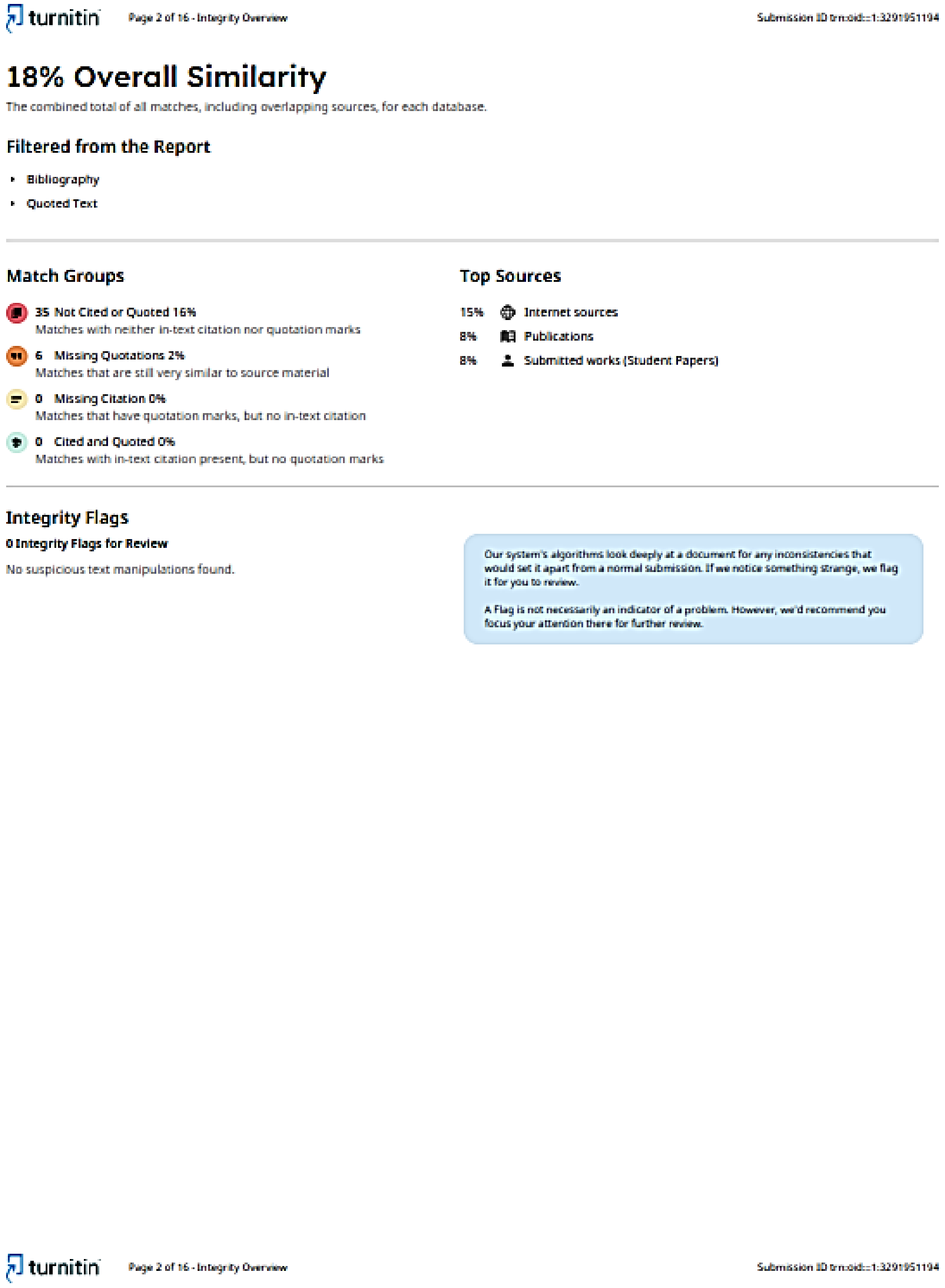


18% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Match Groups

35 Not Cited or Quoted 16%
Matches with neither in-text citation nor quotation marks

6 Missing Quotations 2%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

15%	Internet sources
8%	Publications
8%	Submitted works (Student Papers)

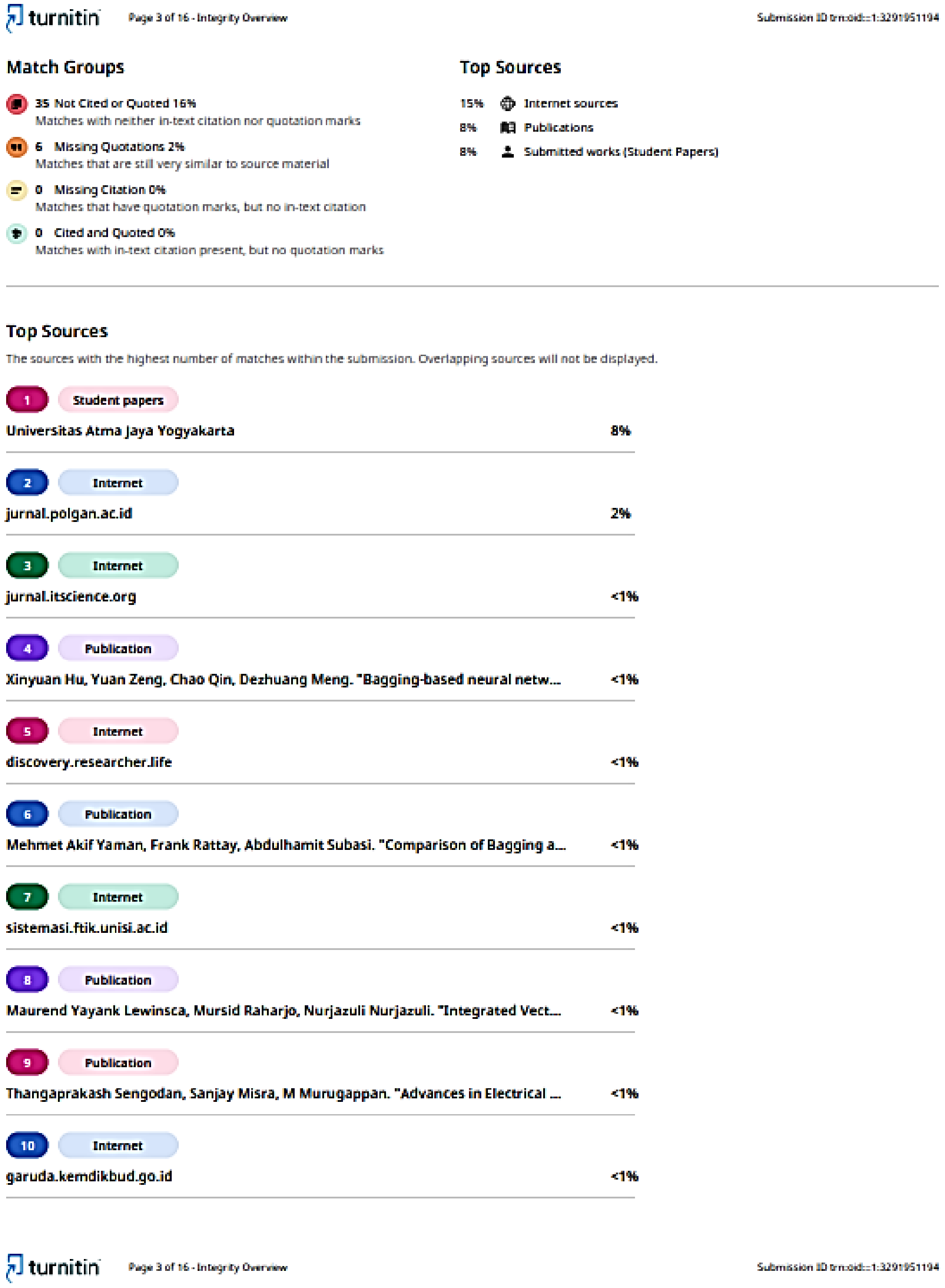
Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.



Match Groups

- 35 Not Cited or Quoted 16%
  Matches with neither in-text citation nor quotation marks
- 6 Missing Quotations 2%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

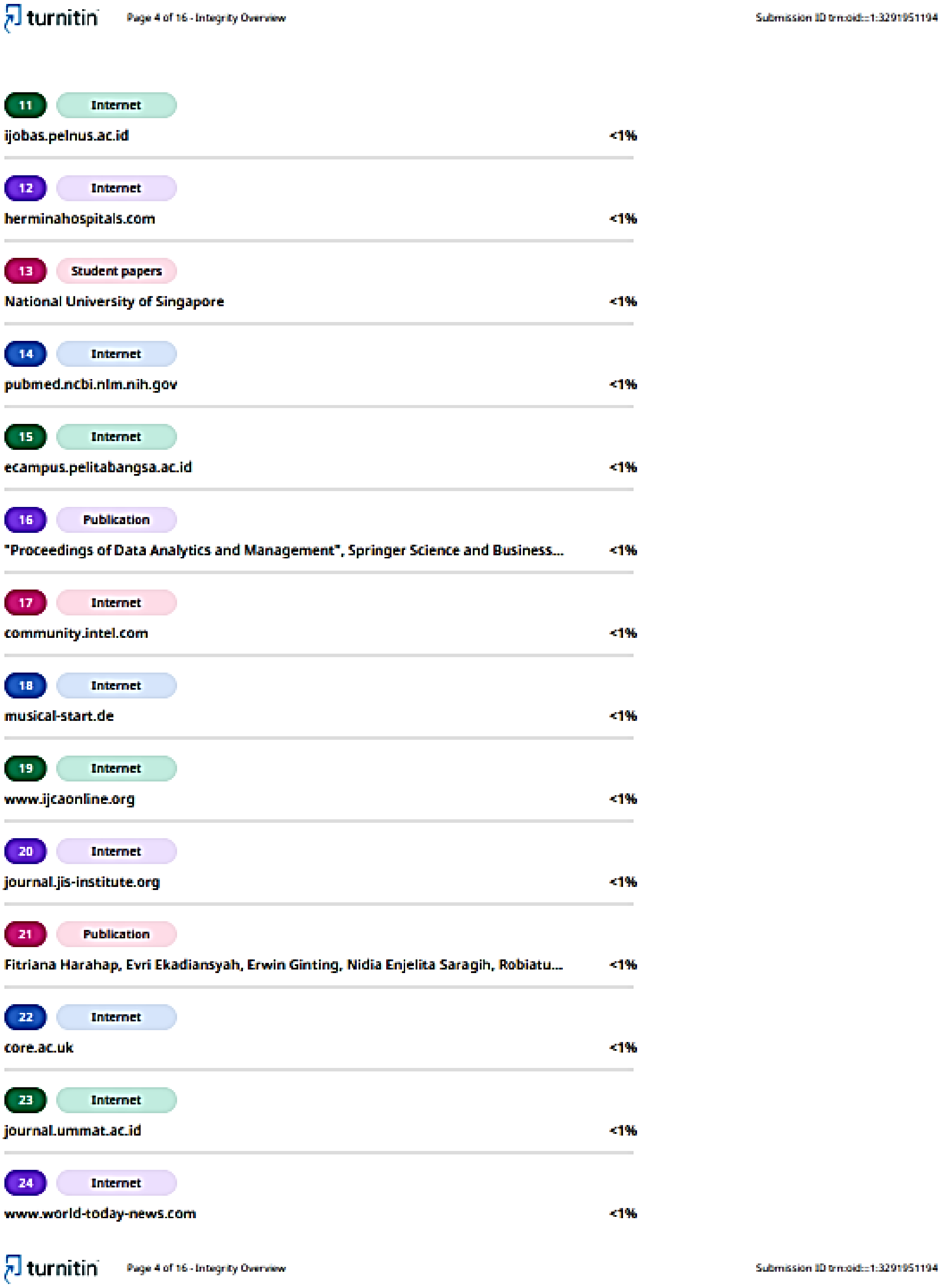
Top Sources

15%	Internet sources
8%	Publications
8%	Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	Universitas Atma Jaya Yogyakarta	8%
2	Internet	jurnal.polgan.ac.id	2%
3	Internet	jurnal.itscience.org	<1%
4	Publication	Xinyuan Hu, Yuan Zeng, Chao Qin, Dezhuang Meng. "Bagging-based neural netw...	<1%
5	Internet	discovery.researcher.life	<1%
6	Publication	Mehmet Akif Yaman, Frank Rattay, Abdulhamit Subasi. "Comparison of Bagging a...	<1%
7	Internet	sistemasi.ftik.unisi.ac.id	<1%
8	Publication	Maurend Yayank Lewinsca, Mursid Raharjo, Nurjazuli Nurjazuli. "Integrated Vect...	<1%
9	Publication	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ...	<1%
10	Internet	garuda.kemdikbud.go.id	<1%



turnitin Page 4 of 16 - Integrity Overview Submission ID trn:oid:::1:3291951194

11	Internet	ijobas.pelnus.ac.id	<1%
12	Internet	herminahospitals.com	<1%
13	Student papers	National University of Singapore	<1%
14	Internet	pubmed.ncbi.nlm.nih.gov	<1%
15	Internet	ecampus.pelitabangsa.ac.id	<1%
16	Publication	"Proceedings of Data Analytics and Management", Springer Science and Business...	<1%
17	Internet	community.intel.com	<1%
18	Internet	musical-start.de	<1%
19	Internet	www.ijcaonline.org	<1%
20	Internet	journal.jis-institute.org	<1%
21	Publication	Fitriana Harahap, Evri Ekadiansyah, Erwin Ginting, Nidia Enjelita Saragih, Robiatu...	<1%
22	Internet	core.ac.uk	<1%
23	Internet	journal.ummat.ac.id	<1%
24	Internet	www.world-today-news.com	<1%

turnitin Page 4 of 16 - Integrity Overview Submission ID trn:oid:::1:3291951194

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Prediction of DHF Disease Using Bagging Algorithm with Decision Tree C4.5

Abdul Halim Mahfudzin[1], **Sriyanto**[2]*, **Sutedi**[3], **Wasilah**[4]
[1][2][3][4]Fakultas Ilmu Komputer, Insitut Informatika dan Bisnis Darmajaya, Lampung, Indonesia
[1]abdul.2321211022@mail.darmajaya.ac.id, [2]paksriyanto@darmajaya.ac.id, [3]sutedi@darmajaya.ac.id, [4]wasilah@darmajaya.ac.id

ABSTRACT

Dengue Fever (DHF) continues to represent a significant public health threat in Indonesia and other tropical regions, with an annual increase in the number of reported cases. The primary aim of this study is to develop a predictive model for DHF by integrating the Bagging technique and the Decision Tree C4.5 algorithm. The goal is to improve prediction accuracy by incorporating key environmental factors such as temperature, humidity, and rainfall. The research adopts a quantitative methodology with a descriptive approach, using publicly available datasets from data.mendeley.com and conducting the analysis using RapidMiner software. The findings of the study demonstrate that the proposed model is highly effective in accurately predicting and classifying DHF cases, achieving significant precision. In addition to this, the model is successful in identifying important patterns and trends linked to the disease's occurrence. These results underscore the efficacy of combining Bagging and Decision Tree C4.5 as a robust tool for detecting and forecasting DHF outbreaks. The research contributes substantially to the field of data-driven prediction models, offering valuable insights for health agencies to develop more effective and proactive strategies for disease prevention. For future research, it is recommended that additional factors such as genetic and medical data be considered, along with the application of triangulation methods to improve the analysis's validity, scope, and overall robustness. This approach would enable a more comprehensive understanding of DHF and its predictive modeling.

Keywords: DHF Prediction; Bagging; Decision Tree C4.5; Machine Learning; Data Mining

INTRODUCTION

Dengue fever is a common infectious disease in tropical countries, including Indonesia, and the number of cases continues to increase every year. Based on data from the Ministry of Health of the Republic of Indonesia, by 2023, there will be more than 60,000 cases of DHF with a high mortality rate. The disease is caused by the Dengue virus, which is transmitted through the bite of the Aedes aegypti mosquito. Data from the World Health Organization (WHO) shows a significant increase in dengue cases in recent years, signifying the importance of early detection efforts to prevent further spread of the disease(Saputri et al., 2022). Dengue fever transmission is faster in urban areas compared to rural areas due to high population density. This increases the chance of contact between infected and susceptible individuals. In addition, Aedes aegypti, as the main vector of Dengue virus, is highly adaptive to the domestic environment, allowing the virus to spread more efficiently from one individual to another(Cholil et al., n.d.).

Dengue fever is an acute febrile illness, especially affecting children, which is accompanied by a direct impact of bleeding and tends to cause a shock effect that can cause death and often causes extraordinary events(Riswanto et al., 2014). Dengue fever is a disease spread by the Aedes aegypti mosquito characterized by sudden, high, and persistent fever that lasts for 2-7 days accompanied by signs of bleeding in the skin, bruising, bleeding gums, vomiting blood. Individuals who have Dengue virus in their blood are the source of dengue disease transmission, who may or may not show symptoms of illness, depending on the body's immunity to the virus. When an individual is bitten by an Aedes aegypti mosquito, the Dengue virus will enter along with the blood that is sucked. Inside the mosquito, the virus multiplies through the process of self-division and spreads to all parts of the mosquito's body. The majority of the virus is found in the mosquito's salivary glands. Within a week, the number can reach tens or hundreds of thousands. Furthermore, when the mosquito bites another individual, after the mosquito's proboscis reaches the blood capillaries, before the blood is sucked, saliva is released first to prevent the blood from clotting. With the mosquito's saliva, the

---

* Corresponding author

Dengue virus is transmitted to another individual(Riswanto et al., 2014).

Although various prevention efforts have been made, such as mosquito nest eradication and socialization to the community, early detection and prediction of DHF outbreaks is still a big challenge. Therefore, a data-driven prediction system is needed that can provide accurate and timely results to design more effective prevention strategies. Recently, machine learning algorithms, such as Decision Tree C4.5 and Bagging methods, have shown great potential in developing data-driven disease prediction models. Previous studies have shown the successful application of these algorithms in various medical classification problems, but few have explicitly combined these two algorithms to predict dengue diseases(Handayani, 2019). Through the use of the C4.5 algorithm in this study, it is expected to produce a model that is not only accurate but also easy to understand and implement(Prayoga & Kadyanan, 2025). However, although the C4.5 and Bagging algorithms have been used in various studies in the health sector for disease classification, the application of these two algorithms, especially in dengue prediction, is still limited. Some previous studies have indeed tried various prediction models for DHF, but the results obtained are still not optimal, both in terms of accuracy and wider application. This research aims to develop a more accurate prediction model by combining the two algorithms, namely Bagging and Decision Tree C4.5, to analyze the factors that affect the spread of DHF in various regions.

Based on the existing problems, the author wishes to classify Dengue Fever (DHF) by implementing the Bagging (Bootstrap Aggregating) algorithm and Decision Tree C4.5 in a study entitled "Prediction of DHF Disease Using Bagging Algorithm with Decision Tree C4.5" (Prediksi Penyakit DBD Menggunakan Algoritma Bagging dengan Decision Tree C4.5). The main objective of this research is to find activity patterns that are reliable, effective, and accurate in predicting the spread of DHF. This research utilizes data obtained from the data.mendeley.com site which provides public datasets with open access, to enable transparent and accessible analysis for anyone interested in further development.

LITERATURE REVIEW

In a general understanding of data science, data mining is a method used in data processing to find hidden patterns from data so that it can produce knowledge. Data mining also has several methods, including association, description, estimation, classification, clustering, and prediction methods(Agustian & Darmawan, 2022). Data Mining is an activity of retrieving new information from a large set of data which is used as material for making decisions or can be said to be the process of collecting important information from large data to be converted into knowledge(Effendi et al., 2021).

Disease prediction can be analyzed using machine learning. Machine Learning (ML) has become a major focus in data analysis in various fields, including manufacturing, healthcare, and social media. The basic concept of ML is its ability to learn from existing data and provide analysis, recommendations, and predictions based on patterns found in those datasets. In this context, ML is essential for optimizing the decision-making process and improving operational efficiency(Duta et al., n.d.)(Maulina & Corry, n.d.).

In the health sector, Machine Learning methods are also used to analyze health data and disease classification. The results of a study that developed a prediction model for dengue cases based on climate data showed that climate variables including air temperature and humidity were significantly related to dengue cases(Rahayuningtyas et al., 2025). Laksono in his research in 2021, on predicting dengue using Markov chain implementation shows how important mathematical models are in understanding the increase in disease cases(Laksono et al., 2021).

Research on the spread of dengue fever by optimizing the K-Nearest Neighbor (KNN) method can predict areas with a high probability of spread (Latuconsina & Matdoan, 2022). This research emphasizes the importance of spatial analysis that can be applied in the C4.5 algorithm to identify risk indicators and deal with epidemics according to geographic location. Lutfianawati in her research in 2024 also contributed with research that used the ARIMA model to predict the incidence of DHF over a certain period of time with a retrospective approach. ARIMA is a temporally oriented model (Lutfianawati et al., 2024).

Dengue fever (DHF) is an infectious disease commonly found in tropical countries, including Indonesia, with an increasing incidence every year. According to data from the Ministry of Health of the Republic of Indonesia, by 2023 there will be more than 60,000 cases of DHF, with a relatively high mortality rate. The disease is caused by the Dengue virus, which is transmitted through the bite of the Aedes aegypti mosquito. The significant increase in the number of DHF cases in recent years, as reported by the World Health Organization (WHO), emphasizes the importance of early detection as a preventive measure so that the spread of the disease can be further controlled (Saputri et al., 2022).

The transmission process of DHF occurs more rapidly in urban areas compared to rural areas, due to the high

---

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

population density. This leads to greater opportunities for infected individuals to interact with susceptible individuals (Cholil et al., n.d.). The Aedes aegypti mosquito, as the main vector of Dengue virus, is highly adaptive to the domestic environment, enabling efficient spread of the virus from one individual to another.

The Bagging (Bootstrap Aggregating) algorithm is one of the applications of machine learning, which is an ensemble learning technique used to improve the performance of prediction models. Ensemble refers to the use of multiple combined models to make predictions, which results in a better final prediction (Lestari & Ilmu Komputer, 2024). Due to the complexity of identification, using a single network has the disadvantage of low accuracy and low reliability, while changing the type of network has no obvious effect, so Bagging-based ensemble learning is needed, to get better identification results (Hu et al., 2022). The Bagging method is used in classification to separate the training data into several new random sampling training data and build a new new model based on the training data (Wahono & Suryana, 2013).

Bagging is an ensemble learning technique using a combination of several network models, which serves to produce better predictions. The results of work on the FERET database using 10-fold cross validation, it can be seen that the Bagging method provides better results than the Adaboost and Multiboosting methods (Yaman et al., 2021). The data distribution of each Bagging data can be different, but the size of the data in Bagging is the same as the original data, because the Bagging technique is done through sampling with replacement(Kurniawan & Prihandono, 2020). Classification can be considered a process or operation that involves finding a model or activity that is able to explain and select and sort different classes of data or concepts. This procedure is carried out with the aim that the model can be used for the purpose of predicting the class of an object whose class is not yet known (Salam Nagalay, 2024).

Decision Tree C4.5 is a Machine Learning algorithm designed for data classification. Decision Tree C4.5 is used as part of a hybrid method for feature selection. This algorithm was chosen for its reliability in finding the best attributes to optimally split the data (best split). The decision tree generated by C4.5 has internal nodes that represent attribute tests, and leaves that represent the final classification or output (Nagra et al., 2020). One of its advantages is its ability to generate near-optimal decision trees even though finding the minimum decision tree is an NP-complete problem, which is theoretically very difficult to solve. Thus, C4.5 is able to reduce the complexity in attribute selection while still producing an effective and accurate decision tree. Another advantage is that the decision tree generated by C4.5 is "white box," meaning that the tree structure is transparent and easy to understand. This makes it easy for experts to validate and assess the results of decisions made by the algorithm, because all steps in the decision-making process can be seen and analyzed(Nagra et al., 2020). Classifying using Bagging in a Decision Tree can improve classification, Bagging performs better than a single decision(Elgimati, 2020).

UTILIZATION OF BAGGING AND C4.5 ALGORITHMS IN CLASSIFICATION AND PREDICTION

Research results related to the Bagging algorithm include: Studies show that the application of the Bagging ensemble algorithm to C4.5 decision trees, provides better results of up to 98.8% in some cases, compared to the accuracy of a single C4.5 which hovers around 95%(OLASUNKANMİ et al., 2020). The use of Bagging algorithms to utilize the power of C4.5 in classification shows significant promise in improving accuracy and reducing the risk of overfitting, which often occurs in simple decision tree implementations(Mijwil & Abttan, 2021). Furthermore, the implementation of C4.5 in different contexts, such as in the selection of industrial training organizations and the evaluation of patients based on disease characteristics, emphasizes the adaptability and effectiveness of this algorithm in various application domains(Ogunde & Idialu, 2019)(Anwar et al., 2018). By using methods that consider the selection of relevant attributes, the accuracy of the C4.5 algorithm can be further improved(Yohannes Cristanto et al., 2024).

METHOD

This research uses a quantitative approach with a descriptive research design that aims to develop a Dengue Fever (DHF) disease prediction model using the Bagging (Bootstrap Aggregating) and Decision Tree C4.5 algorithms. The population in this study is historical data that includes environmental factors such as temperature, rainfall, humidity, and the number of DHF cases that occur in a certain period. The sampling technique used is purposive sampling, where the data taken is public data that is openly available through the data.mendeley.com site. The research instrument used in data collection is a dataset that includes environmental variables that affect the spread of DHF. For

---

* Corresponding author

Journal of Computer Networks, Architecture and
High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

data analysis, this research uses quantitative data analysis techniques by utilizing the Bagging and Decision Tree C4.5 algorithms to build prediction models and optimize accuracy by comparing the performance of the two algorithms in predicting the number of DHF cases.

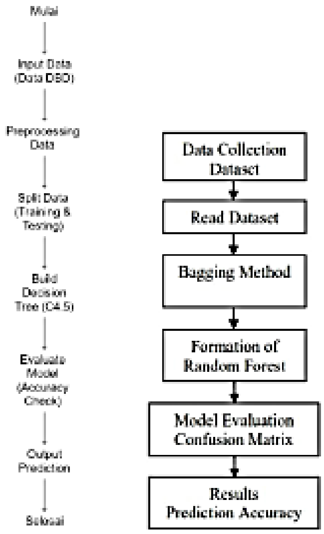


Fig. 1. Research Method of Decision Tree C4.5 (left) and Bagging Algorithm (right)

DATA PREPARATION

The initial stage in data analysis is to prepare a dataset that includes patient attributes such as age, gender, and laboratory test results. These datasets, which are generally available in .csv or .xls format, are imported into RapidMiner Studio using the Import Data feature found on the Repository tab. Afterwards, users can start creating a new project or continue an existing one, allowing the required data to be prepared and analyzed further

DATA PREPROCESSING

Before modeling, a preprocessing stage is essential to improve data quality. This process includes data cleaning through the removal of irrelevant attributes, data normalization to ensure scale consistency, and replacement of missing values. This process is important to ensure that the data used in the analysis is of good quality, so that the prediction results can be more accurate and reliable. In addition, the preprocessing stage also aims to address issues of data imbalance or incomplete data, which can affect the performance of the model.

MODEL PROCESSING AND EVALUATION

Modeling is performed using the Decision Tree algorithm to classify and predict dengue diseases. Bagging technique is applied to improve model accuracy by combining multiple models in an ensemble. In addition, Cross Validation is used to objectively evaluate model performance through cross validation. This process generates evaluation metrics such as accuracy, precision, and recall, which can be used to analyze the effectiveness of the model in recognizing patterns in the DHF dataset.

RESULT

In the preparation of this article, the authors used hardware and software tools that can be used as a foundation in

* Corresponding author

this research writing project. The following specifications are available by the compiler

1. *Hardware:*
- *Laptop Acer Swift SFG14-73*

Specifications:

㉔
- *Processor Intel® Core™ Ultra 5 125H*
- *Graphics Intel® Arc™ Pro Graphics version 32.0.101.5768*
⑰
- *Networking and I/O Intel® Wi-Fi 6E AX211 160MHz*
- *Memory (RAM) 16 GB*
- *Storage NVMe KINGSTON OM8PGP4512Q-AA (476.94 GB)*

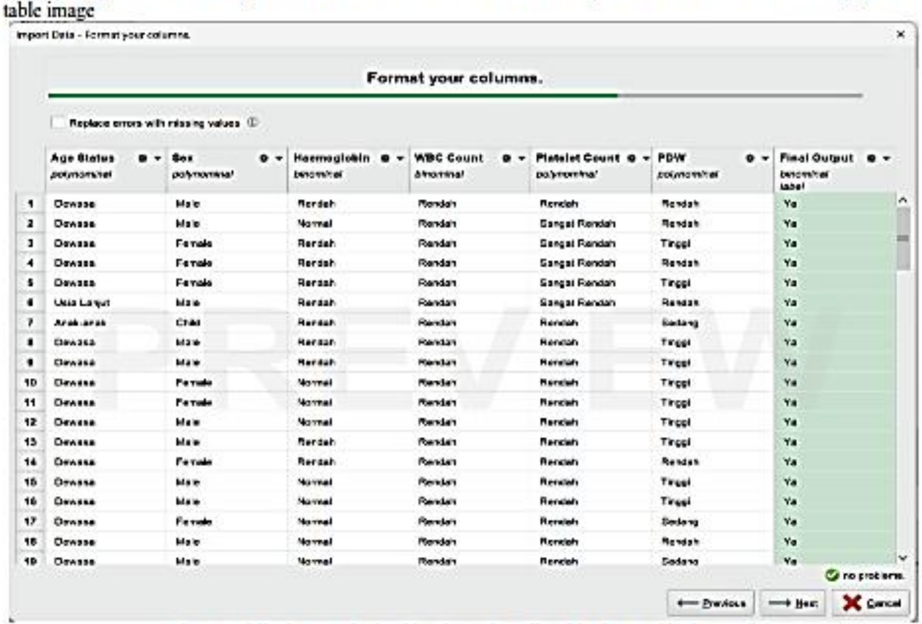
2. *Software*

⑱
- *RapidMiner AI Studio 2025.0.1*

3. *Operating System*
- *Microsoft Windows 11 Pro Version 10.0.26100 Build 26100.*

DATASET

For the public dataset sample that will be used as a research reference, the authors attach the following preview table image



	Age Status polynominal	Sex polynominal	Haemoglobin binominal	WBC Count binominal	Platelet Count polynominal	PDW polynominal	Final Output binominal label
1	Dewasa	Male	Rendah	Rendah	Rendah	Rendah	Ya
2	Dewasa	Male	Normal	Rendah	Sangat Rendah	Rendah	Ya
3	Dewasa	Female	Rendah	Rendah	Sangat Rendah	Tinggi	Ya
4	Dewasa	Female	Rendah	Rendah	Sangat Rendah	Rendah	Ya
5	Dewasa	Female	Rendah	Rendah	Sangat Rendah	Tinggi	Ya
6	Usia Lanjut	Male	Rendah	Rendah	Sangat Rendah	Rendah	Ya
7	Anak-anak	Child	Rendah	Rendah	Rendah	Sedang	Ya
8	Dewasa	Male	Rendah	Rendah	Rendah	Tinggi	Ya
9	Dewasa	Male	Rendah	Rendah	Rendah	Tinggi	Ya
10	Dewasa	Female	Normal	Rendah	Rendah	Tinggi	Ya
11	Dewasa	Female	Normal	Rendah	Rendah	Tinggi	Ya
12	Dewasa	Male	Normal	Rendah	Rendah	Tinggi	Ya
13	Dewasa	Male	Rendah	Rendah	Rendah	Tinggi	Ya
14	Dewasa	Female	Rendah	Rendah	Rendah	Rendah	Ya
15	Dewasa	Male	Normal	Rendah	Rendah	Tinggi	Ya
16	Dewasa	Male	Normal	Rendah	Rendah	Tinggi	Ya
17	Dewasa	Female	Normal	Rendah	Rendah	Sedang	Ya
18	Dewasa	Male	Normal	Rendah	Rendah	Rendah	Ya
19	Dewasa	Male	Normal	Rendah	Rendah	Sedang	Ya

Fig. 2. Public dataset obtained by the authors listed in the RapidMiner tool

㉓ Based on Figure 2, it can be seen that the data to be analyzed consists of several attributes that are relevant for dengue disease prediction. The sample data that the author got from the web data.mendeley.com amounted to 1003 samples and then cleaned the data from empty cells so that the sample became 989 samples. The attributes include ⑭ age status, sex, hemoglobin level, white blood cell count, platelet count, PDW (Platelet Distribution Width), and the final result which is the final output categorized in binary labels. These data are used in the processing process to build the prediction model, where each column is set with the appropriate data type, such as polynomial and binary data.

---

* Corresponding author

**Journal of Computer Networks, Architecture and
High Performance Computing**
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

In addition, the figure also shows that there is no problem importing data because the "Replace errors with missing values" feature has been selected to handle missing or incomplete data. All data fields are well prepared, indicating that the attributes are relevant and ready to be used in further analysis

DISCUSSIONS

After the dataset adjustment stage was carried out, the next step taken by the researcher was to conduct operator experiments using the Bootstrap Aggregating (Bagging) and Decision Tree C4.5 algorithms using the RapidMiner AI Studio software tools.

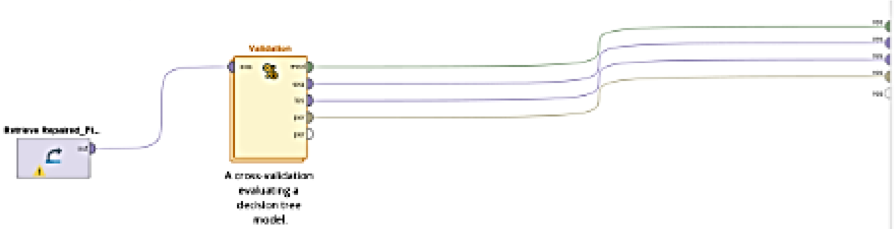


Fig. 3. Working on C4.5 Decision Tree with X-Validation operator

Figure 3 illustrates the use of the X-Validation operator in RapidMiner to evaluate the performance of a C4.5 decision tree model. This process involves dividing the data into subsets for training and testing purposes, which allows for more objective validation. Using X-Validation, the model is tested under various conditions, resulting in a more accurate evaluation of its robustness and predictive ability

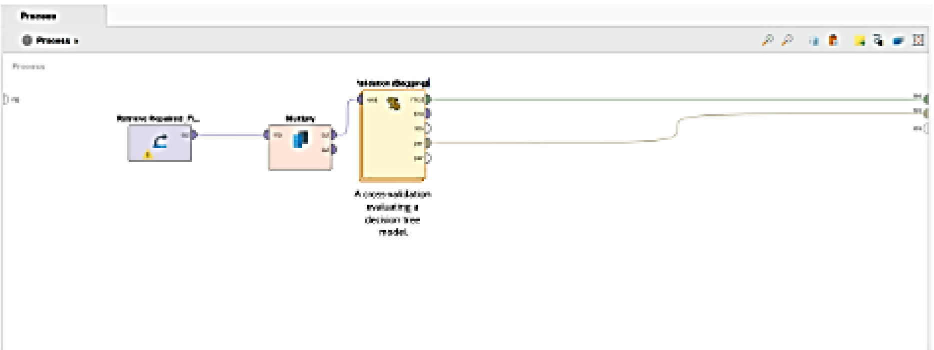


Fig. 4. Bagging + Decision Tree C4.5 with X-Validation and Multiply operators

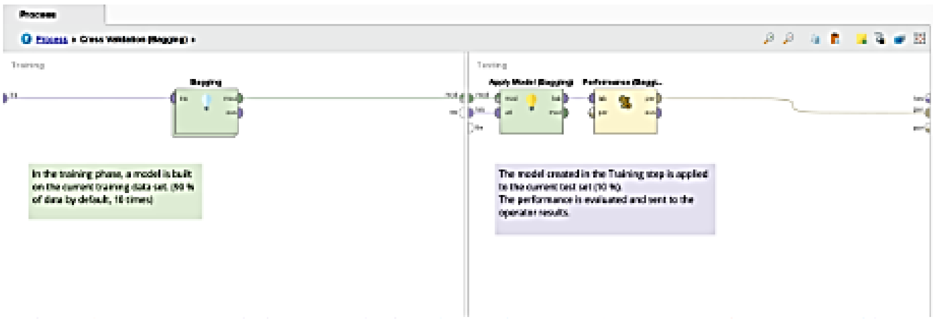
Figure 4 shows the application of the Bagging technique alongside the C4.5 Decision Tree, combined with the X-Validation and Multiply operators. This technique aims to improve model stability and accuracy by combining the results of multiple Decision Tree models trained on different subsets of data. This model fusion allows for a more comprehensive evaluation and reduces the risk of overfitting, which makes the model predictions more reliable

---

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx



In the training phase, a model is built on the current training data set. (90 % of data by default, 10 times)

The model created in the Training step is applied to the current test set (10 %).
The performance is evaluated and sent to the operator results.

Fig. 5. View after double-clicking of X-Validation with Bagging (left) and Performance (right) operators visible

Figure 5 shows the view obtained after double-clicking the X-Validation operator, with the Bagging operator on the left and the Performance operator on the right. The Bagging operator is used to generate multiple models from different data, while the Performance operator measures the predictive quality of the generated models. This performance evaluation is important to understand the extent to which the model can be effectively applied to data that has not been seen before

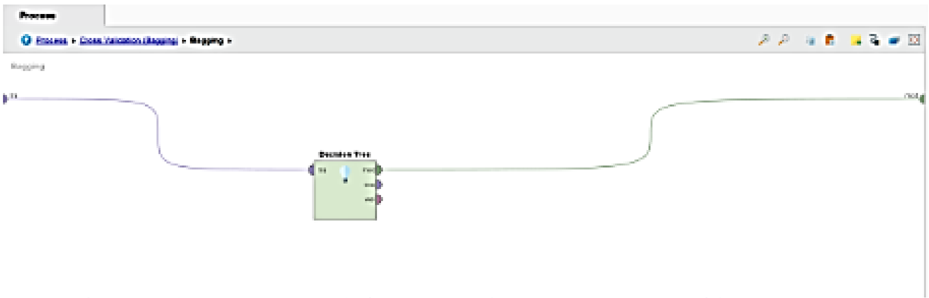


Fig. 6. The display after double-clicking of the Bagging operator shows the Decision Tree operator

Figure 6 displays the results after the Bagging operator has been double-clicked, showing the Decision Tree operator underlying the built model. At this stage, the user can examine the structure of the resulting decision tree and evaluate the selection of relevant attributes for prediction. The use of this operator enables optimization of the classification process by providing a deeper understanding of the data sharing and decision making in the model.

PERFORMANCE RESULT TABLE

Table 1. Performance Results of Bagging + Decision Tree C4.5 using RapidMiner

	True Ya	True Tidak	Class precision
Pred. Ya	666	0	100.00%
Pred. Tidak	3	320	99.07%
Class recall	99.55%	100.00%	

Accuracy: 99.70% +/- 0,49% (micro average: 99.70%)
Precision: 99.09% +/- 1.46% (micro average: 99.07%) (positive class: Tidak)
Recall: 100.00% +/- 0.00% (micro average: 100.00%) (positive class: Tidak)

Table 1 presents the performance evaluation results of the models built using the Bagging technique and the C4.5 Decision Tree algorithm in RapidMiner. The evaluation results include key metrics such as Accuracy, Precision, and

---

* Corresponding author

⑨ Recall. The model showed an Accuracy of 99.70%, with an uncertainty of 0.49%, which signifies a very high level of accuracy in data classification. Precision was recorded at 99.09%, with an uncertainty of 1.46%, indicating the model's excellent ability to accurately classify the "No" class. Recall reached a perfect value of 100%, indicating that the model successfully identified all positive cases without error. In addition, Class Precision showed a value of 100% for the "Yes" class and 99.07% for the "No" class, reinforcing the conclusion that the model has a very reliable performance in predicting DHF disease based on the existing dataset.

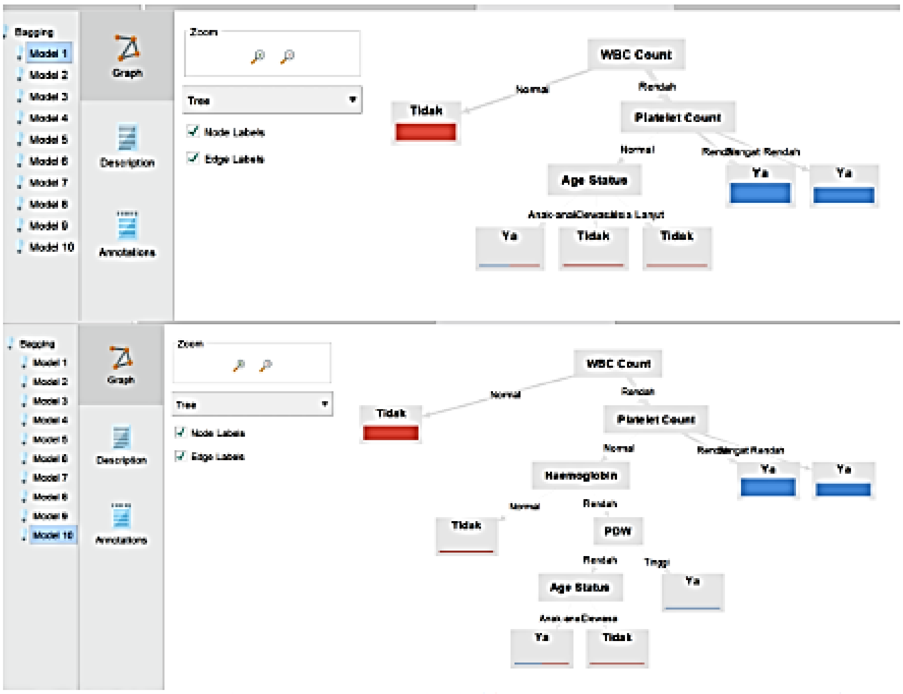


Fig. 7. Model 1 (top) and model 10 (bottom) display decision trees of several sequences, not all of which are attached to this journal article

Based on the display in Figure 6, Model 1 describes the initial decision on WBC Count (white blood cell count), where a Normal count results in a No classification. If the count is Low, the next criterion is Platelet Count, which determines whether the result is Yes or No based on additional classification criteria. The model also incorporates Age Status to refine the prediction. Similarly, Model 10, although following a similar sequence, provides a more detailed classification structure, with checks on all relevant health parameters to provide a more granular final prediction.

C4.5 Decision Tree is a widely utilized classification technique that employs decision trees for data categorization; nonetheless, it frequently encounters issues with overfitting, particularly in the presence of noisy datasets. Incorporating a Bagging (Bootstrap Aggregation) approach can diminish variance and enhance model stability, thus mitigating the danger of overfitting. Bagging constructs numerous Decision Tree models from randomly sampled training data and aggregates the outcomes, enhancing prediction accuracy, particularly in datasets with imbalanced classes.

---

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Comparisons with alternative models, such as Random Forest—an extension of Bagging utilizing numerous Decision Trees—demonstrate that Random Forest frequently excels due to its capacity to manage complicated and irregular data. Random Forest can attain accuracy over 90% on some datasets. The integration of Bagging with boosting algorithms like AdaBoost can enhance performance, but with the potential danger of overfitting on noisy datasets. Ensemble approaches, such as Bagging and Boosting, yield more accurate and predictive models than a solitary Decision Tree model.

This dengue disease prediction model using bagging algorithm and C4.5 decision tree provides various scientific contributions, namely: 1) Bagging improves prediction accuracy by reducing overfitting and increasing the stability of the results through combining multiple decision tree models. 2) C4.5 is effective in classification of numerical and categorical data, separating classes based on informative attributes. 3) The model has applications in public health, particularly in the early detection of DHF, helping the identification of high-risk areas for faster intervention, 4) Besides being applied to DHF, the model can be applied to the prediction of other diseases, improving the performance of the model in general. 5) the development of machine learning algorithms contributes to more accurate modeling of medical data.

C4.5 Decision Tree is one of the classification methods that uses a tree-based approach to classify data. This model is used to classify traffic accident data (Franseda et al., 2020). Research shows that Decision Tree can provide promising results in terms of classification accuracy, although it is often limited by overfitting issues, especially when faced with noisy datasets (Shi et al., 2022). However, the addition of ensemble algorithms such as Bagging has been shown to reduce the variance of these models, supporting more stable decisions and reducing the risk of overfitting (Alali et al., 2022). Bagging models have improved the accuracy of Decision Tree models, resulting in better performance than single Decision Tree models (Pristyanto & Zein, 2023); (Zhang et al., 2021).

In daily life, this model helps health officials predict the likelihood of dengue outbreaks more accurately. By monitoring climate change and mosquito populations, health experts can provide early warnings and take preventive measures, such as insecticide spraying and public awareness campaigns. This can improve the quality of life in dengue-endemic areas and reduce the economic burden associated with handling outbreaks(Dhanam B, 2024). The model is able to capture the complexity in epidemiological data and environmental variables such as temperature, humidity, and rainfall, which have a significant impact on the spread of dengue disease (Mazhar et al., 2024).

Some other studies relevant to this research are the research of Pristyanto & Zein, (2023) that Decision Tree and Balanced-Baggin models can improve accuracy and evaluation metrics such as geometric mean and area under the curve (AUC) in various datasets. In addition, it can be used to overcome datasets that have significant class imbalance in medical data, such as those containing age, sex (gender), hemoglobin levels, white blood cell count (WBC Count), and blood cell panel analysis (RBC PANEL) and platelet count. Bagging decision trees can improve prediction accuracy when compared to conventional models such as logit models. Bagging shows accuracy between 85% to 90% for better prediction (Sadorsky, 2021). Bagging in decision trees can overcome variation and improve classification accuracy (Ly et al., 2019).

CONCLUSION

This research developed a Dengue Fever (DHF) disease prediction model using Bagging and Decision Tree C4.5 algorithms, which resulted in high accuracy with 99.70% accuracy, 99.09% precision, and 100% recall. These findings demonstrate the effectiveness of the model in classifying and detecting dengue cases based on environmental factors such as temperature, humidity, and rainfall.

REFERENCES

Agustian, D. R., & Darmawan, B. A. (2022). Analisis Clustering Demam Berdarah Dengue Dengan Algoritma K-Medoids (Studi Kasus Kabupaten Karawang). *JIKO (Jurnal Informatika Dan Komputer)*, *6*(1), 18. https://doi.org/10.26798/jiko.v6i1.504

Alali, Y., Harrou, F., & Sun, Y. (2022). A proficient approach to forecast COVID-19 spread via optimized dynamic machine learning models. *Scientific Reports*, *12*(1), 2467. https://doi.org/10.1038/s41598-022-06218-3

Anwar, N., Pranolo, A., & Kurnaiwan, R. (2018). Grouping the community health center patients based on the disease characteristics using C4.5 decision tree. *IOP Conference*

* Corresponding author

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Series: Materials Science and Engineering, *403*(1). https://doi.org/10.1088/1757-899X/403/1/012084

Cholil, S. R., Dwijayanto, A. F., Ardianita, T., Studi, P., Informasi, S., & Komunikasi, D. (n.d.). *PREDIKSI PENYAKIT DEMAM BERDARAH DI PUSKESMAS NGEMPLAK SIMONGAN MENGGUNAKAN ALGORITMA C4.5*.

Dhanam B. (2024). Intelligent Systems for Efficiently Predicting and Managing Dengue Fever using Machine Learning Techniques. *Journal of Electrical Systems*, *20*(11s), 2757–2767. https://doi.org/10.52783/jes.7956

Duta, U., Surakarta, B., Fathoni, M. A., & Arta Bhrata, A. (n.d.). *Fakultas Sains dan Teknologi*.

Effendi, H., Syahrial, A., Prayoga, S., & Hidayat, W. D. (2021). Penerapan Metode K-Means Clustering untuk Pengelompokan Lahan Sawit Produktif pada PT Kasih Agro Mandiri. *Teknomatika*, *11*(02), 117–126. https://ojs.palcomtech.ac.id/index.php/teknomatika/article/view/543

Elgimati, Y. (2020). Weighted Bagging in Decision Trees: Data Mining. *JINAV: Journal of Information and Visualization*, *1*(1), 1–14. https://doi.org/10.35877/454ri.jinav149

Franseda, A., Kurniawan, W., Anggraeni, S., & Gata, W. (2020). Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, *8*(3), 282. https://doi.org/10.26418/justin.v8i3.40982

Handayani, I. (2019). Application of K-Nearest Neighbor Algorithm on Classification of Disk Hernia and Spondylolisthesis in Vertebral Column. *Indonesian Journal of Information Systems*, *2*(1), 57–66. https://doi.org/10.24002/ijis.v2i1.2352

Hu, X., Zeng, Y., Qin, C., & Meng, D. (2022). Bagging-based neural network ensemble for load identification with parameter sensitivity considered. *Energy Reports*, *8*, 199–205. https://doi.org/10.1016/j.egyr.2022.08.056

Kurniawan, A., & Prihandono, A. (2020). PENERAPAN TEKNIK BAGGING UNTUK MENINGKATKAN AKURASI KLASIFIKASI PADA ALGORITMA NAIVE BAYES DALAM MENENTUKAN BLOGGER PROFESIONAL. In *Jurnal Bisnis Digitasl dan Sistem Informasi* (Vol. 1).

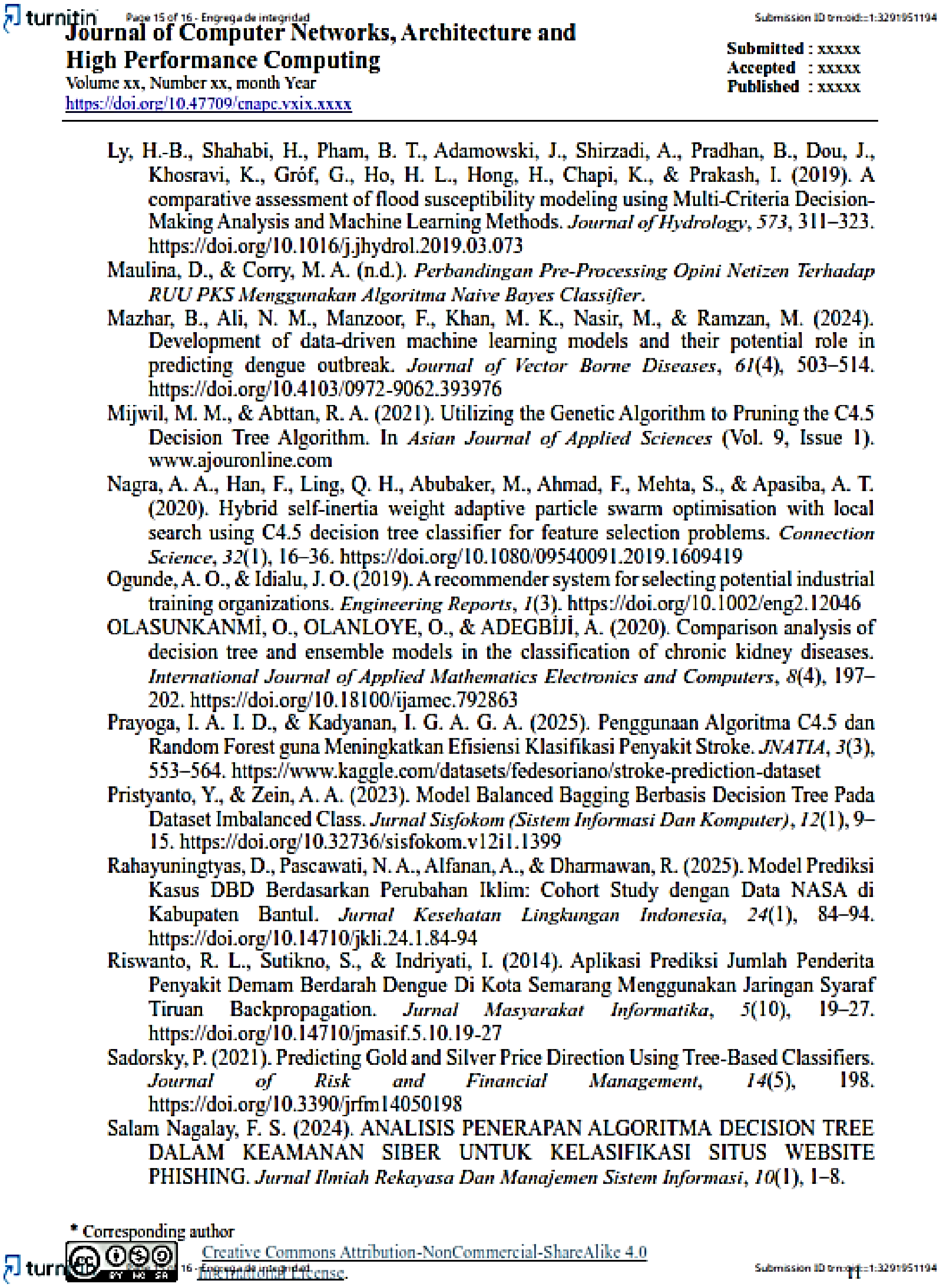
Laksono, T., Rizaldi, S. T., & Arifin, M. (2021). *MALCOM: Indonesian Journal of Machine Learning and Computer Science The Implementation of Markov Chain for Prediction on Chemistry of Hemorrhic Figure Case In Bengkalis District Implementasi Markov Chain untuk Prediksi Kasus Penderita Demam Berdarah Dengue di Kabupaten Bengkalis*. *1*, 17–23.

Latuconsina, S., & Matdoan, M. Y. (2022). Prediksi Penyebaran Kasus Demam Berdarah Dengue (DBD) Dengan Metode K-Nearest Neighbor di Kabupaten Maluku Tengah. *JURNAL ILMIAH MATEMATIKA DAN TERAPAN*, *19*(2), 254–261. https://doi.org/10.22487/2540766x.2022.v19.i2.15984

Lestari, S., & Ilmu Komputer, F. (2024). Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms. *Jurnal Dan Penelitian Teknik Informatika*, *8*(2). https://doi.org/10.33395/v8i2.13448

Lutfianawati, R. F., Ngadino, & Marlik. (2024). Prediksi Kejadian Demam Berdarah Dengue di Kecamatan Papar Kabupaten Kediri Tahun 2016–2021. *ASPIRATOR - Journal of Vector-Borne Diseases Studies*, *14*(1). https://doi.org/10.22435/asp.v14i1.5892

* Corresponding author



Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Ly, H.-B., Shahabi, H., Pham, B. T., Adamowski, J., Shirzadi, A., Pradhan, B., Dou, J., Khosravi, K., Grof, G., Ho, H. L., Hong, H., Chapi, K., & Prakash, I. (2019). A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *Journal of Hydrology*, *573*, 311–323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Maulina, D., & Corry, M. A. (n.d.). *Perbandingan Pre-Processing Opini Netizen Terhadap RUU PKS Menggunakan Algoritma Naive Bayes Classifier*.

Mazhar, B., Ali, N. M., Manzoor, F., Khan, M. K., Nasir, M., & Ramzan, M. (2024). Development of data-driven machine learning models and their potential role in predicting dengue outbreak. *Journal of Vector Borne Diseases*, *61*(4), 503–514. https://doi.org/10.4103/0972-9062.393976

Mijwil, M. M., & Abttan, R. A. (2021). Utilizing the Genetic Algorithm to Pruning the C4.5 Decision Tree Algorithm. In *Asian Journal of Applied Sciences* (Vol. 9, Issue 1). www.ajouronline.com

Nagra, A. A., Han, F., Ling, Q. H., Abubaker, M., Ahmad, F., Mehta, S., & Apasiba, A. T. (2020). Hybrid self-inertia weight adaptive particle swarm optimisation with local search using C4.5 decision tree classifier for feature selection problems. *Connection Science*, *32*(1), 16–36. https://doi.org/10.1080/09540091.2019.1609419

Ogunde, A. O., & Idialu, J. O. (2019). A recommender system for selecting potential industrial training organizations. *Engineering Reports*, *1*(3). https://doi.org/10.1002/eng2.12046

OLASUNKANMİ, O., OLANLOYE, O., & ADEGBİJİ, A. (2020). Comparison analysis of decision tree and ensemble models in the classification of chronic kidney diseases. *International Journal of Applied Mathematics Electronics and Computers*, *8*(4), 197–202. https://doi.org/10.18100/ijamec.792863

Prayoga, I. A. I. D., & Kadyanan, I. G. A. G. A. (2025). Penggunaan Algoritma C4.5 dan Random Forest guna Meningkatkan Efisiensi Klasifikasi Penyakit Stroke. *JNATIA*, *3*(3), 553–564. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

Pristyanto, Y., & Zein, A. A. (2023). Model Balanced Bagging Berbasis Decision Tree Pada Dataset Imbalanced Class. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *12*(1), 9–15. https://doi.org/10.32736/sisfokom.v12i1.1399

Rahayuningtyas, D., Pascawati, N. A., Alfanan, A., & Dharmawan, R. (2025). Model Prediksi Kasus DBD Berdasarkan Perubahan Iklim: Cohort Study dengan Data NASA di Kabupaten Bantul. *Jurnal Kesehatan Lingkungan Indonesia*, *24*(1), 84–94. https://doi.org/10.14710/jkli.24.1.84-94

Riswanto, R. L., Sutikno, S., & Indriyati, I. (2014). Aplikasi Prediksi Jumlah Penderita Penyakit Demam Berdarah Dengue Di Kota Semarang Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Masyarakat Informatika*, *5*(10), 19–27. https://doi.org/10.14710/jmasif.5.10.19-27

Sadorsky, P. (2021). Predicting Gold and Silver Price Direction Using Tree-Based Classifiers. *Journal of Risk and Financial Management*, *14*(5), 198. https://doi.org/10.3390/jrfm14050198

Salam Nagalay, F. S. (2024). ANALISIS PENERAPAN ALGORITMA DECISION TREE DALAM KEAMANAN SIBER UNTUK KELASIFIKASI SITUS WEBSITE PHISHING. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *10*(1), 1–8.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume xx, Number xx, month Year
https://doi.org/10.47709/cnapc.vxix.xxxx

Submitted : xxxxx
Accepted : xxxxx
Published : xxxxx

Saputri, N. D., Khalid, K., & Rolliawati, D. (2022). Comparison of Bagging and Adaboost Methods on C4.5 Algorithm for Stroke Prediction. *SISTEMASI*, *11*(3), 567. https://doi.org/10.32520/stmsi.v11i3.1684

Shi, Z.-Y., Hon, J.-S., Cheng, C.-Y., Chiang, H.-T., & Huang, H.-M. (2022). Applying Machine Learning Techniques to the Audit of Antimicrobial Prophylaxis. *Applied Sciences*, *12*(5), 2586. https://doi.org/10.3390/app12052586

Wahono, R. S., & Suryana, N. (2013). Combining particle swarm optimization based feature selection and bagging technique for software defect prediction. *International Journal of Software Engineering and Its Applications*, *7*(5), 153–166. https://doi.org/10.14257/ijseia.2013.7.5.16

Yaman, M. A., Rattay, F., & Subasi, A. (2021). Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Face Recognition. *Procedia Computer Science*, *194*, 202–209. https://doi.org/10.1016/j.procs.2021.10.074

Yohannes Cristanto, R., Zamzami, E., & Fahmi, F. (2024, February 26). *Improved Decision Tree Accuracy (C4.5) with Attribute Reduction Using Forward Selection in Data Classification*. https://doi.org/10.4108/eai.21-9-2023.2342901

Zhang, T., Fu, Q., Wang, H., Liu, F., Wang, H., & Han, L. (2021). *Bagging Based Machine Learning Algorithms for Landslide Susceptibility Modeling*. https://doi.org/10.21203/rs.3.rs-553248/v1

* Corresponding author