

BAB II

LANDASAN TEORI

2.1 Penelitian Terkait

Beberapa penelitian sebelumnya telah mengaplikasikan teknik *Clustering* untuk memetakan daerah rawan banjir, menunjukkan efektivitas pendekatan ini dalam mengidentifikasi wilayah dengan tingkat risiko bencana yang tinggi.

Penelitian yang dilakukan oleh Xaverius Moruk (2023) menggunakan algoritma *K-Means* untuk mengelompokkan wilayah di Kabupaten Malaka ke dalam beberapa kategori risiko banjir, yaitu Sangat Rawan, Rawan, Terancam, Aman, dan Paling Aman. Studi ini menunjukkan bahwa setelah tiga iterasi, nilai cluster menjadi stabil dan tidak mengalami perubahan, yang mengindikasikan konsistensi dalam pengelompokan data. Temuan ini memperkuat keandalan algoritma *K-Means* dalam mengklasifikasikan wilayah berdasarkan tingkat kerawanan terhadap banjir, serta dapat menjadi referensi bagi pengambilan keputusan dalam upaya mitigasi bencana.

Selain itu, penelitian yang dilakukan oleh Adisanjaya (2021) juga menerapkan teknik *Clustering* untuk mengelompokkan wilayah berdasarkan pola curah hujan yang diamati selama periode tertentu. Wilayah dengan curah hujan tinggi dikategorikan sebagai daerah berisiko banjir lebih besar, sedangkan wilayah dengan curah hujan rendah dimasukkan ke dalam kelompok yang lebih aman. Hasil penelitian ini memberikan panduan yang lebih jelas bagi pihak berwenang dalam

merancang strategi mitigasi yang lebih terarah sesuai dengan tingkat kerawanan masing-masing wilayah.

Penelitian yang dilakukan oleh Febby Arisca Zurfani et al., (2024) menganalisis metode *Clustering K-Means* dalam zonasi daerah terdampak banjir di Kota Medan dengan evaluasi Silhouette Coefficient. Studi ini menunjukkan bahwa metode *K-Means* berhasil mengelompokkan daerah menjadi tiga kategori: aman, rawan, dan sangat rawan banjir. Hasil evaluasi menggunakan Silhouette Coefficient menghasilkan nilai 0,9, yang menunjukkan bahwa *Clustering* cukup signifikan. Namun, penelitian ini juga mencatat bahwa keakuratan model sangat bergantung pada kualitas data yang digunakan, termasuk data topografi dan catatan historis banjir.

Sementara itu, penelitian yang dilakukan oleh Makmun Effendi & Siswandi, (2024) menganalisis prediksi wilayah rawan banjir di Kabupaten Bekasi menggunakan algoritma *K-Means*. Studi ini membagi daerah terdampak menjadi tiga kategori: banjir tinggi, banjir sedang, dan banjir rendah. Hasil evaluasi menggunakan Davies-Bouldin Index menunjukkan nilai -0,452, yang mengindikasikan bahwa metode *Clustering* yang digunakan cukup efektif. Namun, penelitian ini juga mencatat bahwa keakuratan model sangat bergantung pada kualitas data, termasuk jumlah titik banjir dan curah hujan rata-rata

Dari berbagai studi yang ada, dapat disimpulkan bahwa penggunaan algoritma *K-Means* dalam analisis kerawanan banjir telah menunjukkan hasil yang efektif dan efisien dalam berbagai konteks geografis. Namun, meskipun banyak

penelitian telah menggunakan teknik ini, penerapannya di Kota Bengkulu dengan mempertimbangkan karakteristik topografi, pola curah hujan, serta faktor hidrologi lainnya masih sangat terbatas. Oleh karena itu, penelitian lebih lanjut diperlukan untuk menyesuaikan model *Clustering* yang digunakan agar lebih optimal dalam mengidentifikasi daerah berisiko tinggi di wilayah tersebut.

2.2 Banjir

Banjir merupakan bencana alam yang paling sering terjadi di Indonesia dan menjadi ancaman musiman ketika air meluap dari salurannya, menggenangi wilayah sekitarnya. Kejadian ini dapat diprediksi dengan memperhatikan curah hujan serta aliran air, namun dalam beberapa kasus, banjir terjadi secara mendadak akibat badai atau kegagalan tanggul yang dikenal sebagai banjir bandang (Pahleviannur *et al.*, 2023).

Dampak banjir sangat merugikan, baik dari aspek kemanusiaan maupun ekonomi. Penyebabnya beragam, termasuk curah hujan yang tinggi di atas normal, perubahan suhu, bobolnya tanggul atau bendungan, pencairan salju yang cepat, serta terhambatnya aliran air di tempat lain. Selain itu, faktor seperti permukaan tanah yang lebih rendah dari muka air laut, lokasi di daerah cekungan dengan resapan air minim, pembangunan di bantaran sungai, aliran sungai yang tersumbat sampah, serta berkurangnya tutupan lahan di hulu sungai turut memperparah risiko Banjir (Karim Harahap *et al.*, 2022).

Banjir menempati peringkat pertama dalam kategori kejadian bencana di Indonesia, dengan total 4.024 kejadian yang tercatat sejak tahun 1815 hingga 2012. Frekuensi banjir merupakan yang tertinggi, mencapai 39% dari total bencana,

disusul oleh angin puting beliung dan tanah longsor. Dalam periode yang sama, bencana banjir telah menyebabkan 18.569 korban meninggal, menjadikannya salah satu bencana dengan dampak paling besar terhadap kehidupan dan ekonomi masyarakat.

Banjir dapat berupa genangan pada lahan yang biasanya kering, seperti lahan pertanian, permukiman, atau pusat kota. Banjir juga dapat terjadi ketika debit atau volume air yang mengalir di sungai atau saluran drainase melebihi kapasitasnya. Luapan air umumnya tidak menjadi masalah jika tidak menimbulkan kerugian, korban jiwa, atau gangguan signifikan terhadap kehidupan sehari-hari. Namun, jika genangan air terjadi dalam jumlah besar, bertahan lama, dan sering berulang, hal ini dapat menghambat aktivitas masyarakat. Dalam sepuluh tahun terakhir, luas area terdampak dan frekuensi banjir terus meningkat, menyebabkan kerugian yang semakin besar.(Mohtana Kharisma Kadri *et al.*, 2024).

2.2.1 Faktor Penyebab Banjir

Penyebab banjir dapat dikategorikan menjadi dua jenis, yaitu banjir alami dan banjir akibat aktivitas manusia. Banjir alami dipengaruhi oleh faktor-faktor seperti curah hujan, kondisi fisiografi, erosi dan sedimentasi, kapasitas sungai dan drainase, serta pasang surut air (Dian Ramasari and ikhwan, 2024).

Sementara itu, banjir akibat aktivitas manusia terjadi karena perubahan lingkungan yang disebabkan oleh tindakan manusia, seperti perubahan kondisi Daerah Aliran Sungai (DAS), pembangunan permukiman di bantaran sungai, kerusakan sistem drainase, kegagalan infrastruktur pengendali banjir, deforestasi,

serta perencanaan sistem pengendalian banjir yang kurang tepat. Banjir dipengaruhi oleh berbagai faktor yang dapat dikategorikan menjadi faktor alami dan faktor akibat aktivitas manusia (antropogenik) (Nuhun *et al.*, 2024).

A. Faktor Alami

- Curah Hujan Tinggi – Hujan dengan intensitas tinggi dalam waktu lama meningkatkan volume air permukaan dan beban sistem drainase.
- Topografi dan Kondisi Tanah – Dataran rendah lebih rentan terhadap genangan, terutama jika sistem drainasinya buruk. Selain itu, tanah dengan daya serap rendah mempercepat aliran permukaan (runoff), meningkatkan risiko banjir.
- Perubahan Iklim – Anomali iklim global dapat memicu curah hujan ekstrem, memperpanjang musim hujan, dan meningkatkan frekuensi banjir.

B. Faktor Antropogenik

- Sistem Drainase Buruk – Saluran air yang tersumbat sampah atau tidak memiliki kapasitas yang memadai dapat menghambat aliran air dan memicu banjir di perkotaan.
- Alih Fungsi Lahan – Perubahan lahan resapan menjadi permukiman atau industri tanpa perencanaan tata ruang yang baik mengurangi daya serap tanah terhadap air hujan.
- Deforestasi – Penggundulan hutan tanpa reboisasi melemahkan kemampuan tanah dalam menyerap air, sehingga debit air meningkat dan memperbesar potensi banjir.

- Eksploitasi Air Tanah Berlebihan – Pengambilan air tanah yang tidak terkendali menyebabkan penurunan muka tanah, memperburuk risiko banjir, terutama di daerah pesisir yang juga terdampak oleh kenaikan permukaan laut (rob).

2.3 Curah Hujan

Curah hujan adalah jumlah air hujan yang jatuh pada suatu wilayah dalam periode waktu tertentu, yang diukur dalam milimeter (mm). Curah hujan merupakan elemen utama dalam sistem hidrologi, berperan penting dalam siklus air, pengisian sumber daya air tanah, serta mempengaruhi berbagai aspek kehidupan manusia dan lingkungan. Secara umum, curah hujan terjadi melalui proses kondensasi uap air di atmosfer yang membentuk awan, kemudian jatuh sebagai presipitasi ke permukaan bumi (Ruswanti, 2020).

Satuan curah hujan yang digunakan adalah milimeter (mm), di mana 1 mm berarti air yang terkumpul di permukaan datar seluas satu meter persegi setinggi 1 mm atau setara dengan satu liter air. Intensitas curah hujan mengacu pada jumlah hujan yang turun dalam rentang waktu tertentu, sering dinyatakan dalam mm/jam, mm/hari, atau mm/tahun. Data yang digunakan dalam analisis umumnya meliputi nilai maksimum, minimum, dan rata-rata untuk menggambarkan pola hujan di suatu wilayah. Proses pembentukan hujan dipengaruhi oleh berbagai faktor seperti suhu udara, kelembaban, tekanan atmosfer, serta interaksi antara daratan dan lautan (Rafi *et al.*, 2018).

Curah hujan diklasifikasikan berdasarkan durasi dan intensitasnya:

1. Berdasarkan Durasi

- Hujan Harian adalah Hujan yang diukur dalam periode 24 jam.
- Hujan Bulanan – Akumulasi curah hujan yang terjadi dalam satu bulan.
- Hujan Tahunan – Total curah hujan yang terjadi dalam satu tahun, sering digunakan untuk analisis iklim dan perencanaan sumber daya air.

2. Berdasarkan Intensitas

- Hujan Gerimis – Intensitas kurang dari 2,5 mm/jam.
- Hujan Sedang – Intensitas antara 2,5 mm/jam hingga 10 mm/jam.
- Hujan Lebat – Intensitas antara 10 mm/jam hingga 50 mm/jam.
- Hujan Ekstrem – Intensitas lebih dari 50 mm/jam.

Faktor-faktor yang mempengaruhi pola curah hujan di suatu daerah meliputi:

- Letak Geografis Posisi suatu wilayah terhadap garis lintang dan ketinggian mempengaruhi jumlah dan pola curah hujan.
- Angin Muson Angin muson barat dan muson timur berperan dalam menentukan musim hujan dan musim kemarau di Indonesia.
- El Niño dan La Niña Fenomena cuaca global yang dapat menyebabkan curah hujan lebih tinggi atau lebih rendah dari biasanya.
- Topografi Wilayah dengan pegunungan cenderung memiliki curah hujan lebih tinggi akibat efek orografis.

Pemantauan curah hujan menjadi aspek penting dalam mitigasi bencana hidrometeorologi seperti banjir, tanah longsor, dan kekeringan. Oleh karena itu,

data curah hujan yang akurat sangat diperlukan dalam berbagai bidang, termasuk pertanian, pengelolaan air, dan perencanaan kota.

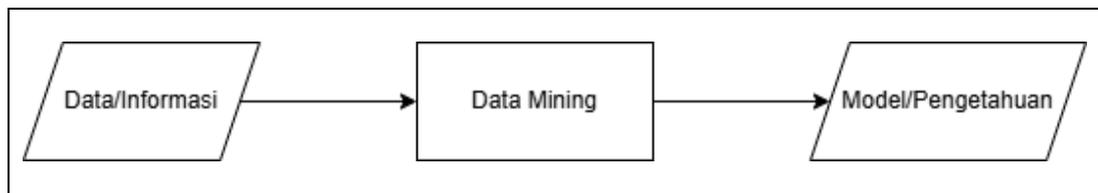
2.4 Clustering

Clustering adalah metode pengelompokan objek atau data berdasarkan tingkat kedekatan antar elemen-elemen yang ada. Perbedaan utama antara *Clustering* dan grup adalah bahwa dalam sebuah grup, anggota-anggota kelompok memiliki kondisi atau karakteristik yang sama; jika tidak sama, maka mereka tidak termasuk dalam grup tersebut. Sedangkan dalam *Clustering*, pengelompokan didasarkan pada kedekatan atau kesamaan karakteristik antar sampel, meskipun tidak harus identik. Salah satu teknik yang umum digunakan dalam *Clustering* adalah perhitungan jarak Euclidean, yang mengukur jarak antara dua titik dalam ruang berdimensi lebih dari satu (Amalina *et al.*, 2022).

Penerapan *Clustering* sangat luas dan bervariasi, karena sering digunakan untuk mengidentifikasi pola atau kesamaan dalam data yang tidak selalu memiliki keseragaman yang jelas. Dalam banyak kasus, baik dalam pengambilan keputusan maupun pemecahan masalah, data yang dianalisis tidak pernah sepenuhnya sama, namun cenderung menunjukkan kemiripan atau kedekatan dalam beberapa aspek. Oleh karena itu, *Clustering* menjadi metode yang sangat berguna untuk menggali wawasan atau informasi yang relevan dari data yang kompleks dan beragam (Najia, 2019).

2.5 Data Mining

Data mining adalah proses eksplorasi dan ekstraksi informasi serta pola yang berguna dari kumpulan data yang besar dan kompleks. Proses ini melibatkan beberapa tahapan, yaitu pengumpulan data, ekstraksi data, analisis data, dan penerapan statistik untuk mengolah data tersebut. Secara umum, data mining juga dikenal dengan istilah-istilah lain seperti *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, dan *information harvesting*. Tujuan utama dari keempat tahapan ini adalah untuk menghasilkan model atau pengetahuan yang dapat mendukung pengambilan keputusan yang lebih tepat (Al Faridzi *et al.*, no date).



Gambar 2. 1 Bagan Data Mining

Gambar 2.1 menggambarkan proses utama dalam data mining, yang terdiri dari tiga tahap: Data/Informasi, Data Mining, dan Model/Pengetahuan. Data/Informasi merupakan kumpulan data mentah yang diperoleh dari berbagai sumber. Kemudian, melalui tahap Data Mining, data ini dianalisis dan diproses menggunakan teknik eksplorasi pola, klasifikasi, clustering, atau prediksi. Hasil akhir dari proses ini adalah Model/Pengetahuan, yaitu pola atau informasi bermakna yang dapat digunakan untuk pengambilan keputusan yang lebih akurat.

Data mining dapat didefinisikan sebagai proses pemecahan kompleksitas sekumpulan data menjadi informasi yang tersembunyi (implisit) dan sebelumnya

tidak diketahui. Proses ini memanfaatkan alat otomatis atau semi-otomatis untuk menganalisis dan menggali pola dari data dengan tujuan menemukan pola yang relevan dan bermakna. Data mining merupakan bagian dari proses yang lebih besar yang dikenal dengan *knowledge discovery in databases* (KDD), yang melibatkan langkah-langkah tambahan untuk mengidentifikasi dan memanfaatkan pengetahuan yang terkandung dalam data (Pujiono, Astuti and Muhamad Basysyar, 2024).

Aplikasi data mining sangat luas dan mencakup berbagai bidang, seperti bisnis, kesehatan, ilmu sosial, hingga penelitian ilmiah. Teknik ini memungkinkan organisasi untuk mengungkap wawasan yang sebelumnya tersembunyi, mendukung prediksi, segmentasi pasar, dan pengambilan keputusan berbasis data yang lebih efisien dan akurat.

Sebagai sebuah bidang, data mining menggabungkan teknik dari berbagai disiplin ilmu, termasuk pembelajaran mesin, pengenalan pola, statistik, basis data, dan visualisasi, untuk menangani permasalahan dalam pengambilan informasi dari kumpulan data besar. Data mining juga dapat dipahami sebagai proses ekstraksi informasi baru yang diperoleh dari data besar, yang selanjutnya digunakan untuk pengambilan keputusan yang lebih baik. Sebagai tambahan, istilah data mining kadang juga merujuk pada *knowledge discovery*.

2.6 K-Means Clustering

K-Means Clustering merupakan salah satu metode *Clustering* non-hierarki yang digunakan untuk mengelompokkan data ke dalam k cluster berdasarkan kesamaan karakteristiknya. Data yang memiliki karakteristik serupa akan dikelompokkan dalam satu cluster, sedangkan data yang berbeda akan

dikelompokkan ke cluster lain. Dengan demikian, setiap cluster memiliki tingkat variasi yang rendah di dalamnya dan perbedaan yang lebih jelas antar cluster (Sibarani, 2018).

K-Means termasuk dalam metode unsupervised learning, yang berarti algoritma ini dapat bekerja dengan data yang tidak memiliki label kategori sebelumnya. Teknik ini sering digunakan dalam berbagai bidang, seperti analisis cuaca, pengelompokan pelanggan, pengolahan citra, serta pemetaan wilayah berdasarkan pola curah hujan untuk mitigasi bencana.

K-Means adalah algoritma yang termasuk dalam metode Unsupervised Learning karena menganalisis dataset berdasarkan pola tanpa memerlukan label pada data masukan. Tahapan *K-Means* adalah sebagai berikut (Buslim, 2020):

1. Menentukan jumlah cluster

Jumlah cluster yang digunakan harus lebih kecil dari jumlah kriteria yang tersedia.

2. Menentukan titik pusat cluster (centroid)

Centroid awal dipilih secara acak dari data atau berdasarkan metode tertentu, misalnya dengan mengambil nilai minimum, rata-rata, dan maksimum untuk tiga cluster.

3. Menghitung jarak antara titik data dan centroid

Pada tahap ini, perhitungan jarak dapat dilakukan menggunakan jarak Euclidean jika data memiliki lebih dari satu variabel. Namun, karena

penelitian ini hanya menggunakan satu variabel, yaitu curah hujan, maka perhitungan jarak dapat disederhanakan menggunakan jarak absolut:

$$d(x, c) = |x - c|$$

Yang dimana:

- $d(x-c)$ adalah jarak antara titik x dan centroid c
- x adalah nilai data curah hujan
- c adalah nilai centroid suatu cluster

4. Mengelompokkan data objek berdasarkan jarak minimum

Setelah perhitungan jarak dilakukan, setiap data akan dimasukkan ke dalam cluster dengan centroid terdekat.

5. Mengulangi tahap sebelumnya hingga centroid stabil

Setelah data dikelompokkan, centroid baru dihitung berdasarkan rata-rata nilai dalam setiap cluster.

6. Mengulangi proses hingga centroid stabil

Langkah-langkah sebelumnya diulang sampai tidak ada lagi perubahan dalam keanggotaan cluster, yang menandakan bahwa proses telah mencapai konvergensi

K-Means bekerja dengan cara menetapkan jumlah cluster, menentukan centroid awal, menghitung jarak antara data dan centroid menggunakan metode yang sesuai, lalu mengelompokkan data ke dalam cluster dengan jarak terdekat. Setelah itu, centroid diperbarui dengan menghitung rata-rata nilai dalam masing-

masing cluster. Proses ini berulang hingga tidak ada perubahan posisi centroid, menandakan bahwa algoritma telah selesai. (Saputra and Nataliani, 2021)

2.7 Python

Python adalah bahasa pemrograman tingkat tinggi yang dibuat oleh Guido Van Rossum dan dirilis pada tahun 1991. Bahasa ini memiliki filosofi perancangan yang menekankan keterbacaan kode, sehingga sintaksisnya sederhana dan mudah dipahami. Python juga bersifat dinamis dan multiguna, menjadikannya pilihan utama dalam berbagai bidang, termasuk Machine Learning dan Deep Learning (Inggit Agustina and Raditya Danar Dana, 2024).

Python memiliki pustaka standar yang luas dan komprehensif, memungkinkan pengembangan aplikasi yang mutakhir dengan kode yang tetap ringkas. Selain itu, karena bersifat open source, Python didukung oleh komunitas yang besar dan aktif. Popularitasnya terus meningkat seiring dengan kemampuannya dalam menangani berbagai kebutuhan pemrograman, mulai dari pengembangan web hingga analisis data (Ljubomir Perkovic, 2011).

Untuk menuliskan kode Python, Anda dapat menggunakan berbagai Integrated Development Environment (IDE), seperti VS Code, Sublime Text, dan PyCharm. Selain itu, tersedia juga IDE berbasis cloud seperti Jupyter Notebook dan Google Colab, yang memudahkan kolaborasi dan eksekusi kode langsung di web. Python sangat ideal untuk pengembangan algoritma Machine Learning, termasuk dalam tugas klasifikasi, regresi, dan *Clustering*. Kombinasi antara fleksibilitas, kelengkapan pustaka, serta dukungan komunitas yang kuat menjadikan Python

sebagai bahasa pemrograman yang terus berkembang dan semakin populer di berbagai bidang teknologi.

Python dianggap sebagai bahasa pemrograman yang paling banyak digunakan dalam bidang Machine Learning dan Deep Learning. Keunggulan Python terletak pada sintaksis yang mudah dipahami dan didorong oleh komunitas yang besar. Selain itu, Python memiliki banyak library yang mendukung pengembangan aplikasi di bidang ini. Beberapa library Python yang populer dan sering digunakan untuk Machine Learning dan Deep Learning antara lain (Riziq sirfatullah Alfarizi *et al.*, 2023):

1. NumPy Untuk perhitungan numerik. Library ini menyediakan fungsi matematika dengan kompleksitas tinggi, berguna untuk memproses array dan matriks multidimensi besar serta menangani aljabar linier, transformasi Fourier, dan bilangan acak.
2. Pandas Untuk manipulasi data. Pandas menyediakan struktur data yang cepat, fleksibel, dan ekspresif yang dirancang untuk bekerja pada data relasional atau berlabel, serta mendukung analisis data dunia nyata.
3. Scikit-learn Untuk implementasi algoritma *K-Means* dan berbagai algoritma pembelajaran mesin lainnya, baik yang diawasi maupun tidak diawasi. Scikit-learn dibangun di atas library seperti NumPy, Pandas, dan Matplotlib, yang memungkinkan pengolahan data dan penerapan algoritma dengan mudah.
4. Matplotlib dan Seaborn Untuk visualisasi hasil *Clustering*. Kedua library ini sangat berguna dalam membuat grafik yang membantu untuk menganalisis dan memahami hasil dari model Machine Learning dan Deep Learning.

Selain itu, Python juga memiliki berbagai library lain yang mendukung pengembangan aplikasi dalam Machine Learning dan Deep Learning, seperti SciPy, Theano, TensorFlow, Keras, dan PyTorch. Setiap library ini menawarkan kemampuan khusus yang memungkinkan pengguna untuk melakukan komputasi matematis yang kompleks, membangun dan mengoptimalkan model pembelajaran mesin, serta melakukan pemrosesan data dan visualisasi dengan lebih efisien.