

Perbandingan Performa Algoritma *Naïve Bayes* dan SVM untuk Klasifikasi Cuaca Ekstrim di Kabupaten Sintang

Dwi Noor Jumiati¹, Putri Yuli Utami^{*2}, Rizki Surtiyan Surya³

Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer,
Universitas Muhammadiyah Pontianak

Email: ¹211230002@unmuhpnk.ac.id, ²putriyuli@unmuhpnk.ac.id, ³rizki.surtiyan@unmuhpnk.ac.id

(Naskah masuk: 7 Januari 2026, diterima untuk diterbitkan: 15 April 2026)

Abstrak: Fenomena perubahan iklim yang semakin kompleks telah memengaruhi pola cuaca di berbagai wilayah, termasuk Kabupaten Sintang, yang ditandai dengan meningkatnya intensitas hujan deras maupun suhu panas tinggi yang berpotensi mengganggu aktivitas sosial dan ekonomi masyarakat. Kondisi tersebut menuntut adanya model klasifikasi yang mampu mengidentifikasi cuaca ekstrim secara akurat guna mendukung mitigasi bencana dan kebijakan daerah. Penelitian ini bertujuan membandingkan kinerja algoritma *naïve bayes* dan Support Vector Machine (SVM) dalam mengklasifikasikan cuaca ekstrim menggunakan parameter meteorologi pada periode 2020–2024. Data diolah melalui normalisasi Min-Max Scaler, pembagian data latih dan uji, serta pengujian performa menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC. Hasil penelitian menunjukkan bahwa SVM menghasilkan performa lebih baik dibandingkan *naïve bayes* dengan nilai akurasi dan metrik evaluasi lainnya yang lebih tinggi. Pembagian data dengan proporsi 80:20 memberikan akurasi terbaik dan lebih stabil dibandingkan proporsi lainnya. Uji validasi variabel menunjukkan bahwa curah hujan (RR) merupakan faktor paling dominan, dibuktikan dengan penurunan akurasi signifikan hingga 0,59 ketika variabel ini dihilangkan. Dengan demikian, algoritma SVM dinyatakan lebih sesuai digunakan untuk klasifikasi cuaca ekstrim di Kabupaten Sintang, serta hasil penelitian ini diharapkan menjadi dasar strategi mitigasi bencana dan perumusan kebijakan berbasis data.

Kata Kunci – Cuaca Ekstrim; Klasifikasi; *Naïve Bayes*; Sintang; Support Vector Machine

Performance Comparison of *Naïve Bayes* and SVM Algorithms for Extreme Weather Classification in Sintang Regency

Abstract: The increasingly complex phenomenon of climate change has affected weather patterns in various regions, including Sintang Regency, which is characterized by the rising intensity of heavy rainfall and high temperatures that potentially disrupt social and economic activities. These conditions require a classification model capable of accurately identifying extreme weather to support disaster mitigation and regional policy-making. This study aims to compare the performance of the *Naïve bayes* and Support Vector Machine (SVM) algorithms in classifying extreme weather using meteorological parameters during the 2020–2024 period. Data processing was carried out through Min-Max Scaler normalization, splitting into training and testing sets, and performance evaluation using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results show that SVM provides better performance than *Naïve bayes*, as indicated by higher accuracy and other evaluation metrics. The 80:20 data split produced the best and most stable accuracy compared to other proportions. Variable validation tests revealed that rainfall (RR) is the most dominant factor, as indicated by a significant decrease in accuracy to 0.59 when this variable was removed. Therefore, the SVM algorithm is considered more suitable for extreme weather classification in Sintang Regency, and the findings of this study are expected to serve as a foundation for disaster mitigation strategies and data-driven policy formulation.

Keywords – Extreme Weather; Classification; *Naïve Bayes*; Sintang; Support Vector Machine

1. PENDAHULUAN

Di Indonesia suhu meningkat sekitar 0,2 hingga 0,3 derajat celsius setiap dekade [1]. Kenaikan suhu ini berdampak pada ketidakstabilan iklim, yang ditandai dengan meningkatnya curah hujan, angin topan yang lebih kuat, perubahan pola musim hujan dan kemarau, serta munculnya cuaca

ekstrim yang sulit diprediksi [2]. Cuaca ekstrim di Indonesia terjadi baik saat musim hujan maupun kemarau. Salah satu daerah yang sering terdampak adalah Kabupaten Sintang.

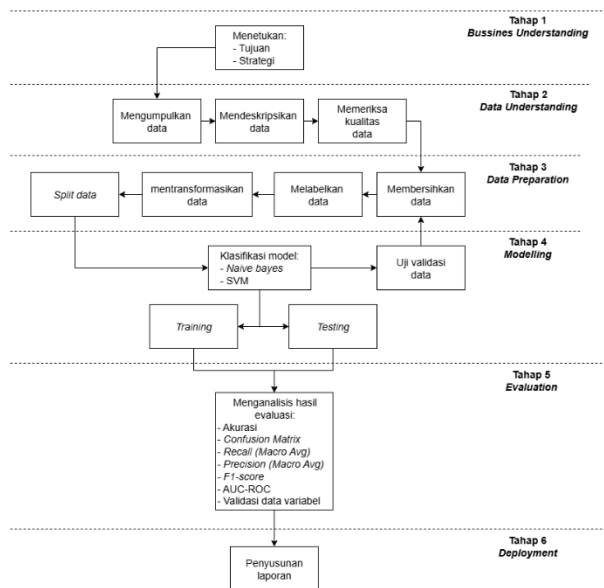
Kabupaten Sintang merupakan wilayah iklim tropis basah yang memiliki curah hujan tinggi sepanjang tahun, sehingga wilayah ini rentan terhadap bencana hidrometeorologis. BPBD Kalimantan Barat mencatat bahwa pada periode 2019–2024 telah terjadi banjir serta kebakaran hutan dan lahan yang berdampak pada kerusakan permukiman, infrastruktur, serta gangguan aktivitas masyarakat, termasuk dampak kesehatan akibat genangan yang berkepanjangan [3]. Selain itu, paparan asap dan gas beracun dari karhutla memicu gangguan pernapasan dan iritasi kesehatan lainnya, terutama pada kelompok rentan [4].

Salah satu cara untuk mengurangi dampak bencana akibat cuaca ekstrim adalah dengan menyediakan informasi prediksi cuaca yang akurat melalui metode klasifikasi data mining. Algoritma yang digunakan dalam penelitian adalah *naïve bayes* dan *Support Vector Machine* (SVM). Metode klasifikasi *naïve bayes* didasarkan pada probabilitas dan perhitungan statistik yang dikembangkan oleh ilmuwan Inggris Thomas sementara metode klasifikasi SVM menitikberatkan pada ide dasar pengklasifikasian data, yaitu sebagai upaya mencari *hyperplane* optimal yang berperan sebagai pemisah antara dua kelas di ruang masukan [5]. Naive bayes unggul dalam kesederhanaan, kecepatan, dan akurasi [6]. Sedangkan SVM dapat memisahkan data *non-linear* yang berdimensi tinggi menggunakan fungsi kernel [7]. Penelitian Nur Kirana menunjukkan akurasi *naïve bayes* sebesar 84,26% [8]. Sedangkan penelitian dari Mahendra melaporkan akurasi SVM sebesar 83% [9].

Tujuan penelitian yaitu untuk membandingkan kinerja algoritma *naïve bayes* dan *Support Vector Machine* (SVM) pada klasifikasi data cuaca ekstrim di Kabupaten Sintang ke dalam lima kategori: hujan normal, hujan ekstrim, panas normal, panas ekstrim, dan mendung. Data diperoleh dari BMKG Tebelian Sintang untuk periode 2020–2024 dan dianalisis menggunakan pendekatan *data mining* guna meningkatkan akurasi prediksi cuaca ekstrim sebagai dasar mitigasi risiko dan sistem peringatan dini. Hasil penelitian diharapkan menjadi referensi dalam pengembangan model prediksi cuaca ekstrim berbasis *naïve bayes* dan SVM pada wilayah beriklim tropis.

2. METODE PENELITIAN

2.1. Tahapan Penelitian



Gambar 1. Tahapan CRISP-DM

Berdasarkan Gambar 1, tahapan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) memiliki 6 tahapan sebagai berikut [10].

2.1.1 Business understanding

Penelitian ini bertujuan menentukan metode terbaik dalam mengidentifikasi faktor cuaca ekstrim di Kabupaten Sintang menggunakan algoritma *naive bayes* dan SVM. Hasilnya diharapkan mendukung BMKG dalam pengambilan keputusan dan mitigasi bencana melalui penyediaan informasi cuaca ekstrim yang tervisualisasi.

2.1.2 Data understanding

Data yang digunakan merupakan data historis selama periode 2020 hingga 2024, dengan total sebanyak 1.842 entri. Dataset ini terdiri atas 11 variabel, yaitu TANGGAL, TN (temperatur minimum), TX (temperatur maksimum), TAVG (temperatur rata-rata), RH_AVG (kelembapan rata-rata), RR (curah hujan), SS (lama penyinaran matahari), FF_X (kecepatan angin maksimum), DDD_X (arah angin maksimum), FF_AVG (kecepatan angin rata-rata), dan DDD_CAR (arah angin terbanyak).

```
file_path = 'stg2020-2024.xlsx'  
df = pd.read_excel(file_path)  
df = df[['TANGGAL', 'TN', 'TX', 'RH_AVG', 'SS', 'RR']]  
df[['TN', 'TX', 'RH_AVG', 'SS', 'RR']] = df[['TN', 'TX', 'RH_AVG', 'SS',  
'RR']].apply(pd.to_numeric, errors='coerce')  
print("Jumlah data kosong (NaN):", df.isna().sum().sum())  
print("Jumlah data duplikat:", df.duplicated().sum())  
print("Jumlah data dengan nilai 8888:", (df[['TN', 'TX', 'RH_AVG',  
'SS', 'RR']] == 8888).any(axis=1).sum())
```

Gambar 2. Kode Pada Data Understanding

Pada Gambar 2 menunjukkan pemeriksaan kualitas data yang dilakukan berdasarkan data iklim pada file *stg2020-2024.xlsx* yang diakses menggunakan fungsi *read_excel()* dari pustaka *pandas*, dengan pemilihan variabel relevan berupa TANGGAL, TN, TX, RH_AVG, SS, dan RR. Seluruh kolom numerik kemudian dikonversi menggunakan *pd.to_numeric(errors='coerce')* untuk memastikan nilai tidak valid berubah menjadi *NaN*. Validasi data mencakup identifikasi nilai hilang melalui *df.isna().sum().sum()*, deteksi duplikasi dengan *df.duplicated().sum()*, serta pemeriksaan nilai tidak wajar khususnya 8888 menggunakan logika komparatif dan fungsi *any()* serta *sum()* untuk menentukan jumlah data bermasalah. Hasil pemeriksaan menunjukkan bahwa terdapat sejumlah data yang tidak memenuhi kriteria kualitas data yang baik. Ditemukan sebanyak 21 entri data kosong (*NaN*), terdapat 109 entri data dengan nilai tidak wajar (8888), dan adanya 15 data duplikat.

2.1.3 Data preparation

```
df = df.dropna()  
df = df[~(df[['TN', 'TX', 'RH_AVG', 'SS', 'RR']] == 8888).any(axis=1)]  
df = df.drop_duplicates()  
  
def klasifikasi_cuaca(row):  
    if row['RR'] > row['SS'] and row['RR'] > 50:  
        return 'Hujan Ekstrim'  
    elif row['RR'] < row['SS'] and row['TX'] >= 35:  
        return 'Panas Ekstrim'  
    elif row['RR'] > row['SS'] and row['RR'] <= 50:  
        return 'Hujan Normal'  
    elif row['RR'] < row['SS'] and row['TX'] < 35:  
        return 'Panas'  
    elif row['RR'] == row['SS']:  
        return 'Mendung Normal'  
    else:  
        return 'Tidak Terdefinisi'  
df['CUACA'] = df.apply(klasifikasi_cuaca, axis=1)  
X = df[['TN', 'TX', 'RH_AVG', 'SS', 'RR']]  
y = df['CUACA']  
scaler = MinMaxScaler()  
X_normalized = scaler.fit_transform(X)  
X_normalized_df = pd.DataFrame(X_normalized, columns=X.columns)  
  
X_train, X_test, y_train, y_test = train_test_split(X_normalized, y,  
                                                    test_size=0.1, random_state=42, stratify=y)
```

Gambar 3. Kode Pada Data Preparation

Pada tahap ini, dilakukan untuk memastikan kualitas dan kesiapan data sebelum dianalisis lebih lanjut. Berdasarkan Gambar 3, proses ini mencakup pembersihan data dengan cara menghapus nilai kosong (*NaN*), data yang tidak valid seperti nilai 8888, serta data duplikat. Selain itu, dilakukan pemilihan atribut. Kemudian dilakukan pelabelan cuaca yang terdiri dari hujan ekstrim, panas ekstrim, hujan normal, panas normal dan mendung. Normalisasi data juga diterapkan menggunakan *Min-Max Scaling* untuk menyetarakan skala antarvariabel. Selanjutnya, pemisahan data menjadi data latih (*training data*) dan data uji (*testing data*) yang akan digunakan pada proses pelatihan dan pengujian model prediktif secara sistematis.

2.1.4 Modelling

Proses ini dilakukan menerapkan algoritma *naive bayes* dan SVM pada data menggunakan tools *Visual Studio Code* dengan bahasa pemrograman *python*.

1. Algoritma *Naive Bayes*

Berikut tahapan dari proses perhitungan metode *naive bayes* [11]:

- 1). Hitung jumlah dan probabilitas awal (*Prior*)

$$P(\text{kelas}) = \frac{\text{jumlah data di kelas } y}{\text{total data}} \quad (1)$$

P : Probabilitas pada kelas data

Y : Kelas data

- 2). Hitung nilai rata-rata (μ) dan *varians* tiap kelas (σ^2).

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (2)$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{n} \quad (3)$$

Berikut keterangan rumus di atas.

μ : Rata-rata kelas

x_1, \dots, x_n : Data kelas y pada fitur X

n : Jumlah kelas y pada fitur X

σ^2 : *Varians* per fitur

- 3). Hitung data *testing* (*Likelihood*)

$$P(x_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \quad (4)$$

- 4). Hitung probabilitas akhir (*Posterior*)

Jumlahkan semua nilai *likelihood* perkelas menggunakan persamaan (5). Kemudian hitung probabilitas akhir untuk menentukan kelas menggunakan persamaan (6).

$$P(C | x_1, \dots, x_k) = P(C | x_1) * \dots * P(C | x_k) \quad (5)$$

$$P(C_i | x) = \frac{P(C_i | x) P(C_i)}{\sum \text{total kelas} * P(C | x) P(C)} \quad (6)$$

Berikut ini keterangan dari rumus di atas.

$P(C | x_1, \dots, x_k)$: Probabilitas kelas C diberikan fitur-fitur x_1, x_2, \dots, x_k

C : Kelas (label) yang ingin diprediksi

$P(C_i | x_i)$: Probabilitas kelas C_i diberikan fitur x_i

$P(C_i | x)$: Probabilitas mengamati x jika kelasnya adalah C_i (*likelihood*).

$P(C_i)$: Probabilitas awal (*prior probability*) dari kelas C_i

$P(C | x)$: *Likelihood* untuk kelas C

$P(C)$: Probabilitas awal dari kelas C

Kode algoritma *naive bayes* menggunakan bahasa *python* dalam melakukan proses *training* dan *testing* ditunjukkan pada Gambar 4.

```

model_nb = GaussianNB()
model_nb.fit(X_train, y_train)
y_test_pred = model_nb.predict(X_test)
y_train_pred = model_nb.predict(X_train)
```

Gambar 4. Kode Algoritma *Naive Bayes*

2. Algoritma SVM

Berikut tahapan dari proses perhitungan algoritma SVM:

- 1). Menghitung kernel RBF data *training* menggunakan persamaan (7).

$$K(x_i, x) = \exp(-\gamma ||x_i - x||^2) \quad (7)$$

Berikut keterangan dari rumus di atas.

- $K(x_i, x)$: Kernel RBF
- γ : Parameter kernel RBF yang mengatur "lebar" pengaruh titik data.
- $\|x_i - x\|^2$: Jarak kuadrat *Euclidean* antara vektor fitur x_i dan x

2). Setelah itu mencari nilai a (alpha) menggunakan rumus *lagrange multiplier* yang ditunjukkan pada persamaan (8) dan b (bias) menggunakan persamaan (9).

$$L(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j x_i^T x_j \quad (8)$$

$$b = y_k - \sum_{i=1}^n a_i y_i K(x_i, x_k) \quad (9)$$

Berikut ini keterangan dari rumus di atas.

- $L(a)$: Fungsi objektif *lagrange multiplier*
- a_i : Variabel *lagrange multiplier* ke- i
- y_i : Label kelas dari data ke- i (+1 atau -1)
- x_i : Vektor fitur dari data ke- i
- Faktor $\frac{1}{2}$: Konstanta dari bentuk kuadrat di fungsi dual SVM.
- b : Nilai bias pada fungsi SVM
- y_k : Label kelas dari support vector ke- k .

3). Kemudian menghitung nilai fungsi keputusan untuk menentukan kelas pada data *testing* menggunakan persamaan (10) dengan kode python pada Gambar 5.

$$f(x) = \sum_{i=1}^n a_i y_i K(x_i, x) + b \quad (10)$$

```
svm_model = SVC(C=1, kernel='rbf', gamma='scale',
probability=True, decision_function_shape='ovr',
random_state=42)
svm_model.fit(X_train, y_train)
y_test_pred = svm_model.predict(X_test)
y_train_pred = svm_model.predict(X_train)
```

Gambar 5. Kode Algoritma SVM

2.1.5 Evaluation

Evaluasi model menggunakan *confusion matrix*, *accuracy*, *macro average recall*, *macro average precision*, *macro average F1-score*, dan AUC-ROC. Persamaan untuk menghitung *accuracy*, *macro average recall*, *macro average precision*, *macro average F1-score*, dan AUC-ROC terdapat pada persamaan (11-17).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$MAP = \frac{\sum_{k=1}^K Precision_k}{K} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$MAR = \frac{\sum_{k=1}^K Recall_k}{K} \quad (15)$$

$$F1 - score = 2 \frac{Presisi * Recall}{Presisi + Recall} \quad (16)$$

$$MAF = \frac{\sum_{k=1}^K F1 - Score_k}{K} \quad (17)$$

3. HASIL DAN PEMBAHASAN

3.1. Hasil Implementasi Algoritma Naïve Bayes dan SVM

Implementasi algoritma *naïve bayes* dan SVM berupa perhitungan manual yang bertujuan untuk memahami proses identifikasi cuaca ekstrim berdasarkan data parameter yang diperoleh. Pada perhitungan manual ini menggunakan sampel 10 data parameter cuaca. Berikut ini perhitungan manual pada algoritma *naïve bayes* dan algoritma SVM.

3.1.1. Langkah-langkah Algoritma Naïve Bayes

1). Hitung jumlah dan probabilitas awal (*Prior*)

Tabel 1. Data Normalisasi *Training* Cuaca Ekstrim

NO	TN	TX	RH-AVG	SS	RR	CUACA
x1	0,1111	0	0,8	0,9219	1	Hujan Normal
x2	0,6667	0,6447	0,7	0,2813	0,3379	Hujan Ekstrim
x3	0,2222	0,5263	0,6	0	0	Mendung
x4	0,8889	0,4475	0,3	0,9844	0,0367	Panas Normal
x5	1	0,4868	0,6	0,9375	0,0153	Panas Normal
x6	0,1667	0,8684	0,4	1	0,0268	Panas Ekstrim
x7	0,7778	0,3421	0,5	0	0	Mendung
x8	0,3889	1	0	0,8594	0,0015	Panas Ekstrim
x9	0,7778	0,5526	0,6	0,9844	0,0734	HujanNormal
x10	0	0,2105	1	0,9531	0,9174	Hujan Ekstrim

Berikut ini perhitungan probabilitas awal berdasarkan data pada Tabel 1.

Panas Normal = 2/10=0,2

Hujan Normal = 2/10=0,2

Panas Ekstrim = 2/10=0,2

Hujan Ekstrim = 2/10=0,2

Mendung = 2/10=0,2

2). Hitung nilai rata-rata (μ) dan *varians* tiap kelas (σ^2).

Misalnya pada variabel TN untuk kelas Hujan Ekstrim, TN (0,1111, 0):

$\mu = (0,1111+0)/2$

=0,05555

$\sigma^2 = ((0,1111 - 0,05555)^2 + (0 - 0,05555)^2)/2$

=0,003085803

Tabel 2. Nilai Rata-Rata Dan *Varians Naïve Bayes* Kelas Hujan Ekstrim

Atribut	μ	σ^2
TN	0,05555	0,003085803
TX	0,10525	0,011077563
RH-AVG	0,9	0,01
SS	0,9375	0,00024336
RR	0,9587	0,00170569

3). Hitung data *testing* (*Likelihood*)

Perhitungan probabilitas *likelihood* pada Gambar 6 untuk kelas Hujan Ekstrim menggunakan data *testing* pada Tabel 3 sebagai berikut.

Tabel 3. Data *Testing* Cuaca Ekstrim

TN	TX	RH-AVG	SS	RR
0,2222	0,8684	0,5	0,9063	0,0229

$$P(TN|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi} (0,003085803)} \exp\left(-\frac{(0,2222 - 0,05555)^2}{2 (0,003085803)}\right) = 0,07978125$$

Gambar 6. Perhitungan Nilai *Likelihood* Data *Testing*

4). Hitung probabilitas akhir (*Posterior*)

Hitung total nilai *likelihood* perkelas seperti Gambar 7, total nilai *likelihood* pada kelas Hujan Ekstrim (HE):

$$P(HE|x) = 0,07978125 * 1,45301E - 11 * 0,001338302 * 3,460959392 * 3,1548E - 111 = 1,6939E - 125$$

Gambar 7. Total Nilai *Likelihood*

Selanjutnya hitung probabilitas akhir seperti Gambar 8 dengan hasil pada Tabel 4:

$$P(HE|x) = \frac{(1,6939E - 125) * 0,2}{(1,65292E - 87 + (4,81847E - 25) + (1005,700812) + (1,6939E - 125) + 0) * 0,2} = 1,6843E - 128$$

Gambar 8. Perhitungan Nilai Probabilitas Akhir (*Posterior*)

Tabel 4. Nilai Probabilitas Akhir (*Posterior*) *Naive Bayes*

Kelas	Nilai probabilitas akhir
Panas Normal	1,64355E-90
Hujan Normal	4,79115E-28
Panas Ekstrim	1
Hujan Ekstrim	1,6843E-128
Mendung	0

Berdasarkan Tabel 4, hasil keseluruhan probabilitas akhir menunjukkan bahwa kelas panas ekstrim memperoleh nilai tertinggi yaitu 1 sehingga kondisi cuaca dari data yang telah diolah adalah panas ekstrim.

3.1.2. Langkah-Langkah Algoritma SVM

1). Hitung kernel RBF (*Radial Basis Function*)

Tabel 5. Data *Training* Cuaca Ekstrim Pada SVM

NO	TN	TX	RH-AVG	SS	RR	CUACA
x1	0,1111	0	0,8	0,9219	1	Hujan Ekstrim
x2	0,6667	0,6447	0,7	0,2813	0,3379	Hujan Normal
x3	0,2222	0,5263	0,6	0	0	Mendung
x4	0,8889	0,4475	0,3	0,9844	0,0367	Panas Normal
x5	1	0,4868	0,6	0,9375	0,0153	Panas Normal
x6	0,1667	0,8684	0,4	1	0,0268	Panas Ekstrim
x7	0,7778	0,3421	0,5	0	0	Mendung
x8	0,3889	1	0	0,8594	0,0015	Panas Ekstrim
x9	0,7778	0,5526	0,6	0,9844	0,0734	Hujan Normal
x10	0	0,2105	1	0,9531	0,9174	Hujan Ekstrim

Misalnya ambil data baris ke 1 (x1) dan ke 2 (x2) pada Tabel 5. Lakukan pelabelan numerik menggunakan teknik OvR dengan kelas target berlabel +1 (x1) dan selain kelas target berlabel -1 (x2). Kemudian cari nilai *gamma scale* menggunakan persamaan (14).

$$\gamma_{scale} = 1 / (n_{feature} * \text{Var}(x)) \quad (14)$$

Diketahui:

Jumlah fitur = 5

Rata-rata varians 5 variabel = 0,112522672

Sehingga:

$$\gamma_{scale} = 1 / (5 * 0,112522672) = 1,777419577 = 1,78$$

Selanjutnya menghitung kernel RBF seperti Gambar 9:

$$\begin{aligned} \|x_1 - x_1\|^2 &= (0,1111 - 0,1111)^2 + (0 - 0)^2 + (0,8 - 0,8)^2 \\ &\quad + (0,9219 - 0,9219)^2 + (1 - 1)^2 \\ &= 0 \\ K(x_1, x_1) &= \exp(-1,78 * 0) \\ &= 1 \\ \|x_1 - x_2\|^2 &= (0,1111 - 0,6667)^2 + (0 - 0,6447)^2 + (0,8 - 0,7)^2 \\ &\quad + (0,9219 - 0,2813)^2 + (1 - 0,3379)^2 \\ &= 1,58307422 \\ K(x_1, x_2) &= \exp(-1,78 * 1,58307422) \\ &= 0,059732913 \\ \|x_2 - x_2\|^2 &= (0,6667 - 0,6667)^2 + (0,6447 - 0,6447)^2 + (0,7 - 0,7)^2 \\ &\quad + (0,2813 - 0,2813)^2 + (0,3379 - 0,3379)^2 \\ &= 0 \\ K(x_2, x_2) &= \exp(-1,78 * 0) \\ &= 1 \end{aligned}$$

Gambar 9. Perhitungan Nilai Kernel RBF

- 2). Mencari nilai a (alpa) menggunakan rumus lagrange *multiplier* seperti Gambar 10 dan nilai b (bias). Diketahui bahwa Hujan Ekstrim berlabel +1 dan lainnya -1, nilai K (x₁, x₁) adalah 1, K(x₁, x₂) adalah 0,059732913, dan K (x₂, x₂) adalah 1.

$$\begin{aligned} L(a) &= a_1 + a_2 - \frac{1}{2} (a_1^2 K(x_1, x_1) + a_2^2 K(x_2, x_2) + 2y_1y_2 K(x_1, x_2)) \\ &= a_1 + a_2 - \frac{1}{2} (a_1^2 (1) + a_2^2 (1) \\ &\quad + 2a_1a_2(1)(-1) (0,059732913)) \\ &= a_1 + a_2 - \frac{1}{2} (a_1^2 + a_2^2 - 0,059732913 a_1a_2) \end{aligned}$$

Gambar 10. Perhitungan Nilai Alpa

Sehingga diperoleh nilai alpa pada kelas Hujan Ekstrim sebagai berikut:

0,648556; -0,704213; -0,633844; -0,567702; -0,568709; -0,483266; -0,572267; -0,483266; -0,572267; -0,483266; -0,679484; dan 0,735141.

Kemudian hitung nilai b (bias) terhadap kelas Hujan Esktrm yang dimulai dari mengumpulkan nilai a kelas Hujan Ekstrim a ≠ 0, lalu berikan label +1 untuk kelas Hujan Ekstrim dan -1 selain kelas tersebut. Selanjutnya hitung kernel K(xi, xk).

Tabel 6. Hitungan Untuk Mencari Nilai b (bias)

No	a _j	y _j	K(xi, xk)	$\sum_{j=1}^n a_j y_j K(x_j, x_i)$
x1	0,648556	1	1	0,648556
x2	-0,704213	-1	0,059732913	0,042064694
x3	-0,633844	-1	0,020670713	0,013102007
x4	-0,567702	-1	0,029100312	0,016520306
x5	-0,568709	-1	0,026626431	0,015142691
x6	-0,483266	-1	0,035815969	0,01730864
x7	-0,572267	-1	0,011649067	0,006666376
x8	-0,483266	-1	0,00792142	0,003828153
x9	-0,679484	-1	0,052803276	0,035878981
x10	0,735141	1	0,830339692	0,610416751
Total				1,409484599

Berdasarkan Tabel 6 berikut perhitungan nilai bias.

$$b = 1 - 1,409484599$$

$$= -0,409484599$$

- 3). Pengujian terhadap data *testing* yang diawali dengan menghitung nilai RBF pada data *testing* terhadap data *training* yang berlabel Hujan Ekstrim ditunjukkan pada Tabel 7.

Tabel 7. Nilai Kernel RBF Data *Testing*

x	Nilai Kernel Data <i>Testing</i>
x1	0,039781847
x10	0,065124411

Kemudian hitung nilai fungsi keputusan disetiap kelas untuk menentukan kelas pada data *testing* ditunjukkan pada Tabel 8. Berikut contoh pada kelas Hujan Ekstrim.

$$[f(x)] \text{ (Hujan Ekstrim)} = (0,648556*(1)*0,039781847) + (0,735141*(1)*0,065124411)+(-0,409484599) = -0,335808219$$

Tabel 8. Nilai Fungsi Keputusan Disetiap Kelas Cuaca Ekstrim

Kelas	f(x)
Panas Normal	-5,244141592
Hujan Normal	-2,179580409
Panas Ekstrim	1,023303226
Hujan Ekstrim	-0,335808219
Mendung	0,485333232

Berdasarkan Tabel 8, hasil keseluruhan probabilitas akhir menunjukkan bahwa kelas panas ekstrim memperoleh nilai tertinggi yaitu 1 sehingga kondisi cuaca dari data yang telah diolah adalah panas ekstrim.

3.2. Pembahasan Evaluasi Algoritma Naïve Bayes dan SVM

3.2.1. Evaluasi Algoritma Naïve Bayes

Tabel 9. Hasil *Testing* Dan *Training* Pada Algoritma *Naive Bayes*

Split Data	Akurasi Testing	Akurasi Training	Mean Recall (Macro Avg)	Mean Precision (Macro Avg)	F1- Score (Macro Avg)	AUC-ROC (Macro Avg)
50:50	0,8672	0,8872	0,7217	0,6896	0,7043	0,9601
60:40	0,8605	0,8893	0,7174	0,6864	0,701	0,9561
70:30	0,8689	0,8816	0,7404	0,6876	0,7106	0,96
80:20	0,871	0,8788	0,7379	0,7086	0,7225	0,9655
90:10	0,8538	0,8785	0,7403	0,6888	0,7123	0,98

Pada Tabel 9, hasil menunjukkan bahwa akurasi pengujian tertinggi diperoleh pada skenario *split data* 80:20 sebesar 0,871, sedangkan akurasi pelatihan tertinggi terdapat pada *split* 60:40 sebesar 0,8893. Nilai Mean Recall tertinggi dicapai pada *split data* 70:30 sebesar 0,7404, sedangkan Mean Precision terbaik terdapat pada *split data* 80:20 sebesar 0,7086. Selain itu, nilai F1-score tertinggi juga diperoleh pada *split data* 80:20 dengan skor 0,7225, yang menunjukkan keseimbangan terbaik antara precision dan recall. Sementara itu, nilai AUC-ROC tertinggi terdapat pada *split data* 90:10 dengan skor 0,980, yang mengindikasikan kemampuan diskriminasi model yang sangat baik pada skenario tersebut. Secara keseluruhan, perbedaan kinerja antar *split data* relatif kecil, namun *split* 80:20 menunjukkan performa yang konsisten baik pada hampir semua metrik evaluasi.

3.2.2. Evaluasi Algoritma SVM

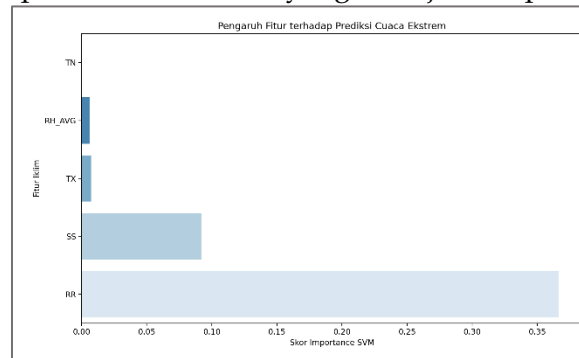
Tabel 10. Hasil *Testing* Dan *Training* Pada Algoritma SVM

Split Data	Akurasi Testing	Akurasi Training	Mean Recall (Macro Avg)	Mean Precision (Macro Avg)	F1- Score (Macro Avg)	AUC-ROC (Macro Avg)
50:50	0,8273	0,8602	0,5384	0,7344	0,54	0,9819
60:40	0,8546	0,8599	0,5551	0,7436	0,5633	0,9859
70:30	0,863	0,8673	0,5637	0,7458	0,5821	0,9905
80:20	0,8768	0,8758	0,5974	0,7488	0,6245	0,9918
90:10	0,8889	0,8818	0,6181	0,7561	0,6363	0,9971

Pada Tabel 10, menunjukkan evaluasi model SVM yang baik pada proporsi data 90:10 dan 80:20. Proporsi 90:10 menghasilkan akurasi testing tertinggi (0,8889), namun dengan selisih akurasi training-testing ($\pm 0,007$) yang lebih besar dibandingkan 80:20 (akurasi testing 0,8768, selisih $\pm 0,001$). Meskipun 90:10 unggul pada sebagian metrik, proporsi 80:20 dipilih karena ukuran data testing yang lebih besar memastikan generalisasi yang lebih stabil, reliabel, dan objektif untuk perbandingan performa algoritma..

3.2. Uji Validasi pada Parameter Cuaca

Berikut ini hasil uji validasi pada variabel cuaca yang ditunjukkan pada Gambar 11.



Gambar 11. Grafik Hasil Uji Validasi

Berdasarkan Gambar 11, menunjukkan bahwa variabel curah hujan (RR) memiliki kontribusi paling besar terhadap klasifikasi cuaca ekstrem dengan skor sekitar 0,37, diikuti oleh lama penyinaran matahari (SS) sebesar 0,10. Sementara itu, kelembapan rata-rata (RH_AVG), suhu maksimum (TX), dan suhu minimum (TN) memiliki pengaruh yang lebih rendah. Temuan ini menegaskan bahwa model SVM lebih mengandalkan faktor presipitasi dan radiasi matahari sebagai penentu utama klasifikasi, sedangkan variabel suhu dan kelembapan berperan sebagai pendukung dalam meningkatkan akurasi prediksi kondisi cuaca ekstrem. Hasil uji validasi pada setiap parameter cuaca ekstrem dapat dilihat pada tabel-tabel sebagai berikut.

Tabel 11. Hasil Pengujian Tanpa TN

Kelas	Hujan Ekstrem	Hujan Normal	Mendung	Panas Ekstrem	Panas Normal
Hujan Ekstrem	18	2	0	0	0
Hujan Normal	0	121	0	0	13
Mendung	0	5	0	0	1
Panas Ekstrem	0	0	0	5	12
Panas Normal	0	6	0	0	158

Tabel 12. Hasil Pengujian Tanpa TX

Kelas	Hujan Ekstrem	Hujan Normal	Mendung	Panas Ekstrem	Panas Normal
Hujan Ekstrem	19	1	0	0	0
Hujan Normal	0	123	0	0	11
Mendung	0	6	0	0	0
Panas Ekstrem	0	0	0	0	17
Panas Normal	0	7	0	0	157

Tabel 13. Hasil Pengujian Tanpa RH-AVG

Kelas	Hujan Ekstrem	Hujan Normal	Mendung	Panas Ekstrem	Panas Normal
Hujan ekstrem	18	2	0	0	0
Hujan normal	0	121	0	0	13
Mendung	0	6	0	0	0
Panas ekstrem	0	0	0	3	14
Panas normal	0	4	0	0	160

Tabel 14. Hasil Pengujian Tanpa SS

Kelas	Hujan Ekstrim	Hujan Normal	Mendung	Panas Ekstrim	Panas Normal
Hujan Ekstrim	19	1	0	0	0
Hujan Normal	0	108	0	0	26
Mendung	0	0	0	0	6
Panas Ekstrim	0	0	0	8	9
Panas Normal	0	6	0	0	158

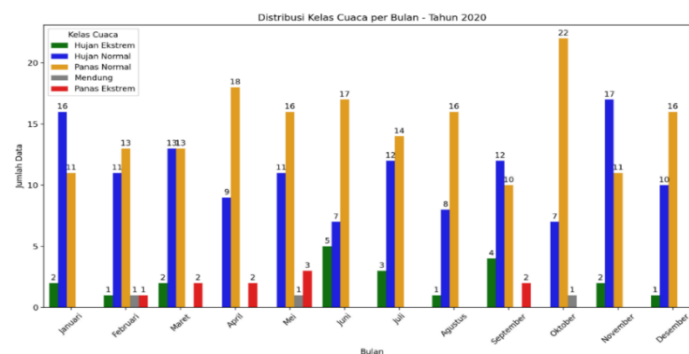
Tabel 15. Hasil Pengujian Tanpa RR

Kelas	Hujan Ekstrim	Hujan Normal	Mendung	Panas Ekstrim	Panas Normal
Hujan Ekstrim	0	15	0	0	5
Hujan Normal	0	81	0	0	53
Mendung	0	6	0	0	0
Panas Ekstrim	0	0	0	7	10
Panas Normal	0	38	0	0	126

Berdasarkan Tabel 11 - Tabel 15, hasil pengujian model SVM tanpa beberapa variabel iklim menunjukkan bahwa penghapusan fitur tertentu berdampak berbeda terhadap performa klasifikasi cuaca. Tanpa variabel TN, TX, dan RG-AVG, model masih mampu mengklasifikasikan sebagian besar kelas dengan baik, meskipun kelas "Mendung" dan "Panas Ekstrim" menunjukkan tingkat kesalahan tinggi. Penghapusan variabel SS menyebabkan penurunan akurasi pada kelas "Hujan Normal" dan "Panas Ekstrim", menandakan bahwa durasi penyinaran matahari berperan penting dalam membedakan kondisi panas ekstrim. Sementara itu, penghapusan variabel RR menghasilkan penurunan kinerja paling signifikan, terutama pada kelas hujan, di mana model gagal mengidentifikasi "Hujan Ekstrim" dengan benar. Semua hasil ini menunjukkan bahwa curah hujan adalah fitur paling krusial pada klasifikasi cuaca ekstrim, diikuti oleh penyinaran matahari sebagai faktor pendukung utama, sedangkan variabel suhu dan kelembapan memiliki kontribusi yang relatif kecil terhadap akurasi model.

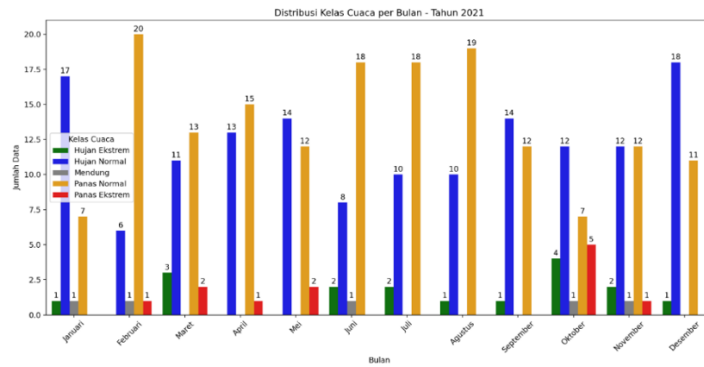
3.3. Visualisasi Data Cuaca Ekstrim

Berikut ini adalah visualisasi dari data cuaca ekstrim di Kabupaten Sintang selama periode 2020-2021.



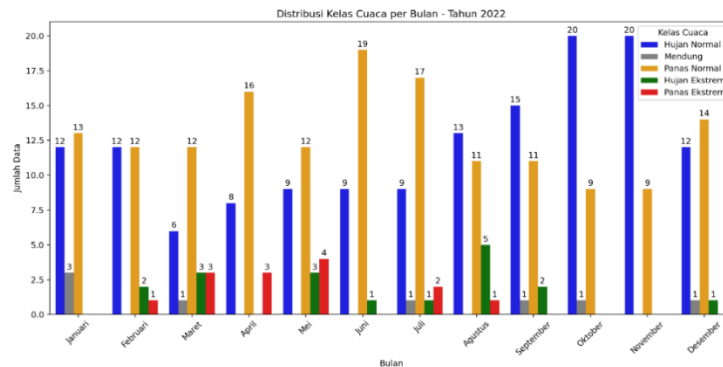
Gambar 12. Kondisi Cuaca Ekstrim Pada Tahun 2020

Berdasarkan Gambar 12, cuaca pada tahun 2020 menunjukkan dominasi kelas "Panas Normal" dan "Hujan Normal", dengan puncak masing-masing terjadi pada Oktober serta Januari dan November. Sebaliknya, kejadian "Hujan Ekstrim" dan "Panas Ekstrim" relatif rendah, sedangkan kategori "Mendung" menjadi yang paling jarang muncul. Hal ini menegaskan bahwa kondisi cuaca sepanjang 2020 lebih banyak berada pada kategori normal.



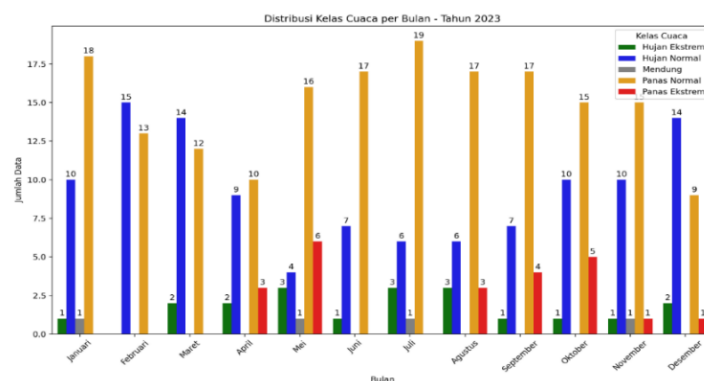
Gambar 13. Kondisi Cuaca Ekstrim Pada Tahun 2021

Berdasarkan Gambar 13, cuaca pada tahun 2021, kondisi cuaca didominasi oleh “Panas Normal” yang mencapai puncak pada Februari dan “Hujan Normal” pada Desember. Cuaca ekstrim hanya terjadi secara sporadis, di mana “Hujan Ekstrem” tertinggi muncul pada Oktober dan “Panas Ekstrem” juga mencapai puncaknya pada Oktober serta muncul di beberapa bulan lain seperti Maret, April, dan Mei. Sementara itu, kelas “Mendung” memiliki frekuensi paling rendah dengan hanya satu kejadian pada sejumlah bulan tertentu.



Gambar 14. Kondisi Cuaca Ekstrim Pada Tahun 2022

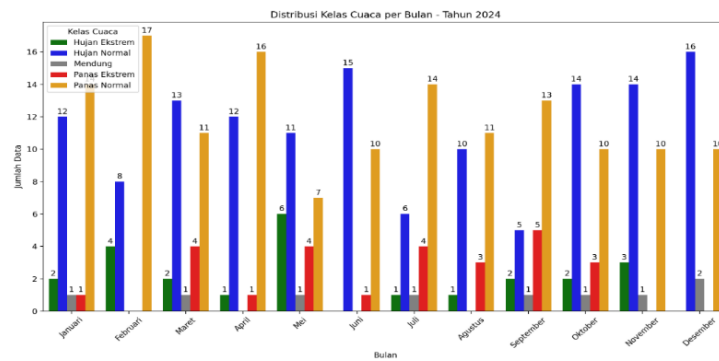
Berdasarkan Gambar 14, cuaca pada tahun 2022 menunjukkan bahwa kelas “Hujan Normal” dan “Panas Normal” mendominasi sepanjang tahun, dengan puncak kejadian masing-masing pada bulan Oktober–November dan Juni. Fenomena cuaca ekstrim, baik “Hujan Ekstrem” maupun “Panas Ekstrem”, hanya terjadi secara sporadis dengan frekuensi tertinggi pada bulan Agustus dan bulan Mei. Sementara itu, kelas “Mendung” merupakan kategori dengan kemunculan paling rendah.



Gambar 15. Kondisi Cuaca Ekstrim Pada Tahun 2023

Berdasarkan Gambar 15, cuaca pada tahun 2023 menunjukkan bahwa kategori “Panas Normal” dan “Hujan Normal” mendominasi sepanjang tahun, dengan puncak masing-masing terjadi pada bulan Juli dan Februari. Sementara itu, cuaca ekstrim seperti “Hujan Ekstrem” dan “Panas Ekstrem” hanya muncul secara sporadis dengan frekuensi yang jauh lebih rendah, di mana

puncak “Panas Ekstrim” tercatat pada bulan Mei. Kategori “Mendung” menjadi kondisi paling jarang terjadi.



Gambar 16. Kondisi Cuaca Ekstrem Pada Tahun 2024

Berdasarkan Gambar 16, cuaca pada tahun 2024 memperlihatkan bahwa kategori “Panas Normal” dan “Hujan Normal” mendominasi sepanjang tahun, dengan puncak kejadian masing-masing terjadi pada bulan Februari dan Desember. Cuaca ekstrem muncul dengan frekuensi rendah, di mana “Hujan Ekstrem” tertinggi terjadi pada bulan Mei dan “Panas Ekstrem” terjadi pada bulan September. Sementara itu, kategori “Mendung” jarang terdeteksi sehingga menjadi kondisi paling jarang terjadi.

4. KESIMPULAN

Kinerja yang lebih unggul ditunjukkan oleh algoritma SVM dibandingkan *Naive Bayes* pada seluruh metrik evaluasi, baik pada data latih maupun data uji, sehingga lebih mampu mengenali pola *non-linear* dan memiliki generalisasi yang lebih baik. Pengujian *split data* juga mengindikasikan bahwa proporsi 80:20 merupakan konfigurasi optimal karena memberikan akurasi tertinggi dibandingkan proporsi lainnya. Proporsi 80:20 dinilai optimal karena mampu menjaga keseimbangan antara jumlah data latih yang cukup untuk membentuk model dan jumlah data uji yang sangat representatif untuk mengevaluasi kinerja model. Hasil uji validasi variabel menunjukkan bahwa setiap parameter cuaca memiliki kontribusi berbeda terhadap akurasi model, di mana penghapusan TN, TX, RH_AVG, dan SS menurunkan akurasi menjadi 0,65; 0,64; 0,62; dan 0,63. Penurunan terbesar terjadi saat RR dihilangkan, dengan akurasi hanya 0,59, sehingga curah hujan dapat dipastikan sebagai fitur paling dominan dalam klasifikasi cuaca ekstrem di Kabupaten Sintang. Penelitian selanjutnya disarankan menggunakan data meteorologi dengan cakupan yang lebih luas serta mengevaluasi model lain untuk perbandingan performa yang lebih komprehensif, serta menambahkan variabel tambahan seperti kecepatan angin, tekanan udara, dan faktor lingkungan lain yang berpotensi memengaruhi kejadian cuaca ekstrem.

DAFTAR PUSTAKA

- [1] Umi Zuhriyah, “Dampak Perubahan Iklim Terhadap Sektor Pertanian di Indonesia,” *Tirto.Id*, pp. 1-8, 2022, [Online]. Available: <https://tirto.id/dampak-perubahan-iklim-terhadap-sektor-pertanian-di-indonesia-gyHf>
- [2] A. S. Mulyani, “Pemanasan Global, Penyebab, Dampak dan Antisipasinya,” *Artik. Pengabd. Masy.*, pp. 1-27, 2021.
- [3] BPBD Provinsi Kalbar, “Rekap Kejadian Bencana Bansor,” pp. 1-14, 2023.
- [4] Erwin, I. R. Mubayanah, R. Mubarok, and A. Mutasyarifin, “Pelatihan Kesiapsiagaan Kebakaran Di Permukiman Desa Tepian Makmur,” *J. Pengabd. Kpd. Masy. P3M STAI Sangatta*, vol. 01, no. 01, pp. 25-37, 2024.
- [5] T. Risanindya, W. Purbasari, and L. Riyandari, “Analisis Sentimen Ulasan Pengguna Aplikasi Maxim pada Google Play Store dengan Metode Support Vector Machine (SVM) dan Naïve Bayes Sentiment Analysis of Maxim App User Reviews on Google Play Store with Support Vector Machine (SVM) and Naïve Bayes Methods,” vol. 14, no. 105, pp. 859-867, 2025.

- [6] S. Sinaga, R. W. Sembiring, and S. Sumarno, "Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru," *J. Mach. ...*, vol. 1, no. 1, pp. 55-64, 2022, [Online]. Available: <https://journal.fkpt.org/index.php/malda/article/view/162%0Ahttps://journal.fkpt.org/index.php/malda/article/download/162/115>
- [7] A. R. I. Pratama, S. A. Latipah, and B. N. Sari, "Optimasi Klasifikasi Curah Hujan Menggunakan Support Vector Machine (Svm) Dan Recursive Feature Elimination (Rfe)," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 2, pp. 314-324, 2022, doi: 10.29100/jipi.v7i2.2675.
- [8] A. Nur Kirana, B. Nurhakim, S. Eka Permana, W. Prihartono, and G. Dwilestari, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1637-1642, 2024, doi: 10.36040/jati.v8i2.8967.
- [9] M. Mahendra and N. Azizah, "Implementation of Machine Learning to Predict the Weather Using a Support Vector Machine: Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine," 2023, [Online]. Available: <http://dx.doi.org/10.21070/ups.2889>
- [10] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103-108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [11] J. V. Wie and M. Siddik, "Penerapan Metode Naive Bayes Dalam Mengklasifikasi Tingkat Obesitas Pada Pria," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 6, no. Desember, pp. 69-77, 2022, [Online]. Available: <https://www.kaggle.com/>,