

BAB 1

PENDAHULUAN

1.1 LATAR BELAKANG

Yogyakarta menjadi salah satu destinasi utama di Indonesia bagi wisatawan domestik dan internasional (Badan Pusat Statistik, 2024). Selain objek wisata budaya dan alam, Yogyakarta juga menawarkan beragam produk ekonomi kreatif yang berasal dari UMKM (Usaha mikro, kecil, menengah) di berbagai desa, seperti kerajinan tangan, kuliner khas, dan seni budaya (Jabawidhiartha et al., 2024).

Berdasarkan data dari Kementerian Pertanahan dan Tata Ruang, terdapat 25.065 UMKM yang tersebar di berbagai wilayah di Yogyakarta (KemenBPN, 2024). Namun, jumlah UMKM yang sangat besar dan tersebar di berbagai wilayah serta data yang belum cukup terstruktur membuat pemerintah daerah dan pemangku kepentingan lainnya menghadapi tantangan dalam pengelompokan dan perumusan strategi pengembangan yang tepat sasaran. Oleh karena itu, diperlukan pendekatan berbasis data untuk mengelompokkan kategori UMKM berdasarkan karakteristik tertentu, untuk mengatasi permasalahan tersebut, diperlukan sebuah visualisasi pengelompokan kategori UMKM yang optimal guna mempermudah akses informasi bagi para pemangku kebijakan dan juga wisatawan dalam melihat kategori UMKM.

Dalam membangun pengelompokan kategori UMKM, digunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan kerangka kerja standar dalam proyek data mining lintas industri untuk memperoleh wawasan berharga melalui teknik seperti klasifikasi dan *clustering* (Sastya & Nugraha, 2023). Diharapkan dengan penerapan metode CRISP-DM dalam pengelompokan kategori UMKM, analisis data dapat dilakukan secara sistematis dan terstruktur, sehingga menghasilkan informasi yang akurat dan relevan bagi industri (Wahyusari & Wardani, 2023).

Salah satu metode yang dapat digunakan untuk melakukan pengelompokan data adalah metode *clustering*. *Clustering* merupakan teknik dalam data mining

yang bertujuan untuk mengelompokkan objek berdasarkan kesamaan karakteristik yang dimiliki (Ramadhan & Astuti, 2024). Dalam penelitian ini, metode *clustering* digunakan sebagai pendekatan utama dalam mengelompokkan kategori UMKM, dengan menerapkan perbandingan algoritma *K-Means* dan *Agglomerative clustering* sebagai teknik yang diterapkan.

K-Means menggunakan pendekatan berbasis *centroid*, di mana data dikelompokkan berdasarkan kedekatan jarak ke pusat kluster yang telah ditentukan sebelumnya. Metode ini lebih efisien untuk dataset besar dan memiliki kecepatan komputasi yang lebih tinggi dibandingkan *Agglomerative clustering* (Faujia et al., 2022). Pada Penelitian yang di lakukan Marcelina et al (2023) dengan judul “Analisis Kluster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma *K-Means Clustering*” menghasilkan akurasi 85,5% dan tingkat error 14,5%.

Sebaliknya, *Agglomerative clustering* menerapkan pendekatan hierarki yang mengelompokkan data secara bertahap berdasarkan kesamaan, tanpa memerlukan jumlah kluster yang telah ditentukan sebelumnya. Hal ini membuatnya lebih cocok untuk dataset yang memiliki struktur hierarkis yang kompleks, tetapi metode ini memiliki kelemahan dalam efisiensi komputasi karena membutuhkan lebih banyak waktu untuk proses penggabungan kluster (Iyan Yulianti et al., 2023). Penelitian yang di lakukan oleh Tinton et al (2021) yang berjudul “*Clustering* pada data sentimen BPJS kesehatan menggunakan algoritma *agglomerative hierarchical clustering average linkage*” menghasilkan 8 cluster dengan akurasi 98%.

Dari kedua penelitian tersebut terdapat fakta bahwa *K-Means* lebih unggul dalam menentukan jumlah kluster optimal untuk data dengan pola distribusi yang jelas, sedangkan *Agglomerative clustering* lebih efektif dalam menangkap struktur hierarkis yang kompleks dalam data. Kemudian penelitian yang di lakukan oleh Ramadhan & Astuti (2024) yang berjudul “ Perbandingan Kinerja Algoritma *K-Means* dan *Agglomerative clustering* Untuk Segmentasi Penjualan Online Pada Customer Retail “, dengan membandingkan dua algoritma tersebut menghasilkan akurasi dimana *K-Means* mendapat *silhouette score* 50% dan *Agglomerative clustering* sebanyak 63% yang menunjukkan *Agglomerative clustering* lebih baik dibandingkan *K-Means* dalam pengelompokan *customer retail*.

Berdasarkan dari hasil penelitian tersebut, penelitian ini melakukan upaya untuk menentukan metode klasterisasi yang paling optimal dengan membandingkan algoritma *K-Means* dan *Agglomerative clustering*. Fokus utama penelitian ini adalah mengevaluasi performa kedua algoritma dalam konteks pengelompokan data kategori UMKM, guna mengetahui metode mana yang lebih efektif dalam menyajikan struktur kelompok yang efektif. Penelitian ini menggunakan hasil evaluasi terbaik sebagai dasar untuk visualisasi pengelompokan kategori UMKM. Visualisasi ini diharapkan dapat membantu para pemangku kepentingan dan juga wisatawan dalam melihat kategori UMKM, sehingga meningkatkan pengalaman berwisata sekaligus mendorong pertumbuhan ekonomi lokal.

1.2 PERUMUSAN MASALAH

Salah satu tantangan dalam pengembangan UMKM di Provinsi Daerah Istimewa Yogyakarta adalah banyaknya jumlah UMKM yang tersebar di berbagai wilayah, namun belum disertai dengan pengelompokan yang terstruktur dan berbasis data. Meskipun data mengenai UMKM telah tersedia, penyajiannya yang belum optimal dan kurang terorganisir menyulitkan pemangku kebijakan dalam melakukan analisis serta merumuskan strategi pengembangan yang tepat sasaran. Selain itu, wisatawan juga mengalami kesulitan dalam melihat kategori UMKM yang sesuai dengan minat dan preferensi mereka akibat keterbatasan informasi. Untuk mengatasi permasalahan tersebut, diperlukan metode pengelompokan data (*clustering*) yang mampu mengelompokkan kategori UMKM berdasarkan karakteristik tertentu secara efektif.

1.3 BATASAN MASALAH

Dalam penelitian ini, ruang lingkup dibatasi agar fokus kajian menjadi lebih terarah dan sesuai dengan tujuan yang ingin dicapai. Penelitian hanya mencakup data UMKM yang berada di wilayah Provinsi Daerah Istimewa Yogyakarta, meliputi lima kabupaten atau kota, yaitu Kota Yogyakarta, Kabupaten Sleman, Kabupaten Bantul, Kabupaten Kulon Progo, dan Kabupaten Gunungkidul. Variabel yang digunakan dalam proses pengelompokan terbatas pada nama desa sebagai

representasi lokasi geografis, dan sektor ekonomi kreatif yang difokuskan pada tiga kategori utama, yaitu kuliner, kerajinan tangan, dan seni budaya.

Algoritma yang digunakan dalam proses clustering dibatasi pada dua metode, yaitu *K-Means* dan *Agglomerative clustering*, tanpa membandingkan dengan algoritma lain seperti *DBSCAN* atau metode berbasis *density* dan *neural network*. Data yang digunakan bersifat sekunder, diperoleh dari instansi atau sumber resmi yang tersedia, dan tidak dilakukan pengumpulan data primer secara langsung di lapangan. Evaluasi hasil clustering dilakukan melalui analisis visual dan metrik seperti *silhouette score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. tanpa melibatkan uji coba langsung terhadap pengguna atau validasi berbasis pengalaman lapangan. Penelitian ini juga dibatasi pada analisis dan interpretasi hasil clustering untuk kepentingan pengambilan keputusan oleh pemerintah daerah serta sebagai informasi pendukung bagi wisatawan, tanpa mencakup pengembangan sistem atau aplikasi digital berbasis hasil klasifikasi tersebut.

1.4 PERTANYAAN PENELITIAN

1. Apa hasil perbandingan algoritma *K-Means* dan *Agglomerative clustering* dalam mengelompokkan kategori UMKM di Provinsi Daerah Istimewa Yogyakarta berdasarkan nama desa, dan sektor ekonomi kreatif?
2. Bagaimana hasil pengelompokan yang dihasilkan oleh algoritma ini dapat membantu pemangku kebijakan dan juga wisatawan dalam melihat kategori UMKM seperti sektor kuliner, kerajinan tangan, dan seni budaya?

1.5 TUJUAN PENELITIAN

Tujuan penelitian ini adalah melihat algoritma mana yang lebih baik dalam menghasilkan pengelompokan kategori UMKM di Provinsi Daerah Istimewa Yogyakarta menggunakan metode *clustering* dengan perbandingan algoritma *K-Means* dan *Agglomerative clustering*, yang dapat dimanfaatkan baik oleh pemangku kebijakan maupun wisatawan.

1.6 MANFAAT HASIL PENELITIAN

Dengan dilakukannya Perbandingan algoritma *K-Means* dan *Aggroremative Clustering* dalam pengelompokan kategori UMKM, diharapkan dapat memberikan manfaat diantaranya:

1. Hasil pengelompokan ini dapat dijadikan sebagai referensi dalam merancang kebijakan strategis yang bertujuan untuk mendukung pertumbuhan sektor ekonomi kreatif dan pariwisata.
2. Hasil pengelompokan diharapkan dapat membantu pemerintah dan juga wisatawan dalam melihat kategori UMKM di provinsi Daerah Istimewa Yogyakarta.