



## Analisis Klasifikasi Cuaca di Kota Batam Menggunakan Algoritma Metode Decision Tree C4.5

**Rizka Fadhillah Arsitania**

Manajemen Rekayasa, Institut Teknologi Batam, Indonesia

**Vidhyah Dhana Reswari**

Manajemen Rekayasa, Institut Teknologi Batam, Indonesia

**Harry Robertson Panggabean**

Manajemen Rekayasa, Institut Teknologi Batam, Indonesia

Alamat: Jl. Gajah Mada, Kompleks Vitka City, Tiban Ayu - Sekupang, Batam 29425,  
Kepulauan Riau - Indonesia

Korespondensi penulis: [2212020@student.iteba.ac.id](mailto:2212020@student.iteba.ac.id).

**Abstract.** An area of Indonesia known as the Riau Islands is made up of numerous islands. Among these is the city of Batam, which has year-round high temperatures and precipitation due to its tropical environment. A number of variables, including temperature, humidity, wind direction, and barometric pressure, have a significant impact on the weather in the Riau Islands. One element that affects it is the weather; variations in the weather can have a big effect on a lot of different industries, like tourism and transportation. The C4.5 decision tree is a tree structure where the nodes stand for decisions and the branches for the decisions' outcomes. The Decision Tree C4.5 algorithm's classification method is employed in this study. The rainfall dataset was acquired from Weather Forecast and split into 104 segments using a 7:3 ratio. 30% of the RapidMiner modeling process is testing, and 70% is training data. There are four types of weather: cloudy, sunny, drizzling, and rainy. We may observe that the qualities "sunny," "cloudy," "drizzly," and "rainy" are meteorological variables in the decision tree pattern.

**Keywords:** Weather, Classification, Decision Tree

**Abstrak.** Kepulauan Riau adalah wilayah Indonesia yang terdiri dari beberapa pulau. Kota Batam adalah salah satunya, dengan iklim tropis dan curah hujan tinggi sepanjang tahun. Beberapa faktor, antara lain suhu, kelembaban, arah angin, dan tekanan barometrik, mempengaruhi cuaca di Kepulauan Riau, dan perubahan cuaca dapat mempengaruhi banyak hal, seperti transportasi dan pariwisata. Metode klasifikasi algoritma Decision Tree C4.5 digunakan dalam metodologi penelitian ini. Pohon keputusan C4.5 adalah struktur pohon yang terdiri dari titik-titik yang mewakili keputusan dan cabang-cabang yang mewakili hasil dari keputusan tersebut. Data curah hujan dikumpulkan dari prakiraan Cuaca dan dibagi menjadi 104 kelompok dengan perbandingan 7:3. 70% data hujan dimodelkan oleh proses pemodelan RapidMiner hingga 30% dari tes atau pengujian. Parameter yang diambil pada Cuaca dikategorikan menjadi empat kategori: cerah, berawan, gerimis, dan hujan. Kita dapat melihat atribut "cerah", "berawan", "gerimis", dan "hujan" sebagai variabel utama cuaca dalam pola keputusan pohon.

**Kata kunci:** Cuaca, Klasifikasi, Decision Tree

### 1. LATAR BELAKANG

Kepulauan Riau merupakan gugusan wilayah Indonesia yang berbentuk provinsi yang terdiri dari rangkaian pulau untuk setiap kota/kabupatennya seperti Kota Batam, Pulau Bintan, Karimun, dan Pulau Lingga. Wilayah ini beriklim tropis, dengan cuaca relatif panas dan lembab sepanjang tahun. Cuaca di Kepulauan Riau sangat dipengaruhi oleh beberapa faktor seperti suhu, kelembaban, arah angin, dan tekanan udara.

Menurut artikel yang ditulis oleh (Kartasapoetra, 2010), cuaca didefinisikan sebagai keadaan atmosfer pada waktu tertentu yang terus berubah secara teratur atau dari waktu ke waktu. Cuaca merupakan salah satu faktor penting yang mempengaruhi banyak aspek kehidupan, seperti transportasi, industri, dan pariwisata. Di kota-kota besar seperti Batam, prediksi cuaca yang akurat sangat penting karena kota Batam sendiri ialah sebagai kota dengan aktivitas ekonomi yang tinggi dan merupakan salah satu pusat industri, seringkali menghadapi tantangan dalam hal perencanaan dan pengambilan keputusan yang dipengaruhi oleh kondisi cuaca.

Kota Batam yang memiliki iklim tropis menjadi salah satu pulau dengan perputaran ekonomi yang tinggi di Indonesia, namun iklim tropis ini menyebabkan curah hujan yang cukup tinggi sepanjang tahun. Cuaca dapat berdampak besar pada banyak industri, seperti transportasi, pariwisata, dan aktivitas sehari-hari masyarakat. Oleh karena itu, kemampuan untuk memprediksi kondisi cuaca secara akurat menjadi sangat penting.

Prediksi cuaca yang akurat membantu masyarakat dan pemerintah kota mengambil keputusan yang lebih baik ketika menghadapi berbagai kondisi cuaca. Cara efektif untuk melakukan prakiraan cuaca adalah dengan menggunakan algoritma klasifikasi. Di antara berbagai algoritma klasifikasi yang tersedia, penggunaan metode *Decision Tree* C4.5 merupakan algoritma yang paling populer dan efektif.

Dikutip dari Jurnal yang ditulis oleh (Prasetio, 2021), *Decision Tree* juga disebut struktur pohon keputusan, adalah struktur pohon di mana hasil pengujian diwakili oleh setiap cabang, kelas atau distribusi kelas diwakili oleh setiap simpul daun, dan atribut pengujian diwakili oleh setiap simpul internal. Pendekatan ini memanfaatkan struktur pohon, di mana keputusan diwakili oleh node dan cabang.

Diharapkan dengan adanya jurnal ini dapat dipahami dan memberi manfaat hasil analisis klasifikasi cuaca ini, Kota Batam dapat lebih mempersiapkan diri dalam menghadapi berbagai tantangan cuaca dan meminimalisir dampak negatif yang mungkin terjadi dan juga dapat menjadi dasar untuk mengembangkan sistem prediksi cuaca yang lebih canggih di masa depan.

## 2. KAJIAN TEORITIS

### A. Analisis Data

Seperti yang dinyatakan oleh Rihai (2016), Analisis data adalah proses mengumpulkan, mengatur, dan menginterpretasikan data. Terlepas dari kenyataan bahwa terminologi ini memiliki fungsi yang berbeda, semuanya memiliki arti yang berbeda. Analisis data terdiri dari beberapa teknologi dan pendekatan yang perlu digabungkan dengan cara-cara baru untuk mengungkap nilai tersembunyi dari kumpulan data yang sangat besar. Kumpulan data ini lebih besar, lebih rumit, dan berisi informasi unik yang tidak ditemukan dalam kumpulan data biasa.

### B. Data Mining

Menurut Sari dkk (2020), data mining bukanlah pengetahuan baru dalam mengolah data. Data mining menggabungkan banyak disiplin ilmu yang ada, yang menjadikannya sulit untuk dijelaskan dalam bentuk kata. Data mining bertujuan untuk mengubah cara tradisional kita melihat data sehingga kita dapat memprosesnya, menurut beberapa disiplin ilmu: 1. Jumlah data yang besar; 2. Dimensi data yang tinggi; dan 3. Data yang heterogen dan beragam. Mengubah data mentah menjadi pengetahuan atau kumpulan informasi yang dapat digunakan sebagai pendukung keputusan yang efektif adalah salah satu dari banyak manfaat yang ditawarkan oleh data mining.

### C. Klasifikasi

Menurut Bafadal (2009:51) klasifikasi merupakan tahap pemilihan serta pengelompokan buku-buku perpustakaan atau bahan Pustaka lainnya berdasarkan aturan tertentu seperti berdasarkan penerbit, nomor seri, dll, kemudian ditempatkan secara bersama-sama di tempat tertentu. Klasifikasi adalah proses pembagian benda atau konsep secara logika ke dalam kelas hirarki, subklas, dan sub kelas berdasarkan karakteristik yang membedakannya. Klasifikasi juga dapat didefinisikan sebagai proses pengaturan pengetahuan secara luas ke dalam kumpulan sistematis.

### D. *Confusion Matrix*

Pada saat mengukur kerja confusion matrix, terdapat empat istilah penggambaran proses klasifikasi. empat istilah tersebut ialah false positive (FP), True Positive (TP), true negative (TN), dan false negative (FN). Nilai TN adalah banyaknya data negatif yang dikenali dengan benar, dan nilai FP adalah data negatif yang dikenali sebagai data

positif. TP berarti data positif yang dikenali dengan benar. FN adalah kebalikan dari positif sebenarnya. Dengan kata lain, meskipun data H. positif, namun dianggap sebagai data negatif. Dengan menggunakan metode pohon keputusan dengan menggunakan atribut pendukung keterampilan berbicara, mendengarkan, membaca dan menulis, diharapkan dapat mengukur keterampilan taruna melalui matriks konfusi dan meningkatkan kemampuan bahasa Inggris maritimnya. Digunakan untuk menentukan presisi, presisi, dan recall untuk memprediksi keterampilan taruna di masa depan.

#### E. *Decision Tree*

Menurut Muzakki dan Wulandari (2016), pohon keputusan adalah kaidah pengelompokan dan prediksi yang terkenal dan ampuh. Teknik pohon keputusan (*Decision Tree*) menjadikan fakta berbentuk data sebagai pohon keputusan yang mewakili data aturan. Aturan mudah dipahami menggunakan bahasa alami. Algoritma C4.5 merupakan algoritma untuk membuat *Decision Tree*.

Teknik yang terkenal untuk prediksi dan pengelompokan adalah pohon keputusan. Teknik ini memfasilitasi eksplorasi data dan mengungkapkan korelasi tersembunyi antara variabel tujuan dan berbagai faktor input potensial.

Rumus pertama adalah:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S: Himpunan Kasus

A: Atribut

n: Jumlah Partisi Atribut A

|S<sub>i</sub>|: Jumlah kasus pada partisi ke-i

|S|: Jumlah Kasus dalam S

Dan rumus yang kedua adalah:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n p_i \log_2 p_i$$

Keterangan:

S: Himpunan kasus

n: Banyaknya partisi S

pi: probabilitas yang didapat dari kasus i dibagi total kasus.

#### F. *Software* Rapidminer

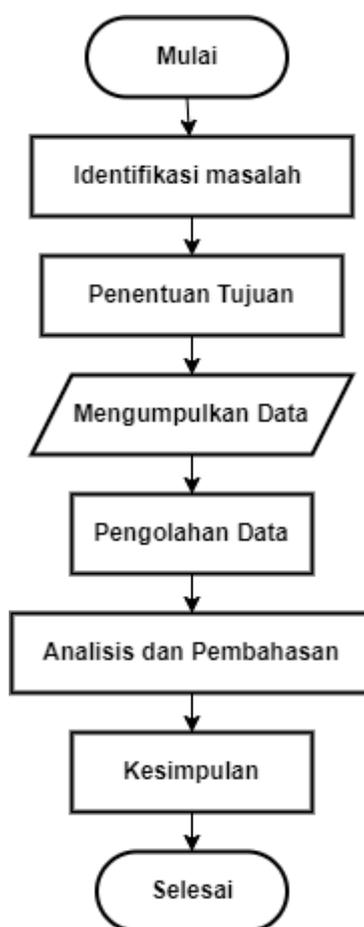
Menurut Sudarsono (2021), perangkat lunak Rapid Miner dikembangkan oleh Ralf Klinkenberg dari rapid-i.com dan Dr. Markus Hofmann dari *Institute of Technology Blanchardstown*. Tampilan GUI-nya membuatnya mudah digunakan. Lisensi publik GNU digunakan untuk membangun perangkat lunak sumber terbuka, yang dikembangkan dengan aplikasi Java. Rapid Miner merupakan perangkat lunak sumber bebas yang dapat digunakan pada sistem operasi apa pun.

#### G. Cuaca

Menurut Puspita dan Yulianti (2016), cuaca adalah keadaan udara pada saat tertentu dan di area tertentu dalam jangka waktu yang relatif singkat. Suhu udara, tekanan udara, angin, dan kelembaban udara merupakan komponen yang mempengaruhi cuaca dan iklim.

## METODE PENELITIAN

Pada metode penelitian ini, digunakan metode Klasifikasi Algoritma *Decision Tree* C4.5 dimana metode ini untuk mengelompokkan atribut atau kelas-kelas pada cuaca di Kota Batam dan membentuk suatu pohon keputusan. Selanjutnya dibawah ini merupakan *flowchart* atau alur dalam penelitian yang telah dilakukan, sebagai berikut.



**Gambar 1.** *Flowchart* Praktikum

## HASIL DAN PEMBAHASAN

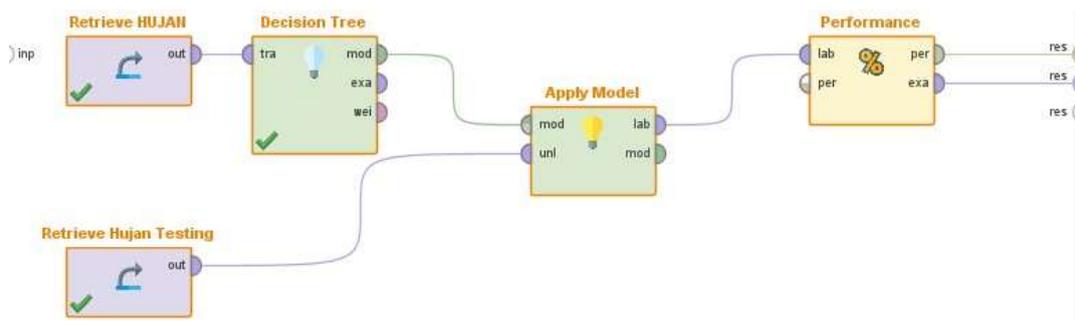
### A. Pengumpulan dan Pengolahan Data

Pada pengumpulan dataset hujan diperoleh dari *weather Forecast*, Tanggal, Waktu, Desember, Cuaca, Suhu (Celcius), Hujan (mm), dan Angin (mph) adalah atribut yang digunakan dalam dataset hujan yang dikumpulkan dari perkiraan Cuaca. Total data yang didapat pada dataset adalah 104 data yang kemudian dalam parameters data dibuat perbandingan sebesar 70:30 sehingga menghasilkan data 72:32 yang akan di prediksi.

**Tabel 1.1** Dataset Hujan

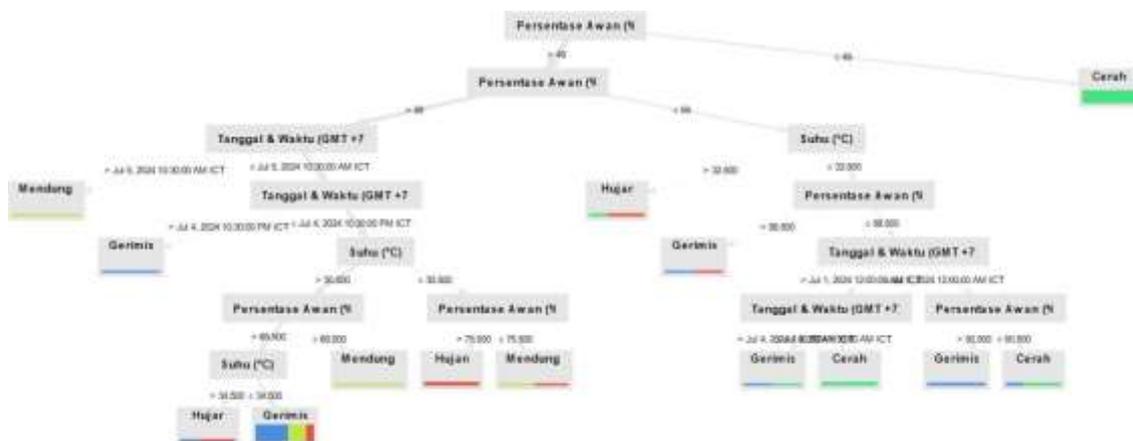
<b>No</b>	<b>Cuaca</b>	<b>Tanggal dan Waktu</b>	<b>Suhu</b>	<b>Tinggi Hujan</b>	<b>Presentase awan</b>
1	Gerimis	27/06/2024 00:00	32	0.6	79
2	Cerah	27/06/2024 03:00	32	0.0	50
3	Mendung	27/06/2024 06:00	32	0.0	78
4	Hujan	27/06/2024 09:00	33	2.2	78
5	Gerimis	27/06/2024 12:00	35	0.4	72
6	Mendung	27/06/2024 15:00	34	0.2	71
7	Cerah	27/06/2024 18:00	33	0.0	22
8	Cerah	27/06/2024 21:00	32	0.0	36
9	Mendung	28/06/2024 00:00	32	0.0	68
10	Mendung	28/06/2024 03:00	31	0.0	60
11	Gerimis	28/06/2024 06:00	31	0.1	69
....					
97	Cerah	07/07/2024 12:00	32	0.0	48
98	Gerimis	07/07/2024 15:00	33	0.2	61
99	Gerimis	07/07/2024 18:00	31	0.1	98
100	Cerah	07/07/2024 21:00	30	0.0	45
101	Cerah	08/07/2024 00:00	30	0.0	44
102	Gerimis	08/07/2024 03:00	30	0.3	73
103	Gerimis	08/07/2024 06:00	29	0.2	72
104	Hujan	08/07/2024 09:00	32	0.8	52

Pada pengolahan data, proses pemodelan menggunakan *software* Rapidminer untuk mengklasifikasikan cuaca yang ada di Kota Batam. Dari proses pemodelan oleh RapidMiner data hujan sejumlah 70% data dibandingkan dengan 30% data testing, klasifikasi cuaca dibagi menjadi 4 yaitu, cerah, mendung, gerimis dan hujan.



Gambar 1.1 Proses Pemodelan Rapidminer

Pada atribut utama yang dapat mempengaruhi proses pengklasifikasian kumpulan data dapat diidentifikasi pada pola pohon keputusan, juga dikenal sebagai decision tree. Sehingga, terbentuklah pola pengetahuan yang dapat digunakan sebagai dasar untuk memprediksi cuaca yang akan datang. Cerah, Mendung, Gerimis, dan Hujan yang menjadi atribut utamanya.



Gambar 1.2 Pohon Keputusan (*Decision Tree*)

B. Analisis

Berikut adalah hasil analisis dari penggunaan *software* rapidminer pada klasifikasi dataset Hujan.

Row No.	Cuaca	prediction(C...	confidence(...	confidence(...	confidence(...	confidence(...	Tanggal & ...	Suhu (°C)	Tinggi Hujan...	Persentase ...
1	Gerimis	Mendung	0	0	1	0	Jul 6, 2024 1...	31	0.3	70
2	Gerimis	Mendung	0	0	1	0	Jul 6, 2024 3...	30	0.4	71
3	Gerimis	Mendung	0	0	1	0	Jul 6, 2024 6...	29	0.4	75
4	Mendung	Mendung	0	0	1	0	Jul 6, 2024 9...	32	0.0	61
5	Gerimis	Mendung	0	0	1	0	Jul 8, 2024 1...	33	0.4	79
6	Gerimis	Mendung	0	0	1	0	Jul 6, 2024 3...	31	0.5	62
7	Cerah	Cerah	0	1	0	0	Jul 6, 2024 6...	31	0.0	46
8	Cerah	Cerah	0	1	0	0	Jul 6, 2024 9...	30	0.0	47
9	Cerah	Cerah	0	1	0	0	Jul 7, 2024 1...	29	0.0	45
10	Cerah	Cerah	0	1	0	0	Jul 7, 2024 3...	29	0.0	41
11	Cerah	Cerah	0	1	0	0	Jul 7, 2024 6...	29	0.0	41
12	Cerah	Cerah	0	1	0	0	Jul 7, 2024 9...	31	0.0	42
13	Cerah	Cerah	0	1	0	0	Jul 7, 2024 1...	32	0.0	48
14	Gerimis	Mendung	0	0	1	0	Jul 7, 2024 3...	33	0.2	61
15	Gerimis	Mendung	0	0	1	0	Jul 7, 2024 6...	31	0.1	68

Gambar 1.3 Hasil Klasifikasi menggunakan *Rapidminer*

Row No.	Cuaca	prediction(C...	confidence(...	confidence(...	confidence(...	confidence(...	Tanggal & ...	Suhu (°C)	Tinggi Hujan...	Persentase ...
16	Cerah	Cerah	0	1	0	0	Jul 7, 2024 9...	30	0.0	45
17	Cerah	Cerah	0	1	0	0	Jul 8, 2024 1...	30	0.0	44
18	Gerimis	Mendung	0	0	1	0	Jul 8, 2024 3...	30	0.3	73
19	Gerimis	Mendung	0	0	1	0	Jul 8, 2024 6...	28	0.2	72
20	Hujan	Gerimis	0.500	0.500	0	0	Jul 8, 2024 9...	32	0.8	52
21	Gerimis	Gerimis	0.500	0.500	0	0	Jul 8, 2024 1...	30	0.1	52
22	Cerah	Cerah	0	1	0	0	Jul 8, 2024 3...	31	0.0	41
23	Mendung	Mendung	0	0	1	0	Jul 8, 2024 6...	30	0.0	63
24	Cerah	Cerah	0	1	0	0	Jul 8, 2024 9...	30	0.0	49
25	Cerah	Cerah	0	1	0	0	Jul 9, 2024 1...	31	0.0	45
26	Hujan	Mendung	0	0	1	0	Jul 9, 2024 3...	31	1.4	83
27	Cerah	Cerah	0	1	0	0	Jul 9, 2024 6...	31	0.0	46
28	Hujan	Mendung	0	0	1	0	Jul 9, 2024 9...	31	1.8	79
29	Mendung	Mendung	0	0	1	0	Jul 9, 2024 1...	33	0.0	67
30	Mendung	Mendung	0	0	1	0	Jul 9, 2024 3...	34	0.0	62

Gambar 1.3 Hasil Klasifikasi menggunakan *Rapidminer* (Lanjutan).

Dari hasil tabel diatas didapati bahwa presentase awan cerah sebesar  $\leq 49\%$ . Pada presentase awan  $\geq 49\%$  terdapat 2 cabang pohon Keputusan dengan pengaruh suhu dan tanggal: Presentase awan  $\geq 49\%$  dengan suhu  $>32^\circ$  adalah hujan, presentase awan  $\geq 56\%$  adalah grimis, kemudian presentase awan  $\leq 56\%$  bisa grimis ataupun cerah. Presentase awan  $> 50\%$ , pada tanggal 5 Juli cuaca mendung, pada 4 Juli grimis, lalu dipengaruhi oleh suhu dengan presentase awan  $>32\%$  cuaca Sebagian besar mendung dan hujan terkadang juga grimis.

## KESIMPULAN DAN SARAN

### A. Kesimpulan

Cara memprediksi cuaca di Kota Batam menggunakan *Decision Tree* C4.5 dengan berawal mengumpulkan data terlebih dahulu, kemudian *cleaning* datanya, pembagian variabel cuaca seperti cerah, mendung, gerimis, dan hujan. Lakukan pengujian menggunakan *software* Rapidminer serta setelahnya di evaluasi dan di implementasikan.

Atribut yang digunakan pada klasifikasi prediksi cuaca di Kota Batam adalah cerah, mendung, gerimis, dan hujan.

### B. Saran

Penelitian ini berupa analisis cuaca menggunakan Algoritma *Decision Tree* C4.5. Oleh karenanya, penelitian serupa dapat diterapkan dengan data di wilayah lain. Kendala berupa data dapat diperbaiki oleh peneliti selanjutnya.

## DAFTAR REFERENSI

- As'ad, B. (2016). *PREDIKSI KEPUTUSAN MENGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES, ONE-R, DAN DECISION TREE*, 10.
- E. T. L, K. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Kaonang, F. J., Rotikan, R., & Tulung, G. S. (2018). *Pemodelan Sistem Prediksi Tanaman Pangan Menggunakan Algoritma Decision Tree*.
- Kartasapoetra, A.G, dan M.M Sutedjo. 2010. *Teknologi Konservasi Tanah dan Air*. Jakarta: Rineka Cipta.
- Kristopher, H., Herwindiati, D. E., & Sutrisno, T. (2023). *PENERAPAN METODE DECISION TREE UNTUK PERKIRAAN CUACA KOTA BEKASI*, 11.
- Manshur, A. (2021). *Satu Data, Big Data dan Analatika Data: Urgensi Pelembagaan, Pembiasaan dan Pembudayaan*, 17.
- Mardi, Y. (n.d.). *Data Mining: Klasifikasi Menggunakan Algoritma C4.5*, 7. Mujiasih, S. (2011). *PEMANFATAN DATA MINING UNTUK PRAKIRAAN CUACA*.
- Muslim, M. A., Prasetyo, B., Harum Marwani, E. L., Herowati, A. J., Mirqotussa'adah, Nurzahputra, A., & Hardiyanti, S. R. (2019). *Data Mining Algoritma C4.5*.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19-26.

Novandya, A., & Oktria, I. (2017). *Penerapan Algoritma Klasifikasi Data Mmining C4.5 Pada Dataset Cuaca Wilayah Bekasi*.

Novitri, R., & Irawati, N. (2019). *ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINE*, 6.

Pascalina, D., Widhiastono, R., & Juliane, C. (2023). *Technomedia Journal (TMJ)*.

*Pengukuran Kesiapan Transformasi Digital Smart City Menggunakan Aplikasi Rapid Miner*, 10.

Prasetio, A., Hasibuan, M. H., & Sitompul, P. (2021). *Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan status Gizi Balita*, 4.

Prasetyo, V. R., Lazuardi, H., Mulyono, A. A., & Lauw, C. (2021). *Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier*, 16.

Rofani, R., Oktavina, L., & Vernanda, D. (2023). *Penerapan Metode Klasifikasi Decision Tree dalam Prediksi kanker Paru-paru menggunakan Algoritma C4.5. Jurnal Tekno Kompak*, 126-139.

Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). *Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. CESS (Journal Comput. Eng. Syst. Sci., vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess. v5i2. 18519*.

Siregar, A. M., Faisal, S., Cahyana, Y., & Priyatna, B. (2020). *PERBANDINGAN ALGORITME KLASIFIKASI UNTUK PREDIKSI CUACA*, 3.

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). *Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. JBASE-Journal of Business and Audit Information Systems*, 4

Wibawa, A. P., Aji Purnama, M. G., Akbar, M. F., & Dwiyanto, F. A. (2018). *Metode-metode Klasifikasi*, 3, 134.