BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Relevan

Kajian penelitian terdahulu menjadi salah satu acuan penting untuk memperkaya landasan teori sekaligus memberikan gambaran mengenai perkembangan studi yang sudah ada. Bagian ini memuat ringkasan beberapa penelitian sebelumnya yang memiliki keterkaitan dengan topik serta metode yang digunakan, sehingga dapat dijadikan bahan perbandingan dengan penelitian yang sedang dilakukan. Beberapa sumber literatur yang dirangkum pada table berikut:

2.1.1 Matriks Penelitian

Berikut adalah beberapa penelitian terkait yang menjadi acuan dalam penyusunan skripsi ini. Penjabaran ini bertujuan untuk menunjukkan keterkaitan antara studi sebelumnya dengan penelitian yang sedang dilakukan:

Penelitian pertama dilakukan oleh F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, dan M. Jajuli (2022) dengan judul *"Implementasi K-Means Clustering untuk Pengelompokkan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM"*. Studi ini menggunakan data titik kebakaran di wilayah Jawa Barat periode 2019-2021. Dengan menerapkan algoritma *K-Means* dan model CRISP-DM, penelitian ini berhasil mengelompokkan wilayah menjadi tiga *cluster* kerawanan (tinggi, sedang, dan rendah) yang divalidasi menggunakan *Elbow Method* serta *Silhouette Score*. Hasilnya menunjukkan bahwa tingkat urbanisasi dan kepadatan penduduk sangat memengaruhi pola kejadian kebakaran. Keterkaitan penelitian ini dengan penelitian penulis adalah pada penggunaan metode *K-Means*, namun penulis melakukan pengembangan dengan menambahkan variabel data *rescue* (penyelamatan) yang belum dibahas dalam studi tersebut (Dhewayani dkk., 2022).

Penelitian kedua dilakukan oleh Wisnu Priyo Jatmiko, M. Gillang Ramadhani, M. Gilang Romadhon, Gilang Adhmadani, Rahmad Fardian, dan Wawan Joko Pranoto (2024) dengan judul *"Penerapan Metode K-Means Clustering Terhadap Bencana Kebakaran Di Kota Samarinda"*. Studi ini menggunakan data kejadian kebakaran per tahun di Kota Samarinda pada periode 2021-2023 yang bersumber dari Dinas Pemadam Kebakaran Kota Samarinda. Dengan menerapkan algoritma *K-Means Clustering*, penelitian ini berhasil mengelompokkan kecamatan di Samarinda ke dalam beberapa tingkatan risiko, di mana tercatat kejadian terbanyak pada tahun 2021 sebanyak 230 kasus. Hasil penelitian menunjukkan bahwa terdapat beberapa kecamatan yang secara konsisten masuk ke dalam klaster risiko tinggi, sehingga informasi ini direkomendasikan bagi pemerintah daerah dalam mengoptimalkan penempatan pos pemadam dan perencanaan sumber daya. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode *K-Means* untuk penentuan strategi mitigasi, namun penelitian penulis memiliki cakupan data yang lebih luas karena menggabungkan data kebakaran dengan data *rescue* (Jatmiko dkk., 2024).

Penelitian ketiga dilakukan oleh Indah Rosaliyah dan Bani Nurhakim (2023) dengan judul *"Clustering Kejadian Bencana Alam di Jawa Barat Berdasarkan Jenis Bencana Menggunakan K-Means"*. Studi ini menggunakan data berbagai jenis kejadian bencana alam seperti gempa bumi, tanah longsor, dan banjir di wilayah administratif Jawa Barat yang bersumber dari laporan instansi resmi regional. Dengan menerapkan algoritma *K-Means Clustering* melalui proses *Knowledge Discovery in Database* (KDD), penelitian ini menghasilkan enam tingkatan klaster wilayah mulai dari kategori "sangat rendah" hingga "sangat tinggi". Hasil penelitian divalidasi menggunakan *Davies-Bouldin Index* (DBI) dengan nilai 9.20 untuk menilai kualitas pemisahan antar klaster. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode

evaluasi DBI untuk mengukur performa klaster, namun perbedaannya penelitian penulis secara spesifik menggabungkan insiden kebakaran dengan tindakan penyelamatan (*rescue*) dalam satu kesatuan data (Rosaliyah dan Nurhakim, 2023).

Penelitian keempat dilakukan oleh Dede Rohman, Rhima Annisa, Deny Indriyana Efendi, dan Dodi Solahudin (2024) dengan judul *"Clustering Bencana Alam Menggunakan K-Means Pada Wilayah Jawa Barat"*. Studi ini menggunakan data frekuensi kejadian bencana alam seperti gempa bumi, tanah longsor, dan banjir di berbagai kabupaten/kota di Provinsi Jawa Barat. Proses penelitian dilakukan melalui tahapan *data mining* menggunakan bantuan *tool* RapidMiner, di mana penentuan jumlah klaster optimal dilakukan dengan *Elbow Method*. Hasil penelitian membagi wilayah Jawa Barat ke dalam tiga klaster berdasarkan tingkat kejadiannya, yaitu klaster 0 (tinggi), klaster 2 (sedang), dan klaster 1 (rendah). Kualitas pengelompokan diuji menggunakan *Davies-Bouldin Index* (DBI) yang menghasilkan nilai sangat rendah yaitu 0,004, menunjukkan bahwa klaster yang terbentuk sangat kompak. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode *K-Means* dan evaluasi DBI untuk mengukur kepadatan internal klaster, namun perbedaannya penelitian penulis berfokus pada data spesifik kebakaran dan *rescue* di lingkup Kabupaten Bandung (Rohman dkk., 2024).

Penelitian kelima dilakukan oleh Farras Salsabila, Ika Fitrianti, Yuyun Umaidah, dan Nono Heryana (2024) dengan judul *"Penerapan Metode CRISP-DM Untuk Analisa Pendapatan Bersih Bulanan Pekerja Informal di Provinsi Jawa Barat Dengan Algoritma K-Means"*. Studi ini menggunakan data pendapatan bersih bulanan pekerja informal per kabupaten/kota di Jawa Barat dengan menggunakan kerangka metodologi CRISP-DM yang meliputi tahapan *business understanding* hingga

deployment. Dengan menerapkan algoritma *K-Means*, data berhasil dikelompokkan menjadi tiga klaster pendapatan, yaitu rendah, sedang, dan tinggi, guna membantu pemerintah dalam mengarahkan kebijakan ekonomi dan subsidi secara lebih merata. Kualitas pengelompokan dioptimalkan menggunakan *Davies-Bouldin Index* (DBI) dan divalidasi melalui visualisasi geografis. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penerapan kerangka kerja CRISP-DM dan penggunaan evaluasi DBI untuk menentukan jumlah klaster yang optimal, meskipun objek penelitian ini berbeda yakni pada sektor ekonomi pekerja informal (Salsabila dkk., 2024).

Penelitian keenam dilakukan oleh Muhamad Firman Al Halik dan Laila Septiana (2022) dengan judul *"Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Menggunakan Algoritma K-Means Clustering"*. Studi ini menggunakan data kejadian bencana alam di 27 kabupaten/kota di Jawa Barat yang bersumber dari portal Open Data Jawa Barat serta dokumen instansi terkait seperti Dinas Lingkungan Hidup. Dengan menggunakan algoritma *K-Means Clustering*, penelitian ini mengelompokkan wilayah ke dalam tiga klaster kerawanan, yaitu rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa wilayah seperti Bandung dan Garut masuk ke dalam klaster tinggi karena memiliki kombinasi frekuensi bencana yang sering terjadi. Penentuan jumlah klaster optimal dalam studi ini dilakukan dengan *Elbow Method* yang kemudian divalidasi melalui visualisasi spasial. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode *K-Means* untuk memetakan wilayah rawan berdasarkan frekuensi kejadian, namun perbedaannya penelitian penulis secara spesifik mengintegrasikan variabel *rescue* ke dalam analisisnya (Al Halik dan Septiana, 2022).

Penelitian ketujuh dilakukan oleh Arye Fandia Kusuma, Nana Suarna, Irfan Ali, dan Dodi Solihudin (2024) dengan judul *"Pengelompokan Regu Penyelamat Non-Kebakaran di Kabupaten Cirebon dengan K-Means Clustering"*. Studi ini menggunakan data layanan darurat non-kebakaran sebanyak 874 kejadian di Kabupaten Cirebon periode 2023-2024, yang mencakup variabel lokasi, jenis penyelamatan, tingkat keparahan, hingga waktu respons. Dengan menerapkan algoritma *K-Means* melalui pendekatan *Knowledge Discovery in Database* (KDD), penelitian ini berhasil menentukan empat klaster optimal untuk efisiensi distribusi regu penyelamat. Kualitas pengelompokan divalidasi menggunakan *Davies-Bouldin Index* (DBI) yang menghasilkan nilai 0,080. Keterkaitan penelitian ini dengan penelitian penulis sangat erat karena sama-sama membahas aspek penyelamatan (*rescue*) dan penggunaan evaluasi DBI, namun perbedaannya penelitian penulis menggabungkan insiden kebakaran dan penyelamatan dalam satu lingkup analisis di Kabupaten Bandung (Kusuma dkk., 2024).

Penelitian kedelapan dilakukan oleh Muhammad Rezki, Siti Nurdiani, Rizky Ade Safitri, Muhammad Ifan Rifani Ihsan, dan Muhammad Iqbal (2022) dengan judul *"Segmentasi Api dan Asap Pada Kebakaran Dengan Metode K-Means Clustering"*. Berbeda dengan penelitian sebelumnya, studi ini menggunakan data berupa citra (gambar) kebakaran untuk memisahkan area api dan asap dari latar belakang melalui teknik segmentasi *pixel*. Dengan menerapkan algoritma *K-Means Clustering* setelah tahap pra-proses citra, penelitian ini berhasil mengidentifikasi area terdampak kebakaran sebagai input sistem deteksi dini, terutama pada kondisi pencahayaan yang kontras. Evaluasi dilakukan melalui perbandingan visual dan metrik kesamaan segmen untuk memastikan akurasi deteksi. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode *K-Means* sebagai algoritma

utama dalam pengolahan data terkait kebakaran, namun perbedaannya penelitian penulis menggunakan data tabular berupa frekuensi insiden dan layanan *rescue*, bukan berbasis pengolahan citra digital (Rezki dkk., 2022).

Penelitian kesembilan dilakukan oleh Yessy Fitriani, M. Farid Rifai, dan M. Yoga Distra S. (2019) dengan judul *"Penentuan Jumlah Kelas Matakuliah Menggunakan Fuzzy Tsukamoto Dan Metode K-Means Cluster"*. Studi ini menggunakan data akademik mahasiswa dan mata kuliah yang bersumber dari Bagian Akademik STT-PLN untuk mengoptimalkan penentuan kuota kelas. Metodologi yang digunakan adalah penggabungan metode *Fuzzy Tsukamoto* untuk memprediksi jumlah mahasiswa yang mengulang dan algoritma *K-Means Clustering* untuk mengelompokkan mata kuliah berdasarkan jumlah peminat menjadi dua klaster. Hasil penelitian ini berupa sistem aplikasi yang membantu pihak akademik meminimalisir kesalahan dalam pembukaan kuota kelas saat pengisian KRS. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penggunaan metode *K-Means* sebagai alat pendukung keputusan dalam manajemen sumber daya, namun perbedaannya penelitian penulis berfokus pada manajemen sumber daya penyelamatan bencana dan kebakaran di Kabupaten Bandung (Fitriani dkk., 2019).

Penelitian kesepuluh dilakukan oleh Wibowo A., Rohman N., Rusdah Achadi A.H., dan Amri I. (2023) dengan judul *"Clustering Indonesian Provinces by Disaster Intensity using K-Means Algorithm: a Data Mining Approach"*. Studi ini menggunakan data intensitas bencana pada tingkat provinsi di Indonesia dengan menggabungkan variabel kejadian bencana, jumlah populasi, dan luas wilayah yang terdampak. Menggunakan algoritma *K-Means*, penelitian ini berhasil mengelompokkan provinsi-provinsi di Indonesia ke dalam tiga klaster kerawanan: tinggi, sedang, dan rendah. Hasil penelitian menunjukkan

bahwa provinsi pada klaster tinggi memiliki karakteristik frekuensi bencana yang sering serta populasi yang padat, sehingga diperlukan prioritas mitigasi dari pemerintah pusat dan daerah. Keterkaitan penelitian ini dengan penelitian penulis terletak pada penerapan metode *K-Means* untuk penentuan prioritas mitigasi bencana, namun perbedaannya penelitian penulis memiliki skala wilayah yang lebih spesifik pada tingkat kecamatan di Kabupaten Bandung serta penambahan variabel layanan *rescue* (Wibowo dkk., 2023).

Tabel 2.1Matriks Penelitian

No	Penulis	Tahun	Jenis Data	Metode	Keterkaitan
1.	F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, M. Jajuli	2022	Titik kebakaran Jawa Barat	K-Means, CRISP-DM	Persamaan pada metode K-Means, perbedaan pada penambahan variabel data *rescue*.
2.	Wisnu Priyo Jatmiko, M. Gillang Ramadhani, M. Gilang Romadhon, Gilang Adhmadani, Rahmad Fardian, Wawan Joko Pranoto	2024	Kasus kebakaran di Samarinda	K-Means Clustering	Persamaan pada penggunaan K-Means, perbedaan pada cakupan integrasi data *rescue*.
3.	Indah Rosaliyah, Bani Nurhakim	2023	Jenis bencana alam di Jawa Barat	K-Means,KDD	Persamaan pada evaluasi DBI, perbedaan pada fokus

					integrasi data kebakaran dan *rescue*.
4.	Dede Rohman, Rhima Annisa, Deny Indriyana Efendi, Dodi Solahudin	2024	Frekuensi bencana alam Jawa Barat	K-Means, RapidMiner	Persamaan pada evaluasi DBI, perbedaan pada fokus wilayah spesifik Kabupaten Bandung.
5.	Farras Salsabila, Ika Fitrianti, Yuyun Umaidah, Nono Heryana	2024	Pendapatan bersih pekerja informal	K-Means, CRISP-DM	Persamaan pada kerangka CRISP-DM dan DBI, perbedaan pada objek penelitian (ekonomi).
6.	Muhamad Firman Al Halik, Laila Septiana	2022	Kejadian bencana alam di Jawa Barat	K-Means Clustering	Persamaan pada pemetaan wilayah rawan, perbedaan pada penggabungan variabel *rescue*.
7.	Arye Fandia Kusuma, Nana Suarna, Irfan Ali, Dodi Solihudin	2024	Layanan darurat non-kebakaran	K-Means, KDD	Persamaan pada aspek *rescue* dan DBI, perbedaan pada penggabungan data insiden kebakaran.

8.	Muhammad Rezki, Siti Nurdiani, Rizky Ade Safitri, Muhammad Ifan Rifani Ihsan, Muhammad Iqbal	2022	Citra (gambar) api dan asap kebakaran	K-Means Clustering	Persamaan pada algoritma K-Means, perbedaan pada jenis data (tabular vs citra digital).
9.	Yessy Fitriani, M. Farid Rifai, M. Yoga Distra S.	2019	Data akademik dan mata kuliah	K-Means, Fuzzy Tsukamoto	Persamaan pada penggunaan K-Means, perbedaan pada bidang manajemen bencana.
10.	Wibowo A., Rohman N., Rusdah Achadi A.H., Amri I.	2023	Intensitas bencana provinsi di Indonesia	K-Means Algorithm	Persamaan pada penentuan prioritas mitigasi, perbedaan pada skala wilayah dan variabel *rescue*.

Pada penelitian ini bertujuan untuk meningkatkan ketepatan dalam pengelompokan data kebakaran dan keadaan darurat dengan memanfaatkan pendekatan CRISP-DM yang dipadukan dengan algoritma clustering. Evaluasi hasil model dilakukan menggunakan indikator Davies-Bouldin Index (DBI), Silhouette Score, dan Elbow Method yang berfungsi untuk menilai kualitas klaster yang terbentuk. Adapun berikut disajikan perbandingan antara penelitian ini dengan studi-studi sebelumnya:

a. Metode Evaluasi yang Digunakan

   Pada penelitian sebelumnya, umumnya evaluasi hasil clustering masih terbatas pada satu atau dua metrik sederhana. Berbeda dengan penelitian ini yang

mengombinasikan Davies-Bouldin Index, Silhouette Score, serta Elbow Method untuk menilai kualitas klaster sekaligus menentukan jumlah klaster yang paling optimal.

b. Dataset yang Digunakan

Studi-studi sebelumnya biasanya menggunakan data bencana secara umum, misalnya banjir, tanah longsor, atau kebakaran hutan. Sedangkan penelitian ini secara khusus menggunakan data historis kejadian kebakaran dan keadaan darurat (rescue) di Kabupaten Bandung tahun 2022-2024, sehingga lebih fokus pada konteks pelayanan pemadam kebakaran.

c. Gap Penelitian

Sebagian besar penelitian terdahulu masih menekankan pada analisis deskriptif atau pengelompokan bencana secara umum, tanpa mengarah langsung pada kebutuhan operasional pemadam kebakaran. Penelitian ini hadir untuk mengisi kekosongan tersebut dengan menerapkan pendekatan CRISP-DM dan metode K-Means Clustering secara langsung pada kasus kebakaran dan keadaan darurat, sehingga hasilnya dapat dimanfaatkan untuk strategi mitigasi dan perencanaan sumber daya di lapangan landasan teori.

2.2 Landsasan Teori

Landasan teori berfungsi sebagai dasar konseptual dalam penelitian, yang memberikan gambaran mengenai konsep, metode, serta pendekatan yang digunakan. Pada penelitian mengenai penerapan metode *K-Means* untuk *clustering* daerah rawan kejadian kebakaran dan keadaan darurat di Kabupaten Bandung, diperlukan pemahaman mengenai konsep dasar *clustering*, kerangka kerja penelitian *data mining*, serta metrik evaluasi yang digunakan untuk menilai hasil pengelompokan.

2.2.1 K-Means Clustering

Clustering merupakan salah satu teknik dalam *data mining* yang berfungsi untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Data yang berada dalam satu klaster memiliki tingkat kesamaan yang

tinggi, sedangkan antar klaster memiliki perbedaan yang jelas. Salah satu algoritma yang paling populer adalah *K-Means* karena kesederhanaannya, efisiensi, serta kemampuannya dalam menangani data berukuran besar (Wulandari, Ansori, & Khadijah, 2022).

Dalam proses penelitian berbasis *data mining*, *CRISP-DM (Cross-Industry Standard Process for Data Mining)* sering digunakan sebagai kerangka kerja standar. CRISP-DM terdiri dari enam tahapan utama, yaitu *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment*. Tahapan ini membantu peneliti agar penelitian berjalan lebih sistematis mulai dari memahami masalah hingga menghasilkan rekomendasi yang relevan dengan kebutuhan. (Wulandari, Ansori, & Khadijah, 2022).

Evaluasi dalam *clustering* menjadi aspek penting untuk menilai kualitas klaster yang dihasilkan. Beberapa metrik evaluasi yang umum digunakan antara lain *Silhouette Score, Davies-Bouldin Index (DBI), dan Elbow Method*. *Silhouette Score* mengukur seberapa baik objek berada pada klaster yang sesuai dibandingkan dengan klaster lain, DBI digunakan untuk menilai pemisahan antar klaster, sedangkan *Elbow Method* membantu menentukan jumlah klaster optimal berdasarkan perubahan nilai *within-cluster sum of squares* (Syahkur, Hartama, & Solikhun, 2024).

Hasil penelitian sebelumnya menunjukkan bahwa kombinasi metrik evaluasi tersebut dapat memberikan gambaran lebih akurat dalam menentukan jumlah dan kualitas klaster. Misalnya, pada penelitian pengelompokan daerah rawan bencana, penggunaan *Elbow*, DBI, dan *Silhouette* terbukti mampu menghasilkan klaster yang lebih signifikan dan bermakna, dengan nilai DBI rendah dan *Silhouette* mendekati 1 sebagai indikator hasil clustering yang optimal (Hartama, Wanayumini, & Damanik, 2024).

Dengan demikian, penggunaan algoritma *K-Means* yang dikombinasikan dengan kerangka kerja CRISP-DM serta evaluasi hasil melalui DBI, *Silhouette Score*, dan *Elbow Method* dianggap relevan untuk penelitian pengelompokan data kebakaran dan keadaan darurat. Hal ini sejalan dengan tujuan penelitian untuk memberikan hasil analisis yang

tidak hanya terstruktur tetapi juga memiliki validitas kuat dalam mendukung pengambilan keputusan di bidang penanggulangan bencana dan darurat.

2.2.2 Machine Learning

Machine Learning adalah studi berkelanjutan tentang konsep pengenalan pola dan pembelajaran komputasi dalam kecerdasan buatan yang menggunakan algoritma pembelajaran seperti diawasi dan tidak diawasi untuk memprediksi dan mendukung pengambilan keputusan otomatis berdasarkan sekumpulan data" (Wardhana, Wang & Sibuea, 2023).

Machine Learning tidak hanya diaplikasikan dalam bidang prediksi, tetapi juga banyak digunakan dalam teknik pengelompokan data atau *clustering*. Seperti yang dijelaskan dalam penelitian Bhustomy, Fergie, Cornelius, Jonathan, Reynaldi (2025)*"Algoritma K-Means merupakan salah satu metode clustering yang cukup populer karena kemampuannya dalam mengelompokkan data ke dalam cluster-cluster dengan tingkat kemiripan yang tinggi antar anggota dalam cluster yang sama, dan perbedaan yang signifikan antar cluster yang berbeda."* Penerapan K-Means dalam penelitian ini berhasil mengelompokkan data musik berdasarkan fitur audio seperti tempo, ritme, dan pitch, sehingga menghasilkan tiga kelompok utama yang menunjukkan perbedaan karakteristik genre musik. Hal ini menegaskan bahwa pendekatan clustering dalam machine learning dapat diterapkan secara luas, termasuk untuk analisis pola data yang kompleks dan beragam. Relevansi metode ini terhadap penelitian kebakaran dan rescue terletak pada prinsip pengelompokan pola yang sama, di mana *K-Means* dapat digunakan untuk mengelompokkan wilayah berdasarkan frekuensi kejadian darurat, tingkat risiko, atau faktor lingkungan. Dengan demikian, hasil clustering dapat membantu instansi pemadam kebakaran dalam memahami pola kejadian dan merancang strategi penanggulangan yang lebih efektif.

Machine learning, khususnya pada *teknik clustering* dengan metode *K-Means*, sering digunakan dalam berbagai aplikasi lokal untuk mengelompokkan data berdasarkan karakteristik tertentu. Misalnya, dalam penelitian *Penerapan Data Mining pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu*

Kakikaki) dituliskan bahwa “K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok” (Wahyu Wijaya Kristianto & Christ Rudianto, 2022). Hasil penelitian menunjukkan bahwa data transaksi penjualan dapat dibagi ke dalam cluster “Laku”, “Cukup”, dan “Kurang”, yang membantu toko dalam menentukan produk mana yang paling diminati pasar pada periode tertentu.

2.2.3 Deep Learning

Deep Learning adalah bagian dari machine learning yang berfokus pada penggunaan jaringan saraf tiruan dengan banyak *hidden layer* untuk menangkap pola kompleks dari data. Contohnya, penelitian “Pemodelan Wilayah Titik Api Kebakaran Hutan Menggunakan Deep Learning” oleh Saruni Dwiasnati dkk. menggunakan Convolutional Neural Network (CNN) untuk klasifikasi citra wilayah titik api. Penelitian ini melibatkan feedforward dan backpropagation, dengan hasil akurasi sekitar 89%, menunjukkan Deep Learning efektif dalam mendeteksi dan memetakan wilayah rawan kebakaran hutan di pulau-pulau seperti Sumba dan Timor di Nusa Tenggara Timur. CNN berhasil mempelajari fitur gambar seperti intensitas cahaya dan tekstur yang membedakan area terbakar dari area tidak terbakar. (Saruni Dwiasnati, Yudo Devianto, Sutan Mohammad Arif, Reza Avrizal, 2024)

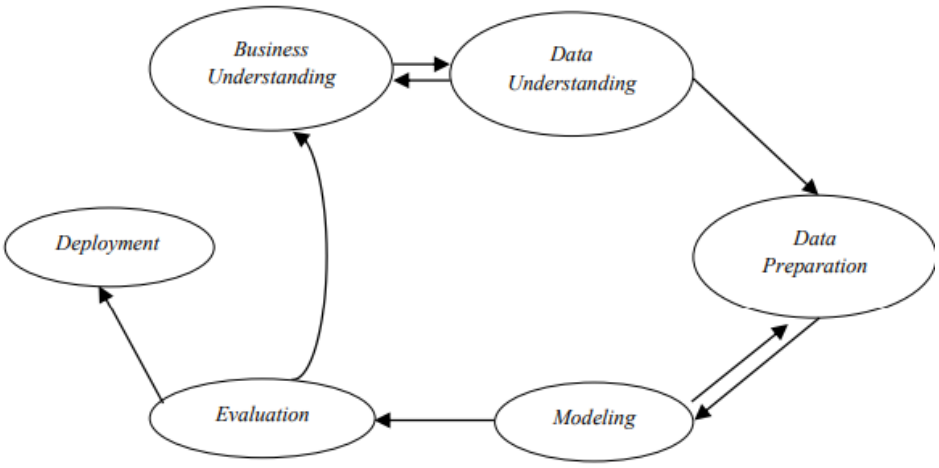
Deep Learning merupakan salah satu cabang dari Machine Learning yang memanfaatkan jaringan saraf tiruan berlapis (artificial neural networks) untuk memproses data dalam jumlah besar dan kompleks. Teknologi ini terbukti efektif digunakan dalam bidang deteksi citra, terutama yang berkaitan dengan identifikasi kebakaran. (Saputra dan Adhinata, 2023) menjelaskan bahwa penggunaan arsitektur DenseNet201 dengan metode transfer learning mampu mencapai akurasi hingga 99% dalam mengklasifikasikan citra kebakaran hutan dan lahan. Hal ini menunjukkan bahwa Deep Learning memiliki potensi yang sangat besar untuk diaplikasikan dalam sistem deteksi otomatis, khususnya dalam konteks keadaan darurat seperti kebakaran yang membutuhkan respon cepat dan akurat.

2.2.4 Cross-Industry Standard Process for Data Mining (CRISP-DM)

Metodologi CRISP-DM merupakan salah satu kerangka kerja yang paling banyak digunakan dalam penelitian data mining. Seperti dijelaskan oleh (Hasanah, Soim, dan Handayani.,2022). *"Model CRISP-DM memiliki enam tahapan utama, yaitu business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Keenam tahapan tersebut memberikan panduan sistematis bagi peneliti untuk memahami permasalahan, mengolah data, membangun model, mengevaluasi hasil, hingga mengimplementasikan solusi."* Kutipan ini menegaskan bahwa CRISP-DM tidak hanya berfungsi sebagai metodologi, tetapi juga sebagai standar yang membantu konsistensi penelitian data mining, termasuk dalam bidang prediksi kebakaran maupun keadaan darurat.

Implementasi CRISP-DM Model Menggunakan Metode Decision Tree menegaskan bahwa CRISP-DM banyak digunakan dalam penelitian berbasis data karena fleksibilitas dan kemampuan adaptasinya terhadap berbagai jenis proyek. Penelitian tersebut menyatakan bahwa CRISP-DM digunakan sebagai metodologi data mining yang terdiri dari tahap business understanding, data understanding, data preparation, modeling, evaluation, dan deployment untuk memecahkan masalah klasifikasi curah hujan dengan algoritma.

Penelitian *CRISP-DM Method On Indonesian Micro Industries (UMKM) Using K-Means Clustering Algorithm* menggunakan data UMKM dari tahun 2014-2018 di berbagai sektor industri (seperti industri logam, kerajinan, makanan & minuman), dengan kerangka kerja CRISP-DM untuk memandu seluruh proses analisis (dari business understanding sampai deployment). Hasilnya menunjukkan bahwa pada $k = 5$ cluster, nilai Davies-Bouldin Index (DBI) untuk tiap cluster berkisar di angka 0,259 sampai 0,369, yang dikenal sebagai DBI rendah, sehingga cluster-cluster yang terbentuk cukup baik. Hal ini membuktikan bahwa penggunaan CRISP-DM + K-Means dalam penelitian-data UMKM mampu memberikan pengelompokan wilayah industri yang representatif serta dapat dimanfaatkan oleh pembuat kebijakan untuk menentukan prioritas pengembangan industri (Ansori & Wulandari, 2022).



Gambar 2.1 METODE CRISP-DM (Dhewayani et al.,2022)

Model *Cross Industry Standard Process for Data Mining (CRISP-DM)* terdiri dari enam tahapan utama, yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation,* dan *Deployment.* Keenam tahapan tersebut saling berhubungan dan membentuk siklus proses analisis data yang sistematis seperti ditunjukkan pada Gambar 2.1.

2.2.4.1 Business Understanding

Tahap *Business Understanding* merupakan langkah awal dalam metode CRISP-DM yang berfokus pada pemahaman tujuan bisnis atau kebutuhan organisasi yang menjadi dasar dilakukannya *data mining*. Menurut Dhewayani et al., (2022), pada tahap ini dilakukan identifikasi terhadap masalah utama, penetapan tujuan analisis, serta penentuan arah dan ruang lingkup penelitian. Tahap ini penting untuk memastikan bahwa hasil analisis data nantinya dapat menjawab kebutuhan yang ada dan memberikan nilai manfaat yang konkret.

2.2.4.2 Data Understanding

Tahap *Data Understanding* bertujuan untuk memperoleh pemahaman awal mengenai data yang akan digunakan. Proses ini mencakup pengumpulan data awal, evaluasi kualitas data, dan analisis karakteristik data untuk menentukan apakah data tersebut layak digunakan dalam proses analisis selanjutnya. Dhewayani et al., (2022) menyebutkan bahwa kegiatan utama pada tahap ini meliputi pemeriksaan struktur data, tipe atribut, nilai yang hilang (*missing values*), serta pencarian pola awal yang mungkin muncul dari data.

2.2.4.3 Data Preparation

Tahap *Data Preparation* dilakukan untuk mempersiapkan data agar dapat digunakan dalam proses pemodelan. Menurut Pyle (1999), tahap ini terdiri atas beberapa kegiatan penting seperti *data selection*, *data cleaning*, *data transformation*, dan *data reduction*. Dhewayani et al., (2022) juga menegaskan bahwa tahap ini bertujuan untuk memastikan data sudah dalam format yang tepat dan bebas dari kesalahan, duplikasi, atau nilai ekstrem yang dapat memengaruhi hasil analisis. Tahap ini sangat krusial karena kualitas data berpengaruh langsung terhadap akurasi hasil model yang dihasilkan.

2.2.4.4 Modeling

Tahap *Modeling* merupakan proses penerapan metode *data mining* yang sesuai dengan tujuan analisis. Pada tahap ini, dilakukan pemilihan algoritma, penentuan parameter model, serta proses pengujian terhadap data yang telah disiapkan sebelumnya. Menurut Dhewayani et al., (2022), tahap *Modeling* menghasilkan pola atau struktur data yang dapat digunakan untuk menemukan hubungan antaratribut. Salah satu metode yang umum digunakan pada tahap ini adalah *K-Means Clustering*, yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik.

2.2.4.5 Evaluation

Tahap *Evaluation* bertujuan untuk menilai sejauh mana hasil pemodelan sudah memenuhi tujuan analisis yang telah ditetapkan sebelumnya. Menurut Larose (2015), tahap ini dilakukan dengan meninjau kembali hasil model serta mengevaluasi tingkat

akurasi dan relevansi dari hasil pengelompokan data. Dalam konteks umum *clustering*, metode evaluasi yang sering digunakan adalah *Silhouette Coefficient* dan *Davies-Bouldin Index (DBI)* untuk mengukur kualitas pembentukan cluster (Dhewayani et al., 2022).

2.2.4.6 Deployment

Tahap *Deployment* merupakan tahap terakhir dalam metode CRISP-DM, di mana hasil analisis diimplementasikan dan digunakan untuk pengambilan keputusan. Menurut Wirth dan Hipp (2000), hasil dari tahap ini dapat berupa laporan analisis, visualisasi data, atau sistem pendukung keputusan berbasis hasil *data mining*. Dhewayani et al., (2022) menjelaskan bahwa tahap Deployment menjadi penting agar hasil penelitian tidak berhenti pada proses analisis, tetapi dapat memberikan manfaat nyata bagi pihak yang membutuhkan informasi tersebut.

2.2.5 Elbow Method

Metode *Elbow* digunakan untuk menentukan jumlah *cluster* optimal dalam algoritma *K-Means*. Menurut Dhewayani et al. (2022), metode ini bekerja dengan menghitung nilai *Within-Cluster Sum of Squares (WCSS)* pada berbagai nilai *k* dan memvisualisasikannya dalam bentuk grafik. Titik siku (*elbow point*) pada grafik menunjukkan jumlah *cluster* terbaik, karena setelah titik tersebut, penurunan nilai WCSS menjadi tidak signifikan.

Rumus perhitungan *Within-Cluster Sum of Squares (WCSS)* digunakan pada metode *Elbow* untuk menentukan jumlah *cluster* optimal seperti ditunjukkan pada Persamaan (2.1).

$$WCSS = \sum_{i=1}^{k} \sum_{x_j \in C_i} (x_j - \mu_i)^2$$

(Persamaan 2.1)

Keterangan:

k = jumlah *cluster*

Ci = jumlah *Cluster*

xj =*data ke-j dalam cluster*

μ_i= pusat (*centroid*) dari *cluster* ke-i

Semakin kecil nilai WCSS, maka semakin baik hasil pengelompokan data, karena menunjukkan bahwa data dalam satu *cluster* memiliki tingkat kemiripan yang tinggi (Dhewayani et al., 2022).

2.2.6 Silhouette Coefficient

Silhouette Coefficient merupakan metode evaluasi yang digunakan untuk mengukur kualitas pengelompokan data hasil *clustering*. Nilai *Silhouette* menunjukkan seberapa baik setiap objek data berada di dalam *cluster*-nya sendiri dibandingkan dengan *cluster* lain. Menurut Rousseeuw (1987), nilai *Silhouette* berkisar antara -1 hingga 1, dengan ketentuan:

- Nilai mendekati **1** → data sudah dikelompokkan dengan sangat baik.
- Nilai **0** → data berada di batas antara dua *cluster*.
- Nilai mendekati **-1** → data salah kelompok.

Perhitungan *Silhouette Coefficient* ditunjukkan pada Persamaan (2.2).

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

(Persamaan 2.2)

Keterangan:

$a(i)$ = rata-rata jarak antara data $i$dan seluruh data dalam *cluster*-nya sendiri.
$b(i)$ = rata-rata jarak antara data $i$dan seluruh data dalam *cluster* terdekat.

Nilai *Silhouette Coefficient* yang tinggi menunjukkan bahwa objek berada dalam *cluster* yang sesuai. Sebaliknya, nilai negatif menandakan bahwa objek cenderung lebih dekat dengan *cluster* lain dibandingkan dengan *cluster*-nya sendiri (Dhewayani et al., 2022).

2.2.7 Davies-Bouldin Index (DBI)

Selain menggunakan metode *Elbow* dan *Silhouette Coefficient*, evaluasi hasil *clustering* juga dapat dilakukan dengan metode *Davies–Bouldin Index (DBI)*. Metode ini digunakan untuk mengukur tingkat kemiripan antar *cluster* dengan

mempertimbangkan jarak antar pusat *cluster* (*centroid*) dan sebaran data dalam setiap *cluster*. DBI memberikan gambaran seberapa terpisah antar *cluster* dan seberapa padat data di dalam satu *cluster*.

Menurut Tarigan (2023) dalam penelitiannya *Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification*, semakin kecil nilai DBI yang dihasilkan, maka semakin baik hasil pengelompokan data. Nilai DBI yang rendah menandakan bahwa *cluster* memiliki jarak yang jauh antar satu sama lain dan tingkat kepadatan internal yang tinggi. Persamaan perhitungan *Davies–Bouldin Index* ditunjukkan pada Persamaan (2.3).

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$$

(Persamaan 2.3)

Keterangan:

k = jumlah *cluster*

σ_i = rata-rata jarak antara setiap titik dalam *cluster* ke-i terhadap pusatnya (*centroid*).

$d(c_i, c_j)$ = jarak antara *centroid* dari *cluster* ke-i dan *cluster* ke-j

Nilai *Davies–Bouldin Index* yang rendah menunjukkan bahwa hasil *clustering* memiliki pemisahan yang baik antar *cluster* serta tingkat homogenitas yang tinggi di dalam *cluster* (Tarigan, 2023).

2.2.8 Python

Bahasa pemrograman Python merupakan salah satu bahasa yang paling banyak digunakan dalam bidang *data science*, *machine learning*, dan analisis data karena sintaksnya sederhana, fleksibel, serta memiliki dukungan pustaka (*library*) yang sangat luas. Python memiliki kemampuan untuk melakukan pengolahan data, visualisasi, hingga penerapan berbagai algoritma seperti *clustering* dan *classification* secara efisien (Pebralia, 2022). Penggunaan Python juga dinilai efektif dalam lingkungan penelitian yang membutuhkan proses analisis data cepat dan hasil yang mudah divisualisasikan.

Menurut Nur & Cahyani (2024), Python menjadi salah satu bahasa utama dalam pengembangan sistem berbasis data karena memiliki berbagai library populer seperti *pandas*, *numpy*, *matplotlib*, dan *scikit-learn* yang memudahkan peneliti dalam proses pra-pemrosesan data (*data preprocessing*) dan pemodelan algoritma. Dalam penelitian mereka yang berfokus pada pembelajaran *data science* berbasis *Jupyter Notebook*, Python terbukti mampu meningkatkan efisiensi dan kejelasan hasil analisis data

Dalam penelitian ini, Python dipilih karena kemampuannya yang komprehensif dalam menangani seluruh tahapan analisis data, mulai dari pengumpulan dan pembersihan data, pemodelan algoritma *K-Means Clustering*, hingga visualisasi hasil pengelompokan wilayah rawan kejadian darurat. Selain itu, Python memungkinkan integrasi dengan *Jupyter Notebook*, sehingga seluruh proses analisis dapat dilakukan secara interaktif dan terdokumentasi dengan baik. Dengan dukungan pustaka seperti *pandas* untuk pengelolaan data, *matplotlib* untuk visualisasi, dan *scikit-learn* untuk penerapan algoritma *machine learning*, Python menjadi pilihan yang tepat untuk menghasilkan proses analisis yang efisien, transparan, dan mudah direplikasi oleh peneliti lain.

2.2.9 Jupyter Notebook

Jupyter Notebook merupakan platform yang memadukan kemampuan pemrograman dengan dokumentasi interaktif, sehingga sangat membantu peneliti dalam melakukan analisis data sekaligus mencatat proses penelitian secara sistematis. Dalam penggunaannya, Jupyter Notebook memungkinkan peneliti menulis kode Python, menampilkan hasil analisis, membuat visualisasi data, dan menyisipkan penjelasan atau catatan penting secara langsung dalam satu dokumen yang sama. Hal ini membuat proses penelitian menjadi lebih transparan dan mudah diikuti, baik oleh peneliti itu sendiri maupun oleh pihak lain yang ingin memahami metode dan hasil yang diperoleh.

Selain itu, Jupyter Notebook memberikan fleksibilitas tinggi karena memungkinkan eksperimen, percobaan, dan dokumentasi dilakukan secara bersamaan. Peneliti tidak hanya fokus pada eksekusi kode, tetapi juga dapat menilai hasil analisis secara real-time, mengoreksi kesalahan, dan menyajikan temuan dengan cara yang lebih komunikatif. Kemampuan ini membuat Jupyter Notebook sangat berguna dalam berbagai

bidang penelitian, terutama yang membutuhkan analisis data yang kompleks, pemodelan, serta visualisasi yang jelas. Dengan demikian, Jupyter Notebook bukan sekadar alat untuk menulis dan menjalankan kode, tetapi juga merupakan media yang efektif untuk mendukung proses penelitian ilmiah secara menyeluruh (Asyrofi & Asyrofi, 2023).

2.2.10 Visual Studio Code

Visual Studio Code merupakan editor kode buatan Microsoft yang dapat diunduh secara gratis dan digunakan di berbagai sistem operasi, seperti Windows, Mac OS, hingga Linux. Perangkat lunak ini berfungsi untuk menyusun sintaks aplikasi serta mendukung modifikasi kode sumber dalam beragam bahasa pemrograman. Sebagai editor teks yang populer, Visual Studio Code menawarkan ekosistem dan ekstensi yang sangat luas, sehingga sangat kompatibel dengan lingkungan pemrograman seperti Python dan JavaScript (Nur dan Cahyani, 2024).

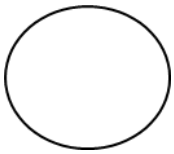
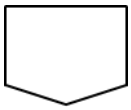
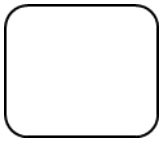
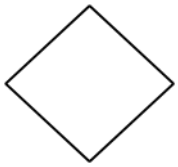
2.2.11 Streamlit

Streamlit adalah sebuah kerangka kerja (*framework*) berbasis Python yang bersifat *open source* dan dirancang untuk memudahkan pengembang dalam membangun aplikasi web yang interaktif, khususnya untuk kebutuhan visualisasi data dan *machine learning*. Keunggulan Streamlit terletak pada kemudahannya yang memungkinkan pembuatan antarmuka web yang responsif tanpa harus menulis kode HTML, CSS, atau JavaScript secara manual (Muhammad Haris M.K.A, 2025).

2.2.12 Flowchart

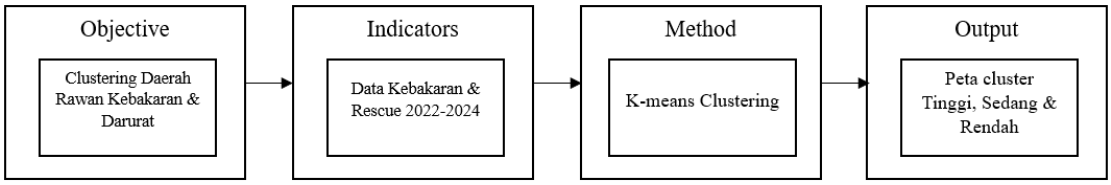
Flowchart merupakan simbol-simbol tertentu untuk menggambarkan urutan proses secara mendetail serta hubungan antarproses dalam suatu sistem. Flowchart berfungsi untuk mempermudah analisis dan perancangan sistem karena memungkinkan perancang melihat keseluruhan alur kerja sebelum implementasi. Dengan demikian, flowchart menjadi alat penting dalam pemrograman dan pengelolaan proses di bidang teknik informatika, membantu memastikan setiap langkah prosedural dapat dipahami dengan jelas dan terstruktur (Everaldo et al., 2023).

Tabel 2.2 Tabel Flowchart

No	Simbol	Nama	Keterangan
1.		Alur Proses	Simbol untuk mempresentasikan urutan atau tahapan proses dari suatu objek.
2.		*On-Page Reference*	Simbol untuk menggambarkan keterkaitan antar objek dalam satu halaman yang sama.
3.		*Off-Page Reference*	Simbol untuk menunjukan hubungan antar objek yang berada di halaman berbeda.
4.		*Terminator*	Simbol yang menandakan titik awal maupun akhir dari sebuah program
5.		*Process*	Simbol yang menjelaskan proses yang di jalankan di dalam komputer.
6.		*Decision*	Simbol yang memberikan dua alternatif jawaban, yaitu YA atau TIDAK, sesuai dengan kondisi tertentu.
7.		*Input/Output*	Simbol yang menggambarkan

			aktivitas input maupun output data.
8.		Manual *Operation*	Simbol yang menunjukan adanya peoses yang di lakukan secara manual, bukan oleh computer.
9.		Document	Simbol menandakan aktivitas pencetakan dokumen.
10.		Predefine Proses	Simbol untuk merepresentasikan pemanggilan bagian program lain (sub-program) atau prosedur tertentu.
11.		*Display*	Simbol menggambarkan keluaran berupa tampilan pada layar monitor.

2.3 Kerangka Pemikiran



Gambar 2.2 Kerangka Pemikiran

Gambar di atas menampilkan tahapan penelitian pada studi ini. Penjelasan untuk tiap tahap disajikan di bawah ini:

1. Objective

   Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan mengelompokkan wilayah rawan kebakaran serta penyelamatan di Kabupaten Bandung berdasarkan data kejadian darurat tahun 2022-2024. Hasil pengelompokan ini diharapkan dapat membantu dinas terkait dalam menentukan prioritas penanganan, penyebaran armada, serta strategi mitigasi yang lebih efektif sesuai dengan tingkat risiko masing-masing wilayah.

2. Indicators

   Data yang digunakan dalam penelitian ini merupakan data historis kejadian kebakaran dan penyelamatan (rescue) di Kabupaten Bandung selama tahun 2022-2024. Dataset ini mencakup variabel utama seperti frekuensi kejadian, response time, dan total kerugian. Data tersebut digunakan untuk mengidentifikasi pola kejadian darurat yang kemudian dijadikan dasar dalam proses pengelompokan wilayah rawan.

3. Method

   Metode yang digunakan adalah algoritma *K-Means Clustering* dengan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Tahapan ini meliputi *business understanding, data understanding, data preparation, modeling, evaluation,* dan *deployment*. Selain itu, dilakukan pengujian untuk menentukan jumlah klaster optimal dan validasi model guna memastikan stabilitas hasil pengelompokan.

4. Output

   Hasil akhir dari penelitian ini adalah sebuah sistem dashboard yang menampilkan peta sebaran dan pengelompokan wilayah berdasarkan tingkat kerawanan (Cluster Tinggi, Sedang, dan Rendah). Hasil ini dievaluasi menggunakan metrik seperti *Silhouette Score* dan *Davies-Bouldin Index* (DBI) untuk memastikan bahwa model *clustering* yang dihasilkan memiliki tingkat keakuratan yang baik dalam menggambarkan kondisi aktual di lapangan.