

BAB I

PENDAHULUAN

I.1 Latar Belakang

Negara Indonesia adalah negara yang beriklim tropis karena posisinya dilalui oleh garis khatulistiwa, hal ini menyebabkan perubahan iklim di sehingga dapat meningkatkan curah hujan. Curah hujan dihitung dengan satuan (mm) yang berisi total volume air yang jatuh ke tanah pada periode waktu tertentu. Intensitas curah hujan yang sangat tinggi di wilayah Indonesia cenderung disebut dengan cuaca ekstrem, karena hal ini dapat menyebabkan bencana banjir. (Marthin Luter Laia & Yudi Setyawan, 2020) (Prakoso Indaryono, 2024).

Diantara bencana alam yang telah terjadi di Indonesia, bencana hidrometeorologi adalah bencana yang cukup sering terjadi hingga persentasenya mencapai 92%. Bencana hidrometeorologi yang terjadi yaitu banjir, tanah longsor dan angin puting beliung. Dari keseluruhan persentase bencana hidrometeorologi yang terjadi, bencana banjir menjadi salah satu bencana yang paling sering terjadi yaitu sebanyak 43,19% (Adel Andira Putri, 2023). Pada tahun 2022 lalu Indonesia mengalami bencana banjir yaitu sebanyak 1.531 kejadian, hal tersebut menyebabkan 128 korban meninggal dunia, 22 korban hilang, 90 korban terluka, 5.300.042 korban menderita, 422.309 korban mengungsi, 2.012 rumah dengan rusak berat, 1.334 rumah dengan rusak sedang, 6.996 rumah dengan rusak ringan, 1.120.885 rumah terendam dan 67.368 kerusakan fasilitas umum dan lainnya (BNPB, 2023).

Pulau Jawa memiliki kepadatan penduduk yang tinggi di Indonesia karena menjadi pusat perekonomian, pemerintahan dan pendidikan (RENIKA, I., & Amin, C., 2021). Provinsi Jawa Timur adalah salah satu provinsi yang memiliki Kabupaten/Kota terbanyak di Indonesia yaitu berjumlah 38 Kabupaten/Kota dan menjadi provinsi terluas di Pulau Jawa dengan luas 48.036,84 km². Banyaknya kabupaten/kota yang ada di Jawa Timur pastinya memiliki karakteristik tiap wilayah yang berbeda salah satunya curah hujan, sehingga hal ini dapat menyebabkan bencana banjir terutama saat musim hujan. Berdasarkan data yang

dikumpulkan dari BPS, rata-rata curah hujan tahun 2022 di Jawa Timur mencapai angka 254 mm. Dari tingginya curah hujan tersebut dapat dipastikan bahwa Jawa Timur memiliki rata-rata curah hujan yang tergolong tinggi. Akibat dari rata-rata curah yang tinggi ini wilayah Jawa Timur tergolong masuk dalam cuaca ekstrem yang dapat menyebabkan bencana alam salah satunya yaitu banjir. Peristiwa banjir di Jawa Timur tahun 2022 terdapat 153 kejadian, hal ini mengakibatkan banyaknya korban dan kerusakan. Menurut data BNPB (Badan Nasional Penanggulangan Bencana) korban bencana banjir mencapai angka 478.559 sedangkan kerusakan yang dialami mencapai angka 95.992 kerusakan rumah, tempat ibadah dan fasilitas umum. (BNPB, 2023)

Wilayah Jawa Timur yang memiliki persentase tinggi terjadinya banjir pada tahun 2022 adalah Kabupaten Jember dengan banyak kejadian mencapai angka 20. Dampak bencana banjir mengakibatkan terjadinya korban, masalah kesehatan, kerusakan hunian masyarakat dan kerusakan fasilitas umum. Terdapat 3 korban meninggal dunia, 1 korban terluka 17.341 korban menderita, 157 korban mengungsi, 51 rumah rusak, 3.528 rumah terendam dan 6 kerusakan fasilitas umum. (BNPB, 2023). Bencana banjir juga terjadi di Kabupaten Banyuwangi tepatnya pada bulan November. Kejadian ini terjadi di Kelurahan Kertosari, Kelurahan Pakis dan Kelurahan Sobo. Terdapat sebanyak 3.245 jiwa warga yang terdampak, dampak yang paling parah terjadi di Kelurahan Sobo dengan sejumlah total 65 rumah rusak. Banjir ini diakibatkan hujan deras sejak minggu hingga hari senin di minggu 2 bulan Oktober 2022. Dampak banjir tersebut mengakibatkan 4 sungai yang meluap yaitu Sungai Kali Sobo, Sungai Kilo, Sungai Kali Bagong dan Sungai Kaligulung (Kompas.com).

Faktor yang menjadi penyebab dari bencana banjir salah satunya adalah curah hujan yang terjadi cukup tinggi sehingga membuat volume air meningkat. Meningkatnya volume air saat hujan dapat menyebabkan bencana banjir, karena tanah atau sistem drainase tidak mampu menahan dan menyerap air sehingga terjadi genangan di permukaan tanah (Nugraha, & M. D. R., 2023). Kepadatan penduduk yang tinggi menimbulkan kebutuhan lahan yang tinggi, sehingga akan banyak lahan yang digunakan untuk hunian masyarakat. Hal ini menyebabkan semakin

berkurangnya ketersediaan lahan (Putri, 2022). Peningkatan jumlah penduduk menjadi faktor bencana banjir karena kebutuhan lahan yang meningkat sehingga berkurangnya area resapan lahan. Selain itu perilaku masyarakat juga dapat meningkatkan bencana banjir, seperti membuang sampah sembarangan serta menebang pohon dan penggundulan hutan secara ilegal (Effendi & Siswandi, 2024). Luas wilayah menjadi indikator yang signifikan dalam bencana banjir, karena semakin sempit luas wilayah maka area resapan air semakin kecil. Kepadatan penduduk dapat meningkatkan kebutuhan infrastruktur, pemukiman, air bersih, fasilitas dan peningkatan produksi sampah. Jika permasalahan ini tidak diselesaikan dengan baik dapat berpengaruh pada pengelolaan lahan untuk resapan air (Sulaiman et al., 2020). Luas wilayah juga berpengaruh dengan kemampuan area dalam menyerap air. Jika suatu wilayah memiliki luas yang besar namun mengubah lahannya menjadi permukaan yang kedap air seperti beton atau aspal dapat mempengaruhi penyerapan dan pengelolaan air untuk masuk ke tanah yang berakibat air hujan mengalir keluar dari permukaan tanah dan meningkatkan resiko terjadinya banjir (Geografi et al., 2023).

Berbagai faktor diatas, seperti curah hujan, kepadatan penduduk, dan luas wilayah berkaitan erat dengan jumlah kejadian banjir karena mempengaruhi tingkat frekuensi terjadinya banjir di wilayah tersebut (Nizar, Y. A., Susilawati, M., & Srinadi, 2023). Pemilihan variabel didasarkan pada penelitian terdahulu, seperti penelitian dengan judul Pemodelan Jumlah Kejadian Banjir di Kabupaten dan Kota Provinsi Jawa Timur dengan Metode Geographically Weighted Regression (GWR) yang menganalisis jumlah kejadian banjir di Jawa Timur dengan mempertimbangkan variabel kepadatan penduduk, dan curah hujan. Hasilnya menunjukkan bahwa variabel-variabel tersebut signifikan dalam mempengaruhi jumlah kejadian banjir di wilayah tersebut (Nizar, Y. A., Susilawati, M., & Srinadi, 2023). Peneliti lain dengan judul Analisis Penyebab Banjir Di Kota Samarinda yang menganalisis luas wilayah berpengaruh terhadap bencana banjir karena mempengaruhi kapasitas resapan air, tata guna lahan, dan distribusi drainase. Wilayah dengan area resapan air yang terbatas atau mengalami alih fungsi lahan secara masif menjadi lebih rentan terhadap genangan, terutama saat intensitas curah hujan tinggi (Sulaiman et al., 2020).

Banyaknya kerugian materi, korban jiwa dan kerusakan akibat bencana banjir tersebut perlu dilakukan pengelompokan bencana banjir menurut daerah kabupaten kota yang rawan atau tidak. Metode data mining sangat tepat digunakan dalam penelitian ini karena menggunakan pendekatan statistik, matematik, kecerdasan buatan dan juga melalui pembelajaran mesin untuk mendapatkan informasi penting dari sekumpulan data yang cukup besar. Effendi & Siswandi, 2024). Data mining menjadi sebuah proses pengambilan informasi dari sejumlah data yang banyak untuk mendapatkan informasi baru yang bisa digunakan secara langsung. Dalam proses sebuah data mining terdapat beberapa metode analisis data yaitu *Clustering*, *Association*, Klasifikasi dan lainnya. Effendi & Siswandi, 2024). *Clustering* merupakan metode yang cukup efektif untuk melakukan analisis data dengan cara mengamati secara rinci yang dibagi menjadi 3 kategori yaitu partisi grid, hirarki dan model (Effendi & Siswandi, 2024). Algoritma K-Means menjadi metode pengelompokan data yang dikelompokkan menjadi banyak kelompok menurut jarak antara elemen data dengan pusat kelompoknya. Sehingga dapat menciptakan kelompok-kelompok yang memiliki kemiripan atau karakteristik sama (Andi Syahrul Ramdana, Kusri, 2020). Data mining dengan metode *Clustering* digunakan dalam pengelompokan data agar menjadi terpisah secara kelompok atau *Cluster* yang memiliki kondisi kedekatan karakteristik sampel (Wathoriq, 2023). Penerapan data mining dengan metode *Clustering* dapat membantu dalam menentukan pola-pola dan hubungan antara variabel.

Penelitian sebelumnya sudah melakukan penelitian untuk menganalisis potensi kerawanan banjir berdasarkan dampak yang terjadi (Halik & Septiana, 2022), memprediksi daerah rawan bencana alam dengan pengklasifikasian 3 *Cluster* yaitu rendah, sedang dan tinggi (Sandiwarno, 2024). Kemudian terdapat penelitian untuk mengelompokkan (*Clustering*) Kecamatan berdasarkan tingkat rawan bencana (Wahidah et al., 2023) dan meneliti dampak yang ditimbulkan karena banjir pada suatu wilayah yang rawan terhadap terjadinya bencana banjir dengan menampilkan perhitungan melalui sistem (Wathoriq, 2023).

Penelitian ini menggunakan empat variabel diantaranya yaitu Curah Hujan, Kepadatan Penduduk, Luas Wilayah dan Banyak Kejadian. Metode yang akan dipakai dalam penelitian ini adalah metode K-Means *Clustering* seperti yang dijelaskan sebelumnya. Berdasarkan pembahasan tersebut penelitian ini memiliki tujuan untuk mengelompokkan bencana banjir di Jawa Timur berdasarkan tingkat kerawannya menggunakan metode *Clustering*. Judul penelitian ini adalah “Penerapan Data Mining Untuk Memprediksi Daerah Rawan Banjir Di Jawa Timur Menggunakan Metode K-Means *Clustering*”. Hasil dari penelitian ini divisualisasikan dalam visualisasi *dashboard* daerah rawan banjir di Jawa Timur. Hasil visualisasi *dashboard* dapat memberikan informasi visual yang mudah dipahami dengan hasil pengelompokan daerah tingkat berdasarkan kerawannya. Diharapkan dengan ini dapat membantu pemerintah untuk mengantisipasi dan menanggulangi bencana banjir sehingga dapat meminimalisir dampak bencana banjir yang akan terjadi.

I.2 Perumusan Masalah

Berdasarkan latar belakang yang dijabarkan, maka rumusan permasalahan untuk penelitian ini ada tiga yaitu :

1. Bagaimana mengelompokkan wilayah di Jawa Timur berdasarkan tingkat kerawanan banjir?
2. Bagaimana evaluasi dari hasil *Cluster* menggunakan k-means untuk mengelompokkan wilayah banjir yang ada di Jawa Timur?
3. Bagaimana visualisasi dari hasil *Clustering* bencana banjir di Jawa Timur berdasarkan tingkat kerawannya?

I.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, tujuan yang dilakukan untuk penelitian ini ada tiga yaitu :

1. Mengelompokkan wilayah banjir di Jawa Timur berdasarkan tingkat kerawannya menggunakan metode Algoritma K-Means *Clustering*

2. Mengetahui evaluasi dari *Cluster* banjir di wilayah Jawa Timur berdasarkan tingkat kerawannya dengan menerapkan 3 perbandingan jarak Euclidean, Manhattan, Chebysev dan PCA.
3. Memvisualisasi hasil *Clustering* bencana banjir di Jawa Timur berdasarkan tingkat kerawannya menggunakan Looker Studio.

I.4 Batasan Penelitian

Berdasarkan rumusan masalah serta tujuan diatas, batasan untuk penelitian ini ada empat yaitu :

1. Penelitian ini menggunakan data tahun 2022.
2. Variabel yang digunakan ada 4 yaitu rata-rata curah hujan, luas wilayah, banyak kejadian banjir dan kepadatan penduduk.
3. Fokus objek penelitian ini adalah Provinsi Jawa Timur
4. Data yang digunakan berasal dari data sekunder website resmi BPS, BNPB dan informasi yang terkait.

I.5 Manfaat Penelitian

1. Untuk Universitas Telkom, penelitian ini bermanfaat untuk memperbanyak literatur teori atau penelitian di bidang data mining khususnya K-Means *Clustering*.
2. Untuk pemerintah, penelitian ini dapat menjadi bahan sumber informasi untuk mengetahui kerawanan bencana alam banjir yang ada di Jawa Timur. Sehingga dapat memudahkan pemerintah untuk mengambil kebijakan yang tepat dalam menanggulangi bencana banjir.
3. Untuk penelitian selanjutnya, penelitian ini dapat sumber teori, acuan atau literatur untuk penelitian serupa. Serta, hasil dan metodologi yang digunakan dalam penelitian ini juga dapat menjadi bahan evaluasi untuk meningkatkan kualitas dan akurasi penelitian di masa mendatang. Dengan demikian, penelitian ini memiliki potensi untuk menjadi kontribusi yang berarti dalam pengembangan ilmu pengetahuan di bidang terkait.

I.6 Sistematika Penulisan

Uraian sistematika penulisan penelitian ini sebagai berikut :

Bab I Pendahuluan

Pada bab ini menjelaskan terkait latar belakang permasalahan, rumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian dan sistematika penulisan penelitian.

Bab II Tinjauan Pustaka

Bab ini berisi landasan teori mengenai penelitian terdahulu, Bencana Alam, Banjir, Curah Hujan, Luas Wilayah, Kepadatan Penduduk, CRISP-DM, Data Mining, *Standar Scaler*, *Clustering*, K-Means *Clustering*, *Elbow Method*, Visualisasi Data dan Alasan pemilihan metode. .

Bab III Metodologi Penelitian

Bab ini menjelaskan tentang metode yang digunakan, penelitian ini menggunakan alur penelitian CRISP-DM yang terdiri 6 alur yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Metode yang digunakan adalah K-Means *Clustering* untuk mengelompokkan data-data bencana banjir.

Bab IV Pengumpulan dan Pengolahan Data

Pada bab ini dibahas proses pengumpulan dan pengolahan data yang digunakan dalam penelitian. Data yang digunakan meliputi curah hujan, kepadatan penduduk, luas wilayah, dan jumlah kejadian banjir di Provinsi Jawa Timur, yang diperoleh dari sumber resmi seperti BPS dan BNPB. Dilakukan pembersihan data, standarisasi menggunakan metode *Standar Scaler* dan reduksi dimensi dengan PCA. Selanjutnya dilakukan proses pengelompokan wilayah menggunakan metode K-Means *Clustering*. Pengelompokan dengan tiga jenis perhitungan jarak, yaitu *Euclidean*, *Manhattan*, dan *Chebyshev*.

Bab V Hasil dan Pembahasan

Bab ini memuat hasil pengujian dan analisis terhadap pengelompokan tingkat kerawanan banjir menggunakan metode K-Means *Clustering* dengan tiga jenis perhitungan jarak, yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Chebyshev Distance*. Analisis mencakup penentuan jumlah *Cluster* optimal menggunakan Elbow Method, evaluasi kualitas *Clustering* menggunakan *Silhouette Coefficient*, dan interpretasi hasil visualisasi pengelompokan.

Bab VI Kesimpulan dan Saran

Bab ini berisi kesimpulan hasil penelitian yang menjawab rumusan masalah. Selain itu, diberikan saran untuk penelitian selanjutnya, seperti penggunaan variabel tambahan atau cakupan data dengan periode waktu yang lebih panjang, serta penerapan metode visualisasi yang lebih interaktif untuk mendukung pengambilan keputusan dalam mitigasi bencana.