

PREDIKSI CURAH HUJAN DI KOTA PADANG DENGAN METODE NAÏVE BAYES

Angela Chow¹, Teny Handhayani², Leonard Augusto Lienry³

³ Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: ¹angela.535220170@stu.untar.ac.id, ²tenyh@fti.untar.ac.id,

³leonard.535210045@stu.untar.ac.id

Abstrak

Kemampuan untuk memprediksi curah hujan merupakan suatu kapabilitas yang penting untuk dikembangkan karena kegunaannya dalam mendukung berbagai sektor kehidupan. Namun, variabilitas cuaca di berbagai daerah menyebabkan perlunya upaya memprediksi curah hujan yang disesuaikan dengan berbagai faktor cuaca seperti temperatur dan kelembapan pada daerah masing-masing. Dalam studi ini, penulis menggunakan algoritma supervised machine learning dalam bentuk metode klasifikasi Naïve Bayes untuk memprediksi curah hujan di Kota Padang. Data yang digunakan untuk melatih model Naïve Bayes adalah data harian iklim di Kota Padang dalam rentang waktu periode tahun 2010-2023. Berdasarkan hasil prediksi dari model Naïve Bayes yang telah dilatih, nilai akurasi prediksi model Naïve Bayes sebesar 55,87%. Kinerja dari model Naïve Bayes dinilai cukup terbatas. Model Naïve Bayes dapat memprediksi kriteria curah hujan berawan dan hujan ringan dengan baik, tetapi model belum dapat memprediksi kriteria hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem secara akurat. Hal ini diakibatkan kurangnya data untuk melatih model Naïve Bayes dalam memprediksi curah hujan di Kota Padang. Oleh karena itu, penulis menyarankan untuk menambah data dari sumber lain untuk meningkatkan akurasi dari model Naïve Bayes.

Kata kunci— Kota Padang, Temperatur, Kelembapan, Curah Hujan, Naïve Bayes

Abstract

The ability to predict rainfall is an important capability that needs to be developed due to its usefulness in supporting various sectors in life. However, weather variability in different regions causes the need to predict rainfall based on weather factors such as temperature and humidity in its respective region. In this study, the author used supervised machine learning algorithms in the form of the Naïve Bayes classification method to predict rainfall in Padang City. The data used to train the Naïve Bayes model consisted of daily climate data from Padang City during the period of 2010 to 2023. Based on the prediction results of the trained Naïve Bayes model, the accuracy of the predictions was 55.87%. The performance of the Naïve Bayes model was considered limited. The model was able to predict cloudy and light rainfall criteria accurately, but it is unable to accurately predict moderate, heavy, very heavy, and extreme rainfall criteria. This limitation was due to the lack of data to train the Naïve Bayes model in predicting rainfall in Kota Padang. Therefore, the authors suggest increasing the number of data from other sources to improve the accuracy of the Naïve Bayes model.

Keywords—Padang City, Temperature, Humidity, Rainfall, Naïve Bayes

1. PENDAHULUAN

Kota Padang merupakan ibukota Provinsi Sumatera Barat yang terletak di pesisir barat pulau Sumatera. Secara astronomi terletak di antara $0^{\circ}44'$ dan $01^{\circ}08'$ Lintang Selatan serta antara $100^{\circ}05'$ dan $100^{\circ}34'$ Bujur Timur [1]. Wilayah Kota Padang berada di daerah dengan iklim tropis basah yang dipengaruhi oleh angin Barat dan memiliki musim kemarau yang cenderung singkat. Studi ini akan berfokus dalam menganalisis tren temperatur rata-rata dan kelembapan rata-rata untuk memprediksi curah hujan di Kota Padang.

Cuaca adalah kondisi atmosfer yang terjadi pada waktu dan tempat tertentu [2]. Variabilitas cuaca dan dampaknya pada pola cuaca merupakan topik yang penting untuk dipelajari. Faktor-faktor yang mempengaruhi pola cuaca antara lain temperatur, kelembapan, dan curah hujan. Pemahaman yang lebih baik akan faktor-faktor ini juga akan meningkatkan akurasi dalam memprediksi perubahan cuaca. Informasi yang diperoleh dari model prediksi cuaca yang akurat akan mendukung pengambilan keputusan yang lebih tepat, meminimalkan resiko, dan meningkatkan efisiensi dalam berbagai sektor kehidupan. Prediksi curah hujan di Kota Padang akan dilakukan dengan menggunakan metode klasifikasi Naïve Bayes.

Metode klasifikasi Naïve Bayes merupakan algoritma *supervised machine learning* yang dibangun dari teorema Bayes untuk menyelesaikan permasalahan klasifikasi. Naïve Bayes membuat prediksi berdasarkan probabilitas suatu entitas untuk melakukan klasifikasi. Biasanya metode ini bekerja secara baik dan efektif pada dataset yang berukuran besar [3].

Sebelumnya, pernah dilakukan penelitian mengenai hubungan temperatur dan kelembapan udara terhadap curah hujan di Kabupaten Sleman dan Kota Pontianak [4, 5]. Studi mengenai klasifikasi curah hujan dengan metode Naïve Bayes sudah pernah dilakukan di Kota Bogor dan Provinsi Bali [6, 7]. Studi mengenai prediksi curah hujan di Kota Padang bertujuan untuk mengetahui pengaruh temperatur dan kelembapan terhadap curah hujan di kota Padang, menggunakan metode Naïve Bayes untuk memprediksi curah hujan di Kota Padang, dan menilai akurasi dari hasil prediksi yang telah dibuat dengan Metode Naïve Bayes.

2. METODE PENELITIAN

Prediksi curah hujan akan dilakukan menggunakan metode klasifikasi Naïve Bayes. Metode penelitian, alur kerja penelitian, data yang diperlukan, serta dasar teori dari penelitian ini akan dijelaskan dalam subbab-subbab berikut.

2.1 Data

Penelitian dilakukan terhadap cuaca di Kota Padang, Provinsi Sumatera Barat, Indonesia dengan data iklim harian periode waktu 1 Januari 2010 - 31 Maret 2023. Data iklim harian Kota Padang yang diperlukan untuk melakukan analisis bersumber dari aplikasi DATA ONLINE - PUSAT DATABASE – BMKG [8]. Badan Meteorologi Klimatologi dan Geofisika (BMKG) adalah lembaga pemerintah yang mengamati dan mengolah informasi mengenai cuaca, iklim, serta gempa dan tsunami. Data iklim diperoleh dari Stasiun Meteorologi Maritim Teluk Bayur yang terletak pada lintang -0.99639 , bujur 100.37222 , serta elevasi 2 m. Parameter yang dipilih untuk dianalisis adalah temperatur rata-rata ($^{\circ}\text{C}$), kelembapan rata-rata (%), dan curah hujan (mm).

Curah hujan adalah ketinggian air hujan yang jatuh pada tempat yang datar jika air tidak meresap, tidak mengalir, dan tidak menguap. Kriteria nilai curah hujan yang digunakan untuk mengklasifikasi intensitas hujan dalam tabel berikut [9].

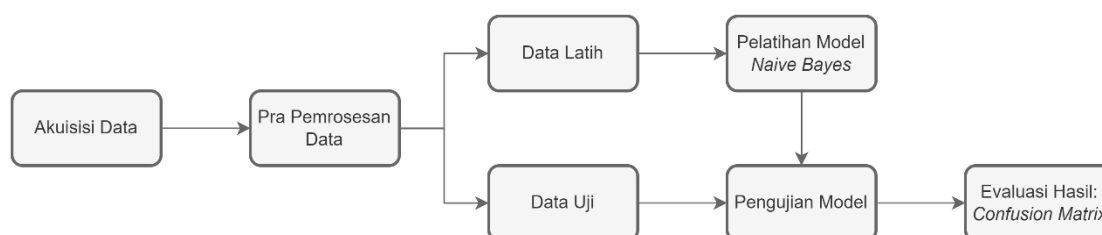
Tabel 1 Kriteria klasifikasi curah hujan berdasarkan intensitas curah hujan

Nilai Curah Hujan	Klasifikasi Curah Hujan
0 mm/hari	berawan
0,5 – 20 mm/hari	hujan ringan
20 – 50 mm/hari	hujan sedang
50 – 100 mm/hari	hujan lebat
100 – 150 mm/hari	hujan sangat lebat
>150 mm / hari	hujan ekstrem

Temperatur adalah ukuran panas atau dinginnya suatu benda atau cairan. Temperatur yang tinggi menimbulkan lebih banyak proses penguapan yang secara langsung meningkatkan proses kondensasi sehingga terjadi curah hujan yang tinggi. Kelembapan adalah jumlah uap air yang berada di udara, dan biasanya tidak kasat mata. Uap air diproduksi dari hasil evaporasi air laut, sungai, danau, dan tanaman.

2.2 Alur Kerja Penelitian

Alur kerja penelitian untuk melakukan prediksi curah hujan melalui beberapa tahapan, dimulai dari akuisisi data, prapemrosesan data, pembagian data menjadi data latih dan data uji, pelatihan model Naïve Bayes, pengujian model, dan evaluasi hasil. Alur kerja penelitian secara keseluruhan dapat diamati dalam Gambar 1.



Gambar 1 Diagram Alur Kerja Penelitian

Pada tahap akuisisi data, dataset parameter temperatur rata-rata, kelembapan rata-rata, dan curah hujan harian diunduh dari aplikasi DATA ONLINE - PUSAT DATABASE – BMKG. Seluruh parameter tersebut bersifat numerik. Dataset yang diperoleh sudah dalam bentuk data terstruktur yang berarti seluruh parameter sudah berada dalam kolom yang sesuai. Namun, dataset masih terpisah dalam file excel yang berbeda-beda per bulan dari tahun 2010 sampai 2023.

Selanjutnya, dilakukan tahap prapemrosesan data untuk menggabungkan data tahun 2010-2023 dan menangani data hilang serta nilai-nilai yang tidak sesuai. Dataset terdiri dari 4838 baris berisi data parameter temperatur rata-rata, kelembapan rata-rata, serta curah hujan harian. Parameter curah hujan kemudian dibagi menjadi beberapa kriteria berdasarkan intensitas curah hujan. Berdasarkan Tabel 1, curah hujan akan diklasifikasi ke dalam kriteria intensitas curah hujan: berawan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem. Dataset terdiri dari 1948 data curah hujan kriteria berawan, 1856 data curah hujan kriteria hujan ringan, 640 data curah hujan kriteria hujan sedang, 287 data curah hujan kriteria hujan lebat, 71 data curah hujan kriteria hujan sangat lebat, dan 36 data curah hujan kriteria hujan ekstrim.

Setelah melalui tahap prapemrosesan, dataset siap digunakan untuk kebutuhan analisis. Untuk lebih memahami parameter dalam dataset, dilakukan perhitungan korelasi menggunakan korelasi koefisien Pearson. Kemudian, dataset dipisah menjadi 75% data latih dan 25% data uji.

Data latih akan digunakan melatih model Naïve Bayes dan data uji akan digunakan untuk menguji model Naïve Bayes dalam memprediksi curah hujan.

Pelatihan model Naïve Bayes dalam memprediksi curah hujan menggunakan parameter temperatur rata-rata dan kelembapan rata-rata sebagai masukan dan kolom kriteria curah hujan sebagai target prediksi. Pengujian model Naïve Bayes dilakukan dengan menggunakan data uji. Model Naïve Bayes membuat prediksi berdasarkan parameter temperatur rata-rata dan kelembapan rata-rata dalam data uji. Hasil prediksi dari model Naïve Bayes kemudian dievaluasi untuk menilai kinerja model Naïve Bayes. Evaluasi model Naïve Bayes menggunakan *confusion matrix* dan dihitung akurasi dari hasil prediksi model Naïve Bayes.

2.3 Korelasi Koefisien

Metode analisis korelasi koefisien yang akan digunakan untuk mengetahui hubungan antara temperatur rata-rata, kelembapan rata-rata, dan curah hujan di Kota Padang adalah metode korelasi koefisien Pearson.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

Dalam rumus korelasi koefisien Pearson diatas, r adalah korelasi koefisien; X_i dan Y_i adalah nilai individual dari X dan Y ; \bar{X} dan \bar{Y} adalah nilai rata-rata dari X dan Y . Korelasi antara parameter dapat dinilai dari derajat korelasi dengan indikator: jika $r = 0$, maka tidak ada hubungan korelasi sama sekali; jika $0 < r < 0,3$, maka hubungan korelasi kurang berarti; jika $0,3 < r < 0,5$, maka hubungan korelasi rendah; jika $0,5 < r < 0,8$, maka hubungan korelasi sedang; jika $0,8 < r < 1$, maka hubungan korelasi tinggi; jika $r = 1$, maka hubungan korelasi sempurna [10]. Nilai korelasi positif menunjukkan adanya hubungan yang positif diantara kedua parameter, sedangkan nilai korelasi negatif menunjukkan adanya hubungan yang negatif diantara kedua parameter.

2.4 Metode Klasifikasi Naïve Bayes

Metode klasifikasi Naïve Bayes adalah metode klasifikasi statistik yang dapat digunakan untuk memprediksi kelas target berdasarkan probabilitas. Metode Naïve Bayes dibangun dari teorema Bayes untuk menyelesaikan permasalahan klasifikasi. Metode Naïve Bayes sudah terbukti mampu diimplementasi pada database besar dengan hasil prediksi kelas berakurasi tinggi dalam kecepatan yang tinggi [11]. Dalam permasalahan klasifikasi, tujuan dari perhitungan untuk menentukan $P(C_i|X)$. Berikut adalah rumus dasar dari teorema Bayes.

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (2)$$

X adalah data kelas target yang akan diprediksi; C_i adalah hipotesis data X merupakan suatu kelas; $P(C_i|X)$ adalah probabilitas hipotesis C_i berdasarkan X ; $P(C_i)$ adalah probabilitas hipotesis C_i ; $P(X|C_i)$ adalah probabilitas X berdasarkan hipotesis C_i ; $P(X)$ adalah probabilitas dari X . Rumus klasifikasi Naïve Bayes berasumsi bahwa keberadaan suatu parameter tidak berhubungan dengan keberadaan parameter yang lain. Secara matematis, metode klasifikasi Naïve Bayes dirumuskan sebagai berikut [12].

$$P(X|C_i) \approx \prod_{k=1}^n P(X_k|C_i) \quad (3)$$

Jika X_k merupakan nilai kontinu, maka diasumsikan nilai-nilai tersebut berbentuk distribusi Gauss dengan μ sebagai mean dan σ sebagai standar deviasi.

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

$$p(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) \quad (5)$$

Dalam memprediksi kelas target X , $P(X/C_i)P(C_i)$ dievaluasi untuk setiap kelas C_i . Metode klasifikasi Naïve Bayes memprediksi bahwa target kelas X adalah C_i hanya jika kelas tersebut memaksimalkan $P(X/C_i)P(C_i)$.

2.5 Confusion Matrix

Evaluasi kinerja dari model Naïve Bayes yang sudah dilatih dilakukan dengan menggunakan *confusion matrix*. *Confusion matrix* mengandung informasi tentang data asli dan hasil prediksi dari model Naïve Bayes. Tabel 3 menunjukkan confusion matriks untuk sistem klasifikasi dua kelas [13].

Tabel 2 *Confusion matrix* sistem klasifikasi dua kelas

		<i>Predicted</i>	
		<i>Negative (N)</i>	<i>Positive (P)</i>
<i>Actual</i>	<i>Negative (N)</i>	<i>True Negative (TN)</i>	<i>False Positive (FP)</i>
	<i>Positive (P)</i>	<i>False Negative (FN)</i>	<i>True Positive (TP)</i>

Dimana TN adalah jumlah prediksi yang benar bahwa suatu data berkelas negatif; FP adalah jumlah prediksi yang salah bahwa suatu data berkelas positif; FN adalah jumlah prediksi yang salah bahwa suatu data berkelas negatif; TP adalah jumlah prediksi yang benar bahwa suatu data berkelas positif.

Perhitungan akurasi mengukur sejauh mana model Naïve Bayes mampu memprediksi secara benar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Perhitungan presisi mengukur jumlah prediksi positif yang benar dari model Naïve Bayes.

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Perhitungan *recall* mengukur sejauh mana model Naïve Bayes mampu memprediksi dengan benar kasus positif dari seluruh data positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

Skor-F1 merupakan penggabungan dari perhitungan presisi dan *recall* menjadi satu nilai tunggal untuk mengukur kinerja model Naïve Bayes secara keseluruhan.

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (9)$$

Confusion matrix akan memberikan pemahaman tentang kinerja model Naïve Bayes dalam memprediksi curah hujan di Kota Padang.

3. HASIL DAN PEMBAHASAN

Hasil penelitian yang telah dijalankan sesuai metode penelitian yang telah ditentukan akan dibahas dalam bab ini, dimulai dari analisis korelasi koefisien Pearson, model Naïve Bayes, hingga evaluasi kinerja model Naïve Bayes menggunakan *confusion matrix*.

3.1 Korelasi Koefisien Pearson

Perhitungan korelasi dengan metode korelasi koefisien Pearson dilakukan untuk mengetahui hubungan korelasi antara parameter temperatur rata-rata, kelembapan rata-rata, serta curah hujan di Kota Padang. Hasil perhitungan korelasi ditampilkan dalam Tabel 3.

Tabel 3 Korelasi antara temperatur rata-rata, kelembapan rata-rata, dan curah hujan

	Temperatur Rata-Rata	Kelembapan Rata-Rata	Curah Hujan
Temperatur Rata-Rata	1,000000	-0,630816	-0,323673
Kelembapan Rata-Rata	-0,630816	1,000000	0,240923
Curah Hujan	-0,323673	0,240923	1,000000

Berdasarkan hasil perhitungan yang telah dilakukan dapat disimpulkan bahwa

- parameter temperatur rata-rata dan kelembapan rata-rata mempunyai hubungan korelasi bernilai -0,630816, yang berarti terdapat hubungan korelasi negatif yang sedang;
- parameter temperatur rata-rata dan curah hujan mempunyai hubungan korelasi bernilai -0,323673, yang berarti terdapat hubungan korelasi negatif yang rendah;
- parameter kelembapan rata-rata dan curah hujan mempunyai hubungan korelasi bernilai 0,240923, yang berarti terdapat hubungan korelasi positif yang kurang berarti.

3.2 Model Naïve Bayes

Model Naïve Bayes dilatih dalam memprediksi kriteria curah hujan di Kota Padang menggunakan data latih. Setelah model Naïve Bayes selesai dilatih, kemudian model digunakan untuk memprediksi kriteria curah hujan dengan data uji. Data latih merupakan 75% dari data dalam dataset, sedangkan 25% data lainnya dijadikan data uji. Dalam Tabel 4 terdapat lima baris dari nomor indeks data, parameter temperatur rata-rata, kelembapan rata-rata, dan kriteria curah hujan dalam data latih yang akan digunakan untuk melatih model Naïve Bayes.

Tabel 4 Data latih

Indeks	Temperatur Rata-Rata	Kelembapan Rata-Rata	Kriteria Curah Hujan
1892	28,6	75	Berawan
4140	28,3	84	Berawan
4551	26,5	87	Hujan sedang
3370	27,6	77	Hujan sedang
1760	26,3	86	Berawan
...

Dalam tabel 5 berikut terdapat lima baris dari nomor indeks data, parameter temperatur rata-rata, kelembapan rata-rata, dan kriteria curah hujan dalam data uji yang akan digunakan untuk menguji kinerja model Naïve Bayes yang telah dilatih sebelumnya.

Tabel 5 Data uji

Indeks	Temperatur Rata-Rata	Kelembapan Rata-Rata	Kriteria Curah Hujan
2339	27,5	82	Hujan ekstrem
2797	26,3	76	Hujan ringan
3817	27,7	78	Berawan
4429	26,1	93	Hujan ringan
4794	28,1	73	Berawan
...

Dalam Tabel 6 ditampilkan perbandingan jumlah data dari setiap kriteria curah hujan dalam dataset, data latih, dan data uji. Kriteria curah hujan dalam dataset, data latih, dan data uji diurutkan dari jumlah terbesar hingga terkecil adalah kriteria berawan, hujan ringan, hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem.

Tabel 6 Jumlah data berdasarkan kriteria curah hujan dalam dataset, data latih, dan data uji

Kriteria Curah Hujan	Dataset	Data Latih	Data Uji
Berawan	1948	1444	504
Hujan ringan	1856	1382	474
Hujan sedang	640	197	143
Hujan lebat	287	222	65
Hujan sangat lebat	71	56	15
Hujan ekstrem	36	27	9
Total	4838	3628	1210

Setelah model Naïve Bayes selesai dilatih menggunakan data latih, model tersebut memprediksi kriteria curah hujan menggunakan data uji yang telah disiapkan. Hasil prediksi dari model tersebut akan dibandingkan dengan data uji. Dalam Tabel 7 berikut terdapat lima baris dari hasil gabungan data uji dan data hasil prediksi dari model Naïve Bayes.

Tabel 7 Hasil gabungan data uji dan data hasil prediksi

Indeks	Kriteria Curah Hujan Asli	Kriteria Curah Hujan Prediksi
2339	Hujan ekstrem	Hujan ringan
2797	Hujan ringan	Hujan ringan
3817	Berawan	Berawan
4429	Hujan ringan	Hujan ringan
4794	Berawan	Berawan
...

Jika kriteria data uji sama dengan hasil prediksi, artinya data berhasil diprediksi dengan benar. Sebaliknya jika kriteria data uji berbeda dengan hasil prediksi, artinya model Naïve Bayes gagal membuat prediksi yang akurat.

Tabel 8 Perbandingan jumlah data uji asli dengan data hasil prediksi

Kriteria Curah Hujan	Data Uji	Hasil Prediksi
Berawan	504	581
Hujan ringan	474	589
Hujan sedang	143	37

Hujan lebat	65	0
Hujan sangat lebat	15	0
Hujan ekstrem	9	3

Tabel 8 menampilkan perbandingan antara jumlah data untuk setiap kriteria curah hujan dalam data uji dengan jumlah data untuk setiap kriteria curah hujan hasil prediksi. Hasil prediksi kemudian akan dievaluasi menggunakan *confusion matrix* untuk menilai kinerja dari model Naïve Bayes dalam memprediksi curah hujan di Kota Padang.

3.3 Evaluasi Confusion Matrix

Berdasarkan hasil prediksi 1210 data uji curah hujan di Kota Padang menggunakan model Naïve Bayes, didapatkan 581 hasil prediksi dalam kriteria berawan, 589 hasil prediksi dalam kriteria hujan ringan, 37 hasil prediksi dalam kriteria hujan sedang, 0 prediksi dalam kriteria hujan lebat, 0 hasil prediksi dalam kriteria hujan sangat lebat, dan 3 hasil prediksi dalam kriteria hujan ekstrem. Hasil prediksi tersebut kemudian digunakan untuk mengevaluasi kinerja model Naïve Bayes. Seluruh hasil prediksi beserta data uji asli dimuatkan dalam *confusion matrix* Gambar 2.

		Confusion Matrix					
Actual	Berawan	365	135	4	0	0	0
	Hujan ringan	157	303	14	0	0	0
	Hujan sedang	42	92	8	0	0	1
	Hujan lebat	12	44	7	0	0	2
	Hujan sangat lebat	2	10	3	0	0	0
	Hujan ekstrem	3	5	1	0	0	0
		Berawan	Hujan ringan	Hujan sedang	Hujan lebat	Hujan sangat lebat	Hujan ekstrem
		Predicted					

Gambar 2 *Confusion Matrix* untuk mengevaluasi kinerja model Naive Bayes

Dari *confusion matrix* dalam Gambar 2, maka kinerja dari model Naïve Bayes dapat diukur dengan menghitung nilai akurasi, presisi, recall, dan skor-F1. Hasil perhitungan yang sudah di konversikan ke dalam bentuk persentase terdapat dalam Tabel 9 berikut.

Tabel 9 Hasil perhitungan akurasi, presisi, *recall*, dan skor-F1 pada setiap kriteria curah hujan

Kriteria Curah Hujan	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Berawan	72,42%	62,82%	72,42%	67,28%
Hujan ringan	63,92%	51,44%	63,92%	57,01%
Hujan sedang	5,59%	21,62%	5,59%	8,89%
Hujan lebat	0,00%	0,00%	0,00%	0,00%
Hujan sangat lebat	0,00%	0,00%	0,00%	0,00%
Hujan ekstrem	0,00%	0,00%	0,00%	0,00%

Secara keseluruhan, model Naïve Bayes diukur memiliki nilai akurasi 55,87%, nilai presisi 48,87%, nilai recall 55,87%, skor-F1 51,41%. Berdasarkan perhitungan nilai akurasi yang dilakukan pada setiap kriteria curah hujan, kinerja model Naïve Bayes dalam memprediksi kriteria curah hujan dapat dikatakan cukup terbatas.

Hasil prediksi kriteria curah hujan berawan sudah bernilai akurasi 72,42% dan kriteria curah hujan ringan bernilai akurasi 63,92%, artinya model Naïve Bayes mampu memprediksi kriteria-kriteria tersebut dengan cukup baik. Namun, nilai akurasi dari hasil prediksi kriteria curah hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem sangat rendah. Hal ini diakibatkan dari kurangnya data dalam kriteria curah hujan tersebut untuk melatih model Naïve Bayes dalam memprediksi curah hujan di Kota Padang.

Perbaikan lebih lanjut diperlukan untuk meningkatkan kinerja model Naïve Bayes dalam memprediksi curah hujan secara akurat untuk semua kriteria curah hujan. Dengan demikian, perlu ditambahkan data dari sumber lain untuk melatih model Naïve Bayes, terutama untuk data kriteria curah hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem.

4. KESIMPULAN

Setelah penelitian dilakukan, dapat disimpulkan bahwa metode Naïve Bayes kurang efektif dalam memprediksi curah hujan di Kota Padang pada rentang waktu periode 2010-2023. Hal ini dikarenakan nilai akurasi dari hasil prediksi curah hujan di Kota Padang secara keseluruhan hanya 55,87%. Model Naïve Bayes sudah dapat memprediksi kriteria curah hujan berawan dan hujan ringan dengan cukup baik dengan nilai akurasi berturut-turut 72,42% dan 63,92%. Namun, hasil prediksi kriteria curah hujan sedang, hujan lebat, hujan sangat lebat, dan hujan ekstrem sangat rendah dengan nilai akurasi berturut-turut 5,59%, 0%, 0%, dan 0%.

Meskipun terdapat beberapa kriteria curah hujan yang memiliki akurasi yang rendah, model ini tetap memberikan prediksi yang dapat digunakan. Untuk meningkatkan akurasi prediksi curah hujan, perlu dilakukan pengumpulan lebih banyak data untuk kriteria curah hujan yang kurang diwakilkan. Selain itu, penelitian lebih lanjut untuk mengeksplorasi dan menambahkan faktor-faktor lainnya ke dalam model Naïve Bayes dapat meningkatkan kinerjanya. Dengan mengatasi batasan ini, diharapkan dapat meningkatkan akurasi prediksi curah hujan menggunakan metode Naïve Bayes di Kota Padang, yang akan mendukung pengambilan keputusan yang lebih baik dan persiapan menghadapi cuaca di wilayah tersebut.

UCAPAN TERIMA KASIH

Ucapan terima kasih diberikan kepada dosen mata kuliah Big Data serta asisten dosen mata kuliah Big Data, karena penelitian ini tidak dapat diselesaikan tanpa bantuan dari mereka.

DAFTAR PUSTAKA

- [1] Diskominfo, "Sejarah Kota Padang," 1 Januari 2016. [Online]. Available: <https://padang.go.id/sejarah-kota-padang>.
- [2] K. Kurniawan, "Pengertian Cuaca Menurut BMKG," 7 Juni 2020. [Online]. Available: <https://maritim.kalbar.bmkg.go.id/konten/pengertian-cuaca/>. [Accessed 1 Juli 2023].
- [3] JavaTpoint, "Naïve Bayes Classifier Algorithm," JavaTpoint, [Online]. Available: <https://www.javatpoint.com/machine-learning-naive-bayes-classifier>. [Accessed 1 Juli 2023].
- [4] N. H. ., Z. A. K. E. W. Muhammad Wildan Al Azkia, "ANALISIS TEMPERATURE DAN KELEMBABAN TERHADAP CURAH HUJAN DI KABUPATEN SLEMAN PROVINSI DAERAH ISTIMEWA YOGYAKARTA," in *PROSIDING-M21*.
- [5] M. I. J. Marnia, "Analisis Hubungan Kelembaban Udara dan Suhu Udara Terhadap," *PRISMA FISIKA*, vol. IV, no. 3, pp. 80-83, 2016.
- [6] D. K. Azriel Alfian Rizqi, "Klasifikasi Curah Hujan di Kota Bogor Provinsi Jawa Barat dengan Menggunakan Metode Naïve Bayes," in *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi*, Jakarta, 2022.
- [7] A. A. K. D. I Gede Aris Gunadi, "Klasifikasi Curah Hujan di Provinsi Bali Berdasarkan Metode Naïve Bayesian," *Wahana Matematika dan Sains: Jurnal Matematika, Sains, dan Pembelajarannya*, vol. 12, no. 1, 2018.
- [8] "DATA ONLINE - PUSAT DATABASE - BMKG," [Online]. Available: <http://dataonline.bmkg.go.id/home>. [Accessed 1 April 2023].
- [9] BMKG, "Probabilitas Curah Hujan 20 mm (tiap 24 jam)," BMKG, [Online]. Available: [https://www.bmkg.go.id/cuaca/probabilitas-curah-hujan.bmkg#:~:text=0.5%20%20E2%80%93%2020%20mm%20Fhari%20\(,merah\)%20%3A%20Hujan%20sangat%20lebat](https://www.bmkg.go.id/cuaca/probabilitas-curah-hujan.bmkg#:~:text=0.5%20%20E2%80%93%2020%20mm%20Fhari%20(,merah)%20%3A%20Hujan%20sangat%20lebat). [Accessed 3 Juli 2023].
- [10] Y. G. H. Y. Wang Yongjun, "Research on Correlation analysis of industry electricity quantity," in *International Conference on Information Technology and Management Innovation*, 2015.
- [11] R. T. V. W. L. Y. S. Hakam Febtadianrano Putro, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *Jurnal TIKomSiN*, vol. 8, no. 2, pp. 19-24, 2020.
- [12] K. M. Leung, Naive Bayesian Classifier, 2007.
- [13] C. J. C. A. K. Santra, "Genetic Algorithm and Confusion Matrix for Document Clustering," *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 1, pp. 322-328, 2012.