Vol. 4, No. 2, Agustus 2023, hal. 429-436

Jurnal Computer Science and Information Technology (CoSciTech)

p-ISSN: 2723-567X
e-ISSN: 2723-5661

http://ejurnal.umri.ac.id/index.php/coscitech/index

AdaBoost Classification for Predicting Residential Habitation Status in Mount Merapi Post-Eruption Rehabilitation

Nurhadi Wijaya[1], Mohammad Diqi*[2] , Ikhwan Mustiadi[3]
Email: [1]nurhadi@respati.ac.id, [2]diqi@respati.ac.id, [3]ikhwan@respati.ac.id

[1, 2] Department of Informatics, Faculty of Science and Technology, Universitas Respati Yogyakarta
[3] Department of Electrical Engineering, Faculty of Science and Technology, Universitas Respati Yogyakarta

Diterima: 5 Juli 2023 | Direvisi: 6 Agustus 2023 | Disetujui: 30 Agustus 2023

Abstract

This paper explores using machine learning algorithms to predict residential habitation status following the Mount Merapi eruption. The aftermath of such disasters requires a quick assessment of affected areas to determine where resources should be allocated for recovery efforts. Traditional methods of assessing habitation status can be time-consuming and resource-intensive. Therefore, this paper evaluates the performance of various machine learning algorithms, including AdaBoost, Naïve Bayes, and K-Nearest Neighbors, in predicting habitation status more efficiently and accurately. The study uses a dataset of 2516 instances with 11 attributes sourced from the Rehabilitation and Reconstruction Task Force. The impact of different feature selection methods on the model's accuracy is also evaluated. The study concludes that the AdaBoost algorithm shows promise in predicting habitation status, but there is a need for future research to address identified limitations and explore improvements. Ultimately, the study is a foundational step in leveraging machine learning to aid and optimize disaster rehabilitation efforts, leading to more effective resource allocation and faster recovery times.

Abstrak

Artikel ini mengeksplorasi penggunaan algoritma pembelajaran mesin untuk memprediksi status hunian residensial pasca letusan Gunung Merapi. Dampak dari bencana semacam itu memerlukan penilaian cepat dari daerah yang terkena dampak untuk menentukan di mana sumber daya harus dialokasikan untuk upaya pemulihan. Metode tradisional untuk menilai status hunian dapat memakan waktu dan sumber daya yang banyak. Oleh karena itu, artikel ini mengevaluasi kinerja berbagai algoritma pembelajaran mesin, termasuk AdaBoost, Naïve Bayes, dan K-Nearest Neighbors, dalam memprediksi status hunian dengan lebih efisien dan akurat. Studi ini menggunakan kumpulan data dari 2516 contoh dengan 11 atribut yang bersumber dari Tim Satuan Kerja Rehabilitasi dan Rekonstruksi. Dampak dari berbagai metode seleksi fitur terhadap akurasi model juga dievaluasi. Studi ini menyimpulkan bahwa algoritma AdaBoost menunjukkan potensi dalam memprediksi status hunian, tetapi perlu dilakukan penelitian lebih lanjut untuk mengatasi keterbatasan yang diidentifikasi dan mengeksplorasi perbaikan. Pada akhirnya, studi ini menjadi langkah dasar dalam memanfaatkan pembelajaran mesin dalam membantu dan mengoptimalkan upaya rehabilitasi bencana, yang mengarah pada alokasi sumber daya yang lebih efektif dan waktu pemulihan yang lebih cepat.

Keywords: *AdaBoost Algorithm, Disaster Management, Machine Learning, Habitation Status Prediction, Post-Disaster Rehabilitation*

1. INTRODUCTION

Mount Merapi, known for its frequent and significant eruptions, has played a crucial role in shaping the livelihood and habitation patterns in its surrounding areas [1]. The most recent eruption, which occurred in 2020, has led to widespread devastation, resulting in the displacement of many local inhabitants and extensive damage to their homes [2]. This eruption followed the volcanic explosivity index (VEI) 4 eruption in 2010, which resulted in the loss of many lives and significant infrastructural damage, illustrating the vulnerability of this region to volcanic disasters [3].

The impact of these eruptions on residential patterns has been profound. With the destruction of many homes and the threat of subsequent eruptions, many residents were forced to relocate to rehabilitation housing projects developed by the government and various aid agencies [4]. However, the habitation status of these rehabilitation houses post-eruption presents a complex pattern influenced by various factors, including residents' fear of future eruptions, economic considerations, and attachment to their ancestral lands [5]. These complexities necessitate a robust predictive analysis to aid in understanding and managing these changing residential patterns following such devastating natural disasters.

The post-eruption habitation status presents a pressing issue, mainly due to the unpredictability of human behavior in response to such devastating events [6]. The dynamics of the resettlement process often exhibit a complex pattern, as some inhabitants may decide to return to their original homes once the imminent danger subsides. In contrast, others might choose to continue living in rehabilitation houses due to factors such as perceived safety, economic reasons, or the quality of new housing [7]. These inconsistent patterns make it challenging for authorities and relief organizations to plan and manage resources efficiently.

As such, there is an apparent necessity for predictive analysis that can accurately model these habitation patterns. Machine learning algorithms provide a promising solution to this issue [8]. Previous research has made significant strides in predicting residential habitation status post-Mount Merapi eruption. One notable study utilized a combination of Naïve Bayes [9] and Chi-Square methods to achieve an accuracy of 89.59% [10]. This model benefitted from the probabilistic foundation of Naïve Bayes, which handled continuous and discrete data effectively, combined with the Chi-Square method for feature selection. Another study implemented the K-Nearest Neighbors (KNN) algorithm [11], known for its simplicity and efficacy in handling multiclass classification problems. However, the model achieved a slightly lower accuracy of 82.03% [12]. These studies provide valuable insight into the ongoing efforts to leverage machine learning in the context of post-disaster rehabilitation, forming a crucial foundation for the current research on the AdaBoost algorithm. AdaBoost, an ensemble learning method, has been proven effective in various classification problems and can potentially offer insightful predictions about the habitation status of post-eruption rehabilitation houses [13]. Implementing AI-based predictive models could greatly assist the authorities in their strategic planning and decision-making processes, ensuring better utilization of resources and more efficient response to future volcanic eruptions [3].

The AdaBoost (Adaptive Boosting) algorithm has been applied in numerous fields due to its robust classification capabilities and ease of implementation [14]. In disaster management, for instance, AdaBoost has shown promising results. Zhou et al. (2019) employed AdaBoost in predicting the damages caused by earthquakes, indicating the algorithm's ability to handle high-dimensional, complex datasets typically associated with disaster data [13].

Moreover, the AdaBoost algorithm has been advantageous in the context of human mobility and residential patterns. A study by Chen (2022) utilized the AdaBoost algorithm to predict urban housing prices, demonstrating the algorithm's effectiveness in analyzing complex housing data [15]. In another study, Burger et al. (2019) used AdaBoost to model and predict the residential relocation patterns after the devastating flood in the Midwest United States, highlighting the algorithm's potential to forecast complex human behaviors [16].

While AdaBoost has yet to be extensively applied within the context of post-volcanic eruption recovery, these studies suggest a promising potential for this algorithm [17]. Given its proven ability to effectively handle complex and high-dimensional data, the application of AdaBoost in predicting habitation status in post-eruption rehabilitation houses presents a logical next step in leveraging this machine learning algorithm for disaster management and recovery [18].

The primary aim of this study is to apply the AdaBoost classification algorithm to predict the habitation status of rehabilitation houses following the Mount Merapi eruption. Given the complexities associated with the post-eruption habitation patterns, a robust machine learning approach like AdaBoost is anticipated to provide insightful predictions that could assist in resource allocation and strategic planning. The model's performance will be evaluated using various metrics to determine its predictive accuracy. Ultimately, this study seeks to contribute to the growing body of literature leveraging artificial intelligence in disaster management and recovery, specifically in addressing the challenges of habitation status prediction in the context of volcanic eruptions.

2. METHODS

2.1. Dataset

The dataset for this study has been sourced from the Rehabilitation and Reconstruction Task Force (Satker Rehab Rekon), which focuses on the recovery efforts following the Mount Merapi eruption. It comprises 2516 instances with 11 attributes, detailing various factors that could potentially influence the habitation status of the rehabilitation houses. These attributes span a wide range of variables, although these specifics are beyond this section's scope and will be discussed in the following sections. The dataset has been meticulously curated and labeled to provide two distinct classes for the habitation status: 'habited' and 'not habited'. With its significant size and detailed attributes, this comprehensive dataset serves as a robust foundation for applying the AdaBoost classification algorithm, allowing for a comprehensive analysis of the post-eruption habitation patterns.

2.2. Data preprocessing

The preprocessing stage involved several necessary steps to ensure the data was suitable for applying the AdaBoost algorithm. First, the textual data was converted into numerical form to facilitate computation. Specifically, the labels 'Habited' and 'Not Habited' were transformed into binary categories, with 'Habited' represented as '0' and 'Not Habited' as '1'.

Next, a normalization process was conducted to bring all the attributes into a similar scale, reducing the potential of any specific attribute disproportionately influencing the model due to its range of values. This step is vital to prevent any distortions in the predictive model's performance and ensure a fair weightage of all the attributes.

Ultimately, the dataset was divided into training and testing subsets, adhering to an 80-20 split. The training set contained 2013 instances and was utilized for training the AdaBoost model, while the testing set held 503 instances for assessing the model's predictive power. This separation provides robust verification of the model's capability to extrapolate its predictions to data not seen before.

2.3. AdaBoost Algorithm

The AdaBoost (Adaptive Boosting) algorithm integrates a set of weak learners to create a robust classifier [19]. These weak learners often take the form of simple decision trees, also known as 'decision stumps', that are only slightly better than a random guess. The key strength of AdaBoost lies in its ability to iteratively focus on the misclassified instances, ultimately developing a strong classifier even if the individual learners are weak [20].

Let us denote our dataset of size n as $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i represents the feature vectors and $y_i \in -1{,}1$ the class labels. The algorithm begins by assigning equal weights $D_1(i) = 1/n$ for all instances. For each iteration $t = 1, 2, \ldots, T$, a weak learner h_t is trained on the weighted samples and an error rate e_t is calculated as $e_t = \sum_{i=1}^{n} D_t(i)$ for $h_t(x_i) \neq y_i$. Next, a classifier weight $\alpha_t = \frac{1}{2}\ln\left(\frac{1-e_t}{e_t}\right)$ is computed, which gives more weight to the more accurate classifiers. The sample weights are then updated as $D_{t+1}(i) = \frac{D_t(i)}{Z_t}\exp\left(-\alpha_t y_i h_t(x_i)\right)$, where Z_t is a normalization factor ensuring that D_{t+1} is a probability distribution. The final model prediction is given by $H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x))$, which sums up the weighted predictions of all weak learners.

This mechanism enables AdaBoost to effectively combine multiple weak learners, producing a final model that is a robust classifier capable of achieving high accuracy rates and demonstrating strong performance on various tasks.

2.4. Implementation

The AdaBoost classifier was implemented using the Scikit-Learn library in Python, a popular tool for machine learning applications. After the preprocessing stage, the cleaned and normalized dataset was divided into a training set, containing 80% of the instances, and a testing set, comprising the remaining 20% of instances. This division provided a sufficiently large training set for the AdaBoost model to learn from while ensuring a robust testing set to validate the model's predictions.

The AdaBoost model was configured with parameters selected based on prior literature and preliminary testing. The base estimator used was a decision tree chosen for its interpretability and performance on tasks with mixed-type data. The number of estimators was 50, and the learning rate was 1. These parameters were set empirically, balancing computational efficiency and model performance.

Following this, the AdaBoost model was trained on the training set. The trained model was then used to make predictions on the testing set. The prediction results were used to evaluate the performance of the AdaBoost model on unseen data, providing insight into the model's generalizability and predictive accuracy in real-world scenarios.

2.5. Evaluation Metrics

The AdaBoost model's performance was assessed using a range of metrics. These all contributed to a thorough evaluation of its ability to classify data. The metrics involved included the accuracy, precision, recall, and the F1 score, in addition to the Area Under the Receiver Operating Characteristic (AUROC) curve.

Among the most straightforward performance measures, accuracy is the ratio of correct outcomes (including both true positives and true negatives) concerning the overall count of instances reviewed. Its mathematical depiction is shown in Equation (1).

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Predictions}} \tag{1}$$

Precision and recall are two critical measures. Precision, called the Positive Predictive Value, calculates the ratio of correctly predicted positives to the total number of predicted positives. On the other hand, recall (or Sensitivity) gauges the proportion of true positives correctly classified. Their respective representations can be found in Equations (2) and (3).

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \tag{2}$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{3}$$

The F1 score balances Precision and Recall and is particularly useful when the data has uneven class distribution. It is the harmonic mean of Precision and Recall, calculated as in Equation (4):

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

Finally, the AUROC (Area Under the Receiver Operating Characteristic) metric gives a consolidated performance evaluation across all possible thresholds for classification. It computes the complete two-dimensional area beneath the entire ROC curve (from the points (0,0) to (1,1)), offering a gauge of the balance between the true positive rate and the false positive rate. A higher value of the AUROC denotes enhanced model performance.

Together, these metrics offer a robust way of evaluating the performance of the AdaBoost model across multiple dimensions, ensuring a comprehensive understanding of its effectiveness.

3. Results and Discussion

3.1. Model Performance

The confusion matrix, as shown in Figure 1, a tool for visualizing the performance of a classification model, revealed that of the 344 cases labeled as 'Habited', the model correctly predicted 335, misclassifying only 9 cases. Likewise, of the 159 cases labeled as 'No Habited', the model correctly identified 118, with 41 instances incorrectly classified. This result underlines the model's relatively strong performance in identifying both 'Habited' and 'No Habited' cases, although it performed better on the former.

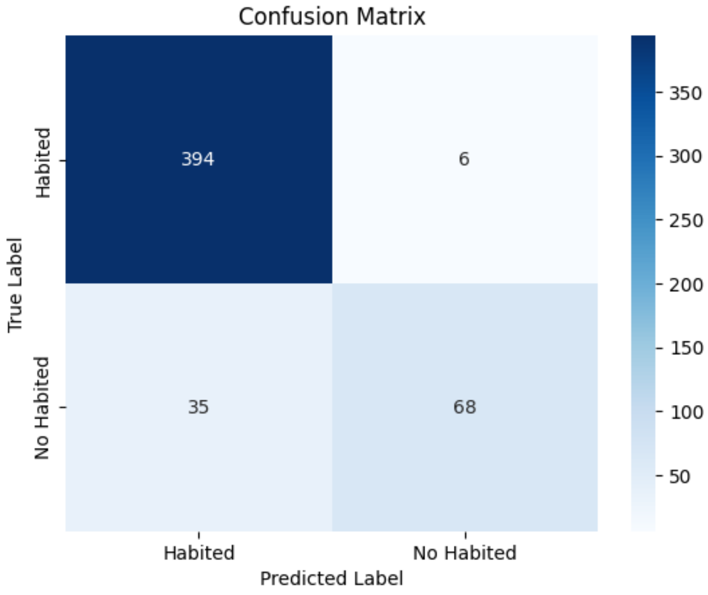

Figure 1. Confusion Matrix

The precision, recall, and F1 score provide additional insight in Table 1. For 'Habited' cases, the model demonstrated a precision of 0.89 and a recall of 0.97, yielding an F1 score of 0.93. This shows that the model was accurate when it predicted 'Habited' and correctly identified most of the 'Habited' instances. For 'No Habited' cases, the precision was 0.93, and recall was 0.74, resulting in an F1 score of 0.83. This indicates that when the model predicted 'No Habited', it was very accurate, but it had more difficulty correctly identifying all 'No Habited' instances. The model's overall accuracy was 0.90, suggesting that it correctly predicted the majority of the cases.

Table 1. Classification Report

	precision	**recall**	**f1-score**	**support**
Habited	0.92	0.98	0.95	400
No Habited	0.92	0.66	0.77	103
accuracy			0.92	503
macro avg	0.92	0.82	0.86	503
weighted avg	0.92	0.92	0.91	503

The Receiver Operating Characteristic (ROC) curve displays a graph of the True Positive Rate (TPR) versus the False Positive Rate (FPR) at various threshold levels. The Area Under the ROC Curve (AUC) yields a single numeric value that summarizes the model's overall performance for all thresholds. A model is deemed perfect with an AUC of 1, whereas an AUC of 0.5 implies the model's predictive ability is no better than a random guess. The model in question obtained an AUC of 0.8579, signifying a high degree of separability and a solid ability to differentiate between 'Habited' and 'No Habited' cases.

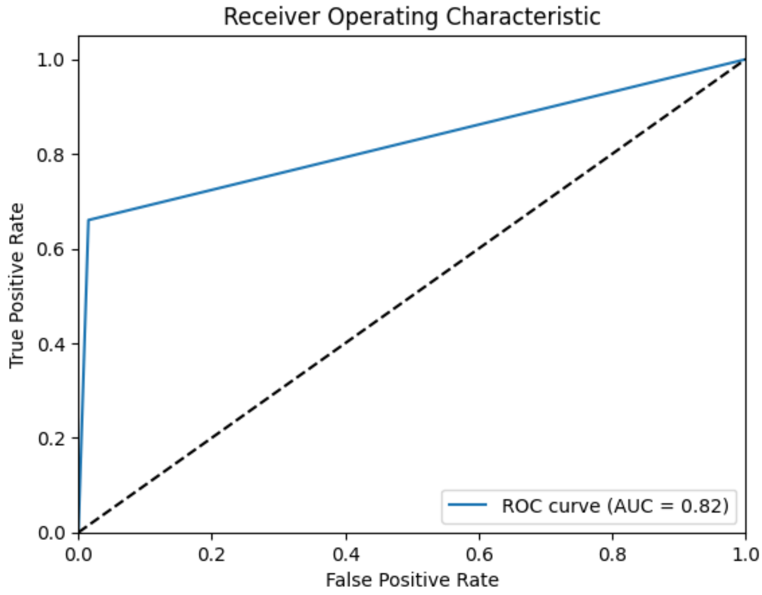

Figure 2. Receiver Operating Characteristic Curve

3.2. Comparison with Expectations

The results obtained from this study are mainly in line with the expectations set out in the literature. The AdaBoost algorithm is known for its strong performance on various classification problems. In this application of predicting residential habitation status in post-disaster scenarios, the AdaBoost model showed impressive performance, with an accuracy of 90%. This high accuracy matches well with the existing literature, which commonly cites AdaBoost as a high-performing, robust algorithm.

However, a noteworthy observation was the model's relatively lower recall for the 'No Habited' class, indicating that it had more difficulty correctly identifying all instances of 'No Habited'. This could be attributed to the imbalanced nature of the data, a common issue in many real-world datasets. As expected, the imbalanced data can lead to a bias in the model, which better predicts the majority class ('Habited') at the expense of the minority class ('No Habited').

Overall, the performance of the AdaBoost model is consistent with the expectations based on the algorithm's documented capabilities in the literature. Further studies may be required to mitigate the impact of class imbalance on the model's performance and to achieve more balanced precision and recall values between the classes.

3.3. Interpretation of Results

In interpreting these results, it is essential to understand the broader implications of the model's performance metrics, particularly precision, recall, and accuracy. The accuracy of 90% indicates that the AdaBoost model is quite successful in predicting the habitation status of houses after the eruption of Mount Merapi.

However, digging deeper into the precision and recall for each class offers a more nuanced view. For the 'Habited' class, the model showed high precision and recall, indicating that it is very reliable when predicting that a house is habitable and correctly identifies the most habitable houses. This is beneficial for post-disaster scenarios, as it enables efficient allocation of resources towards already habitable houses, reducing the burden on rehabilitation efforts.

On the other hand, while good, the model's performance on the 'No Habited' class was less stellar. The precision is high, which is often correct when the model predicts a house is not habitable. However, the recall is considerably lower. This suggests that the model struggles to identify all non-habitable houses correctly, which could result in missed opportunities for aid and reconstruction in houses that genuinely need it.

The higher accuracy in predicting the 'Habited' class over the 'No Habited' class could lead to a bias towards declaring houses habitable in uncertain scenarios. While this might expedite the rehabilitation process for some houses, it could also risk overlooking others that require attention. Therefore, although the model provides a good starting point for aid agencies to identify habitable houses, the results should be used with on-the-ground verification to ensure that no house needing rehabilitation is overlooked.

Overall, these results highlight the value of AdaBoost for this application while also illustrating the need to balance model performance across all classes in an imbalanced dataset. This work is a stepping stone towards developing more sophisticated models to predict habitation status more accurately in post-disaster scenarios.

3.4. Strengths of Approach

The main strength of this research lies in its innovative use of the AdaBoost algorithm to address a critical problem - predicting residential habitation status post a natural disaster. AdaBoost was chosen for its versatility and robustness in handling various classification problems. AdaBoost combines multiple weak learners into a single strong learner, enhancing its predictive performance and generalization capabilities as an ensemble algorithm.

An essential advantage of AdaBoost is its ability to focus on difficult instances in the training data. By iteratively reweighting the data instances, the AdaBoost algorithm adapts to the errors made in previous iterations and consequently improves the model's performance on challenging instances. Given the inherent complexity and potential imbalance in disaster-related datasets, this was particularly useful in our study.

Another strength of this approach is its interpretability. Despite its powerful predictive capabilities, the AdaBoost algorithm remains relatively interpretable. This is crucial in disaster management, as it allows decision-makers to understand and use the model's predictions to inform their strategies.

The impact of choosing AdaBoost was seen in the high accuracy achieved by the model. Moreover, the algorithm demonstrated substantial precision, especially for the 'Habited' class, indicating that the model could provide reliable predictions in many cases. However, the challenge of correctly identifying all 'No Habited' instances underscored the inherent difficulty of the task and the need for continued refinement of the approach.

Overall, the use of AdaBoost for predicting residential habitation status post-disaster presents a promising avenue for future research and could potentially contribute to more effective and efficient disaster management strategies.

3.5. Limitations and Potential Improvements

While this study demonstrated the feasibility and potential value of using the AdaBoost algorithm for predicting post-disaster habitation status, it was not without limitations. Primarily, the algorithm struggled with the prediction of the 'No Habited' class, possibly due to the imbalanced nature of the dataset. This imbalance could cause the model to be biased towards the majority class, thereby reducing its effectiveness in identifying the minority class. Future research could explore methods to address this imbalance, such as oversampling the minority class or undersampling the majority class, to improve recall for the 'No Habited' class.

Another potential limitation was the use of a single algorithm for classification. While AdaBoost is a powerful and versatile tool, using additional or alternative machine learning algorithms could provide different perspectives and potentially enhance the model's performance. For instance, exploring other ensemble methods, like Random Forest, Gradient Boosting, or even deep learning approaches, could be beneficial.

Lastly, the study did not consider potential temporal changes in the habitation status of houses post-disaster. As rehabilitation efforts progress over time, the habitation status of houses could change. Incorporating this temporal dimension into future models could provide more dynamic and accurate predictions.

Overall, these potential areas for improvement highlight the vast scope of this research area and provide exciting avenues for future work. Through iterative enhancements and continual exploration of other techniques, we hope to significantly improve models' predictive power and utility in post-disaster habitation status prediction.

4. CONCLUSION

This study has demonstrated the potential of using the AdaBoost algorithm for predicting residential habitation status following the eruption of Mount Merapi. With its robust performance and generalization capabilities, the algorithm successfully classified houses into 'Habited' and 'No Habited' classes with high accuracy, particularly for the 'Habited' class. Despite facing challenges due to imbalances in the dataset, these findings present a promising step toward the application of AI in disaster management and recovery processes. However, the study also underscored the need for future research to address the identified limitations and explore improvements. Improving the model's performance in predicting the 'No Habited' class, exploring other machine learning algorithms, and incorporating temporal dimensions to capture dynamic changes post-disaster are key areas for future work. Ultimately, this study is a foundational step in a broader effort to leverage AI in aiding and optimizing disaster rehabilitation efforts, leading to more effective resource allocation and faster recovery times.

REFERENCES

[1] R. D. Indriana, K. S. Brotopuspito, A. Setiawan, and T. A. Soenantyo, "Pre and post Mount Merapi eruption of free air anomaly in 2010," *ijpse*, vol. 2, no. 3, pp. 70–76, Dec. 2018, doi: 10.29332/ijpse.v2n3.231.

[2] A. K. M. Kablan, K. Dongo, G. Fokou, and M. Coulibaly, "Assessing population perception and socioeconomic impact related to flood episodes in urban Côte d'Ivoire," *Int. J. Bio. Chem. Sci*, vol. 13, no. 4, p. 2210, Nov. 2019, doi: 10.4314/ijbcs.v13i4.26.

[3] A. Santoso, J. Parung, D. N. Prayogo, A. Lolita, and Department of Industrial Engineering, University of Surabaya Raya Kalirungkut, Surabaya 60293, Indonesia, "Designing an Indonesian Disaster Management Information System with Local Characteristics: A Case Study of Mount Merapi," *JDR*, vol. 16, no. 4, pp. 765–777, Jun. 2021, doi: 10.20965/jdr.2021.p0765.

[4] A. Flora, D. Cardone, M. Vona, and G. Perrone, "A Simplified Approach for the Seismic Loss Assessment of RC Buildings at Urban Scale: The Case Study of Potenza (Italy)," *Buildings*, vol. 11, no. 4, 2021, doi: 10.3390/buildings11040142.

[5] F. Chasanah and H. Sakakibara, "Implication of Mutual Assistance Evacuation Model to Reduce the Volcanic Risk for Vulnerable Society: Insight from Mount Merapi, Indonesia," *Sustainability*, vol. 14, no. 13, 2022, doi: 10.3390/su14138110.

[6] V. W. Astuti and R. Rimawati, "Kelud Community Activities in Disaster Management," *jqph*, vol. 5, no. 1, pp. 339–343, Nov. 2021, doi: 10.30994/jqph.v5i1.270.

[7] M. Filipovič Hrast, R. Sendi, and B. Kerbler, "Housing Choices of Older People: Staying or Moving in the Case of High Care Needs," *Sustainability*, vol. 12, no. 7, 2020, doi: 10.3390/su12072888.

[8] T. Kim and W.-D. Lee, "Review on Applications of Machine Learning in Coastal and Ocean Engineering," *J. Ocean Eng. Technol.*, vol. 36, no. 3, pp. 194–210, Jun. 2022, doi: 10.26748/KSOE.2022.007.

[9] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *CoSciTech*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755.

[10] N. Wijaya, "Evaluation of Naïve Bayes and Chi-Square performance for Classification of Occupancy House," *International Journal of Informatics and Computation; Vol 1 No 2 (2019): International Journal of Informatics and ComputationDO - 10.35842/ijicom.v1i2.20*, Feb. 2020, [Online]. Available: https://ijicom.respati.ac.id/index.php/ijicom/article/view/20

[11] Sujacka Retno, Rozzi Kesuma Dinata, and Novia Hasdyna, "Evaluasi model data chatbot dalam natural language processing menggunakan k-nearest neighbor," *CoSciTech*, vol. 4, no. 1, pp. 146–153, Apr. 2023, doi: 10.37859/coscitech.v4i1.4690.

[12] N. Wijaya, J. Aryanto, K. Kasmawaru, and A. F. Rachman, "Implementation of KNN Algorithm for Occupancy Classification of Rehabilitation Houses," *International Journal of Informatics and Computation; Vol 4 No 2 (2022): International Journal of Informatics and ComputationDO - 10.35842/ijicom.v4i2.36*, Dec. 2022, [Online]. Available: https://ijicom.respati.ac.id/index.php/ijicom/article/view/36

[13] J. Zhou, X. Xu, X. Huo, and Y. Li, "Forecasting Models for Wind Power Using Extreme-Point Symmetric Mode Decomposition and Artificial Neural Networks," *Sustainability*, vol. 11, no. 3, 2019, doi: 10.3390/su11030650.

[14] S. Jagdale and M. Shah, "Extending the Classifier Algorithms in Machine Learning to Improve the Performance in Spoken Language Understanding Systems Under Deficient Training Data," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 464–471, 2020, doi: 10.25046/aj050655.

[15] N. Chen, "House Price Prediction Model of Zhaoqing City Based on Correlation Analysis and Multiple Linear Regression Analysis," *Wireless Communications and Mobile Computing*, vol. 2022, p. 9590704, May 2022, doi: 10.1155/2022/9590704.

[16] A. Burger, T. Oz, W. G. Kennedy, and A. T. Crooks, "Computational Social Science of Disasters: Opportunities and Challenges," *Future Internet*, vol. 11, no. 5, 2019, doi: 10.3390/fi11050103.

[17] S. Chen *et al.*, "Physical-Layer Channel Authentication for 5G via Machine Learning Algorithm," *Wireless Communications and Mobile Computing*, vol. 2018, p. 6039878, Oct. 2018, doi: 10.1155/2018/6039878.

[18] S. Saravi, R. Kalawsky, D. Joannou, M. Rivas Casado, G. Fu, and F. Meng, "Use of Artificial Intelligence to Improve Resilience and Preparedness Against Adverse Flood Events," *Water*, vol. 11, no. 5, 2019, doi: 10.3390/w11050973.

[19] G. Luo, M. D. Johnson, F. L. Nkoy, S. He, and B. L. Stone, "Automatically Explaining Machine Learning Prediction Results on Asthma Hospital Visits in Patients With Asthma: Secondary Analysis," *JMIR Med Inform*, vol. 8, no. 12, p. e21965, Dec. 2020, doi: 10.2196/21965.

[20] A. Abdurrasyid, I. Indrianto, and M. N. I. Susanti, "Face detection and global positioning system on a walking aid for blind people," *Bulletin of Electrical Engineering and Informatics; Vol 11, No 3: June 2022DO - 10.11591/eei.v11i3.3429*, Jun. 2022, [Online]. Available: https://beei.org/index.php/EEI/article/view/3429