

BAB I PENDAHULUAN

1.1. Latar Belakang

Indonesia secara geografis dan iklim memiliki kemungkinan terbesar terhadap bencana banjir [1]. Laporan tahun 2025 dari Badan Penanggulangan Bencana Nasional menunjukkan angka 4.727 peristiwa bencana alam dengan banjir sebagai bencana yang mendominasi, yakni 2.009 kejadian atau 42,5% dari total kejadian bencana nasional [2]. Fenomena banjir ini berdampak signifikan pada berbagai aspek, baik dari segi ekonomi, lingkungan, maupun keselamatan publik [3]. Hal ini diperparah oleh perubahan topologi lahan yang berubah tahun ke tahun [4] dan curah hujan yang tinggi [5]. Kompleksitas pencegahan bencana ini menuntut adanya mitigasi bencana berbasis teknologi secara sistematis, terencana, dan berbasis data [6].

Salah satu upaya mitigasi yang dapat dilakukan yaitu, dengan memetakan kerawanan banjir [7]. Dalam beberapa tahun terakhir, berbagai penelitian telah dilakukan untuk memetakan kerawanan dan prediksi banjir dengan beragam pendekatan metodologi. Salah satu pendekatan konvensional yang digunakan adalah metode pembobotan berbasis pengetahuan ahli, seperti *Analytic Hierarchy Process* [8]. Namun, metode tersebut dikritik dikarenakan kerentanannya terhadap subjektivitas peneliti atau para ahli [9]. Seiring berkembangnya teknologi, pendekatan statistik berbasis algoritma *Machine Learning* semakin banyak diterapkan, seperti algoritma K-Means dan Random Forest, yang diterapkan untuk memprediksi banjir di DKI Jakarta dan menghasilkan akurasi sebesar 95%. [10]. Di sisi lain, penelitian yang menggunakan algoritma Regresi Logistik telah diterapkan pada berbagai daerah seperti pada daerah Kuala Lumpur, yang menghasilkan akurasi sebesar 84% [11], dan kota Charlotte dengan akurasi sebesar 97.92% [12]. Selain itu, pendekatan *hybrid*, yang menggabungkan metode Regresi Logistik dengan *neural network*, juga dilakukan. Meskipun menghasilkan tingkat akurasi yang lebih rendah, yaitu sebesar 79%, metode ini dinilai memiliki sensitivitas yang lebih baik dalam mengidentifikasi kelas banjir [13]. Penelitian terdahulu telah menunjukkan adanya perkembangan dalam pemilihan metode untuk pemetaan kerawanan banjir. Meskipun demikian, sebagian besar penelitian tersebut masih terbatas pada ruang lingkup yang relatif sempit, dengan penerapan pada wilayah studi tertentu seperti pada daerah kota, kecamatan, atau hanya di area spesifik, seperti di sekitar daerah aliran sungai. Sementara menurut Dawson (2023), meskipun model *Machine Learning* berhasil diterapkan untuk lokasi tertentu, model tersebut masih bisa digunakan

dalam skala yang lebih luas, dikarenakan biaya komputasi yang lebih hemat dibandingkan model banjir fisik [14].

Salah satu algoritma yang sesuai untuk memetakan kerawanan banjir adalah Regresi Logistik. Meskipun banyak algoritma *Machine Learning* yang dapat digunakan dan berpotensi memberikan akurasi yang lebih baik, Regresi Logistik dipilih karena tetap relevan dan banyak digunakan dalam penelitian terdahulu, seperti pada penelitian [15], [16], dan [17]. Selain karena model dapat menghasilkan probabilitas kejadian yang dapat digunakan untuk memetakan kerawanan banjir [18], Regresi Logistik merupakan model transparan yang memiliki keunggulan dalam interpretasi, sehingga memungkinkan manusia untuk memahami alasan model dalam melakukan prediksi. Hal ini menjadi nilai tambah dibandingkan model yang bersifat *black-box*, yang sulit dipahami manusia, dan dapat menurunkan tingkat kepercayaan dalam proses pengambilan keputusan [19], [20]. Di sisi lain, algoritma ini juga relatif efisien dalam komputasi, sehingga lebih sesuai untuk kondisi dengan keterbatasan sumber daya, atau kebutuhan prediksi secara *real-time*. Dari uraian tersebut, maka penelitian ini menggunakan algoritma Regresi Logistik untuk memetakan kerawanan banjir.

Dalam pemetaan kerawanan banjir, akurasi model juga bergantung pada pemilihan variabel independen dan data *training* [21]. Berdasarkan beberapa penelitian, terdapat beberapa faktor yang mempengaruhi probabilitas banjir, seperti parameter elevasi dan jarak ke sungai yang dianggap penting dalam penelitian [12], [21], beserta parameter lainnya, seperti curah hujan [17], [21]. Selain itu, parameter kemiringan juga konsisten digunakan dalam penelitian [11], [12], [17], [21], [22]. Berdasarkan uraian tersebut, maka penelitian ini memilih 4 parameter, yaitu elevasi, jarak ke sungai, curah hujan, dan kemiringan.

Ratusan kabupaten/kota yang tersebar di Indonesia dipilih sebagai ruang lingkup studi dalam penelitian ini. Pemilihan ruang lingkup ini didasarkan pada dua pertimbangan strategis. Pertama, Keadaan geografis di Indonesia sangatlah beragam mulai dari daerah yang terletak di dataran tinggi, dataran rendah, tetapi juga ada suatu daerah yang memiliki dataran rendah dan juga dataran tinggi dan Curah hujan yang tinggi akan langsung berpengaruh terhadap meluasnya daerah genangan banjir di dataran rendah [23]. Kedua, ketersediaan basis data kejadian bencana nasional yang dikelola oleh Badan Nasional Penanggulangan Bencana (BNPB) melalui Data Informasi Bencana Indonesia (DIBI) menyediakan rekam jejak historis yang ekstensif, sehingga memungkinkan prediksi banjir secara komprehensif di berbagai karakteristik wilayah.

Penelitian ini bertujuan untuk menerapkan dan menguji performa metode tersebut dalam memetakan probabilitas banjir dengan mengambil data empiris dari ratusan kabupaten/kota yang tersebar di Indonesia . Berdasarkan uraian diatas, maka dilakukan penulisan tugas akhir dengan judul: **"IMPLEMENTASI METODE REGRESI LOGISTIK DALAM PEMETAAN KERAWANAN BANJIR"**.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, permasalahan yang ditemukan yaitu sebagian besar penelitian dalam pemetaan kerawanan banjir masih terbatas pada ruang lingkup yang sempit, seperti pada satu kota, kecamatan ataupun di area spesifik seperti sekitar daerah aliran sungai.

1.3. Tujuan

Tujuan dari penyusunan tugas akhir ini adalah membangun sistem pemetaan kerawanan banjir pada ruang lingkup yang lebih luas, dengan melibatkan ratusan kabupaten/kota yang tersebar di Indonesia, serta menguji performa Regresi Logistik pada studi kasus dengan ruang lingkup yang lebih luas.

1.4. Manfaat

Penelitian ini diharapkan dapat memberikan kontribusi positif baik dari aspek pengembangan ilmu pengetahuan maupun penerapan praktis di lapangan, yaitu:

1. Sebagai kontribusi bagi peneliti selanjutnya yang ingin mengembangkan model Regresi Logistik atau model banjir lainnya dalam pemetaan kerawanan banjir.
2. Bagi Pemerintah Daerah: Hasil berupa titik rawan banjir dapat digunakan sebagai data pendukung dalam evaluasi Rencana Tata Ruang Wilayah (RTRW) dan perumusan kebijakan mitigasi bencana yang lebih presisi dan tepat sasaran.
3. Bagi BNPB/Badan Penanggulangan Bencana Daerah (BPBD): Memberikan informasi spasial mengenai titik-titik prioritas yang memerlukan infrastruktur pengendalian banjir atau sistem peringatan dini berdasarkan tingkat probabilitas kejadian.
4. Bagi Masyarakat: Meningkatkan kesadaran mengenai tingkat kerentanan lingkungan tempat tinggal mereka terhadap bahaya banjir, sehingga dapat meningkatkan kesiapsiagaan menghadapi bencana.

1.5. Ruang Lingkup

Ruang lingkup dari tugas akhir sebagai berikut:

1. *Dataset* yang digunakan berjumlah 18.047 kejadian, terdiri dari 7.912 kejadian yang teridentifikasi banjir, dan 10.135 kejadian teridentifikasi tidak banjir. *Dataset* ini diperoleh dari kaggle <https://www.kaggle.com/datasets/yusufginanjar7/banjir-jabodetabek> yang merupakan hasil kompilasi data kejadian BNPB dan penggabungan parameter melalui Google Earth Engine (GEE). Meskipun *dataset* ini berjudul Banjir Jabodetabek, cakupan datanya tidak hanya terbatas di wilayah Jabodetabek. *Dataset* ini mencakup sekitar 365 nama kabupaten/kota, baik didalam maupun diluar Jabodetabek, dengan 9.242 kejadian di wilayah Jabodetabek dan 8.805 kejadian di luar wilayah Jabodetabek.
2. *Dataset* tersebut kemudian diproses untuk ditambahkan variabel baru seperti jarak terhadap sungai (*distance to river*).
3. Data jaringan sungai diambil dari *OpenStreetMap*.
4. Input sistem berupa parameter prediktor seperti elevasi (*elevation*), kemiringan tanah (*slope*), curah hujan (*rainfall*), dan jarak terhadap sungai (*distance to river*).
5. *Hyperparameter Tuning* menggunakan metode *GridSearchCV*
6. Metode Evaluasi: Performa model diukur menggunakan matriks evaluasi Confusion Matrix yang meliputi:
 - a. *Accuracy*: Mengukur tingkat ketepatan prediksi total.
 - b. *Precision, Recall & Specificity*: Mengukur kemampuan model dalam mengidentifikasi kelas banjir secara spesifik.
 - c. *F1-Score*: Keseimbangan antara precision dan recall.
 - d. *Area Under Curve of ROC* (AUC-ROC): Untuk melihat kemampuan model dalam membedakan antara kelas banjir dan tidak banjir pada berbagai ambang batas (*threshold*).
7. Output berupa:
 - a. Probabilitas banjir
 - b. Hasil prediksi yang menunjukkan apakah terjadi banjir atau tidak