

BAB I

PENDAHULUAN

1.1. Latar Belakang

Indonesia adalah negara kepulauan yang terletak di wilayah tropis dan terbentang di sepanjang garis khatulistiwa, yaitu antara 6° LU–11° LS dan 95° BT–141°BT. Kondisi geografis ini bersama dengan intensitas curah hujan tinggi dan musim kemarau panjang, menjadikan Indonesia sangat rentan terhadap berbagai bencana alam, seperti banjir, tanah longsor, dan kekeringan [1]. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) tahun 2023, tercatat sebanyak 1.255 peristiwa banjir di seluruh provinsi Indonesia dengan total 3.860.136 korban dan 1.083.004 kerusakan [2]. Bahkan pada tahun 2024, banjir menduduki peringkat pertama sebagai bencana alam yang paling sering terjadi di Indonesia dengan 1.420 peristiwa yang mencapai 6.391.056 korban dan 2.005.567 kerusakan [3]. Banjir tidak hanya menimbulkan kerugian material seperti kerusakan infrastruktur, tetapi juga mengganggu aktivitas sosial dan ekonomi masyarakat serta menimbulkan ancaman terhadap keselamatan jiwa.

Berdasarkan data dari BNPB tersebut, terjadi peningkatan yang signifikan pada jumlah korban dan kerusakan yang ditimbulkan akibat bencana banjir dibandingkan tahun sebelumnya. Kondisi ini menunjukkan bahwa bencana banjir masih menjadi permasalahan yang serius karena berdampak besar terhadap keselamatan jiwa dan kerusakan infrastruktur di berbagai wilayah Indonesia [4]. Selain itu, setiap provinsi memiliki karakteristik dampak banjir yang berbeda-beda, baik dari segi korban jiwa maupun kerusakan yang ditimbulkan. Perbedaan karakteristik tersebut menyebabkan pendekatan penanganan dan mitigasi bencana banjir pada setiap wilayah tidak dapat disamaratakan dan perlu disesuaikan dengan kondisi masing-masing wilayah [5]. Apabila perbedaan karakteristik dampak ini tidak dianalisis secara mendalam, maka alokasi sumber daya dalam upaya mitigasi bencana banjir berpotensi menjadi kurang efektif dan tidak tepat sasaran [6]. Oleh karena itu, dibutuhkan suatu pendekatan untuk mengidentifikasi karakteristik dampak banjir sekaligus menentukan wilayah prioritas dalam upaya mitigasi

bencana banjir. Klasterisasi dapat menjadi solusi yang efektif untuk mengelompokkan wilayah berdasarkan karakteristik tingkat keparahan dampak banjir yang serupa. Hasil klasterisasi yang diperoleh dapat menjadi dasar penyusunan strategi mitigasi banjir yang lebih terarah dan sesuai dengan kondisi setiap provinsi untuk mengurangi risiko kerugian, kerusakan, dan korban jiwa.

Beberapa penelitian terdahulu telah mengkaji metode klasterisasi terhadap indikator dampak banjir. Sebagai contoh, penelitian [5] melakukan klasterisasi provinsi di Indonesia berdasarkan kejadian dan dampak bencana banjir menggunakan metode *Fuzzy Possibilistic C-Means*. Penelitian ini menggunakan data sekunder dari BPBD yang terdiri dari 34 provinsi dengan variabel frekuensi kejadian dan dampak banjir pada periode 2017-2021. Penelitian ini menghasilkan 2 klaster terbaik dengan indeks *Modified Partition Coefficient* (MPC) sebesar 0,6506. Dari hasil klasterisasi tersebut, klaster 1 terdiri dari 19 provinsi dengan tingkat risiko rendah dan klaster 2 terdiri dari 15 provinsi dengan tingkat risiko tinggi. Sementara itu, penelitian serupa [7] dilakukan menggunakan metode *Fuzzy C-Means*. Hasil klasterisasi dari penelitian tersebut adalah 2 klaster dengan indeks MPC sebesar 0,6536, di mana klaster 1 terdiri dari 27 provinsi dengan tingkat risiko rendah dan klaster 2 terdiri dari 7 provinsi dengan tingkat risiko tinggi.

Berdasarkan beberapa penelitian sebelumnya, terlihat bahwa metode klasterisasi yang sering digunakan untuk analisis dampak banjir adalah metode *Fuzzy Possibilistic C-Means* dan *Fuzzy C-Means*. Namun, kedua metode tersebut memiliki keterbatasan dalam menangani data dengan kepadatan atau distribusi yang bervariasi maupun menangani penciran (*noise*) yang dapat memengaruhi hasil klasterisasi [8]. Berbeda dengan kedua metode tersebut, *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) mampu mengatasi keterbatasan ini melalui pendekatan struktur hierarkis berbasis kepadatan yang efektif untuk mengelompokkan data dengan kepadatan yang bervariasi sekaligus mengidentifikasi *noise* secara otomatis [9]. Selain itu, metode HDBSCAN juga secara otomatis dapat menentukan jumlah klaster optimal, tidak seperti metode-metode lain yang memerlukan penentuan jumlah klaster secara apriori [10]. Dalam proses pembentukan klaster, metode ini menggunakan metrik

Mutual Reachability Distance untuk membentuk struktur hierarki yang biasa disebut *Minimum Spanning Tree* berdasarkan tingkat kepadatan antartitik data. Titik yang tidak memenuhi ambang batas kepadatan maupun titik yang berada di dalam kelompok dengan jumlah anggota di bawah batas minimum yang ditentukan akan dikategorikan sebagai *noise* [11].

Berdasarkan penelitian [12], metode HDBSCAN terbukti mampu mengidentifikasi klaster dengan kepadatan yang bervariasi dan tahan terhadap *noise* sehingga dapat menghasilkan klaster yang lebih konsisten dibandingkan dengan metode *Fuzzy C-Means* dan *Fuzzy Possibilistic C-Means*. Selain itu, penelitian [13] juga membuktikan bahwa HDBSCAN sebagai pengembangan dari metode DBSCAN, lebih unggul dalam mengidentifikasi klaster dengan ukuran, bentuk, dan kepadatan yang berbeda tanpa memerlukan parameter epsilon secara manual seperti pada DBSCAN. Keunggulan-keunggulan tersebut menjadikan metode HDBSCAN sebagai solusi yang lebih sesuai untuk analisis data kebencanaan yang tidak berdistribusi seragam dan bervariasi antarwilayah karena variasi tersebut berpotensi membentuk kelompok data dengan tingkat kepadatan yang berbeda-beda. Oleh karena itu, inovasi penelitian ini terletak pada penerapan metode HDBSCAN untuk mengelompokkan provinsi di Indonesia berdasarkan indikator dampak banjir, yang saat ini masih belum diterapkan pada studi kasus serupa.

Metode HDBSCAN memiliki sejumlah parameter penting yang dapat disesuaikan, seperti *minimum samples* dan *minimum cluster size* yang secara signifikan memengaruhi kinerja model. Pemilihan parameter yang tidak tepat dapat menyebabkan hasil klasterisasi menjadi kurang representatif. Salah satu metode yang dapat digunakan untuk menentukan kombinasi parameter optimal adalah *Bayesian Optimization*, yaitu metode optimasi berbasis model probabilistik yang mampu mengeksplorasi ruang parameter secara adaptif [14]. Berbeda dengan metode pencarian konvensional seperti *Grid Search* atau *Random Search* yang cenderung memerlukan banyak percobaan, *Bayesian Optimization* menggunakan Teorema Bayes dan *Gaussian Process* untuk memodelkan fungsi objektif sehingga proses pencarian parameter menjadi lebih efisien, baik dari sisi akurasi maupun komputasi [15]. Penelitian [16] menunjukkan bahwa *Bayesian Optimization*

berhasil menemukan kombinasi parameter optimal pada model klasterisasi dan menghasilkan performa klaster yang jauh lebih baik dibandingkan dengan tanpa optimasi. Hasilnya terlihat dari *silhouette score* yang meningkat dari 0,54 menjadi 0,71. Oleh karena itu, penggunaan *Bayesian Optimization* ke dalam model klasterisasi dapat menjadi strategi yang efektif untuk memperoleh kombinasi parameter optimal dan meningkatkan performa klasterisasi.

Penelitian ini berfokus pada penerapan metode HDBSCAN dengan *Bayesian Optimization* untuk mengelompokkan provinsi di Indonesia berdasarkan indikator dampak banjir. Melalui penerapan metode HDBSCAN, penelitian ini menghasilkan klasterisasi yang lebih akurat melalui pendekatan hierarkis berbasis kepadatan (*hierarchical density-based*) dibandingkan dengan pendekatan partisi maupun hierarkis konvensional. *Bayesian Optimization* juga diterapkan untuk memperoleh kombinasi parameter optimal sehingga kinerja HDBSCAN dalam mengidentifikasi struktur klaster dapat ditingkatkan. Hasil analisis klasterisasi akan disajikan dalam bentuk *website* interaktif berbasis *Streamlit* yang menampilkan peta wilayah Indonesia berdasarkan tingkat keparahan dampak banjir. Peta tersebut dilengkapi dengan fitur label klasterisasi dan pewarnaan wilayah berdasarkan tingkat keparahan dampak banjir, serta tampilan nilai setiap variabel ketika pengguna memilih atau menekan suatu provinsi. Selain fitur visualisasi, pengguna juga dapat mengunduh hasil klasterisasi dalam format CSV untuk keperluan analisis lanjutan. Hasil penelitian ini diharapkan dapat memberikan gambaran yang komprehensif bagi pemerintah maupun instansi terkait dalam menyusun perencanaan strategi mitigasi bencana dan prioritas penanganan wilayah berdasarkan tingkat keparahan dampak banjir.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, rumusan masalah yang akan dibahas dalam penelitian ini meliputi:

1. Bagaimana penerapan metode HDBSCAN dengan *Bayesian Optimization* dapat mengelompokkan setiap provinsi di Indonesia berdasarkan indikator dampak banjir?

2. Bagaimana hasil klasterisasi provinsi di Indonesia berdasarkan indikator dampak banjir menggunakan metode HDBSCAN dengan *Bayesian Optimization*?
3. Bagaimana hasil rancangan *Graphical User Interface* (GUI) untuk menampilkan hasil klasterisasi provinsi di Indonesia berdasarkan indikator dampak banjir?

1.3. Batasan Masalah

Penetapan batasan masalah diperlukan untuk menegaskan fokus dan ruang lingkup kajian. Batasan-batasan masalah yang diterapkan adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari publikasi Badan Nasional Penanggulangan Bencana pada tahun 2024. Variabel yang digunakan meliputi jumlah korban meninggal dan hilang, jumlah korban terluka dan mengungsi, jumlah rumah rusak berat, jumlah rumah rusak sedang, jumlah rumah rusak ringan, dan jumlah rumah terendam.
2. Analisis klaster difokuskan pada pengelompokan wilayah berdasarkan indikator dampak banjir, tanpa menggunakan variabel penyebab banjir yang bersifat geografis maupun hidrologis.
3. Analisis klaster dilakukan pada tingkat provinsi sebagai fokus utama penelitian sehingga tidak mempresentasikan secara rinci tingkat dampak banjir pada wilayah administratif yang lebih kecil.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, tujuan utama dari penelitian ini adalah sebagai berikut:

1. Untuk menerapkan metode HDBSCAN dengan *Bayesian Optimization* dalam mengelompokkan setiap provinsi di Indonesia berdasarkan indikator dampak banjir.
2. Untuk mengetahui hasil klasterisasi provinsi di Indonesia berdasarkan indikator dampak banjir menggunakan metode HDBSCAN dengan *Bayesian Optimization*.

3. Untuk mengembangkan *Graphical User Interface* (GUI) berbasis *website* yang menampilkan hasil klasterisasi setiap provinsi di Indonesia berdasarkan indikator dampak banjir.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak, baik bagi bidang keilmuan, masyarakat, maupun pemerintah, yaitu:

1. Bagi bidang keilmuan

Penelitian ini dapat memberikan pemahaman tentang metode HDBSCAN dengan *Bayesian Optimization* dalam studi kasus pengelompokan setiap provinsi di Indonesia berdasarkan indikator dampak banjir. Selain itu, penelitian ini juga dapat memperkaya literatur terkait klasterisasi, khususnya dalam penerapan HDBSCAN yang dikombinasikan dengan *Bayesian Optimization* sebagai metode klasterisasi.

2. Bagi masyarakat

Penelitian ini dapat membantu meningkatkan pemahaman mengenai tingkat keparahan dampak banjir di masing-masing wilayah sehingga masyarakat dapat memahami kerugian yang dihadapi serta mendorong kesiapsiagaan dan partisipasi aktif dalam upaya mitigasi bencana banjir.

3. Bagi pemerintah atau instansi terkait

Penelitian ini dapat memberikan gambaran atau pemahaman yang lebih mendalam mengenai pola dampak banjir di setiap provinsi di Indonesia. Hasil penelitian ini dapat mendukung pengambilan keputusan dalam upaya penanggulangan dan mitigasi bencana banjir yang lebih tepat dan efisien, termasuk dalam pengembangan infrastruktur pencegah banjir dan prioritas penanganan pada wilayah dengan tingkat keparahan dampak banjir yang tinggi.