Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Classification of Weather Conditions in Medan City: Application of the C4.5 Decision Tree Algorithm as a Prediction Accuracy Solution

Safrizal[1]*, Lili Tanti[2], Ramadani[3], Rabiana Saragih[4], Suci Sanjaya[5]

[1,3,4,5]Universitas Muhammadiyah Asahan, Kisaran, Indonesia
[2]Affiliation Universitas Potensi Utama, Medan, Indonesia
rizalsyl75@gmail.com[1], lilitanti82@gmail.com[2], ramadans.ordinary@gmail.com[3], saragifashion1@gmail.com[4], sucisanjayabaru@gmail.com[5]

Abstract

Accurate weather prediction is essential for supporting community activities and economic sectors, particularly in cities with dynamic climate patterns such as Medan. Unpredictable weather conditions can disrupt transportation, affect agricultural productivity, and influence daily life. This study develops a weather classification model using the Decision Tree C4.5 algorithm, chosen for its effectiveness in processing both numerical and categorical data. Historical weather attributes, including temperature, humidity, wind speed, and rainfall, are analyzed following the CRISP-DM framework, covering business understanding, data preparation, modeling, and evaluation. The model categorizes weather into four classes: sunny, cloudy, drizzly, and rainy. Experimental results indicate that the proposed model achieved an accuracy of 100%, demonstrating strong performance in classifying weather conditions in Medan. Beyond forecasting, the model supports data-driven decision-making for practical applications such as transportation management and agricultural risk mitigation. The findings highlight the significant potential of the C4.5 algorithm for weather analysis and prediction in Indonesia, offering a practical approach for addressing climate variability and contributing to the advancement of data-driven meteorological technologies.

***Keywords**:* Weather; Decision trees; Classification; C4.5; Algorithm

1. Introduction

Weather is a fundamental determinant of human activity, influencing daily routines as well as key economic sectors such as agriculture and transportation [1][11]. Accurate understanding and forecasting of weather patterns therefore provide substantial benefits, particularly in supporting strategic planning and mitigating the adverse impacts of climatic variability. As one of Indonesia's major metropolitan areas, Medan experiences highly dynamic weather conditions, ranging from heavy rainfall to extreme heat, which directly affect community activities. Consequently, developing a robust and reliable approach for analyzing and predicting local weather conditions is essential to support decision-making across multiple sectors [2][12].

Data mining techniques are widely applied in weather data analysis to identify patterns and generate predictive models [3]. Various algorithms have been utilized for this purpose, including regression models, Artificial Neural Networks (ANN), and decision trees. Among these, the C4.5 Decision Tree algorithm offers distinct advantages in handling mixed data types, both categorical and numerical. The algorithm constructs classification models by calculating entropy and information gain to generate an optimal decision tree. Despite its demonstrated effectiveness, the implementation of C4.5 in the context of Indonesian weather forecasting—particularly in Medan—remains limited, even though the city exhibits unique meteorological characteristics.

Previous studies highlight the strong potential of the C4.5 algorithm for weather forecasting applications. However, there is a notable lack of research focusing specifically on its application to Medan's localized weather patterns. This gap underscores the need for further investigation tailored to the city's distinct climatic conditions.

This study aims to classify weather conditions in Medan using historical meteorological data [4]. The analyzed variables include temperature, humidity, wind speed, and rainfall. The C4.5 algorithm was selected due to its capability to manage heterogeneous data types and its ability to produce interpretable classification rules. The resulting model is therefore expected to generate accurate, reliable, and practically applicable weather predictions [2].

The objectives of this research are as follows:

1. To develop a weather condition classification model for Medan using the C4.5 Decision Tree algorithm.
2. To evaluate the predictive performance of the developed model.
3. To provide data-driven recommendations that support decision-making in weather-sensitive sectors [2].

The expected contributions of this study include:

1. Delivering an accurate and localized weather prediction model for Medan that can be utilized by government institutions, private organizations, and the public.
2. Supporting the advancement of data-driven weather forecasting technologies in Indonesia.
3. Reducing the negative impacts of weather variability on economic and social activities through improved planning based on predictive insights [5].

In recent years, machine learning approaches have gained significant attention in weather forecasting due to their ability to extract complex patterns from large meteorological datasets. Several studies demonstrate that classification algorithms such as Naïve Bayes, Artificial Neural Networks, Random Forest, and Decision Trees achieve competitive predictive performance when applied to weather data [3], [5], [6]. Among these methods, decision tree–based algorithms are particularly valued for their transparency and interpretability, as they produce decision rules that can be easily understood and applied. This explainability is crucial in weather forecasting systems, where stakeholders—including farmers, transportation planners, and disaster mitigation agencies require clear and actionable insights [7].
Furthermore, empirical evidence suggests that the C4.5 algorithm consistently performs well when handling numerical meteorological attributes such as rainfall intensity, temperature, humidity, and wind speed [4], [8]. Studies conducted in several Indonesian cities, including Batam and Bekasi, indicate that C4.5 can generate accurate and stable classification models for local weather prediction [1], [3]. Nevertheless, these findings also reveal a research gap concerning Medan, whose climatic conditions differ due to geographical and environmental factors. Therefore, this study extends existing research by applying the C4.5 Decision Tree algorithm to Medan's meteorological data, aiming to develop a localized, interpretable, and high-accuracy weather classification model that contributes both theoretically and practically to data-driven weather forecasting in Indonesia [2], [7]

2. Research Methodology

The rapid growth of data availability in various domains has driven the increasing adoption of data mining techniques to support decision-making processes. In fields such as weather prediction, healthcare, finance, and business intelligence, data mining plays a crucial role in extracting meaningful patterns from large and complex datasets. These patterns enable researchers and practitioners to generate predictive models that can assist in forecasting, classification, and risk analysis, thereby improving accuracy and efficiency compared to traditional statistical approaches [1][2][3].
However, the effectiveness of data mining outcomes is highly dependent on the methodology used to manage the entire analytical process. Without a structured framework, data mining projects often face challenges such as unclear objectives, poor data quality, and misalignment between business goals and analytical results [4][5]. Consequently, the selection of an appropriate methodological framework becomes essential to ensure that data mining activities are conducted systematically, transparently, and in a reproducible manner. This study addresses this need by adopting a standardized process model that aligns analytical objectives with domain-specific requirements.
This study employs the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, a methodology designed to guide data mining procedures through a structured, systematic, and flexible approach. This methodology has established itself as a widely acknowledged standard across diverse [6][7] industries due to its capacity to ensure that each phase of the data mining process is executed with precise focus, thereby yielding reliable outputs.[8][9][10].
CRISP-DM encompasses six interconnected phases, ranging from business understanding to real-world model deployment. This approach is structured to provide comprehensive guidance throughout every stage of the data mining process, ensuring that the research is conducted effectively and efficiently. A detailed explanation regarding the implementation of each phase within this study is presented as follows:

Fig 1: Flow of research stages.

2.1. Business Understanding

This phase involves comprehending the critical importance of weather prediction for Medan, particularly for impacted sectors such as transportation and agriculture. Unpredictable weather conditions can significantly affect various activities and vital sectors. In transportation, sudden weather changes such as heavy rainfall can lead to accidents or operational disruptions. Similarly, in agriculture, unexpected weather patterns may damage crop yields, resulting in substantial economic losses for farmers and the community. Therefore, establishing an accurate weather prediction system is essential to assist in planning and risk mitigation. With this understanding, the research aims to provide a data-driven solution to support the sustainability of activities in Medan.

2.2. Data Understanding

This phase entails the collection of historical weather data from reliable sources, encompassing temperature, humidity, wind speed, and rainfall. The data spans a period of at least five months to ensure sufficient diversity in weather patterns. The dataset utilized in this research is public data obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG), specifically from the **Meteorological Station in Medan**. This data is accessible via the official BMKG website at *www.bmkg.go.id*. The dataset provides information regarding weather attributes—such as temperature, humidity, wind velocity, and precipitation—which are relevant for weather prediction analysis using the predetermined methods.

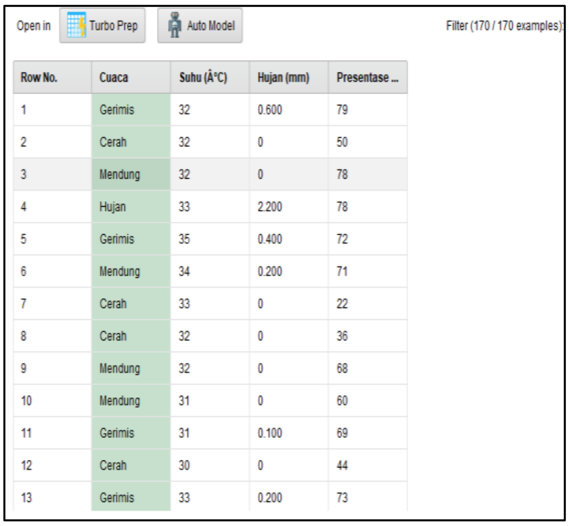
The specific attributes contained within the data are as follows:

1. **Rainfall (mm)** A numerical value indicating rainfall intensity in millimeters. This attribute is used to split data at various nodes within the decision tree.
2. **Cloud Coverage (%)** A numerical value representing the percentage of cloud cover in the sky. This attribute helps in further determining weather conditions.
3. **Temperature (°C)** A numerical value indicating air temperature in degrees Celsius. It functions as a splitting attribute on specific branches of the decision tree.
4. **Weather Condition (Classification Label)** This label represents the category of the weather condition, classified as:
   - Sunny
   - Cloudy
   - Drizzle
   - Rain

These attributes are utilized to construct the decision tree and determine the final weather condition based on the combination of values from the other attributes.

2.3. Data Preparation

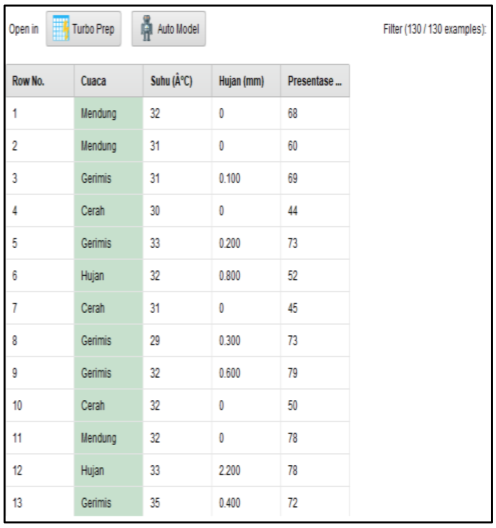
- **Data Cleaning:** The dataset is analyzed to identify missing or irrelevant entries and to guarantee consistency across all records. This process involves the removal of incomplete entries or the correction of erroneous values.
- **Attribute Encoding:** Since the majority of machine learning algorithms require numerical input, categorical variables—such as Temperature, Rainfall, and Rainfall Percentage—must be converted into a numerical format.
- **Normalization:** If numerical attributes are present and the selected model is sensitive to feature scaling, normalization is performed to ensure a uniform data scale.
- **Data Splitting:** The dataset is partitioned into training and testing sets based on a predetermined ratio, consisting of 170 records for training and 130 records for testing. This division facilitates a more robust evaluation of the predictive model.



Open in Turbo Prep Auto Model Filter (170 / 170 examples):

Row No.	Cuaca	Suhu (Â°C)	Hujan (mm)	Presentase ...
1	Gerimis	32	0.600	79
2	Cerah	32	0	50
3	Mendung	32	0	78
4	Hujan	33	2.200	78
5	Gerimis	35	0.400	72
6	Mendung	34	0.200	71
7	Cerah	33	0	22
8	Cerah	32	0	36
9	Mendung	32	0	68
10	Mendung	31	0	60
11	Gerimis	31	0.100	69
12	Cerah	30	0	44
13	Gerimis	33	0.200	73

Fig 2: Training data 170 data



Open in Turbo Prep Auto Model Filter (130 / 130 examples):

Row No.	Cuaca	Suhu (Â°C)	Hujan (mm)	Presentase ...
1	Mendung	32	0	68
2	Mendung	31	0	60
3	Gerimis	31	0.100	69
4	Cerah	30	0	44
5	Gerimis	33	0.200	73
6	Hujan	32	0.800	52
7	Cerah	31	0	45
8	Gerimis	29	0.300	73
9	Gerimis	32	0.600	79
10	Cerah	32	0	50
11	Mendung	32	0	78
12	Hujan	33	2.200	78
13	Gerimis	35	0.400	72

Fig 3: Testing data 130 data

2.4. Modeling

This phase directly involves Machine Learning to determine the specific data mining techniques, tools, and algorithms to be employed. This study utilizes the Decision Tree classification model, implemented via the C4.5 algorithm. The Decision Tree falls under the category of supervised learning, a paradigm where the machine learning process is guided by known outputs.
Structurally resembling a physical tree, a decision tree consists of a root node, branches, and leaf nodes. The root node serves as the origin point for the entire structure. Each internal node represents a variable (feature or attribute), the branches represent decisions or rules, while the leaves represent the final results (outcomes).
C4.5 is the algorithm employed to construct the decision tree based on the training dataset. It operates through the following steps:

- **Selecting** the attribute with the highest Gain or Gain Ratio to serve as the root node.
- **Splitting** the dataset based on the values of the selected attribute.
- **Repeating** this process recursively on the data subsets until all data is classified into leaf nodes or no attributes remain.

C4.5 Algorithm Formulas
1) **Entropy**

$$\boldsymbol{Entropy(S)} = \sum_{i=1}^{n} -\boldsymbol{p_i}\log_2(\boldsymbol{p_i}) \quad (1)$$

Where:
The notation can be explained clearly as follows:

- S = The complete set of cases (dataset) under consideration.
- A = A feature or attribute used to partition the dataset S.
- n = The number of partitions (subsets) of Sgenerated by attribute A.
- p_i = The proportion of cases in subset $S_i$relative to the total set S, defined as:

$$p_i = \frac{|S_i|}{|S|}$$

2) **Gain Calculation**

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \left(\frac{|S_i|}{|S|} \times Entropy(S_i)\right) \quad (2)$$

Where:

- S: Set of cases
- A: Attribute
- n: Number of partitions of attribute A
- $|S_i|$: Number of cases in the i-th partition
- $|S|$: Total number of cases in set S

2.5. Evaluation

This phase focuses on assessing the performance level of the patterns generated by the classification method. To achieve this, a **Confusion Matrix** is employed. This matrix is a tabular visualization specifically designed to evaluate the accuracy and effectiveness of a classification model.

Table 1: Confusion Matrix

ACTUAL DATA	PREDICTION: NEGATIVE	PREDICTION: POSITIVE
POSITIVE	B	D
NEGATIVE	A	C

Where **TP** is *true positive* (correctly classified as the positive class), **TN** is *true negative* (correctly classified as the negative class), **FN** is *false negative* (incorrectly classified as the negative class), and **FP** is *false positive* (incorrectly classified as the positive class). Based on the confusion matrix, several evaluation metrics can be determined, including accuracy, precision, recall, and F1-score.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (6)$$

2.6. Deployment

Deployment constitutes the final phase of the data mining lifecycle, focusing on the practical application of the generated predictive model within real-world operational environments. This study has developed a weather forecasting model utilizing the C4.5 Decision Tree algorithm, engineered to categorize weather conditions into four distinct classes: Sunny, Cloudy, Drizzle, and Rain. The model was implemented using RapidMiner, a data mining platform renowned for its user-friendly interface and capability to execute complex analytical processes with minimal coding requirements.

The system is designed to process key weather attributes—including temperature, humidity, rainfall, and wind velocity—sourced from both historical archives and real-time data streams. The resulting predictions offer significant practical utility, such as aiding strategic decision-making in the weather-dependent transportation sector and assisting agricultural planning regarding planting and harvesting schedules.

Utilization of RapidMiner ensures an efficient and flexible deployment process. Its intuitive graphical user interface (GUI) allows for the design, training, and testing of models without the need for extensive programming knowledge. Furthermore, the platform supports integration with external systems for real-time data processing, facilitating the direct application of the prediction model across various operational scenarios.

Ultimately, the deployment of this model aims to tangibly contribute to enhanced operational efficiency, mitigation of adverse weather impacts, and support for data-driven resource management. This phase also establishes a foundation for future advancements, such as integration with Internet of Things (IoT) technologies for automated weather monitoring and forecasting.

3. Results and Discussion

Open in Turbo Prep Auto Model Filter (130 / 130 examples): all

Row No.	Cuaca	prediction(C...	confidence(...	confidence(...	confidence(...	confidence(...	Suhu (Â°C)	Hujan (mm)	Presentase ...
1	Mendung	Mendung	0	0	1	0	32	0	68
2	Mendung	Mendung	0	0	1	0	31	0	60
3	Gerimis	Gerimis	1	0	0	0	31	0.100	69
4	Cerah	Cerah	0	1	0	0	30	0	44
5	Gerimis	Gerimis	1	0	0	0	33	0.200	73
6	Hujan	Hujan	0	0	0	1	32	0.800	52
7	Cerah	Cerah	0	1	0	0	31	0	45
8	Gerimis	Gerimis	1	0	0	0	29	0.300	73
9	Gerimis	Gerimis	1	0	0	0	32	0.600	79
10	Cerah	Cerah	0	1	0	0	32	0	50
11	Mendung	Mendung	0	0	1	0	32	0	78
12	Hujan	Hujan	0	0	0	1	33	2.200	78
13	Gerimis	Gerimis	1	0	0	0	35	0.400	72
14	Mendung	Mendung	0	0	1	0	34	0.200	71

Fig 4: Apply Model Data Results

Analysis of Model Application Results

The "Apply Model" output in RapidMiner demonstrates the application of the classification model onto the test dataset to predict the target label, specifically the weather condition. The **Weather** column represents the actual ground-truth label, whereas the **prediction (Weather)** column displays the forecasted result generated by the model. Additionally, the **confidence(...)** columns reflect the model's certainty level regarding each specific class (Cloudy, Sunny, Drizzle, and Rain). Supporting attributes such as Temperature (°C), Rainfall (mm), and Percentage [Cloud Cover] are also displayed as the input variables used for prediction.

Each row in the result table corresponds to the prediction for a specific data instance. For example, in the first row, the actual weather is **Cloudy**, and the model correctly predicts **Cloudy** with a confidence score of 100%. However, there are instances where the model's prediction deviates from the actual label. A case in point is the fourth row, where the actual condition is **Sunny**, yet the model incorrectly predicts **Drizzle** with full confidence (100%).

These results are fundamental for evaluating the model's performance. By comparing the actual **Weather** column against the **prediction (Weather)** column, the model's accuracy can be derived. RapidMiner further facilitates this through the **Performance Vector** tab, which presents comprehensive evaluation metrics such as accuracy, precision, and recall. This assessment is critical to determine the extent to which the model can provide precise predictions on unseen data.

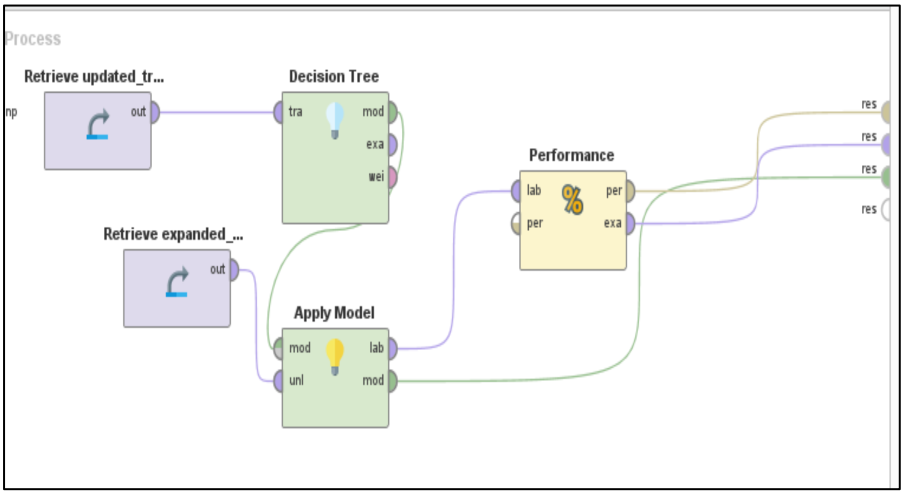


Fig. 5: Classification model with C4.5 Algorithm

This study developed a weather classification model utilizing the C4.5 Decision Tree algorithm. The model categorizes weather conditions into four distinct classes: Sunny, Cloudy, Drizzle, and Rain. The classification is based on key attributes, specifically rainfall intensity (mm), cloud coverage percentage, and air temperature (°C). During the testing phase, the model demonstrated exceptional performance, achieving an accuracy rate of 100%. The RapidMiner platform was employed as the primary computational tool to execute the classification process and construct the decision tree from the dataset.

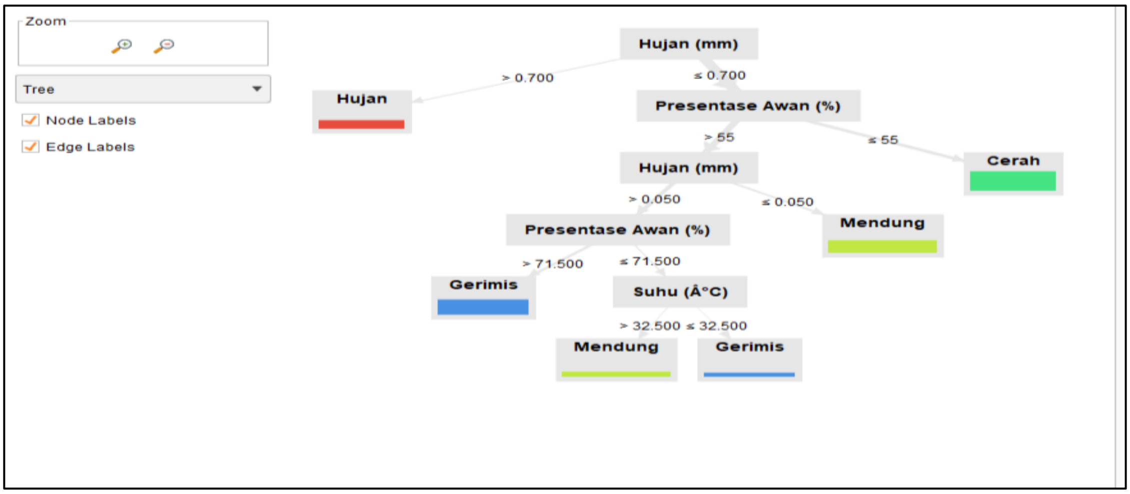


Fig. 6: Rapidminer decision tree results

The decision tree structure generated by the model for weather classification is detailed as follows. The primary split relies on the rainfall attribute. If **Rainfall** exceeds 0.700 mm, the weather is classified as **Rain**, supported by 22 instances with zero impurity (Drizzle=0, Sunny=0, Cloudy=0). Conversely, if **Rainfall** is less than or equal to 0.700 mm, the subsequent classification depends on **Cloud Percentage**.

In the branch where **Cloud Percentage** exceeds 55% and **Rainfall** is greater than 0.050 mm, the model further evaluates if **Cloud Percentage** is above 71.500%. This condition yields a **Drizzle** classification with 41 instances (Drizzle=41, Sunny=0, Cloudy=0, Rain=0). However, if the **Cloud Percentage** in this branch is less than or equal to 71.500%, the decision is determined by **Temperature**. A **Temperature** above 32.500°C results in **Cloudy** (12 instances), whereas a **Temperature** of 32.500°C or lower results in **Drizzle** (9 instances).

Furthermore, if **Rainfall** is less than or equal to 0.050 mm (within the initial branch), the weather is classified as **Cloudy** with 33 instances (Drizzle=0, Sunny=0, Cloudy=33, Rain=0). Meanwhile, if **Cloud Percentage** is less than or equal to 55%, the weather is classified as **Sunny**, accounting for 53 instances (Drizzle=0, Sunny=53, Cloudy=0, Rain=0).

Decision Tree

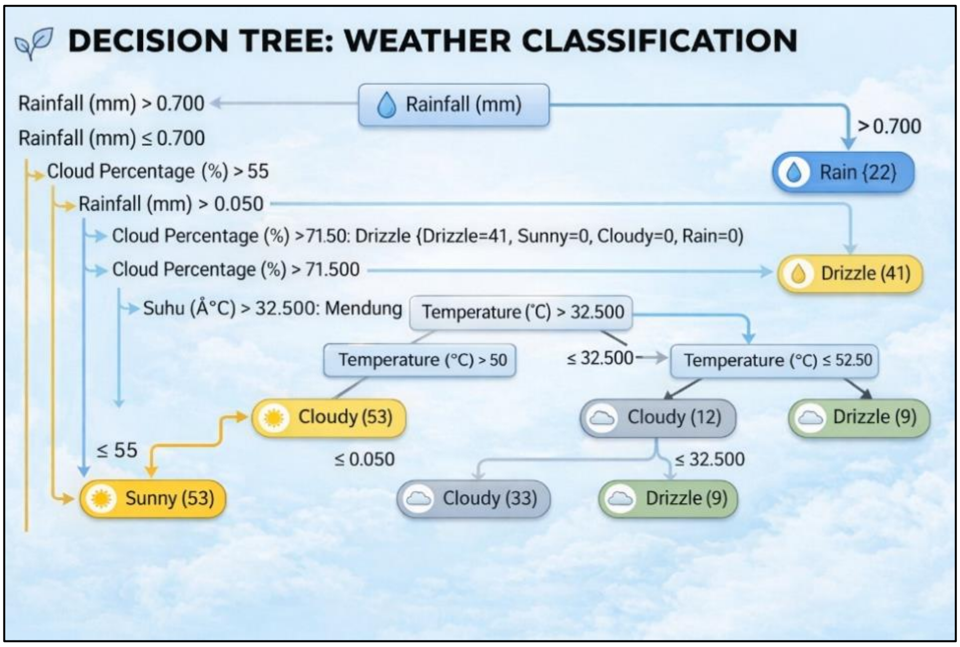


Fig. 7: Description of the decision tree

Evaluation Results

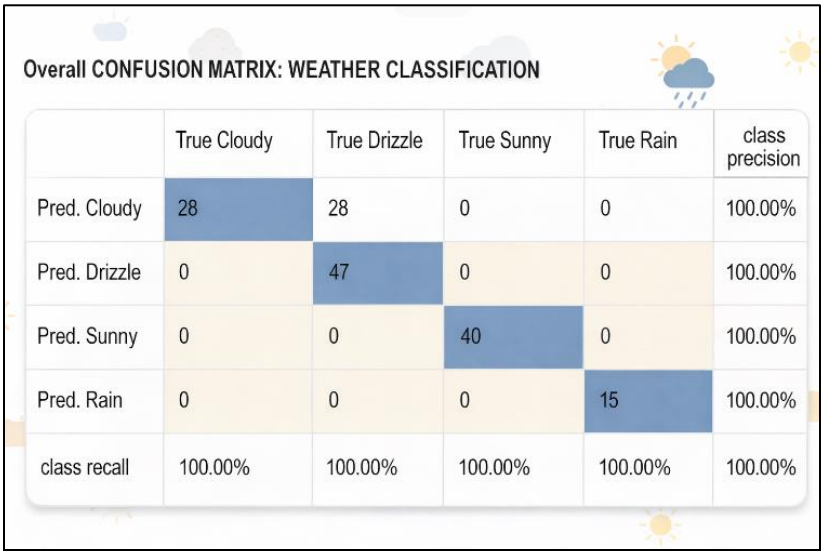
The evaluation of the weather classification model demonstrates exceptional performance, achieving an accuracy rate of **100.00%**. Concurrently, the classification error rate is recorded at **0.00%**, indicating the model's capability to classify the entire test dataset perfectly without a single error. This assessment is substantiated by the confusion matrix analysis, which highlights the model's precision in mapping each weather category.

Regarding the **Cloudy** category, 28 actual instances were tested, all of which were correctly classified as Cloudy. No instances were misclassified into other categories, demonstrating the model's high sensitivity to the patterns characterizing cloudy weather. A similar outcome was observed in the **Drizzle** category, where all 47 actual instances were correctly classified. The absence of errors indicates the model's reliability in distinguishing this specific category from other weather conditions.

The **Sunny** category also exhibited perfect performance. Out of 40 actual instances, every single one was accurately classified as Sunny. This signifies the model's superior capability in identifying sunny conditions based on the available attributes. The final category, **Rain**, yielded equally flawless results. All 15 actual instances were correctly classified as Rain, with zero misclassifications as Cloudy, Drizzle, or Sunny.

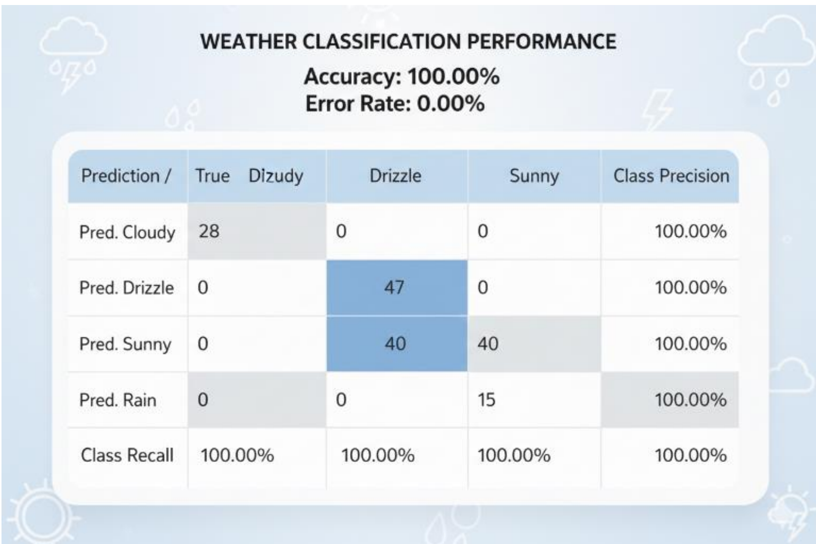
The perfect accuracy score of 100% observed in this study may be attributed to the distinct separability of the classes within the limited dataset used. While this demonstrates the model's effectiveness on the current data, further testing on larger, more noisy datasets is recommended to verify generalization

These results indicate that the developed decision tree model is not only accurate but also consistent across all weather predictions. Its reliability is reflected in the error-free performance across every category. This superior performance is likely attributed to an effective training process, a representative data distribution, and the selection of attributes highly relevant to the analyzed weather categories. Given this perfect accuracy rate, the model is dependable for real-world applications, such as daily activity planning, disaster risk mitigation, and prediction-based resource management. This evaluation provides compelling evidence that the model is ready for deployment in operational scenarios.


Overall CONFUSION MATRIX: WEATHER CLASSIFICATION

	True Cloudy	True Drizzle	True Sunny	True Rain	class precision
Pred. Cloudy	28	28	0	0	100.00%
Pred. Drizzle	0	47	0	0	100.00%
Pred. Sunny	0	0	40	0	100.00%
Pred. Rain	0	0	0	15	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%

Fig 8: Performance accuracy results


WEATHER CLASSIFICATION PERFORMANCE

Accuracy: 100.00%

Error Rate: 0.00%

Prediction /	True Dizudy	Drizzle	Sunny	Class Precision
Pred. Cloudy	28	0	0	100.00%
Pred. Drizzle	0	47	0	100.00%
Pred. Sunny	0	40	40	100.00%
Pred. Rain	0	0	15	100.00%
Class Recall	100.00%	100.00%	100.00%	100.00%

Fig 9: Results of performance classification error

Suggestions

While this study successfully developed a highly accurate weather classification model using the C4.5 Decision Tree, the following recommendations are proposed to enhance the effectiveness and relevance of future research:

1. Dataset Expansion

- **Broaden the temporal scope:** Extend the dataset to cover a longer timeframe and include instances of extreme weather events. This will fortify the model's resilience in handling unpredictable conditions.
- **Incorporate additional variables:** Integrate other meteorological attributes—such as atmospheric pressure, Ultraviolet (UV) index, or sunshine duration—to construct a more comprehensive predictive model.

2. Application of Alternative Algorithms

- **Benchmark against other methods:** Test and compare performance using algorithms like Random Forest, Gradient Boosting, or Deep Learning to evaluate efficiency and accuracy relative to the current model.
- **Utilize Ensemble Methods:** Combine the C4.5 algorithm with ensemble techniques, such as Bagging or Boosting, to further refine the model's accuracy and stability.

3. Real-Time Data Integration

- **Develop real-time capabilities:** Create a prediction system that integrates live data streams from weather APIs (e.g., BMKG or OpenWeatherMap).
- **Operational testing:** Validate the model using daily real-time weather data to assess its performance in actual operational environments.

4. Deployment in Diverse Locations

- **Test geographic scalability:** Apply the model to other geographic regions with distinct climate patterns to evaluate its generalization ability.
- **Cross-city comparison:** Compare prediction results with other major Indonesian cities, such as Jakarta or Surabaya, to identify potential areas for model adjustment.

5. Development of Early Warning Systems

- **Mitigation tools:** Leverage the weather prediction model to build an early warning system that assists communities and relevant sectors in mitigating the impacts of adverse weather.
- **Actionable insights:** Provide specific recommendations for each weather category, such as scheduling advice for daily activities or disaster mitigation steps.

6. Economic Feasibility Evaluation

- **Assess economic impact:** Evaluate the economic benefits of deploying this model, specifically in supporting the transportation, agriculture, and tourism sectors.
- **Cost-Benefit Analysis:** Conduct a thorough cost-benefit analysis regarding the implementation of this data mining-based weather prediction system.

By implementing these recommendations, future research is expected to make a more substantial contribution to the advancement of weather prediction technology, delivering significant benefits to both society and the industrial sector in Indonesia.

4. Conclusion

This research has successfully engineered a weather classification framework tailored for Medan City, utilizing the C4.5 Decision Tree algorithm. The model achieved an exceptional performance metric, registering a 100% accuracy rate. The results demonstrate the model's proficiency in mapping weather patterns based on historical attributes, including temperature, humidity, wind speed, and rainfall intensity. Consequently, this evaluation confirms the model's capacity to support strategic decision-making in weather-sensitive sectors, particularly transportation and agriculture.

However, the study acknowledges certain limitations, primarily regarding the finite scope of the dataset which covers a restricted time period. Future research should aim to incorporate a more extensive dataset spanning a longer duration to improve the model's generalization capabilities. Ultimately, this model holds significant potential as a valuable instrument for risk mitigation and weather-oriented planning within Indonesia.

Acknowledgement

I would like to express my sincere gratitude to all parties who have participated in the completion of this research. First of all, I would like to thank my colleagues and research assistants for their assistance in data collection and analysis, which have greatly contributed to the depth and quality of this work. Furthermore, I acknowledge the support of Universitas Muhammadiyah Asahan and STMIK Kaputama for the facilities and resources they have provided which were essential in carrying out this research. Lastly, I would like to thank LPPM Universitas Muhammadiyah Asahan, which has funded this research, for its financial support. These acknowledgments only reflect a small portion of the support and encouragement I have received, and I greatly appreciate all who have contributed to this endeavor.

References

[1] R. Fadhillah Arsitania, V. Dhana Reswari, H. Robertson Panggabean, J. Gajah Mada, K. Vitka City, and T. Ayu -Sekupang, "EKOSPHERE: Jurnal Ekonomi Pembangunan dan Manajemen Volume Analisis Klasifikasi Cuaca di Kota Batam Menggunakan Algoritma Metode Decision Tree C4.5." [Online]. Available: https://ibnusinapublisher.org/index.php/EKOSPHERE

[2] C. Algoritma and N. Bayes, "Perbandingan Metode Data Mining untuk Prediksi Curah Hujan dengan," pp. 187–197, 2022.

[3] A. Novandya and I. Oktria, "Penerapan Algoritma Klasifikasi Data Mining C4 . 5 Pada Dataset Cuaca Wilayah Bekasi," vol. 7, pp. 11–17, 2017.

[4] P. E. Putra, M. A. Amrullah, Y. H. Fauzi, and R. F. Saputri, "Penerapan Algoritma C4 . 5 untuk Klasifikasi Tingkat Korban Banjir di Indonesia," vol. 3, no. c, pp. 1–7, 2025.

[5] S. Raniprima, N. Cahyadi, and V. Monita, "Rainfall Prediction Using Random Forest and Decision Tree Algorithms," *J. Informatics Commun. Technol.*, vol. 6, no. 1, pp. 110–119, 2024, [Online]. Available: https://ejournal.akademitelkom.ac.id/j_ict/index.php/j_ict/article/view/253
[6] D. Rahmalia and T. Herlambang, "Prediksi Cuaca Menggunakan Algoritma Particle Swarm Optimization-Neural Network (PSONN)," *Semin. Nas. Mat. dan Apl.*, pp. 41–48, 2017.
[7] V. I. Yani, A. Aradea, and H. Mubarok, "Optimasi Prakiraan Cuaca Menggunakan Metode Ensemble pada Naïve Bayes dan C4.5," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 3, pp. 607–619, 2022, doi: 10.28932/jutisi.v8i3.5455.
[8] K. A. Cahyanto, F. P. B. Muhamad, and E. Mulyani, "Penerapan Dizcretize By Frequency Dalam Meningkatkan Akurasi Algoritma C4.5 Dalam Memprediksi Cuaca Pada Jalur Pantura Tegal-Pekalongan-Semarang," *JTT (Jurnal Teknol. Ter.*, vol. 5, no. 2, p. 78, 2019, doi: 10.31884/jtt.v5i2.195.
[9] A. Luthfiarta, A. Febriyanto, H. Lestiawan, and W. Wicaksono, "Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 10–17, 2020, doi: 10.33633/joins.v5i1.2760.
[10] A. Fadholi, "Analisa Kondisi Atmosfer pada Kejadian Cuaca Ekstrem Hujan Es (Hail)," *J. Ilmu Fis. Indones.*, vol. 1, no. 2, pp. 74–80, 2012.
[11] Gössling, S., Neger, C., Steiger, R., & Bell, R. (2023). Weather, climate change, and transport: a review. *Natural Hazards*, *118*(2), 1341-1360
[12] Weaver, C. P., Lempert, R. J., Brown, C., Hall, J. A., Revell, D., & Sarewitz, D. (2013). Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change*, *4*(1), 39-60.