International Journal of Artificial Intelligence in Medical Issues

Volume 1 Issue 2
ISSN 3025-4167
https://doi.org/10.56705/ijaimi.v1i2.96

Research Article

Performance Evaluation of Bagging Meta-Estimator in Lung Disease Detection: A Case Study on Imbalanced Dataset

Rezania Agramanisti Azdy [1,*], Rahmat Fuadi Syam [2], Edi Faizal [3], Sumiyatun [4]
[1] Universitas Bina Darma, rezania.agramanisti.azdy@binadarma.ac.id
[2] Universitas Pancasakti Makassar, rahmat@unpacti.ac.id
[3] Universitas Teknologi Digital Indonesia, edifaizal@utdi.ac.id
[4] Universitas Teknologi Digital Indonesia, sumiyatun@utdi.ac.id
Correspondence should be addressed to Rezania Agramanisti Azdy; rezania.agramanisti.azdy@binadarma.ac.id

Received 05 July 2023; Revised 19 August 2023; Accepted 28 September 2023; Published 30 November 2023

Abstract:
In this study, titled "Performance Evaluation of Bagging Meta-Estimator in Lung Disease Detection: A Case Study on Imbalanced Dataset," we explore the effectiveness of the Bagging Meta-Estimator in diagnosing lung diseases, focusing on the challenges of imbalanced datasets. Utilizing a dataset segmented and characterized by Hu moments and encompassing categories of Normal, Bacterial Pneumonia, and Tuberculosis, the algorithm's performance was assessed through a 5-fold cross-validation. Results indicated moderate effectiveness with an average accuracy of 60.574%, precision of 60.749%, recall of 59.753%, and F1-Score of 59.416%, highlighting variable performance across folds. These findings suggest that while the Bagging Meta-Estimator has potential in medical imaging, further refinement is needed for consistent and reliable lung disease detection, especially in imbalanced datasets.

Keywords: Lung Disease Detection, Bagging Meta-Estimator, Imbalanced Datasets, Medical Imaging, Machine Learning, Hu Moments, Cross-Validation.

Dataset link: https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types/

1. Introduction

Lung diseases, including bacterial pneumonia and tuberculosis, are major global health concerns. In recent years, the advancement of medical imaging technologies has offered new avenues for early diagnosis and treatment. However, the interpretation of lung images remains a challenging task due to the complex nature of lung diseases and the subtlety of their manifestations in medical images. The advent of machine learning and image processing techniques has opened up possibilities for automated and accurate diagnosis. Among these, the application of feature extraction methods, such as Hu moments from segmented lung images, has shown promise. Yet, the effectiveness of these techniques in a real-world clinical setting, especially when dealing with imbalanced datasets [1], is still under investigation.

One significant challenge in the automated diagnosis of lung diseases from medical images is the imbalanced nature of datasets. In many real-world scenarios, certain conditions, like tuberculosis, are underrepresented compared to more common conditions, such as bacterial pneumonia or normal lung images. This imbalance can lead to biased models that perform poorly in detecting less common diseases. Furthermore, the choice of machine learning algorithms and their ability to handle such imbalances effectively remains a critical area for exploration.

The primary objective of this research is to evaluate the performance of the Bagging Meta-Estimator, a robust machine learning approach, in the context of lung disease detection. This study aims to determine whether this algorithm can effectively handle the challenges posed by imbalanced datasets in medical imaging. By focusing on a dataset comprising Normal, Bacterial Pneumonia, and Tuberculosis classes, the study seeks to provide insights into the algorithm's ability to distinguish between these conditions accurately.

This study is guided by the hypothesis that the Bagging Meta-Estimator [2], [3] will demonstrate superior performance in handling imbalanced datasets for lung disease detection compared to traditional machine learning models. Specifically, it hypothesizes that the Bagging Meta-Estimator will yield higher accuracy, precision, recall, and F-measure scores in distinguishing between normal lungs, bacterial pneumonia, and tuberculosis. These metrics are crucial in medical diagnosis, where the cost of false negatives and false positives can be high.

The scope of this research is confined to the evaluation of the Bagging Meta-Estimator on a segmented and feature-extracted dataset comprising three classes of lung conditions. The study utilizes Hu moments extracted from segmented lung images using thresholding techniques. It's important to note that while this approach is promising, the findings may not be generalizable to other types of lung diseases or imaging techniques. Additionally, the study's reliance on a specific dataset means that the results may be influenced by the peculiarities of this data, including its size and quality.

This research aims to contribute to the field of medical image analysis by providing a comprehensive evaluation of the Bagging Meta-Estimator's performance on an imbalanced lung disease dataset. The findings are expected to offer valuable insights for healthcare professionals and researchers in the field of medical diagnostics. Furthermore, this study contributes to the broader domain of machine learning by exploring the application of advanced algorithms in handling imbalanced datasets, a common challenge in various domains, including healthcare.

2. Method

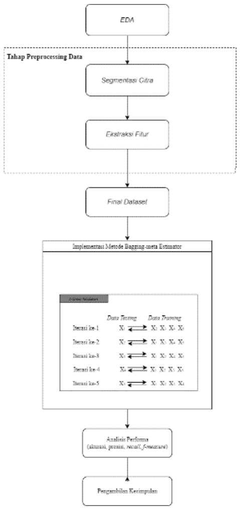

Figure 1: Bagging-meta Estimator Evaluation Workflow

This study adopts a quantitative research design, focusing on the evaluation of the Bagging Meta-Estimator algorithm for lung disease detection. The research design encompasses the collection of lung image data, application of image segmentation and feature extraction techniques, followed by the implementation of the Bagging Meta-Estimator algorithm. To validate the performance of the model, the study employs a 5-fold cross-validation approach, analyzing metrics such as accuracy, precision, recall, and F-measure. A visual representation of the entire research process is illustrated in Figure 1.

Sample or Data Selection:

The dataset used in this study consists of lung images classified into three categories: Normal, Bacterial Pneumonia, and Tuberculosis. These images have undergone preprocessing stages including segmentation and feature extraction. The selection criteria for the images were based on their diagnostic clarity and representativeness of each condition. The dataset is noted for its imbalanced nature, with varying representation across the three categories, reflecting real-world scenarios in medical diagnostics.

Tools and Technology Used:

The research harnesses Python as the primary programming language, leveraging libraries such as OpenCV for image processing, Scikit-learn for machine learning, and NumPy for numerical computations.

Thresholding Edge Detection

The lung images were segmented using a thresholding technique. Thresholding is a simple, yet effective method for separating objects from the background in an image [4], [5]. The formula for basic thresholding is given by Equation (1) [6]–[8]:

$$I(x,y) = \{ 1, \text{if } \mathrm{I(x,y)} > T, 0, if\ I(x,y) \leq T\} \tag{1}$$

$I(x,y)$ is the intensity of the pixel at coordinate (x,y) in the image, and T is the specified threshold value. When the pixel intensity $(I(x,y))$ exceeds the threshold value (T), the pixel will be classified as a pixel with a value of 1 (white), whereas if the intensity is less than or equal to the threshold value, the pixel will be classified as a pixel with a value of 0 (black). In Figure 2, 3 and 4 the results of image segmentation using Thresholding features on the dataset are shown.

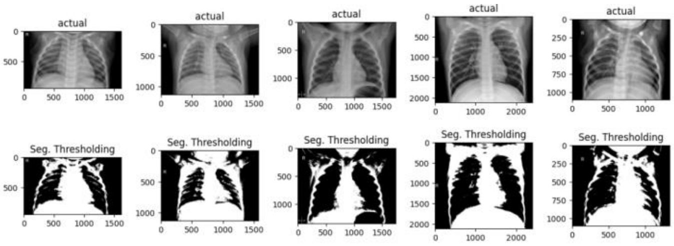

Figure 2: Thresholding Edge Detection Results for Healthy Class

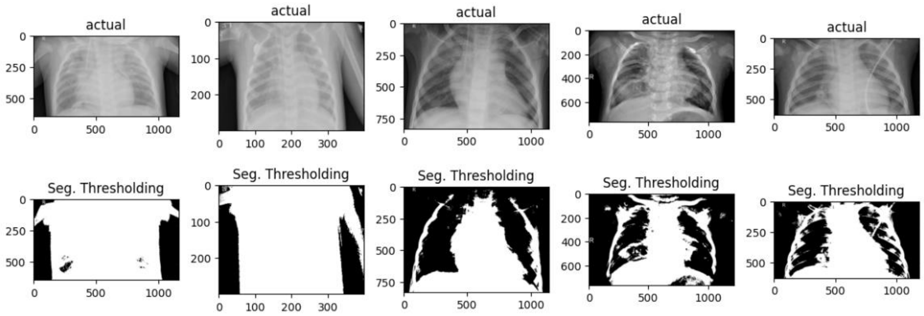

Figure 3: Thresholding Edge Detection Results for Bacterial Pneumonia Class

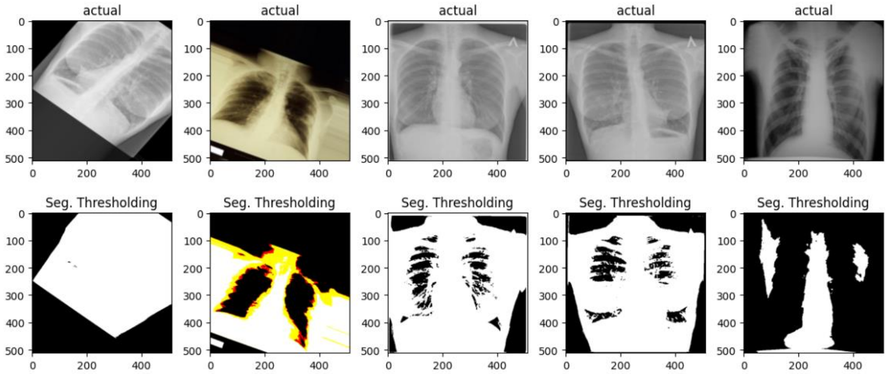

Figure 4: Thresholding Edge Detection Results for Tuberculosis Class

Hu-Moments

Hu moments were extracted from the segmented images [9]. Hu moments are a set of seven moment invariants derived from image moments, providing a basis for shape description. The Hu moments are defined as Equation (2):

$$H = h_1, h_2, \ldots, h_7 \tag{2}$$

Where each is a function of the image's central moments, providing a translation, scale, and rotation invariant description of the image.

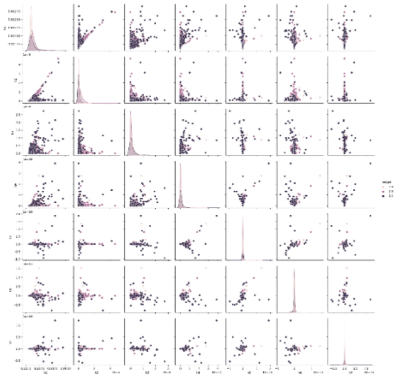

Figure 4: Scatter Plot Visualization of Extracted Hu Moments Features

Model Training and Testing

The Bagging Meta-Estimator algorithm was applied to the dataset. Bagging, or Bootstrap Aggregating, involves training multiple instances of a base estimator on random subsets of the original dataset and then aggregating their individual predictions to form a final prediction [10]–[12]. The Bagging Meta-Estimator formula can be represented as Equation (3):

$$y_{pred} = \frac{1}{N}\sum_{i=1}^{N} y_{pred_i} \tag{3}$$

Where y_{pred} is the final prediction, N is the number of base estimators, and y_{pred_i} is the prediction made by the i^{th} base estimator.

Performance Evaluation

For model evaluation, 5-fold cross-validation was utilized. In this approach, the dataset is divided into five equal parts. In each iteration, four parts are used for training, and the remaining part is used for testing. This process is repeated five times, with each part being used exactly once as the test set. The performance metrics—accuracy, precision, recall, and F-measure [13]–[15] are calculated for each fold and then averaged to provide a comprehensive evaluation of the model's performance. he formulas for these metrics are as follow Equation (4) [16]–[20]:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

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

3. Result and Discussion

The evaluation of the Bagging Meta-Estimator's performance in lung disease detection was conducted using a 5-fold cross-validation approach on an imbalanced dataset. The dataset comprised lung images, segmented and characterized by Hu moments, classified into three categories: Normal, Bacterial Pneumonia, and Tuberculosis.

Visualization of the Results

The detailed results are presented in Table 1 and visualized in Figure 5 for a clearer understanding and comparison of the metrics across different iterations.

Table 1: Performance Metrics Across 5-Fold Cross-Validation for the Bagging-meta Estimator Algorithm

K-n	**Performa**			
	Akurasi	***Presisi***	***Recall***	***F-Measure***
K-1	64.609%	62.367%	62.963%	60.957%
K-2	63.786%	59.884%	65.432%	61.982%
K-3	53.909%	56.659%	48.148%	53.444%
K-4	61.317%	61.699%	60.905%	58.603%
K-5	59.259%	63.137%	61.317%	62.095%
$\sum \boldsymbol{Avg}$	60.574%	60.749%	59.753%	59.416%

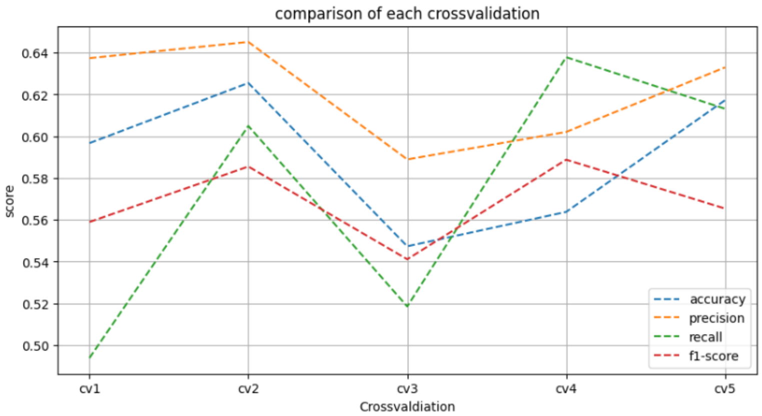

Figure 5: Visualization of Performance Metrics Across 5-Fold Cross-Validation for the Bagging-meta Estimator Algorithm

On average, the model achieved an Accuracy of 60.574%, Precision of 60.749%, Recall of 59.753%, and F-Measure of 59.416%.

Interpretation of the Results

The accuracy shows moderate performance, with a notable dip in the third fold. This variation might be attributed to the imbalanced nature of the dataset or particularities within the folds. The precision and recall values are relatively consistent, with recall demonstrating more variability across folds. The lower recall in the third fold suggests difficulties in correctly identifying all positive cases of lung diseases. The F1-score, an important metric in the context of imbalanced datasets, presents a balanced view between precision and recall. Its fluctuation across folds points to challenges in achieving a consistent trade-off between false positives and false negatives.

Discussion

The moderate performance of the Bagging Meta-Estimator in this study suggests that while the algorithm can handle imbalanced datasets to a certain extent, there are limitations in its efficacy. The fluctuation in recall rates across different folds indicates potential inconsistencies in the model's ability to identify all positive cases accurately. Previous research has emphasized the challenges posed by imbalanced datasets in medical imaging. Our results align with these findings, highlighting the complexities of achieving high accuracy in such conditions. The use of Bagging

Meta-Estimators has been seen as a promising approach in other domains, and this research extends its application to lung disease detection.

The findings have practical implications for the application of machine learning in medical imaging. While the Bagging Meta-Estimator shows potential, the moderate results suggest a need for further refinement when applied in clinical settings for lung disease detection. The main limitation of this research lies in its reliance on a specific dataset, which may not be representative of all types of lung diseases. Additionally, the inherent imbalances in the dataset could have impacted the model's performance.

Recommendations for Further Research

Alternative machine learning algorithms or advanced ensemble methods that might offer better handling of imbalanced datasets. Different feature extraction techniques to enhance model accuracy. Expansion of the dataset to include a broader range of lung diseases and conditions.

4. Conclusion

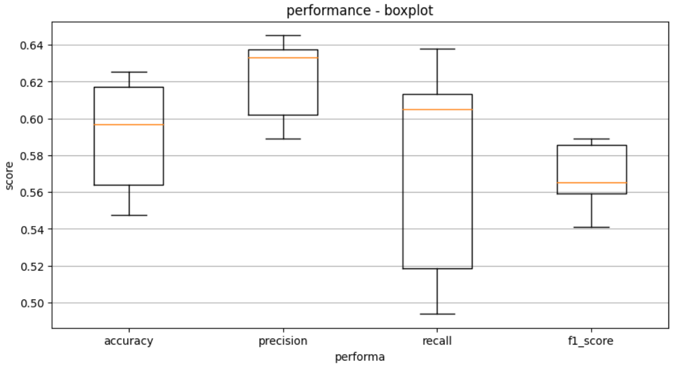

Figure 6: Boxplot of Performance Metrics Across 5-Fold Cross-Validation for the Bagging-meta Estimator Algorithm

This study set out to evaluate the performance of the Bagging Meta-Estimator in detecting lung diseases within an imbalanced dataset. Our results revealed moderate effectiveness, with average accuracy, precision, recall, and F1-Scores of approximately 60.574%, 60.749%, 59.753%, and 59.416% respectively. The notable variability in performance across different folds, especially a dip in recall rates, underscores the challenges inherent in working with imbalanced medical imaging datasets. This research supports the hypothesis that while Bagging Meta-Estimators can be effective in managing imbalanced datasets, their performance in complex medical imaging tasks, such as lung disease detection, may require further optimization. These findings contribute to the growing body of knowledge in the application of machine learning algorithms in medical imaging, particularly in addressing the challenges posed by imbalanced datasets.

Moving forward, it is recommended that future research explores more advanced ensemble techniques or alternative machine learning algorithms that might offer better performance in handling imbalanced datasets. Additionally, investigating the impact of different feature extraction methods could potentially enhance model accuracy and reliability. From a practical standpoint, expanding the dataset to include a wider range of lung conditions and integrating these findings into clinical practice would be beneficial. These steps are essential to improve the diagnostic accuracy and efficacy of automated systems, ultimately aiding in early and accurate detection of lung diseases in clinical settings.

References:

[1] J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai, "Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting," *Inf. Fusion*, vol. 54, pp. 128–144, Feb. 2020, doi: 10.1016/j.inffus.2019.07.006.

[2] A. Mohanta and S. Panigrahi, "Health Insurance Fraud Detection Using Feature Selection and Ensemble Machine Learning Techniques," in *Advances in Distributed Computing and Machine Learning*, 2023, pp. 197–207.

[3] J. Dutta, Y. W. Kim, and D. Dominic, "Comparative Analysis of Various Ensemble Approaches for Web Page Classification," in *Data Engineering and Data Science*, Wiley, 2023, pp. 137–172.

[4] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 1828–1837, doi: 10.1109/CVPR.2018.00196.

[5] M. Zhao, S. Zhong, X. Fu, B. Tang, and M. Pecht, "Deep Residual Shrinkage Networks for Fault Diagnosis," *IEEE Trans. Ind. Informatics*, vol. 16, no. 7, pp. 4681–4690, Jul. 2020, doi: 10.1109/TII.2019.2943898.

[6] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology," *IEEE Trans. Med. Imaging*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017, doi: 10.1109/TMI.2017.2677499.

[7] D. Park, Y. Hoshi, and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based Variational Autoencoder," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1544–1551, Jul. 2018, doi: 10.1109/LRA.2018.2801475.

[8] M. A. El Aziz, A. A. Ewees, and A. E. Hassanien, "Whale Optimization Algorithm and Moth-Flame Optimization for multilevel thresholding image segmentation," *Expert Syst. Appl.*, vol. 83, pp. 242–256, Oct. 2017, doi: 10.1016/j.eswa.2017.04.023.

[9] X. Lu *et al.*, "An Outdoor Support Insulator Surface Defects Segmentation Approach via Image Adversarial Reconstruction in High-Speed Railway Traction Substation," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–19, 2022, doi: 10.1109/TIM.2022.3211558.

[10] J. Treliński and B. Kwolek, "Ensemble of Multi-channel CNNs for Multi-class Time-Series Classification. Depth-Based Human Activity Recognition," in *Mobile Computing and Sustainable Informatics*, 2020, pp. 455–466.

[11] A. Singh and V. Gutte, "Classification of Breast Tumor Using Ensemble Learning," in *Mobile Computing and Sustainable Informatics*, 2022, pp. 491–507.

[12] R. R. Kadhim and M. Y. Kamil, "Comparison of breast cancer classification models on Wisconsin dataset," *Int. J. Reconfigurable Embed. Syst.*, vol. 11, no. 2, p. 166, Jul. 2022, doi: 10.11591/ijres.v11.i2.pp166-174.

[13] M. Ziacchi *et al.*, “Bipolar active fixation left ventricular lead or quadripolar passive fixation lead? An Italian multicenter experience,” *J. Cardiovasc. Med.*, vol. 20, no. 4, pp. 192–200, Apr. 2019, doi: 10.2459/JCM.0000000000000778.

[14] E. K. Shea and R. S. Hess, “Assessment of postprandial hyperglycemia and circadian fluctuation of glucose concentrations in diabetic dogs using a flash glucose monitoring system,” *J. Vet. Intern. Med.*, vol. 35, no. 2, pp. 843–852, Mar. 2021, doi: 10.1111/jvim.16046.

[15] A. Sindy, “Pattern of Patients and Diseases During Mass Transit: The Day of Arafat Experience,” *Pakistan J. Med. Sci.*, vol. 31, no. 5, Sep. 2015, doi: 10.12669/pjms.315.8017.

[16] M. M. Baharuddin, T. Hasanuddin, and H. Azis, “Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca,” *Ilk. J. Ilm.*, vol. 11, no. 28, pp. 269–274, 2019, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Baharuddin, Hasanuddin, Azis - 2019 - Analisis Performa Metode K-Nearest Neighbor untuk Identifikasi Jenis Kaca.pdf.

[17] H. Azis, F. T. Admojo, and E. Susanti, “Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah,” *Techno.Com*, vol. 19, no. 3, 2020, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Azis, Admojo, Susanti - 2020 - Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah.pdf.

[18] A. Nurul, Y. Salim, and H. Azis, “Analisis performa metode Gaussian Naïve Bayes untuk klasifikasi citra tulisan tangan karakter arab,” *Indones. J. Data Sci.*, vol. 3, no. 3, pp. 115–121, 2022, doi: https://doi.org/10.56705/ijodas.v3i3.54.

[19] H. Azis, F. Fattah, and P. Putri, “Performa Klasifikasi K-NN dan Cross-validation pada Data Pasien Pengidap Penyakit Jantung,” *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, [Online]. Available: file:///Users/kbh/Downloads/507-2012-5-PB.pdf.

[20] A. A. Karim, H. Azis, and Y. Salim, “Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana 1,” *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 84–87, 2018, [Online]. Available: file:///Users/kbh/Library/Application Support/Mendeley Desktop/Downloaded/Karim, Azis, Salim - 2018 - Kinerja Metode C4.5 dalam Penyaluran Bantuan Dana Bencana 1.pdf.