Jurnal Responsif: Riset Sains & Responsif
Vol.7 No 1 Februari 2025, Halaman. 19~25
E-ISSN**:** 2685-6964

ANALISIS LOWONGAN PEKERJAAN DI BIDANG TEKNOLOGI INFORMASI BERDASARKAN LOKASI MENGGUNAKAN TEKNIK KLASTERING K – MEANS

Hendy D. Siswaja[1], Rizki T. Prasetio[2]

[1] SATU University
e-mail: hendy.siswaja@univ.satu.ac.id

[2] SATU University
e-mail: rizki.prasetio@univ.satu.ac.id

Abstract

This study analyzes job vacancy distribution in the Technology Information sector based on location using the K-Means clustering technique. In the current era of Industry 4.0, job seekers and employers face challenges with data overload and time-consuming recruitment processes. Identifying hidden patterns in job vacancy data is crucial to understanding the concentration of job opportunities across regions. The methodology involves data preprocessing, including data cleaning, transformation, and normalization, followed by clustering using K-Means and evaluation with the Davies-Bouldin Index (DBI) to determine optimal clustering. The results reveal that clustering with 6 groups provides the most meaningful separation of job vacancies based on location and category, with a significant dominance of one cluster. The findings suggest that more detailed analysis of smaller clusters could uncover niche opportunities. This approach can assist policymakers and job seekers in making more informed decisions regarding career opportunities
Keywords*: K-Means, job vacancies, clustering, Davies-Bouldin Index, location distribution*

Abstrak

Penelitian ini menganalisis distribusi lowongan pekerjaan di Teknologi Informasi berdasarkan lokasi menggunakan teknik klastering K-Means. Di era Industri 4.0 saat ini, pencari kerja dan pemberi kerja menghadapi tantangan dalam mengelola data yang berlebihan dan proses rekrutmen yang memakan waktu. Mengidentifikasi pola tersembunyi dalam data lowongan pekerjaan sangat penting untuk memahami konsentrasi peluang kerja di berbagai wilayah. Metodologi yang digunakan mencakup prapemrosesan data, termasuk pembersihan data, transformasi, dan normalisasi, diikuti dengan klastering menggunakan K-Means serta evaluasi dengan Davies-Bouldin Index (DBI) untuk menentukan jumlah klaster optimal. Hasil penelitian menunjukkan bahwa klastering dengan 6 kelompok memberikan pemisahan data yang paling berarti berdasarkan kategori pekerjaan dan lokasi, dengan satu klaster yang dominan. Temuan ini menunjukkan bahwa analisis lebih lanjut pada klaster-klaster kecil dapat mengungkap peluang pekerjaan yang lebih spesifik. Pendekatan ini dapat membantu pembuat kebijakan dan pencari kerja dalam membuat keputusan yang lebih tepat terkait peluang karir.
Kata Kunci : *K-Means, lowongan pekerjaan, klastering, Davies-Bouldin Index, distribusi lokasi*

1. Pendahuluan

Perkembangan teknologi informasi (TI) yang pesat telah menciptakan banyak peluang kerja di berbagai lokasi, baik di tingkat nasional maupun global. Permintaan tenaga kerja di bidang ini terus meningkat seiring dengan transformasi digital yang dilakukan oleh perusahaan di berbagai sektor. Namun, distribusi lowongan pekerjaan tidak merata di setiap daerah. Beberapa wilayah memiliki konsentrasi peluang kerja yang lebih tinggi dibandingkan yang lain, bergantung pada faktor seperti perkembangan industri, ketersediaan tenaga kerja terampil, serta kebijakan pemerintah terkait teknologi dan investasi.

Analisis lowongan pekerjaan di bidang teknologi informasi berdasarkan

Diterima February 19, 2025; Revisi February 25 , 2025; Disetujui February 28, 2025

lokasi menjadi penting untuk memahami bagaimana peluang kerja tersebar di berbagai daerah. Dengan meneliti faktor-faktor yang mempengaruhi distribusi ini, pencari kerja dapat mengidentifikasi lokasi dengan prospek karier terbaik, sementara perusahaan dapat merancang strategi rekrutmen yang lebih efektif.

Seiring dengan pesatnya perkembangan teknologi informasi, pola pencarian pekerjaan juga mengalami pergeseran menjadi pencarian pekerjaan melalui internet. Pencarian pekerjaan melalui internet berarti memanfaatkan internet untuk menemukan peluang kerja dan kandidat yang sesuai. Proses ini dilakukan baik oleh pencari kerja maupun pemberi kerja. Saat ini, banyak pengguna internet, terutama remaja, menggunakan internet untuk mencari pekerjaan yang mereka inginkan. Bagi pencari kerja, metode ini lebih praktis dibandingkan dengan cara tradisional seperti mencari informasi melalui koran, selebaran, atau iklan cetak. Oleh karena itu, pencarian kerja secara *online* lebih cepat dan menawarkan lebih banyak pilihan. Selain itu, pencarian pekerjaan melalui internet juga memberikan wawasan empiris tentang hubungan antara kualitas dan jumlah informasi yang tersedia (Samah dkk., 2022).

Saat ini pemberi kerja dan pencari kerja dihadapkan pada masalah *overload* data dan proses rekrutmen yang memakan waktu. profil kandidat yang beragam menyulitkan perekrut menemukan kompetensi yang tepat. Sistem rekomendasi pekerjaan berbasis pembelajaran mesin dapat membantu mencocokkan kandidat dengan pekerjaan yang relevan, berdasarkan perilaku pencari kerja dan kebutuhan pemberi kerja. Dalam pasar kerja yang berkembang, khususnya di bidang teknologi dan informasi, lowongan pekerjaan sangat dipengaruhi oleh faktor seperti lokasi, industri, dan tingkat pendidikan. Mengidentifikasi pola tersembunyi dalam data lowongan pekerjaan menjadi penting untuk memahami konsentrasi peluang kerja di berbagai wilayah. Oleh karena itu, analisis lowongan pekerjaan yang lebih mendalam dapat membantu pengambil kebijakan, pencari kerja, dan perusahaan dalam merancang strategi yang lebih efektif (Mhamdi dkk., 2020).

Pencarian pekerjaan melalui media *online* menawarkan metode yang praktis namun kemudahan akses informasi lowongan kerja yang tersedia secara *online* tidak menjamin kemudahan seorang pencari kerja mendapatkan pekerjaan. Terkadang hal tersebut membuat pencari kerja semakin bingung karena banyaknya pilihan jenis pekerjaan (Agustyani & Santoso, 2019).

Data mining adalah proses mencari pola tersembunyi dalam suatu kumpulan data yang terdapat dalam gudang data atau basis data, berdasarkan pengetahuan yang sebelumnya tidak diketahui. Teknik ini memungkinkan berbagai kemampuan, termasuk menjelaskan bagaimana analisis dapat mengidentifikasi pola serta memahami proses yang terjadi dalam data (Ramdhani & Alamsyah, 2023).

Clustering merupakan salah satu teknik dalam data mining yang digunakan untuk menyelesaikan berbagai masalah kompleks di bidang ilmu komputer dan statistik. Metode ini berfungsi untuk mengelompokkan sekumpulan data ke dalam dua atau lebih kelompok, di mana data dalam satu kelompok memiliki karakteristik yang lebih mirip dibandingkan dengan data di kelompok lain. Pengelompokan ini dilakukan secara otomatis berdasarkan informasi yang tersedia dalam data tanpa memerlukan klasifikasi awal. Salah satu metode *clustering* yang populer adalah K-Means, yaitu teknik pengelompokan data non-hierarkis yang dapat membagi data ke dalam dua atau lebih kelompok. Metode ini bekerja dengan mengelompokkan data yang memiliki karakteristik serupa ke dalam satu kelompok yang sama, sementara data dengan karakteristik berbeda akan dimasukkan ke dalam kelompok lain. Tujuan utama dari proses pengelompokan ini adalah untuk meminimalkan fungsi objektif yang telah ditetapkan, yaitu dengan mengurangi variasi dalam kelompok (*intra-cluster*) dan meningkatkan variasi antar kelompok (*inter-cluster*) (Aldino dkk., 2021).

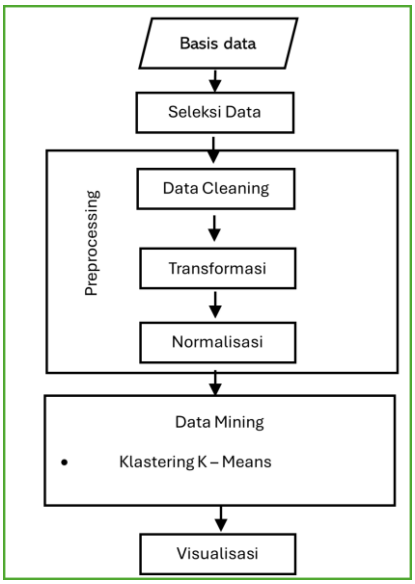
Indeks Davies-Bouldin (DBI) adalah salah satu metode yang digunakan untuk mengukur validitas hasil *clustering* dalam suatu metode pengelompokan. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak antar-klaster (*inter-cluster*) sekaligus meminimalkan jarak antara titik-titik dalam satu klaster (*intra-cluster*). Jika jarak antar-klaster maksimal, maka kesamaan karakteristik dalam setiap klaster cenderung kecil, sehingga perbedaan antara klaster menjadi lebih jelas. Sebaliknya, jika jarak dalam klaster minimal, berarti setiap objek dalam klaster memiliki tingkat

kesamaan karakteristik yang tinggi (Mughnyanti dkk., 2020).

Dalam penelitian ini, metode klastering K-Means dengan evaluasi Davies-Bouldin Index dipergunakan untuk mengidentifikasi pola distribusi lowongan pekerjaan di sektor tersebut, serta memahami faktor-faktor yang memengaruhi lokasi pekerjaan yang ditawarkan. Dengan pendekatan ini, diharapkan dapat memberikan wawasan mengenai peluang karir yang tersebar di berbagai wilayah, serta membantu pencari kerja untuk menyesuaikan pencarian lowongan berdasarkan preferensi lokasi yang lebih relevan.

2. Metode Penelitian

Penelitian ini dimulai dengan mengidentifikasi masalah yang dilakukan dengan melakukan pengamatan terhadap data set terbuka dari Kaggle, lalu melakukan pengolahan data dan pembahasan data. Pengolahan data mengunakan konsep *preprocessing*. Tahapan *preprocessing* adalah mengubah data ke suatu format sehingga *data mining* lebih mudah dan efektif sesuai kebutuhan, dalam penelitian ini atribut *category* dan *description* diubah dari *text* menjadi *numeric*. Tahapan *preprocessing* juga dapat membantu dalam mendapatkan hasil yang lebih akurat, dan mengurangi waktu dalam penghitungan terutama dalam skala besar, dan dapat membuat nilai data menjadi lebih kecil tanpa mengubah informasi di dalamnya (Nugroho dkk., 2022).



Gambar 1. Tahapan Metode Penelitian
Sumber: (Sopyan dkk., 2022)

2.1. Seleksi Data

Penelitian ini menggunakan *dataset* yang diperoleh dari *platform* Kaggle yang menyediakan data terbuka dalam format CSV. Rentang data lowongan kerja yang diperoleh dari Januari 2024 sampai September 2024. *Dataset* ini terdiri dari 9.919 entri data dengan 21 atribut yang mencakup berbagai informasi terkait lowongan pekerjaan, seperti *Website Domain, Ticker, Job Opening Title, Job Opening URL, First Seen At. Last Seen At, Location, Location Data, Category, Seniority. Keywords, Description, Salary, Salary Data, Contract Types, Job Status, Job Language, Job Last Processed At, O*NET Code, O*NET Family, O*NET Occupation Name*. Data ini merupakan data mentah yang belum melalui proses pembersihan (*data cleaning*) yang kemudian akan dilakukan seleksi fitur data dengan memilih kolom-kolom atribut yang diperlukan.

2.2. Data Cleaning

Pada tahap ini data yang akan digunakan ternyata terdapat kendala atau masalah ketika akan dilakukan klastering, seperti data yang hilang (*missing values*), inkonsistensi format (misalnya format tanggal atau tipe data yang tidak konsisten), serta duplikasi entri, sebelum digunakan untuk analisis lebih lanjut, data-data tersebut kemudian dihapus dalam *dataset* Penelitian ini memilih 3 atribut utama yaitu *Location, Category, O*NET Family* karena dinilai mewakili apa yang akan diteliti yaitu lokasi dan jenis lowongan pekerjaan. Setelah melakukan pembersihan data, data yang tersisa menjadi 7.925.

2.3. Transformasi Data

Pada atribut*, Category* dan *O*NET Family* perlu dilakukan transformasi data dari *text* ke *numeric*, agar memberikan nilai yang valid terhadap variabel pada saat melakukan proses klastering. Penelitian ini menggunakan *text to nominal* di Rapidminer*.* Untuk dapat mengubah *text* menjadi *numeric.*

2.4. Normalisasi Data

Dalam penelitian ini teknik *processing text to nominal* di Rapidminer*,* pada atribut *Category* dan *O*NET Family (Description)* untuk menghasilkan nilai valid 1 pada kategori yang sesuai dan nilai 0 pada kategori yang tidak sesuai, sehingga saat dilakukannya proses klastering dapat disimpulkan nilai klasteringnya.

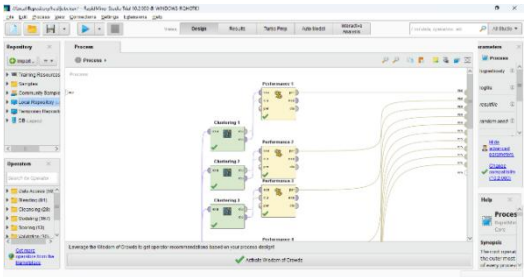
2.5. Data Mining

Merupakan proses untuk menggali pola atau informasi tersembunyi dalam kumpulan data yang besar. Salah satu teknik yang sering digunakan dalam data mining adalah klastering, yaitu metode yang bertujuan untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kemiripan tertentu. Klastering tidak memerlukan label atau kategori sebelumnya dan membantu menemukan struktur dalam data yang tidak terlihat secara langsung. Dalam konteks ini, K-Means adalah salah satu algoritma klastering yang populer, yang mengelompokkan data berdasarkan kedekatannya dengan titik pusat klaster (*centroid*).

Untuk menentukan jumlah klaster yang optimal dalam K-Means, salah satu metode yang digunakan adalah *Davies-Bouldin Index* (DBI). DBI merupakan sebuah metrik yang digunakan untuk menilai kualitas klaster dalam analisis data. Tujuan utama dalam penggunaan teknik ini adalah untuk mengukur sejauh mana algoritma klastering dapat memisahkan kelompok data yang berbeda serta mendekatkan titik data ke pusat klasternya. Nilai BDI yang semakin rendah menunjukkan bahwa klastering yang dihasilkan semakin baik (Umagapi dkk., 2023). Dengan menggunakan metode DBI, kita dapat mengevaluasi berbagai jumlah klaster yang dihasilkan oleh K-Means dan memilih jumlah klaster yang optimal untuk menghasilkan pengelompokan data yang lebih bermakna dan efektif.

2.6. Visualisasi

Pola informasi yang dihasilkan dari tahap sebelumnya kemudian ditampilkan dalam bentuk tabel dan grafik agar lebih mudah dipahami. Visualisasi ini dapat digunakan untuk mengetahui lowongan pekerjaan seperti apa dan lokasi pekerjaan yang terdapat dia suatu daerah tertentu
Klastering dilakukan bedasarkan kategori dan jenis pekerjaan terhadap lokasi lowongan kerja itu berada. Penelitian ini melakukan 10 kali pengujian secara bersamaan menggunakan Rapidminer dengan menggunakan *cluster distance perform Davies Bouldin*. Berikut gambar pengujian melalui *rapid miner.*



Gambar 2. Proses Klastering Pada *Rapid Miner*

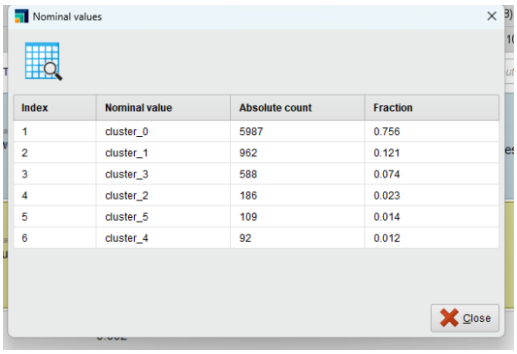
Dari proses didapat hasil data sebagai berikut :

K- Means	Performance Vector									
	1	2	3	4	5	6	7	8	9	10
Average Centroid	-1,881	-1,891	-1,799	-1,750	-1,732	-1,725	-1,644	-1,628	-1,650	-1,58
0	-1,956	-1,955	-1,967	-1,966	-1,967	-1,965	-1,969	-1,967	-1,970	-1,974
1	-0,955	-0,976	-0,806	-0,744	-0,869	-0,963	-0,955	-0,976	-0,843	-0,473
2			-0,962	-1,030	-0,963	-1,373	-0,976	-1,200	-0,955	-0,963
3				-0,963	-0,955	-0,955	-1,175	-0,901	-0,925	-0,521
4					-0,708	-0,931	-0,965	-0,925	-0,976	-1,045
5						-0,678	-0,783	-0,963	-0,808	-0,976
6							-0,881	-1,030	-0,963	-0,955
7								-1,046		-0,915
8										-0,887
9										-0,773
10										
DBI	-2,288	-2,317	-2,237	-2,279	-2,207	-2,491	-2,365	-2,435	-2,253	-2,157

Gambar 3. Hasil uji *K – Means*

Dari hasil uji diatas maka dapat disimpulkan bahwa nilai *centroid* klaster terlihat bervariasi dan fluktuatif, terutama pada iterasi-iterasi awal, dengan beberapa klaster menunjukkan nilai lebih tinggi dibandingkan yang lainnya. Seiring bertambahnya jumlah *cluster*, nilai *centroid* mulai lebih stabil dan cenderung semakin negatif. Hal ini mengindikasikan bahwa posisi pusat klaster semakin terkonsolidasi, meskipun ada beberapa klaster yang mengalami perubahan signifikan pada beberapa iterasi. Fluktuasi nilai *centroid* tersebut menunjukkan bahwa algoritma *K-Means* masih berusaha menemukan pemisahan terbaik antar data, terutama pada jumlah *cluster* yang lebih sedikit. Analisis menggunakan *Davies-Bouldin Index* (DBI) juga memberikan wawasan terkait kualitas klastering yang dihasilkan. DBI yang lebih rendah menunjukkan klastering yang lebih baik, yaitu pemisahan antar klaster yang lebih jelas dan *compact*. Nilai DBI yang tercatat berkisar antara -2,491 hingga -2,157, dengan nilai terendah tercatat pada iterasi ke-6, yaitu -2,491, yang menunjukkan bahwa pemisahan antar klaster pada iterasi tersebut relatif lebih baik. Namun, DBI yang cenderung sedikit meningkat pada iterasi-iterasi berikutnya menunjukkan adanya penurunan kualitas klastering, meskipun masih berada dalam kisaran yang dapat

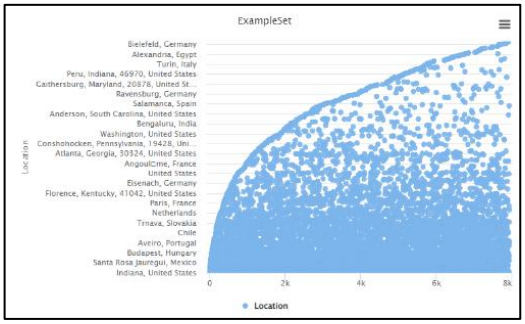
diterima. Oleh karena itu, dapat disimpulkan bahwa pemisahan klaster yang optimal terjadi sekitar jumlah klaster 6. Dari hasil yang diperoleh, terlihat bahwa jumlah klaster yang lebih sedikit cenderung menghasilkan DBI yang lebih tinggi, menandakan pemisahan klaster yang kurang optimal. Sebaliknya, pada jumlah klaster yang lebih banyak, meskipun DBI tidak menunjukkan penurunan yang signifikan, kualitas klastering cenderung lebih baik. Oleh karena itu, berdasarkan analisis DBI dan posisi *centroid*, jumlah klaster yang ideal kemungkinan berada pada angka 6, di mana pemisahan antar klaster tampak lebih jelas dan terstruktur.



Index	Nominal value	Absolute count	Fraction
1	cluster_0	5987	0.756
2	cluster_1	962	0.121
3	cluster_3	588	0.074
4	cluster_2	186	0.023
5	cluster_5	109	0.014
6	cluster_4	92	0.012

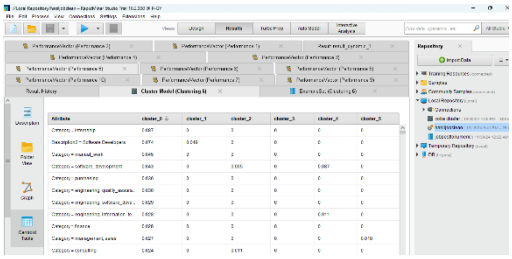
Gambar 4. *Nominal Values* K-Means 6

Dalam hasil *Cluster Model K – Means 6* distribusi data di antara cluster ini tidak merata. *Cluster* 0 mendominasi dengan lebih dari 75% data (5,987 dari 7,924), sementara *cluster* lainnya memiliki jumlah item yang jauh lebih kecil, Klaster yang kecil (seperti klaster 4 dan 5) ini mengindikasikan adanya kelompok data yang sangat berbeda (*outliers* atau kelompok minoritas) atau klastering yang kurang optimal.



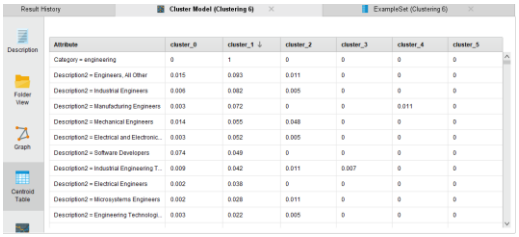
Gambar 5. Visualisasi Hasil Klastering

Pada gambar grafik diatas lowongan pekerjaan banyak terdapat di area Indiana, United State dan pekerjaan mendominasi disetiap klaster.



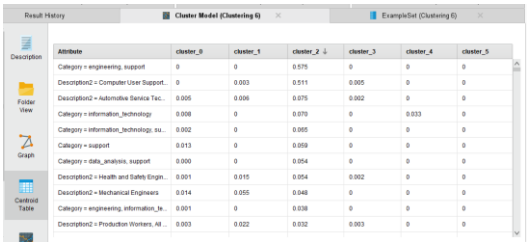
Gambar 6. Nilai Centroid Klaster 0

Pada Gambar 6 pengujian di klaster 0 pekerjaan yang tersedia di kategori *internship* dengan nilai 0.087 (nol koma nol delapan tujuh).



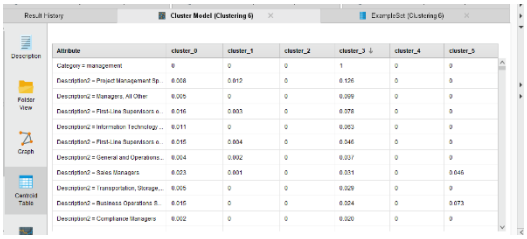
Gambar 7. Nilai Centroid Klaster 1

Di klaster 1 kategori yang dominan pekerjaan di bidang *engineering* dengan nilai centroid 1 (satu).



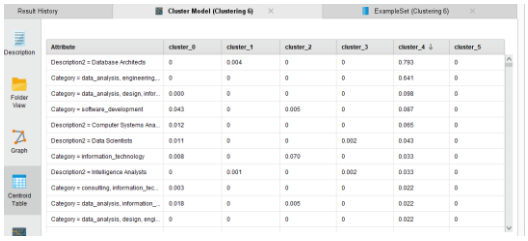
Gambar 8. Nilai Centroid Klaster 2

Untuk di klaster 2 dominasi pekerjaan yang dibutuhkan di bagian *Support Engineering* dengan nilai centroid 0.575 (nol koma lima tujuh lima).



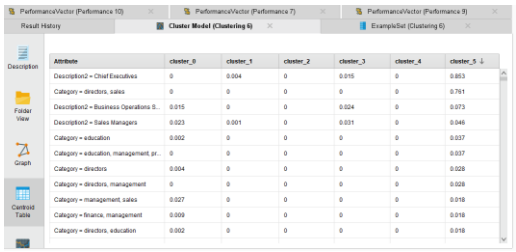
Gambar 9. Nilai Centroid Klaster 3

Pada klaster 3 bidang pekerjaan di bidang *Management* dengan nilai centroid 1 (satu).



Gambar 10. Nilai Centroid Klaster 4

Pada klaster 4 pekerjaan sebagai *Database Architects* dengan nilai centroid 0,793 (nol koma tujuh sembilan tiga).



Gambar 11. Nilai Centroid Klaster 5

Sedangkan pada di klaster 5 adalah *Chief Executive* dengan nilai centroid 0.853 (nol koma delapan lima tiga)

4. Kesimpulan

Dari hasil klastering dengan K – means 6 dapat disimpulkan bahwa lowongan pekerjaan di bidang teknologi. Sebagian besar data cenderung terpusat dalam satu *cluster* besar, sementara beberapa *cluster* kecil menunjukkan adanya data yang lebih jarang atau lebih spesifik. Hal ini dapat menunjukkan bahwa dalam sektor pekerjaan yang dianalisis, sebagian besar lowongan pekerjaan mungkin terfokus di lokasi tertentu atau kategori yang lebih umum, menandakan bahwa distribusi lowongan pekerjaan tidak merata, namun klaster terpisah dengan baik karena jarak antar klaster cukup besar karena klastering berhasil memisahkan data dengan baik berdasarkan kategori pekerjaan dan lokasi.

Untuk menyeimbangkan distribusi data antara klaster besar dan kecil, perlu dilakukan beberapa penyesuaian misalnya dengan menggabungkan kategori pekerjaan atau lokasi yang sangat dominan, atau memperhatikan klaster kecil yang mungkin berisi data minoritas yang penting. Klaster kecil seperti *cluster* 4 dan 5 dapat menjadi area yang menarik untuk analisis lebih lanjut. Mungkin klaster ini mewakili jenis pekerjaan atau lokasi yang sedang berkembang atau jarang ditemukan, yang dapat membuka peluang baru dalam pasar kerja. Penelitian ini juga membuka peluang penggunaan atribut lainnya yang ada pada *dataset* untuk menganalisis jauh lebih dalam seperti pendidikan dan pengalaman kerja,.

Referensi

Agustyani, E. M., & Santoso, I. (2019). Analisis Lowongan Pekerjaan Studi Kasus: Portal Lowongan Kerja Jobstreet. *Seminar Nasional Official Statistics*, 1–10.

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency. *Journal of Physics: Conference Series*, *1751*(1), 012038.

Mhamdi, D., Moulouki, R., El Ghoumari, M. Y., Azzouazi, M., & Moussaid, L. (2020). Job recommendation based on job profile clustering and job seeker behavior. *Procedia Computer Science*, *175*, 695–699.

Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. *IOP Conference Series: Materials Science and Engineering*, *725*(1), 012128.

Nugroho, J. R., Suprapto, Y. K., & Setijadi, E. (2022). Clustering Tingkat Risiko Klasifikasi Lapangan Usaha (KLU) Menggunakan Metode K-Means. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *9*(3), 533–540.

Ramdhani, Y., & Alamsyah, D. P. (2023). Enhancing Sustainable Rice Grain Quality Analysis with Efficient SVM Optimization Using Genetic Algorithm. *E3S Web of Conferences*, *426*, 01035.

Samah, K. A. F. A., Wirakarnain, N. S. D., Hamzah, R., Moketar, N. A., Riza, L. S., & Othman, Z. (2022). A linear regression approach to predicting salaries with visualizations of job vacancies: a case study of Jobstreet Malaysia. *Int J Artif Intell ISSN*, *2252*(8938), 1131.

Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1464–1470.

Umagapi, I. T., Umaternate, B., Hazriani, H., & Yuyun, Y. (2023). Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa. *Prosiding Sisfotek*, *7*(1), 303–308.