

Penerapan Data Mining dengan Metode Decision Tree untuk Prediksi Cuaca di Kota Seattle menggunakan Aplikasi Weka

Syarla Shalsabilla Putri Rachmawati¹, Kharisma Vidya Prakusa², Siti Rihastuti³

^{1,2}Program Studi Informatika, ³Program Studi Manajemen Informatika
STMIK Amikom Surakarta, Sukoharjo

Email : ¹syarla.10311@mhs.amikomsolo.ac.id,

²kharisma.10312@mhs.amikomsolo.ac.id, ³siti@dosen.amikomsolo.ac.id

Abstract

The application of the Decision Tree (J48) method through the WEKA application succeeded in producing a weather prediction model in the city of Seattle with an accuracy of around 86,016%. This model has a simple structure and is easy to understand, but there are limitations in classifying some weather categories. The use of ensemble methods or a combination of several models can increase the accuracy of weather predictions. This research contributes to the application of Decision Trees for weather prediction in the city of Seattle using WEKA, and realizes the potential for further development to improve the performance and accuracy of weather predictions. For further research, it is recommended to examine relevant weather attributes, use weather data from different sources, test models with diverse data, and consider ensemble methods. Model development for long-term weather prediction with historical data also needs to be considered. With these steps, it is hoped that we can improve the accuracy and performance of weather prediction models in the city of X using the Decision Tree method and the WEKA application.

Keywords: Data Mining, Classification, Decision Tree, Weather, J48 Algorithm, Weka

Abstraksi

Penerapan metode Decision Tree (J48) melalui aplikasi WEKA berhasil menghasilkan model prediksi cuaca di kota Seattle dengan akurasi sekitar 86.016%. Model ini memiliki struktur sederhana dan mudah dipahami, tetapi ada keterbatasan dalam mengklasifikasikan beberapa kategori cuaca. Penggunaan metode ensemble atau kombinasi beberapa model dapat meningkatkan akurasi prediksi cuaca. Penelitian ini berkontribusi dalam penerapan Decision Tree untuk prediksi cuaca di kota Seattle menggunakan WEKA, dan menyadari potensi pengembangan lebih lanjut untuk meningkatkan performa dan akurasi prediksi cuaca. Untuk penelitian lebih lanjut, dianjurkan untuk mengkaji atribut cuaca yang relevan, menggunakan data cuaca dari sumber yang berbeda, menguji model dengan data yang beragam, serta mempertimbangkan metode ensemble. Pengembangan model untuk prediksi cuaca jangka panjang dengan data historis juga perlu dipertimbangkan. Dengan langkah-langkah ini, diharapkan dapat meningkatkan akurasi dan performa model prediksi cuaca di kota X menggunakan metode Decision Tree dan aplikasi WEKA.

Kata Kunci: Data Mining, Klasifikasi, Decision Tree, Cuaca, Algoritma J48, Weka

1. PENDAHULUAN

Teknologi Informasi telah mengalami perkembangan pesat dalam beberapa dekade terakhir dan telah menjadi pendorong utama transformasi dalam berbagai sektor. Salah satu bidang yang mendapat manfaat signifikan dari perkembangan teknologi informasi adalah Pembelajaran Data (*Data Mining*). Pembelajaran Data adalah proses mengeksplorasi dan menganalisis data besar untuk mengidentifikasi pola, hubungan, dan tren yang tersembunyi, serta mengambil informasi berharga untuk pengambilan keputusan.

Analisis data menjadi sangat penting karena cuaca adalah salah satu aspek alam yang memiliki dampak langsung pada kehidupan manusia dan berbagai aktivitasnya. Informasi tentang cuaca yang akurat dan tepat waktu membantu masyarakat, perusahaan, dan pemerintah untuk mengambil langkah-langkah yang diperlukan untuk menghadapi perubahan cuaca dan memitigasi risiko yang mungkin terjadi.

Prediksi cuaca adalah salah satu aplikasi utama dari analisis data cuaca. Dengan menggunakan teknologi informasi dan algoritma pembelajaran mesin seperti *Decision Tree*, prediksi cuaca dapat dilakukan dengan lebih akurat dan efisien. *Decision Tree* adalah metode klasifikasi berstruktur pohon yang digunakan untuk memprediksi nilai dari variabel target berdasarkan fitur-fitur yang ada[1]. Algoritma J48 membangun sebuah pohon keputusan berdasarkan pada seperangkat input data yang berlabel[2]. Model yang dihasilkan dengan Algoritma C4.5 (algoritma J48 dalam WEKA) yang dihasilkan dalam proses *training* dari data pelatihan berupa sebuah pohon keputusan.

Decision Tree adalah struktur *flowchart* yang menyerupai *Tree* (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes[3], dan simpul daun merepresentasikan kelas atau distribusi kelas[4]. *Decision Tree* ialah metode analisis keputusan dengan mendapatkan dan membandingkan probabilitas dari *leave* dan *node* yang berbeda yang kemudian dievaluasi[5]. Akurasi hasil perhitungan menelusuri potensi dan minat calon mahasiswa menggunakan *decision tree* sebesar 90,3%[6]

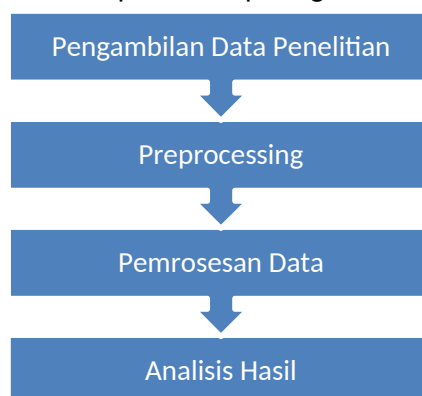
Analisa perhitungan cuaca menggunakan *decision tree* dapat menghasilkan akurasi sebesar 84,19%[1]. Hasil prediksi cuaca dari perhitungan menggunakan *decision tree* membantu bagi para pelintas di pulau Jawa untuk memperhatikan perubahan cuaca yang terjadi yang dapat berpengaruh terhadap keselamatan dan kenyamanan perjalanan, [7], *decision tree* dan pengujian aplikasi rapidminer mampu menghasilkan prediksi perubahan cuaca disuatu wilayah[8]. Dari tiga metode yang digunakan dalam memprediksi curah hujan, yang terdiri dari metode *Decision tree* C.45, KNN dan Naïve Bayes, menghasilkan akurasi yang paling tinggi yaitu metode *Decision tree* sebesar 88,03%[9]. *Decision tree* mampu menghasilkan klasifikasi bidang kesehatan seperti klasifikasi pada penyakit stroke[10].

Secara keseluruhan, perbedaan penelitian yang dilakukan dengan penelitian sebelumnya adalah terletak pada lokasi penelitian, dataset yang digunakan, algoritma

Data Mining yang diterapkan dan software pengujian yang digunakan. Penelitian ini menggunakan atribut tanggal, curah hujan, suhu maksimum, suhu minimum, kecepatan angin, dan kondisi cuaca. Berbeda dengan penelitian[7] atribut yang digunakan dalam penelitian tersebut adalah wilayah, tanggal, jam, suhu, arah angin, angin, kelembapan, dengan target berupa kondisi prakiraan cuaca dengan pengujian menggunakan software Rapidminer. Peneliti tentang prediksi cuaca bertujuan untuk mengetahui tingkat akurasi dari hasil analisa data dan performa model prediksi cuaca menggunakan metode *decision tree* algoritma J48 dan software pengujian Weka di kota Seattle. Kota Seattle dipilih karena memiliki variasi cuaca yang signifikan sepanjang tahun, sehingga analisis data dari kota ini akan memberikan wawasan yang berharga tentang penggunaan *decision tree* dalam prediksi cuaca.

2. METODE PENELITIAN

Pada penelitian ini dibagi menjadi beberapa tahapan yaitu tahap pencarian data, tahap *preprocessing*, tahap pemrosesan data dan tahap evaluasi hasil penelitian[11]. Alur penelitian dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

1. Pengumpulan data. Data diambil di laman web kaggle.com yaitu *dataset seattle weather* yang berisi 1461 cuaca di kota Seattle, dengan atribut yang ada dalam data tersebut antara lain seperti tanggal, curah hujan, suhu maksimum, suhu minimum, kecepatan angin, dan kondisi cuaca.
2. Pada tahap *preprocessing*, dilakukan *cleaning* (pembersihan) data menggunakan aplikasi Weka. *Cleaning data* adalah sebuah proses untuk mengidentifikasi, mengatasi, dan menghapus data yang salah atau tidak sempurna agar data lebih berkualitas dan siap digunakan dalam proses pemodelan data.
3. Pada tahap pemrosesan data, peneliti melakukan pemodelan data menggunakan aplikasi Weka dengan metode klasifikasi. Metode klasifikasi yang digunakan adalah *Decision Tree* (pohon keputusan) berbasis algoritma J48 untuk memprediksi cuaca di kota Seattle berdasarkan atribut cuaca seperti tanggal, curah hujan, suhu maksimum, suhu minimum, kecepatan angin, dan kondisi cuaca.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan data prediksi cuaca di kota Seattle ditahun 2012 sebanyak 1461 data yang diambil dari laman kaggle.com, yang memiliki 6 atribut.

3.2. Preprocessing Data

Seluruh atribut yang terdapat dalam dataset digunakan untuk memprediksi cuaca yaitu tanggal, curah hujan, suhu maksimum, suhu minimum, kecepatan angin dankondisi cuaca sehingga pada tahap preprocessing data ini tidak ada atribut yang dihilangkan.

3.3. Data Mining (Klasifikasi)

Algoritma J48 merupakan algoritma perkembangan dari C4.5 yaitu versi C4.8 yang ditulis ulang dalam Java (bahasa yang digunakan di Weka) karena hal tersebut, maka di aplikasi Weka bernama J48. Karena peneliti menggunakan *training set* dan *tes set*, maka seluruh dataset (1461 data) digunakan sebagai *use training set*. Berikut hasil pengujian menggunakan aplikasi Weka tersaji pada gambar 2.

```
Classifier output
Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    seattle-weather - SYARLA 2113010311
Instances:   1461
Attributes:  6
             date
             precipitation
             temp_max
             temp_min
             wind
             weather
Test mode:   evaluate on training data

=== Classifier model (full training set) ===

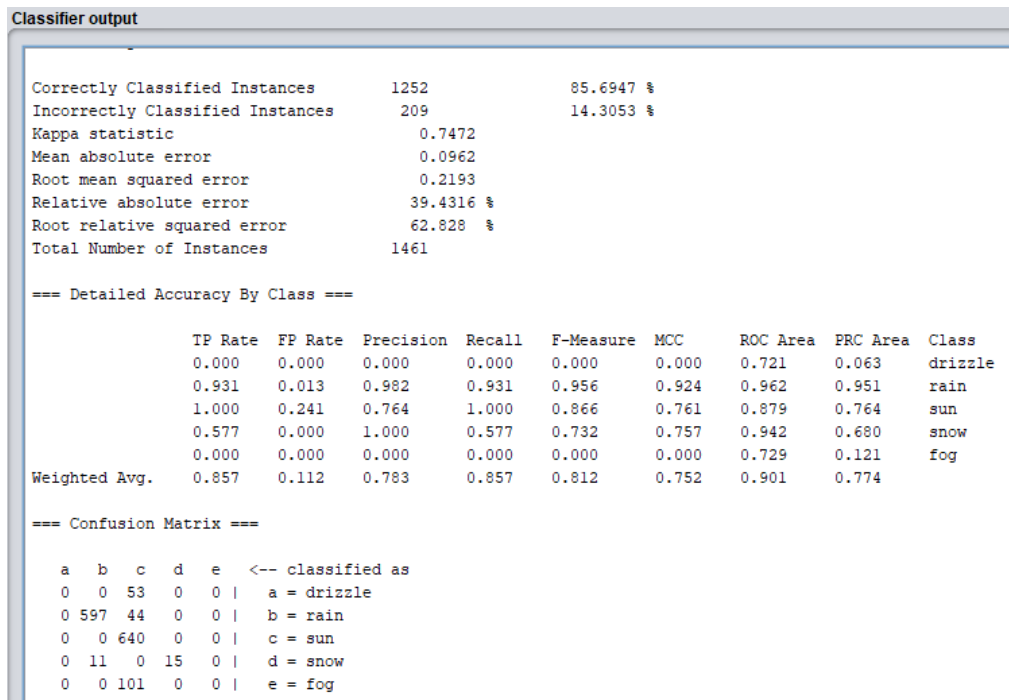
J48 pruned tree
-----

precipitation <= 0: sun (838.0/198.0)
precipitation > 0
|   temp_min <= 0.6
|   |   wind <= 3.7
|   |   |   temp_max <= 3.3: snow (3.0)
|   |   |   temp_max > 3.3: rain (17.0/4.0)
|   |   |   wind > 3.7: snow (12.0)
|   |   |   temp_min > 0.6: rain (591.0/7.0)
|   |   temp_min > 0.6: rain (591.0/7.0)

Number of Leaves :    5
Size of the tree :    9
```

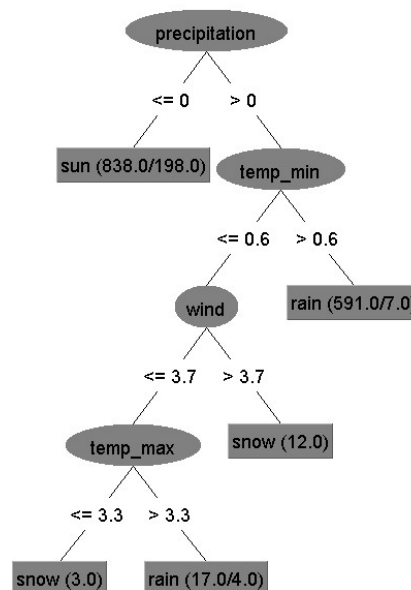
Gambar 2. Hasil Pengujian menggunakan aplikasi Weka

Jika curah hujan ≤ 0 maka cuaca cerah. Jika curah hujan > 0 , suhu minimum $\leq 0,6$, kecepatan angin $3,7$ dan suhu maksimum $\leq 3,3$ maka salju akan turun. Jika curah hujan > 0 , suhu minimum $\leq 0,6$, kecepatan angin $3,7$ dan suhu maksimum $\leq 3,3$ maka cuaca hujan. Jika curah hujan > 0 , suhu minimum $\leq 0,6$, kecepatan angin $> 3,7$ maka cuaca bersalju. Jika curah hujan > 0 dan suhu minimum $> 0,6$ maka turun hujan (Gambar 2).



Gambar 3. Hasil akurasi menggunakan Weka

Hasil klasifikasi menunjukkan terdapat 5 jenis cuaca yaitu gerimis, hujan, cerah, bersalju dan berkabut. Hasil akurasi menggunakan aplikasi Weka adalah sebesar 85,6946% atau sebanyak 1252 data yang diklasifikasikan dengan benar dan 14,3053% atau sebanyak 209 data diklasifikasikan salah ditunjukkan pada gambar 3 diatas. Struktur pohon keputusan sebanyak 5 node daun terbentuk dan 9 node untuk ukuran pohon keputusan tersebut tersaji pada gambar 4.



Gambar 4. Struktur pohon keputusan prediksi kondisi cuaca

3.4. Analisis Hasil

Pengujian data di penelitian ini menggunakan aplikasi WEKA untuk memprediksi cuaca di kota Seattle. Program ini menggunakan metode yang disebut Decision Tree (pohon keputusan) algoritma J48 untuk membantu memprediksi cuaca. Dari hasil penelitian, didapatkan bahwa model prediksi cuaca yang dibuat berhasil dengan baik, dengan tingkat keakuratan sekitar 85,69%. Dari 1461 data cuaca yang dipelajari, 1252 data berhasil diprediksi dengan benar, dan 209 data diprediksi dengan kesalahan. Ini menunjukkan bahwa modelnya bekerja dengan baik, tapi masih ada beberapa kasus yang tidak tepat.

Pohon keputusan ini memiliki total 9 simpul (*nodes*) dan 5 daun (*leaves*). Daun dan simpul pada pohon keputusan ini mewakili keputusan atau aturan yang digunakan untuk mengklasifikasikan cuaca. Setiap simpul mengandung sebuah kondisi berdasarkan atribut cuaca tertentu, seperti curah hujan (*precipitation*), suhu minimum (*temp_min*), dan kecepatan angin (*wind*). Pada awal pohon keputusan, terdapat simpul yang berhubungan dengan curah hujan. Jika curah hujan kurang dari atau sama dengan 0, cuaca diprediksi sebagai "sun" (matahari). Namun, jika curah hujan lebih dari 0, model melanjutkan ke simpul berikutnya yang terkait dengan suhu minimum.

Selanjutnya, pada simpul berikutnya, model mempertimbangkan suhu minimum (*temp_min*). Jika suhu minimum kurang dari atau sama dengan 3.3, model akan melihat simpul berikutnya terkait dengan kecepatan angin (*wind*). Pada simpul ini, jika kecepatan angin kurang dari atau sama dengan 3.7, cuaca diprediksi sebagai "rain" (hujan). Namun, jika kecepatan angin lebih dari 3.7, cuaca diprediksi sebagai "snow" (salju).

4. KESIMPULAN

1. Penggunaan metode *Decision Tree* (Algoritma J48) melalui aplikasi WEKA berhasil menghasilkan model prediksi cuaca di kota Seattle dengan tingkat akurasi sebesar 85,69%. Model ini memiliki struktur yang sederhana dan mudah dipahami, namun perlu diperhatikan adanya keterbatasan dalam mengklasifikasikan beberapa kelas cuaca.
2. Dibandingkan dengan penelitian sebelumnya [7] yang menggunakan dataset sebanyak 1.680 data dengan pengujian menggunakan aplikasi Rapidminer dan hasil akurasi 81,94%, maka penelitian tentang prediksi kondisi cuaca di kota Seattle ini menghasilkan akurasi lebih baik yaitu sebesar 85,69%.
3. Model ini bisa dikembangkan untuk analisis prediksi cuaca dalam jangka waktu yang lebih panjang dengan mengintegrasikan data cuaca historis. Dengan demikian, penelitian ini memberikan kontribusi penting dalam penerapan *Decision Tree* J48 untuk prediksi cuaca di kota Seattle menggunakan WEKA, sambil menyadari potensi

pengembangan lebih lanjut untuk meningkatkan performa dan akurasi prediksi cuaca.

5. SARAN

1. Dalam penelitian lebih lanjut, perlu mengkaji atribut cuaca yang relevan, gunakan data cuaca dari sumber berbeda, uji model dengan data yang beragam, dan kembangkan model untuk prediksi cuaca jangka panjang dengan data historis.
2. Diharapkan hal ini meningkatkan akurasi dan performa model prediksi cuaca di kota Seattle menggunakan metode Decision Tree J48 dan aplikasi WEKA.

DAFTAR PUSTAKA

- [1] F. F. Kusuma, "Penerapan Data Mining Untuk Akurasi Analisis Cuaca di Australia Menggunakan Algoritma J48 Decision Tree," *J. Comput. Sci. Inf. Syst. J-Cosys*, vol. 3, no. 2, pp. 65–68, 2023, doi: 10.53514/jco.v3i2.396.
- [2] K. Sinaga, E. Buulolo, and B. Nadeak, "Implementasi Algoritma Decision Tree_J48 untuk Memprediksi Resiko Kredit pada Koperasi Simpan Pinjam (Studi Kasus : Kofipindo Lubuk Pakam)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 20–24, 2019, doi: 10.30865/komik.v3i1.1561.
- [3] D. A. P. Situmorang and G. L. Ginting, "Penerapan Data Mining Algoritma J48 Untuk Mengidentifikasi Faktor-Faktor Kecelakaan Kerja," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 5, pp. 530–536, 2020, doi: 10.30865/jurikom.v7i4.2277.
- [4] I. Rukmana, A. Rasheda, F. Fathulhuda, M. R. Cahyadi, and F. Fitriyani, "Analisis Perbandingan Kinerja Algoritma Naïve Bayes, Decision Tree-J48 dan Lazy-IBK," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1038, 2021, doi: 10.30865/mib.v5i3.3055.
- [5] R. S. Lutfiyani and N. Retnowati, "Implementasi Pendeteksian Spam Email Menggunakan Metode Text Mining Dengan Algoritma Naïve Bayes Dan Decision Tree J48," *J. Komput. dan Inform.*, vol. 9, no. 2, pp. 244–252, 2021, doi: 10.35508/jicon.v9i2.5304.
- [6] B. Budiman, "Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru," *Nuansa Inform.*, vol. 15, no. 2, pp. 37–52, 2021, doi: 10.25134/nuansa.v15i2.4162.
- [7] D. K. Pramudito, "DATA MINING IMPLEMENTATION ON JAVA NORTH COAST WEATHER FORECAST DATASET USING C4.5 ALGORITHM," *J. Teknol. Pelita Bangsa*, vol. 13, no. September, pp. 139–148, 2022.
- [8] F. Indriaharti Harida and N. Khazizah, "Analisis Cuaca Di Kota Jakarta Bulan Januari Tahun 2018 Menggunakan Algoritma Decision Tree," *Poros Tek.*, vol. 14, no. 1, pp. 33–37, 2022, [Online]. Available: <https://www.kaggle.com/datasets/msf1203/pr>
- [9] A. Al Arif, M. Firdaus, Rahmadden, and Y. Maruhawa, "Perbandingan Metode Data Mining untuk Prediksi Curah Hujan dengan Algoritma C4.5, Naïve Bayes, dan KNN," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 187–197, 2022.

- [10] I. Lishania, R. Goejantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *J. Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/571>
- [11] R. Ramadhani, Y. Hendriyani, and I. Pendahuluan, "Prediksi Prestasi Siswa Berbasis Data Mining Menggunakan Algoritma Decision Tree (Studi Kasus : SMKN 2 Padang)," vol. 9, no. 3, 2021.