

KLASTERISASI UMKM MENGGUNAKAN SPECTRAL-KMEANS BERDASARKAN FINANSIAL, TENAGA KERJA, DAN LEGALITAS

Wina Witanti ¹⁾, Gunawan Abdillah ²⁾, Edvin Ramadhan ³⁾

1) Program Studi Teknik Informatika Universitas Jenderal Achmad Yani

E-mail: witanti@gmail.com

2) Program Studi Teknik Informatika Universitas Jenderal Achmad Yani

E-mail: gna@unjani.ac.id

3) Program Studi Teknik Informatika Universitas Jenderal Achmad Yani

E-mail: edvin.ramadhan@gmail.com

Abstrak

Indonesia memiliki lebih dari 64 juta UMKM yang berperan strategis dalam perekonomian namun sebagian besar kajian klasterisasi UMKM di Indonesia masih berfokus pada aspek tunggal seperti omset atau aset, dan belum menggabungkan indikator finansial, tenaga kerja, serta legalitas secara terintegrasi. Penelitian ini menganalisis 13.564 data UMKM dari berbagai sektor menggunakan pendekatan Spectral Embedding untuk reduksi dimensi dan KMeans untuk klasterisasi, setelah melalui tahap pembersihan data, transformasi logaritma pada variabel finansial, label encoding atribut kategorikal, rekayasa fitur rasio keuangan dan tenaga kerja, serta standarisasi. Klasterisasi menghasilkan tiga kelompok utama, yakni klaster pertama berisi UMKM skala kecil dengan aset, omset, dan jumlah tenaga kerja terbatas serta mayoritas belum berstatus legal, klaster kedua terdiri atas UMKM menengah dengan performa finansial dan tenaga kerja sedang, sementara klaster ketiga mencakup UMKM mapan yang didukung legalitas dan indikator finansial yang tinggi. Evaluasi menggunakan Silhouette Score sebesar 0,403 dan Davies-Bouldin Index sebesar 0,780 menandakan pemisahan klaster cukup baik dan valid. Temuan ini menunjukkan bahwa kombinasi Spectral Embedding dan KMeans efektif dalam mengidentifikasi segmentasi UMKM multidimensi, sehingga dapat menjadi dasar bagi penyusunan kebijakan pembinaan dan pengembangan UMKM secara lebih terarah di Indonesia.

Kata kunci: UMKM, Klasterisasi, Spectral Embedding, KMeans, Data Mining.

Pendahuluan

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu pilar utama perekonomian Indonesia karena kontribusinya yang signifikan dalam penyerapan tenaga kerja, penciptaan produk domestik, serta pemerataan pembangunan ekonomi [1]. Data Kementerian Koperasi dan UKM menunjukkan bahwa jumlah UMKM terus mengalami peningkatan setiap tahun, namun perkembangan tersebut tidak selalu diikuti dengan kinerja yang merata di seluruh sektor [2]. Banyak UMKM masih menghadapi tantangan serius, seperti keterbatasan akses permodalan, hambatan dalam memperluas pasar, rendahnya kapasitas manajerial, serta perbedaan status legalitas yang memengaruhi daya saing [3]. Kondisi ini menunjukkan perlunya pendekatan analitis yang mampu mengelompokkan UMKM berdasarkan karakteristik mendasar, sehingga strategi pengembangan dan kebijakan yang diterapkan dapat lebih tepat sasaran sesuai dengan realitas lapangan [4].

Permasalahan utama yang muncul adalah bagaimana membangun pemetaan klaster UMKM yang mempertimbangkan indikator relevan, meliputi kondisi finansial, jumlah tenaga kerja, dan status legalitas. Selama ini, pengelompokan UMKM umumnya masih dilakukan secara administratif atau berdasarkan kategori usaha konvensional, sehingga belum mampu merepresentasikan heterogenitas dan kompleksitas yang terdapat dalam populasi UMKM [5]. Pengelompokan yang lebih mendalam sangat diperlukan untuk menjadi dasar penyusunan strategi intervensi, pemberian dukungan, serta perumusan kebijakan yang efektif [6].

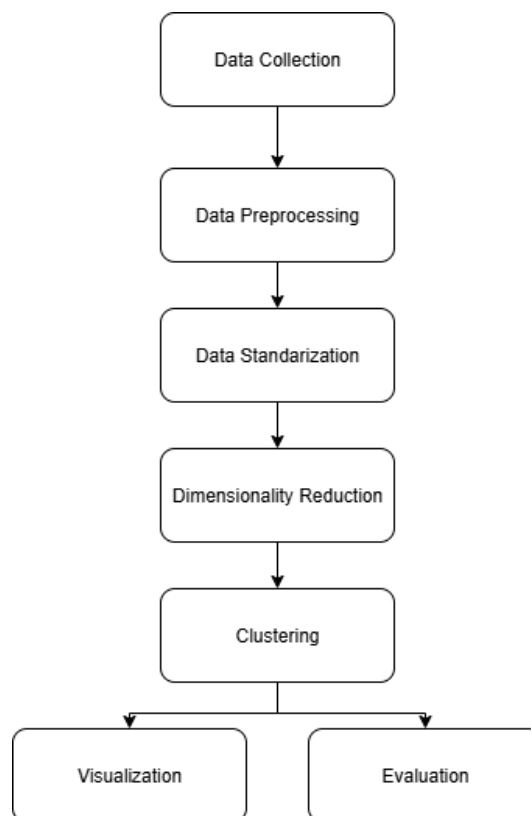
Berbagai studi sebelumnya telah menekankan pentingnya klasterisasi UMKM. Penelitian di Kabupaten Kutai Timur menggunakan metode *Analytic Hierarchy Process* untuk merumuskan strategi pengembangan berbasis klaster, yang menghasilkan tiga kelompok strategi sesuai dengan kriteria produksi, permodalan, dan pemasaran [7]. Penelitian lain di Jawa Tengah menerapkan

algoritma K-Means untuk menganalisis pembiayaan UMKM melalui program Kredit Usaha Rakyat (KUR), dengan temuan bahwa sektor perdagangan merupakan kluster dominan [8]. Sementara itu, studi di Kota Yogyakarta menggunakan metode *K-Medoids* untuk memetakan UMKM bidang kerajinan dan fashion, dengan hasil berupa dua kluster yang dievaluasi melalui *silhouette coefficient* dan menunjukkan pola struktural yang sedang [9]. Meskipun memberikan kontribusi penting, sebagian besar penelitian terdahulu masih mengandalkan metode klasterisasi konvensional dan belum banyak mengeksplorasi pendekatan berbasis graf spektral yang lebih adaptif terhadap distribusi data nonlinier.

Untuk mengisi celah tersebut, penelitian ini mengimplementasikan pendekatan Spectral Embedding untuk merepresentasikan data UMKM dalam ruang berdimensi rendah, yang kemudian diklusterkan menggunakan algoritma K-Means. Pendekatan ini dipandang lebih unggul karena mampu menangkap struktur data yang kompleks dibandingkan algoritma konvensional yang cenderung terbatas pada pola kluster berbentuk sferis [10]. Dengan memanfaatkan representasi graf, metode ini ditujukan untuk menggambarkan variasi UMKM yang dipengaruhi oleh kombinasi faktor finansial, tenaga kerja, dan legalitas secara lebih komprehensif. Tujuan penelitian ini adalah menghasilkan pemetaan kluster UMKM yang akurat dan informatif sehingga dapat digunakan sebagai dasar rekomendasi bagi pengembangan dan perumusan kebijakan yang lebih tepat sasaran.

Metodologi Penelitian

Alur penelitian ini disusun secara sistematis mulai dari pengumpulan data hingga visualisasi hasil. Tahapan utama penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram alur penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis sebagaimana ditunjukkan pada kerangka metode. Tahap pertama adalah pengumpulan data (*data collection*), yaitu dengan memanfaatkan dataset UMKM yang berisi informasi finansial, jumlah tenaga kerja, serta status legalitas usaha. Data tersebut kemudian melalui proses *preprocessing* untuk memastikan kualitasnya, mencakup penghapusan nilai yang hilang atau tidak valid, pembersihan data dari entri anomali, serta penghilangan atribut yang tidak relevan seperti identitas unik maupun nama usaha.

Selanjutnya dilakukan proses *encoding* terhadap variabel kategorikal, misalnya jenis usaha, status legalitas, dan keterlibatan marketplace, dengan menggunakan teknik Label Encoding. Untuk variabel numerik seperti aset, omset, laba, dan biaya karyawan, dilakukan transformasi logaritmik agar distribusinya lebih normal serta konversi ke bentuk numerik untuk menghindari inkonsistensi tipe data. Selain itu, penelitian ini juga melakukan *feature engineering* dengan membangun variabel turunan, seperti rasio laba terhadap aset, rasio biaya terhadap omset, omset per tenaga kerja, dan umur usaha, yang dipandang mampu memberikan informasi tambahan dalam membedakan karakteristik antar UMKM.

Tahap berikutnya adalah standarisasi data menggunakan *StandardScaler* agar seluruh variabel berada pada skala yang sebanding dan tidak menimbulkan bias dalam proses pembentukan kluster. Setelah itu dilakukan reduksi dimensi dengan metode *Spectral Embedding*, yang berfungsi untuk memproyeksikan data berdimensi tinggi ke dalam ruang berdimensi rendah dengan mempertahankan struktur hubungan antar data. Pendekatan ini dipilih karena mampu merepresentasikan pola nonlinier yang lebih kompleks dibandingkan teknik reduksi konvensional.

Hasil transformasi kemudian dianalisis menggunakan algoritma *K-Means* untuk membentuk kluster UMKM. Jumlah kluster ditentukan berdasarkan pengujian nilai k yang optimal, dan pada penelitian ini digunakan tiga kluster sebagai representasi utama. Untuk memastikan kualitas hasil pengelompokan, dilakukan evaluasi menggunakan dua metrik, yaitu *Silhouette Score* dan *Davies-Bouldin Index*. *Silhouette Score* digunakan untuk menilai seberapa baik objek berada dalam klusternya, sedangkan *Davies-Bouldin Index* mengukur tingkat kemiripan antar kluster.

Tahap terakhir adalah visualisasi hasil klusterisasi dengan *scatter plot* dua dimensi, yang memudahkan interpretasi pola distribusi UMKM sesuai dimensi hasil *Spectral Embedding*. Visualisasi ini digunakan sebagai pendukung analisis agar perbedaan antar kluster dapat dipahami dengan lebih jelas, sekaligus menjadi dasar dalam menarik kesimpulan mengenai karakteristik kelompok UMKM yang terbentuk.

Hasil dan Pembahasan

Dataset penelitian ini terdiri dari 13.564 baris data UMKM yang memuat informasi mengenai identitas usaha, karakteristik tenaga kerja, indikator finansial, platform pemasaran (marketplace), kapasitas produksi, status legalitas, hingga tahun berdiri. Data yang digunakan memiliki variasi cukup besar antar UMKM, baik dari sisi aset, omset, maupun jumlah tenaga kerja. Untuk memberikan gambaran, sebagian kecil dari dataset ditampilkan pada Tabel 1.

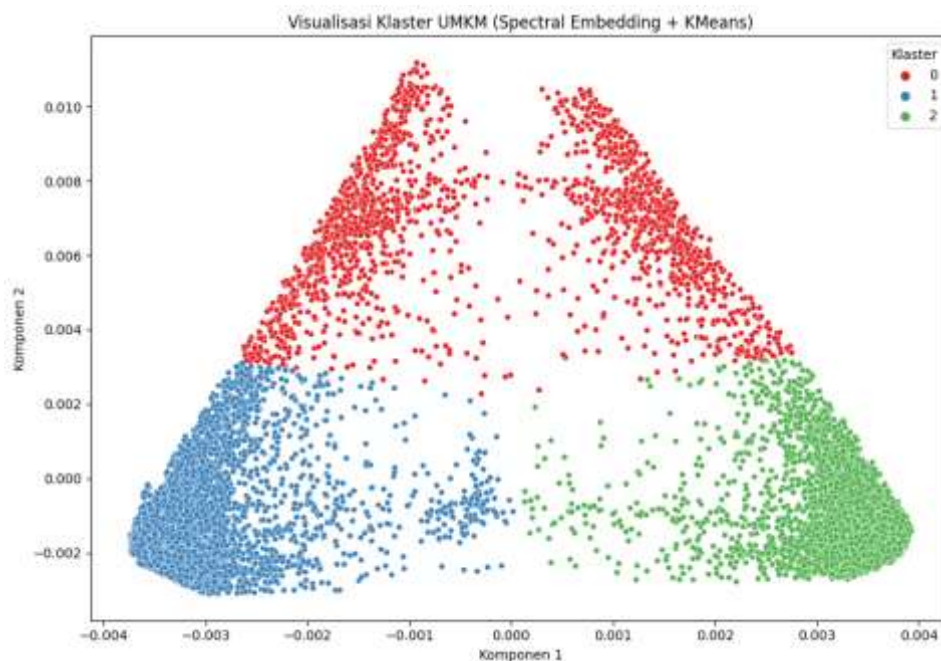
Tabel 1. Tabel Sampel Dataset UMKM

Nama_Usaha	Jenis_Usaha	Aset	Legalitas	Tahun_Berdiri
UD. Alif Pamungkas	Kesehatan	Rp5.497.149	Belum Terdaftar	2015
UD. Zidanar Panji	Perdagangan	Rp7.398.384	Terdaftar	2014
UD. Mika Bagaskara	Perdagangan	Rp8.132.674	Terdaftar	2003
UD. Farhan Mulyo	Pendidikan	Rp7.888.385	Belum Terdaftar	2009

Sebelum dilakukan klusterisasi, data tersebut melalui *preprocessing* sesuai rancangan metode penelitian. Proses ini mencakup penghapusan nilai kosong dan data anomali, transformasi variabel numerik dengan logaritma, pengkodean variabel kategorikal dengan *Label Encoding*, serta

standarisasi menggunakan *StandardScaler*. Selain itu, dilakukan pula rekayasa fitur seperti perhitungan rasio laba terhadap aset, rasio biaya terhadap omset, omset per tenaga kerja, omset per pelanggan, serta umur usaha.

Tahapan selanjutnya adalah reduksi dimensi menggunakan *Spectral Embedding* dan klasterisasi dengan algoritma *K-Means*. Jumlah klaster optimal ditentukan sebanyak tiga kelompok. Evaluasi model menunjukkan bahwa kualitas klasterisasi cukup baik, dengan nilai *Silhouette Score* sebesar 0,740 dan *Davies-Bouldin Index* sebesar 0,394. Visualisasi hasil klasterisasi dapat dilihat pada Gambar 1, yang memperlihatkan distribusi tiga klaster UMKM berdasarkan representasi hasil *Spectral Embedding*.



Gambar 1. Visualisasi klaster UMKM menggunakan Spectral Embedding + KMeans

Dari hasil pengelompokan, terbentuk tiga klaster utama dengan karakteristik yang berbeda. Ringkasan hasil klasterisasi dapat dilihat pada Tabel 2. Klaster pertama didominasi oleh UMKM dengan aset dan omset relatif kecil, tenaga kerja terbatas, serta sebagian besar belum memiliki legalitas. Klaster kedua berisi UMKM dengan kapasitas menengah, legalitas bervariasi, dan proporsi biaya terhadap omset yang lebih besar. Klaster ketiga berisi UMKM dengan skala lebih besar, tenaga kerja lebih banyak, omset dan aset yang tinggi, serta mayoritas sudah memiliki legalitas formal.

Tabel 2. Tabel Ringkasan Karakteristik Klaster UMKM

Klaster	Karakteristik Umum
0 (Merah)	UMKM skala besar, omset tinggi, aset besar, tenaga kerja banyak, mayoritas legal
1 (Biru)	UMKM skala kecil/mikro, aset dan omset rendah, tenaga kerja sedikit, mayoritas belum legal
2 (Hijau)	UMKM skala menengah, variasi legalitas, biaya relatif besar terhadap omset

Temuan ini memperlihatkan bahwa metode *Spectral-KMeans* mampu menangkap heterogenitas UMKM secara lebih baik. Setiap kluster dapat digunakan sebagai dasar kebijakan: kluster kecil diarahkan pada dukungan permodalan dan legalisasi, kluster menengah pada peningkatan efisiensi manajerial, sedangkan kluster besar pada ekspansi pasar dan kemitraan strategis. Dengan demikian, hasil penelitian ini tidak hanya memberikan kontribusi metodologis tetapi juga implikasi praktis dalam pengembangan UMKM di Indonesia

Kesimpulan

Penelitian ini berhasil mengelompokkan 13.564 data UMKM di Indonesia berdasarkan indikator finansial, tenaga kerja, dan legalitas dengan menggunakan pendekatan *Spectral Embedding* yang dipadukan dengan algoritma *K-Means*. Hasil evaluasi menunjukkan kualitas klusterisasi yang baik, dengan nilai *Silhouette Score* sebesar 0,740 dan *Davies-Bouldin Index* sebesar 0,394. Dari proses tersebut terbentuk tiga kluster utama, yaitu kluster UMKM skala kecil dengan aset dan omset rendah serta mayoritas belum memiliki legalitas, kluster menengah dengan variasi legalitas dan beban biaya yang relatif besar, serta kluster besar dengan aset dan omset tinggi, tenaga kerja lebih banyak, dan legalitas yang lebih mapan.

Temuan ini memperlihatkan bahwa metode *Spectral-Kmeans* mampu menangkap heterogenitas UMKM secara lebih akurat dibandingkan pendekatan klusterisasi konvensional. Secara praktis, hasil klusterisasi dapat menjadi dasar bagi perumusan kebijakan pengembangan UMKM yang lebih tepat sasaran, seperti dukungan permodalan dan legalisasi bagi UMKM kecil, peningkatan efisiensi manajerial pada UMKM menengah, serta dorongan ekspansi pasar dan kemitraan strategis bagi UMKM besar. Dengan demikian, penelitian ini memberikan kontribusi metodologis sekaligus implikasi praktis dalam upaya penguatan peran UMKM sebagai pilar utama perekonomian Indonesia.

Ucapan Terima kasih

Ucapan terima kasih disampaikan kepada LPPM Universitas Jenderal Achmad Yani (Unjani) atas dukungan pembiayaan penelitian ini.

Daftar Pustaka

- [1] Yulya Ammi Hapsari, Putri Apriyanti, Aldi Hermiyanto, and Fahrur Rozi, “Analisa Peran UMKM Terhadap Perkembangan Ekonomi di Indonesia,” *Jurnal Manajemen dan Ekonomi Kreatif*, vol. 2, no. 4, pp. 53–62, Oct. 2024, doi: 10.59024/jumek.v2i4.464.
- [2] T. N. Anisah, “Analisis Potensi dan Tantangan Pengembangan UMKM Angkringan sebagai Komponen Ekonomi Kreatif di Kecamatan Kemantren, Yogyakarta,” *ADARMA: Jurnal Pengabdian Masyarakat Universitas Janabadra*, vol. 10, no. 2, pp. 54–59, Jan. 2024, doi: 10.37159/jad.v10i2.7.
- [3] N. Hafizah, A. Lia Hananto, F. Nurapriani, and E. Novalia, “SEGMENTASI NASABAH UMKM BERDASARKAN KINERJA DAN KEUNTUNGAN MENGGUNAKAN K-MEANS CLUSTERING,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, pp. 8661–8665, Jul. 2025, doi: 10.36040/jati.v9i5.15056.
- [4] C. B. Ajarwiro, M. Maimunah, and P. Sukmasetya, “Analisis Pengelompokan UMKM Berdasarkan Kategori Menggunakan Algoritma K-Means dan K-Medoids,” *CESS*

- [5] L. F. Saskya and R. A. N. Apriyanto, “IMPLEMENTASI FUZZY C-MEANS CLUSTERING DALAM PENGELOMPOKAN UMKM DI KELURAHAN PANGONGANGAN KOTA MADIUN,” *Power Elektronik : Jurnal Orang Elektro*, vol. 11, no. 2, p. 204, Jul. 2022, doi: 10.30591/polektro.v12i1.3713.
- [6] I. A. Romdoni, N. Purwanti, A. Hermawan, R. N. Haryadi, L. Pusvisasari, and I. Irawati, “Pelatihan Keuangan dan Literasi Bisnis untuk UMKM: Pengabdian yang Berdampak pada Stabilitas Usaha Mikro di Sumedang,” *Jabdimas: Jurnal Pengabdian Masyarakat*, vol. 1, no. 2, pp. 39–43, Dec. 2023, doi: 10.56457/jabdimas.v1i2.83.
- [7] M. Y. Bustomi and J. Suryanto, “Strategi Pengembangan Klaster UMKM Menggunakan Metode Analytic Hierarchy Process Di Kabupaten Kutai Timur,” *Jurnal Teknologi Agro-Industri*, vol. 9, no. 2, Nov. 2022, doi: 10.34128/jtai.v9i2.162.
- [8] A. L. Mu’afifi, D. Alfareza, J. P. Natarendra, M. A. Pratama, M. R. Khafabih, and R. Vindua, “Analisis Klaster Pembiayaan UMKM dan Sektor Ekonomi Menggunakan Metode K-Means di Jawa Tengah,” *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 5469–5473, Jul. 2025, doi: 10.31004/riggs.v4i2.1457.
- [9] L. Zahrotun, A. A. E. Putra, U. Linarti, D. Normawati, and I. Arfiani, “Analisis UMKM Bidang Kerajinan dan Fashion di Kota Yogyakarta Menggunakan Metode K-Medoids,” *CESS (Journal of Computer Engineering, System and Science)*, vol. 10, no. 1, p. 206, Jan. 2025, doi: 10.24114/cess.v10i1.67270.
- [10] F. Tang, X. Zhao, and C. Li, “Community Detection in Multilayer Networks Based on Matrix Factorization and Spectral Embedding Method,” *Mathematics*, vol. 11, no. 7, p. 1573, Mar. 2023, doi: 10.3390/math11071573.