

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor usaha terbesar di Indonesia dan pilar utama ekonomi nasional adalah usaha mikro, kecil, dan menengah (UMKM). UMKM memainkan peran penting dalam menciptakan lapangan kerja, meningkatkan pendapatan masyarakat, dan mendorong pertumbuhan ekonomi baik di tingkat regional maupun nasional [1]. Ada 64,19 juta usaha kecil dan menengah (UMKM) di seluruh Indonesia, menurut data yang dirilis pada bulan Maret 2021 oleh Kementerian Koperasi dan Usaha Kecil Menengah (Kemenkop UKM). Sektor manufaktur kecil dan menengah (UMKM) menempati 97% dari tenaga kerja nasional dan berkontribusi sebesar 61,07% terhadap PDB nasional, yang setara dengan Rp8.573,89 triliun. Data ini menunjukkan betapa pentingnya UMKM dalam mendukung stabilitas ekonomi, terutama di tengah tantangan ekonomi global yang semakin kompleks [2]. Namun, meskipun UMKM memiliki kontribusi besar terhadap perekonomian, sektor ini juga menghadapi berbagai kendala. Beberapa di antaranya adalah keterbatasan modal kerja dan investasi, masalah dalam pemasaran produk, dan masalah dalam pengadaan dan distribusi bahan baku dan input lainnya. Selain itu, banyak UMKM yang menghadapi kesulitan mendapatkan informasi tentang peluang pasar dan kesulitan dalam mengidentifikasi pasar yang tepat serta kurangnya tenaga kerja dengan keahlian tinggi yang berdampak pada sumber daya manusia yang kurang berkualitas dan kemampuan teknologi yang terbatas, yang menghambat kemajuan dan daya saing usaha [3]. Kendala ini dapat menghambat potensi UMKM untuk tumbuh lebih optimal dan memberikan dampak ekonomi yang lebih signifikan. Di sinilah kemajuan teknologi dan analisis data dapat memberikan solusi [4].

Kabupaten Sidoarjo, sebagai salah satu pusat industri utama di Jawa Timur, memiliki jumlah UMKM yang signifikan dan beragam. Wilayah ini dikenal sebagai kawasan industri yang terus berkembang, dengan banyak UMKM yang beroperasi di berbagai sektor ekonomi. Meskipun UMKM memiliki potensi besar, sektor ini masih menghadapi berbagai tantangan baik dari dalam maupun dari luar. Salah satu kendala utama ialah jumlah UMKM yang terus berkembang pesat tanpa adanya

pengelompokan yang jelas. Hal ini menyulitkan Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo dalam menentukan bantuan yang tepat dan merancang program kerja yang sesuai untuk mendukung pengembangan ekonomi lokal. Tanpa adanya pengelompokan yang tepat berdasarkan karakteristik dan kebutuhan masing-masing UMKM, sulit untuk memberikan intervensi yang efektif, baik dalam hal permodalan, pelatihan, maupun peningkatan kapasitas usaha [5][6]. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengelompokkan UMKM secara lebih spesifik, sehingga kebijakan dan program yang dirancang dapat lebih terarah dan berdampak positif bagi perekonomian daerah.

Berdasarkan permasalahan yang dihadapi oleh UMKM, studi ini secara spesifik meneliti penerapan metodologi *CRISP-DM* dan algoritma *clustering DBSCAN* sebagai pendekatan yang efektif untuk mengelompokkan UMKM di Kabupaten Sidoarjo. Metode *CRISP-DM* (*Cross Industry Standard Process for Data Mining*) yaitu salah satu metodologi yang dapat digunakan selama proses analisis data perusahaan kecil dan menengah (UMKM). *CRISP-DM* menawarkan sistem yang teratur yang terdiri dari enam tahap utama: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan. Dengan menggunakan metode ini, proses penelitian dapat dilakukan secara terstruktur, dimulai dari memahami kebutuhan pengembangan UMKM hingga menghasilkan rekomendasi strategis berdasarkan hasil analisis data. Pendekatan ini juga memungkinkan para peneliti untuk memanfaatkan data UMKM yang tersedia secara lebih efektif, mengidentifikasi peluang, dan mengatasi kendala yang ada [7].

Sebuah algoritma yang dapat digunakan dalam fase pemodelan adalah *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*). *DBSCAN* dirilis pertama kali pada tahun 1996 sebagai algoritma klasterisasi yang berfokus pada kepadatan data. Algoritma ini memiliki keunggulan dibandingkan metode klasterisasi lainnya karena mampu menangani data dengan distribusi yang tidak teratur dan mengandung *noise* atau *outlier*. Konsep utama dari *DBSCAN* melibatkan tiga elemen penting yaitu titik inti (*core point*), titik tepi (*border point*), dan *noise*. Titik inti terbentuk ketika jumlah data di sekitarnya memenuhi jumlah minimal poin yang telah

ditetapkan. Titik tepi adalah data yang berada di sekitar titik inti tetapi tidak cukup untuk membentuk kluster baru. Data yang tidak masuk dalam kategori ini disebut *noise* [8][9].

Metode perhitungan jarak dalam penelitian ini yang digunakan sangat memengaruhi hasil klasterisasi. Oleh karena itu, dilakukan perbandingan antara Tiga pendekatan untuk perhitungan jarak utama, yaitu *Euclidean Distance*, *Manhattan Distance*, dan *Hamming Distance*. *Euclidean Distance* adalah pendekatan pengukuran jarak antara dua titik dalam ruang *Euclidean*, dihitung dengan akar kuadrat dari perbedaan jumlah kuadrat setiap dimensi. Metode ini sering digunakan dalam *clustering* untuk menentukan tingkat kemiripan data. Kemudian *Manhattan Distance*, juga sering disebut sebagai "jarak blok kota," menghitung berapa jauh dua titik terpisah dengan menjumlahkan nilai absolut perbedaan pada tiap dimensinya. Jarak ini cocok digunakan untuk dataset yang lebih besar atau ketika keakuratan dimensi tertentu lebih penting daripada yang lain. Sedangkan *Hamming Distance* mengukur jumlah posisi di mana dua vektor biner berbeda. Digunakan pada data kategorikal atau biner, metode ini berguna untuk mengidentifikasi perbedaan antara dua string atau array [10][11]. Selanjutnya hasil *clustering* menggunakan *algoritma DBSCAN* dievaluasi dengan metrik *silhouette coefficient*. Metrik ini digunakan untuk mengukur sejauh mana setiap data sesuai dengan kluster tempatnya di antara kluster lainnya. Nilai *silhouette coefficient* berkisar dari nol hingga satu, dengan nilai yang lebih tinggi di sekitar satu menunjukkan kualitas kluster yang lebih baik [12].

Berdasarkan penelitian terkait, klasterisasi wilayah Pulau Sumatera yang rawan bencana menggunakan pendekatan DBSCAN bertujuan mengidentifikasi wilayah dengan tingkat kerawanan berbeda. Hasil analisis menunjukkan dua *cluster* dengan parameter $MinPts = 4$ dan $Eps = 0,18$, yaitu *cluster* 1 dengan kerawanan sedang (102 kabupaten/kota), *cluster* 2 dengan kerawanan rendah (8 kabupaten/kota), dan 44 kabupaten/kota yang dianggap *noise* karena kerawanan bencana tinggi [13].

Tinjauan penelitian terkait berikutnya yaitu, pengelompokan Metode Self Organizing Maps (SOM) dan DBSCAN telah digunakan untuk menilai kebutuhan Keluarga Berencana (KB) di Nusa Tenggara Barat. Penelitian ini bertujuan untuk

mengidentifikasi daerah dengan tingkat *unmet need* tinggi, sedang, dan rendah. Hasil analisis menggunakan DBSCAN dengan parameter *minpts* 3 dan *epsilon* 1.20 menghasilkan dua *cluster*, yaitu *cluster* 1 memiliki 97 kecamatan, dan *cluster* 2 memiliki 4 kecamatan dan 16 kecamatan yang termasuk dalam kategori *noise* karena memiliki data yang tidak terkelompok [14].

Dengan memanfaatkan teknologi analisis data, hasil pengelompokan UMKM dapat digunakan secara lebih efektif dalam menyusun kebijakan pengembangan yang tepat sasaran. Hal ini mencakup alokasi dana yang lebih efisien, penyediaan pelatihan yang sesuai dengan kebutuhan pelaku usaha, serta strategi penyaluran kredit yang lebih optimal. Dengan demikian, kebijakan yang dihasilkan dapat mendukung pertumbuhan UMKM secara berkelanjutan dan meningkatkan persaingan ekonomi daerah.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Dari uraian latar belakang yang telah dijelaskan, Kabupaten Sidoarjo menghadapi kesulitan dalam mengelola dan mendukung perkembangan UMKM secara efektif, akibat ketiadaan pengelompokan yang jelas berdasarkan karakteristik dan kebutuhan masing-masing unit usaha. Dengan jumlah UMKM yang terus berkembang pesat, tanpa adanya pengelompokan yang tepat, sulit bagi Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo untuk memahami perbedaan potensi, dan tantangan yang dihadapi oleh setiap kelompok UMKM. Hal ini menyulitkan dalam perencanaan kebijakan yang sesuai, serta memperlambat pemberian dukungan yang efektif, baik dalam hal permodalan, pelatihan, maupun pengembangan kapasitas usaha yang lebih terfokus pada kebutuhan spesifik tiap kelompok UMKM.

1.2.2 Metode Usulan

Penelitian ini menyarankan untuk menggunakan algoritma DBSCAN, karena keunggulannya dalam mengelompokkan data berdasarkan kepadatan dan kemampuannya dalam mendeteksi outlier tanpa memerlukan jumlah klaster yang ditentukan sebelumnya. Dalam penelitian ini, pengelompokan dilakukan dengan menggunakan tiga metode perhitungan jarak, yaitu jarak euclidean, jarak manhattan, dan jarak hamming, yang akan dibandingkan untuk menilai dampaknya terhadap hasil

klasterisasi. Kualitas klasterisasi akan dievaluasi menggunakan *silhouette coefficient* dan *davies-bouldin index*, yang mengukur sejauh mana data dalam klaster memiliki kesamaan dibandingkan klaster lainnya.

1.2.3 Pertanyaan Penelitian

Dengan mempertimbangkan kasus yang telah diidentifikasi dan pendekatan yang telah diajukan, berikut adalah beberapa pertanyaan yang muncul:

1. Bagaimana hasil eksperimen nilai parameter *epsilon* dan jumlah minimum tetangga (*minpts*) pada algoritma DBSCAN terhadap hasil klasterisasi dengan metode jarak *Euclidean*, *Manhattan*, dan *Hamming*?
2. Apa perbedaan hasil klasterisasi UMKM di Kabupaten Sidoarjo yang diterapkan dengan algoritma DBSCAN menggunakan metode jarak *Euclidean*, *Manhattan*, dan *Hamming*, berdasarkan evaluasi dengan *silhouette coefficient* dan *davies-bouldin index*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

1. Melakukan analisis hasil eksperimen nilai parameter *epsilon* dan jumlah minimum tetangga (*minpts*) pada algoritma DBSCAN terhadap hasil klasterisasi UMKM di Kabupaten Sidoarjo dengan menggunakan metode jarak *Euclidean*, *Manhattan*, dan *Hamming*.
2. Membandingkan hasil klasterisasi UMKM di Kabupaten Sidoarjo berdasarkan metode jarak *Euclidean*, *Manhattan*, dan *Hamming* menggunakan *silhouette coefficient* dan *davies-bouldin index* untuk menentukan metode terbaik, sehingga mempermudah Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo dalam menganalisis UMKM untuk mendukung perencanaan program bantuan dan perancangan program kerja yang lebih tepat sasaran.

1.3.2 Manfaat

1. Studi ini dapat mengembangkan wawasan terkait penerapan algoritma DBSCAN dalam analisis klasterisasi, terutama dalam menganalisis hasil eksperimen nilai parameter *epsilon* dan jumlah minimum tetangga (*minpts*)

terhadap hasil klasterisasi, serta memperluas pemahaman konsep data mining melalui perbandingan metode jarak *Euclidean*, *Manhattan*, dan *Hamming*.

2. Hasil penelitian ini diharapkan dapat membantu Dinas Koperasi dan Usaha Mikro Kabupaten Sidoarjo dalam memberikan rekomendasi yang sesuai dengan kebutuhan setiap kelompok UMKM untuk mendukung pengembangan ekonomi lokal serta menyusun kebijakan berbasis data yang lebih efektif.

1.4 Batasan Masalah

Batasan masalah studi ini adalah sebagai berikut:

1. Data yang digunakan ialah data UMKM di Kabupaten Sidoarjo pada tahun 2024 yang diperoleh dari Dinas Koperasi dan Usaha mikro Kabupaten Sidoarjo.
2. Data yang digunakan memiliki 5 atribut utama yaitu modal, omset per bulan, jumlah tenaga kerja, izin usaha, marketplace.
3. Data diolah menggunakan algoritma *DBSCAN* untuk klasterisasi.
4. *Dataset* yang digunakan terdiri dari 1479 data UMKM yang ada di Kabupaten Sidoarjo.
5. Membandingkan tiga metode pengukuran jarak yaitu jarak euclidean, jarak manhattan, dan jarak hamming untuk klasterisasi *DBSCAN*.
6. Hasil klasterisasi dievaluasi dengan memanfaatkan pendekatan *silhouette coefficient* dan *davies-bouldin index* untuk mengukur kualitas kluster.

1.5 Sistematika Penulisan

Sistematika penulisan dari penelitian ini dapat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini memberikan penjelasan terkait latar belakang, perumusan permasalahan, tujuan, manfaat dari penelitian ini, batasan masalah, dan tata cara penyusunan proposal.

BAB II TINJAUAN PUSTAKA

Bab ini mencakup kajian literatur terkait konsep UMKM, data mining, *CRISP-DM*, *algoritma DBSCAN*, metode perhitungan

jarak *Euclidean*, *Manhattan*, dan *Hamming*, metode evaluasi model *silhouette coefficient* dan *davies-bouldin index* serta penelitian terdahulu sebagai dasar teoritis penelitian.

BAB III METODOLOGI PENELITIAN

Tahapan penelitian diuraikan dalam bab ini menggunakan kerangka *CRISP-DM*, mulai dari pemahaman bisnis, pengumpulan dan pengolahan data UMKM, penerapan *algoritma DBSCAN* dengan tiga metode jarak, evaluasi hasil klasterisasi, hingga penerapan hasil.

BAB IV HASIL DAN PEMBAHASAN

Hasil penelitian akan diuraikan pada bab ini serta pembahasan terkait melalui temuan-temuan yang diperoleh.

BAB V PENUTUP

Penelitian ini menghasilkan kesimpulan dan saran yang dapat diterapkan.

PUSTAKA Daftar referensi yang digunakan dalam penelitian disertakan dalam bagian ini.