for-age (TB/U), and weight-for-height (BB/TB) to assess nutritional status and guide interventions [1]. In rural areas such as Doko Village, located in Ngasem Sub-district of Kediri Regency, East Java, the challenges of monitoring and improving child nutrition are exacerbated by limited access to health facilities and professional nutritionists. Despite initiatives such as the Posyandu program—a community-based healthcare service for mothers and children—real-time data processing and analysis for identifying children at nutritional risk remain underdeveloped [2]. Consequently, there is a critical need for data-driven systems that can assist health workers in early detection of malnutrition. The Indonesian Ministry of Health has adopted the WHO 2020 anthropometric standards as the national reference for child growth monitoring [3]. However, interpreting this data manually poses a significant burden on health personnel, especially during monthly weighing sessions at Posyandu. A system that integrates computational intelligence with anthropometric data analysis can greatly enhance the speed and accuracy of nutritional assessments.

In response to this need, machine learning (ML) techniques offer promising solutions. ML algorithms are capable of learning patterns from historical data and making accurate predictions on new, unseen data. One such algorithm is Light Gradient Boosting Machine (LightGBM), which is well-known for its high efficiency and performance in both classification and regression problems, particularly on structured tabular datasets [4]. LightGBM stands out due to its use of leaf-wise tree growth strategy, which allows deeper and more complex trees to be constructed compared to traditional level-wise approaches. This capability makes it ideal for detecting subtle patterns in data, such as the relationships between age, gender, body weight, and height in nutritional classification tasks [5]. In the context of child nutrition, LightGBM can be trained to recognize patterns associated with different nutritional statuses. This research aims to build a classification model using LightGBM to predict children's nutritional status based on Weight-for-Age (BB/U), using a dataset collected from Doko Village that includes key anthropometric indicators. The study limits the classification task to three categories only—Normal, Underweight, and Overweight—to align with WHO standards and simplify the model output for field application [6]. The proposed model is developed and evaluated using four different data splitting ratios (60:40, 70:30, 80:20, and 90:10) to determine the most effective balance between training and testing performance. By evaluating each model using metrics such as accuracy, precision, recall, F1-score, and confusion matrix, the research seeks to identify the best configuration for practical deployment in healthcare settings [7].

Moreover, the trained model is integrated into a Streamlit-based web application, designed with a simple user interface to support real-time nutritional status prediction. This approach enhances usability for non-technical users such as village midwives, Posyandu volunteers, and parents, enabling them to receive instant classification results and tailored recommendations [8]. Previous studies have utilized machine learning to classify nutritional status, but few have focused specifically on rural contexts with limited infrastructure. This study fills that gap by providing a localized solution based on real-world data, which is both scalable and interpretable [9]. The ability to automate and personalize nutritional monitoring supports Indonesia's broader goals for reducing stunting prevalence to below 14% by 2024, as stated in the National Medium-Term Development Plan [10]. In summary, this research presents a novel integration of LightGBM classification and web-based visualization to address malnutrition among children under five in rural Indonesia. The system not only automates nutritional assessments but also contributes to the national agenda for improving child health outcomes.

II. RESEARCH METHOD

The nutritional status of children under five years old is widely recognized as a critical indicator of a nation's public health and development. According to WHO standards, children are assessed using anthropometric indices such as weight-for-age (BB/U), height-for-age (TB/U), and weight-for-height (BB/TB), each serving a unique purpose in identifying malnutrition [1]. These indicators are particularly important during the first five years of life, a period characterized by rapid physical, cognitive, and emotional development. Several studies have identified that malnutrition, including stunting, wasting, and underweight, can severely affect a child's long-term development and productivity. The Indonesian Ministry of Health reported that in 2022, the prevalence of stunting in children under five was still above 20%, exceeding the national target [2]. Various government programs, such as Posyandu and Gerakan Nasional Percepatan Perbaikan Gizi, aim to reduce this figure through routine monitoring and community-based interventions.

Anthropometric measurements are the most commonly used tools to determine nutritional status due to their simplicity and cost-effectiveness. These include weight, height, and age, which are compared against reference growth charts to determine whether a child's development is on track [3]. However, traditional methods of interpreting these measurements require manual plotting and analysis, which are prone to human error and often inefficient in high-volume field settings. Recent advances in data science and machine learning have opened new avenues for automating and improving the accuracy of nutritional status classification. Machine learning (ML) is a branch of artificial intelligence that allows systems to learn from historical data and make predictions or decisions without being explicitly programmed [4]. In the health sector, ML has been successfully applied in predicting disease outcomes, classifying patient risks, and assisting in medical diagnoses.

Among various ML algorithms, ensemble methods such as Gradient Boosting have gained popularity for their high accuracy in classification tasks. LightGBM (Light Gradient Boosting Machine), developed by Microsoft, is a gradient boosting framework that uses tree-based learning algorithms. It is known for its speed and efficiency, particularly on large datasets with many features [5]. LightGBM uses a leaf-wise growth strategy and supports techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce training time and memory usage. Previous studies have shown the successful implementation of LightGBM in various domains, including finance, energy, and healthcare. For instance, in medical research, LightGBM has been used to predict diabetes, heart disease, and mortality risk with considerable accuracy [6]. At the *t*-th iteration, the predicted value for the *i*-th sample is denoted as:

$$y_i^{(t)} = y_i^{(t-1)} + f_i\,(x_i)$$

Where $y_i^{(t)}$ is the prediction result at iteration t, $y_i^{(t-1)}$ is the result from the previous iteration, and $f_i(x_i)$ is the new model built at the current iteration. This concept reflects the principle of boosting, which aims to correct the errors from the previous iteration. Unlike the level-wise approach that expands the tree evenly at each level, LightGBM adopts a more flexible leaf-wise approach. Leaf-wise growth adds branches only to the leaf node with the greatest potential to reduce loss. This enables LightGBM to achieve higher accuracy with the same tree depth. Its ability to handle missing values and categorical variables without the need for heavy preprocessing makes it suitable for real-world health data applications. In the context of child nutrition, a few studies have adopted ML techniques to classify children into different nutritional status categories. For example, Putra et al. used decision trees and random forest algorithms to classify stunting status based on demographic and nutritional factors [4]. However, these models often face limitations in accuracy and generalization, especially when applied to rural datasets with high variability and noise.

Confusion matrix-based performance evaluation is commonly employed to assess classification models in healthcare-related research. Metrics such as accuracy, precision, recall, and F1-score provide a comprehensive picture of the model's ability to distinguish between different classes, especially in imbalanced datasets [8]. For multiclass problems like BB/U classification, these metrics are essential to ensure that minority classes such as “Underweight” or “Overweight” are accurately predicted. Furthermore, web-based systems such as Streamlit have enabled the deployment of machine learning models in a user-friendly environment. Streamlit simplifies the development of interactive applications by allowing developers to combine Python code and frontend elements within a single framework [9]. For public health applications, this means that complex predictive models can be made accessible to non-technical users, such as midwives or community health volunteers. Given these developments, integrating LightGBM with a web interface like Streamlit can serve as a practical solution for real-time nutritional assessment in rural communities. By leveraging accurate classification models and interactive interfaces, health workers can identify at-risk children earlier and take necessary interventions, thereby supporting national efforts to reduce malnutrition rates [10].

III. METHODOLOGY

This research applies a quantitative approach using supervised machine learning to classify the nutritional status of children under five years old based on the BB/U (Weight-for-Age) indicator. The dataset used contains anthropometric data such as age (in months), weight (kg), height (cm), and gender. These features are used as input variables (features), while the BB/U classification (Normal, Underweight, Overweight) is used as the target output (label). The LightGBM algorithm is chosen due to its speed and ability to handle large-scale data efficiently. The research was conducted in Desa Doko, Ngasem Subdistrict, Kediri Regency, where community health centers (Posyandu) routinely collect anthropometric data. Data collection was based on real measurements conducted over a 10-month period from June 2024 to March 2025. Each child's data was recorded monthly, allowing the dataset to include temporal changes in weight and height. This data is then cleaned, normalized, and converted into a format suitable for machine learning modeling.

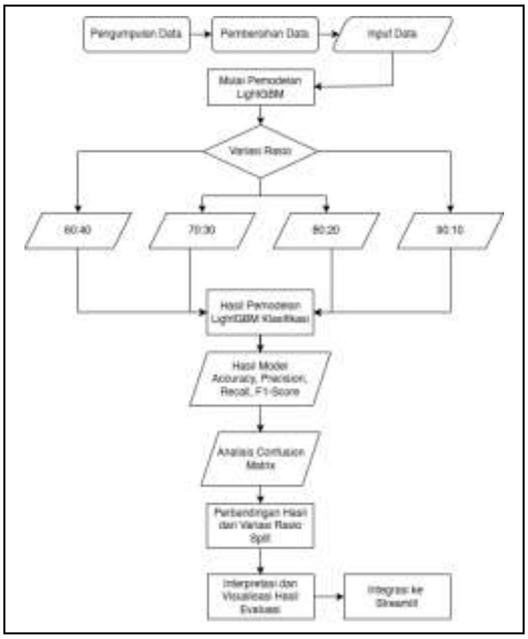

Figure 1. Research Flowchart

The data preprocessing phase begins with handling missing values and converting categorical variables into numerical form. For example, gender is encoded as 1 for male and 0 for female. Only numeric features (age, weight, height, gender) are retained for model input. The BB/U column is retained as the classification target, while other nutritional indicators (TB/U, BB/TB, and their z-scores) are excluded from the feature set to maintain focus and model simplicity. After preprocessing, the dataset is divided into training and testing subsets using four different ratios: 60:40, 70:30, 80:20, and 90:10. This step ensures the model is evaluated under various data distribution scenarios and is not overfitted to a particular subset. The division is performed randomly but consistently using a fixed random seed (e.g., random_state=42) for reproducibility.

Model training is carried out using the LightGBM classifier. This algorithm builds decision trees sequentially, using a gradient boosting technique to reduce prediction error at each iteration. Unlike traditional algorithms that grow trees level-wise, LightGBM uses a leaf-wise growth strategy, which selects the leaf with the maximum loss reduction to split. This approach increases accuracy but may lead to overfitting if not controlled with parameters like max_depth or min_data_in_leaf. Hyperparameter tuning is not explicitly performed in this study to maintain a consistent training process across data ratios. However, default parameters such as learning_rate, n_estimators, and verbosity are applied, which already produce reliable performance based on prior research. The evaluation uses key metrics including accuracy, precision, recall, and F1-score to assess the classification results.

Each trained model is evaluated on its respective test set using a confusion matrix, which allows visualization of prediction accuracy for each class label. This helps identify which classes are being correctly or incorrectly predicted. The matrix is displayed for each data ratio, allowing the researcher to compare the effect of training data proportion on prediction performance. Among the four tested ratios, the 80:20 configuration produced the highest accuracy and F1-score. Therefore, this model is selected as the final model and saved using the pickle library for future use. The saved model (lgbm_terlatih_80.pkl) is then integrated into a web-based application using Streamlit to allow real-time nutritional status prediction.

The Streamlit application allows users, particularly healthcare workers, to input data such as gender, age, weight, and height. The system then displays the predicted BB/U status, along with additional information about ideal growth indicators and health recommendations based on the child's nutritional classification. This implementation aims to simplify access to AI-based decision support tools in rural and community health settings. The final output of this research is not only a predictive model but also an interactive application that facilitates the early detection of nutritional problems. Through this integration, the research contributes to public health monitoring efforts by offering a practical tool that supports interventions at the grassroots level.

IV. Result and Discussion

The classification model was implemented using the LightGBM algorithm with a focus on the BB/U (Weight-for-Age) indicator. After data preprocessing and feature selection, the dataset was split into training and testing subsets using four different ratios: 60:40, 70:30, 80:20, and 90:10. The features used for training include gender, age in months, weight, and height. The label or target variable is the BB/U status with three classes: "Normal," "Underweight," and "Overweight." For each split ratio, a LightGBM model was trained and evaluated. The classification performance was measured using standard evaluation metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive overview of how well the model performs in distinguishing between the three BB/U categories. In addition, confusion matrices were used to examine the number of correct and incorrect predictions for each class.

At a 60:40 ratio, the model achieved an accuracy of 96%, but the precision, recall, and F1-score were relatively lower at 0.68. This indicates that while the model was generally accurate, it struggled with balance across the three classes due to less training data. A larger test set may also introduce more variance that the model has not been trained to handle. At the 70:30 ratio, the model's accuracy improved to 97%, with precision, recall, and F1-score rising to 0.94. This configuration provided a more balanced performance, demonstrating better generalization from the training data to unseen test data. The results suggest that increasing the training set size enhances the model's ability to learn meaningful patterns. The best performance was obtained at the 80:20 ratio, where the model achieved an accuracy of 97%, precision of 0.94, recall of 0.95, and an F1-score of 0.94. This split offered an optimal balance between training depth and test set variability, indicating that the model could effectively differentiate between the three BB/U categories without overfitting. Interestingly, the 90:10 ratio also yielded high accuracy (97%), but due to the limited test data size, the performance evaluation metrics were slightly more variable. While the training performance was strong, a small test set may not provide a reliable measure of real-world performance. Thus, although this split is efficient in training, it may not offer the best validation scenario.

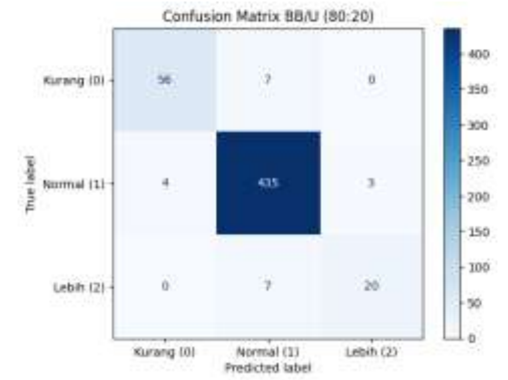

Figure 2 Confusion Matrix BB/U (80:20)

Based on these findings, the 80:20 ratio in Figure 2 was selected as the final model configuration. This model was then saved using the Python pickle library to allow for deployment without retraining. Saving the trained model enables integration into practical applications and makes it accessible for use in real-time prediction systems. The saved model was deployed in a Streamlit-based web application. This interface allows users to input the child's age, gender, weight, and height to obtain an immediate BB/U classification. The application also provides additional information based on age category and gender, including ideal weight and height, along with personalized nutritional advice. To improve user experience, the application displays a table of ideal values for different age groups and gender categories. The table is divided into male and female references, derived from pediatric growth standards, providing a comparison between the user's input and recommended anthropometric ranges.

This contextual feedback aids caregivers in understanding their child's health status more comprehensively. Visual elements such as colored notifications and interactive predictions enhance usability for non-technical users, such as health workers or caregivers at Posyandu (integrated health posts). This user-centric design ensures that the system is not only technically effective but also accessible and practical for frontline use. In conclusion, the results show that LightGBM can successfully classify the nutritional status of children based on BB/U indicators. With the deployment of this model into a web-based interface, the research bridges the gap between machine learning and public health service, providing a digital solution that can assist in early detection and intervention in child malnutrition.

V. Conclusion

This research implemented the Light Gradient Boosting Machine (LightGBM) algorithm to classify the nutritional status of children under five based on the Weight-for-Age (BB/U) indicator. The system utilized anthropometric features, including gender, age in months, body weight, and height. Four different training-to-testing ratios (60:40, 70:30, 80:20, and 90:10) were tested to determine the most optimal model performance. Among these, the 80:20 ratio yielded the best results, with accuracy, precision, recall, and F1-score all reaching above 94%. The model was saved and deployed in a web-based application using Streamlit, enabling real-time predictions of children's nutritional status based on input data.

The final system provides not only BB/U predictions but also supportive information, including ideal anthropometric benchmarks per age and gender, as well as actionable nutritional advice. This empowers health workers and caregivers to monitor and address early signs of malnutrition. The application's intuitive design ensures accessibility for users in community health centers, especially Posyandu, and supports the broader public health objective of early stunting prevention. However, this study is limited to a single target classification (BB/U). Future research could extend the classification to include other indices such as TB/U (Height-for-Age) and BB/TB (Weight-for-Height), or even perform multi-output classification. In addition, hyperparameter tuning, which was not deeply explored in this study, could further improve model accuracy and generalizability.

Another potential improvement lies in integrating broader datasets across regions and time periods to enhance the robustness and applicability of the model on a national scale. Incorporating longitudinal data could also enable not just classification, but forecasting of nutritional trends among children over time. In conclusion, this research successfully demonstrates the effectiveness of LightGBM in addressing public health challenges through digital solutions. The integration of machine learning with a web interface opens up new possibilities for preventive nutrition services. It is hoped that the model and application developed here can serve as a practical tool in the early identification and management of nutritional problems among Indonesian children.

References

[1] Ministry of Health of the Republic of Indonesia. (2020). *Regulation of the Minister of Health No. 2 of 2020 on Child Anthropometric Standards*. Jakarta: Ministry of Health.

[2] Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 4023–4031.

[3] Siloam Hospitals. (2022). *Ideal Weight Standards for Toddlers*.

[4] Andreev, V. (2022). *Gradient Boosting Algorithms Explained: LightGBM, XGBoost, and CatBoost*.

[5] WHO. (2022). *Child Growth Standards*. World Health Organization.

[6] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.

[7] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 3146–3154).

[8] Hafid, F., Ansar, A., Nasrul, N., Ramadhan, K., & Ardiansyah, A. (2023). Prevalence and Determinants Stunting Among Children Under Two Years in Indonesian District. Poltekita: Jurnal Ilmu Kesehatan, 16(4), 553-563.

[9] Prakoso, A. D., Azmiardi, A., Febriani, G. A., & Anulus, A., "Studi case control: pemantauan pertumbuhan, pemberian makan dan hubungannya dengan stunting pada anak panti asuhan di kota semarang: studi case control," *Jurnal Ilmu Kesehatan Bhakti Husada: Health Sciences Journal,* pp. 160-172, 2021.

[10] Suhartatik, S., & Al Faiqoh, Z, "Peran kader posyandu dalam pemantauan status gizi balita: Literature review," *Journal of Health, Education and Literacy (J-Healt),* pp. 19-25, 2022.

[11] Duran, F., Wijaya, F., Hulu, Y. R., Harahap, M., & Prabowo, A, "PERBANDINGAN KINERJA ALGORITMA RANDOM FOREST CLASSIFIER DAN LIGHTGBM CLASSIFIER UNTUK PREDIKSI PENYAKIT JANTUNG," Data Sciences Indonesia (DSI), pp. 98-103, 2023.

[12] F. Riza, "Analisis dan prediksi data penjualan menggunakan machine learning dengan pendekatan ilmu data," Data Sciences Indonesia (DSI), pp. 62-68, 2021.

[13] Toharudin, T., Caraka, R. E., Pratiwi, I. R., Kim, Y., Gio, P. U., Sakti, A. D., ... & Pardamean, B, "Boosting Algorithm to Handle Unbalanced Classification of PM 2.5 Concentration Levels by Observing Meteorological Parameters in Jakarta-Indonesia Using AdaBoost, XGBoost, CatBoost, and LightGBM. IEEE Access, 11, 35680-35696.," 2023.

[14] Ivanoti, V. I., Triyono, G., & Utami, D. P, "Decision Support System for Predicting Employee Leave Using the Light Gradient Boosting Machine (Lightgbm) and K-Means Algorithm," Jurnal Teknik Informatika (JUTIF), pp. 657-667, 2023.

[15] Setyorini, C., & Lieskusumastuti, A. D., "Gambaran Status Gizi Bayi Dan Balita Pada Masa Covid-19 Di Kalurahan Jetis. Avicenna," *Journal of Health Research,* 2021.

[16] I. Aristiyani, "Dampak Status Ekonomi pada Status Gizi Balita," *Jurnal Keperawatan Widya Gantari Indonesia,* 2023.

[17] Tanuwijaya, R. R., Djati, W. P. S. T., & Manggabarani, S, "Hubungan pengetahuan pemberian makanan bayi dan anak (PMBA) ibu terhadap status gizi pada balita," *Jurnal Dunia Gizi,* pp. 74-79, 2020.

[18] A. Subayu, "Penerapan Metode K-Means untuk Analisis Stunting Gizi pada Balita: Systematic Review," *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi,,* pp. 42-50, 2022.

[19] W. M. P. Dhuhita, "Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita," Jurnal Informatika, pp. 160-174, 2015.

[20] Utami, M. T.,& Widiyaningsih, E. N., “Status berat badan lahir rendah dengan kejadian stunting pada balita di Desa Sukadadi, Lampung. Ilmu Gizi Indonesia,” pp. 95-104, 2023.

[21] Helmyati, S., Dipo, D. P., Adiwibowo, I. R., Wigati, M., Safika, E. L., Hariawan, M. H., ... & Trisnantoro, L. (2021). Monitoring continuity of maternal and child health services, Indonesia. Bulletin of the World Health Organization, 100(2), 144.

[22] Motee, A., & Jeewon, R. (2014). Importance of exclusive breastfeeding and complementary feeding among infants. Current Research in Nutrition and Food Science Journal, 2(2), 56-72.

[23] Riyantoko, P. A., Fahrudin, T. M., & Hindrayani, K. M. (2021, October). Analisis Sederhana Pada Kualitas Air Minum Berdasarkan Akurasi Model Klasifikasi Dengan Menggunakan Lucifer Machine Learning. In PROSIDING SEMINAR NASIONAL SAINS DATA (Vol. 1, No. 01, pp. 12-18).

[24] Tse, A. D. P., Suprojo, A., & Adiwidjaja, I. (2017). Peran kader posyandu terhadap pembangunan kesehatan masyarakat. Jurnal Ilmu Sosial dan Ilmu Politik (JISIP), 6(1).