



PERBANDINGAN PERFORMA NAÏVE BAYES DAN MODIFIKASI NAIVE BAYES DENGAN LAPLACE CORRECTION PADA KLASIFIKASI POTENSI BANJIR MENGGUNAKAN DATA CURAH HUJAN

Wan Dinul Aqli¹, Nurdin²

^{1,2}Magister Teknologi Informasi, Universitas Malikussaleh

Email: wan.257110201012@mhs.unimal.ac.id¹, nurdin@unmal.ac.id²

Informasi Artikel	ABSTRACT
<p>Riwayat artikel : Disubmit : 28 Mei 2026 Direvisi : 10 Juni 2026 Diterima : 15 Juni 2026 Dipublikasi : 30 Juni 2026</p>	<p><i>Flood is one of the most frequent natural disasters in Kerala, India, causing significant social, environmental, and economic impacts. Machine learning methods have been widely applied for flood prediction and classification; however, studies comparing standard Naïve Bayes and Modified Naïve Bayes with Laplace Correction on Kerala rainfall data remain limited. Therefore, this study aims to compare the performance of Naïve Bayes and Modified Naïve Bayes with Laplace Correction in classifying flood potential based on rainfall data. The dataset used is the Kerala Flood Prediction Dataset consisting of 118 historical rainfall records from 1901–2018. The data were preprocessed through label transformation and quartile-based discretization, then divided into 82 training data (70%) and 36 testing data (30%). Model performance was evaluated using accuracy, precision, recall, and F1-score. The results show that both Naïve Bayes and Modified Naïve Bayes achieved the same performance, with an accuracy of 88.89%, precision of 85.00%, recall of 94.44%, and F1-score of 89.47%. Although Laplace Correction successfully addressed the zero-probability problem found in the Annual Rainfall attribute, it did not improve classification performance. These findings indicate that Laplace Correction improves probability stability but does not necessarily increase classification performance on the Kerala rainfall dataset.</i></p>
<p>Keywords: Naïve Bayes, Laplace Correction, Flood Classification, Rainfall, Machine Learning</p>	
	ABSTRAK
<p>Kata Kunci: Naïve Bayes, Laplace Correction, Klasifikasi Banjir, Curah Hujan, Machine Learning</p>	<p>Banjir merupakan salah satu bencana alam yang sering terjadi di Kerala, India, dan menimbulkan dampak sosial, lingkungan, serta ekonomi yang signifikan. Berbagai metode machine learning telah digunakan untuk klasifikasi banjir, namun penelitian yang membandingkan Naïve Bayes standar dan Modifikasi Naïve Bayes dengan <i>Laplace Correction</i> pada data curah hujan Kerala masih terbatas. Oleh karena itu, penelitian ini bertujuan membandingkan performa kedua metode dalam mengklasifikasikan potensi banjir berdasarkan data curah hujan. Dataset yang digunakan adalah <i>Kerala Flood Prediction Dataset</i> yang terdiri atas 118 data historis periode 1901–2018. Data diproses melalui transformasi label dan diskretisasi berbasis kuartil, kemudian dibagi menjadi 82 data pelatihan (70%) dan 36 data pengujian (30%). Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction menghasilkan performa yang sama, yaitu accuracy 88,89%, precision 85,00%, recall 94,44%, dan F1-score 89,47%. Meskipun berhasil mengatasi permasalahan zero probability pada atribut Annual Rainfall, Laplace Correction tidak meningkatkan performa klasifikasi. Temuan ini menunjukkan bahwa Laplace Correction mampu meningkatkan stabilitas perhitungan probabilitas tanpa memberikan peningkatan performa klasifikasi pada dataset curah hujan Kerala.</p>





PENDAHULUAN

Banjir merupakan salah satu bencana alam yang sering terjadi dan menimbulkan dampak yang signifikan terhadap kehidupan masyarakat. Tingginya intensitas curah hujan, perubahan cuaca, serta berkurangnya daerah resapan air menjadi faktor utama yang meningkatkan risiko banjir. Selain menyebabkan kerusakan infrastruktur dan kerugian ekonomi, banjir juga berdampak pada aktivitas sosial, kesehatan, dan lingkungan sehingga diperlukan upaya mitigasi yang mampu mengidentifikasi potensi banjir secara cepat dan akurat (Intan et al., 2021; Maulita & Nurdin, 2023; Meladiar & Indra, 2024).

Perkembangan teknologi machine learning memberikan peluang dalam pengembangan sistem prediksi dan klasifikasi banjir berbasis data historis. Berbagai algoritma telah digunakan untuk mengidentifikasi pola curah hujan guna mendukung mitigasi bencana. Penelitian sebelumnya menunjukkan bahwa algoritma seperti Random Forest dan Naïve Bayes mampu memberikan performa yang kompetitif dalam klasifikasi kejadian banjir serta meningkatkan efektivitas sistem peringatan dini dibandingkan pendekatan konvensional (Indaryono et al., 2024; Nurdin et al., 2021; Perdana Putra et al., 2025; Sinatrya et al., 2025).

Berbagai penelitian telah menerapkan machine learning untuk prediksi banjir. Natzir (2023) melaporkan bahwa Naïve Bayes mampu menghasilkan akurasi yang kompetitif meskipun masih berada di bawah Random Forest. Sementara itu, Pitaloka et al. (2024) mengembangkan model berbasis CNN dan BiLSTM yang menunjukkan performa prediksi yang lebih baik dibandingkan beberapa metode sebelumnya. Pada dataset Kerala, Triyanto et al. (2021) menunjukkan bahwa Naïve Bayes mampu mengklasifikasikan potensi banjir dengan tingkat akurasi yang cukup baik, sedangkan Refida Septiana Putri et al. (2025) meningkatkan sensitivitas deteksi banjir melalui kombinasi Naïve Bayes dan SMOTE (Natzir, 2023; Pitaloka et al., 2024; Refida Septiana Putri et al., 2025; Triyanto et al., 2021).

Meskipun memiliki proses komputasi yang sederhana dan efisien, Naïve Bayes memiliki kelemahan berupa asumsi independensi antaratribut dan rentan terhadap permasalahan zero probability. Kondisi ini terjadi ketika suatu nilai atribut tidak muncul pada kelas tertentu sehingga probabilitas posterior menjadi nol dan berpotensi menurunkan kualitas klasifikasi. Untuk mengatasi permasalahan tersebut, beberapa penelitian mengembangkan modifikasi Naïve Bayes, salah satunya melalui penerapan Laplace Correction yang terbukti mampu menghilangkan probabilitas nol dan meningkatkan stabilitas perhitungan probabilitas (Chen et al., 2021; Kurniawan et al., 2021).

Selain Laplace Correction, Kim dan Lee (2022) mengembangkan Weighted Naïve Bayes melalui pemberian bobot pada atribut. Namun pendekatan tersebut memerlukan mekanisme optimasi tambahan yang relatif lebih kompleks. Oleh karena itu, Laplace Correction menjadi alternatif yang lebih





sederhana karena tidak memerlukan proses pembobotan atribut dan tetap mampu mengatasi permasalahan zero probability (Kim & Lee, 2022).

Meskipun Naïve Bayes telah banyak diterapkan pada klasifikasi banjir berbasis data curah hujan, penelitian yang secara khusus membandingkan performa Naïve Bayes standar dan Modifikasi Naïve Bayes dengan Laplace Correction pada dataset Kerala masih terbatas. Oleh karena itu, penelitian ini bertujuan membandingkan performa kedua metode dalam mengklasifikasikan potensi banjir menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengembangan model klasifikasi banjir yang lebih robust terhadap permasalahan probabilitas nol.

METODE PENELITIAN

3.1 Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk membandingkan performa Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction pada klasifikasi potensi banjir menggunakan data curah hujan. Dataset yang digunakan adalah Kerala Flood Prediction Dataset yang terdiri atas 118 data historis periode 1901–2018. Tahapan penelitian meliputi preprocessing data, pembagian data training dan testing, pembangunan model Naïve Bayes dan Modifikasi Naïve Bayes, serta evaluasi menggunakan accuracy, precision, recall, dan F1-score.

	SUBDIVISION	YEAR	JAN	FEB	MAR	APR	...	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOOD
DS													
0	KERALA	1901	28.7	44.7	51.6	160.0	...	197.7	266.9	350.8	48.4	3248.6	Y
ES													
1	KERALA	1902	6.7	2.6	57.3	83.9	...	491.6	358.4	158.3	121.5	3326.6	Y
ES													
2	KERALA	1903	3.2	18.6	3.1	83.6	...	341.8	354.1	157.0	59.0	3271.2	Y
ES													
3	KERALA	1904	23.7	3.0	32.2	71.5	...	222.7	328.1	33.9	3.3	3129.7	Y
ES													
4	KERALA	1905	1.2	22.3	9.4	105.9	...	217.2	383.5	74.4	0.2	2741.6	Y
NO													
..

Gambar 1. Dataset Penelitian

	JAN	FEB	MAR	...	DEC	ANNUAL RAINFALL	FLOODS
count	118.000000	118.000000	118.000000	...	118.000000	118.000000	118.000000
mean	12.218644	15.633898	36.670339	...	40.009322	2925.405085	0.508475
std	15.473766	16.406290	30.063862	...	36.676330	452.169407	0.502060
min	0.000000	0.000000	0.100000	...	0.100000	2068.800000	0.000000
25%	2.175000	4.700000	18.100000	...	10.350000	2613.525000	0.000000
50%	5.800000	8.350000	28.400000	...	31.100000	2934.300000	1.000000
75%	18.175000	21.400000	49.825000	...	54.025000	3170.400000	1.000000
max	83.500000	79.000000	217.200000	...	202.300000	4473.000000	1.000000

Gambar 2. Deskripsi Statistik Dataset Kerala





Diskretisasi dilakukan menggunakan metode kuartil berdasarkan distribusi masing-masing atribut. Nilai di bawah Q1 dikategorikan sebagai Rendah, nilai antara Q1 dan Q3 sebagai Sedang, dan nilai di atas Q3 sebagai Tinggi. Pendekatan ini dipilih karena setiap atribut memiliki rentang nilai yang berbeda sehingga kategorisasi menjadi lebih representatif terhadap distribusi data.

```

df['FLOODS'].replace({'YES':1, 'NO':0}, inplace=True)

```

	JAN	FEB	MAR	APR	MAY	...	OCT	NOV	DEC	ANNUAL RAINFALL	FLOODS
0	Tinggi	Tinggi	Tinggi	Tinggi	Sedang	...	Sedang	Tinggi	Sedang	Tinggi	1
1	Sedang	Rendah	Tinggi	Sedang	Sedang	...	Tinggi	Sedang	Tinggi	Tinggi	1
2	Sedang	Sedang	Rendah	Sedang	Sedang	...	Sedang	Sedang	Tinggi	Tinggi	1
3	Tinggi	Rendah	Sedang	Rendah	Sedang	...	Sedang	Rendah	Rendah	Sedang	1
4	Rendah	Tinggi	Rendah	Sedang	Sedang	...	Tinggi	Rendah	Rendah	Sedang	0
...
113	Sedang	Sedang	Rendah	Sedang	Sedang	...	Tinggi	Sedang	Sedang	Sedang	1
114	Sedang	Sedang	Tinggi	Tinggi	Sedang	...	Sedang	Tinggi	Tinggi	Rendah	0
115	Sedang	Rendah	Sedang	Tinggi	Sedang	...	Sedang	Sedang	Sedang	Rendah	0
116	Rendah	Sedang	Rendah	Rendah	Sedang	...	Rendah	Rendah	Sedang	Rendah	0
117	Tinggi	Tinggi	Sedang	Sedang	Sedang	...	Tinggi	Sedang	Tinggi	Tinggi	1

[118 rows x 14 columns]

Gambar 3. Dataset Kerala Setelah Diskretisasi Data

3.2 Objek Penelitian

Objek penelitian adalah Kerala Flood Prediction Dataset yang diperoleh dari Kaggle dan terdiri atas 118 data historis curah hujan wilayah Kerala, India periode 1901–2018. Atribut yang digunakan meliputi curah hujan bulanan (JAN–DEC) dan Annual Rainfall sebagai fitur, sedangkan Floods digunakan sebagai variabel target dengan kelas YES dan NO.

Tabel 3.1 Definisi Operasional Variabel

Variabel	Jenis Variabel	Keterangan
Jan – Dec	Independen (X)	Data curah hujan bulanan yang digunakan sebagai fitur klasifikasi
Annual Rainfall	Independen (X)	Total curah hujan tahunan yang digunakan sebagai fitur klasifikasi
Flood	Dependen (Y)	Status kejadian banjir dengan kelas YES dan NO

3.3 Bahan dan Alat Penelitian

Bahan penelitian berupa Kerala Flood Prediction Dataset dan literatur pendukung terkait machine learning, Naïve Bayes, serta Laplace Correction. Alat yang digunakan meliputi komputer berbasis Windows, bahasa pemrograman Python, Visual Studio Code, Jupyter Notebook, serta pustaka Pandas, NumPy, Scikit-learn, dan Matplotlib.

3.4 Teknik Pengumpulan Data

Teknik pengumpulan data dilakukan melalui studi dokumentasi dan studi pustaka. Studi dokumentasi dilakukan dengan mengumpulkan Kerala Flood Prediction Dataset dari Kaggle, sedangkan studi pustaka dilakukan melalui pengumpulan jurnal, prosiding, buku, dan artikel ilmiah yang relevan dengan penelitian.





3.5 Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan melalui beberapa tahapan, yaitu preprocessing data, pembagian data, pembangunan model klasifikasi, serta evaluasi performa model. Tahapan analisis dilakukan untuk membandingkan performa algoritma Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction dalam mengklasifikasikan potensi banjir berdasarkan data curah hujan.

Dataset yang telah diproses dibagi menjadi data training (70%) dan testing (30%). Model Naïve Bayes dibangun menggunakan Teorema Bayes, sedangkan Modifikasi Naïve Bayes menerapkan Laplace Correction untuk mengatasi zero probability. Evaluasi dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score.

3.5.1 Implementasi Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen terhadap atribut lainnya. Probabilitas posterior dihitung menggunakan persamaan berikut:

$$P(C | X) = \frac{P(X | C)P(C)}{P(X)}$$

dengan $P(C|X)$ menyatakan probabilitas hipotesis terhadap data, $P(X|C)$ menyatakan probabilitas data terhadap hipotesis, $P(C)$ menyatakan probabilitas awal (prior probability), dan $P(X)$ menyatakan probabilitas data.

3.5.2 Modifikasi Naïve Bayes dengan Laplace Correction

Laplace Correction merupakan teknik smoothing yang digunakan untuk mengatasi permasalahan zero probability pada algoritma Naïve Bayes dengan menambahkan konstanta pada frekuensi kemunculan atribut (Chen et al., 2021).

Masalah zero probability:

$$P(x_i | C) = 0$$

Salah satu kelemahan Naïve Bayes adalah terjadinya zero probability, yaitu kondisi ketika suatu nilai atribut tidak muncul pada kelas tertentu sehingga probabilitas bersyarat bernilai nol. Kondisi tersebut dapat menyebabkan probabilitas posterior menjadi nol dan memengaruhi hasil klasifikasi. Untuk mengatasi permasalahan tersebut digunakan Laplace Correction (*Laplace Smoothing*), yaitu dengan menambahkan nilai 1 pada setiap frekuensi kemunculan atribut sehingga probabilitas nol dapat dihindari. Probabilitas bersyarat dihitung menggunakan persamaan berikut:

Modifikasi naïve bayes dengan laplace correction:

$$P(x_i | C) = \frac{N_{ic} + 1}{N_c + k}$$





dengan N_c menyatakan jumlah kemunculan atribut pada kelas tertentu, N menyatakan jumlah data pada kelas tersebut, dan k menyatakan jumlah kategori atribut. Pada penelitian ini nilai $k = 3$ karena setiap atribut curah hujan dikategorikan menjadi Rendah, Sedang, dan Tinggi.

3.5.3 Evaluasi Model

Evaluasi performa model dilakukan menggunakan confusion matrix. Berdasarkan hasil confusion matrix, dihitung nilai accuracy, precision, recall, dan F1-score untuk mengukur kinerja masing-masing model klasifikasi.

Nilai accuracy dihitung menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Nilai precision dihitung menggunakan persamaan:

$$Precision = \frac{TP}{TP + FP}$$

Nilai recall dihitung menggunakan persamaan:

$$Recall = \frac{TP}{TP + FN}$$

Nilai f1-score dihitung menggunakan persamaan:

$$F1 - Score = \frac{2 \times (Precision \times recall)}{Precision + Recall}$$

dengan TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative). Selain accuracy, penelitian ini juga menggunakan precision, recall, dan F1-score untuk memberikan evaluasi yang lebih komprehensif terhadap performa model.

3.5.4 Perbandingan Performa Model

Hasil evaluasi Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction dibandingkan berdasarkan nilai accuracy, precision, recall, dan F1-score. Model yang memiliki nilai evaluasi lebih tinggi dianggap memiliki performa yang lebih baik dalam mengklasifikasikan potensi banjir berdasarkan data curah hujan.

HASIL DAN PEMBAHASAN

4.1 Hasil Preprocessing Data

Tahap preprocessing dilakukan untuk mempersiapkan dataset sebelum proses klasifikasi menggunakan algoritma Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction. Tahapan preprocessing meliputi pengecekan missing value, transformasi label kelas, serta diskretisasi atribut





curah hujan menggunakan metode kuartil. Hasil preprocessing bertujuan untuk memastikan kualitas data dan menghasilkan representasi atribut yang sesuai dengan kebutuhan algoritma klasifikasi.

4.1.1 Hasil Kondisional Probability

Setelah proses diskretisasi, dilakukan perhitungan frekuensi kemunculan setiap kategori atribut pada masing-masing kelas (Floods = 0 dan Floods = 1) sebagai dasar pembentukan conditional probability pada algoritma Naïve Bayes. Frekuensi tersebut menunjukkan jumlah kemunculan kategori Rendah, Sedang, dan Tinggi pada setiap atribut curah hujan terhadap kelas banjir dan tidak banjir. Hasil perhitungan ditunjukkan pada Tabel 4.1.

Tabel 4.1 Frekuensi Kategori Atribut Berdasarkan Kelas Banjir

Atribut	Floods = 0			Floods = 1		
	Rendah	Sedang	Tinggi	Rendah	Sedang	Tinggi
JAN	12	21	7	10	15	17
FEB	11	19	10	12	23	7
MAR	10	22	8	8	24	10
APR	11	18	11	10	18	14
MAY	13	22	5	8	19	15
JUN	18	17	5	5	23	14
JUL	14	20	6	5	23	14
AUG	16	19	5	5	23	14
SEP	18	19	3	6	21	15
OCT	10	18	12	9	21	12
NOV	13	17	10	9	17	16
DEC	13	20	7	10	20	12
Annual Rainfall	21	19	0	0	21	21

Berdasarkan Tabel 4.1, setiap atribut curah hujan memiliki distribusi kategori yang berbeda pada masing-masing kelas. Beberapa atribut menunjukkan perbedaan distribusi yang cukup jelas antara kelas banjir dan tidak banjir, seperti atribut Annual Rainfall yang memiliki kategori Tinggi hanya muncul pada kelas banjir (Floods = 1) dan kategori Rendah hanya muncul pada kelas tidak banjir (Floods = 0). Kondisi ini menunjukkan bahwa atribut tersebut memiliki keterkaitan yang kuat terhadap kejadian banjir. Frekuensi kategori pada Tabel 4.1 selanjutnya digunakan untuk menghitung probabilitas bersyarat (conditional probability) pada proses klasifikasi menggunakan Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction.

4.1.2 Hasil Pegecekan Missing Value

Pengecekan missing value dilakukan untuk memastikan tidak terdapat data yang kosong atau hilang pada dataset yang dapat memengaruhi proses klasifikasi. Hasil pengecekan menunjukkan bahwa seluruh atribut memiliki jumlah missing value sebesar nol sehingga tidak diperlukan proses penanganan data hilang (missing value handling). Hasil pengecekan missing value ditunjukkan pada Gambar 4.





	JAN	FEB	MAR	APR	MAY	JUN	...	SEP	OCT	NOV	DEC	ANNUAL RAINFALL	FLOOD
5	False	False	False	False	False	False	...	False	False	False	False	False	False
0	False	False	False	False	False	False	...	False	False	False	False	False	False
1	False	False	False	False	False	False	...	False	False	False	False	False	False
2	False	False	False	False	False	False	...	False	False	False	False	False	False
3	False	False	False	False	False	False	...	False	False	False	False	False	False
4	False	False	False	False	False	False	...	False	False	False	False	False	False
...

Gambar 4. Hasil Pengecekan Missing Value

4.1.3 Hasil Transformasi Label Kelas

Tahap transformasi label dilakukan dengan mengubah atribut Floods yang semula berbentuk kategori YES dan NO menjadi nilai numerik 1 dan 0. Transformasi ini dilakukan untuk memudahkan proses pengolahan data dan evaluasi model klasifikasi. Hasil transformasi menunjukkan bahwa dataset terdiri atas dua kelas, yaitu kelas banjir (Floods = 1) dan kelas tidak banjir (Floods = 0). Distribusi jumlah data pada masing-masing kelas setelah transformasi ditunjukkan pada Gambar 5.

```
df['FLOODS'].replace({'YES':1,'NO':0}, inplace=True)
D:\python3.13\kecerdasanKomputasional\tugas2\kerala.py:38: FutureWarning: Downcasting behavior in `replace` is deprecated and will be removed in a future version. To retain the old behavior, explicitly call `result.infer_objects(copy=False)`. To opt-in to the future behavior, set `pd.set_option('future.no_silent_downcasting', True)`
df['FLOODS'].replace({'YES':1,'NO':0}, inplace=True)
FLOODS
1    60
0    58
Name: count, dtype: int64
```

Gambar 5. Hasil Transformasi Label dan Distribusi Kelas Dataset

4.2 Implementasi Model

Setelah melalui tahap preprocessing, dataset yang terdiri dari 118 data dibagi menjadi data pelatihan (training data) sebesar 70% dan data pengujian (testing data) sebesar 30%. Hasil pembagian data menghasilkan 82 data training dan 36 data testing. Distribusi kelas pada data training terdiri atas 42 data banjir (Floods = YES) dan 40 data tidak banjir (Floods = NO), sedangkan pada data testing masing-masing terdiri atas 18 data banjir dan 18 data tidak banjir.

4.2.1 Probabilitas Prior

Probabilitas prior dihitung berdasarkan proporsi masing-masing kelas pada data training. Hasil perhitungan menunjukkan bahwa probabilitas awal kelas banjir sebesar 0,5122 dan probabilitas awal kelas tidak banjir sebesar 0,4878 sebagaimana ditunjukkan pada Tabel 4.2.

Tabel 4.2 Probabilitas Prior Kelas

Kelas	Probabilitas
FLOOD = YES	0,5122
FLOOD = NO	0,4878





Berdasarkan Tabel 4.2 terlihat bahwa distribusi kedua kelas relatif seimbang sehingga tidak terjadi dominasi yang signifikan pada salah satu kelas.

4.2.2 Identifikasi Zero Probability

Pada tahap perhitungan probabilitas bersyarat ditemukan adanya kasus zero probability pada atribut Annual Rainfall. Hasil tabulasi silang antara kategori curah hujan tahunan dan status banjir ditunjukkan pada Tabel 4.3.

Tabel 4.3 Distribusi Kategori Annual Rainfall terhadap Status Banjir

Annual Rainfall	Flood = No	Flood = Yes
Rendah	21	0
Sedang	19	21
Tinggi	0	21

Berdasarkan Tabel 4.3 terdapat kategori yang memiliki frekuensi nol, yaitu kategori Rendah pada kelas Flood = YES dan kategori Tinggi pada kelas Flood = NO. Kondisi tersebut menyebabkan probabilitas bersyarat bernilai nol sehingga berpotensi memengaruhi hasil klasifikasi apabila menggunakan Naïve Bayes standar.

4.2.3 Implementasi Laplace Correction

Untuk mengatasi permasalahan zero probability, penelitian ini menerapkan Laplace Correction dengan menambahkan konstanta sebesar satu pada setiap frekuensi kemunculan atribut. Penerapan Laplace Correction menghasilkan probabilitas bersyarat yang tidak bernilai nol sehingga seluruh atribut tetap dapat berkontribusi dalam proses klasifikasi. Model Naïve Bayes standar dan Modifikasi Naïve Bayes dengan Laplace Correction selanjutnya digunakan untuk melakukan klasifikasi terhadap data testing yang sama sehingga hasil evaluasi yang diperoleh dapat dibandingkan secara objektif.

4.3 Hasil Pengujian Model Naïve Bayes

Setelah proses pelatihan (*training*) selesai dilakukan, model Naïve Bayes diuji menggunakan 36 data pengujian (*testing data*) yang tidak terlibat pada proses pembentukan model. Pengujian dilakukan untuk mengetahui kemampuan model dalam mengklasifikasikan potensi banjir berdasarkan atribut curah hujan yang telah melalui proses diskretisasi. Hasil pengujian model Naïve Bayes ditampilkan dalam bentuk confusion matrix sebagaimana ditunjukkan pada Tabel 4.4.

Tabel 4.4 Confusion Matrix Naïve Bayes

Actual / Prediksi	Flood = No	Flood = Yes
Flood = No	15	3
Flood = Yes	1	17





Berdasarkan Tabel 4.4 diperoleh sebanyak 15 data berhasil diklasifikasikan dengan benar sebagai tidak banjir (True Negative), sedangkan 17 data berhasil diklasifikasikan dengan benar sebagai banjir (True Positive). Selain itu terdapat 3 data yang diprediksi sebagai banjir padahal kondisi sebenarnya tidak banjir (False Positive) dan 1 data yang diprediksi sebagai tidak banjir padahal kondisi sebenarnya banjir (False Negative). Berdasarkan nilai confusion matrix, dilakukan perhitungan metrik evaluasi berupa accuracy, precision, recall, dan F1-score. Perhitungan ini dilakukan untuk memberikan gambaran yang lebih komprehensif mengenai performa model.

4.3.1 Perhitungan Accuracy

Nilai accuracy menunjukkan tingkat ketepatan model dalam melakukan klasifikasi terhadap seluruh data pengujian. Berdasarkan hasil confusion matrix diperoleh nilai: TP = 17, TN = 15, FP = 3, dan FN = 1. Maka nilai accuracy dihitung sebagai berikut:

$$Accuracy = \frac{17 + 15}{17 + 15 + 3 + 1} = \frac{32}{36} = 0,8889 = 88,89\%$$

4.3.2 Perhitungan Precision

Nilai precision menunjukkan kemampuan model dalam memberikan prediksi banjir yang benar dari seluruh data yang diprediksi sebagai banjir.

$$Precision = \frac{17}{17 + 3} = \frac{17}{20} = 0,85 = 85,00\%$$

4.3.3 Perhitungan Recall

Nilai recall menunjukkan kemampuan model dalam mendeteksi seluruh kejadian banjir yang sebenarnya terjadi.

$$Recall = \frac{17}{17 + 1} = \frac{17}{18} = 0,9444 = 94,44\%$$

4.3.4 Perhitungan F1-Score

Nilai *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall*.

$$F1\ Score = \frac{2 \times 0,85 \times 0,9444}{0,85 + 0,9444} = 0,8947 = 89,47\%$$

4.3.5 Analisis Hasil Pengujian Naïve Bayes

Berdasarkan hasil pengujian, model Naïve Bayes memperoleh nilai accuracy sebesar 88,89%, precision sebesar 85,00%, recall sebesar 94,44%, dan F1-score sebesar 89,47%. Nilai recall yang tinggi menunjukkan bahwa model memiliki kemampuan yang baik dalam mendeteksi kejadian banjir sehingga sebagian besar data banjir berhasil diklasifikasikan dengan benar.

4.4 Hasil Pengujian Modifikasi Naïve Bayes Dengan Laplace Correction





Modifikasi Naïve Bayes dengan Laplace Correction diterapkan untuk mengatasi permasalahan zero probability yang ditemukan pada atribut Annual Rainfall. Pengujian dilakukan menggunakan data testing yang sama dengan model Naïve Bayes standar sehingga hasil evaluasi yang diperoleh dapat dibandingkan secara objektif. Hasil pengujian model Modifikasi Naïve Bayes dengan Laplace Correction ditunjukkan pada Tabel 4.5.

Tabel 4.5 Confusion Matrix Modifikasi Naïve Bayes dengan Laplace Correction

Actual / Prediksi	Flood = No	Flood = Yes
Flood = No	15	3
Flood = Yes	1	17

Berdasarkan Tabel 4.5 diperoleh sebanyak 15 data berhasil diklasifikasikan dengan benar sebagai tidak banjir (True Negative) dan 17 data berhasil diklasifikasikan dengan benar sebagai banjir (True Positive). Selain itu terdapat 3 data yang diprediksi sebagai banjir padahal kondisi sebenarnya tidak banjir (False Positive) dan 1 data yang diprediksi sebagai tidak banjir padahal kondisi sebenarnya banjir (False Negative).

4.4.1 Analisis Hasil Pengujian Modifikasi Naïve Bayes

Berdasarkan hasil pengujian, model Modifikasi Naïve Bayes dengan Laplace Correction memperoleh nilai accuracy sebesar 88,89%, precision sebesar 85,00%, recall sebesar 94,44%, dan F1-score sebesar 89,47%. Meskipun pada tahap implementasi ditemukan permasalahan zero probability pada atribut Annual Rainfall, penerapan Laplace Correction tidak menghasilkan perubahan nilai evaluasi pada data pengujian.

4.5 Perbandingan Performa Model

Setelah dilakukan pengujian terhadap Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction, diperoleh hasil evaluasi berdasarkan metrik accuracy, precision, recall, dan F1-score. Perbandingan performa kedua model ditunjukkan pada Tabel 4.6.

Tabel 4.6 Perbandingan Performa Model

Metric Evaluasi	Naïve Bayes	Modifikasi Naïve Bayes
Accuracy	88,89%	88,89%
Precision	85,00%	85,00%
Recall	94,44%	94,44%
F1-Score	89,47%	89,47%

Berdasarkan Tabel 4.6, Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction menghasilkan performa yang sama, yaitu accuracy 88,89%, precision 85,00%, recall 94,44%, dan F1-score 89,47%. Hasil ini menunjukkan bahwa penerapan Laplace Correction tidak meningkatkan





performa klasifikasi pada dataset yang digunakan karena kasus zero probability pada atribut Annual Rainfall tidak memengaruhi hasil prediksi data pengujian.

4.6 Analisis Pengaruh Atribut Curah Hujan terhadap Klasifikasi Banjir

Analisis atribut dilakukan untuk mengidentifikasi keterkaitan curah hujan terhadap kejadian banjir berdasarkan distribusi probabilitas data pelatihan. Hasilnya menunjukkan bahwa Annual Rainfall memiliki keterkaitan paling kuat dengan kelas banjir, dengan probabilitas sebesar 0,5000 (50%), lebih tinggi dibandingkan atribut lainnya sebagaimana ditunjukkan pada Tabel 4.7.

Tabel 4.7 Probabilitas Kemunculan Kategori Tinggi pada Kelas Banjir

No	Atribut	Probabilitas
1	Annual Rainfall	0,5000
2	JAN	0,4048
3	NOV	0,3810
4	MAY	0,3571
5	SEP	0,3571
6	APR	0,3333
7	JUN	0,3333
8	JUL	0,3333
9	AUG	0,3333
10	OCT	0,2857
11	DEC	0,2857
12	MAR	0,2381
13	FEB	0,1667

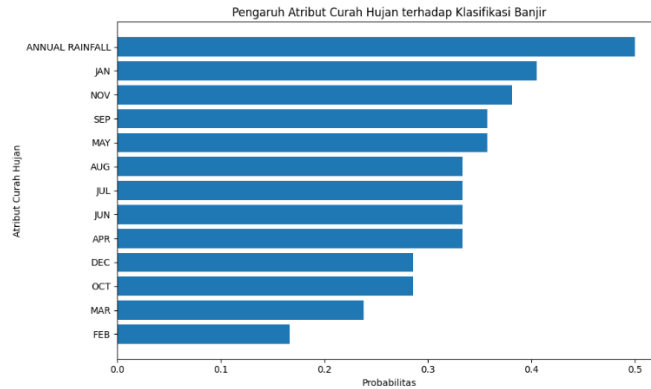
Berdasarkan Tabel 4.7, atribut Annual Rainfall memiliki probabilitas tertinggi sehingga menjadi indikator yang paling dominan dalam menentukan kelas banjir. Temuan ini diperkuat oleh distribusi data pada atribut tersebut sebagaimana ditunjukkan pada Tabel 4.8.

Tabel 4.8 Distribusi Annual Rainfall terhadap Status Banjir

Annual Rainfall	Flood = No	Flood = Yes
Rendah	21	0
Sedang	19	21
Tinggi	0	21

Berdasarkan Tabel 4.8 terlihat bahwa seluruh data dengan kategori Annual Rainfall Tinggi berada pada kelas banjir (Flood = YES), sedangkan seluruh data dengan kategori Annual Rainfall Rendah berada pada kelas tidak banjir (Flood = NO). Pola tersebut menunjukkan bahwa curah hujan tahunan merupakan faktor yang sangat menentukan dalam proses klasifikasi potensi banjir pada dataset Kerala. Selain Annual Rainfall, beberapa atribut curah hujan bulanan seperti JAN, NOV, MAY, dan SEP juga menunjukkan hubungan yang cukup kuat terhadap kejadian banjir.





Gambar 6. Pengaruh Atribut Curah Hujan terhadap Klasifikasi Banjir

4.7 Pembahasan Hasil Penelitian

Penelitian ini bertujuan membandingkan performa Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction dalam klasifikasi potensi banjir berdasarkan data curah hujan. Hasil pengujian menunjukkan bahwa kedua model menghasilkan performa yang sama, yaitu accuracy 88,89%, precision 85,00%, recall 94,44%, dan F1-score 89,47%. Temuan ini menunjukkan bahwa Laplace Correction tidak memberikan peningkatan performa klasifikasi pada dataset Kerala karena kasus zero probability yang ditemukan tidak memengaruhi hasil prediksi data pengujian.

Meskipun demikian, Laplace Correction terbukti efektif mengatasi permasalahan zero probability dan meningkatkan stabilitas perhitungan probabilitas. Hasil ini sejalan dengan penelitian Chen et al. (2021) dan Kurniawan et al. (2021) yang menunjukkan bahwa modifikasi Naïve Bayes dapat meningkatkan keandalan model dalam menghadapi distribusi data tertentu. Namun, berbeda dengan penelitian tersebut, karakteristik dataset Kerala memungkinkan Naïve Bayes standar menghasilkan performa yang sama baiknya dengan model yang dimodifikasi.

Analisis atribut menunjukkan bahwa Annual Rainfall merupakan atribut yang memiliki keterkaitan paling kuat terhadap kejadian banjir dengan probabilitas sebesar 0,5000. Seluruh data dengan kategori Annual Rainfall Tinggi berada pada kelas banjir, sedangkan kategori Annual Rainfall Rendah berada pada kelas tidak banjir. Selain itu, atribut JAN, NOV, MAY, dan SEP juga menunjukkan kontribusi terhadap proses klasifikasi.

SIMPULAN

5.1 Kesimpulan

Penelitian ini berhasil membandingkan performa Naïve Bayes dan Modifikasi Naïve Bayes dengan Laplace Correction dalam klasifikasi potensi banjir menggunakan dataset curah hujan Kerala. Hasil pengujian menunjukkan bahwa kedua model menghasilkan performa yang sama,





yaitu accuracy 88,89%, precision 85,00%, recall 94,44%, dan F1-score 89,47%. Laplace Correction terbukti mampu mengatasi zero probability pada atribut Annual Rainfall, namun tidak memberikan peningkatan performa klasifikasi pada dataset yang digunakan.

5.2 Saran

Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam untuk mengevaluasi pengaruh Laplace Correction pada karakteristik data yang berbeda. Selain itu, perbandingan dapat diperluas dengan algoritma lain seperti Decision Tree, Random Forest, dan Support Vector Machine, serta menambahkan variabel pendukung seperti tinggi muka air, kelembapan udara, penggunaan lahan, dan kondisi drainase untuk mendukung pengembangan sistem peringatan dini banjir.

DAFTAR RUJUKAN

- Chen, H., Hu, S., Hua, R., & Zhao, X. (2021). Improved naive Bayes classification algorithm for traffic risk management. *EURASIP Journal on Advances in Signal Processing*, 6, 1–12. <https://doi.org/10.1186/s13634-021-00742-6>
- Indaryono, N. A. P., Saedudin, R. R., & Hamami, F. (2024). Analisa Perbandingan Algoritma Random Forest Dan Naïve Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(1), 158–167. <https://doi.org/10.29100/jipi.v9i1.4421>
- Intan, I., Rismayani, Aminah Dinayati Ghani, S., Nurdin, & Koswara, A. T. (2021). Analisis Performansi Prakiraan Cuaca Menggunakan Algoritma Machine Learning. *Jurnal Pekommas*, 6(2), 1–8. <https://doi.org/10.30818/jpkm.2021.2060221>
- Kim, T., & Lee, J. (2022). Exponential Loss Minimization for Learning Weighted Naive Bayes Classifiers. *IEEE Access*, 10, 22724–22736. <https://doi.org/10.1109/ACCESS.2022.3155231>
- Kurniawan, Y. I., Razi, F., Wijayanto, B., & Hidayat, M. L. (2021). Naive Bayes modification for intrusion detection system classification with zero probability. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2751–2758. <https://doi.org/10.11591/eei.v10i5.2833>
- Maulita, M., & Nurdin, N. (2023). Pendekatan Data Mining Untuk Analisa Curah Hujan Menggunakan Metode Regresi Linear Berganda (Studi Kasus: Kabupaten Aceh Utara). *IDEALIS : InDonEsiA Journal Information System*, 6(2), 99–106. <https://doi.org/10.36080/idealism.v6i2.3034>
- Meladiar, O., & Indra, I. (2024). Penerapan Text Mining Untuk Klasifikasi Informasi Banjir Di Jakarta Berdasarkan Data Twitter Implementation of Text Mining for Flood Information Classification in





- Jakarta Based on Twitter Data Using. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, 3(September), 761–768. <https://senafti.budiluhur.ac.id/senafti/article/view/1493>
- Natzir, S. M. (2023). Perbandingan Kinerja Model Pembelajaran Mesin Dalam Prediksi Banjir Menggunakan KNN, Naive Bayes Dan Random Forest. *Jurnal Teknologi Informasi*, 14(2), 59–64. <https://doi.org/DOI> : <https://doi.org/10.52972/hoaq.vol14no1>
- Nuridin, N., Suhendri, M., Afrilia, Y., & Rizal, R. (2021). SISTEMASI: Jurnal Sistem Informasi Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (Nbc). *SISTEMASI: Jurnal Sistem Informasi*, 10(2), 268–279. <https://doi.org/10.32520/stmsi.v10i2.1193>
- Perdana Putra, M. R., Rama Ashari, Muhirin, Azib Widad Zuhaily Imam, & Kusriani. (2025). Flood Prediction Using Machine Learning Model Integrated with Geographical Information System. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, 10(2), 121–126. <https://doi.org/10.23917/khif.v10i2.3723>
- Pitaloka, E., Hartanto, T. B., & Sandiwarno, S. (2024). Penerapan Machine Learning Untuk Prediksi Bencana Banjir. *Jurnal Sistem Informasi Bisnis*, 14(1), 62–76. <https://doi.org/10.21456/vol14iss1pp62-76>
- Refida Septiana Putri, Randika, R. P., Atiqah Noor Zhaafirah, Febi Dwi Sasmita, Adhisa Nanda Kurnia, & Albab Muzaki. (2025). Optimasi Prediksi Bencana Banjir Menggunakan Teknik Smote Berbasis Algoritma Naive Bayes. *Journal of Informatics and Interactive Technology*, 2(2), 354–360. <https://doi.org/10.63547/jiite.v2i2.68>
- Sinatrya, I. M., Pohan, A. B., Yunita, Y., Amalia, H., & Lestari, A. F. (2025). Penerapan Integrasi Algoritma K-Means Dan Naive Bayes Untuk Klasifikasi Wilayah Rawan Banjir Di Jakarta. *Computer Science (CO-SCIENCE)*, 5(2), 67–76. <https://doi.org/10.31294/coscience.v5i2.6900>
- Triyanto, S., Sunyoto, A., & Arief, M. R. (2021). Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naive Bayes. *JOISIE Journal Of Information System And Informatics Engineering*, 5(2), 109–117. <https://ejournal.pelitaIndonesia.ac.id/ojs32/index.php/JOISIE/article/download/1785/790>

