

BAB I

PENDAHULUAN

1.1 Latar Belakang

UMKM adalah salah satu pilar utama dalam perekonomian Indonesia. Data Kementerian Koperasi dan UKM tahun 2023 mencatat ada lebih dari 65 juta unit UMKM yang menyumbang sekitar 61% PDB nasional dan menyerap sekitar 97% tenaga kerja (Yoni Efilia et al., 2024). Peran ini menjadikan UMKM sebagai tulang punggung ekonomi yang terbukti mampu bertahan bahkan saat pandemi Covid-19 melanda. Namun, di balik kontribusi besar tersebut, UMKM menghadapi tantangan serius. Keterbatasan modal, akses pasar yang terbatas, serta data usaha yang beragam membuat kebijakan pengembangan seringkali meleset dari sasaran (Harahap et al., 2025). Kabupaten Sidoarjo, sebagai salah satu pusat UMKM di Jawa Timur, dikenal dengan keragaman sektor usaha mulai dari kuliner, kerajinan, hingga industri olahan ikan bandeng asap. Pemerintah daerah bahkan menempatkan Sidoarjo sebagai Kota UMKM Indonesia karena jumlah dan variasi UMKM yang tinggi (Karinayah, 2018). ini menegaskan pentingnya pemetaan UMKM yang akurat dan representatif, agar kebijakan pembinaan dan pengembangan usaha dapat lebih efektif serta sesuai dengan karakteristik local (Rahmadani Dyah Rannu & Subroto Tjipto Waspodo, 2022).

Meskipun UMKM memiliki kontribusi besar terhadap perekonomian, pengelolaan dan pemetaan UMKM masih menghadapi berbagai kendala. Salah satu masalah utama adalah data UMKM yang beragam dan berbeda dalam hal modal, omzet, jumlah tenaga kerja, dan sektor usaha sehingga sulit untuk dikelompokkan secara konsisten. Pemetaan UMKM yang selama ini dilakukan pemerintah daerah umumnya masih bersifat administratif atau manual, sehingga belum mampu memberikan gambaran yang akurat mengenai karakteristik usaha. Akibatnya, kebijakan pembinaan dan pengembangan seringkali tidak tepat sasaran karena tidak didukung oleh analisis berbasis data yang kuat (Yoni Efilia et al., 2024)(Juliansyah et al., 2024).

Permasalahan utama penelitian ini terletak pada ketidak konsistenan hasil klusterisasi UMKM di Kabupaten Sidoarjo ketika proses pengelompokan dilakukan

berulang menggunakan metode *Clustering* tunggal. Metode seperti K-means dan K-medoids memang banyak digunakan karena sederhana dan efisien, namun hasilnya sangat bergantung pada inisialisasi awal dan sensitif terhadap outlier. Akibatnya, pemetaan UMKM seringkali tidak konsisten, terutama ketika data diperbarui atau parameter berubah. Kondisi ini membuat hasil analisis sulit dijadikan dasar kebijakan jangka panjang, karena peta UMKM yang dihasilkan bisa berbeda pada setiap percobaan (Catur Pamungkas et al., 2025)(wahid et al., 2024). Pada wilayah penelitian, yaitu UMKM di Sidoarjo, kelemahan metode tunggal semakin terasa karena jumlah usaha yang terus berkembang pesat. Pemetaan berbasis K-means atau *agglomerative Clustering* belum mampu menangkap kompleksitas data yang heterogen, sehingga kelompok yang terbentuk kurang representatif. Hal ini menimbulkan risiko kebijakan pembinaan yang tidak tepat sasaran, misalnya bantuan modal atau pelatihan yang tidak sesuai dengan kebutuhan kelompok usaha (Agustin et al., 2024)(Mustaniroh et al., 2020).

Berbagai penelitian sebelumnya telah mencoba menggunakan metode *Clustering* untuk pemetaan UMKM. Metode K-means banyak dipakai karena sederhana dan efisien dalam mengelompokkan data berdasarkan atribut seperti omzet, modal, dan tenaga kerja. Selain itu, *Agglomerative Clustering* juga digunakan untuk mengidentifikasi pola usaha di tingkat regional, sedangkan Fuzzy C-means dan Self-Organizing Map (SOM) dipakai untuk menangkap karakteristik data yang lebih kompleks. Upaya ini menunjukkan bahwa analisis berbasis *Clustering* dapat membantu memahami keragaman UMKM dan mendukung kebijakan pembinaan (Catur Pamungkas et al., 2025)(wahid et al., 2024). Namun, penerapan metode tersebut masih terbatas pada penggunaan algoritma tunggal. Misalnya, penelitian tentang pengelompokan UMKM berbasis K-means hanya menghasilkan cluster yang sensitif terhadap inisialisasi awal, sementara penelitian dengan SOM dan Fuzzy C-means menunjukkan hasil yang berbeda ketika parameter diubah. Hal ini menegaskan bahwa meskipun metode *Clustering* sudah digunakan dalam pemetaan UMKM, hasilnya belum cukup stabil dan konsisten untuk dijadikan dasar kebijakan jangka Panjang (Agustin et al., 2024)(Mustaniroh et al., 2020).

Untuk mengatasi kelemahan metode *Clustering* tunggal, penelitian ini menggunakan pendekatan ensemble *Clustering* berbasis weighted voting. Ensemble *Clustering* merupakan metode yang menggabungkan hasil dari beberapa algoritma untuk menghasilkan kelompok yang lebih stabil dan representative (Banerjee et al., 2021). Dalam penelitian ini dipilih kombinasi K-means, K-medoids dan *Agglomerative Hierarchical Clustering* (AHC) di mana hasil cluster dari masing-masing algoritma digabungkan melalui mekanisme weighted voting. Pendekatan ini terbukti mampu mengurangi sensitivitas terhadap inisialisasi dan outlier, sekaligus meningkatkan konsistensi hasil pemetaan UMKM (Husin & Ku-Mahamud, 2018). Selain itu, tren penelitian terkini menekankan pentingnya kombinasi algoritma dalam analisis data kompleks. Hal ini menunjukkan bahwa ensemble *Clustering* semakin relevan untuk diterapkan pada data UMKM yang heterogeny (Fu et al., 2022).

Dengan penerapan ensemble *Clustering* berbasis *weighted voting*, penelitian ini diharapkan mampu menghasilkan pemetaan UMKM yang lebih stabil, konsisten, dan representatif dibandingkan metode *Clustering* tunggal. Hasil pemetaan yang lebih baik untuk membantu pemerintah daerah maupun pemangku kebijakan dalam menentukan strategi pembinaan dan pengembangan UMKM secara tepat sasaran (Alqurashi & Wang, 2019). Selain itu, hasil penelitian ini diharapkan dapat menjadi referensi akademis bagi pengembangan metode *Clustering* di bidang ekonomi lokal, khususnya dalam pemetaan usaha kecil dan menengah. Pendekatan ensemble juga dapat membuka peluang penerapan lebih luas pada data sosial-ekonomi lain yang bersifat heterogen, sehingga memberikan kontribusi terhadap pengembangan ilmu data science terapan (Burton et al., 2023). Secara praktis, pemetaan UMKM yang lebih konsisten akan mendukung kebijakan berbasis data, seperti penyaluran bantuan modal, pelatihan keterampilan, dan pengembangan klaster usaha. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi metodologis, tetapi juga berdampak langsung pada pemberdayaan UMKM dan peningkatan ekonomi daerah.

1.2 Perumusan Masalah

Pada bagian perumusan masalah, diuraikan mengenai permasalahan, metode usulan serta pertanyaan penelitian yang diuraikan sebagai berikut:

1.2.1 Permasalahan

UMKM di Kabupaten Sidoarjo membutuhkan metode *Clustering* yang mampu menghasilkan kelompok data yang konsisten dan stabil. Metode *Clustering* tunggal meskipun umum digunakan, cenderung menghasilkan hasil yang bervariasi saat proses pengelompokan diulang, sehingga dapat mengurangi kepercayaan terhadap stabilitas hasil *Clustering*. Ketidak stabilan ini dapat menghambat pengelolaan UMKM yang efektif, seperti dalam perencanaan kebijakan ekonomi lokal atau alokasi sumber daya, karena pemetaan yang tidak andal sulit dijadikan dasar keputusan yang akurat.

1.2.2 Metode Ulasan

Untuk menjawab permasalahan tersebut, penelitian ini mengusulkan penerapan ensemble *Clustering* berbasis weighted voting dengan mengombinasikan algoritma K-means, K-medoids dan Agglomerative Hierarchical Clustering (AHC). Ensemble *Clustering* merupakan pendekatan yang menggabungkan hasil dari beberapa algoritma untuk memperoleh cluster yang lebih baik. Dalam konteks penelitian ini, hasil *Clustering* dari K-means, K-medoids dan Agglomerative Hierarchical Clustering (AHC) akan digabungkan melalui mekanisme weighted voting, sehingga kelemahan masing-masing metode dapat diminimalisasi.

Pendekatan ini diharapkan mampu menghasilkan pemetaan UMKM yang lebih akurat dan konsisten. K-means tetap memberikan efisiensi dalam pengolahan data besar, K-medoids memberikan ketahanan terhadap outlier, sementara AHC mampu mengidentifikasi hubungan antar cluster secara lebih jelas. Dengan adanya mekanisme weighted voting, hasil akhir akan lebih stabil karena mempertimbangkan bobot dari masing-masing algoritma. Metode usulan ini tidak hanya relevan secara teoritis, tetapi juga memiliki potensi penerapan praktis dalam mendukung kebijakan berbasis data di tingkat daerah.

1.2.3 Pertanyaan Penelitian

Berdasarkan identifikasi masalah dan metode yang diusulkan, maka pertanyaan penelitian dalam studi ini dirumuskan sebagai berikut:

1. Berapa hasil k optimal ensemble *Clustering* menggunakan Silhouette Score?

2. Bagaimana hasil analisis dari metode ensemble *Clustering* untuk pemetaan UMKM berbasis weighted voting?

1.3 Tujuan dan Manfaat

Pada bagian tujuan dan manfaat, diuraikan sasaran penelitian serta kontribusi yang diharapkan sebagai berikut:

1.3.1 Tujuan

Penelitian ini bertujuan untuk menerapkan metode ensemble *Clustering* dengan pendekatan weighted voting dalam proses pemetaan UMKM di Kabupaten Sidoarjo. Tujuan utamanya adalah untuk:

1. Menentukan nilai k optimal pada metode ensemble clustering menggunakan Silhouette Score sebagai metrik evaluasi internal.
2. Menganalisis hasil pemetaan UMKM di Kabupaten Sidoarjo yang diperoleh dari penerapan metode ensemble clustering berbasis weighted voting

Dengan tercapainya tujuan tersebut, penelitian ini diharapkan dapat menghasilkan pemetaan data UMKM yang konsisten dan berbasis analisis data, sehingga dapat menjadi dasar informasi dalam mendukung pengelolaan UMKM di Kabupaten Sidoarjo.

1.3.2 Manfaat

1. Penelitian ini diharapkan dapat memperkaya kajian ilmiah di bidang data mining, khususnya terkait penerapan metode ensemble *Clustering* berbasis weighted voting pada pemetaan data UMKM, serta menjadi referensi bagi penelitian selanjutnya yang membahas analisis klaster dan pengelompokan data usaha mikro, kecil, dan menengah.
2. Hasil penelitian ini diharapkan dapat memberikan Gambaran pengelompokan UMKM yang lebih objektif dan berbasis data, sehingga dapat menjadi bahan pertimbangan bagi pemerintah daerah, khususnya Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo, dalam memahami karakteristik UMKM dan mendukung perumusan program pembinaan, pendampingan, dan pemberdayaan UMKM secara lebih tepat sasaran.

1.4 Batasan Masalah

Batasan masalah studi ini adalah sebagai berikut:

1. Data yang digunakan ialah data UMKM di Kabupaten Sidoarjo tahun 2025 yang diperoleh dari Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo, dengan total sebanyak 3.320 data UMKM.
2. Atribut yang digunakan dalam proses analisis terdiri dari lima variabel utama, yaitu modal, omzet per bulan, jumlah tenaga kerja, izin usaha, dan marketplace, yang mencerminkan karakteristik utama dari masing-masing pelaku UMKM.
3. Penelitian ini menggunakan pendekatan *ensemble Clustering* dengan metode weighted voting, yang menggabungkan hasil *Clustering* dari tiga algoritma dasar yaitu K-Means, K-Medoids dan AHC.
4. Pembobotan weighted voting ditentukan berdasarkan nilai Silhouette Score masing-masing algoritma clustering dan dikonversi menggunakan fungsi Softmax sebelum digunakan pada proses Weighted Voting.
5. Evaluasi hasil klasterisasi dilakukan menggunakan metrik evaluasi internal, yaitu *Silhouette Score*, untuk menilai kualitas hasil pengelompokan UMKM.
6. Penelitian ini tidak membahas aspek non-teknis, seperti kebijakan ekonomi, perilaku sosial pelaku UMKM, atau implementasi sistem aplikasi, melainkan berfokus pada analisis data dan pemodelan pengelompokan UMKM menggunakan metode *ensemble Clustering*.

1.5 Sistematika Penulisan

Sistematika penulisan proposal ini penulis susun dengan mengikuti aturan penulisan proposal pada program studi S1 Sistem Informasi Universitas Trunodjoyo Madura. Berikut merupakan metodologi penulisan proposal pada penelitian ini:

BAB I PENDAHULUAN

Bab ini memberikan penjelasan terkait latar belakang, perumusan permasalahan, tujuan dan manfaat penelitian, batasan masalah, serta sistematika penulisan proposal.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori yang meliputi konsep Usaha Mikro, Kecil, dan Menengah (UMKM), data mining, kerangka kerja CRISP-DM, algoritma K-Means, algoritma K-Medoids, Algoritma Agglomerative Hierarchical Clustering (AHC), metode ensemble *Clustering* berbasis weighted voting, serta metrik evaluasi internal klasterisasi seperti *Silhouette Score*. Selain itu, bab ini juga memuat penelitian terdahulu yang relevan sebagai dasar teoritis penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian menggunakan kerangka kerja CRISP-DM yang meliputi pemahaman bisnis dan data UMKM, pengumpulan dan pengolahan data, penerapan algoritma K-Means, K-Medoids dan Agglomerative Hierarchical Clustering (AHC). proses penggabungan hasil menggunakan metode ensemble *Clustering* berbasis weighted voting, serta evaluasi hasil klasterisasi menggunakan metrik evaluasi internal untuk menilai kualitas pengelompokan UMKM di Kabupaten Sidoarjo.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian, analisis data, dan evaluasi menggunakan *Silhouette Coefficient*, serta pembahasan terkait temuan penelitian

BAB V KESIMPULAN DAN SARAN

Penelitian ini menghasilkan kesimpulan dan saran yang dapat diterapkan.

REFRENSI

Bab berisi tentang keterangan mengenai sumber rujukan yang digunakan dalam penyusunan Tugas Akhir. Pada penelitian ini referensi disusun secara otomatis dengan menggunakan aplikasi mendelay