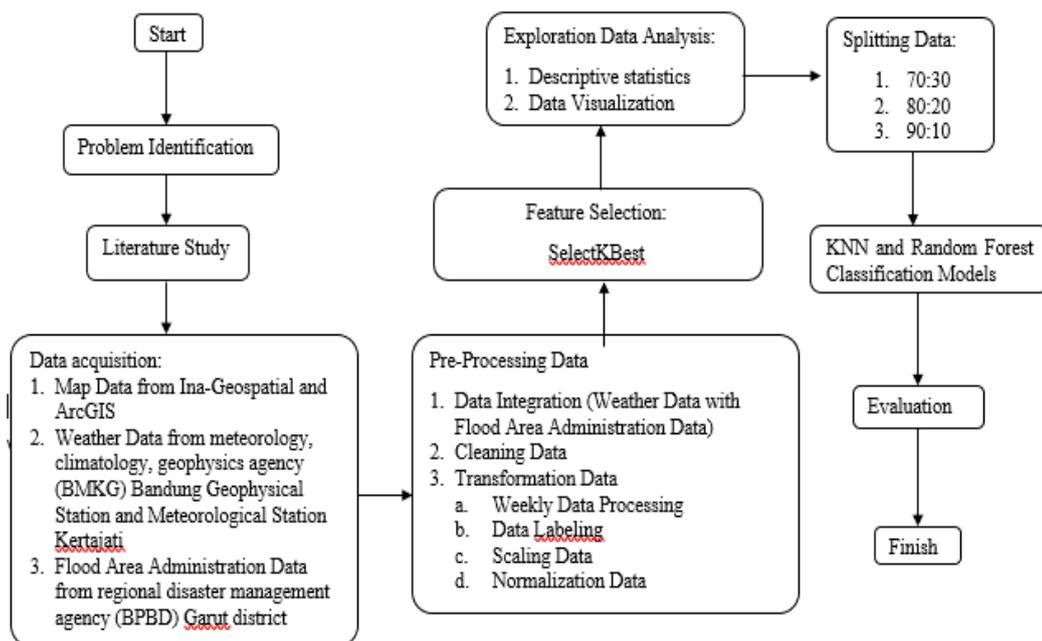


BAB III

METODE PENELITIAN

Alur metode penelitian yang dilakukan pada penelitian ini menggunakan kerangka penelitian seperti pada Gambar 3.1 berikut:



Gambar 3. 1 Kerangka Penelitian Terkait

3.1 Identifikasi Masalah

Dalam indentifikasi masalah pada penelitian ini, terdapat dua fokus utama. Pertama, penelitian ini bertujuan untuk mengusulkan penggunaan pendekatan multivariable dalam deteksi banjir dengan memanfaatkan metode *Machine learning* seperti *K-Nearest Neighbor* dan *Random Forest*. Kedua, penelitian ini membandingkan performa kedua algoritma tersebut dalam mendeteksi banjir, dengan tujuan untuk menentukan metode yang lebih efektif dan efisien.

3.2 Studi Literatur

Studi literatur ini bertujuan memahami metode prediksi banjir menggunakan data cuaca dan algoritma Machine Learning seperti *K-Nearest Neighbors* (KNN) dan *Random Forest* (RF). Data cuaca dari BMKG dan peta wilayah dari Ina-Geospasial mendukung analisis ini. Kajian mengacu pada penelitian sebelumnya dengan variasi parameter KNN ($k=1-10$) dan RF ($n_estimators=10$ dan 100) serta evaluasi model melalui akurasi, precision, recall, F1-score, dan Confusion Matrix.

3.3 Pengumpulan Data (Akuisisi)

Penelitian ini menggunakan data peta wilayah dari Ina-Geoportal, termasuk koordinat bujur, lintang, dan elevasi, untuk memetakan area terdampak banjir dari Kabupaten Garut dan menentukan stasiun cuaca terdekat dari BMKG melalui perhitungan centroid. Data cuaca dari BMKG Stasiun Geofisika Bandung (periode 1 Januari 2022 – 7 Juli 2024) serta data banjir dari BPBD Garut pada periode yang sama dikumpulkan, mencakup informasi kejadian banjir, korban, dan kerusakan infrastruktur. Tahap pengumpulan data mencakup pemetaan wilayah, pemilihan stasiun cuaca terdekat, dan integrasi data cuaca serta banjir untuk analisis prediksi dengan model KNN dan *Random Forest*.

3.4 Pre-processing Data

Setelah data terkumpul, dilakukan beberapa tahap pemrosesan awal, yaitu:

3.4.1 Data Integrasi (Penggabungan Data)

Pengintegrasian data cuaca dari BMKG di Stasiun Geofisika Bandung dengan data administrasi daerah banjir dari BPBD Kabupaten Garut untuk menghasilkan satu dataset yang relevan.

3.4.2 Data Cleaning (Pembersihan Data)

Menghapus data yang tidak lengkap atau tidak relevan, serta menangani *missing values* agar hasil analisis lebih akurat.

3.4.3 Data Transformation (Transformasi Data)

Data yang telah dibersihkan kemudian ditransformasikan melalui beberapa proses seperti berikut :

a. Pengolahan Data Mingguan

Proses ini mengubah data harian menjadi format mingguan melalui metode agregasi yang disesuaikan untuk setiap kolom. Penelitian (Gideon Namlea Lesnusa¹, 2024) penggunaan data mingguan memberikan tingkat kedetailan yang lebih baik untuk menganalisis pola kejadian banjir, yang dapat merespons lebih cepat terhadap perubahan cuaca dibandingkan data bulanan. Pendekatan mingguan dapat mendeteksi dan menganalisis dampak lebih rinci dan akurat dalam kejadian banjir yang seringkali terjadi dalam rentang waktu yang singkat.

b. Data Selection

Data Selection adalah proses memilih subset data yang relevan dari kumpulan data yang lebih besar untuk dianalisis. Ini bisa melibatkan

pemilihan baris, kolom, atau nilai tertentu berdasarkan kriteria atau kondisi tertentu, seperti rentang waktu, jenis variabel, atau nilai tertentu. Tujuannya adalah menyaring informasi yang penting dan mengeliminasi data yang tidak diperlukan untuk analisis lebih lanjut.

c. Pelabelan data

Labeling data adalah proses memberi label atau kategori pada data yang ada, untuk tujuan klasifikasi atau pengenalan pola. Dalam konteks analisis data, labeling berarti memberi penilaian atau label tertentu pada data berdasarkan karakteristik. Dalam penelitian ini, labeling data dilakukan dengan memberikan label pada data iklim dan banjir. Proses labeling ini melibatkan pengumpulan data historis terkait parameter iklim (seperti curah hujan, suhu, kelembapan) dan data banjir (seperti jumlah kejadian banjir, kerusakan yang terjadi, jumlah korban). Berdasarkan parameter-parameter ini, data yang terkait dengan kejadian banjir akan diberi label "terjadi banjir" (Label 1), sedangkan data yang tidak terkait dengan kejadian banjir akan diberi label "tidak terjadi banjir" (Label 0). Parameter-parameter yang digunakan dalam data pelabelan, diantaranya:

1. Parameter Data Iklim

Tabel 3. 1 Parameter Data Iklim

Parameter	Kondisi "Banjir"	Kondisi "Tidak Banjir"	Penjelasan
Tn (Suhu minimum)	$< 20^{\circ}\text{C}$	$\geq 20^{\circ}\text{C}$	Suhu minimum di bawah 20°C mengindikasikan potensi cuaca ekstrem

Parameter	Kondisi “Banjir”	Kondisi “Tidak Banjir”	Penjelasan
			(misalnya hujan deras di malam/pagi hari).
T _x (Suhu Maksimum)	> 35°C	≤ 35°C	Suhu maksimum di atas 35°C menunjukkan suhu siang ekstrem yang dapat memicu hujan konvektif.
T _{avg} (Suhu Rata-rata)	> 30°C	≤ 30°C	Suhu rata-rata tinggi mengindikasikan cuaca panas yang ekstrem.
RH _{avg} (Kelembapan Rata-rata)	> 85%	≤ 85%	Kelembapan tinggi (> 85%) dapat menjadi tanda potensi hujan lebat.
RR (Curah Hujan)	> 50 mm	≤ 50 mm	Curah hujan lebih dari 50 mm/hari digolongkan hujan sangat lebat, yang berpotensi menyebabkan banjir.
ss (Durasi Penyinaran Matahari)	< 2 jam	≥ 2 jam	Penyinaran matahari rendah (< 2 jam) mengindikasikan hari mendung atau hujan.
ff _x (Kecepatan Angin Maksimum)	> 10 m/s	≤ 10 m/s	Kecepatan angin maksimum yang tinggi menunjukkan cuaca buruk, seperti badai atau hujan lebat.
ff _{avg} (Kecepatan Angin Rata-rata)	> 5 m/s	≤ 5 m/s	Kecepatan angin rata-rata tinggi menunjukkan angin yang konsisten kuat, sering terjadi saat cuaca buruk.
ddd _{car_label} (Arah Angin Tenang)	1 (Tidak Tenang)	0 (Tenang)	Angin yang tidak tenang mengindikasikan gangguan atmosfer, sering kali dikaitkan dengan cuaca buruk atau badai.

2. Parameter Data Banjir

Tabel 3. 2 Parameter Data Banjir

Parameter	Kondisi "Banjir"	Kondisi "Tidak Banjir"	Penjelasan
jumlah kejadian	> 0 kejadian	0 kejadian	Jika ada kejadian banjir, ini dianggap sebagai indikator banjir.
Jumlah Korban	> 5 orang	0 orang	Korban lebih dari 5 jiwa menunjukkan dampak signifikan dari banjir.
Total Kerusakan Fasilitas	> 5 fasilitas	0 fasilitas	Kerusakan fasilitas umum lebih dari 5 unit menunjukkan dampak parah dari banjir.
Total Kerusakan Unit	> 5 unit	0 unit	Kerusakan unit properti/peralatan lebih dari 5 menunjukkan tingkat kerusakan tinggi akibat banjir.
TPT (m) (Unit)	> 10 unit	0 unit	Kerusakan talud, perkuatan tebing, atau tanggul lebih dari 10 unit menunjukkan banjir besar.
Jalan (m) (Unit)	> 3 unit	0 unit	Kerusakan jalan signifikan sering terjadi akibat banjir besar.
Sawah/Kebun/Lahan (m ²)	> 100 m ²	0 m ²	Luas sawah atau kebun yang rusak lebih dari 100 m ² menunjukkan dampak besar banjir pada sektor pertanian.
Kolam (ha) (Unit)	> 1 hektar	0 hektar	Kerusakan kolam ikan atau budidaya yang luas menunjukkan dampak signifikan pada sektor perikanan atau ekonomi setempat.

d. Scaling Data

Scaling data adalah proses yang digunakan untuk mengubah rentang nilai dari variabel dalam dataset sehingga lebih konsisten dan tidak memiliki skala yang berbeda-beda. Hal ini penting dalam analisis data dan pembelajaran mesin karena model dapat lebih efektif dan cepat beradaptasi jika fitur-fitur yang digunakan memiliki skala yang seragam.

e. Normalisasi Data

Normalisasi data dilakukan untuk menstandarkan skala fitur-fitur pada atribut yang digunakan. Hal ini penting agar setiap fitur memiliki rentang nilai yang sama, terutama ketika fitur tersebut memiliki satuan atau skala yang berbeda.

3.5 Feature Selection

Pemilihan fitur dilakukan untuk meningkatkan kinerja model klasifikasi dengan memilih atribut paling relevan dalam dataset. Penelitian ini menggunakan metode SelectKBest dengan fungsi penilaian f_{classif} (ANOVA F-value) untuk mengidentifikasi dan memilih fitur-fitur yang paling relevan dalam memprediksi kejadian dan dampak banjir. Pada penelitian (Fitri et al., 2023) Metode filter seperti SelectKBest digunakan untuk memilih fitur terbaik dengan meningkatkan akurasi prediksi. SelectKBest menilai fitur berdasarkan uji statistik, memilih fitur dengan skor tertinggi yang paling relevan dengan variabel target.

3.6 Exploration Data Analysis (EDA)

Dilakukan analisis deskriptif untuk memahami distribusi variabel serta hubungan antar variabel. Visualisasi data membantu dalam melihat pola yang ada dalam dataset.

3.7 Splitting Data

Setelah fitur-fitur penting dipilih, data kemudian dibagi menjadi data latih (training data) dan data uji (testing data) dengan beberapa skema pembagian, yaitu: 70:30, 80:20, 90:10. Skema ini digunakan untuk melatih model dan mengevaluasi performa model. Rasio 70:30 memberikan keseimbangan yang baik, 80:20 meningkatkan generalisasi dengan lebih banyak data training, dan 90:10 memaksimalkan pelatihan model. Dengan mencoba ketiga rasio, peneliti dapat menentukan rasio terbaik untuk akurasi optimal dan keandalan prediksi. (Galang Paksi Permana, 2024)

3.8 Klasifikasi Data (*K-Nearest Neighbor* dan *Random Forest*)

3.8.1 K-Neareast Neighbor

Algoritma knn diberi nilai $k=1-10$, maka nilai k berfungsi sebagai nilai *neighbour* yang akan digunakan untuk klasifikasi. Nilai k ini berfungsi sebagai menentukan jarak antar data atau tetangga dari data tersebut. Pada penelitian bahwa pemilihan rentang $k=1$ hingga 10 pada *K-Nearest Neighbors* (KNN) umum digunakan untuk menentukan parameter terbaik, guna mencapai akurasi tinggi dalam prediksi data cuaca. Variasi nilai k dalam rentang ini membantu mengidentifikasi jumlah tetangga optimal untuk klasifikasi, menyeimbangkan antara underfitting (terlalu sederhana)

dan overfitting (terlalu kompleks), sehingga k yang tepat dapat meningkatkan kinerja model (Rangkuti et al., 2021).

3.8.2 *Random Forest*

Algoritma ini digunakan dengan variasi parameter $n_estimators = 10$ (Aji et al., 2020) dan 100 untuk menentukan jumlah pohon keputusan yang digunakan dalam model. penggunaan 100 estimator membuat model lebih akurat dan andal, khususnya untuk kategori-kategori yang dinilai (Anwar et al., 2021).

3.9 Evaluasi

Setelah model dilatih, dilakukan evaluasi menggunakan data uji untuk mengukur akurasi, presisi, recall, dan F1-score dari masing-masing model. Evaluasi ini membantu menentukan model mana yang paling tepat digunakan dalam memprediksi kejadian banjir di Kabupaten Garut.