



Perbandingan Algoritma *Random Forest* dan *Decision Tree* untuk Memprediksi Cuaca di Kalimantan Barat

Nur Ayu Fitri Maharani^{1*}, Putri Yuli Utami², Rizki Surtiyan Surya³

¹⁻³ Program Studi Sistem Informasi, Fakultas Teknik Dan Ilmu Komputer, Universitas Muhammadiyah Pontianak, Indonesia

*Penulis Korespondensi: nurayufitrimaharani03@gmail.com¹

Abstract. This research aims to determine the results of weather predictions in West Kalimantan. This research was conducted to compare the Random Forest and Decision Tree algorithms in predicting weather in West Kalimantan. The data used comes from the BMKG West Kalimantan Climate Station from 2022-2024 including attributes, namely, date, average temperature (Tavg), average humidity (RH_avg), rainfall (RR), duration of sunlight (ss), and average wind speed (ff-avg). The initial stage of research using the CRISP-DM method is to collect climate data obtained from the official website of the Meteorology, Climatology and Geophysics Agency (BMKG), then process the data, normalize the data, and then apply the Random Forest and decision tree algorithms using the RapidMiner tool. Based on the results of weather prediction research with a comparison of Random Forest algorithm test data with an accuracy of 94.64% and Random Forest algorithm training data with an accuracy of 94.05%. And to compare the decision tree algorithm test data with an accuracy of 93.45% and the decision tree algorithm training data with an accuracy of 92.26%.

Keywords: BMKG; Decision Tree; Prediction; Random Forest; Weather.

Abstrak. Penelitian ini bertujuan untuk mengetahui hasil prediksi cuaca di Kalimantan Barat. Penelitian ini dilakukan untuk membandingkan algoritma random forest dan decision tree dalam memprediksi cuaca di Kalimantan Barat. Data yang digunakan berasal dari BMKG Stasiun Iklim Kalimantan Barat dari tahun 2022-2024 mencakup atribut yaitu, tanggal, temperatur rata-rata (Tavg), kelembapan rata-rata (RH_avg), curah hujan (RR), lamanya penyinaran matahari (ss), dan kecepatan angin rata-rata (ff-avg). Tahapan penelitian yang menggunakan metode CRISP-DM untuk langkah awalnya adalah dengan mengumpulkan data iklim yang diperoleh dari website resmi Badan Meteorologi Klimatologi dan Geofisika (BMKG) kemudian dilakukan proses data, normalisasi data, dan selanjutnya melakukan penerapan algoritma random forest dan decision tree menggunakan tools RapidMiner. Berdasarkan hasil penelitian prediksi cuaca dengan perbandingan data uji algoritma Random Forest hasil accuracy sebesar 94.64% dan data latih algoritma Random Forest hasil accuracy sebesar 94.05%. Dan untuk perbandingan data uji algoritma decision tree hasil accuracy sebesar 93.45% dan data latih algoritma decision tree hasil accuracy sebesar 92.26%.

Kata kunci: BMKG; Cuaca; Hutan Acak; Pohon Keputusan; Prediksi.

1. LATAR BELAKANG

Garis Khatulistiwa (lintang 0°) melintasi Kalimantan Barat, yang terletak di atas kota Pontianak. Akibatnya, wilayah ini memiliki lingkungan tropis lembap dengan suhu dan kelembapan tinggi sepanjang tahun. Saat ini, Kalimantan Barat sebagian besar berawan dengan suhu tertinggi sekitar 25°C. Terdapat dua musim utama: musim kemarau, yang berlangsung dari Mei hingga September, dan musim hujan, yang berlangsung dari Oktober hingga April dan mencapai puncaknya pada bulan Desember. Wilayah Kalimantan Barat berpotensi mengalami hujan sedang hingga lebat, disertai badai petir dan hembusan angin kencang dalam waktu singkat (Agus Setyaningsih, 2017).

Sebagian besar Kalimantan Barat rentan terhadap kondisi cuaca ekstrem, termasuk hujan sedang hingga lebat, angin kencang, dan badai yang dapat menyebabkan banjir, tanah longsor,

dan kerusakan infrastruktur, menurut prakiraan cuaca Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Kalimantan Barat. Wilayah Kalimantan Barat sering terjadi banjir yang meluas akibat curah hujan tinggi, terutama di daerah yang rawan banjir seperti di sekitar sungai-sungai besar, dan kasus karhutla sering muncul selama periode kekeringan ekstrem di musim tanpa hujan berkepanjangan.

Cuaca sangat penting dalam setiap aspek kehidupan. Aktivitas manusia dipengaruhi oleh fluktuasi cuaca yang signifikan. Aktivitas yang berkaitan dengan cuaca meliputi hortikultura, penerbangan, dan pertanian. Memantau cuaca sangatlah penting agar siap menghadapi kemungkinan kejadian seperti badai petir dan hujan lebat. Masyarakat dapat memperkirakan kemungkinan kejadian dan mengambil tindakan pencegahan yang diperlukan untuk melindungi diri dari bahaya yang tidak diinginkan dengan memperkirakan kondisi cuaca di masa mendatang secara akurat (Hamami & Dahlan, 2022). Fluktuasi cuaca ditentukan oleh sejumlah elemen alam, mencakup suhu udara, tekanan udara, kadar kelembapan, intensitas angin, serta posisi dan waktu tertentu. Berdasarkan faktor-faktor itu, dapat muncul beragam kondisi cuaca seperti cerah, mendung, atau turun hujan.

Probabilitas perubahan cuaca dapat diantisipasi lebih tepat dengan penggunaan algoritma kategorisasi cuaca yang efisien. Teknik pembelajaran terawasi yang digunakan dalam klasifikasi antara lain jaringan saraf tiruan, mesin vektor pendukung, pohon keputusan, hutan acak, dan bayes naif. Metode yang digunakan dalam studi ini, termasuk hutan acak dan pohon keputusan, dapat digunakan untuk membuat prakiraan cuaca. Berkat struktur pohon keputusannya yang lugas, Hutan Acak dapat menangani outlier secara efektif dan meningkatkan akurasi model bahkan ketika terdapat data yang hilang. Namun, efektivitas pohon keputusan dalam konstruksinya menjadikannya bermanfaat. Pohon keputusan mudah dipahami, mengelola properti diskrit dan numerik, serta mencapai tingkat akurasi yang memadai (Rahma Ferika Shaumi, n.d. 2023).

Mengingat permasalahan yang telah disebutkan, peneliti bermaksud menggunakan perangkat RapidMiner Studio untuk melakukan studi prakiraan cuaca di Kalimantan Barat menggunakan dua algoritma: hutan acak dan pohon keputusan. Tanggal, suhu rata-rata, kelembapan rata-rata, curah hujan, lama penyinaran matahari, dan kecepatan angin rata-rata merupakan beberapa parameter kategoris yang termasuk dalam data klimatologi yang dikumpulkan dari BMKG. Data dari tahun 2022 hingga 2024 merupakan dataset yang digunakan.

2. KAJIAN TEORITIS

Badan Meteorologi Klimatologi dan Geofisika (BMKG)

Badan Meteorologi, Klimatologi, dan Geofisika adalah lembaga federal yang beroperasi di luar departemen. Sesuai dengan peraturan perundang-undangan yang berlaku, fungsi BMKG meliputi pelaksanaan kewajiban pemerintah di bidang meteorologi, klimatologi, kualitas udara, dan geofisika.

Cuaca

Keadaan atmosfer, langit, atau udara bumi disebut cuaca. Cuaca mencakup perubahan sinar matahari, curah hujan, angin, dan suhu. Konferensi Iklim Dunia mendefinisikan cuaca sebagai keadaan atmosfer secara keseluruhan, yang mencakup variasi, perkembangan, dan hambatan dari kejadian-kejadian tertentu. Di seluruh dunia, cuaca dapat berubah drastis dalam hitungan menit, jam, hari, atau minggu. Perbedaan suhu dan kelembapan di berbagai lokasi biasanya menjadi sumber fluktuasi cuaca. Kemiringan sumbu Bumi terhadap suatu wilayah dan pemanasan matahari adalah dua contoh faktor yang memengaruhi fluktuasi cuaca (Indriaharti Harida & Khazizah, 2022).

Algoritma Random Forest

Random Forest merupakan turunan dari teknik decision tree, di mana sejumlah pohon keputusan dibentuk secara kolektif dengan pemilihan variabel acak pada setiap percabangan. Kelebihanannya antara lain mampu menangani data yang tidak lengkap, tahan terhadap outlier, efisien dalam penyimpanan data, serta dapat meningkatkan performa klasifikasi melalui seleksi fitur yang efektif, bahkan pada data besar dan kompleks (Supriyadi et al., 2020).

Algoritma Decision Tree

Data dikonversi menjadi pohon keputusan dan aturan keputusan menggunakan algoritma pohon keputusan. Penambangan data dan masalah klasifikasi merupakan aplikasi ideal untuk teknik ini. Kelas-kelas yang dapat digunakan untuk klasifikasi baru dipetakan ke nilai atribut melalui metode pohon keputusan.

3. METODE PENELITIAN

Dalam penelitian ini, digunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) yang terdiri dari enam tahapan utama. Metodologi ini memberikan kerangka kerja yang terstruktur dan sistematis untuk melakukan analisis data dan membangun model prediksi cuaca yang akurat (Ibrahim, n.d.). Tujuan memprediksi cuaca adalah untuk memberikan informasi yang berguna kepada masyarakat tentang kondisi cuaca yang akan datang, misalnya meningkatkan akurasi, memperluas cakupan, atau menghasilkan prediksi

jangka panjang. Pada penelitian ini, langkah awalnya adalah dengan melakukan wawancara dengan Ibu Septikasari, S.Si staff BMKG di Stasiun Klimatologi Kalimantan Barat dan mengumpulkan data iklim yang didapat dari website resmi Badan Meteorologi Klimatologi dan Geofisika (BMKG) yaitu <https://dataonline.bmkg.go.id> dengan data online yang dapat diakses. Data tersebut merupakan data iklim dari 2022-2024. Parameter yang diterapkan yang meliputi suhu (T), kelembapan (RH), lama penyinaran matahari (ss), kecepatan angin (ff), dan curah hujan (RR). teknik pengumpulan data adalah observasi, teknik analisis data yang di gunakan adalah Normalisasi Data, Modelling, Evalution, Deployment,

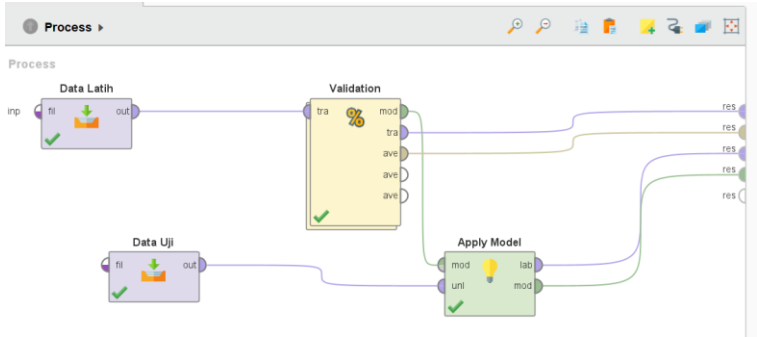
Implementasi Model Algoritma Random Forest

Keterangan masing-masing operator beserta parameter yang digunakan pada prediksi cuaca menggunakan algoritma random forest pada tabel 1.

Tabel 1. Operator Pemodelan Algoritma Random Forest.

Operator	Parameter	Nilai
Split Validation	Split	Relative
	Split Ratio	0.7
	Sampling Type	Automatic (Default)
Apply Model	Application Parameters	Default
	Number of Trees	100
	Criterion	Gain-ratio (Default)
Random Forest	Maximal Depth	10 (Default)
	Voting Strategy	Confidence
		(Default)
		vote
Performance	Accuracy	True (Default)
	Precision	False (Default)
	Recall	True (Default)
		True (Default)

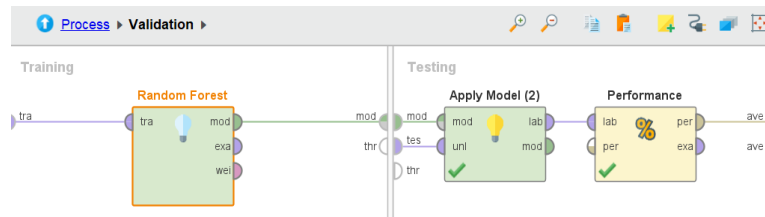
Setelah diketahui operator beserta parameter yang digunakan, kemudian akan dilakukan pemodelan menggunakan data latih untuk diterapkan pada data uji menggunakan algoritma random forest. Proses pemodelan algoritma random forest dikaji dengan tools RapidMiner yang ditampilkan pada gambar1



Gambar 1. Tampilan Proses Pada Algoritma Random Forest.

Gambar 1. ialah tahapan prosedur yang dilakukan pada algoritma random forest yaitu input data operator read excel untuk mengimpor dataset yang akan digunakan sebagai

data latih lalu setelah itu menambahkan *operator validation (split validation)* dan drag data uji tahun 2024 lalu tambahkn *operator apply model* kemudian sambungkan *output mod (model)* dari *operator validation* ke input mod (model) pada *operator apply model* sambungkan *output* dari data uji ke input unl (*unlabeled examples*) pada *operator apply model*. Setelah itu sambungkan ke masing-masing *port result* dari *operator operator* yang digunakan. Selanjutnya klik 2 kali pada *validation* maka akan menuju ke tampilan proses *training-testing* pada algoritma seperti gambar 2



Gambar 2. Tampilan training testing pada algoritma *Random Forest*.

Gambar 2 setelah menyelesaikan proses pemodelan, tambahkan *operator random forest*. Selanjutnya, tambahkan *operator apply model* dan *performance*. *Operator apply model* digunakan untuk menerapkan model prediksi yang telah dilatih pada data uji yang labelnya belum diketahui sedangkan *Operator performance* digunakan untuk mengkaji performa model prediksi dengan menampilkan skor *confusion matrix*.

Implementasi Model Algoritma *Decision Tree*

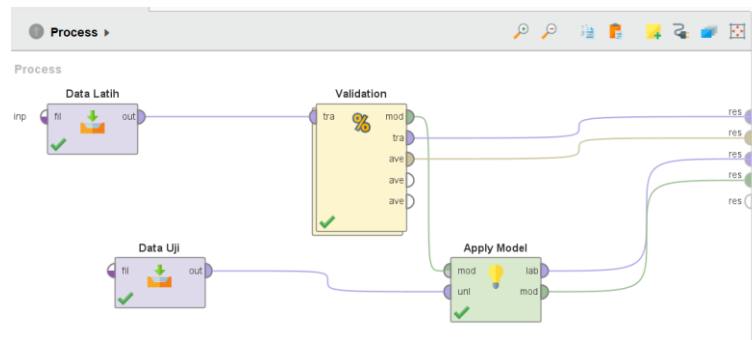
Keterangan masing-masing *operator* beserta parameter yang digunakan pada prediksi cuaca menggunakan algoritma *decision tree* pada tabel berikut ini

Tabel 2. *Operator* Pemodelan Algoritma *Decision Tree*.

<i>Operator</i>	<i>Parameter</i>	<i>Nilai</i>
<i>Split Validation</i>	<i>Split</i>	<i>Relative</i>
	<i>Split Ratio</i>	0.7
	<i>Sampling Type</i>	<i>Automatic (Default)</i>
<i>Apply Model</i>	<i>Application Parameters</i>	<i>Default</i>
	<i>Criterion</i>	<i>Gain ratio (Default)</i>
	<i>Maximal Depth</i>	10 (<i>Default</i>)
	<i>Confidence</i>	0.1 (<i>Default</i>)
<i>Decision Tree</i>	<i>Minimal gain</i>	0.01 (<i>Default</i>)
	<i>Minimal leaf size</i>	2 (<i>Default</i>)
	<i>Minimal size for split</i>	4 (<i>Default</i>)
	<i>Number of prepruning alternatives</i>	3 (<i>Default</i>)
	<i>Accuracy</i>	<i>True (Default)</i>
<i>Performance</i>		<i>False (Default)</i>
	<i>Precision</i>	<i>True (Default)</i>
	<i>Recall</i>	<i>True (Default)</i>

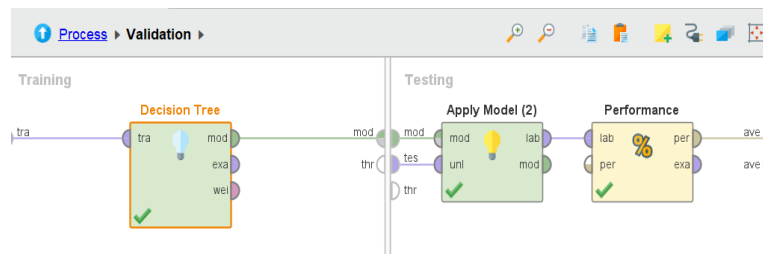
Setelah diketahui *operator* beserta parameter yang digunakan, kemudian akan dilakukan pemodelan menggunakan data latih untuk dijalankan pada data uji menggunakan

algoritma Decision Tree. Proses pembuatan model dilakukan melalui aplikasi RapidMiner yang divisualisasikan pada Gambar 3.



Gambar 3. Tampilan Proses Pada Algoritma *Decision Tree*.

Gambar 3 ialah tahapan prosedur yang dilakukan pada algoritma *decision tree* yaitu *input data operator read excel* untuk mengimpor *dataset* yang akan digunakan sebagai data latih lalu setelah itu menambahkan *operator validation (split validation)* dan drag data uji tahun 2024 lalu tambahkan *operator apply model* kemudian sambungkan *output mod (model)* dari *operator validation* ke *input mod (model)* pada *operator apply model* sambungkan *output* dari data uji ke *input unl (unlabeled examples)* pada *operator apply model*. Setelah itu sambungkan ke masing-masing *port result* dari *operator* yang digunakan. Selanjutnya klik 2 kali pada *validation* maka akan menuju ke tampilan proses *training-testing* pada algoritma seperti gambar 4



Gambar 4. Tampilan Training Testing Pada Algoritma *Decision Tree*.

Gambar 4 Setelah menyelesaikan proses pemodelan, tambahkan *operator decision tree*. Selanjutnya, tambahkan *operator apply model* dan *performance*. *Operator apply model* digunakan untuk menerapkan model prediksi yang telah dilatih pada data uji yang labelnya belum diketahui sedangkan *operator performance* digunakan untuk mengevaluasi kinerja *model* prediksi dengan menampilkan nilai *confusion matrix*.

4. HASIL DAN PEMBAHASAN

Perbandingan Data Latih dan Data Uji *Decision Tree*

Tabel 3. Perbandingan Data Latih Dan Data Uji Algoritma *Decision Tree*.

Perbandingan	Data Latih	Data Uji
60:40	92.44%	89.29%
70:30	92.20%	88.10%
80:20	93.45%	92.86%
90:10	92.86%	86.21%

Pada tabel 1 menunjukan dengan 4 kali percobaan bahwa *accuracy* perbandingan 80:20 pada algoritma *decision tree* memiliki tingkat *accuracy* yang optimal dengan nilai data latih 93.45% dan data uji 92.86% dibandingkan perbandingan 60:40, 70:30, dan 90:10.

Evaluasi Kinerja Algoritma

Evaluasi Kinerja Algoritma Random Forest

Hasil pengujian data uji klasifikasi pada algoritma *random forest* dalam melakukan prediksi digunakan operator *Performance* untuk memperoleh nilai akurasi, yang ditampilkan pada Gambar

accuracy: 94.64%

	true Cerah	true Hujan	true Berawan	class precision
pred. Cerah	110	0	6	94.83%
pred. Hujan	1	46	0	97.87%
pred. Berawan	2	0	3	60.00%
class recall	97.35%	100.00%	33.33%	

Gambar 5. Hasil Pengujian Kinerja Data Uji Algoritma *Random Forest*.

Gambar 5 menunjukkan bahwa algoritma *random forest* me *accuracy* sejumlah 94.64% dan mendapatkan 110 data diprediksi cerah dan kenyataannya benar cerah, 6 data diprediksi cerah padahal sebenarnya berawan. 1 data diprediksi hujan padahal sebenarnya cerah, 46 data diprediksi hujan dan kenyataannya benar, 2 data diprediksi berawan padahal sebenarnya cerah, 3 data diprediksi berawan dan kenyataannya berawan.

Perhitungan *accuracy* menggunakan rumus (3) yaitu :

$$Accuracy = \frac{110+46+3}{110+46+3+6+1+2} \times 100\% = 0,94642 = 94.64\%$$

Perhitungan *precision* menggunakan rumus (4) yaitu :

$$Precision (Cerah) = \frac{110}{110+6} \times 100\% = 0,94827 = 94.83\%$$

$$Precision (Hujan) = \frac{46}{46+1} \times 100\% = 0,97872 = 97.87\%$$

$$Precision (Berawan) = \frac{3}{3+2} \times 100\% = 0,6 = 60.00\%$$

Perhitungan *recall* menggunakan rumus (5) yaitu :

$$Recall (Cerah) = \frac{110}{110+1+2} \times 100\% = 0,97345 = 97.35\%$$

$$\text{Recall (Hujan)} = \frac{46}{46+0} \times 100\% = 1 = 100.00\%$$

$$\text{Recall (Berawan)} = \frac{3}{3+6} \times 100\% = 0,33333 = 33.33\%$$

Pada tahap pengujian data latih, algoritma Random Forest diterapkan untuk klasifikasi dan prediksi dengan bantuan operator *Performance* sehingga diperoleh nilai akurasi, seperti ditampilkan pada Gambar 6

accuracy: 94.05%

	true Cerah	true Hujan	true Berawan	class precision
pred. Cerah	111	0	7	94.07%
pred. Hujan	1	46	1	95.83%
pred. Berawan	1	0	1	50.00%
class recall	98.23%	100.00%	11.11%	

Gambar 6 . Hasil Pengujian Kinerja Data Latih Algoritma *Random Forest*.

Gambar 6 mendapati bahwa algoritma *random forest* mempunyai *accuracy* sejumlah 94.05% dan mendapatkan 111 data diprediksi cerah dan kenyataannya benar cerah, 7 data diprediksi cerah padahal sebenarnya berawan. 1 data diprediksi hujan padahal sebenarnya cerah, 46 data diprediksi hujan dan kenyataannya benar, 1 data diprediksi hujan padahal sebenarnya berawan, 1 data diprediksi berawan padahal sebenarnya cerah, 1 data diprediksi berawan dan kenyataannya berawan.

Perhitungan *accuracy* menggunakan rumus (3) yaitu :

$$\text{Accuracy} = \frac{111+46+1}{111+46+1+7+1+1+1} \times 100\% = 0,94047 = 94.05\%$$

Perhitungan *precision* menggunakan rumus (4) yaitu :

$$\text{Precision (Cerah)} = \frac{111}{111+7} \times 100\% = 0,94067 = 94.07\%$$

$$\text{Precision (Hujan)} = \frac{46}{46+2} \times 100\% = 0,95833 = 95.83\%$$

$$\text{Precision (Berawan)} = \frac{1}{1+1} \times 100\% = 0,5 = 50.00\%$$

Perhitungan *recall* menggunakan rumus (5) yaitu :

$$\text{Recall (Cerah)} = \frac{111}{111+2} \times 100\% = 0,9823 = 98.23\%$$

$$\text{Recall (Hujan)} = \frac{46}{46+0} \times 100\% = 1 = 100.00\%$$

$$\text{Recall (Berawan)} = \frac{1}{1+8} \times 100\% = 0,1111 = 11.11\%$$

Evaluasi Kinerja Algoritma Decision Tree

Pada tahap pengujian data uji, algoritma Decision Tree digunakan untuk melakukan klasifikasi dan prediksi dengan bantuan operator *Performance* sehingga diperoleh nilai akurasi, seperti ditampilkan pada Gambar

accuracy: 93.45%

	true Cerah	true Hujan	true Berawan	class precision
pred. Cerah	108	1	5	94.74%
pred. Hujan	1	45	0	97.83%
pred. Berawan	4	0	4	50.00%
class recall	95.58%	97.83%	44.44%	

Gambar 7. Hasil Pengujian Kinerja Data Uji Algoritma *Decision Tree*.

Gambar 7 mendapati bahwa algoritma *decision tree* mempunyai *accuracy* sejumlah 93.45% dan mendapatkan 108 data diprediksi cerah dan kenyataannya benar cerah, 1 data diprediksi cerah padahal hujan, 5 data diprediksi cerah padahal sebenarnya berawan. 1 data diprediksi hujan padahal sebenarnya cerah, 45 data diprediski hujan dan kenyataannya benar, 4 data diprediksi berawan padahal sebenarnya cerah, 4 data diprediksi berawan dan kenyataannya berawan.

Perhitungan *accuracy* menggunakan rumus (3) yaitu :

$$Accuracy = \frac{108+45+4}{108+45+4+1+5+1+4} \times 100\% = 0,93452 = 93.45\%$$

Perhitungan *precision* menggunakan rumus (4) yaitu :

$$Precision (Cerah) = \frac{108}{108+1+5} \times 100\% = 0,94736 = 94.74\%$$

$$Precision (Hujan) = \frac{45}{45+1} \times 100\% = 0,97826 = 97.83\%$$

$$Precision (Berawan) = \frac{4}{4+4} \times 100\% = 0,5 = 50.00\%$$

Perhitungan *recall* menggunakan rumus (5) yaitu :

$$Recall (Cerah) = \frac{108}{108+1+4} \times 100\% = 0,9557 = 95.58\%$$

$$Recall (Hujan) = \frac{45}{45+1} \times 100\% = 0,97826 = 97.83\%$$

$$Recall (Berawan) = \frac{4}{4+5+4} \times 100\% = 0,44444 = 44.44\%$$

Pengujian data uji klasifikasi pada algoritma *Decision Tree* dilakukan dengan operator *Performance* untuk menentukan tingkat akurasi, yang ditampilkan pada Gambar 8

accuracy: 92.26%

	true Cerah	true Hujan	true Berawan	class precision
pred. Cerah	108	2	6	93.10%
pred. Hujan	1	44	0	97.78%
pred. Berawan	4	0	3	42.86%
class recall	95.58%	95.65%	33.33%	

Gambar 8. Hasil Pengujian Kinerja Data Latih Algoritma *Decision Tree*.

Gambar 8 mendapati bahwa algoritma *decision tree* mempunyai *accuracy* sejumlah 92.26% dan memperoleh 108 data diprediksi cerah dan kenyataannya benar cerah, 2 data diprediksi cerah padahal hujan, 6 data diprediksi cerah padahal sebenarnya berawan. 1 data

diprediksi hujan padahal sebenarnya cerah, 44 data diprediski hujan dan kenyataannya benar, 4 data diprediksi berawan padahal sebenarnya cerah, 3 data diprediksi berawan dan kenyataannya berawan.

Perhitungan *accuracy* menggunakan rumus (3) yaitu :

$$Accuracy = \frac{108+44+3}{108+44+3+2+6+1+4} \times 100\% = 0,92261 = 92.26\%$$

Perhitungan *precision* menggunakan rumus (4) yaitu :

$$Precision (Cerah) = \frac{108}{108+2+6} \times 100\% = 0,93103 = 93.10\%$$

$$Precision (Hujan) = \frac{44}{44+1} \times 100\% = 0,9778 = 97.78\%$$

$$Precision (Berawan) = \frac{3}{3+4} \times 100\% = 0,42857 = 42.86\%$$

Perhitungan *recall* menggunakan rumus (5) yaitu :

$$Recall (Cerah) = \frac{108}{108+1+4} \times 100\% = 0,9557 = 95.58\%$$

$$Recall (Hujan) = \frac{44}{44+2} \times 100\% = 0,95652 = 95.65\%$$

$$Recall (Berawan) = \frac{3}{3+6} \times 100\% = 0,33333 = 33.33\%$$

Confusion Matrix

Untuk mengevaluasi dan membandingkan performa kedua algoritma, dilakukan pengujian menggunakan *confusion matrix*, dan hasil perbandingan tersebut ditampilkan pada Tabel 2

Tabel 4. *Confusion matrix.*

Algoritma	Accuracy	Cerah	Precision			Recall	
			Berawan	Hujan	Cerah	Berawan	Hujan
Random Forest	94.64%	94.83%	60.00%	97.87%	97.35%	100.00%	33.33%
Decision Tree	93.45%	94.74%	50.00%	97.83%	95.58%	44.44%	97.83%

Untuk menentukan metode mana yang lebih baik untuk prediksi cuaca, Tabel 2 menunjukkan hasil perbandingan performa kedua algoritma. Dengan perbandingan akurasi 93,45% berbanding 94,64%, hasilnya menunjukkan bahwa algoritma pohon keputusan berkinerja lebih buruk daripada pendekatan hutan acak dalam hal akurasi, presisi, dan perolehan kembali. Angka presisi kategori Cerah adalah 94,74% dan 94,83%, kategori Berawan adalah 50,00% dan 60,00%, dan kategori Hujan adalah 97,87% dan 97,83%. Terakhir, perbandingan perolehan kembali menunjukkan bahwa kategori Cerah memiliki angka presisi 95,58% dan 97,35%, kategori Berawan memiliki angka presisi 44,44% dan 100,00%, dan kategori Hujan memiliki angka presisi 33,33% dan 97,83%.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang dilakukan menggunakan algoritma Random Forest dan Decision Tree, diperoleh beberapa kesimpulan meliputi:

Menurut temuan pengujian menggunakan algoritma Random Forest diperoleh bahwa curah hujan (RR) merupakan atribut paling berpengaruh dalam klasifikasi cuaca. Pada kedua sampel, nilai $RR \leq 5.700$ umumnya memprediksi cuaca cerah, sedangkan $RR > 5.700$ menghasilkan hujan atau berawan bergantung pada kombinasi nilai kelembapan rata-rata (RH_avg), temperatur rata-rata (Tavg), dan kecepatan angin rata-rata (Ff_avg). Hasil perhitungan Decision Tree menunjukkan bahwa RR memiliki nilai gain tertinggi sebesar 0,6687, diikuti oleh RH_avg, durasi penyinaran matahari (ss), Ff_avg, dan Tavg, sehingga memperkuat temuan bahwa curah hujan (RR) merupakan variabel kunci dalam penentuan kondisi cuaca.

Alat RapidMiner digunakan untuk membagi hasil prediksi cuaca, yang didasarkan pada 706 titik data, dengan rasio 80:20. Hasilnya adalah 143 data uji dan 563 data latih. Pendekatan hutan acak, yang menghasilkan prakiraan, memiliki tingkat akurasi 94,64%, sedangkan hasil prediksi menggunakan metode decision tree memiliki nilai accuracy sebesar 93.45%. Hal ini menunjukkan bahwa metode yang lebih baik dan optimal digunakan untuk memprediksi cuaca di Kalimantan Barat adalah algoritma random forest.

DAFTAR REFERENSI

- Akbar, H., & Sanjaya, W. K. (2023). Kajian performa metode class weight random forest pada klasifikasi imbalance data kelas curah hujan. *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, 3(1). <https://doi.org/10.20885/snati.v3i1.30>
- Algoritma, A., Pada, K., Rapidminer, S., & Ainurrohman, W. (2021). Akurasi algoritma klasifikasi pada software Rapidminer dan Weka. *Prosiding Seminar Nasional Matematika*, 4, 493–499. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Dwiyanti, Z. A., & Prianto, C. (2023). Prediksi cuaca Kota Jakarta menggunakan metode random forest. *Jurnal Tekno Insentif*, 17(2), 127–137. <https://doi.org/10.36787/jti.v17i2.1136>
- Faid, M., Jasri, M., & Rahmawati, T. (2019). Perbandingan kinerja tool data mining Weka dan Rapidminer dalam algoritma klasifikasi. *Teknika*, 8(1), 11–16. <https://doi.org/10.34148/teknika.v8i1.95>
- Ferika Shaumi, R. (n.d.). *ABSTRACT WEATHER FORECAST FOR BANDAR LAMPUNG CITY USING RANDOM FOREST AND C4.5*.
- Hamami, F., & Dahlan, A. (2022). Klasifikasi cuaca Provinsi DKI Jakarta menggunakan algoritma random forest dengan teknik oversampling. *Jurnal TEKNOINFO*, 16(1). <https://doi.org/10.33365/jti.v16i1.1533>

- Harida, I., F., & Khazizah, N. (2022). Analisis cuaca di Kota Jakarta bulan Januari tahun 2018 menggunakan algoritma decision tree. 14(1), 33–37.
<https://www.kaggle.com/datasets/msf1203/pr>
- Ibrahim, D. (n.d.). 2nd Seminar Nasional IPTEK Terapan (SENIT) 2017 Tegal-Indonesia.
<http://conference.poltektegal.ac.id/index.php/senit2017>
- Pramudito, D. K. (2021). Data mining implementation on Java North Coast weather forecast dataset using C4.5 algorithm. 12.
- Ranti, N., M., & Hanif, K. H. (2022). Klasifikasi penyakit kanker payudara menggunakan perbandingan algoritma machine learning. 3(1), 1–6.
<https://creativecommons.org/licences/by/4.0/> <https://doi.org/10.35960/ikomti.v3i1.766>
- Setyaningsih, A. (2017). Implementasi metode Kohonen untuk prediksi curah hujan (Studi Kasus: Kota Pontianak). *Kumpulan JurnaL Ilmu Komputer (KLIK)*, 04(02).
<https://doi.org/10.20527/klik.v4i2.105>
- Shalsabilla, S., Rachmawati, P., Prakusa, K. V., & Rihastuti, S. (n.d.). Penerapan data mining dengan metode decision tree untuk prediksi cuaca di Kota Seattle menggunakan aplikasi Weka.
- Siregar, A. M. (n.d.). *Accounting Information System PERBANDINGAN ALGORITME KLASIFIKASI UNTUK PREDIKSI CUACA*.
- Supriyadi, R., Gata, W., Maulidah, N., Fauzi, A., Komputer, I., & Nusa Mandiri. (2020). Penerapan algoritma random forest untuk menentukan kualitas anggur merah. 13(2), 67–75.
<http://journal.stekom.ac.id/index.php/E-Bisnispage67>
<https://doi.org/10.51903/e-bisnis.v13i2.247>
- Taqiyuddin, M., & Sasongko, B. (2024). Prediksi cuaca Kabupaten Sleman menggunakan algoritma random forest. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1683.
<https://doi.org/10.30865/mib.v8i3.7897>