

## Rainfall Classification Based on El-Niño and La-Niña Climate Phenomenon Using Naive Bayes Classifier Algorithm

Mely Erlinda<sup>1\*</sup>, Andrianingsih Andrianingsih<sup>2</sup>

<sup>1,2</sup> Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Jakarta Selatan, 12520, Indonesia

### Informasi Artikel

Diterima : 28 Februari 2025  
Revisi : 3 Maret 2025  
Publikasi : 20 Maret 2025

### Kata Kunci:

Curah Hujan  
El-Niño  
La-Niña  
Naive Bayes Classifier  
CRISP-DM

### ABSTRAK

Indonesia sebagai negara tropis menghadapi dampak signifikan dari fenomena *El Niño* dan *La Niña* terhadap pola curah hujan. Penelitian ini bertujuan untuk mengklasifikasikan curah hujan harian di kota-kota besar Indonesia seperti, DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung, ke dalam tiga kategori: tidak hujan, hujan sedang, dan hujan ekstrem, serta mengidentifikasi kondisi iklim berdasarkan kategori *El Niño*, *La Niña*, dan Normal menggunakan algoritma *Naive Bayes Classifier*. Dengan menerapkan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka pemrosesan dataset curah hujan harian periode Januari hingga Desember 2023, yang diperoleh dari BMKG. Hasil analisis menunjukkan bahwa algoritma *Naive Bayes Classifier* memiliki performa tinggi dengan akurasi 93,15%, *precision* 98%, *recall* 93%, dan *F1-score* 94%. Analisis lebih lanjut, menunjukan bahwa *El Niño* menyebabkan penurunan curah hujan yang signifikan, sedangkan *La Niña* meningkatkan curah hujan ekstrem, terutama di Medan. Studi ini berkontribusi dalam pengembangan model klasifikasi curah hujan yang dapat membantu pemerintah untuk mengantisipasi dampak perubahan iklim dan meningkatkan efisiensi pengelolaan sumber daya air di perkotaan.

### ABSTRACT

Indonesia as a tropical country faces significant impacts of *El Niño* and *La Niña* phenomena on rainfall patterns. This research aims to classify daily rainfall in major Indonesian cities such as DKI Jakarta, Surabaya, Medan, Makassar, and Bandung into three categories: no rain, moderate rain, and extreme rain, and identify climate conditions based on *El Niño*, *La Niña*, and Normal categories using the *Naive Bayes Classifier* algorithm. By applying the CRISP-DM (*Cross-Industry Standard Process for Data Mining*) method as a framework for processing daily rainfall datasets for the period January to December 2023, obtained from BMKG. The analysis results show that the *Naive Bayes Classifier* algorithm has high performance with 93.15% accuracy, 98% precision, 93% recall, and 94% *F1-score*. Further analysis showed that *El Niño* causes a significant decrease in rainfall, while *La Niña* increases extreme rainfall, especially in Medan. This study contributes to the development of rainfall classification models that can help the government to anticipate climate change impacts and improve the efficiency of water resources management in urban areas.

This is an open-access article under the [CC BY-SA](#) license



\*Penulis Koresponden

Email: [melyerlinda2021@student.unas.ac.id](mailto:melyerlinda2021@student.unas.ac.id)

---

Cara sitasi IEEE:

M. Erlinda, dan Andrianingsih, "Rainfall Classification Based on El-Niño and La-Niña Climate Phenomenon Using Naive Bayes Classifier Algorithm," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 1, pp. 239-251, Maret 2025. doi:10.30811/jaise.v5i1.6552

---

## 1. PENDAHULUAN

Indonesia merupakan negara kepulauan beriklim tropis yang memiliki dua musim utama, yaitu musim hujan dan musim kemarau. Selain itu, Indonesia merupakan negara dengan fluktuasi cuaca yang besar sepanjang tahun. Secara umum, musim kemarau berlangsung dari Juni hingga November, sementara musim hujan terjadi antara Desember hingga Mei. Namun, akibat perubahan iklim global, pola musim menjadi semakin sulit diprediksi [1]. Ketidakpastian dalam transisi antara musim kemarau dan musim hujan menjadi tantangan yang signifikan. Dalam sektor pertanian, ketidakpastian curah hujan mempengaruhi jadwal tanam dan panen. Sementara itu, kota-kota besar seperti DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung menghadapi tantangan besar akibat tingginya curah hujan, seperti banjir, kemacetan, dan kerusakan infrastruktur [2]. Banjir yang kerap melanda DKI Jakarta disebabkan oleh kombinasi antara tingginya curah hujan, buruknya sistem *drainase*, dan urbanisasi yang tidak terkendali. Selain itu, ketidakteraturan distribusi curah hujan yang tidak memengaruhi pengelolaan sumber daya air.

Cuaca dan iklim Indonesia sangat dipengaruhi oleh fenomena iklim global seperti *El Niño* dan *La Niña*, yang merupakan dua peristiwa alam dengan karakteristik berlawanan dan terjadi secara berkala di wilayah tropis Samudra Pasifik. Kedua fenomena ini merupakan bagian dari siklus besar, yaitu *El Niño-Southern Oscillation* (ENSO), dan berdampak besar pada pola distribusi curah hujan di berbagai wilayah di Indonesia. *El Niño* terjadi ketika suhu permukaan laut di bagian tengah dan timur Samudra Pasifik meningkat lebih dari  $0,5^{\circ}\text{C}$  di atas rata-rata normal, sedangkan *La Niña* terjadi saat suhu di wilayah yang sama menurun lebih dari  $-0,5^{\circ}\text{C}$  di bawah rata-rata normal [3].

Fenomena *El-Niño* biasanya mengakibatkan kondisi kekeringan di sejumlah wilayah di Indonesia, sering kali dikaitkan dengan musim kemarau yang lebih panjang dan potensi terjadinya kebakaran hutan. Sedangkan *La-Niña* dapat memicu peningkatan curah hujan ekstrem dengan mengakibatkan banjir dan tanah longsor. Fenomena cuaca yang ekstrem seperti ini memerlukan metode analisis yang akurat untuk mengklasifikasikan pola curah hujan harian, khususnya pada kota besar di Indonesia yang mempunyai kerentanan tinggi terhadap efek perubahan cuaca [4]. Fenomena iklim *El-Niño* dan *La-Niña* memiliki dampak besar terhadap dinamika atmosfer serta pola distribusi curah hujan di Indonesia. Sementara itu, *Indian Ocean Dipole* (IOD) merupakan fenomena *oseanografi* di Samudra Hindia yang ditandai oleh perbedaan suhu permukaan laut antara wilayah barat dan timur samudra [5].

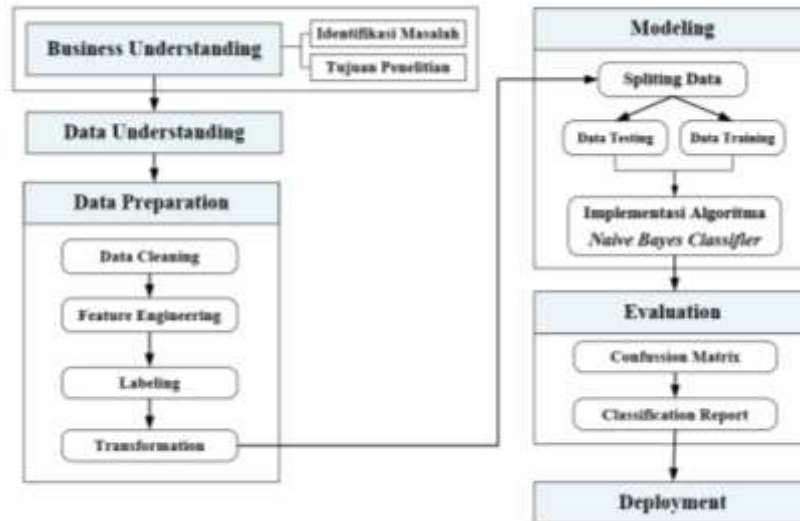
Dalam beberapa tahun terakhir, pemodelan dan prediksi cuaca berkembang pesat karena teknologi dan algoritma berbasis data. *Naïve Bayes Classifier* (NBC) digunakan dalam penelitian ini sebagai metode klasifikasi berbasis probabilitas dari *teorema Bayes*. Algoritma ini dikenal karena kesederhanaannya dan kemampuannya menangani data yang beragam, termasuk curah hujan harian [6]. Namun, penerapannya di kota-kota besar Indonesia seperti DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung menghadapi tantangan karena perbedaan karakteristik cuaca dan faktor geografis. Oleh karena itu, klasifikasi curah hujan harian menjadi penting untuk membantu pemerintah dan masyarakat dalam mengantisipasi dampak perubahan pola cuaca.

Penelitian ini bertujuan untuk mengaplikasikan algoritma *Naïve Bayes Classifier* dalam klasifikasi curah hujan harian di kota-kota besar Indonesia, yaitu DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung. Studi ini menganalisis pola curah hujan dengan mengelompokkan data ke dalam tiga kategori utama, yaitu tidak hujan, hujan sedang dan hujan ekstrem. Dataset historis curah hujan harian yang digunakan diperoleh dari BMKG, kemudian diklasifikasikan berdasarkan tiga kondisi iklim utama, yaitu *El Niño*, *La Niña*, dan Normal. Melalui analisis ini, penelitian diharapkan dapat mengembangkan model klasifikasi curah hujan yang lebih akurat, yang nantinya dapat digunakan untuk membantu pemerintah dalam mendukung mitigasi risiko bencana, seperti banjir akibat hujan ekstrem dan kekeringan akibat curah hujan rendah, di kota-kota besar Indonesia.

## 2. METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan pola curah hujan di kota-kota besar di Indonesia berdasarkan fenomena iklim, seperti *El-Niño* dan *La-Niña*. Dataset yang digunakan berasal dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) mencakup data curah hujan harian dari Bulan Januari hingga Desember 2023. Dalam penelitian ini, menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahap utama yaitu, *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Metode ini digunakan sebagai kerangka kerja untuk membimbing proses klasifikasi curah hujan

berdasarkan fenomena *El Niño* dan *La Niña* menggunakan *Naïve Bayes Classifier* [7]. Penelitian ini juga akan menghasilkan visualisasi data yang interaktif untuk membantu pemangku kepentingan, seperti pemerintah dan perencanaan kota, dalam memahami pola curah hujan untuk mitigasi risiko bencana banjir atau kekeringan. Pada bagian ini, akan dijelaskan metodologi yang digunakan serta tahapan penelitian secara rinci.



Gambar 1. Tahapan Penelitian

## 2.1 Business Understanding

Pada tahap ini, peneliti mengidentifikasi permasalahan yang akan diselesaikan serta memahami tujuan penelitian, yaitu klasifikasi curah hujan berdasarkan fenomena iklim *El-Niño*, *La-Niña*, dan kondisi Normal. Dengan tujuan penelitian ini, memberikan pola curah hujan untuk membantu pengambilan keputusan, misalnya dalam bidang pertanian, mitigasi bencana, atau pengelolaan sumber daya air. Dataset yang digunakan yaitu, data iklim harian di kota-kota besar di Indonesia seperti, DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung pada bulan Januari hingga Desember 2023.

Tabel 1. Struktur Data dalam Dataset

Kolom
Tanggal
Kode_wilayah
Wilayah
Stasiun
Temperatur_rata-rata
Kelembapan_rata-rata
Curah_hujan
Lamanya_penyinaran_matahari
Kecepatan_angin_rata-rata

## 2.2 Data Understanding

Tahap ini berfokus pada eksplorasi dan analisis awal dataset. Dataset iklim harian yang digunakan diperoleh dari *website* online BMKG. Dalam era digital saat ini, BMKG juga telah memanfaatkan teknologi informasi untuk menyediakan dataset melalui *platform online*, yang dapat diakses oleh publik maupun peneliti untuk keperluan analisis lebih lanjut [8]. Dataset yang digunakan mencakup parameter seperti tanggal, kode wilayah, wilayah, stasiun, temperatur rata-rata, kelembapan rata-rata, curah hujan, lamanya penyinaran matahari, dan kecepatan angin rata-rata. Dataset yang diperoleh dengan total 2190 yang terdiri dari 9 kolom. Pada tahap ini, dilakukan analisis deskriptif terhadap dataset untuk memahami distribusi nilai curah hujan. Visualisasi seperti grafik curah hujan bulanan per wilayah atau hubungan fenomena iklim dengan curah hujan juga dilakukan untuk mengetahui pola awal dalam dataset.

## 2.3 Data Preparation

Tahap ini merupakan proses persiapan data untuk analisis lebih lanjut dan pembangunan model, yang terdiri dari beberapa tahapan sebagai berikut:

### 1. *Cleaning Data*

Langkah awal untuk membersihkan dan menyiapkan dataset awal agar layak untuk di analisis dan pemodelan. Tahap pembersihan dilakukan untuk mengatasi masalah data duplikasi, menangani missing values menggunakan metode seperti imputasi *mean* atau *median* untuk data numerik. interpolasi untuk data deret waktu, atau menghapus baris atau kolom dengan data hilang yang signifikan.

### 2. *Feature Engineering*

Proses dalam *Machine Learning* untuk membuat, memodifikasi, atau memilih fitur (variabel) yang relevan dan berguna dari data mentah agar model dapat bekerja lebih baik. Data curah hujan dibuat fitur baru yaitu, “kategori curah hujan” berdasarkan kategori curah hujan, seperti tidak hujan, hujan sedang, hujan ekstrem, dan fitur “klasifikasi” sebagai kelas iklim seperti *El-Niño*, *La-Niña*, atau Normal.

Tabel 2. Kategori Curah Hujan

Curah Hujan (mm)	Kategori Curah Hujan
Curah hujan 0 mm/hari	Tidak Hujan
Curah hujan $\leq 50$ mm/hari	Hujan Sedang
Curah hujan $> 50$ mm/hari	Hujan Ekstrem

Kategori curah hujan diklasifikasikan ke dalam tiga kelompok berdasarkan nilai curah hujan yang terukur. Tidak Hujan ditandai dengan curah hujan 0 mm, menunjukkan tidak adanya hujan signifikan. Hujan mencakup curah hujan kurang dari 50 mm, yang meliputi hujan ringan hingga sedang. Sementara itu, Hujan Ekstrem terjadi ketika curah hujan lebih dari 50 mm, yang dapat mengindikasikan hujan lebat atau ekstrem dengan potensi menyebabkan banjir.

Tabel 3. Variabel *El-Niño*, *La-Niña*, dan Normal

Fenomena Iklim	Rentan Curah Hujan (mm)	Klasifikasi	Keterangan
<i>El-Niño</i>	$< (\text{Rata-rata} - \text{Standar Deviasi})$	Rendah	Curah hujan lebih rendah dari rata-rata, sering terjadi selama fenomena El Niño.
<i>La-Niña</i>	$> (\text{Rata-rata} + \text{Standar Deviasi})$	Ekstrem	Curah hujan lebih ekstrem dari rata-rata, sering terjadi selama fenomena La Niña.
Normal	$(\text{Rata-rata} - \text{Standar Deviasi}) \leq \text{Curah Hujan} \leq (\text{Rata-rata} + \text{Standar Deviasi})$	Normal	Curah hujan berada dalam rentang rata-rata, tidak terlalu rendah atau tinggi.

Perubahan curah hujan di Indonesia dipengaruhi secara signifikan oleh kondisi atmosfer global, seperti fenomena *El-Niño* dan *La-Niña*. Oleh karena itu, mengetahui pola curah hujan menjadi sangat krusial, terutama di kota-kota besar yang lebih rentan terhadap dampak perubahan iklim [9]. Untuk data lebih mudah disajikan, hasil klasifikasi akan dikelompokkan ke dalam tiga kategori fenomena iklim, yaitu *El Niño*, *La Niña*, dan Normal. Batas nilai untuk setiap kategori akan ditentukan menggunakan metode standar deviasi.

### 3. *Labeling*

Proses *labeling* dalam kode yang diberikan bertujuan untuk mengonversi nilai-nilai kategorikal menjadi bentuk numerik sehingga dapat digunakan dalam algoritma *Machine Learning*. Pada tahap ini, Konversi teks menjadi angka menggunakan *LabelEncoder*. Fungsi melakukan *encoding*, yang secara otomatis mengubah nilai kategorikal menjadi angka berdasarkan urutan alfabetis atau urutan data asli. Kolom dengan nilai kategorikal (teks) yaitu, “Wilayah”, “Stasiun”, “Kategori\_Curah\_Hujan”, “Klasifikasi”. Selain itu pada proses ini melakukan konversi kolom “Tanggal” menjadi tipe *datetime*.

### 4. *Data Transformation*

Proses mengubah data menjadi format yang sesuai untuk algoritma *Naive Bayes Classifier*. *Feature scaling*, digunakan untuk normalisasi atau standarisasi diterapkan untuk memastikan variabel dalam dataset memiliki skala yang sama. Data kemudian dinormalisasi agar memiliki skala seragam, misalnya menggunakan *Min-Max Scaling* atau standarisasi, sehingga membantu mengurangi bias dalam algoritma.

## 2.4 Modeling Algoritma Naïve Bayes Classifier

Tahap ini melibatkan Pembangunan model klasifikasi menggunakan algoritma *Naïve Bayes Classifier*. Algoritma ini dipilih karena kesederhanaannya, efisiensi komputasi, dan kemampuannya untuk bekerja dengan baik pada dataset. Tahap pertama, melakukan *splitting data* menjadi *data training* dan *data testing*, dengan proporsi umum 80% untuk *data training* dan 20% untuk *data testing* dengan tujuan mengevaluasi kinerja model secara objektif. Pada tahap ini model dilatih menggunakan data training untuk mempelajari hubungan antara curah hujan dengan indikator fenomena iklim. Selain itu, proses evaluasi model dilakukan dengan menggunakan metrik seperti *accuracy*, *precision*, *recall*, dan *F1-Score* untuk memastikan bahwa model dapat mengklasifikasikan data curah hujan dengan baik ke dalam kategori *El-Niño*, *La-Niña*, dan Normal. Berikut formula yang digunakan dalam *Naive Bayes Classifier* (NBC):

$$P(Y | X) = \frac{P(X | Y) P(Y)}{P(X)} \quad (1)$$

Keterangan:

- X : Data yang kelasnya belum diketahui
- Y : Hipotesis bahwa data X termasuk dalam kelas tertentu
- $P(Y|X)$  : Probabilitas dari hipotesis Y, berdasarkan kondisi X
- $P(Y)$  : Probabilitas dari hipotesis Y
- $P(X|Y)$  : Probabilitas dari X, berdasarkan kondisi hipotesis Y benar
- $P(X)$  : Probabilitas dari X



Gambar 2. Tahapan Algoritma *Naïve Bayes Classifier*

Berdasarkan gambar 2 merupakan alur proses penerapan Algoritma *Naïve Bayes Classifier* dalam analisis data. Secara keseluruhan, tahapan ini menggambarkan tahapan dalam *Machine Learning* menggunakan *Naive Bayes Classifier*, mulai dari pemisahan data, penerapan model, hingga evaluasi kinerjanya [10].

## 2.5 Evaluation

Tahap evaluasi digunakan untuk menilai kinerja model secara menyeluruh. Model diuji menggunakan *data testing* yang terpisah dari *data training*. Jika hasil evaluasi menunjukkan bahwa model tidak cukup akurat, maka dilakukan tuning *hyperparameter*, eksplorasi fitur tambahan, atau bahkan pemilihan ulang algoritma. Visualisasi hasil klasifikasi, seperti *confusion matrix*, digunakan untuk memeriksa kesalahan prediksi. Selain itu, analisis tambahan dilakukan untuk memahami bagaimana model berperilaku terhadap wilayah tertentu atau kondisi iklim tertentu

## 2.6 Deployment

Langkah akhir dalam proses membuat model *Machine Learning* yang bertujuan untuk menerapkan model yang telah dibangun ke dalam lingkungan nyata agar dapat digunakan oleh pengguna atau sistem lain. Pada tahap ini, hasil analisis dan model klasifikasi curah hujan berdasarkan iklim fenomena *El Niño*, *La Niña* dalam bentuk *dashboard* dengan menggunakan *framework Streamlit*.

## 3. HASIL DAN PEMBAHASAN

Setiap tahapan dalam penelitian ini dijelaskan secara ringkas dan jelas untuk memastikan bahwa langkah-langkah yang dilakukan selaras dengan tujuan yang telah ditetapkan. Penelitian ini terdiri dari beberapa tahap utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Berikut ini adalah penjelasan detail mengenai implementasi dari masing-masing tahapan tersebut.

### 3.1 Business Understanding

Mendefinisikan tujuan utama penelitian, yaitu untuk mengklasifikasikan curah hujan berdasarkan fenomena iklim *El-Niño* dan *La-Niña* dengan menggunakan algoritma *Naive Bayes Classifier*. Selain itu, dilakukan identifikasi kebutuhan, seperti jenis data yang diperlukan dan parameter yang relevan untuk analisis. Data diperoleh dari *situs web* Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), dengan menggunakan data iklim harian dari masing-masing stasiun. di DKI Jakarta, Surabaya, Medan, Makassar, dan Bandung. Dari data yang sudah diperoleh dari *website* data online BMKG berdasarkan indikator temperatur<sub>minimum</sub>,

temperatur\_maksimum, temperatur\_rata-rata, kelembapan\_rata-rata, curah\_hujan, kecepatan\_angin\_maksimum, dan kecepatan\_angin\_rata-rata. Selanjutnya data disimpan dengan format CSV (*Comma Separated Value*).

### 3.2 Data Understanding

Tahap ini, dataset dianalisis untuk memahami struktur, pola, dan distribusi data. Statistik deskriptif digunakan untuk menggambarkan karakteristik data, serta dilakukan eksplorasi terhadap hubungan antar variabel menggunakan visualisasi seperti *scatter plot* atau *heatmap*. Tahap yang pertama yaitu import file dataset ke dalam *Google Colaboratory*. Setelah itu masukan import library python yang akan digunakan untuk analisis, *import numpy, pandas, seaborn, matplotlib.pyplot, sklearn, sklearn.naive\_bayes*.

```

Informasi Dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2190 entries, 0 to 2189
Data columns (total 9 columns):
 #   Column                               Non-Null Count  Dtype
---  ---                               ---
 0   Tanggal                             2190 non-null   object
 1   Kode_wilayah                        2190 non-null   float64
 2   Wilayah                             2190 non-null   object
 3   Stasiun                             2190 non-null   object
 4   Temperatur_rata-rata                2174 non-null   float64
 5   Kelembapan_rata-rata                2174 non-null   float64
 6   Curah_hujan                         2190 non-null   float64
 7   Lamanya_penyinaran_matahari         2173 non-null   float64
 8   Kecepatan_angin_rata-rata           2183 non-null   float64
dtypes: float64(6), object(3)
memory usage: 154.1+ KB
    
```

Gambar 3. Informasi Dataset

Berdasarkan gambar 3 tersebut menunjukkan hasil dari perintah `data.info()` dalam `pandas`, yang memberikan ringkasan informasi tentang dataset yang sedang digunakan. Dataset tersebut memiliki 2190 baris dengan indeks mulai dari 0 hingga 2189, dataset terdiri dari 9 kolom. Menunjukkan bahwa terdapat tipe data pada setiap kolom yaitu, *object*, dan *float*.

### 3.3 Data Preparation

Tahap ini merupakan proses persiapan data untuk analisis lebih lanjut dan pembangunan model, yang terdiri dari beberapa tahapan sebagai berikut:

#### 1. *Cleaning Data*

Tahap pertama dalam *cleaning data*, yaitu melakukan proses pembersihan data dilakukan untuk mengatasi masalah data duplicated, missing values, imputasi nilai yang hilang dan melakukan klasifikasi berdasarkan kategori curah hujan, seperti tidak hujan, hujan sedang, dan hujan ekstrem, serta klasifikasi fenomena iklim seperti *El-Niño*, *La-Niña*, atau Normal.

Tabel 4. Cek *Missing Value*

Kolom	Precentage (null)
Tanggal	0
Kode_wilayah	0
Wilayah	0
Stasiun	0
Temperatur_rata-rata	16
Kelembapan_rata-rata	16
Curah_hujan	0
Lamanya_penyinaran_matahari	17
Kecepatan_angin_rata-rata	7

Tabel 5. Hasil Pemrosesan *Missing Value*

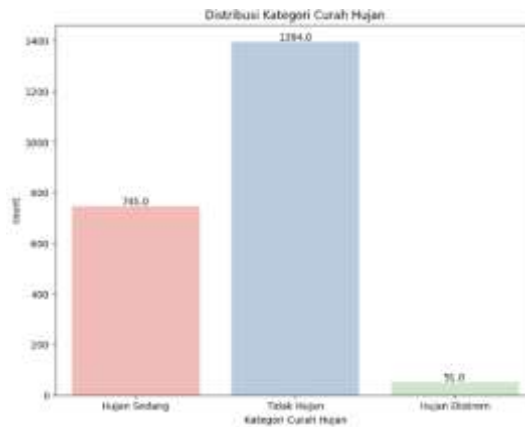
Kolom	Precentage (null)
-------	-------------------

Tanggal	0
Kode_wilayah	0
Wilayah	0
Stasiun	0
Temperatur_rata-rata	0
Kelembapan_rata-rata	0
Curah_hujan	0
Lamanya_penyinaran_matahari	0
Kecepatan_angin_rata-rata	0

Berdasarkan tabel 4 dan 5 menunjukan bahwa hasil cek *missing value* dan Imputasi pada dataset sudah berhasil dah tidak ada lagi nilai kosong pada dataset. Untuk menangani *missing value* pada dataset dilakukannya imputasi nilai hilang dengan *SimpleImputer* dengan *strategi median*. Tahap selanjutnya melakukan pengecekan data *duplicate*. Pada pemrosesan cek data *duplicate*, menunjukan bahwa tidak terdapat data *duplicate* pada dataset. Tahap selanjutnya mengatur kategori curah hujan berdasarkan kategori tidak hujan hujan sedang, hujan ekstrem, serta klasifikasi fenomena iklim seperti *El-Niño*, *La-Niña*, atau Normal.

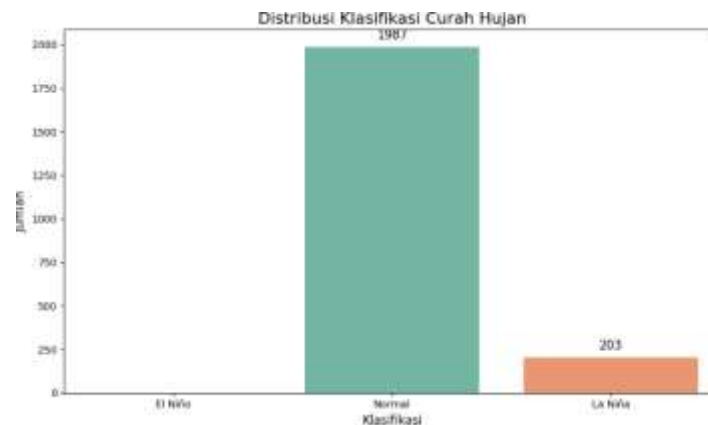
2. *Feature Engineering*

Proses menambahkan kolom baru yang bertujuan untuk mengelompokkan data curah hujan ke dalam beberapa kategori berdasarkan nilai curah hujan. Fungsi *categorize\_rainfall* digunakan untuk menentukan “kategori curah hujan”. Selanjutnya kolom baru yang ditambahkan yaitu, “Klasifikasi” ke dataset berdasarkan kriteria yang dihitung dari data curah hujan berdasarkan statistik wilayah.



Gambar 4. Distribusi Kategori Curah Hujan

Berdasarkan gambar 4 pada distribusi kategori curah hujan dibagi menjadi tiga, yaitu tidak hujan, hujan sedang, dan hujan ekstrem. Pada grafik tersebut terlihat bahwa kategori curah hujan “tidak hujan” memiliki grafik tertinggi dibandingkan dengan kategori “hujan sedang” dan “hujan ekstrem”. Pada klasifikasi ini terlihat bahwa kategori “Tidak Hujan” lebih tinggi. Tidak Hujan memiliki jumlah tertinggi dengan 1.394 kejadian, menunjukkan bahwa sebagian besar hari dalam data yang digunakan tidak mengalami hujan dan hubungan variabel curah hujan dengan variael lainnya.



Gambar 5. Distribusi Klasifikasi Curah Hujan

Berdasarkan gambar 5 merupakan grafik diagram batang yang menggambarkan distribusi curah hujan berdasarkan tiga kategori utama, yaitu *El Niño*, *La Niña*, dan Normal. Berdasarkan grafik, terlihat bahwa kategori Normal memiliki jumlah kejadian tertinggi, mencapai 1987 kejadian yang menunjukkan dominasi kondisi curah hujan normal dalam data yang dianalisis. Sebaliknya, kategori *La Niña* hanya mencatat hanya 203 kejadian, sementara kategori *El Niño* memiliki jumlah yang sangat kecil, hampir mendekati nol. Hal ini menunjukkan bahwa kondisi ekstrem seperti *El Niño* dan *La Niña* jarang terjadi dibandingkan dengan kondisi normal.

3. Label Encoding

Pada data dalam penelitian atau analisis, terutama ketika data memiliki atribut kategorikal yang tidak dapat digunakan langsung dalam algoritma machine learning, maka dataset pada, kolom seperti "Wilayah", "Stasiun", "Kategori\_Curah\_Hujan," dan "Klasifikasi" sehingga tidak dapat diolah langsung oleh algoritma yang membutuhkan data numerik. Label Encoding mengubah kategori tersebut menjadi angka, sehingga algoritma dapat memproses data dengan lebih efisien.

Hari	Bulan	Tanggal	Kode_wilayah	Wilayah	Musim	Temperatur_rata-rata	Bekas hujan_rata-rata	Curah_hujan	Luas_pemukiman_malahari	Kecepatan_angin_rata-rata	Kategori_Curah_Hujan	Klasifikasi
0	1	1	2000-08-01	31.71	0	0	36.0	87.0	14.0	0.0	1.0	0
1	2	1	2000-08-02	31.71	0	0	37.0	81.0	14.0	0.7	2.0	0
2	3	1	2000-08-03	31.71	0	0	36.0	82.0	0.0	0.0	1.0	0
3	4	1	2000-08-04	31.71	0	0	36.4	86.0	3.4	0.0	1.0	0
4	5	1	2000-08-05	31.71	0	0	37.4	83.0	36.0	0.0	2.0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
1986	12	12	2000-12-07	12.71	3	3	38.4	84.0	0.0	2.1	1.0	1
1986	12	12	2000-12-08	12.71	3	3	39.0	84.0	3.0	3.7	2.0	0
1987	01	12	2000-12-09	12.71	3	3	37.0	91.0	3.0	2.6	1.0	0
1986	12	12	2000-12-06	12.71	3	3	37.7	87.0	0.0	0.0	1.0	0
1986	12	12	2000-12-01	12.71	3	3	39.0	86.0	19.7	1.8	2.0	0

Gambar 6. Hasil Label Encoding

4. Data Transformation

Pada tahap *data transformation* menggunakan *Feature scaling*, untuk normalisasi atau standarisasi diterapkan untuk memastikan variable dalam dataset memiliki skala yang sama. Data kemudian dinormalisasi agar memiliki skala seragam, misalnya menggunakan *Min-Max Scaling* atau standarisasi.

Tanggal	Kode_wilayah	Wilayah	Musim	Temperatur_rata-rata	Bekas hujan_rata-rata	Curah_hujan	Luas_pemukiman_malahari	Kecepatan_angin_rata-rata	Kategori_Curah_Hujan	Klasifikasi
0	0	31.71	0	0	0.780769	0.960209	14.0	0.0	0.766667	0
1	1	31.71	0	0	0.885113	0.643708	14.0	0.7	0.333333	0
2	2	31.71	0	0	0.807209	0.824787	0.0	0.0	0.766667	0
3	3	31.71	0	0	0.795181	0.693333	2.4	0.0	0.766667	0
4	4	31.71	0	0	0.882091	0.864883	36.0	0.0	0.333333	0
...	...	...	...	...	...	...	...	...	...	...
1986	361	12.71	3	3	0.895422	0.670889	0.0	2.1	0.766667	1
1986	361	12.71	3	3	0.961446	0.670889	3.0	2.7	0.333333	0
1987	362	12.71	3	3	0.813233	0.647917	4.6	2.6	0.766667	0
1986	361	12.71	3	3	0.814117	0.660289	4.2	0.6	0.466667	0
1986	361	12.71	3	3	0.882436	0.833333	19.7	1.8	0.333333	0

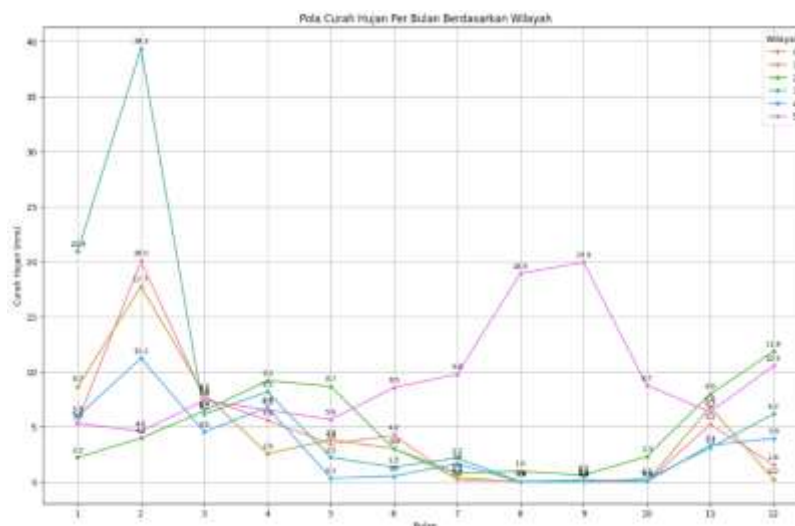
Gambar 7. Hasil Feature scaling

3.4 Modeling



Gambar 8. Grafik Pola Curah Hujan Berdasarkan Wilayah

Berdasarkan gambar 8 merupakan grafik pola curah hujan berdasarkan wilayah. Grafik tersebut menunjukkan pola curah hujan berdasarkan wilayah dengan sumbu horizontal mewakili kode wilayah dan sumbu vertikal menunjukkan curah hujan dalam milimeter (mm). Dari data yang ditampilkan, wilayah dengan curah hujan tertinggi dengan kode wilayah 12.71 (Medan) dengan 9.4 mm, sementara wilayah dengan curah hujan terendah dengan kode wilayah 35.78 (Surabaya) dengan 3.3 mm. Setelah mengalami penurunan drastis di kode wilayah 31.71 (Jakarta Pusat) dengan 3.3 mm, curah hujan sedikit meningkat di kode wilayah 31.72 (Jakarta Pusat) dengan 4.2 mm dan kode wilayah 32.73 (Bandung) dengan 4.8 mm. Namun, kembali mengalami penurunan ke titik terendah di kode wilayah 35.78 (Surabaya) sebelum akhirnya meningkat signifikan di kode wilayah 73.71 (Makassar) dengan 7.3 mm. Pola ini menunjukkan fluktuasi curah hujan yang cukup besar di berbagai wilayah, yang dapat menjadi dasar analisis lebih lanjut dalam perencanaan lingkungan, pertanian, atau mitigasi bencana terkait curah hujan.



Gambar 9. Grafik Pola Curah Hujan Per Bulan

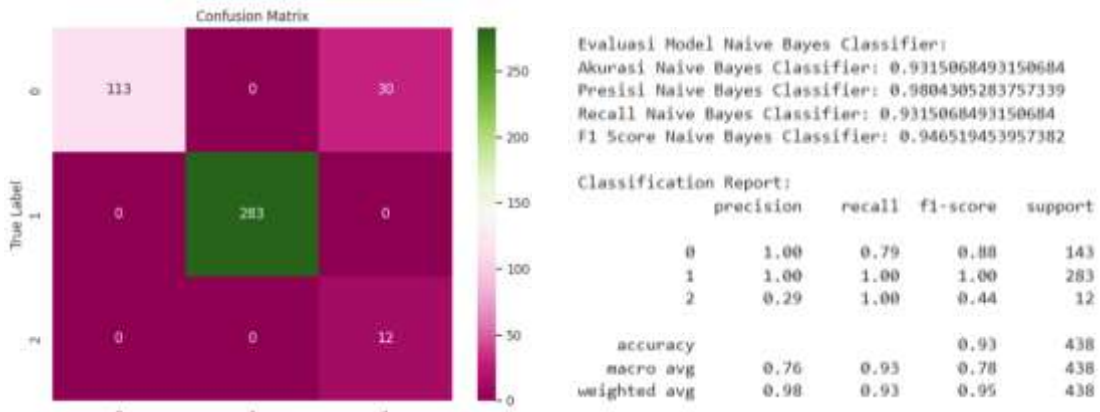
Berdasarkan gambar 9 grafik terlihat bahwa kode wilayah 73.71 (Makassar) mengalami lonjakan curah hujan yang sangat tinggi pada bulan Februari dengan nilai mencapai 39.3 mm, yang merupakan curah hujan tertinggi di antara semua wilayah dan bulan. Kemudian kode wilayah 12.71 (Medan) juga mengalami peningkatan signifikan pada bulan Agustus dan September dengan curah hujan mencapai 18.9 mm dan 19.9 mm. Sebaliknya, kode wilayah lain seperti 31.71 (Jakarta Pusat), 31.72 (Jakarta Utara), 32.73 (Bandung), dan 35.78 (Surabaya) memiliki pola curah hujan yang lebih stabil, dengan nilai yang lebih rendah dibandingkan dua wilayah tersebut. Secara umum, sebagian besar wilayah mengalami curah hujan rendah pada pertengahan tahun, terutama antara bulan Mei hingga September, sebelum kembali meningkat menjelang akhir tahun. Beberapa wilayah menunjukkan pola fluktuatif yang cukup signifikan, terutama 12.71 (Medan) dan 73.71 (Makassar), yang mungkin disebabkan oleh faktor geografis atau iklim mikro di daerah tersebut. Tren ini menunjukkan bahwa musim hujan lebih dominan terjadi pada awal dan akhir tahun, sementara pertengahan tahun cenderung mengalami curah hujan yang lebih rendah. Pola ini dapat digunakan untuk perencanaan mitigasi banjir, pertanian, serta pengelolaan sumber daya air di masing-masing wilayah.

### 3.5 Evaluasi

Klasifikasi curah hujan berdasarkan fenomena iklim *El-niño* dan *La-niña* menggunakan algoritma *Naive Bayes Classifier*. Implementasi algoritma dilakukan menggunakan bahasa pemrograman *Python* di lingkungan *Google Colab*, dimana hasil pengujian dari model meliputi *classification report*, *recall*, *precision*, *f1-score* dan *accuracy* model dalam memprediksi curah hujan. Pada pengujian dataset, hasil klasifikasi divisualisasikan menggunakan *Confusion Matrix* yang menggambarkan perbandingan antara hasil prediksi model dengan klasifikasi sebenarnya.

Secara keseluruhan, model *Naive Bayes Classifier* memiliki tingkat akurasi yang cukup tinggi, yaitu 93.15% dari prediksi yang dihasilkan sesuai dengan data aktual, yang menunjukkan bahwa sebagian besar prediksi model sesuai dengan data sebenarnya. Dengan nilai *precision* 98%, *recall* 93% dan *F1-score* mencapai 94%. Hasil evaluasi ini menunjukkan bahwa model *Naive Bayes Classifier* bekerja sangat baik dalam mendeteksi kelas "*La Niña*", namun masih memerlukan peningkatan untuk menangani prediksi kelas "*El-Niño*" yang memiliki performa rendah. Untuk meningkatkan performa model, beberapa langkah yang dapat dilakukan

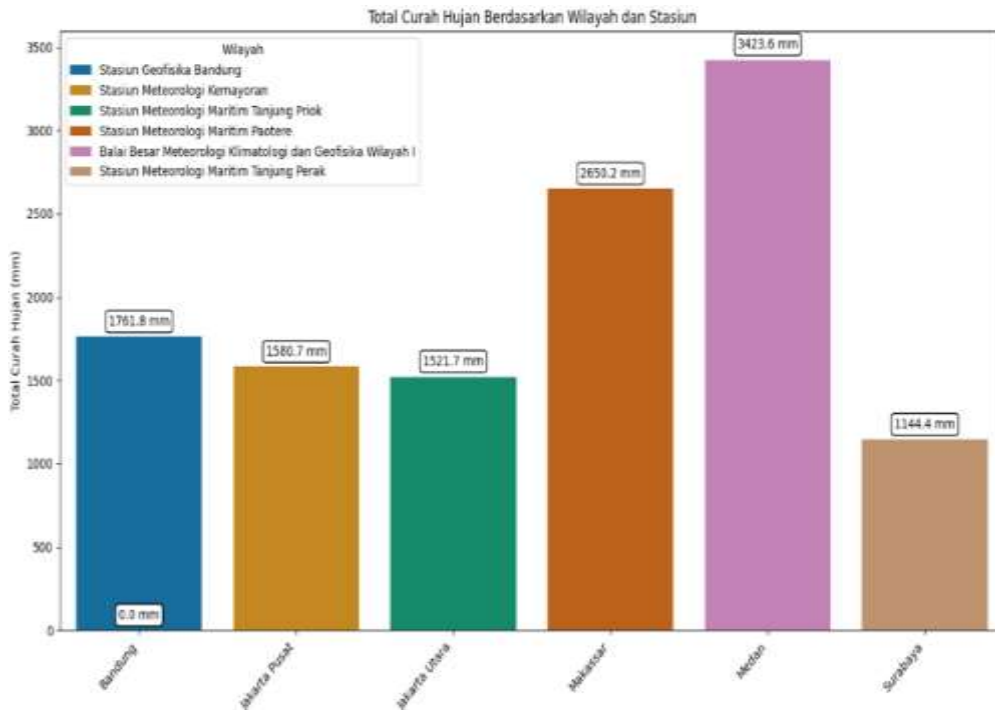
adalah dengan meningkatkan jumlah data pelatihan, melakukan optimasi fitur, atau mencoba metode klasifikasi lain yang lebih kompleks.



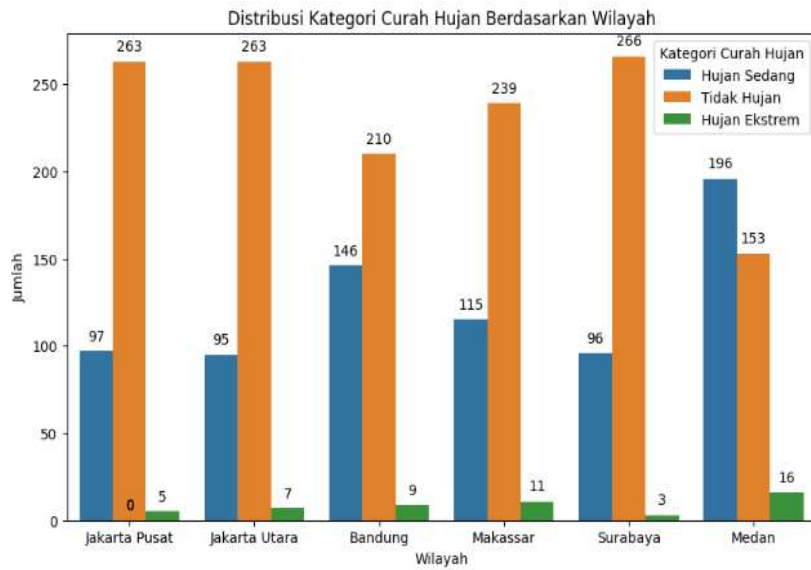
Gambar 10. Confusion Matrix

### 3.6 Hasil Analisis

Berdasarkan gambar 11 menunjukkan curah hujan (mm) berdasarkan wilayah dan stasiun pengamatan. Terlihat bahwa wilayah Medan mencatat curah hujan tertinggi dibandingkan wilayah lainnya, dengan kontribusi data berasal dari Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah I. Wilayah Makassar juga memiliki curah hujan yang relatif tinggi, yang diamati oleh Stasiun Meteorologi Maritim Paotere. Sementara itu, wilayah seperti Jakarta Pusat dan Jakarta Utara mencatat curah hujan yang lebih rendah, dengan data yang berasal dari Stasiun Meteorologi Kemayoran dan Stasiun Meteorologi Maritim Tanjung Priok. Wilayah Surabaya memiliki curah hujan yang rendah, dengan data dari Stasiun Meteorologi Tanjung Perak, sedangkan Bandung memiliki curah hujan moderat yang diamati oleh Stasiun Geofisika Bandung.

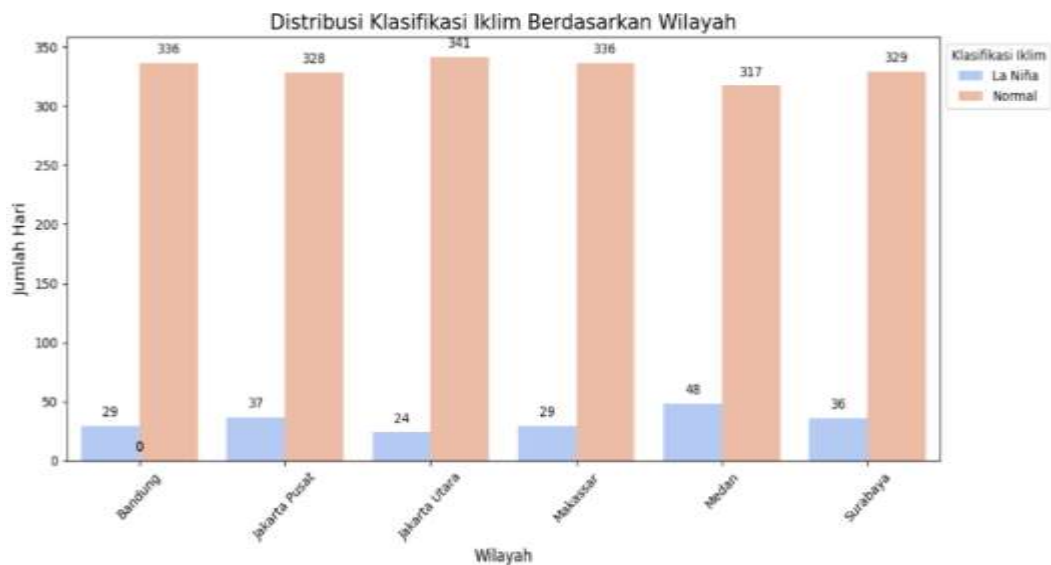


Gambar 11. Grafik Total CH Berdasarkan Wilayah dan Stasiun



Gambar 12. Grafik Distribusi CH Berdasarkan Wilayah

Berdasarkan gambar 12 dapat dilihat bahwa grafik kategori Tidak Hujan mendominasi di semua wilayah, dengan jumlah kejadian tertinggi, khususnya di Jakarta Pusat, Jakarta Utara, dan Surabaya, yang mencatat lebih dari 250 kejadian. Sementara itu, kategori Hujan Sedang memiliki jumlah yang cukup signifikan di beberapa wilayah, terutama Bandung, Medan, dan Makassar, meskipun tetap lebih rendah dibandingkan kategori Tidak Hujan. Sebaliknya, kategori Hujan Ekstrem hanya memiliki jumlah kejadian yang sangat kecil di seluruh wilayah, dengan angka yang relatif seragam dan mendekati nol. Kesimpulannya, kondisi tanpa hujan lebih sering terjadi dibandingkan kondisi hujan sedang atau ekstrem di wilayah-wilayah yang dianalisis. Informasi ini memberikan gambaran penting mengenai pola curah hujan yang dapat dimanfaatkan untuk memahami karakteristik iklim lokal serta membantu dalam perencanaan mitigasi risiko terkait curah hujan.



Gambar 13. Grafik Distribusi Klasifikasi Berdasarkan Wilayah

Berdasarkan gambar 13 menggambarkan grafik distribusi klasifikasi iklim berdasarkan wilayah untuk dua jenis iklim, yaitu *La Niña* dan Normal. Hasil analisis menunjukkan bahwa jumlah hari dengan klasifikasi Normal (diwakili oleh warna oranye) secara signifikan lebih tinggi dibandingkan dengan hari-hari beriklim *La Niña* (ditandai dengan warna biru) di semua wilayah. Secara umum, jumlah hari Normal hampir mencapai 350 hari, sedangkan hari *La Niña* hanya muncul dalam jumlah yang jauh lebih kecil. Dari grafik di atas, dapat disimpulkan bahwa iklim Normal lebih dominan di setiap wilayah, sementara fenomena *La Niña* terjadi dalam

durasi yang lebih terbatas. Hal ini memberikan gambaran tentang pola kestabilan iklim di berbagai wilayah yang dianalisis.

### 3.7 Deployment

Tahap *deployment*, penelitian ini memanfaatkan *Streamlit*, sebuah *framework Python* yang dirancang untuk mempermudah pembuatan aplikasi *web* berbasis *data science* dan *Machine Learning*. Tujuan dari *deployment* ini adalah menyajikan hasil klasifikasi curah hujan berdasarkan fenomena iklim *El-Niño* dan *La-Niña* menggunakan algoritma *Naïve Bayes Classifier* dalam bentuk *user interface* (UI) yang interaktif dan mudah digunakan.



Gambar 14, Home Page Dashboard

Berdasarkan gambar 4.14 menunjukkan tampilan *design home page* dari *dashboard* untuk mengklasifikasikan data curah hujan berdasarkan fenomena iklim *El-Niño* dan *La-Niña* dengan algoritma *Naïve Bayes Classifier*. Pada bagian kiri *dashboard* terdapat menu navigasi yang terdiri dari beberapa bagian, yaitu menu *home*, yang berisi pengantar mengenai tujuan dan lokasi penelitian. menu *data overview* yang memberikan ringkasan dataset yang digunakan, menu *classification* sebagai modul utama untuk melakukan klasifikasi curah hujan berdasarkan input pengguna, serta *modeling* yang menjelaskan hasil dari metode pemodelan, termasuk algoritma yang digunakan.

## 4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa pola curah hujan harian di kota-kota besar Indonesia dapat diklasifikasikan ke dalam tiga kategori utama, yaitu *El-Niño*, *La-Niña*, dan Normal. Hasil klasifikasi menunjukkan bahwa kota Jakarta Pusat, Jakarta Utara, Bandung, Surabaya, dan Makassar mengalami kondisi Normal, sedangkan kota Medan memiliki jumlah *La-Niña* tertinggi dan Jakarta Utara paling sedikit. Hasil penelitian menunjukkan bahwa *La-Niña* meningkatkan curah hujan, sementara *El-Niño* mengurangi curah hujan, berpotensi menyebabkan kekeringan. Model *Naïve Bayes Classifier* digunakan dalam klasifikasi ini dan mencapai akurasi 93.15%, dengan *precision* 98%, *recall* 93%, dan *F1-score* 94%. Secara keseluruhan, penelitian ini memberikan wawasan tentang pola hujan di kota-kota besar serta dampak fenomena iklim global, yang dapat digunakan untuk mitigasi bencana dan pengelolaan sumber daya air.

## UCAPAN TERIMAKASIH

Penulis menyampaikan apresiasi yang mendalam kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atas penyediaan data yang sangat berperan dalam mendukung pelaksanaan penelitian ini. Ucapan terima kasih juga ditujukan kepada Universitas Nasional (UNAS) atas dukungan akademik serta bimbingan yang diberikan selama proses penelitian. Selain itu, penulis menghargai setiap bantuan dan kontribusi dari berbagai pihak yang turut berperan dalam menyelesaikan penelitian ini.

## REFERENSI

- [1] A. W. C. Isnatul Mahmuda, "Jurnal ilmiah sinteks issn : 1907-2007 e-issn : 2579-7115," vol. 10, no. 2, pp. 10–16, 2024.
- [2] M. Alviriza Ramadhan, F. Tri Anggraeny, and C. Aji Putra, "Klasifikasi Curah Hujan Harian Menggunakan Metode K-Nearest Neighbor," *JATI (Jurnal Mhs. Tek. Inform.,* vol. 8, no. 3, pp. 3863–3869, 2024.
- [3] B. Yuniasih, W. N. Harahap, and D. A. S. Wardana, "Anomali Iklim El Nino dan La Nina di Indonesia pada 2013-2022," *AGROISTA J. Agroteknologi,* vol. 6, no. 2, pp. 136–143, 2023.
- [4] M. Ariska, H. Akhsan, M. Muslim, M. Romadoni, and F. S. Putriyani, "Prediksi Perubahan Iklim Ekstrem di Kota Palembang dan Kaitannya dengan Fenomena El Niño-Southern Oscillation (ENSO) Berbasis Machine Learning," *JIPFRI (Jurnal Inov. Pendidik. Fis. dan Ris. Ilmiah),* vol. 6, no. 2, pp. 79–86, 2022.
- [5] J. Ina, Ruminta, B. H. K. Tjasyono, A. L., and B. Harijono Sriworo, "Pengaruh El Niño, La Niña Dan Indian Ocean Dipole Terhadap Curah Hujan Pentad Di Wilayah Indonesia," *Encycl. Environ. Chang.,* pp. 168–177, 2014.
- [6] H. R. Burhani, I. Fitri, and A. Andrianingsih, "Perbandingan Naïve bayes dan Certainty factor pada Sistem Pakar Untuk

- Mendiagnosa Dini Penyakit Glaukoma,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 5, no. 3, p. 291, 2020.
- [7] S. Santiastry, W. Apriandari, T. Informatika, U. M. Sukabumi, K. Sukabumi, N. Bayes, T. B. Inggris, and U. M. Sukabumi, “PENERAPAN ALGORITMA NAIVE BAYES DAN METODE CRISP-DM DALAM,” vol. 8, no. 5, pp. 10432–10439, 2024.
- [8] A. S. Agung, A. A. Fauzi, A. A. Nur Risal, and F. Adiba, “Implementasi Teknik Data Mining terhadap Klasifikasi Data Prediksi Curah Hujan BMKG Di Sulawesi Selatan,” *J. Tekno Insentif*, vol. 17, no. 1, pp. 22–23, 2023.
- [9] A. T. Z. Irma Nurmaulida, Aswan S. Sunge, “Penggunaan Naïve Bayes dalam Implementasi Prediksi Tingkat Curah Hujan,” vol. 8, no. 3, pp. 3149–3157, 2023.
- [10] M. Fauzi, R. Mahendra, N. Lutvi, F. Sains, U. Muhammadiyah, S. Kampus, J. Raya, G. No, K. Candi, K. Sidoarjo, and J. Timur, “Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine,” *J. Ilm. Komputasi*, vol. 23, no. 1, pp. 45–50, 2024.