



Penerapan Algoritma *K-Means* untuk Klasterisasi Kabupaten/Kota Berdasarkan Karakteristik Ketenagakerjaan Menggunakan *Rapidminer* di Provinsi Lampung

Nindi Khoirunisa^{1*}, Hafin Nurhidayat², Hyang Lizant Berlyan Yanzant³, Anggayuh Prasetyo⁴, Nuari Anisa Sivi⁵

¹⁻⁵Sistem Informasi, Universitas Nahdlatul Ulama Lampung, Indonesia

Email: nindikhoirunnisa7@gmail.com¹, hafinganteng9137@gmail.com², hyanglizant03@gmail.com³, anggayuhp12@gmail.com⁴, nuarianisasivi@gmail.com⁵

*Penulis Korespondensi: nindikhoirunnisa7@gmail.com

Abstract. *Employment is an important indicator that reflects the economic condition and development level of a region. Variations in employment characteristics among regencies and cities in Lampung Province indicate the need for a grouping method to identify areas that share similar employment profiles. This research aims to classify regencies and cities in Lampung Province by applying the K-Means clustering algorithm. The study utilized secondary data obtained from the Central Statistics Agency (BPS) of Lampung Province, consisting of the number of registered job seekers, registered job vacancies, workforce placements, the ratio of vacancies to job seekers, the ratio of placements to job seekers, the Open Unemployment Rate (OUR), and the Labor Force Participation Rate (LFPR). The research stages included data collection, data preprocessing through Z-Transformation normalization, and cluster formation using the K-Means algorithm in RapidMiner. The results demonstrate that employment conditions vary across regencies and cities in Lampung Province, leading to the formation of several clusters with distinct characteristics. The resulting clusters provide an overview of regional employment patterns and may serve as supporting information for policymakers in formulating employment programs and strategies that are better aligned with local needs.*

Keywords: *Clustering; Data Mining; Employment; K-Means; Lampung Province; RapidMiner.*

Abstrak. Kondisi ketenagakerjaan merupakan salah satu indikator yang dapat digunakan untuk menggambarkan kondisi serta perkembangan perekonomian suatu wilayah. Adanya perbedaan karakteristik ketenagakerjaan pada setiap kabupaten/kota di Provinsi Lampung menunjukkan perlunya dilakukan pengelompokan wilayah guna mengetahui daerah-daerah yang memiliki kesamaan kondisi. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Lampung berdasarkan karakteristik ketenagakerjaannya dengan memanfaatkan algoritma K-Means. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Lampung, yang meliputi jumlah pencari kerja terdaftar, jumlah lowongan kerja terdaftar, jumlah tenaga kerja yang berhasil ditempatkan, rasio lowongan kerja terhadap pencari kerja, rasio penempatan tenaga kerja terhadap pencari kerja, Tingkat Pengangguran Terbuka (TPT), dan Tingkat Partisipasi Angkatan Kerja (TPAK). Proses penelitian diawali dengan pengumpulan data, dilanjutkan dengan normalisasi menggunakan metode Z-Transformation, kemudian dilakukan proses *cluster* ing menggunakan algoritma K-Means melalui perangkat lunak RapidMiner. Hasil penelitian menunjukkan adanya keragaman karakteristik ketenagakerjaan antar kabupaten/kota yang menghasilkan beberapa kelompok wilayah berdasarkan tingkat kemiripan data. Hasil pengelompokan tersebut diharapkan dapat memberikan gambaran kondisi ketenagakerjaan di Provinsi Lampung serta menjadi masukan bagi pemerintah daerah dalam merumuskan kebijakan dan program ketenagakerjaan yang lebih tepat sasaran.

Kata kunci: *Cluster ing; Data Mining; K-Means; Ketenagakerjaan; Provinsi Lampung; RapidMiner.*

1. LATAR BELAKANG

Ketenagakerjaan merupakan salah satu indikator yang dapat digunakan untuk menggambarkan kondisi pembangunan ekonomi suatu wilayah. Tingkat penyerapan tenaga kerja yang tinggi menunjukkan bahwa aktivitas ekonomi berjalan dengan baik dan mampu menyediakan kesempatan kerja bagi masyarakat (Khatib & Dalam, 2023). Sebaliknya,

keterbatasan lapangan pekerjaan dapat memicu peningkatan jumlah pengangguran yang pada akhirnya berdampak pada kesejahteraan masyarakat dan pertumbuhan ekonomi daerah (Kholifah et al., 2025). Oleh karena itu, informasi mengenai kondisi ketenagakerjaan menjadi penting untuk mendukung perumusan kebijakan pembangunan yang lebih tepat sasaran.

Keragaman kondisi ketenagakerjaan antar kabupaten/kota di Provinsi Lampung menunjukkan adanya perbedaan karakteristik pasar kerja pada masing-masing wilayah. Perbedaan tersebut dapat diukur melalui berbagai indikator, seperti jumlah pencari kerja terdaftar, jumlah lowongan kerja terdaftar, jumlah penempatan tenaga kerja, rasio lowongan terhadap pencari kerja, rasio penempatan terhadap pencari kerja, Tingkat Pengangguran Terbuka (TPT), dan Tingkat Partisipasi Angkatan Kerja (TPAK) (Nashrullah et al., 2025). Variasi pada indikator-indikator tersebut menunjukkan bahwa setiap daerah memiliki kondisi pasar kerja yang berbeda sehingga memerlukan pendekatan analisis yang mampu mengidentifikasi pola dan kesamaan karakteristik antarwilayah. Pengelompokan wilayah berdasarkan karakteristik ketenagakerjaan dapat membantu pemerintah dalam menentukan prioritas program dan kebijakan yang sesuai dengan kondisi masing-masing daerah (Aziz et al., 2023).

Kemajuan teknologi informasi telah membuka peluang untuk memanfaatkan teknik data mining dalam menggali informasi bernilai dari data yang tersimpan dalam jumlah besar (Lazuardi et al., 2025). *Cluster* ini merupakan salah satu teknik yang banyak digunakan dalam data mining untuk mengelompokkan data berdasarkan tingkat kemiripan antar objek. Melalui teknik ini, objek yang memiliki karakteristik yang hampir sama akan ditempatkan dalam kelompok yang sama, sedangkan objek dengan karakteristik yang berbeda akan berada pada kelompok yang berbeda (Terbuka et al., 2024). Dalam konteks ketenagakerjaan, *cluster* ing dapat digunakan untuk memetakan kondisi wilayah dan mengidentifikasi daerah yang memiliki karakteristik ketenagakerjaan yang sejenis (Kholifah et al., 2025).

Pemilihan algoritma K-Means dalam penelitian ini didasarkan pada kemampuannya untuk mengelompokkan data secara efektif melalui proses yang relatif sederhana. Selain mudah diimplementasikan, algoritma ini juga mampu memberikan hasil pengelompokan yang cukup baik pada berbagai jenis data. Metode ini bekerja dengan membagi data ke dalam sejumlah *cluster* berdasarkan kedekatan jarak terhadap pusat *cluster* (*centroid*) (Muningsih et al., 2021). Berbagai penelitian sebelumnya menunjukkan bahwa algoritma K-Means efektif digunakan untuk melakukan pengelompokan pada bidang ketenagakerjaan, seperti pengelompokan pencari kerja, indikator pengangguran, hingga pemetaan kondisi ketenagakerjaan antarwilayah (Arifin et al., 2025).

Penelitian terkait pengelompokan wilayah menggunakan indikator ketenagakerjaan sebagai dasar analisis telah banyak dilakukan di berbagai daerah di Indonesia. Namun, sebagian besar penelitian masih berfokus pada tingkat pengangguran, dampak ketenagakerjaan akibat pandemi, maupun karakteristik pencari kerja pada wilayah tertentu. Penelitian yang secara khusus mengombinasikan indikator pencari kerja terdaftar, lowongan kerja terdaftar, penempatan tenaga kerja, rasio lowongan terhadap pencari kerja, rasio penempatan terhadap pencari kerja, TPT, dan TPAK untuk mengelompokkan kabupaten/kota di Provinsi Lampung masih relatif terbatas. Kondisi ini menunjukkan adanya peluang penelitian untuk menghasilkan informasi yang lebih komprehensif mengenai karakteristik ketenagakerjaan di Provinsi Lampung (Nashrullah et al., 2025).

Dengan adanya perbedaan karakteristik ketenagakerjaan antar kabupaten/kota, penelitian ini menggunakan algoritma K-Means untuk membentuk kelompok wilayah yang memiliki tingkat kemiripan tertentu. Hasil pengelompokan diharapkan dapat memberikan informasi yang bermanfaat mengenai kondisi ketenagakerjaan di Provinsi Lampung dan mendukung penyusunan kebijakan yang lebih efektif dalam upaya peningkatan kualitas tenaga kerja.

2. KAJIAN TEORITIS

Ketenagakerjaan

Keadaan ketenagakerjaan sering digunakan sebagai indikator untuk menggambarkan tingkat aktivitas dan perkembangan ekonomi di suatu wilayah (Kholifah et al., 2025). Gambaran kondisi ketenagakerjaan suatu wilayah dapat ditinjau melalui berbagai indikator, seperti jumlah pencari kerja terdaftar, jumlah lowongan kerja terdaftar, jumlah penempatan tenaga kerja, rasio lowongan terhadap pencari kerja, rasio penempatan terhadap pencari kerja, Tingkat Pengangguran Terbuka (TPT), serta Tingkat Partisipasi Angkatan Kerja (TPAK). Menurut data yang dipublikasikan oleh Badan Pusat Statistik (BPS), berbagai indikator ketenagakerjaan dapat digunakan untuk menggambarkan tingkat penyerapan tenaga kerja serta kondisi pasar kerja pada suatu wilayah. (BPS Provinsi Lampung BPS-Statistics Lampung Province, 2025) Semakin baik kondisi ketenagakerjaan, maka semakin tinggi pula tingkat kesejahteraan masyarakat (Nashrullah et al., 2025).

Data Mining dan Clustering

Data mining merupakan proses pengolahan data dalam jumlah besar untuk menemukan pola, hubungan, atau informasi penting yang sebelumnya tidak terlihat secara langsung (Terbuka et al., 2024). Teknik ini banyak digunakan dalam berbagai bidang untuk membantu pengambilan keputusan berbasis data (Sudrajat et al., n.d.).

Salah satu metode dalam data *mining* adalah *clustering*. *Clustering* adalah teknik pengelompokan data tanpa label berdasarkan kemiripan karakteristik tertentu. Data yang memiliki kesamaan akan dikelompokkan dalam satu *cluster*, sedangkan data yang berbeda akan dipisahkan ke *cluster* lainnya. Dalam penelitian ini, *clustering* digunakan untuk mengelompokkan kabupaten/kota berdasarkan indikator ketenagakerjaan yang dimiliki (Lazuardi et al., 2025).

Algoritma K-Means dan RapidMiner

K-Means merupakan salah satu algoritma *clustering* yang paling umum digunakan karena memiliki konsep yang sederhana dan efisien dalam pengolahan data. Algoritma ini bekerja dengan cara membagi data ke dalam beberapa kelompok berdasarkan jarak terdekat terhadap titik pusat *cluster* (*centroid*) (Rumah et al., 2024). Proses perhitungan dilakukan secara berulang hingga posisi centroid tidak mengalami perubahan yang signifikan. Sebelum proses *clustering* dilakukan, data biasanya dinormalisasi terlebih dahulu agar memiliki skala yang sama. Salah satu metode yang sering digunakan adalah Z-Score normalization, karena dapat membantu mengurangi bias akibat perbedaan satuan antar variabel (Rahayu Iskandar & Kesumawati, 2024). Dalam penelitian ini, proses pengolahan data dilakukan menggunakan RapidMiner (Anggriani & Arif, 2024). Perangkat lunak ini digunakan untuk melakukan preprocessing data, penerapan algoritma K-Means, serta menampilkan hasil *cluster* ing dalam bentuk visual yang lebih mudah dipahami.

Penelitian Terdahulu

Penelitian oleh Joviandi & Kesumawati, (2024) menggunakan kombinasi Multidimensional *Scaling* dan K-Means *Clustering* untuk mengelompokkan lapangan usaha berdasarkan level keterampilan tenaga kerja. Hasil penelitian menunjukkan bahwa K-Means mampu membentuk beberapa *cluster* yang merepresentasikan karakteristik lapangan usaha secara berbeda-beda berdasarkan variabel yang dianalisis. (Rahayu Iskandar & Kesumawati, 2024)

Selanjutnya, penelitian oleh Kholifah et al. (2024) melakukan pengelompokan kabupaten/kota di Jawa Barat berdasarkan indikator ketenagakerjaan menggunakan K-Means dan Fuzzy C-Means. Hasil penelitian menunjukkan bahwa K-Means dapat digunakan untuk

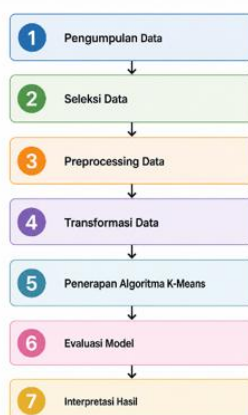
mengidentifikasi perbedaan kondisi ketenagakerjaan antar wilayah serta menghasilkan *cluster* yang cukup stabil dalam menggambarkan karakteristik daerah (Kholifah et al., 2025).

Penelitian lain oleh Saputra et al. (2023) menerapkan K-Means untuk mengelompokkan tingkat pengangguran terbuka di Indonesia berdasarkan data provinsi. Hasil penelitian menunjukkan bahwa K-Means mampu mengelompokkan wilayah menjadi beberapa kategori berdasarkan tingkat pengangguran (tinggi, sedang, dan rendah) sehingga dapat membantu dalam analisis kebijakan ketenagakerjaan (Arifin et al., 2025).

Selain itu, penelitian oleh Lazuardi et al. (2024) juga menunjukkan bahwa K-Means *Cluster* ing dapat digunakan untuk menganalisis karakteristik pencari kerja berdasarkan tingkat pendidikan dan jenis kelamin. Penelitian ini menegaskan bahwa teknik *clustering* sangat berguna dalam memahami pola ketenagakerjaan berbasis data statistik.(Lazuardi et al., 2025)

3. METODE PENELITIAN

Penelitian ini dilakukan untuk mengidentifikasi pola kesamaan karakteristik ketenagakerjaan pada kabupaten/kota di Provinsi Lampung melalui pendekatan *data mining*. Metode yang digunakan adalah *clustering* dengan algoritma K-Means, yang dipilih karena mampu mengelompokkan data berdasarkan tingkat kemiripan antar objek. Seluruh proses pengolahan data dilakukan menggunakan perangkat lunak RapidMiner, mulai dari tahap persiapan data hingga pembentukan *cluster*. Hasil pengelompokan yang diperoleh diharapkan dapat memberikan gambaran mengenai karakteristik ketenagakerjaan pada masing-masing wilayah sehingga perbedaan kondisi antar kabupaten/kota dapat diketahui secara lebih jelas (Syarif, 2023). Untuk memberikan gambaran yang lebih jelas mengenai tahapan penelitian yang dilakukan, alur penelitian disajikan dalam diagram berikut.:



Gambar 1. Tahapan Proses *Data Mining* Menggunakan Algoritma K-Means.

Pengumpulan Data

Tahap awal dalam penelitian ini adalah pengumpulan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Lampung dalam Angka 2025 (BPS Provinsi Lampung BPS-Statistics Lampung Province, 2025). Data yang digunakan mencakup beberapa indikator ketenagakerjaan, yaitu jumlah pencari kerja terdaftar, jumlah lowongan kerja terdaftar, jumlah penempatan tenaga kerja, rasio lowongan terhadap pencari kerja, rasio penempatan terhadap pencari kerja, Tingkat Pengangguran Terbuka (TPT), serta Tingkat Partisipasi Angkatan Kerja (TPAK) (Nashrullah et al., 2025). Seluruh data tersebut dikumpulkan sebagai dasar dalam proses analisis klasterisasi.

Seleksi Data

Pada tahap ini dilakukan proses pemilihan variabel yang dianggap relevan dengan tujuan penelitian (Data et al., 2025). Tidak semua data dari sumber digunakan, melainkan hanya variabel yang berkaitan langsung dengan kondisi ketenagakerjaan yang telah ditentukan sebelumnya. Data yang tidak memiliki keterkaitan dengan tujuan analisis diabaikan agar proses pengolahan data menjadi lebih terarah dan fokus.

Preprocessing Data

Sebelum data dianalisis lebih lanjut, dilakukan proses pembersihan data untuk memastikan kualitas data yang digunakan. Tahap ini mencakup pemeriksaan terhadap data yang kosong, tidak lengkap, atau tidak konsisten. Jika ditemukan ketidaksesuaian data, maka dilakukan penyesuaian atau penghapusan agar data yang digunakan benar-benar valid. Proses ini penting untuk menghindari kesalahan dalam hasil *clustering*.

Trnsformasi Data

Karena algoritma K-Means sangat dipengaruhi oleh skala data, maka seluruh variabel numerik terlebih dahulu dilakukan transformasi. Proses ini dilakukan dengan normalisasi atau standarisasi agar setiap variabel memiliki skala yang seimbang. Dengan demikian, tidak ada variabel yang mendominasi hasil perhitungan hanya karena perbedaan satuan atau rentang nilai.

Penerapan Algoritma K-Means

Setelah data siap, proses analisis dilakukan menggunakan algoritma K-Means melalui aplikasi RapidMiner. Pada tahap ini ditentukan jumlah *cluster* yang akan dibentuk, kemudian sistem akan mengelompokkan kabupaten/kota berdasarkan tingkat kemiripan data pada setiap variabel ketenagakerjaan. Hasil dari proses ini berupa pembentukan beberapa kelompok (*cluster*) beserta titik pusat (*centroid*) masing-masing *cluster* yang menggambarkan karakteristik tiap kelompok (Syarif, 2023).

Evaluasi Model

Hasil *clustering* kemudian dievaluasi untuk mengetahui kualitas pengelompokan yang dihasilkan. Evaluasi dilakukan menggunakan metode seperti *Davies Bouldin Index* (DBI) yang berfungsi untuk mengukur tingkat kepadatan dan pemisahan antar *cluster*. Selain itu, juga dilakukan pengamatan terhadap nilai jarak antar *cluster* serta jarak dalam *cluster* (*within cluster distance*) (Alfian & Hidayat, 2024). Semakin kecil nilai DBI, maka semakin baik kualitas *cluster* ing yang dihasilkan karena menunjukkan *cluster* yang lebih terpisah dan homogen.

Interpretasi Hasil

Tahap terakhir adalah memberikan interpretasi terhadap hasil *clustering* yang diperoleh. Pada tahap ini, setiap *cluster* dianalisis untuk mengetahui karakteristik masing-masing kelompok kabupaten/kota. Analisis dilakukan dengan melihat pola dari variabel ketenagakerjaan seperti tingkat pengangguran, penyerapan tenaga kerja, serta rasio lowongan kerja. Dari hasil tersebut kemudian ditarik kesimpulan mengenai pengelompokan wilayah di Provinsi Lampung, apakah termasuk kategori dengan kondisi ketenagakerjaan tinggi, sedang, atau rendah, serta bagaimana gambaran umum kondisi ketenagakerjaan di setiap kelompok wilayah.

4. HASIL DAN PEMBAHASAN

Hasil Pengolahan Data di RapidMiner

Data yang digunakan pada penelitian ini berasal dari sumber data sekunder yang diperoleh melalui publikasi *Provinsi Lampung Dalam Angka 2025* yang diterbitkan oleh Badan Pusat Statistik (BPS) Provinsi Lampung. Publikasi tersebut menyediakan berbagai informasi statistik yang digunakan sebagai dasar dalam analisis karakteristik ketenagakerjaan pada kabupaten/kota di Provinsi Lampung. Seluruh data kemudian diproses menggunakan aplikasi RapidMiner melalui metode K-Means *Clustering*. Tahapan pengolahan dimulai dari pembersihan data, transformasi, hingga proses pengelompokan dengan jumlah *cluster* sebanyak tiga kelompok.

Data

Tabel 1. Data dari BPS.

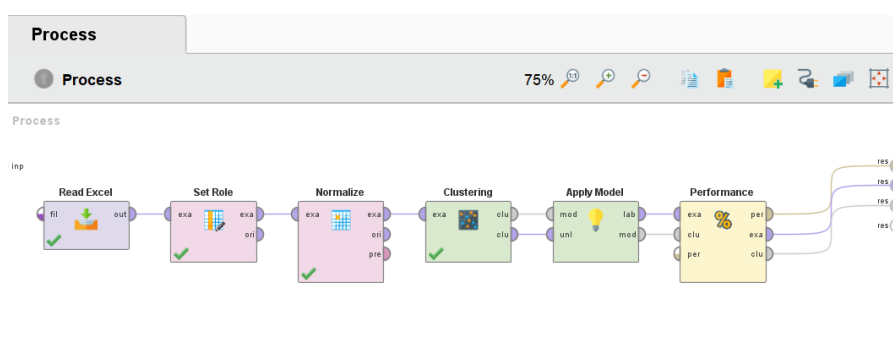
Kabupaten_ Kota	X1_Pencari Kerja	X2_Lowon gan	X3_Penemp atan	X4_Rasio_Low ongan	X5_Rasio_Penem patan	X6_T PT	X7_TP AK
Lampung Barat	111	20	12	18.02	10.81	2.09	82.83
Tanggamus	428	85	80	19.86	18.69	3.19	70.75
Lampung Selatan	770	362	181	47.01	23.51	4.84	67.34
Lampung Timur	2757	136	130	4.93	4.71	3.02	68.91
Lampung Tengah	1421	97	94	6.83	6.62	3.33	73.51
Lampung Utara	1025	142	136	13.85	13.27	5.85	66.55
Way Kanan	498	43	38	8.63	7.63	3.27	73.71
Tulang Bawang	360	149	145	41.39	40.28	3.23	71.54
Pesawaran	459	128	125	27.89	27.23	4.36	70.69
Pringsewu	676	55	44	8.14	6.51	4.39	73.55
Mesuji	285	40	32	14.04	11.23	2.85	70.20
Tulang Bawang Barat	602	61	59	10.13	9.80	4.12	70.90
Pesisir Barat	258	9	4	3.49	1.55	3.04	72.41
Bandar Lampung	2451	229	103	9.34	4.20	7.44	66.25
Metro	189	117	5	61.90	2.65	3.71	70.08

Tabel di atas menunjukkan data yang digunakan dalam penelitian yang diperoleh dari 15 kabupaten/kota di wilayah Provinsi Lampung dengan tujuh variabel ketenagakerjaan, yaitu jumlah pencari kerja, jumlah lowongan kerja, jumlah penempatan tenaga kerja, rasio lowongan terhadap pencari kerja, rasio penempatan terhadap pencari kerja, TPT, dan TPAK.

Secara umum, terlihat adanya perbedaan nilai antar wilayah pada setiap variabel. Variasi kondisi ketenagakerjaan yang terjadi pada masing-masing kabupaten/kota dimanfaatkan sebagai dasar dalam proses pengelompokan menggunakan algoritma K-Means, sehingga wilayah dengan karakteristik yang serupa dapat diidentifikasi dan ditempatkan dalam kelompok yang sama.

Sebelum dilakukan proses *clustering*, data terlebih dahulu melalui tahap normalisasi untuk menyeragamkan skala antar variabel. Langkah ini diperlukan agar setiap variabel dapat memberikan kontribusi yang proporsional dalam pembentukan *cluster* dan tidak menyebabkan bias pada hasil analisis.

Proses



Gambar 1. Proses Pengolahan Data Menggunakan RapidMiner.

Gambar tersebut menunjukkan rangkaian proses pengolahan data pada RapidMiner. Proses dimulai dari *input* data, kemudian dilanjutkan dengan pengaturan atribut, normalisasi, dan penerapan algoritma K-Means untuk pembentukan *cluster*. Setelah model dihasilkan, dilakukan penerapan model serta evaluasi menggunakan fitur *performance* untuk menilai kualitas hasil *clustering* yang diperoleh.

Hasil *Clustering* K-Means

Setelah proses *clustering* selesai dijalankan, RapidMiner menghasilkan pengelompokan kabupaten/kota ke dalam beberapa *cluster* berdasarkan tingkat kemiripan karakteristik ketenagakerjaan. Pengelompokan ini dilakukan dengan mempertimbangkan seluruh variabel yang digunakan dalam penelitian, sehingga wilayah yang memiliki karakteristik serupa akan berada pada kelompok yang sama. Hasil pembentukan *cluster* dapat dilihat pada gambar berikut.

Tabel 2. Hasil *Cluster* ing Menggunakan RapidMiner.

N o.	Kabupaten/ Kota	Clust er	X1_Pencari_Kerja	X2_Lowongan	X3_Penempatan	X4_Rasio_Lowongan	X5_Rasio_Penempatan	X6_TPT	X7_TPAK
3	Lampung Selatan	clust er_0	-0.062	2.762	1.803	1.570	1.022	0.688	-0.987
8	Tulang Bawang	clust er_0	-0.573	0.413	1.165	1.247	2.590	-0.510	0.062
9	Pesawaran	clust er_0	-0.450	0.182	0.811	0.471	1.370	0.331	-0.148
1	Lampung Barat	clust er_1	-0.884	-1.009	-1.190	-0.096	-0.165	-1.357	2.894

2	Tanggamus	<i>cluster_1</i>	-0.488	-0.293	0.014	0.009	0.571	-0.539	-0.133
5	Lampung Tengah	<i>cluster_1</i>	0.751	-0.160	0.262	-0.740	-0.557	-0.435	0.559
7	Way Kanan	<i>cluster_1</i>	-0.401	-0.756	-0.730	-0.636	-0.463	-0.480	0.609
10	Pringsewu	<i>cluster_1</i>	-0.179	-0.623	-0.623	-0.664	-0.568	0.353	0.569
11	Mesuji	<i>cluster_1</i>	-0.667	-0.789	-0.836	-0.325	-0.126	-0.792	-0.271
12	Tulang Bawang Barat	<i>cluster_1</i>	-0.271	-0.557	-0.358	-0.550	-0.260	0.152	-0.095
13	Pesisir Barat	<i>cluster_1</i>	-0.701	-1.131	-1.332	-0.932	-1.031	-0.651	0.283
15	Metro	<i>cluster_1</i>	-0.787	0.060	-1.314	2.427	-0.928	-0.153	-0.301
4	Lampung Timur	<i>cluster_2</i>	2.418	0.270	0.900	-0.849	-0.736	-0.666	-0.594
6	Lampung Utara	<i>cluster_2</i>	0.257	0.336	1.006	-0.336	0.065	1.439	-1.185
14	Bandar Lampung	<i>cluster_2</i>	2.036	1.295	0.422	-0.595	-0.784	2.621	-1.261

Hasil pengelompokan yang ditunjukkan pada gambar di atas menghasilkan tiga *cluster*, yaitu *cluster 0*, *cluster 1*, dan *cluster 2*. Setiap *cluster* terdiri atas sejumlah wilayah dengan karakteristik yang berbeda satu sama lain sesuai dengan pola data ketenagakerjaan yang dimiliki oleh setiap kabupaten/kota. Pengelompokan tersebut menjadi dasar untuk menganalisis kondisi ketenagakerjaan di Provinsi Lampung secara lebih terstruktur.

Cluster 0

Cluster 0 terdiri atas Kabupaten Lampung Selatan, Tulang Bawang, dan Pesawaran. Wilayah dalam kelompok ini memiliki nilai yang relatif tinggi pada variabel jumlah lowongan kerja, penempatan tenaga kerja, serta rasio lowongan dan rasio penempatan terhadap pencari kerja. Kondisi tersebut menunjukkan bahwa kesempatan kerja dan kemampuan penyerapan tenaga kerja di wilayah-wilayah ini cenderung lebih baik dibandingkan kelompok lainnya.

Dengan demikian, *cluster* 0 dapat menggambarkan daerah dengan kondisi ketenagakerjaan yang cukup baik.

Cluster 1

Cluster 1 terdiri atas Lampung Barat, Tanggamus, Lampung Tengah, Way Kanan, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, dan Metro.

Kelompok ini memiliki jumlah anggota paling banyak. Sebagian besar variabel ketenagakerjaan pada *cluster* ini berada pada kisaran nilai menengah sehingga mencerminkan kondisi yang relatif stabil. Meskipun demikian, masih terdapat beberapa indikator yang menunjukkan bahwa peningkatan kualitas ketenagakerjaan tetap diperlukan.

Cluster 2

Cluster 2 terdiri atas Lampung Timur, Lampung Utara, dan Bandar Lampung.

Karakteristik utama *cluster* ini adalah tingginya jumlah pencari kerja dibandingkan kelompok lain. Selain itu, beberapa wilayah dalam *cluster* ini juga memiliki tingkat pengangguran terbuka yang lebih tinggi. Hal tersebut menunjukkan bahwa jumlah tenaga kerja yang tersedia belum sepenuhnya seimbang dengan peluang kerja yang ada, sehingga *cluster* 2 dapat dianggap sebagai kelompok wilayah yang menghadapi tantangan ketenagakerjaan lebih besar.

Evaluasi Model (*Performance RapidMiner*)

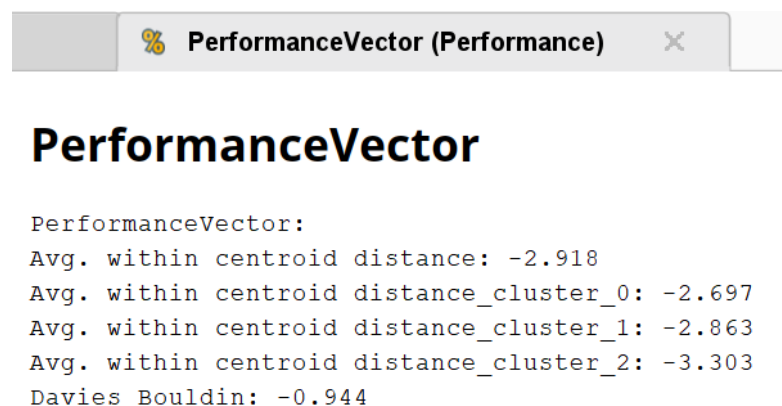
Setelah proses pengelompokan data menggunakan algoritma K-Means selesai dilakukan, langkah selanjutnya adalah mengevaluasi hasil *cluster* ing yang telah terbentuk. Evaluasi dilakukan melalui operator Performance pada RapidMiner untuk melihat kualitas *cluster* berdasarkan tingkat kedekatan data dalam satu kelompok dan perbedaan antar kelompok.

Berdasarkan hasil evaluasi, diperoleh nilai *Average Within Centroid Distance* sebesar -2,918. Nilai rata-rata jarak terhadap centroid pada masing-masing *cluster* yaitu -2,697 untuk *cluster* 0, -2,863 untuk *cluster* 1, dan -3,303 untuk *cluster* 2. Nilai tersebut menggambarkan seberapa dekat anggota *cluster* dengan pusat kelompoknya (*centroid*). Semakin dekat jarak antar data dengan *centroid*, maka semakin seragam karakteristik anggota dalam *cluster* tersebut.

Dari hasil yang diperoleh, *cluster* 2 memiliki nilai rata-rata jarak yang paling kecil dibandingkan *cluster* lainnya. Hal ini menunjukkan bahwa anggota pada *cluster* 2 memiliki tingkat kesamaan karakteristik yang relatif lebih tinggi sehingga membentuk kelompok yang lebih kompak.

Selain itu, hasil evaluasi juga menghasilkan nilai Davies Bouldin sebesar -0,944. Nilai ini digunakan untuk memberikan gambaran mengenai kualitas pengelompokan yang dihasilkan. Semakin rendah nilai Davies Bouldin, maka pemisahan antar *cluster* cenderung semakin baik dan karakteristik setiap *cluster* menjadi lebih mudah dibedakan.

Secara umum, hasil evaluasi menunjukkan bahwa algoritma K-Means berhasil mengelompokkan kabupaten/kota yang memiliki karakteristik ketenagakerjaan yang cenderung serupa dalam satu *cluster*, serta memperlihatkan perbedaan yang cukup jelas jika dibandingkan dengan *cluster* lainnya. Dengan demikian, hasil *cluster* ing yang diperoleh dapat digunakan untuk menggambarkan pola ketenagakerjaan di Provinsi Lampung berdasarkan variabel yang dianalisis.



Gambar 2. Hasil Evaluasi Clustering Menggunakan PerformanceVector pada RapidMiner.

5. KESIMPULAN DAN SARAN

Penelitian ini dilakukan untuk mengelompokkan kabupaten/kota di Provinsi Lampung berdasarkan karakteristik ketenagakerjaan dengan menerapkan algoritma K-Means. Berdasarkan hasil analisis data dari *Provinsi Lampung Dalam Angka 2025*, diperoleh tiga *cluster* yang menunjukkan perbedaan karakteristik ketenagakerjaan antarwilayah. Hasil pengelompokan memperlihatkan adanya wilayah dengan tingkat penyerapan tenaga kerja yang relatif baik, wilayah dengan kondisi yang cukup stabil, serta wilayah yang masih menghadapi tantangan dalam penyediaan kesempatan kerja. Temuan ini menunjukkan bahwa algoritma K-Means dapat digunakan untuk mengidentifikasi kemiripan karakteristik ketenagakerjaan sehingga memberikan gambaran yang lebih jelas mengenai kondisi ketenagakerjaan di Provinsi Lampung. Hasil evaluasi model yang ditunjukkan oleh nilai *Average Within Centroid Distance* sebesar -2,918 dan *Davies Bouldin* sebesar -0,944 mengindikasikan bahwa proses *cluster* ing mampu membentuk kelompok wilayah dengan karakteristik yang relatif serupa.

Hasil penelitian ini diharapkan dapat memberikan kontribusi sebagai bahan pertimbangan bagi pemerintah daerah serta instansi terkait dalam merumuskan kebijakan ketenagakerjaan yang lebih sesuai dengan kondisi masing-masing wilayah. Untuk penelitian selanjutnya, disarankan agar menggunakan data dengan rentang waktu yang lebih panjang serta menambahkan variabel lain yang berkaitan dengan aspek sosial dan ekonomi sehingga hasil pengelompokan menjadi lebih komprehensif. Selain itu, penerapan metode *cluster* ing yang berbeda sebagai pembanding juga dapat dilakukan untuk memperoleh hasil analisis yang lebih beragam.

DAFTAR REFERENSI

- Alfian, W., & Hidayat, T. (2024). Analisis *cluster* ing pegawai berdasarkan tingkat kedisiplinan menggunakan algoritma K-Means dan Davies-Bouldin Index. *JEECOM*, 6(2), 437–448. <https://doi.org/10.33650/jeecom.v4i2>
- Anggriani, Y. P., & Arif, A. (2024). Implementasi algoritma K-Means *cluster* ing dalam menentukan blok tanaman sawit produktif pada PT Arta Prigel. *Komtika*, 8(1), 22–32. <https://doi.org/10.31603/komtika.v8i1.11192>
- Arifin, S., Sa, Y., Firdaus, A. A., Afriansyah, M., & Saputra, J. (2025). *Cluster* ing daftar pemilih tetap (DPT) di Desa Kerembong pada Pilkada 2024 menggunakan algoritma K-Means dengan RapidMiner. 8(November).
- Aziz, A., Norhasanah, S., Ali, U. D., Hulu, M. B., Timur, K. K., Tengah, K., Informasi, J. S., Darwan, U., & Sampit, A. (2023). Klastrisasi data tenaga kerja terbuka menurut provinsi dengan penggunaan algoritma K-Means. *10*(3), 444–456.
- BPS Provinsi Lampung. (2025). *Provinsi Lampung dalam angka 2025*. BPS Provinsi Lampung.
- Data, P., Kerja, S., & Nasional, A. (2025). Analisis K-Means *cluster* ing angkatan kerja berdasarkan jenis kelamin, usia, pendidikan, dan pendapatan pada data satuan kerja angkatan nasional tahun 2021. 2(1), 942–951.
- Khatib, J., & Dalam, S. (2023). *Indonesian Journal of Computer Science*, 12(1), 714–728.
- Kholifah, R. I. N., Sarah, A. D., Masichah, F., & Haris, M. A. (2025). Pengelompokan kabupaten/kota di Jawa Barat berdasarkan indikator ketenagakerjaan menggunakan metode K-Means dan fuzzy C-Means dengan evaluasi rasio SW/SB sebagai validasi klaster. *Jurnal Statistika dan Komputasi*, 4(2), 68–82. <https://doi.org/10.32665/statkom.v4i2.5196>
- Lazuardi, S. I., Putra, J. L., & Setiyorini, T. (2025). Klasterisasi data mining pencari kerja DKI Jakarta menggunakan metode K-Means *cluster* ing. *Detika*, 6(1), 1–5. <https://doi.org/10.24176/detika.v6i1.15471>
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan metode K-Means dan optimasi jumlah *cluster* dengan indeks Davies-Bouldin untuk *cluster* ing provinsi berdasarkan potensi desa. *Jurnal Sains dan Manajemen*, 9(1), 96.

- Nashrullah, A. F., Mahardhika, R. D., Rusdiyanto, N. R., & Shella, S. (2025). Klasterisasi produktivitas daerah di Jawa Tengah berdasarkan ketenagakerjaan menggunakan K-Means dan average linkage. 7(2), 138–149. <https://doi.org/10.35508/jd.v7i2.22516>
- Rahayu Iskandar, A. N. J., & Kesumawati, A. (2024). Pengelompokan lapangan usaha berdasarkan level skill menggunakan multidimensional scaling dan K-Means *cluster ing*. *Emerging Statistics and Data Science Journal*, 2(1), 67–76. <https://doi.org/10.20885/esds.vol2.iss.1.art7>
- Rumah, P., Royal, S., Hidayah, A., Dulisep, D., & Angga, B. (2024). Implementasi algoritma K-Means menggunakan RapidMiner untuk klasterisasi data obat. 7(2), 200–211.
- Sudrajat, W., Cholid, I., Petrus, J., Studi, P., Informatika, M., Multi, U., Palembang, D., Manajemen, P. S., Multi, U., Palembang, D., Informatika, P. S., Multi, U., Palembang, D., & Phase, M. (n.d.). Penerapan algoritma K-Means *cluster ing* untuk pengelompokan UMKM menggunakan RapidMiner, 27–36.
- Syarif, Z. N. (2023). Penerapan information gain dan algoritma K-Means untuk klasterisasi kedisiplinan pegawai menggunakan RapidMiner. *TeknoIS: Jurnal Ilmiah Teknologi Informasi dan Sains*, 13(1), 1–12. <https://doi.org/10.36350/jbs.v13i1.165>
- Terbuka, P., Menurut, T. P. T., & Di, P. (2024). Algoritma K-Means untuk mengelompokkan tingkat. 15(2), 75–81. <https://doi.org/10.36982/jiig.v15i2.4359>