



Analisis Ramalan Cuaca di Sekupang, Kota Batam Menggunakan Algoritma Decision Tree dan Confusion Matrix

Fitrianingsih

Institut Teknologi Batam, Indoensia

Bella Zuraeni

Institut Teknologi Batam, Indoensia

Alamat: Jl Gajah Mada, Kompleks Vitka City, Tiban Ayu – Sekupang, Batam 29425,

Kepulauan Riau – Indonesia

penulis: 2212025@student.iteba.ac.id

Abstract. Weather is an atmospheric condition consisting of variables such as temperature, humidity, precipitation, wind, and air pressure and describes the state of the environment at a particular time and place. This research discusses the use of the Decision Tree algorithm in predicting conditions. Decision trees can model in the form of a tree structure, where each internal node represents a trait or feature, each branch represents a decision rule, and each leaf or leaflet represents a category or outcome. This research aims to predict the weather by applying data classification methods using the decision tree algorithm. The research method involves classifying weather data based on group attributes such as wind speed, temperature, rainfall, and weather. Decision tree is a model that maps decisions in the form of a tree. This research uses 104 historical weather data obtained from weather online, the dataset used includes time, temperature, cloud percent, rainfall, and weather. Test results with 70% of the dataset are useful for testing the model, and 30% are used to train model performance. Model evaluation is done using accuracy, precision, recall. The results show a model accuracy of 96.77%, indicating good model performance in predicting weather conditions.

Keywords: Weather, Decision Tree, Prediction.

Abstrak. Cuaca adalah kondisi atmosfer yang terdiri dari variabel seperti suhu, kelembaban, curah hujan, angin, dan tekanan udara dan menggambarkan keadaan lingkungan pada waktu dan tempat tertentu.. Penelitian ini membahas penggunaan algoritma Decision Tree dalam memprediksi kondisi. Decision tree dapat membuat model dalam bentuk struktur pohon, di mana setiap simpul internal mewakili suatu sifat atau ciri, setiap cabang mewakili aturan keputusan, dan setiap daun atau daun mewakili kategori atau hasil. Penelitian ini bertujuan untuk memprediksi cuaca dengan menerapkan metode klasifikasi data menggunakan algoritma decision tree. Metode penelitian melibatkan pengklasifikasian data cuaca berdasarkan atribut kelompok seperti kecepatan angin, suhu, curah hujan, dan cuaca. Decision tree yaitu model yang memetakan keputusan dalam bentuk pohon. Penelitian ini menggunakan 104 data histori cuaca yang di peroleh dari weather online, dataset yang digunakan meliputi waktu, suhu, persen awan, curah hujan, dan cuaca. Hasil pengujian dengan 70% dari dataset berguna untuk menguji model, dan 30% digunakan untuk melatih performa model. Evaluasi model dilakukan menggunakan akurasi, precision, recall. Hasil menunjukkan akurasi model sebesar 96,77%, menunjukkan kinerja model yang baik dalam memprediksi kondisi cuaca.

Kata kunci: Cuaca, Pohon Keputusan, Prediksi

1. LATAR BELAKANG

Indonesia terletak di wilayah beriklim tropis yang 81% airnya hangat, sehingga suhu di wilayah pesisir stabil pada angka 28°C, sedangkan di daerah pedalaman dan pegunungan suhunya 26°C, dan di pegunungan tinggi biasanya 23°C. Suhu bervariasi dari musim ke musim, namun perubahannya relatif kecil. Karena Indonesia adalah negara tropis, Oleh karena itu, tidak ada musim dingin, musim gugur, musim panas, atau musim semi; satu-satunya yang ada adalah musim hujan dan musim panas. Meskipun berbeda di beberapa daerah, sebagian besar Indonesia (termasuk Jawa dan Bali) mengalami musim kemarau bulan April hingga Oktober. Saat ini, musim hujan terjadi dari bulan November - Maret. Namun dampak pemanasan global membuat prediksi perubahan musim di Indonesia agak sulit.

Menurut Undang-Undang Nomor 25 Tahun 2002, Kepulauan Riau adalah provinsi ke-32 di Indonesia. Terdiri dari Kota Tanjung Pinang, Kota Batam , Kabupaten Bintan, Kabupaten Karimun, Kabupaten Natuna, dan Kabupaten Lingga. Ada 4 provinsi, 2 kota , 42 kelurahan, dan 256 kelurahan dan desa. Salah satu contohnya adalah kecamatan Sekupang di Batam, yang memiliki iklim tropis. Musim hujan terjadi dari November hingga April, sering kali disertai dengan curah hujan tinggi dan kelembapan yang tinggi. Musim kemarau berlangsung dari Mei hingga Oktober, suhu yang cenderung stabil dan curah hujan rendah.

Untuk meningkatkan prediksi cuaca di Sekupang, serta memahami pola cuaca, Decision Tree dapat memberikan solusi yang efektif. Decision Tree mampu mengelola data cuaca yang kompleks dan memberikan interpretabilitas yang tinggi dalam prediksi. Dengan menganalisis data cuaca historis dan variabel lingkungan, Decision Tree dapat membantu dalam memprediksi kondisi cuaca secara lebih akurat.

2. KAJIAN TEORITIS

A. Cuaca

Istilah ketentuan "cuaca " lebih mengacu pada keadaan atmosfer , di mana lapisan udara kemudian memperluas planet Bumi . Oleh karena itu , "cuaca" cuaca " biasanya mengacu pada keadaan atmosfer di dunia kita sendiri .lebih mengacu pada keadaan atmosfer , dimana lapisan udara kemudian membesarplanet Bumi. Oleh karena itu, istilah " cuaca " biasanya merujuk pada keadaan atmosfer di dunia kita sendiri Menurut Kamus Besar Bahasa Indonesia (KBBI), keadaan, udara (tentang suhu, cahaya matahari, kelembaban, kecepatan angin, dan sebagainya) pada tempat tertentu dalam jangka waktu tertentu. Cuaca adalah keadaan atmosfer pada suatu waktu tertentu yang ciri- cirinya dapat berubah seiring berjalannya waktu

atau dari waktu ke waktu. Faktor ini sangat mempengaruhi banyak aspek kehidupan, termasuk transportasi, industri, dan pariwisata.

B. Confusion Matrix

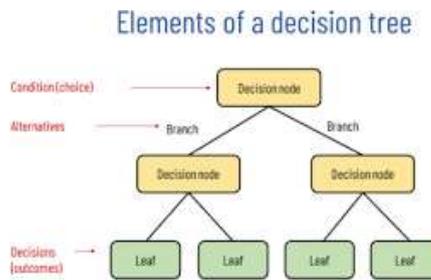
Confusion matrix bertujuan untuk mengevaluasi model dan ditampilkan secara visual dalam format tabel. Hal ini memberikan praktisi data wawasan yang lebih mendalam tentang kinerja model, kesalahan, dan kelemahan. Hal ini memungkinkan praktisi data untuk menganalisis model lebih lanjut melalui penyesuaian.

C. Rapid Miner

Sudarsono (2021) menyatakan bahwa Dr. Markus Hofmann dari Blanchardstown Institute of Technology dan Ralf Klinkenberg dari rapid-i.com mengembangkan perangkat lunak yang disebut RapidMiner. Perangkat lunak ini memiliki antarmuka pengguna grafis, atau GUI, yang membuatnya lebih mudah digunakan. RapidMiner adalah perangkat lunak analisis data independen yang membantu Anda belajar ilmu penambangan data. Platform ini dibuat oleh perusahaan yang mengkhususkan diri dalam data besar dan membantu berbagai proses yang berkaitan dengan bisnis, penelitian, pendidikan, pelatihan, dan pembelajaran. Sekitar seratus solusi pembelajaran untuk pengelompokan, klasifikasi, dan analisis regresi disediakan oleh RapidMiner, yang juga mendukung dua puluh dua format file, termasuk format seperti.xls dan.csv.

D. Decision Tree

Decision tree adalah algoritma diagram yang menggunakan struktur probabilitas untuk membantu merumuskan kesimpulan dan memprioritaskan tugas terkait dengan regresi dan klasifikasi. Struktur pohon keputusan berasal dari satu simpul akar, dengan cabang-cabang, simpul daun (juga disebut simpul terminal), dan simpul internal (atau simpul keputusan). Simpul akar tunggal merespons pertanyaan atau masalah yang perlu diselesaikan. Cabang adalah jalur keputusan yang pada akhirnya terhubung ke beberapa simpul internal. Setiap pohon keputusan mungkin memiliki beberapa simpul internal sebagai rencana cadangan dalam situasi darurat atau tak terduga. Node internal juga dapat memiliki simpul daun, yang menunjukkan akhir dari proses tersebut.[2]



Gambar 1. Element of Decision Tree

Menurut Rofani jika kita ingin menghitung *entropi* suatu sistem maka kita gunakan rumus seperti berikut,

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

Entropy (si)

Keterangan:

- S : Himpunan Kasus
- A : Atribut
- N : JumlahPartisiAtributA
- |S_i| : Jumlahkasus pada partisike-i
- |S| : Jumlah Kasus dalam S

E. Klasifikasi

Menurut Bafadal klasifikasi bertujuan pemilihan dan pengelompokan buku perpustakaan serta bahan perpustakaan lainnya, berdasarkan kriteria tertentu, dengan tujuan mengumpulkannya dalam satu tempat. Klasifikasi adalah proses memprediksi kelas atau kategori suatu entitas berdasarkan karakteristik atau atributnya. Ini merupakan jenis pembelajaran yang diawasi di mana model pembelajaran mesin dapat dilatih menggunakan

kumpulan data berlabel. Dalam pelatihan, dataset diberi label kelas untuk membantu model mengidentifikasi kelas-kelas dari data baru yang belum pernah dilihat sebelumnya.

F. Konsep Klasifikasi

Menurut Yael Keshet dalam makalahnya yang berjudul "*Classification Systems in the Sociology of Knowledge*," klasifikasi adalah tindakan alami manusia yang membuat dunia di sekitarnya lebih mudah dipahami dan masuk akal. Pendapat ini diterima dengan baik ketika pakaian dan perlengkapannya disusun ke dalam beberapa kelompok berdasarkan bentuk, ukuran, atau fungsinya, untuk mempermudah penyimpanan. memungkinkan aplikasi untuk menemukan sesuatu yang berguna saat mencari hal lain (deteksi acak).

G. Data Mining

Data mining bertujuan untuk mengumpulkan data dalam jumlah besar dan mengekstraksi informasi yang digunakan. Tujuannya adalah mencari tren yang berguna dalam sebuah database untuk mendukung pengambilan keputusan di masa yang akan datang. (F. Zaki, 2014). Metode yang digunakan meliputi teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (Dwi & Slamet, 2012).

3. METODE PENELITIAN

Data yang digunakan 104 data sekunder dari histori world weather online pada wilayah Sekupang, kota Batam Kepulauan Riau. Data mencakup periode dari 26 Juni – 8 Juli 2024 dengan menggunakan atribut waktu, suhu, persen awan, curah hujan dan data yang sudah di dapatkan di olah menggunakan software rapidminer. Pada penelitian ini, flowchart visual yang menunjukkan proses atau alur kerja sangat penting. Untuk merencanakan, menganalisis, dan memahami langkah-langkah yang diperlukan untuk menyelesaikan suatu tugas atau masalah, diagram alur digunakan. Untuk membantu memahami alur kerja, simbol dalam flowchart memiliki arti khusus.



Gambar 2. Flowchart

Langkah pertama dalam proses penelitian adalah menentukan topik penulisan yang spesifik dan jelas. Setelah topik ditentukan, langkah berikutnya adalah mengidentifikasi masalah yang akan diteliti dengan mendetail dan membuat penjelasan yang dapat diukur. Setelah masalah diidentifikasi, peneliti menetapkan tujuan penelitian yang jelas untuk memberikan panduan sepanjang proses penelitian. Data diperoleh dari World Weather Online yang berisi 104 data cuaca di Sekupang, Kota Batam, Kepulauan Riau dengan atribut seperti tanggal, waktu, suhu, persen awan, dan curah hujan. Data diolah menggunakan software RapidMiner untuk analisis lebih lanjut. Setelah data terkumpul, dilakukan analisis untuk memproses dan menginterpretasi informasi sesuai dengan landasan teori yang telah ditetapkan. Hasil penelitian disimpulkan dalam kesimpulan yang merangkum temuan utama untuk pencapaian tujuan penelitian. Proses penelitian dianggap selesai setelah laporan akhir disusun dan hasilnya dijelaskan dengan jelas, memastikan semua langkah penelitian terpenuhi dan tujuan penelitian tercapai.

4. HASIL DAN PEMBAHASAN

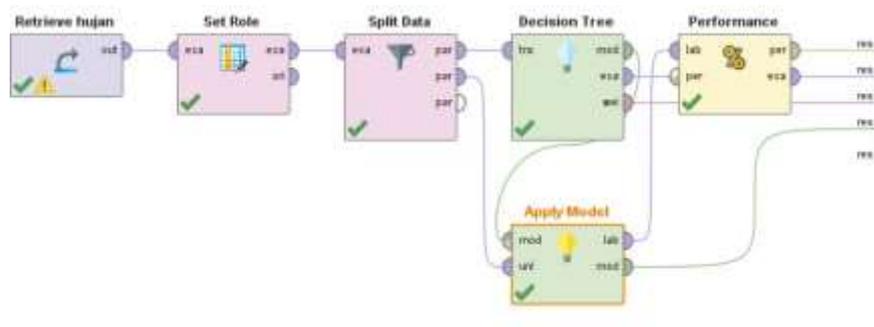
Pengumpulan Data

Pada penelitian ini menggunakan 104 data sekunder dari histori world weather online pada wilayah Sekupang, kota Batam Kepulauan Riau. Data mencakup periode dari 26 Juni – 8 Juli 2024 dengan atribut yang ada dalam data tersebut antara lain seperti tanggal, waktu, suhu, persen awan, curah hujan dan data yang sudah di dapatkan di olah menggunakan software rapidminer.

Tabel 1. Dataset Cuaca di Sekupang

No	Tanggal	Waktu	Suhu	Persen Awan	Curah Hujan	Cuaca
1	Wednesday, June 26, 2024	0:00	28	37.00%	0.6	Gerimis
2	Wednesday, June 26, 2024	3:00	28	84.00%	0	Cerah
3	Wednesday, June 26, 2024	6:00	27	71.00%	0	Cerah
4	Wednesday, June 26, 2024	9:00	29	76.00%	2.2	Hujan
5	Wednesday, June 26, 2024	12:00	29	70.00%	0.4	Mendung
6	Wednesday, June 26, 2024	15:00	30	83.00%	0.2	Mendung
7	Wednesday, June 26, 2024	18:00	29	24.00%	0	Cerah
8	Wednesday, June 26, 2024	21:00	28	100.00%	0	Cerah
9	Thursday, June 27, 2024	0:00	27	79.00%	0	Cerah
...						
96	Sunday, July 07, 2024	21:00	28	45.00%	0	Cerah
97	Monday, July 08, 2024	0:00	27	44.00%	0	Cerah
98	Monday, July 08, 2024	3:00	27	73.00%	0.3	Mendung
99	Monday, July 08, 2024	6:00	26	72.00%	0.2	Mendung
100	Monday, July 08, 2024	9:00	29	52.00%	0.8	Gerimis
101	Monday, July 08, 2024	12:00	27	52.00%	0.1	Mendung
102	Monday, July 08, 2024	15:00	28	41.00%	0	Cerah
103	Monday, July 08, 2024	18:00	27	93.00%	0	Cerah
104	Monday, July 08, 2024	21:00	27	49.00%	0	Cerah

Pengolahan Data



Gambar 3. Operator pada RapidMiner

Pada RapidMiner, operator adalah komponen dasar yang digunakan untuk melakukan berbagai tugas dalam proses analisis data. Berikut adalah tujuan dan penggunaan dalam analisis data:

Read Data: Mengimpor dataset ke dalam RapidMiner.

Set Role: Menandai atribut tertentu sebagai label target (variabel dependen). Pada parameter attribute name menggunakan data cuaca dan pada target role menggunakan label.

Split Data: Membagi dataset menjadi subset untuk testing dan latihan. Pada praktikum ini data dibagi menjadi 30% data latihan dan 70% data pengujian.

Apply Model: Menerapkan model yang telah dilatih ke data baru (biasanya data pengujian).

Decision Tree: Bertujuan untuk memvisualisasikan model pohon keputusan.

Confusion Matrix: Menyediakan tabel yang menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas.

Decision Tree



Gambar 4. Decision Tree

Menurut Muzakki dan Wulandari (2016), Decision tree adalah teknik klasifikasi dan prediksi yang terkenal dan sangat efektif. Teknik pohon keputusan mengorganisir kumpulan data yang sangat besar menjadi struktur pohon keputusan yang berisi aturan. Hal ini memungkinkan pengguna memahami aturan menggunakan bahasa alami. Alat yang dapat digunakan untuk mengklasifikasikan kondisi cuaca berdasarkan nilai curah hujan adalah pilihan pohon. Berikut adalah analisis dari keputusan pohon:

Struktur Decision Tree

1. Node Root (Akar)

Curah Hujan > 1.85 : Kondisi cuaca diprediksi " Hujan" jika curah hujan lebih besar dari 1.85

Curah Hujan ≤ 1.85 : Jika curah hujan kurang dari 1.85, lanjut ke node berikutnya.

2. Node 2

Curah Hujan > 0.05 : Jika curah hujan lebih besar dari 0.05, lanjut ke node berikutnya.

Curah Hujan ≤ 0.05 : Jika curah hujan kurang dari 0.05, kondisi cuaca diprediksi sebagai "Cerah".

3. Node 3

Curah Hujan > 0.45 : Jika curah hujan lebih besar dari 0.45, kondisi cuaca diprediksi sebagai "Gerimis".

Curah Hujan ≤ 0.45 : Jika curah hujan kurang dari 0.45, kondisi cuaca diprediksi sebagai "Mendung".

Confusion Matrix

accuracy: 96.77%

	true Gerimis	true Cerah	true Hujan	true Mendung	class precision
pred Gerimis	4	0	0	1	80.00%
pred Cerah	0	17	0	0	100.00%
pred Hujan	0	0	1	0	100.00%
pred Mendung	0	0	0	8	100.00%
class recall	100.00%	100.00%	100.00%	88.89%	

Gambar 5. Confusion Matriks

Dengan analisis ini, kita dapat melihat bagaimana decision tree memproses nilai curah hujan untuk memberikan prediksi kondisi cuaca. Model ini sangat berguna untuk situasi di mana keputusan harus dibuat berdasarkan beberapa aturan sederhana dan berjenjang.

Precision

Gerimis: 80% (dari 5 prediksi Gerimis, 4 benar) Cerah: 100% (dari 17 prediksi Cerah, semua benar) Hujan: 100% (dari 1 prediksi Hujan, semua benar)

Mendung: 100% (dari 9 prediksi Mendung, semua benar) Recall

Gerimis: 80% (dari 5 instance Gerimis, 4 diprediksi benar) Cerah: 100% (dari 17 instance Cerah, semua diprediksi benar) Hujan: 100.00% (dari 1 instance Hujan, semua diprediksi benar) Mendung: 88.89% (dari 9 instance Mendung, 8 diprediksi benar) Akurasi: 96.77% (total prediksi benar dari keseluruhan instance)

Secara keseluruhan, model menunjukkan akurasi yang sangat baik dan performa yang baik dalam klasifikasi sebagian besar kelas.

Tree

```
Curah Hujan > 1.850: Hujan {Gerimis=0, Cerah=0, Hujan=3, Mendung=0}
Curah Hujan ≤ 1.850
|   Curah Hujan > 0.050
|   |   Curah Hujan > 0.450: Gerimis {Gerimis=10, Cerah=0, Hujan=0, Mendung=1}
|   |   Curah Hujan ≤ 0.450: Mendung {Gerimis=0, Cerah=0, Hujan=0, Mendung=19}
|   Curah Hujan ≤ 0.050: Cerah {Gerimis=0, Cerah=40, Hujan=0, Mendung=0}
```

Gambar 4. Visualisasi Decision Tree

Jika total curah hujan lebih besar dari 1,85, diperkirakan akan turun hujan. Jika curah hujan kurang dari 1,85, pindah ke node berikutnya.

Jika curah hujan kurang dari 0,05, diperkirakan langit akan cerah. Jika curah hujan lebih besar dari 0,05, lanjutkan ke node berikutnya. Jika curah hujan di atas 0,45, akan terjadi hujan ringan.

Jika curah hujan di bawah 0,45, akan berawan.

5. KESIMPULAN DAN SARAN

Kesimpulan

Model decision tree yang dianalisis menggunakan RapidMiner menunjukkan kinerja yang sangat baik dalam memprediksi kondisi cuaca berdasarkan curah hujan. Dengan membagi 104 data menjadi 30% untuk pelatihan dan 70% untuk pengujian, model menghasilkan aturan prediksi yang jelas.

Analisis confusion matrix menunjukkan bahwa prediksi kondisi cuaca di Sekupang kota Batam menghasilkan akurasi lebih baik yaitu sebesar 96,77% menunjukkan efektivitasnya dalam klasifikasi cuaca berdasarkan curah hujan

Saran

Penelitian selanjutnya sebaiknya dikembangkan dengan menggunakan bahan dan topik lain. Hasil penelitian ini dapat dijadikan acuan untuk mengembangkan modul untuk prakiraan cuaca berdasarkan parameter curah hujan.

DAFTAR REFERENSI

- Amril Mutoi Siregar, T. S. (2020). Klasifikasi untuk prediksi cuaca menggunakan ensemble learning. *Jurnal Pengkajian dan Penerapan Teknik Informatika*, 13, 138-147.
- As'ad, B. (2016). Prediksi keputusan menggunakan metode klasifikasi Naïve Bayes, One-R, dan Decision Tree. 10.
- DataCamp. (2024, June 5). What is a confusion matrix in machine learning. Retrieved July 15, 2024, from <https://www.datacamp.com/tutorial/what-is-a-confusion-matrix-in-machine-learning>
- Dwi, M. B. (2012). Klasifikasi data karyawan untuk menentukan jadwal kerja menggunakan metode decision tree. *Jurnal IPTEK*, 17-23.
- Kartasapoetra, A. d. (2010). *Teknologi konservasi tanah dan air*. Jakarta: Rineka Cipta.
- Keshet, Y. (2011). Classification system in the light of sociology of knowledge. *Journal of Documentation*, 144-158.
- Kristopher, H., Herwindiati, D. E., & Sutrisno, T. (2023). Penerapan metode decision tree untuk perkiraan cuaca Kota Bekasi. 11.
- Lorio. (2022, December 8). RapidMiner. Retrieved July 15, 2024, from <https://binus.ac.id/entrepreneur/2022/12/08/rapidminer/>
- Muzakir, A., & Wulandari, R. A. (2016). Jurusan Informatika. Model data mining sebagai prediksi penyakit hipertensi kehamilan dengan teknik decision tree. 6.

- RevoU. (2024). Decision tree. Retrieved July 15, 2024, from <https://revou.co/kosakata/decision-tree>
- Riau, B. P. (2022, August 17). Provinsi Kepulauan Riau. Retrieved July 24, 2024, from <https://kepri.bpk.go.id/provinsi-kepulauan-riau/>
- Rofani, R. O. (2023). Penerapan metode klasifikasi decision tree dalam prediksi kanker paru-paru menggunakan algoritma. *Jurnal Tekno Kompak*, 126-136.
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis data mining data Netflix menggunakan aplikasi Rapid Miner. *Journal of Business and Audit Information Systems*, 9.
- Syariffudin, F. (2022, July 21). Klasifikasi artikel-artikel jurnal Pustakaloka berdasarkan skema JITA. *Jurnal Ilmu Perpustakaan dan Informasi*, 20-37.
- Telkom University. (2024, April 18). Flowchart pengertian: Fungsi, jenis, simbol, dan contohnya. Retrieved July 22, 2024, from <https://bee.telkomuniversity.ac.id/pengertian-flowchart-fungsi-jenis-simbol-dan-contohnya/>
- Wibawa, A. P., Aji Purnama, M. G., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode klasifikasi. 3, 134.
- Wonderful Indonesia. (2023, July 20). Musim, iklim dan cuaca. Retrieved July 23, 2024, from <https://www.indonesia.travel/id/id/informasi-umum/musim-iklim-dan-cuaca.html>
- World Weather Online. (2024, June 26). Sekupang weather. Retrieved July 8, 2024, from <https://www.worldweatheronline.com/sekupang-weather/riau-island/id.aspx>
- Zaki, F. (2014). Konsep data mining algoritma.