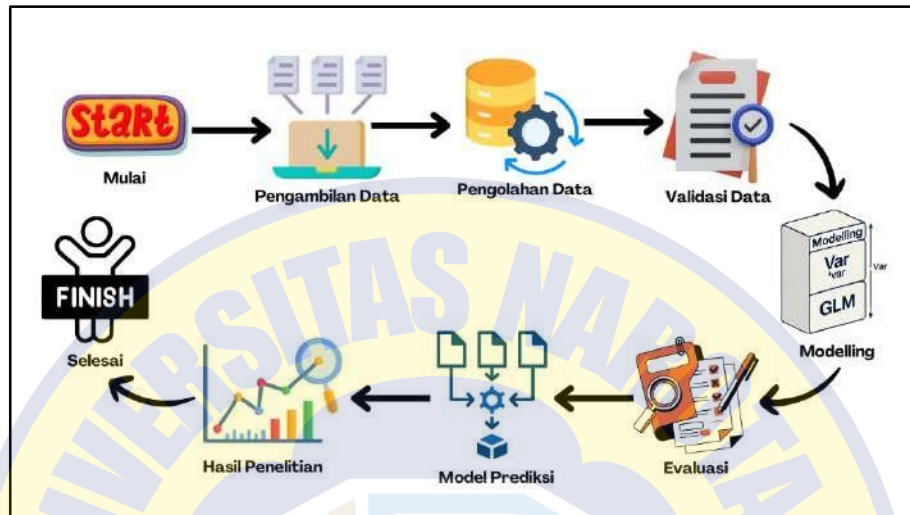


## BAB III

### METODOLOGI PENELITIAN



Gambar 3.1 Alur Metodologi Penelitian

Metode Penelitian ini menggunakan Gambar 3.1 sebagai ilustrasi alur proses yang dilakukan dalam penelitian. Langkah-langkahnya adalah sebagai berikut:

1. Mulai (Start)

Penelitian dimulai dengan perencanaan yang matang untuk memastikan setiap langkah berjalan sistematis dan efisien. Dalam tahap ini, dilakukan identifikasi masalah penelitian, tujuan yang ingin dicapai, serta metode yang akan digunakan. Peneliti menyusun jadwal kegiatan, menentukan alat dan sumber daya yang dibutuhkan, serta mempelajari literatur relevan untuk memperkuat landasan teori. Perencanaan ini penting untuk memastikan bahwa setiap tahap saling terhubung dan mendukung tujuan akhir penelitian. Dengan pendekatan yang terstruktur, risiko kesalahan dalam eksekusi dapat diminimalkan.

2. Pengambilan Data

Tahap ini melibatkan pengumpulan data dari BMKG, lembaga terpercaya yang menyediakan informasi meteorologi dan klimatologi. Data yang dikumpulkan meliputi curah hujan, suhu udara, kelembapan, lama penyinaran matahari, serta kecepatan angin. Data ini dipilih karena memiliki korelasi tinggi terhadap pola curah hujan. Pengambilan data dilakukan dengan memastikan keakuratan dan kelengkapannya, termasuk rentang

waktu yang relevan dengan tujuan penelitian. Keberhasilan tahap ini sangat menentukan kualitas analisis selanjutnya.

### 3. Pengolahan Data

Setelah data terkumpul, tahap selanjutnya adalah pengolahan data. Proses pengolahan ini dilakukan dengan menggunakan dua alat, yaitu Microsoft Excel dan Python. Excel digunakan untuk proses awal, seperti membersihkan data dari nilai yang hilang atau tidak valid, serta melakukan perhitungan dasar. Sementara itu, Python digunakan untuk proses analisis lebih lanjut, seperti normalisasi data dan transformasi variabel agar dapat digunakan dalam pemodelan statistik

### 4. Validasi Data

Pada tahap ini, proses validasi data dilakukan untuk memastikan bahwa data yang akan digunakan dalam analisis atau pemodelan memenuhi kriteria kualitas tertentu, yaitu konsistensi, akurasi, dan validitas. Validasi data mencakup pemeriksaan terhadap :

#### 1. Pola Outlier

Data diperiksa untuk mengidentifikasi nilai-nilai ekstrem yang secara signifikan menyimpang dari pola data lainnya. Outlier dapat memengaruhi hasil analisis, sehingga perlu ditangani, misalnya dengan transformasi, penghapusan, atau penggantian nilai.

## 2. Distribusi Data

Distribusi data dianalisis untuk memastikan apakah data sesuai dengan asumsi metode statistik atau pemodelan yang akan digunakan. Jika distribusi data tidak normal, transformasi seperti transformasi logaritma dapat diterapkan untuk menormalkan distribusi.

## 3. Hubungan Antar Variabel

Hubungan antar variabel dalam dataset diperiksa untuk mengidentifikasi pola korelasi atau ketergantungan yang mungkin memengaruhi hasil analisis. Dalam validasi data ini, transformasi logaritma digunakan untuk menangani distribusi data yang tidak normal atau memiliki skala yang terlalu luas. Transformasi logaritma membantu menormalkan data dan mengurangi pengaruh nilai-nilai ekstrem (outlier). Selain itu, transformasi ini juga efektif untuk memperbaiki hubungan non-linear antara variabel. Untuk data dengan nilai mendekati nol, digunakan penyesuaian :

$$X' = \log(X + 1) \quad (1)$$

Penjelasan Rumus :

- Rumus ini digunakan jika data asli (X) mengandung nilai nol atau mendekati nol.
- Penambahan angka 1 (+1) bertujuan untuk menghindari kesalahan karena  $\log(0)$  tidak terdefinisi (tak hingga).

- Transformasi ini membuat data lebih stabil, terutama untuk nilai kecil atau nol, sehingga analisis menjadi lebih akurat.

Pada tahap ini, data yang telah diolah diperiksa untuk memastikan konsistensi, akurasi, dan validitasnya. Validasi dilakukan dengan memeriksa pola outlier, distribusi data, dan hubungan antar variabel. Hal ini bertujuan untuk memastikan bahwa data yang digunakan dalam pemodelan dapat menghasilkan hasil yang valid.

## 5. Pemodelan (Modelling)

Setelah data diproses, tahap selanjutnya adalah membangun model untuk menganalisis hubungan antara variabel iklim dan curah hujan. Penelitian ini menggunakan tiga model utama: Generalized Linear Model (GLM), Random Forest (RF), dan Vector Autoregression (VAR). Setiap model memiliki pendekatan spesifik yang disesuaikan dengan karakteristik data dan tujuan analisis.

### 1. Generalized Linear Model (GLM)

Data diatur dengan membedakan variabel dependen (curah hujan) dan variabel independen (faktor iklim seperti suhu, kelembapan, kecepatan angin, dan durasi sinar matahari). Selanjutnya, dataset dibagi menjadi data latih dan data uji untuk memastikan evaluasi yang akurat. Fungsi link log sering digunakan untuk menangkap hubungan eksponensial yang umum dalam data iklim. Model dibangun dengan mengestimasi koefisien regresi



menggunakan metode estimasi maksimum likelihood. Koefisien ini memberikan wawasan tentang pengaruh masing-masing variabel independen terhadap curah hujan. Evaluasi model dilakukan dengan metrik seperti Mean Squared Error (MSE) dan deviance untuk memastikan keakuratan prediksi dan kemampuan model dalam menangkap pola data.

## 2. Random Forest (RF)

Merupakan algoritma ensemble berbasis pohon keputusan yang efektif untuk menangkap hubungan kompleks antar variabel. Model ini bekerja dengan membangun banyak pohon keputusan dari subset data yang diambil secara acak, kemudian menggabungkan hasil prediksi dari semua pohon dengan cara rata-rata. Parameter seperti jumlah pohon ( $n_{\text{estimators}}$ ), kedalaman maksimum ( $\text{max\_depth}$ ), dan jumlah fitur yang digunakan pada setiap split ( $\text{max\_features}$ ) disesuaikan untuk meningkatkan kinerja model. Data latih dan data uji digunakan untuk memastikan evaluasi yang konsisten. Performa model diukur menggunakan MSE dan Mean Absolute Error (MAE), sementara validasi silang digunakan untuk memastikan model stabil dan dapat diandalkan.

## 3. Vector Autoregression (VAR)

Dirancang untuk menganalisis hubungan dinamis antar variabel dengan mempertimbangkan efek temporal. Model ini memanfaatkan lag, yaitu nilai-nilai sebelumnya dari variabel, untuk memprediksi nilai saat ini. Lag optimal dipilih menggunakan kriteria informasi seperti Akaike Information Criterion (AIC) atau Bayesian Information Criterion (BIC).

Semua variabel iklim dan lag-nya dimasukkan ke dalam model untuk memprediksi variabel target. Koefisien yang dihasilkan mencerminkan pengaruh antar variabel dalam konteks waktu. Evaluasi dilakukan menggunakan MSE untuk memastikan keakuratan prediksi, dan hasil VAR dibandingkan dengan model lainnya untuk menentukan pendekatan terbaik.

Ketiga model ini dievaluasi secara menyeluruh berdasarkan performa prediksi, sehingga dapat ditentukan metode yang paling efektif untuk memahami hubungan antara variabel iklim dan curah hujan serta memprediksi pola curah hujan di masa depan.

#### 6. Evaluasi

Setelah model selesai dibangun, langkah evaluasi dilakukan untuk menilai sejauh mana model dapat memprediksi curah hujan secara akurat. Salah satu metrik yang digunakan adalah Mean Squared Error (MSE).

MSE adalah metrik evaluasi yang menghitung rata-rata dari kuadrat selisih antara nilai aktual dan nilai prediksi. Karena kesalahan dikuadratkan, MSE memberikan bobot lebih besar pada kesalahan prediksi yang besar, sehingga sangat sensitif terhadap outlier.

Rumus MSE :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

Penjelasan Rumus :

- $y_i$ : Nilai aktual.
- $\hat{y}_i$  : Nilai prediksi.
- $n$  : Jumlah data.

MSE membantu mengidentifikasi seberapa baik model dapat menangkap pola dalam data. Nilai MSE yang lebih kecil menunjukkan bahwa model menghasilkan prediksi yang lebih akurat. Metrik ini sangat cocok digunakan jika tujuan penelitian adalah meminimalkan kesalahan prediksi, terutama pada kasus di mana kesalahan yang besar memiliki dampak signifikan.

## 7. Model Prediksi

Berdasarkan evaluasi, model yang paling akurat dipilih untuk memprediksi curah hujan pada periode mendatang. Model ini dapat memberikan prediksi yang lebih realistis mengenai pola curah hujan, yang penting untuk perencanaan mitigasi bencana, seperti banjir, dan untuk pengelolaan sumber daya alam yang lebih efektif.

## 8. Hasil Penelitian

Pada tahap ini, hasil dari pemodelan dan evaluasi disusun dalam bentuk laporan yang merangkum temuan utama dari penelitian. Laporan ini berisi analisis tentang hubungan variabel iklim dan curah hujan, serta hasil



prediksi yang dapat digunakan untuk perencanaan di masa depan. Hasil penelitian ini juga dapat memberikan rekomendasi untuk mitigasi risiko cuaca ekstrem dan perencanaan sumber daya alam berbasis iklim.

## 9. Selesai (Finish)

Penelitian selesai setelah hasil telah disajikan. Hasil ini memberikan wawasan yang berguna tentang prediksi cuaca dan pola yang dapat digunakan untuk perencanaan atau pengambilan keputusan terkait cuaca dan iklim.

### 3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif, yang bertujuan untuk menganalisis hubungan antara variabel-variabel iklim dengan curah hujan di Kota Kediri. Metode kuantitatif dipilih karena pendekatan ini memungkinkan peneliti untuk mengeksplorasi pola hubungan menggunakan data numerik. Sebagai contoh, pengaruh kenaikan suhu terhadap intensitas curah hujan dapat dianalisis secara mendalam menggunakan regresi linier dan metode statistik lainnya. Penelitian ini juga memungkinkan evaluasi keakuratan prediksi menggunakan metrik Mean Absolute Error (MAE) untuk memberikan hasil yang valid dan reliabel.

### 3.2 Desain Penelitian

Desain penelitian ini bersifat korelasional, di mana hubungan antara variabel bebas (seperti suhu, kelembapan, kecepatan angin) dan variabel terikat (curah hujan) dianalisis untuk mengidentifikasi pola. Sebagai contoh, peneliti akan

memeriksa bagaimana peningkatan kelembapan udara pada bulan-bulan tertentu berkorelasi dengan peningkatan curah hujan. Dengan pendekatan ini, peneliti tidak hanya memahami hubungan antar variabel tetapi juga mengidentifikasi faktor dominan yang memengaruhi curah hujan di Kota Kediri.

### 3.3 Variabel Penelitian

- Variabel Bebas (Independent Variables) yaitu Temperatur udara, kelembapan, kecepatan angin, dan penyinaran matahari. Variabel ini dipilih berdasarkan literatur yang menunjukkan bahwa faktor-faktor ini memiliki pengaruh signifikan terhadap curah hujan. Sebagai ilustrasi, data dari BMKG menunjukkan bahwa peningkatan suhu sering diikuti oleh peningkatan penguapan yang berkontribusi pada intensitas curah hujan.
- Variabel Terikat (Dependent Variable) yaitu Curah hujan. Variabel ini menjadi fokus penelitian untuk membantu memprediksi potensi banjir berdasarkan analisis klimatologi.

### 3.4 Prosedur Penelitian

#### 1. Akses Data

Data cuaca harian diperoleh dari database BMKG Kota Kediri. Data ini meliputi catatan historis terkait variabel penelitian selama periode Januari hingga Oktober 2024.

#### 2. Pembersihan Data

Tahap ini melibatkan penghapusan data yang tidak lengkap atau tidak relevan. Sebagai contoh, data yang memiliki nilai kosong pada suhu atau curah hujan akan diisi menggunakan metode imputasi seperti *mean substitution*.

### 3. Penyusunan Dataset

Data yang sudah bersih dikelompokkan berdasarkan variabel yang relevan dan disusun dalam format siap analisis, seperti CSV atau Excel.

### 4. Analisis Awal

Exploratory Data Analysis (EDA) dilakukan untuk memahami pola data, seperti tren musiman curah hujan atau fluktuasi suhu selama periode tertentu.

### 5. Pengembangan Model

Model prediksi dibuat menggunakan algoritma Advanced Regression, yang kemudian diuji untuk keakuratan.

## 3.5 Populasi dan Sampel

- Populasi penelitian mencakup seluruh data cuaca harian Kota Kediri yang dicatat oleh BMKG. Populasi ini penting untuk memastikan bahwa hasil penelitian mencakup seluruh variasi kondisi cuaca di wilayah tersebut.

- Sampel data dipilih dengan teknik sampling sistematis, di mana data dari sejumlah periode waktu tertentu dipilih untuk mewakili variasi kondisi cuaca. Dataset ini terdiri dari 304 kolom, yang masing-masing merepresentasikan variabel atau atribut yang dianalisis untuk mengidentifikasi pola cuaca. Pemilihan data dilakukan agar hasil penelitian bisa menggambarkan kondisi cuaca yang beragam di wilayah penelitian dengan baik.

### 3.6 Teknik Pengumpulan Data

#### 1. Pengunduhan Data

Data cuaca harian dalam format CSV/Excel diunduh secara berkala untuk memastikan kelengkapan dataset.

#### 2. Validasi Data

Validasi dilakukan untuk memeriksa apakah semua variabel yang dibutuhkan tersedia dan memenuhi kriteria kualitas data. Sebagai contoh, data dengan pengukuran anomali, seperti suhu yang terlalu tinggi dibandingkan dengan rata-rata, diperiksa ulang.

#### 3. Pengelompokan Data

Data disusun berdasarkan variabel-variabel penelitian untuk mempermudah analisis, seperti memisahkan data suhu dan kelembapan berdasarkan bulan pengamatan.

### 3.7 Instrumen Penelitian

- Data Sekunder

Catatan cuaca harian dari BMKG yang meliputi curah hujan, suhu, kelembapan, kecepatan angin, dan penyinaran matahari. Data ini dinilai sebagai sumber yang akurat karena berasal dari pengamatan langsung menggunakan alat standar internasional.

- Alat Observasi :
  - *Rain gauge* untuk mengukur curah hujan.
  - *Anemometer* untuk mencatat kecepatan angin.
  - *Termometer* untuk mengukur suhu udara.

### 3.8 Analisis Data

Analisis data pada penelitian ini dilakukan melalui tahapan berikut:

1. Exploratory Data Analysis (EDA)

Langkah awal berupa analisis eksplorasi data dilakukan untuk mendapatkan wawasan awal mengenai dataset. Analisis ini mencakup pemeriksaan statistik deskriptif, visualisasi pola rata-rata bulanan curah hujan, dan identifikasi hubungan antar variabel, seperti korelasi antara suhu, kelembapan, dan curah hujan.

2. Pemodelan Prediksi



Pada tahap ini, model prediktif dikembangkan menggunakan algoritma Advanced Regression untuk memprediksi curah hujan berdasarkan variabel independen yang tersedia. Model ini dibandingkan dengan pendekatan lain seperti regresi linier sederhana untuk menilai apakah algoritma yang digunakan dapat memberikan hasil prediksi yang lebih akurat.

### 3. Evaluasi Kinerja Model

Evaluasi model dilakukan menggunakan metrik MSE (Mean Squared Error) untuk mengukur tingkat kesalahan prediksi. MSE memberikan bobot lebih besar pada kesalahan yang besar, sehingga membantu mendeteksi prediksi yang sangat meleset. Penilaian menggunakan kedua metrik ini memberikan gambaran yang lebih lengkap tentang keandalan dan performa model dalam memprediksi curah hujan.