

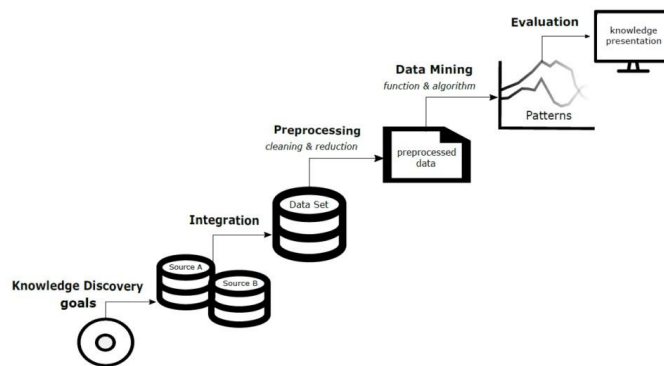
BAB II

KAJIAN PUSTAKA

A. Kajian Teoritis

1. *Data Mining*

Data mining merupakan suatu proses analisis data yang bertujuan untuk menggali informasi dari kumpulan data berukuran besar guna menemukan pola, hubungan, serta pengetahuan baru yang dapat dimanfaatkan dalam pengambilan keputusan. Seiring dengan perkembangan teknologi informasi, data *mining* kini mencakup proses pengumpulan, pengelolaan, dan analisis data digital untuk memahami fenomena yang terjadi atau membuat prediksi (Venturelli et al., 2025:445). Oleh karena itu, data *mining* memegang peran krusial dalam mengolah data perilaku, sehingga mampu menghasilkan keputusan yang lebih terstruktur dan berbasis bukti (*evidence-based*).



Gambar 2.1 Proses *Data Mining*

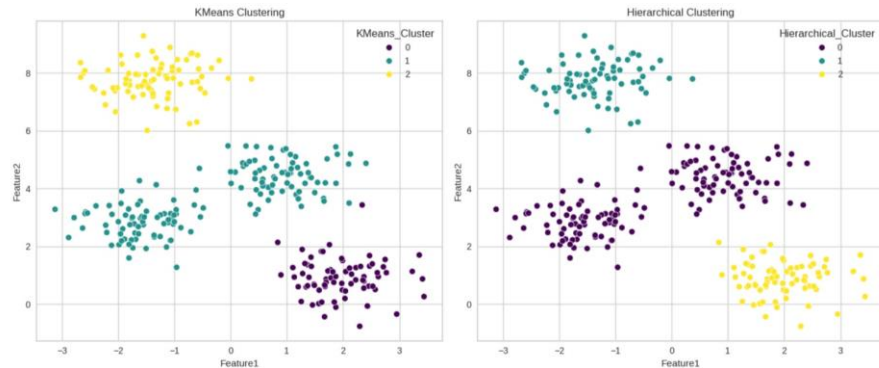
Sumber: Diadaptasi dari Sapuan et al., (2022)

Data mining memungkinkan identifikasi pola tersembunyi dari kumpulan data besar, sehingga data tidak terstruktur dapat diolah menjadi informasi bernilai (Papakyriakou & Barbounakis, 2022). Keberhasilan data mining sangat bergantung pada tahapan pra-pemrosesan data yang tepat guna memastikan kualitas input sebelum algoritma pemodelan dijalankan, sehingga pengetahuan yang diekstrak memiliki validitas dan tingkat akurasi yang tinggi.

Berdasarkan kedua uraian di atas, dapat disimpulkan bahwa *data mining* merupakan serangkaian proses sistematis untuk mengekstrak pola dan pengetahuan berharga dari himpunan data yang kompleks, di mana penerapannya tidak hanya berfokus pada tingkat akurasi komputasi, tetapi juga pada transparansi dan pemahaman model guna mendukung pengambilan keputusan yang terstruktur dan berbasis bukti.

2. *Clustering*

Clustering adalah teknik pengelompokan data tanpa label ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristiknya, sehingga data yang berada dalam satu kelompok memiliki kesamaan dibandingkan dengan data pada kelompok lainnya (Oyewole & Thopil, 2023:6439-6440). Teknik ini banyak dimanfaatkan dalam berbagai bidang karena kemampuannya dalam mengidentifikasi pola tersembunyi pada kumpulan data yang kompleks tanpa memerlukan informasi awal mengenai hubungan antar data.



Gambar 2.2 Ilustrasi *Clustering*

Sumber: Diadaptasi dari Mulyana et al., (2024)

Dalam perkembangannya, terdapat berbagai pendekatan clustering yang digunakan untuk meningkatkan akurasi dan efisiensi pengelompokan. Namun, kinerja *clustering* sangat tergantung pada pusat kluster awal yang tepat, sehingga memerlukan *pre-training* coba-coba (*trial-and-error*) secara manual yang merusak konsistensi hasil dan menyulitkan penerapannya (Liu et al., 2022).

Berdasarkan kedua uraian di atas, dapat disimpulkan bahwa *clustering* merupakan teknik pengelompokan data tanpa label berdasarkan tingkat kemiripan karakteristik untuk mengidentifikasi pola tersembunyi, di mana dalam perkembangannya kinerjanya sangat bergantung pada penentuan pusat *cluster* awal sehingga memerlukan *pre-training* yang tepat agar dapat menghasilkan kelompok data secara konsisten dan efisien.

3. Algoritma *K-Means*

Algoritma *K-Means* merupakan teknik *data mining* yang membagi data ke dalam beberapa kelompok berdasarkan tingkat kedekatan atau jaraknya (Setyaningtyas et al., 2022). Metode ini menggunakan pendekatan *unsupervised learning* karena tidak memerlukan label data dalam proses pengelompokannya dan mengelompokkan data berdasarkan kedekatannya terhadap *centroid*.

Dalam penerapannya, algoritma *K-Means* mengeksekusi serangkaian tahapan secara terstruktur, mulai dari inisialisasi pusat *cluster* hingga pencapaian kondisi konvergen. Selain itu, efektivitas hasil pemodelan *K-Means* sangat bergantung pada pemilihan jumlah *cluster* (k), sehingga proses pengelompokan dapat menggunakan metode *Elbow*, *Davies Bouldin Index* atau *Silhouette Score* untuk menentukan nilai k paling optimal (Orisa, 2022). Berikut tahapan algoritma *K-Means*:

a. Menentukan jumlah *cluster* (k)

Tahapan awal algoritma *K-Means* dimulai dengan menentukan jumlah *cluster* (k) yang akan dibentuk, yang mana nilai k ini menjadi parameter utama dalam menginisialisasi titik pusat *cluster* (*centroid*).

b. Inisialisasi *centroid* awal

Menentukan titik pusat *cluster* (*centroid*) awal sebanyak k secara acak dari *dataset* yang tersedia sebagai acuan awal pengelompokan.

- c. Menghitung jarak setiap data ke *centroid*

Menghitung jarak antara setiap data dengan masing-masing *centroid* menggunakan metode perhitungan jarak rumus *Euclidean Distance*.

Rumus *Euclidean Distance*:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2}$$

Keterangan:

d = Jarak

x = Data

c = *Centroid*

i = Indeks data

j = Indeks *cluster*

n = Jumlah atribut dalam setiap data

k = Indeks atribut

- d. Mengelompokkan data ke dalam *cluster* terdekat

Setiap data akan dimasukkan ke dalam *cluster* yang memiliki jarak paling dekat dengan *centroid* berdasarkan hasil perhitungan jarak.

- e. Menghitung ulang *centroid*

Setelah semua data dikelompokkan, *centroid* baru dihitung dengan cara mencari nilai rata-rata dari seluruh data dalam masing-masing *cluster*.

Rumus *centroid*:

$$c_j = \frac{1}{n} \sum_{i=1}^n x_i$$

Keterangan:

x = Data

c = *Centroid*

i = Indeks data

j = Indeks *cluster*

n = Jumlah atribut dalam setiap data

- f. Menghitung kembali jarak data ke *centroid* baru

Mengulangi proses perhitungan jarak antara setiap data dengan *centroid* yang telah diperbarui.

- g. Melakukan pengelompokkan ulang data

Data akan kembali dikelompokkan berdasarkan *centroid* terbaru untuk melihat apakah terjadi perubahan keanggotaan *cluster*.

- h. Melakukan iterasi hingga konvergen

Proses langkah c sampai g diulang secara terus-menerus hingga tidak ada perubahan anggota *cluster* atau perubahan sangat kecil dan jumlah iterasi maksimum tercapai.

i. Menghasilkan hasil akhir *clustering*

Setelah kondisi konvergen tercapai, diperoleh hasil akhir berupa kelompok *cluster* yang merepresentasikan pengelompokan data berdasarkan kemiripan karakteristik. Dalam sistem ini, kualitas dan pemilihan jumlah *cluster* optimal dievaluasi secara otomatis menggunakan dua metrik evaluasi, yaitu:

- 1) *Davies-Bouldin Index* (DBI): Mengukur rasio kedekatan data di dalam *cluster* dengan jarak antar *cluster*. Semakin kecil nilai DBI (mendekati 0), semakin baik kualitas pemisahan *clusternya*.

Rumus *Davies-Bouldin Index*:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right)$$

Keterangan:

DBI = *Davies-Bouldin Index*

k = Jumlah *cluster*

S_i = Rata-rata jarak data terhadap *centroid* pada *cluster* ke- i

S_j = Rata-rata jarak data terhadap *centroid* pada *cluster* ke- j

M_{ij} = Jarak antara *centroid* ke- i dan *cluster* ke- j

i, j = Indeks *cluster*

- 2) *Silhouette Score*: Mengukur kecocokan suatu data di dalam *clusternya* dibandingkan dengan *cluster* terdekat lain. Nilainya berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan pengelompokan yang sangat ideal.

Rumus *Silhouette Score*:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Keterangan:

$s(i)$ = Nilai *silhouette* data ke- i

$a(i)$ = Rata-rata jarak data ke- i dengan data lain

$b(i)$ = Rata-rata jarak data ke- i dengan *cluster* terdekat lainnya

$\max(a(i), b(i))$ = Nilai maksimum antara $a(i)$ dan $b(i)$

Berdasarkan pembahasan di atas, dapat disimpulkan bahwa algoritma *K-Means* merupakan metode *clustering* yang mengelompokkan data secara iteratif berdasarkan jarak *Euclidean* menuju *centroid*, di mana akurasi dan validitas struktur kelompok yang dihasilkan dapat dievaluasi secara objektif menggunakan kombinasi metrik *Davies-Bouldin Index* dan *Silhouette Score*.

4. Rancang Bangun

Menurut Setiawan et al., (2024), rancang bangun adalah proses mengimplementasikan hasil analisis menjadi kode program guna membangun sistem baru atau memperbaiki sistem lama. Tahapan ini berfokus pada

transformasi gagasan konseptual menjadi bentuk aplikasi fungsional yang siap diimplementasikan secara teknis.

Sejalan dengan pendapat tersebut, Pratiwi et al., (2024) menjelaskan bahwa rancang bangun merupakan tahapan perencanaan yang menyusun alur kegiatan secara sistematis dalam pengembangan suatu program atau aplikasi. Melalui tahapan ini, hasil evaluasi sistem diwujudkan menjadi kode program yang nyata guna menghasilkan aplikasi yang lebih tepat guna bagi penggunaanya.

Di sisi lain, perancangan ini dasarnya bermula dari penyusunan sketsa atau gambaran awal yang belum pernah ada sebelumnya. Rencana awal tersebut kemudian diolah dan dikembangkan secara bertahap hingga menjadi sebuah sistem utuh yang berfungsi sesuai tujuan. Proses ini tidak sekadar membuat ide di atas kertas, tetapi juga menyempurnakan bentuk desain agar implementasinya nanti berjalan lebih efektif dan efisien. Oleh karena itu, rancang bangun juga dapat didefinisikan sebagai fase awal pemetaan sketsa mentah menjadi rancangan yang berorientasi pada fungsi dan fungsionalitas sistem (Ariansyah & Wijaya, 2021).

Berdasarkan ketiga uraian di atas, dapat disimpulkan bahwa rancang bangun merupakan serangkaian proses sistematis dan terencana yang dimulai dari penyusunan gambaran awal hingga penerjemahan hasil analisis ke dalam bahasa pemrograman, guna menghasilkan perangkat lunak baru atau pembaruan sistem yang efektif, efisien, dan sesuai kebutuhan pengguna.

5. Sistem Informasi

Sistem informasi mengintegrasikan berbagai subsistem untuk mengolah data menjadi informasi, yang jika dikendalikan dan dievaluasi dengan baik, dapat dimanfaatkan sebagai dasar pengambilan keputusan (Effendi et al., 2023). Hal ini dipertegas oleh Azis, (2022:20) yang menyatakan bahwa sistem informasi dapat mengolah data dengan cepat, akurat, dan efisien untuk meningkatkan kualitas pelayanan, kinerja perusahaan, serta ketepatan pengambilan keputusan.

Dari sudut pandang fungsional, (Adham, 2024) menjelaskan bahwa sistem informasi berperan dalam mengoptimalkan rangkaian proses pengumpulan, pengolahan, penyimpanan, hingga pendistribusian data secara efektif dan efisien. Komponen yang menyusun sebuah sistem informasi terdiri atas perangkat keras dan lunak, jaringan, data, penggunaanya, serta prosedur kerja yang saling terhubung untuk mencapai efisiensi pengelolaan data.

Pada penerapannya, pembangunan sistem informasi melibatkan antara manusia dan teknologi dalam organisasi untuk mengolah data menjadi informasi bermanfaat, memenuhi kebutuhan transaksi, serta mengendalikan jaringan komunikasi (Arribe & Ryandi, 2023). Melalui kesatuan unsur tersebut, output informasi yang dihasilkan dapat didistribusikan dan dimanfaatkan secara efektif oleh pihak-pihak yang membutuhkan.

Berdasarkan ketiga uraian di atas, dapat disimpulkan bahwa sistem informasi merupakan kesatuan kerangka terintegrasi yang menggabungkan teknologi, manusia, dan prosedur kerja secara terencana untuk mengolah data mentah menjadi informasi yang akurat, relevan, serta mudah didistribusikan guna mendukung efisiensi operasional dan pengambilan keputusan organisasi.

6. Pemetaan Wilayah Provinsi Jawa Timur

Proses pengelompokan karakteristik wilayah tingkat regional menjadi instrumen penting untuk mendukung efisiensi program pemerintah. Pengelompokan kabupaten/kota ke dalam beberapa *cluster* menggunakan kemiripan sifat sangat krusial dalam menentukan sasaran kebijakan pembangunan dan alokasi sumber daya secara tepat. Tujuan pengelompokan wilayah kabupaten/kota menggunakan kesamaan karakteristik ke dalam *cluster* tertentu yaitu untuk menyajikan analisis data yang komprehensif (Mawardi & Saputri, 2025).

Penggambaran data geografis secara visual juga efektif digunakan sebagai alat bantu utama untuk menyajikan informasi sektoral secara nyata di suatu daerah. Pemetaan adalah proses visualisasi dan pengelompokan data wilayah ke dalam bentuk peta, di mana variasi *cluster* ditandai dengan warna berbeda dan disesuaikan dengan skala peta untuk menggambarkan kondisi yang sebenarnya. (Aisah et al., 2022).

Berdasarkan kedua uraian di atas, dapat disimpulkan bahwa pengelompokan dan pemetaan wilayah secara visual merupakan instrumen strategis yang menggabungkan metode klasifikasi data dan tampilan bergradasi warna, guna menggambarkan kondisi nyata daerah serta mempermudah pemerintah maupun pemangku kepentingan dalam menentukan sasaran kebijakan pembangunan yang terukur dan tepat sasaran.

7. *Open Data Analytics*

Open data analytics kini menjadi fondasi penting dalam era transformasi digital karena memfasilitasi transparansi dan kolaborasi informasi secara luas. Saat menyatukan beberapa sumber data, menjaga konsistensi dan keandalan data menjadi kuncinya. Proses ini tidak mudah karena sering dihadapkan pada masalah perbedaan kosa kata dalam data, tidak adanya pengidentifikasi umum, hingga minimnya informasi (Flausino et al., 2024). Integrasi *Open Data Analytics* mempercepat pengambilan keputusan sekaligus membuka peluang inovasi yang lebih inklusif di sektor industri.

Dengan memanfaatkan *Open data analytics* ini, berbagai organisasi mulai dari bisnis lokal hingga instansi pemerintah kini dapat mengolah data besar tanpa terhalang biaya lisensi yang mahal. Kemudahan akses ini menyamakan kesempatan bagi semua pihak untuk berinovasi dan memecahkan masalah secara bersama-sama. Pada akhirnya, keterbukaan tersebut melahirkan keputusan yang lebih cerdas, transparan, dan berdampak langsung pada masyarakat.

8. *Website*

Website merupakan media di internet yang memuat berbagai data dan informasi untuk memudahkan pengguna dalam melakukan pencarian (Mongkau et al., 2023). Kehadirannya memungkinkan pertukaran informasi dapat terjadi secara cepat, fleksibel, dan lintas wilayah tanpa dibatasi oleh batasan geografis.

Untuk mengakses *website*, pengguna memerlukan *browser* sebagai sarana untuk menelusuri halaman yang tersedia. *Website* didefinisikan sebagai aplikasi media interaktif berisi dokumen multimedia seperti teks, gambar, hingga video yang memanfaatkan protokol HTTP dan diakses melalui perangkat lunak menggunakan *browser* (Nurlailah & Wardani, 2023). Fleksibilitas ini menjadikan platform berbasis *website* sebagai sarana utama penyebaran informasi di era digital modern.

Menurut Rahmawati & Sumarsono (2024), Perkembangan *website* kini telah bertransformasi dari media informasi statis menjadi media yang menyajikan interaksi secara lebih aktif. Dalam perkembangannya, peranan *website* kini telah bertransformasi tidak sekadar sebagai media bacaan yang pasif, melainkan menjadi *platform* yang lebih hidup serta interaktif bagi para penggunanya. Dengan demikian, *website* terdiri dari sejumlah halaman yang saling terintegrasi dan dapat diakses secara global, sehingga berperan sebagai sarana komunikasi dan penyampaian informasi yang efektif serta mudah dijangkau oleh masyarakat.

Berdasarkan ketiga uraian di atas, dapat disimpulkan bahwa *website* merupakan platform media berbasis internet yang saling terhubung melalui *hyperlink* dan dapat diakses melalui peramban (*browser*), yang berfungsi sebagai sarana penyampaian informasi interaktif serta komunikasi global yang efisien dan mudah dijangkau oleh berbagai kalangan pengguna.

a. *Python*

Menurut Suharto (2023:1), *Python* merupakan salah satu bahasa pemrograman tingkat tinggi yang bersifat dinamis dan menggunakan sistem interpreter dalam proses eksekusinya. Bahasa pemrograman ini mendukung konsep pemrograman berorientasi objek serta menyediakan berbagai struktur data tingkat tinggi yang memudahkan pengembangan aplikasi. *Python* dikenal sebagai bahasa pemrograman yang mudah dipelajari namun tetap memiliki kemampuan yang kuat dan fleksibel, sehingga banyak digunakan dalam berbagai kebutuhan pengembangan perangkat lunak. Selain dimanfaatkan sebagai media pembelajaran pemrograman, *Python* juga diterapkan secara luas dalam berbagai bidang, seperti bisnis, kesehatan, sains, media, hingga teknologi modern seperti pengendalian robot dan pembuatan animasi film. Proses eksekusi pada *Python* dilakukan melalui interpreter yang menerjemahkan kode program secara langsung ke dalam bahasa mesin saat program dijalankan, sehingga menjadikan *Python* sebagai bahasa yang efisien dan praktis dalam pengembangan aplikasi.

Berdasarkan kajian literatur tersebut, dapat disimpulkan bahwa *Python* merupakan bahasa pemrograman tingkat tinggi yang mudah dipahami, fleksibel, serta memiliki dukungan fitur yang lengkap, sehingga sangat sesuai digunakan dalam pengembangan berbagai sistem dan aplikasi berbagai teknologi informasi.

b. *React (Vite)*

Menurut Chhetri (2024:5), Kini *React* mendominasi sebagai pustaka *frontend* yang paling banyak dipakai, khususnya untuk mengembangkan aplikasi *website* yang membutuhkan daya tampung trafik tinggi. Ketika pustaka tersebut dikolaborasikan dengan *Vite* sebagai instrumen *build tool* modern, ekosistem ini mampu memangkas waktu inisiasi server lokal melalui optimalisasi modul berbasis *Native ESM*.

Keunggulan operasional dari pemanfaatan teknologi mutakhir ini terletak pada efisiensi fitur *Hot Module Replacement* (HMR) yang berjalan sangat presisi selama aktivitas rekayasa kode berlangsung. Setiap modifikasi pada baris skrip komponen visual akan langsung direfleksikan ke peramban secara instan tanpa perlu melakukan penyegaran halaman secara menyeluruh. Pada fase akhir pengembangan, sistem ini akan mengeksekusi pembundelan seluruh aset proyek menggunakan *Rollup* guna menghasilkan dokumen produksi yang sangat ringan, teroptimasi, dan memiliki latensi rendah saat diakses oleh pengguna.

c. *Tailwind CSS*

Menurut Salmi (2023), Keunggulan utama yang membedakan *Tailwind CSS* dari *framework* lainnya terletak pada fleksibilitas kustomisasi yang tinggi. *Framework* ini memberikan kebebasan penuh bagi pengembang untuk merancang halaman *website* tanpa terikat pada *style* bawaan. Pendekatan ini memungkinkan penerapan properti visual secara langsung pada struktur HTML.

Tailwind CSS didefinisikan sebagai *framework* berbasis *utility-first* yang memfasilitasi perancangan antarmuka kustom secara efisien (Rifandi et al., 2022). Integrasi sistem kisi dan *flexbox* bawaan yang dikelola mempermudah standarisasi elemen visual seperti jarak, tipografi, hingga skema warna dalam satu ekosistem pengembangan. Kemudahan ini menjadikannya pilihan ideal dalam mempercepat pengembangan *website* dari skala purwarupa hingga produksi.

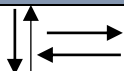
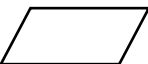
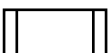
Berdasarkan kedua uraian di atas, dapat disimpulkan bahwa *Tailwind CSS* merupakan *framework* berbasis kelas utilitas dan konfigurasi terpusat yang mempermudah standarisasi elemen visual serta pembuatan tata letak responsif secara efisien, sehingga mampu mempercepat siklus pengembangan *website* secara signifikan.

9. *Flowchart*

Flowchart merupakan instrumen visual yang menyederhanakan alur kerja serta konsep kompleks di berbagai bidang, dengan merangkum informasi rumit menjadi representasi visual yang ringkas guna mengoptimalkan pemahaman dan komunikasi (Pan et al., 2024).

Dalam perancangan sistem, *flowchart* dapat diterapkan untuk memetakan alur identifikasi secara terstruktur. Pengembang dapat menyusun urutan kerja secara runtut mulai input data, *preprocessing*, proses *clusterisasi* dengan algoritma *K-Means*, hingga visualisasi hasil ke dalam peta digital. Pemodelan ini mengefisiensikan alur kerja, mempermudah evaluasi logika dan pelacakan kesalahan, diagram ini menggambarkan urutan operasi secara sistematis menggunakan alur panah serta simbol-simbol standar, seperti bentuk persegi panjang untuk merepresentasikan setiap tahapan proses.

Tabel 2.1 Simbol *Flowchart*

No	Simbol	Nama	Fungsi
1		<i>Flow</i>	Menjelaskan aliran proses dalam sebuah sistem atau program
2		<i>Terminal</i>	Menunjukkan awal dan akhir suatu proses atau program
3		<i>Process</i>	Menyatakan suatu proses
4		<i>Input / Output</i>	<i>Input / Output</i> data
5		<i>Decision</i>	Menyatakan suatu kondisi
6		<i>Predefine Process</i>	Menyatakan sub-program

7		<i>Preparation</i>	Inisialisasi / pemberian nilai awal
8		<i>Display</i>	Menyatakan alat <i>output</i>
9		<i>Document</i>	<i>Input / Output</i> berbentuk dokumen
10		<i>Manual Operation</i>	Proses manual
11		<i>On Page Connector</i>	Penghubung bagian dalam satu halaman
12		<i>Off Page Connector</i>	Penghubung bagian berbeda halaman

Sumber: Tarsini & Anggraeni (2024)

10. *Unified Modeling Language* (UML)

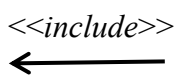
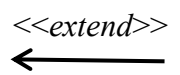
Menurut Hidayati et al., (2023), *Unified Modeling Language* (UML) memiliki berbagai keunggulan, seperti menyediakan bahasa pemodelan visual yang ekspresif untuk berbagai bahasa pemrograman maupun proses rekayasa umum, menyatukan informasi terbaik dalam pemodelan, serta mendukung pemodelan visual. Selain itu, UML mempermudah pengembang dalam memahami sistem dan berfungsi sebagai rekam cetak biru (*blueprint*) yang memberikan panduan detail untuk tahap penyusunan kode program.

UML berperan penting dalam membantu proses pengembangan perangkat lunak agar lebih sistematis, terorganisir, dan mudah dipahami oleh seluruh pihak yang terlibat. Dengan adanya UML, pengembang dapat mendokumentasikan kebutuhan sistem, alur proses, hingga hubungan antar komponen secara lebih jelas sehingga mendukung proses analisis, implementasi, dan pengembangan sistem secara efektif.

a. *Use Case Diagram*

Use case diagram berfungsi untuk menggambarkan fungsi utama yang dijalankan oleh sistem, serta menjelaskan bagaimana aktor dan sistem saling berinteraksi (Ramdany et al., 2024). Diagram ini memvisualisasikan fungsi sistem agar pengguna, klien, dan pengembang lebih mudah memahami kebutuhan serta cara kerja sistem saat dirancang. *Use case diagram* membantu proses komunikasi antar pihak yang terlibat dengan cara merepresentasikan fungsionalitas sistem secara lebih jelas dan terstruktur.

Tabel 2.2 Simbol *Use Case Diagram*

No	Simbol	Nama	Keterangan
1		Aktor	Mewakili peran orang
2		<i>Use Case</i>	Abstraksi antara sistem dan aktor
3		<i>Association</i>	Penghubung <i>use case</i> dengan aktor
4		Generalisasi	Spesialisasi aktor untuk berhubungan dengan <i>use case</i>
5		<i>Include</i>	Simbol yang menunjukkan bahwa suatu <i>use case</i> mengeksekusi fungsi dari <i>use case</i> lain, sebagai bagian proses tersebut.
6		<i>Extend</i>	Simbol yang menunjukkan bahwa suatu <i>use case</i> dapat atau menambahkan fungsi ke <i>use case</i> lain jika kondisi tertentu terpenuhi.

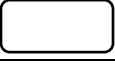
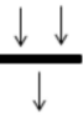
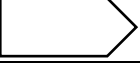
Sumber: Harlina et al., (2025)

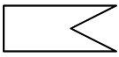
b. *Activity Diagram*

Activity diagram berfungsi untuk menggambarkan jalannya alur kerja dan aktivitas yang terjadi di dalam suatu sistem maupun proses bisnis (Rahmawati et al., 2023). Gambaran visual mengenai jalannya proses di dalam sistem ditunjukkan melalui *activity diagram*, yang menjelaskan tahapan sistem dalam memproses perintah serta menampilkan output-nya secara sistematis.

Tahapan alur kerja mulai dari proses awal, percabangan atau pengambilan keputusan yang dapat terjadi selama proses berjalan, hingga tahapan akhir dari aktivitas sistem.

Tabel 2.3 Simbol *Activity Diagram*

No	Simbol	Nama	Keterangan
1		<i>Start</i>	Status awal merupakan awal dari aktivitas sistem.
2		<i>Activity</i>	Bertujuan untuk menggambarkan aktivitas dari sebuah sistem
3		<i>Connector</i>	Ini menunjukkan alur arah atau alur kontrol suatu aktivitas.
4		<i>Joint / Sinkronisasi</i> <i>ion bar</i>	Penggabungan digunakan untuk menghubungkan aktivitas satu dengan aktivitas yang lain.
5		<i>Fork</i>	Dua operasi bersamaan dipisahkan dari satu aliran aktivitas utama.
6		<i>Decision</i>	<i>Decision</i> / percabangan merupakan aktivitas yang memiliki lebih dari satu kondisi atau pilihan.
7		<i>Send Signal</i>	Menyampaikan ke aktivitas penerima bahwa suatu sinyal sedang dikirim.

8		<i>Receive Signal</i>	Menunjukkan bahwa suatu peristiwa telah diterima. Alur yang dihasilkan dari tindakan ini selesai setelah peristiwa diterima.
9		<i>Option Loop</i>	Hal ini memberi desainer kemampuan untuk menggambarkan urutan berulang didalam simbol <i>loop</i> .
10		<i>Flow final</i>	Menandakan akhir dari suatu aliran proses tertentu. akhir suatu proses harus ditandai dengan simbol akhir aliran.
11		<i>End</i>	Menandakan akhir suatu aktivitas dan akhir semua aliran proses.

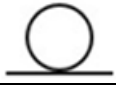
Sumber: Gedam & Meshram (2023)

c. *Sequence Diagram*

Sequence diagram menggambarkan alur interaksi dan komunikasi pesan yang terjadi di antara sejumlah objek secara berurutan (Sutrisno & Karnadi, 2021). *Sequence diagram* berfungsi untuk menjelaskan interaksi antar objek melalui pengiriman pesan sebagai respon terhadap suatu tindakan, sehingga urutan proses yang terjadi dalam sistem dapat digambarkan secara sistematis dan mudah dipahami.

Berdasarkan uraian di atas, dapat disimpulkan bahwa *sequence diagram* merupakan alat pemodelan visual yang krusial untuk menggambarkan alur komunikasi dan pertukaran pesan antar komponen sistem secara kronologis, sehingga mempermudah pemahaman mengenai urutan eksekusi proses dalam suatu perangkat lunak.

Tabel 2.4 Simbol *Sequence Diagram*

Simbol	Elemen	Keterangan
	<i>Actor</i>	Menggambar ilustrasi seseorang yang menggunakan suatu sistem.
	<i>Entity Class</i>	Menggambarkan hubungan yang akan terjadi.
	<i>Boundary Class</i>	Mengelola pertukaran informasi antar lingkungan sistem.
	<i>Control Class</i>	Menjelaskan hubungan antara <i>boundary</i> dan tabel.
	<i>A Focus of Control & A Life Line</i>	Menjelaskan titik awal dan akhir <i>message</i> tersebut.
	<i>A Message</i>	Menjelaskan cara mengirim pesan.

Sumber: Gunawan et al., (2023)

B. Kajian Empiris

Beberapa penelitian terdahulu menunjukkan bahwa algoritma *K-Means* sangat efektif untuk memetakan karakteristik wilayah berdasarkan indikator pembangunan daerah. Penelitian-penelitian tersebut menjadi landasan penting dalam mengintegrasikan *clusterisasi* ke dalam *website open-data analytics*. Melalui pendekatan ini, sistem yang dibangun dapat mengolah kombinasi variabel numerik data dari portal *open-data* indikator daerah secara dinamis dan otomatis, dengan syarat minimal menyertakan 2 variabel numerik dalam satu aspek atau indikator yang sejenis dan tidak dapat digabungkan dengan aspek yang berbeda.

Penelitian yang dilakukan oleh Deviana & Putro (2022) berjudul “Pengelompokan Wilayah di Indonesia Berdasarkan Indikator Kerawanan Ekonomi Pasca Pandemi Covid-19 (*K-Means Cluster Algorithm*)” membahas masalah ketimpangan pemulihan ekonomi antar-daerah setelah pandemi. Penelitian ini bertujuan untuk mengelompokkan tingkat kerawanan ekonomi wilayah di Indonesia menggunakan algoritma *K-Means Clustering* guna menentukan daerah mana yang perlu diprioritaskan. Metode yang digunakan adalah penelitian kuantitatif, di mana jumlah kelompok diuji untuk menemukan hasil yang paling pas. Data yang diolah diambil dari BPS mengenai ekonomi makro daerah. Hasilnya, penelitian ini sukses memetakan wilayah ke dalam beberapa kelompok kerawanan ekonomi yang berbeda, sehingga metode *K-Means* terbukti sangat cocok untuk membantu menyusun rencana kebijakan daerah.

Penelitian selanjutnya oleh Fadila & Putri (2023) dengan judul “Analisis Perkembangan Ketahanan Pangan di Indonesia: Pendekatan Menggunakan *Big Data* dan *Data Mining*” membahas pentingnya pemenuhan kebutuhan pangan nasional yang bergantung pada sektor pertanian lokal. Penelitian ini bertujuan untuk melihat tren sekaligus mengelompokkan stabilitas ketahanan pangan nasional menggunakan gabungan metode *Long ShortTerm Memory* (LSTM) dan *K-Means Clustering*. Metode yang dipakai adalah eksplorasi *big data* dengan memanfaatkan data dari *Google Trends* untuk melihat pergerakan ketahanan pangan masyarakat. Karena belum ada penelitian terdahulu yang menghubungkan *Google Trends* dengan ketahanan pangan, penetapan kata kunci pencarian

disesuaikan secara khusus berdasarkan indikator dan deskripsi masing-masing aspek. Hasilnya menunjukkan bahwa pengelompokan *K-Means* berhasil membagi tingkat ketahanan pangan daerah ke dalam zonasi kelompok yang rapi, sehingga efektif digunakan untuk mendeteksi dini masalah logistik pangan.

Penelitian yang dipublikasikan oleh Asyafiiyah et al., (2025) berjudul “Analisis Wilayah Prioritas Pembangunan di Provinsi Jawa Timur Berdasarkan Indikator Sosial Menggunakan Metode *K-Means Clustering*” menyoroti masalah ketimpangan capaian pembangunan antar kabupaten/kota di Jawa Timur. Penelitian ini bertujuan untuk mengelompokkan wilayah yang perlu diprioritaskan pembangunannya dengan mengukur kemiripan kondisi sosial memakai algoritma *K-Means Clustering*. Algoritma *K-Means Clustering* mengelompokkan data dengan cara menghitung rata-rata posisi objek pada masing-masing *cluster*. Data yang dianalisis mencakup indikator sosial yang berpengaruh bagi pembangunan wilayah kabupaten/kota di Jawa Timur yaitu jumlah penduduk, laju pertumbuhan penduduk, tingkat kemiskinan, Indek Pembangunan Manusia (IPM), Tingkat Pengangguran Terbuka (TPT), dan rata-rata lama sekolah. Hasilnya menunjukkan bahwa metode *K-Means Clustering* teruji efektif sebagai instrumen pendukung keputusan untuk menetapkan kawasan prioritas pembangunan berbasis data.

Penelitian oleh Mawardi & Saputri (2025) dengan judul “Pemetaan Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Luas Panen, Produksi, dan Produktivitas Beras” membahas pengelompokan Kabupaten/Kota di Jawa Timur yang dilakukan berdasarkan tiga variabel, yaitu produksi beras, luas lahan panen,

serta produktivitas. Penelitian ini bertujuan untuk memetakan potensi sektor pertanian secara menyeluruh guna mendorong optimalisasi hasil produksi beras di Jawa Timur. Metode yang digunakan adalah metode *K-Means Clustering*. Data yang digunakan adalah data sekunder yang bersumber dari Badan Pusat Statistik Provinsi Jawa Timur yakni Produksi Beras dan Luas Panen Padi. Hasil penelitian ini, peningkatan produktivitas beras difokuskan pada daerah berproduktivitas rendah yang berada di sekitar wilayah berproduktivitas tinggi.

Penelitian kelima yang dilakukan oleh Wahyudi et al., (2025) berjudul “*Clustering Regencies in Indonesia for Regional Mapping Using the K-Means Algorithm*” membahas masalah dalam memetakan kondisi ekonomi kabupaten/kota secara menyeluruh berdasarkan data PDRB 2024 terbaru. Penelitian ini menerapkan metode *K-Means clustering* guna mengelompokkan kabupaten/kota di Indonesia berdasarkan PDRB 2024 secara objektif, sekaligus menganalisis pengaruh hasil pengelompokan tersebut bagi perencanaan pembangunan daerah. Metode penelitian menggunakan algoritma *K-Means clustering* yaitu sebuah pendekatan *unsupervised learning* yang diterapkan untuk mengelompokkan data berdasarkan kesamaan karakteristiknya. Hasil penelitian menunjukkan bahwa pengelompokan *K-Means* ini berhasil melakukan pengelompokan kabupaten dan kota di Indonesia berdasarkan data PDRB tahun 2024. Hasil ini sangat berguna untuk membantu pemerintah menyalurkan program bantuan pembangunan agar lebih tepat sasaran.

Berdasarkan kelima penelitian di atas, dapat disimpulkan bahwa metode *K-Means Clustering* adalah alat analisis yang paling sering diandalkan untuk mengelompokkan karakteristik dan potensi wilayah. Metode ini mampu mengolah himpunan data berukuran besar berdasarkan kemiripan karakternya secara efektif. Tahapan pengerjaannya meliputi pembersihan data (*preprocessing*), pengubahan variabel, eksekusi fungsi pengelompokan, hingga menguji kualitas kelompok data tersebut. Secara umum, hasil dari studi-studi ini membuktikan bahwa algoritma *K-Means* mampu menyederhanakan struktur data yang kompleks menjadi kelompok wilayah yang terstruktur, sehingga sangat andal untuk membantu mengambil keputusan strategis berdasarkan data ilmiah.

Adapun perbedaan mendasar penelitian ini dengan penelitian-penelitian sebelumnya terletak pada pengembangan *website open-data analytics* yang mengintegrasikan proses komputasi algoritma secara langsung di dalam *server*. Selain itu, *Website open-data analytics* dirancang secara fleksibel untuk menerima pengunggahan variabel numerik dari *dataset open-data*, dengan syarat minimal menyertakan 2 variabel numerik yang berada dalam satu aspek yang sama. Hal ini bertujuan agar sistem dapat melakukan pengelompokan multidimensi secara dinamis sekaligus merender *Grouped Bar Chart* untuk visualisasi perbandingan detail indikator tiap daerah terhadap rata-rata *clusternya*. Sistem ini juga dilengkapi dengan visualisasi peta interaktif yang memanfaatkan penomoran Kode Wilayah guna mencegah kesalahan pembacaan data, fitur navigasi pada peta untuk mengakses halaman detail, serta grafik DBI & *Silhouette Score*.

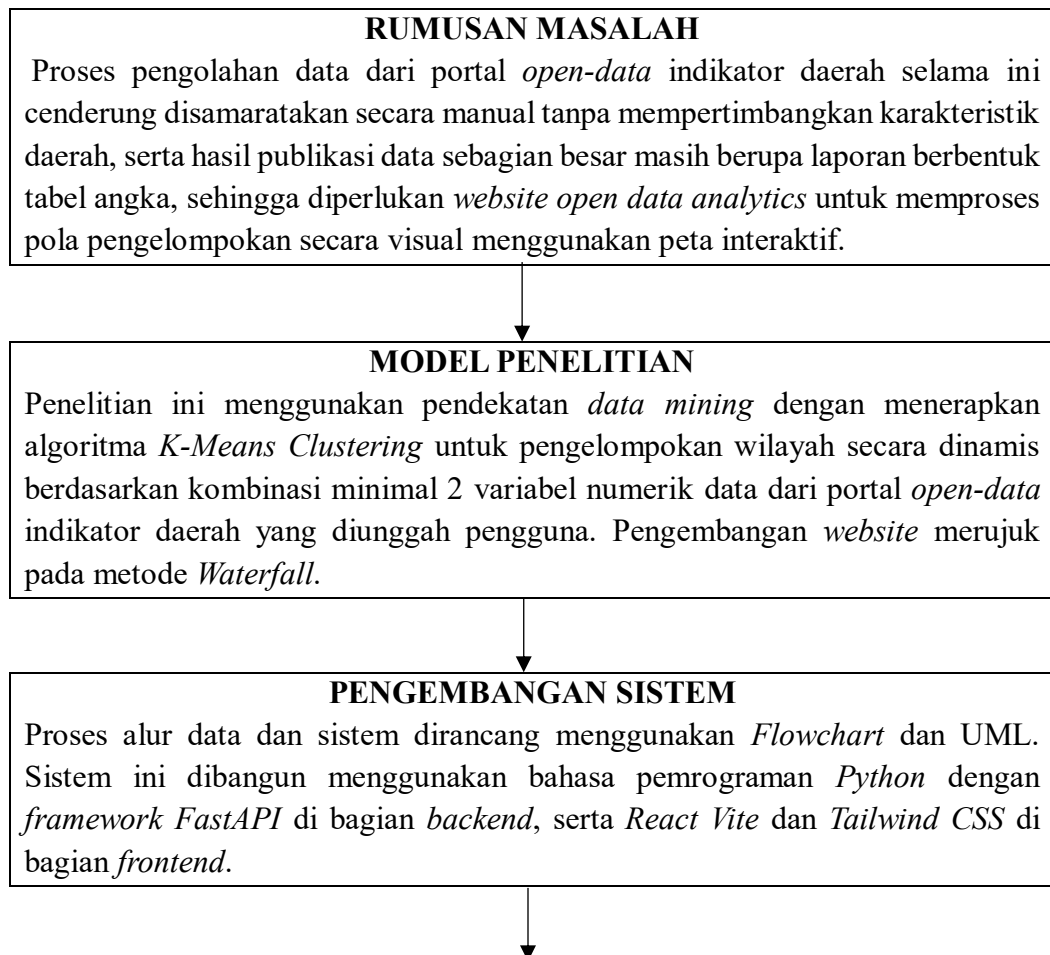
C. Kerangka Berpikir

Permasalahan utama yang mendasari penelitian ini adalah proses pengolahan data dari portal *open-data* indikator daerah Provinsi Jawa Timur yang masih disajikan secara manual dalam bentuk tabel-tabel angka yang terpisah. Kondisi tersebut menyebabkan data sulit dipetakan secara cepat, proses pengelompokan karakteristik daerah memerlukan waktu yang lama, serta membatasi aksesibilitas bagi masyarakat umum, akademisi, maupun analis pemula untuk menarik kesimpulan.

Untuk mengatasi permasalahan tersebut, diperlukan suatu *website open-data analytics* yang mampu mengelompokkan wilayah di Provinsi Jawa Timur secara otomatis berdasarkan data yang diunggah pengguna. Sistem ini dirancang secara fleksibel untuk menerima berkas bertipe .csv, .xlsx, atau .xls dengan syarat minimal menyertakan 2 variabel numerik dalam satu aspek atau indikator yang sejenis dan tidak digabungkan dengan aspek yang berbeda. Guna menyajikan hasil analisis secara informatif, sistem dilengkapi dengan visualisasi peta interaktif berbasis Kode Wilayah, fitur navigasi pada peta untuk mengakses halaman detail, grafik evaluasi performa model (DBI & *Silhouette Score*), serta *Grouped Bar Chart* (Grafik Batang Kelompok).

Dalam merancang dan mengembangkan sistem informasi ini, penelitian menggunakan metode *Waterfall*. Metode *Waterfall* dipilih karena memiliki alur kerja yang teratur dan sistematis, sehingga setiap tahapan pengembangan sistem mulai dari analisis kebutuhan, perancangan antarmuka dan *backend*, pengkodean,

hingga pengujian dapat diselesaikan secara berurutan dan jelas. Evaluasi performa model pengelompokan data diuji menggunakan metrik *Davies-Bouldin Index* (DBI) dan *Silhouette Score* pada rentang nilai $k = 2$ sampai $k = 5$, sedangkan pengujian fungsionalitas sistem menggunakan metode *Black Box Testing* guna memastikan sistem berjalan sesuai rancangan. Selain itu, digunakan metode *System Usability Scale* (SUS) untuk mengukur tingkat kemudahan penggunaan *website* bagi pengguna. Berikut ini merupakan bagian kerangka berpikir yang menggambarkan alur penelitian dari identifikasi masalah hingga luaran hasil penelitian:



IMPLEMENTASI

Tahap ini mewujudkan rancangan menjadi *website open-data analytics* yang dilengkapi visualisasi peta interaktif menggunakan *react-leaflet* berbasis Kode Wilayah, fitur navigasi klik langsung pada peta untuk mengakses halaman detail wilayah, grafik evaluasi performa model (DBI & *Silhouette Score*), serta *Grouped Bar Chart* untuk visualisasi perbandingan detail indikator tiap daerah terhadap rata-rata *clusternya*.



HASIL PENELITIAN

Sistem berhasil mengelompokkan 38 kabupaten/kota di Jawa Timur secara otomatis berdasarkan data yang diunggah. Jumlah kelompok *cluster* optimal ditentukan secara otomatis pada rentang nilai $k = 2$ sampai $k = 5$ melalui validasi metrik *Davies-Bouldin Index* (DBI) dan *Silhouette Score*, sedangkan seluruh fungsionalitas sistem diuji menggunakan metode *Black Box Testing*. Selain itu, digunakan metode *System Usability Scale* (SUS) untuk mengukur tingkat kemudahan penggunaan *website* bagi pengguna.

Gambar 2.3 Kerangka Berpikir