OPEN ACCESS

Volume 9 (2), April-June 2025, 703-712

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i2.3438

Implementasi Data Mining untuk *Clustering* Lowongan Pekerjaan Menggunakan Metode Algoritma *K-Means*

Rifqi Mubarok [1], Akhmal Angga Syahputra [2*], Abdillah Teguh Permana [3], Lifa Sholiah [4], Tarwoto [5]

[1,2,3,4,5] Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Kota Purwokerto, Provinsi Jawa Tengah, Indonesia.*

a r t i c l e i n f o

Article history:
Received 6 December 2024
Received in revised form
20 December 2024
Accepted 10 January 2025
Available online April 2025.

Keywords:
Data Mining; Job Vacancies; K-Means; Clustering; CRISP-DM.

Kata Kunci:
Data Mining; Lowongan Pekerjaan; K-Means; Klustering; CRISP-DM.

a b s t r a c t

The development of digital technology has transformed the way businesses recruit employees online. This study aims to create an interactive dashboard that facilitates job seekers and companies, using clustering methods with the K-Means algorithm to analyze job posting data in the United States. The data from the Kaggle LinkedIn Job Postings 2023 dataset, consisting of 33,000 records, is processed using the CRISP-DM phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The clustering analysis results in four job categories: low-mid-level general jobs, high-level executive jobs, time-based jobs, and mid-high-level professional jobs. Model evaluation shows good clustering quality with a Silhouette Coefficient of 0.78 and a Davies-Bouldin Index of 0.55. The developed dashboard helps companies plan recruitment and job seekers find positions matching their skills and salary expectations. The practical contribution of this study is modernizing the recruitment process, assisting companies and recruitment agencies in screening candidates more efficiently, and improving job matching through deeper data analysis.

a b s t r a k

Perkembangan teknologi digital telah mengubah cara bisnis merekrut karyawan secara online. Penelitian ini bertujuan membuat dashboard interaktif yang mempermudah pencari kerja dan perusahaan, menggunakan metode clustering dengan algoritma K-Means untuk menganalisis data lowongan kerja di Amerika Serikat. Data dari dataset Kaggle LinkedIn Job Postings 2023, mencakup 33.000 data, diproses menggunakan tahapan CRISP-DM: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. Analisis clustering menghasilkan empat kategori pekerjaan: pekerjaan umum tingkat rendah-menengah, pekerjaan eksekutif tingkat tinggi, pekerjaan berbasis waktu, dan pekerjaan profesional tingkat atas-menengah. Evaluasi model menunjukkan kualitas klaster yang baik dengan Silhouette Coefficient (0,78) dan Davies-Bouldin Index (0,55). Dashboard yang dibangun membantu perusahaan merencanakan perekrutan dan pencari kerja menemukan pekerjaan sesuai keterampilan dan ekspektasi gaji. Kontribusi praktis penelitian ini adalah memodernisasi proses rekrutmen, membantu perusahaan dan agensi rekrutmen menyaring kandidat secara lebih efisien, serta meningkatkan kecocokan pekerjaan melalui analisis data yang lebih mendalam.

Corresponding Author. Email: 21sa2129@mhs.amikompurwokerto.ac.id [2].

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

1. Pendahuluan

Perkembangan teknologi yang pesat telah membawa dampak signifikan di berbagai sektor, termasuk pada media daring yang kini menjadi salah satu saluran utama dalam penyebaran informasi lowongan pekerjaan. Di era digital, banyak perusahaan memanfaatkan platform daring tidak hanya untuk memasarkan produk tetapi juga dalam proses rekrutmen (Chandra, 2023). Pendekatan ini lebih efisien dibandingkan metode konvensional, karena pencari kerja dapat mengakses informasi lowongan dengan cepat dan luas (Destiyanti *et al.*, 2023). Platform seperti *LinkedIn*, *JobStreet*, *Karier.com*, dan *Jobs.id* telah memudahkan pencari kerja dalam menemukan peluang yang sesuai serta membantu perusahaan dalam menemukan kandidat yang tepat. Namun, untuk menarik minat pencari kerja yang sesuai, perusahaan harus mampu merancang iklan lowongan pekerjaan yang menarik, baik dari segi susunan kata maupun visual. Setiap elemen dalam iklan lowongan perlu diperhatikan agar dapat menarik perhatian pembaca (Dewi & Nursiyono, 2023). Selain itu, kemampuan pelamar untuk menemukan posisi yang relevan dengan keterampilan mereka juga menjadi faktor penting dalam pencarian kerja.

Meskipun banyak informasi tersedia, data lowongan pekerjaan sering tersebar di berbagai situs, baik di platform penyedia pekerjaan maupun situs resmi perusahaan, yang mengakibatkan kesulitan dalam mengakses informasi secara efisien. Dengan menggunakan *big data*, informasi tersebut dapat dikumpulkan, dikelompokkan, dan dianalisis dari berbagai sumber untuk mempermudah akses dan pemanfaatannya dalam skala besar (Husna, 2024). Penelitian ini menggunakan metode klasterisasi untuk mengelompokkan data lowongan pekerjaan berdasarkan kesamaan karakteristik. Metode klasterisasi mempermudah pencari kerja dalam menemukan posisi yang sesuai dengan keterampilan mereka, serta membantu perusahaan dalam menyaring kandidat secara lebih terarah (Winarta & Kurniawan, 2021). Salah satu algoritma dalam klasterisasi adalah *K-Means*, yang efektif dalam membagi data ke dalam kelompok yang memiliki karakteristik serupa. Algoritma ini merupakan teknik klasterisasi non-hirarki yang mengelompokkan data ke dalam beberapa cluster, dengan setiap cluster berisi data yang memiliki kesamaan karakteristik. Hasil klasterisasi ini memberikan informasi yang berguna bagi pembuat kebijakan dalam proses pengambilan keputusan. Berbagai penelitian sebelumnya telah membuktikan efektivitas algoritma *K-Means* dalam pengelompokan data. Misalnya, penelitian yang dilakukan oleh Ulummuddin dkk., untuk menganalisis *pain points* yang dialami oleh pengguna dengan menggunakan *K-Means* dan beberapa metode evaluasi (Ulummuddin & Sari, 2020). Penelitian lain yang dilakukan oleh Mayasari dan Nugraha (2023) mengelompokkan kabupaten/kota di Provinsi Jawa Tengah berdasarkan data kemiskinan untuk mengidentifikasi karakteristik masing-masing wilayah. Selain itu, penelitian oleh Sembiring dkk., menggunakan *K-Means* untuk memetakan desa yang terjangkit Demam Berdarah Dengue dengan evaluasi *Devies-Bouldin Index* (DBI) (Sembiring, 2021). Penelitian oleh Wahyudi dan Silfia (2022) menggunakan *K-Means* untuk menentukan strategi penjualan yang optimal bagi toko *S&R Baby Store*, dengan hasil evaluasi DBI sebesar 0,560 yang menunjukkan hasil klasterisasi yang cukup baik. Berdasarkan penelitian-penelitian sebelumnya yang mengaplikasikan algoritma *K-Means* dalam berbagai bidang, terdapat potensi besar untuk penerapannya dalam pengelompokan data lowongan pekerjaan.

2. Metodologi Penelitian

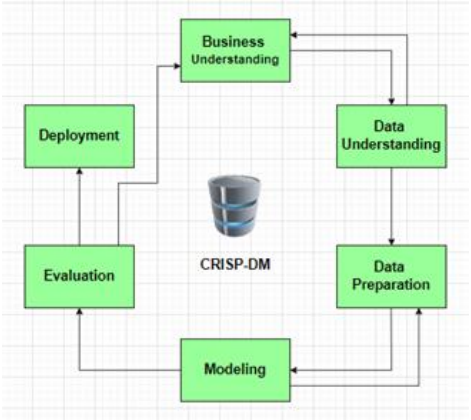
Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset publik *LinkedIn Job Postings* yang tersedia di Kaggle pada tahun 2023, yang berisi informasi lowongan pekerjaan di Amerika Serikat. Dataset ini mencakup 33.000 data lowongan pekerjaan, yang terbagi dalam empat kategori, yaitu data postingan, detail pekerjaan, detail perusahaan, dan mapping. Data postingan terdiri dari 25 kolom, data detail pekerjaan terdiri dari 15 kolom, data detail perusahaan terdiri dari 18 kolom, dan data mapping terdiri dari 4 kolom. Dataset ini dipilih karena relevansinya dengan tujuan penelitian, yang bertujuan untuk menganalisis tren lowongan pekerjaan, keterampilan yang dibutuhkan, serta hubungan antara perusahaan dan jenis pekerjaan. Dataset ini mencakup berbagai sektor dan posisi pekerjaan, memberikan volume data yang

besar dan representatif untuk menggali informasi tentang pasar tenaga kerja di Amerika Serikat. Selain itu, dataset ini berasal dari *LinkedIn*, platform profesional terbesar di dunia, yang menjamin akurasi dan kredibilitas data. Dengan data yang terperinci dan beragam, dataset ini memungkinkan analisis komprehensif mengenai faktor-faktor seperti keterampilan yang dicari, lokasi pekerjaan, dan ukuran perusahaan, yang sangat relevan untuk penelitian ini.

Tahapan Penelitian

Penelitian ini mengadopsi metode *CRISP-DM* (Cross-Industry Standard Process for Data Mining). *CRISP-DM* adalah metode data mining yang dikembangkan oleh Daimler-Chrysler, SPSS, dan NCR, dan dapat diterapkan pada berbagai alat dan lini bisnis (Sarimole & Hakim, 2024). Metode ini terdiri dari enam tahapan, yaitu: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan implementasi (*deployment*) (Singgalen, 2023). Tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:



Gambar 1. Tahapan Metode CRISP-DM

Pemahaman Bisnis (*Business Understanding*)

Pada tahapan ini, dilakukan studi literatur mengenai klasterisasi pekerjaan serta analisis aplikasi pencari kerja. Penelitian ini bertujuan untuk mengidentifikasi masalah yang sering dihadapi oleh pelamar kerja dan sumber daya manusia (HR) dalam proses perekrutan karyawan. Berdasarkan analisis tersebut, ditemukan solusi untuk beberapa masalah yang ada, yaitu dengan menggunakan analisis klasterisasi.

Pemahaman Data (*Data Understanding*)

Pada tahapan ini, dilakukan pemahaman menyeluruh terhadap data yang berjumlah sebanyak 33.000 data. Proses ini meliputi pemahaman mengenai tipe data, atribut data, relasi antar data, jumlah entitas pada setiap tabel, jumlah kolom, dan keterangan atribut data yang terdapat dalam dataset.

Persiapan Data (*Data Preparation*)

Pada tahapan ini, dilakukan persiapan data sebelum memasuki tahap analisis dan pemodelan, untuk menghindari terjadinya anomali dan inkonsistensi pada data. Beberapa tahapan yang dilakukan dalam persiapan data adalah sebagai berikut:

1) Data *Preprocessing*

Tahapan ini bertujuan untuk mengidentifikasi nilai data yang hilang dan deteksi outlier. Penanganan dilakukan dengan cara menghapus baris data yang memiliki nilai hilang atau melakukan imputasi untuk mengisi data yang hilang. Untuk menangani outlier, dilakukan penghapusan dengan menggunakan metode *interquartile range* (Dendi & Sanjaya, 2024). Rumus yang digunakan untuk mendeteksi outlier mengikuti pendekatan yang dijelaskan oleh Sihombing dkk. (2023).

$$IQR = Q3 - Q1$$
$$Fence\ low = Q1 - 1.5 \times IQR$$
$$Fence\ high = Q3 + 1.5 \times IQR$$

Keterangan:
Q1: quartile ke-1
Q3: quartile ke-3
Fence_low: batas bawah
Fence_high: batas atas

2) Transformasi Data

Pada tahapan ini, dilakukan penggabungan data dari berbagai tabel (*data integration*) serta pengkodean ulang (*data encoding*) untuk mempermudah pengolahan data lebih lanjut (Rizquina, 2023). Dataset ini terdiri dari empat data yang terpisah, yang kemudian digabungkan (*data merging*) untuk mempermudah pemodelan. Dataset ini mencakup dua tipe data, yaitu data kategorikal dan numerik. Pada tahapan *data selection*, transformasi dilakukan pada data kategorikal agar menjadi numerik, guna

mempermudah identifikasi data yang memiliki korelasi tinggi.

3) Rekayasa Fitur (*Feature Engineering*)
   Tahapan ini bertujuan untuk membuat beberapa fitur baru yang dapat mendukung proses analisis menggunakan *machine learning* (Sari *et al*., 2024).
4) Seleksi Fitur (*Feature Selection*)
   Tahapan ini bertujuan untuk menghilangkan data yang memiliki korelasi tinggi dan mengidentifikasi data yang mengalami multikolinearitas. Jika suatu variabel tidak memberikan informasi yang signifikan, maka variabel tersebut harus dihapus (Zuguang, 2022). Pada tahapan ini, juga dilakukan seleksi fitur yang paling relevan menggunakan *Heatmap Correlation* untuk mendeteksi fitur dengan korelasi tinggi, serta *Variance Inflation Factor* (VIF) untuk mendeteksi adanya multikolinearitas pada fitur (Khakim *et al*., 2023). Penggunaan *Heatmap Correlation* dan *VIF* dalam seleksi fitur membantu menyaring fitur yang relevan serta menghilangkan redundansi, yang pada gilirannya meningkatkan kualitas model. Dengan memilih fitur yang memiliki korelasi rendah satu sama lain dan memberikan kontribusi signifikan terhadap model, kami dapat memperoleh model yang lebih efisien, cepat, dan akurat.

Pemodelan (*Modeling*)

Pemodelan dimulai dengan menentukan model yang tepat untuk menyelesaikan permasalahan yang ada. Pada tahap ini, beberapa algoritma diuji untuk menentukan yang paling sesuai dengan karakteristik data, dan diputuskan untuk menggunakan metode klasterisasi (*clustering*) dengan algoritma *K-Means*. Pemilihan algoritma ini didasarkan pada kesederhanaannya, sifatnya yang intuitif, serta waktu komputasi yang cepat. Meskipun algoritma lain seperti *DBSCAN* dan *hierarchical clustering* dapat digunakan, keduanya memiliki keterbatasan.

DBSCAN lebih cocok untuk data dengan noise dan klaster yang berbentuk tidak beraturan, namun memerlukan penyesuaian parameter yang rumit; sementara *hierarchical clustering* lebih lambat dan kurang efisien untuk dataset yang besar. *K-Means* lebih unggul karena skalabilitasnya, kecepatan, dan kemampuannya yang efektif dalam mengelompokkan data dengan klaster yang terdefinisi dengan jelas, menjadikannya pilihan yang optimal untuk dataset lowongan pekerjaan ini (Adiputra, 2022). Data yang memiliki karakteristik serupa akan dikelompokkan dalam satu klaster, sedangkan data dengan karakteristik yang berbeda akan dikelompokkan dalam klaster yang lain (Adiputra, 2022). Rumus yang digunakan untuk klasterisasi adalah sebagai berikut:

$$d\,(xi, \mu j) = \sqrt{\Sigma(xi - \mu j)^2}$$
$$\mu j(t+1) = \frac{1}{Nsj}\Sigma j \in sj^{xj}$$

Evaluasi (*Evaluation*)

Evaluasi dilakukan untuk menganalisis hasil dari pemodelan yang telah dibangun. Metode yang dipilih dalam evaluasi ini adalah *silhouette coefficient*. *Silhouette coefficient* digunakan untuk menilai kualitas dan kekuatan klaster, serta mengukur seberapa baik suatu objek ditempatkan dalam klaster yang sesuai (Shoolihah *et al*., 2017). Dengan menggunakan metode ini, dapat dinilai kualitas model yang telah dibangun berdasarkan jumlah klaster yang telah ditentukan.

Penyebaran (*Deployment*)

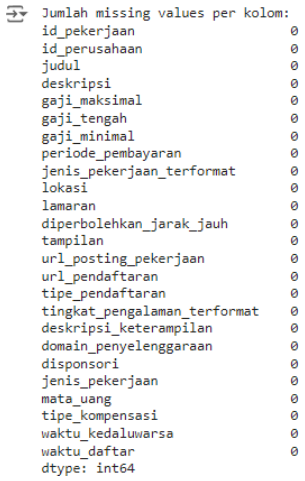
Pada tahap akhir, dilakukan pembuatan dasbor analisis berdasarkan hasil dari analisis dan pemodelan yang telah dilakukan. Dasbor ini dirancang untuk mempermudah akses informasi bagi para pencari kerja, manajemen sumber daya manusia (HR), lembaga pendidikan dan pelatihan kerja, investor, dan perusahaan.

3. Hasil dan Pembahasan

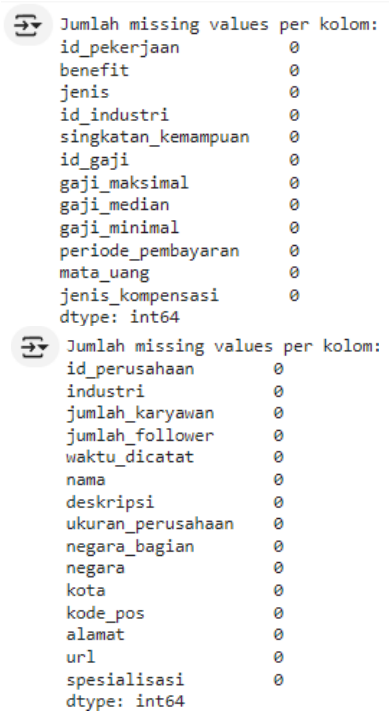
Hasil

Data *Preprocessing*

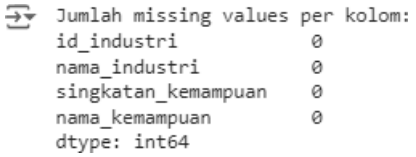
Setelah dilakukan pemahaman terhadap data, ditemukan bahwa beberapa data mengandung masalah seperti nilai yang hilang (*missing value*) dan outliers. Oleh karena itu, dilakukan penanganan sebagai berikut:



Gambar 2. *Handle Missing Value* data postingan lowongan kerja



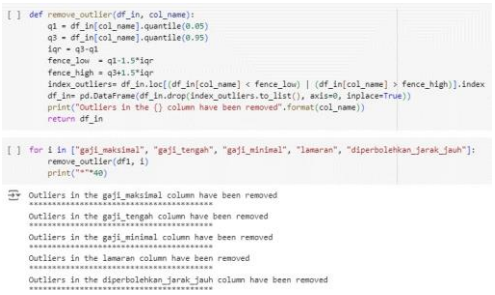
Gambar 3. *Handle Missing Value* data detail pekerjaan



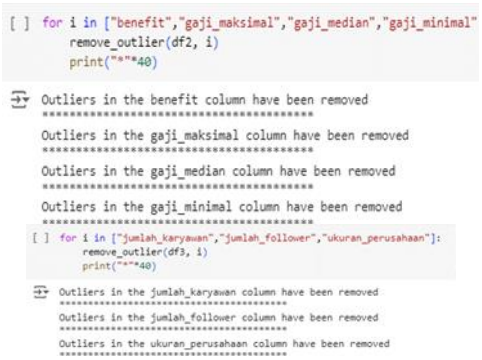
Gambar 4. *Handle Missing Value data mapping*

Tahapan penanganan data hilang (*missing value*) dilakukan dengan menggunakan metode imputasi, yaitu dengan memasukkan nilai rata-rata (*mean*) atau median untuk data numerik, dan memasukkan nilai modus untuk data kategorikal.

Selanjutnya, penanganan outliers dilakukan untuk meningkatkan akurasi model, menjaga kualitas data, dan mendeteksi anomali dalam dataset.



Gambar 5. *Handle Outliers* data postingan lowongan pekerjaan



Gambar 6. *Handle Outliers* data detail pekerjaan

Penanganan outliers dilakukan hanya pada tiga data pertama, sedangkan data terakhir (*mapping*) tidak memiliki outliers yang signifikan.

Transformasi Data

Pada dataset ini, terdapat empat set data yang terpisah. Untuk mempermudah pemodelan, dilakukan penggabungan data (*data merging*). Dataset ini terdiri dari dua tipe data, yaitu data kategorikal dan numerikal. Pada tahapan *data selection*, transformasi data kategorikal menjadi numerik diperlukan untuk mempermudah identifikasi data yang memiliki korelasi tinggi.

Rekayasa Fitur (*Feature Engineering*)

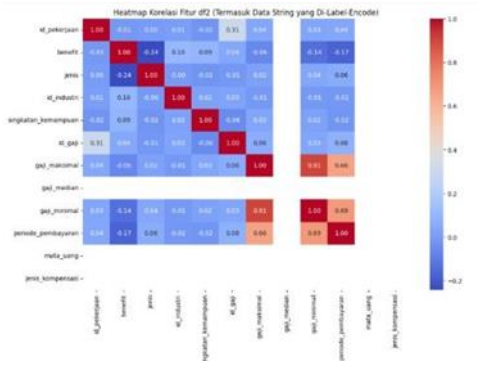
Penambahan fitur dilakukan untuk menyederhanakan model dengan merangkum informasi dari beberapa fitur yang sudah ada, serta mengembangkan fitur baru. Dalam proses ini, dihasilkan dua fitur baru, yaitu rata-rata karyawan dan ukuran perusahaan, yang diperoleh dari fitur jumlah karyawan.

Tabel 1. *Feature Engineering*

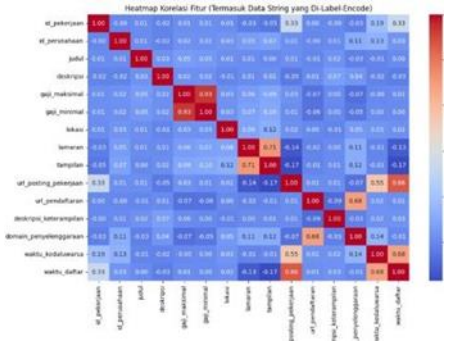
No	Rata-rata Karyawan	Ukuran Perusahaan
1	73	Kecil
2	219	Sedang
3	783	Besar

Feature Selection

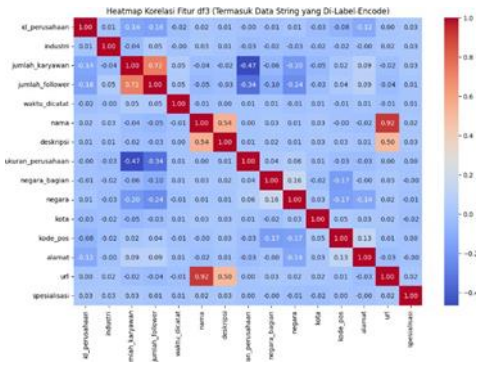
Tahap ini, membantu menghilangkan data yang memiliki korelasi tinggi dan data yang mempunyai multikolinearitas. Menggunakan visualiasi *heatmap correlation* untuk mendeteksi data korelasi tinnggi.



Gambar 7. *Heatmap correlation* postingan lowongan kerja



Gambar 8. *Heatmap correlation* detail pekerjaan



Gambar 9. *Heatmap correlation* detail perusahaan

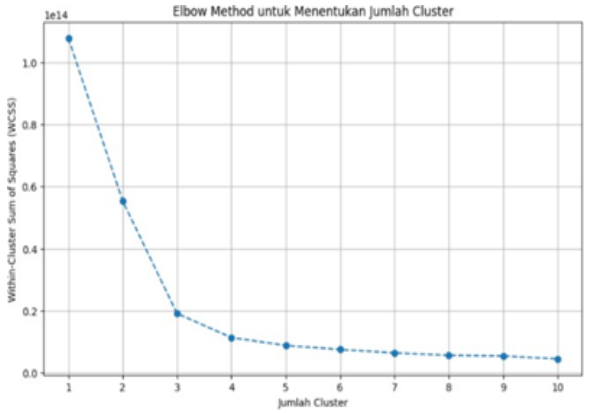
Fitur yang memiliki korelasi tinggi adalah fitur yang memiliki nilai mendekati angka 1. Fitur-fitur dalam kategori ini dapat dipertimbangkan untuk dihapus. Namun, ada beberapa fitur yang mungkin tetap dipertimbangkan untuk digunakan, tergantung pada kebutuhan dalam pemodelan serta pertimbangan dari sisi bisnis. Berdasarkan visualisasi yang telah dilakukan, berikut adalah beberapa fitur yang akan dihapus.

Tabel 2. Fitur yang dihapus

No	Data	Fitur yang dihapus
1	Postingan Lowongan Pekerjaan	Gaji tengah, diperbolehkan jarak jauh.
2	Detail Pekerjaan	Gaji median, mata uang, jenis kompensasi

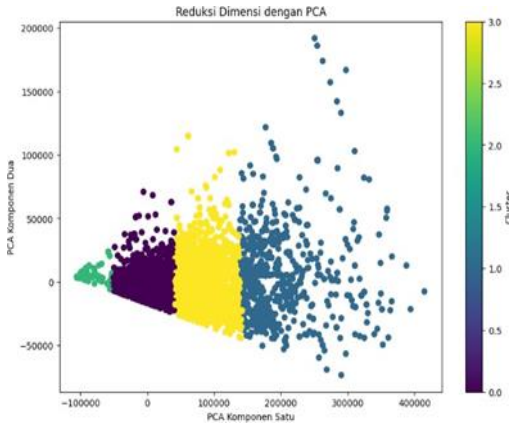
Pemodelan (*Modeling*)

Pada tahap ini, ditentukan teknik data mining yang akan digunakan, yaitu pemodelan klasterisasi (*clustering*) dengan algoritma *K-Means*. Langkah pertama yang dilakukan adalah memilih fitur-fitur yang akan digunakan dalam proses pemodelan. Fitur-fitur yang dipilih untuk digunakan dalam pemodelan ini antara lain: deskripsi, gaji maksimal, gaji minimal, periode pembayaran, lokasi, lamaran, tampilan, tipe pendaftaran, tingkat pengalaman terformat, deskripsi keterampilan, disponsori, dan jenis pekerjaan. Selanjutnya, jumlah klaster yang akan digunakan ditentukan melalui visualisasi menggunakan metode *elbow*. Metode optimasi *elbow* digunakan untuk menentukan nilai *K* optimum, yang ditentukan dengan melihat persentase grafik yang membentuk sudut tajam (Rizquina, 2023).



Gambar 10. *Elbow Method*

Berdasarkan visualisasi diatas, titik yang mengalami penurunan paling akhir adalah di titik cluster 4. Titik ini disebut titik siku, yang bermakna titik jumlah klaster optimal pada titik nomor 4.



Gambar 11. *Cluster Visualization*

Hasil visualisasi klaster dapat dilihat pada Gambar 11, yang menunjukkan pemisahan data menjadi empat klaster yang berbeda. Klaster 0, yang ditandai dengan warna ungu, mencakup data dengan karakteristik tertentu yang membedakannya dari klaster lainnya. Klaster 1, yang berwarna biru, menggambarkan data dengan atribut yang serupa namun berbeda dari klaster 0. Klaster 2, berwarna hijau, mencerminkan kelompok data dengan karakteristik yang lebih spesifik, sementara klaster 3, yang berwarna kuning, menunjukkan kelompok data dengan atribut yang unik namun saling terkait. Pembagian data ke dalam empat klaster ini menunjukkan efektivitas pemodelan dalam mengelompokkan data berdasarkan kesamaan fitur yang relevan. Hasil klasterisasi ini memberikan gambaran yang lebih jelas tentang distribusi data dan hubungan antar kategori lowongan pekerjaan yang dianalisis dalam penelitian ini.

Tabel 3. Hasil *Cluster*

Cluster ke-	Nama Cluster	Deskripsi
Cluster 0	*Low to Mid- Level General Jobs*	Mencakup pekerjaan dengan gaji terendah hingga menengah, mayoritas di lebel mid-senior dan entry
Cluster 1	*High-Level Executive and Specialized Jobs*	Terdiri dari posisi eksekutif dan spesialis dengan gaji sangat tinggi, berlokasi di kota-kota besar seperti New York dan San Francisco, dan memerlukan pengalaman senior atau direktur.
Cluster 2	*Hourly and Temporary Jobs*	Mencakup pekerjaan dengan pembayaran per jam banyak di level mid-senior dan entry, tersebar di kota besar seperti New York dan Los Angeles.
Cluster 3	*Upper Mid-Level Professional Jobs*	Mencakup posisi professional dengan gaji sangat tinggi, berlokasi di pusat-pusat urban utama dan membutuhkan pengalaman mid-senior.

Evaluasi *(Evaluation)*

Model yang sudah dibuat, dilakukan evaluasi untuk mengukur koherensi klaster atau melihat informasi seberapa dekat titik data dengan klaster. Tahapan pada evaluasi ini menggunakan metode *silhouette coefficient* dan *Davies-Bouldin Index* (DBI). Hasilnya nilai *Silhouette* dan DBI pada klaster 4 mencapai 0.78 dan 0.55.

Tabel 4. *Silhoutte coefficient and Davies-Bouldin Index* (DBI)

Nilai Silhoutte Rata-rata	0,78048
Nilai Davies Bouldin Index	0,55457

Nilai *silhouette* pada klaster ini termasuk dalam kategori baik. Semakin tinggi nilai *silhouette* yang mendekati angka 1, semakin baik kualitas klaster tersebut. *Silhouette coefficient* digunakan untuk menilai kualitas dan kekuatan klaster, serta mengukur tingkat kedekatan objek dalam klaster (Khairunnas *et al.*, 2023). Selain itu, nilai *Davies-Bouldin Index* (DBI) juga menunjukkan kualitas yang baik; semakin rendah nilai DBI yang mendekati angka 0, semakin baik kualitas klaster tersebut. Pada penelitian yang dilakukan oleh Manalu dan Gunadi dengan judul "Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada CV Digital Dimensi," nilai DBI terbaik yang ditemukan adalah 0,502146, yang kemudian memutuskan untuk menggunakan jumlah klaster sebanyak 5 (Manalu & Gunadi, 2023). Sementara itu, pada penelitian ini, digunakan 4 klaster yang ditentukan berdasarkan hasil visualisasi dengan

metode *elbow*. Pada tahap terakhir ini, dilakukan pembuatan dasbor yang berisi visualisasi segmentasi pekerjaan dan analisis klaster berdasarkan pemodelan yang telah dibuat, menggunakan alat *Looker Studio*. Hasil visualisasi ini dapat dimanfaatkan oleh para pencari kerja untuk menemukan pekerjaan yang sesuai dengan kemampuan dan gaji yang diharapkan. Selain itu, dasbor ini juga dapat digunakan oleh HR (*Human Resources*) untuk menyusun strategi dalam pembuatan iklan lowongan pekerjaan agar dapat menarik kandidat terbaik.

Gambar 12. *Dashboard* segmentasi pekerjaan

Gambar 13. *Dashboard* segmentasi pekerjaan lanjutan

Gambar 14. *Dashboard* Analisis Klaster

Pembahasan

Penelitian ini memanfaatkan algoritma *K-Means* untuk klasterisasi data lowongan pekerjaan, dengan tujuan mengelompokkan lowongan pekerjaan berdasarkan fitur-fitur yang relevan untuk mempermudah pencari kerja dan perusahaan dalam proses rekrutmen. Algoritma *K-Means* dipilih karena kesederhanaannya dan efisiensinya dalam menangani data besar, serta kemampuannya dalam mengelompokkan data yang memiliki klaster yang jelas, seperti yang dijelaskan oleh Winarta dan Kurniawan (2021). Meskipun algoritma lain seperti *DBSCAN* dan *hierarchical clustering* dapat digunakan, keduanya memiliki keterbatasan, seperti kesulitan dalam menangani data yang mengandung noise dan bentuk klaster yang tidak teratur, serta kinerja yang lebih lambat pada dataset besar (Sembiring, 2021; Manalu & Gunadi, 2023). Dalam evaluasi model, nilai *silhouette coefficient* yang mendekati angka 1 dan *Davies-Bouldin Index* (DBI) yang mendekati angka 0 menunjukkan bahwa klaster yang terbentuk memiliki kualitas yang baik, sesuai dengan temuan Khairunnas *et al.* (2023). Penelitian ini memilih jumlah klaster sebanyak 4 berdasarkan visualisasi dengan metode *elbow*, yang menunjukkan titik optimal pada angka tersebut, meskipun penelitian sebelumnya oleh Manalu dan Gunadi (2023) menggunakan 5 klaster dengan nilai DBI terbaik 0,502146.

Hasil klasterisasi ini diterjemahkan dalam bentuk dasbor visualisasi yang dibangun menggunakan *Looker Studio*. Dasbor ini memungkinkan para pencari kerja untuk menemukan pekerjaan yang sesuai dengan keterampilan dan gaji yang diharapkan, serta membantu HR (*Human Resources*) dalam merancang iklan lowongan yang lebih tepat sasaran, yang dapat menarik kandidat terbaik. Penerapan *data mining* dalam lowongan pekerjaan ini memberikan kontribusi penting, karena mengubah cara data digunakan untuk menyaring dan mengelompokkan informasi yang relevan bagi kedua belah pihak. Selain itu, meskipun algoritma *K-Means* efektif pada dataset ini, ada peluang untuk menguji algoritma lain seperti *DBSCAN* atau *hierarchical clustering* dalam pengelompokan data dengan karakteristik yang lebih kompleks (Sarimole & Hakim, 2024). Penelitian ini menawarkan pendekatan baru dalam penggunaan *clustering* untuk analisis pasar tenaga kerja, yang dapat memberikan wawasan strategis bagi perusahaan dan pencari kerja untuk merencanakan proses rekrutmen yang lebih efisien.

4. Kesimpulan

Berdasarkan analisis dan pengujian pada pemodelan dataset lowongan pekerjaan di Amerika Serikat menggunakan metode klasterisasi dengan algoritma *K-Means*, serta evaluasi menggunakan *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI), dapat disimpulkan bahwa algoritma *K-Means* efektif dalam mengelompokkan data lowongan pekerjaan berdasarkan atribut-atribut utama seperti lokasi, industri, dan keterampilan yang dibutuhkan. Hasil pengujian menunjukkan bahwa jumlah klaster yang optimal dapat ditemukan melalui analisis iteratif, yang memungkinkan pemisahan data menjadi kelompok-kelompok yang lebih homogen dan relevan. Proses evaluasi menggunakan *Silhouette Coefficient* dan DBI menunjukkan bahwa kualitas klasterisasi yang dilakukan cukup baik, dengan pemisahan antar klaster yang jelas serta kecocokan antar data dalam setiap klaster yang relatif tinggi.

Algoritma *K-Means* menunjukkan hasil yang memadai, terdapat beberapa area yang dapat diperbaiki untuk menghasilkan model yang lebih akurat. Sebagai rekomendasi, penggunaan algoritma lain seperti *DBSCAN* atau *Agglomerative Clustering* dapat dipertimbangkan untuk mengatasi kelemahan *K-Means* dalam menangani data dengan distribusi yang tidak terstruktur atau adanya outlier. Selain itu, pengembangan lebih lanjut pada pemilihan fitur, seperti menambahkan variabel terkait pengalaman kerja atau pendidikan, dapat meningkatkan kualitas klasterisasi dan memberikan wawasan yang lebih mendalam bagi pengguna. Hasil penelitian ini dapat diterapkan untuk meningkatkan efisiensi proses rekrutmen di industri dengan mengotomatisasi penyaringan lowongan pekerjaan dan kandidat berdasarkan kecocokan keterampilan, lokasi, dan preferensi. Platform rekrutmen berbasis dashboard interaktif yang dihasilkan dapat memfasilitasi pencari kerja dalam mencari lowongan yang paling relevan dengan keterampilan dan tujuan karir mereka. Di sisi lain, perusahaan dapat memanfaatkan hasil klasterisasi untuk memetakan kebutuhan tenaga kerja mereka dan menemukan kandidat yang lebih tepat dengan lebih cepat. Selain itu, *dashboard* ini dapat diperluas dengan fitur analisis tren industri, yang memberikan wawasan mengenai sektor yang sedang berkembang dan keterampilan yang paling dibutuhkan oleh pasar kerja. Hal ini tidak hanya akan membantu pencari kerja dalam merencanakan pengembangan karir mereka, tetapi juga memberikan nilai tambah bagi perusahaan yang ingin menyesuaikan strategi perekrutan dengan tren pasar. Penelitian ini memberikan kontribusi penting dalam pemanfaatan teknologi data untuk mempermudah proses pencocokan antara pencari kerja dan perusahaan. Dengan penerapan yang tepat, algoritma klasterisasi dan dashboard interaktif ini berpotensi untuk merevolusi cara rekrutmen dilakukan, baik untuk individu maupun organisasi, dalam menghadapi tantangan dunia kerja yang semakin kompetitif.

5. Daftar Pustaka

Adiputra, I. N. M. (2021). Clustering Penyakit Dbd Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means. *INSERT: Information System and Emerging Technology Journal*, *2*(2), 99-105.

Chandra, E. (2023). MANFAAT PERSONAL BRANDING & PROFESSIONAL NETWORKING UNTUK HIRING DECISION PADA MEDIA LINKEDIN BAGI PERUSAHAAN. *AKSES: JOURNAL OF PUBLIK & BUSINESS ADMINISTRATION SCIENCE*, *5*(1), 21-34.

Destiyanti, D., Nugroho, W. B., & Kamajaya, G. APLIKASI LINKEDIN DALAM PERSPEKTIF KONSTRUKSI SOSIAL TEKNOLOGI.

Dewi, D. M., & Nursiyono, J. A. (2023). Pengaruh Online Adversiting terhadap Pencarian Kerja di Indonesia (Studi Kasus: jobs. id dan Google Trends). *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, *3*(1), 8-15. https://doi.org/10.20885/snati.v3i1.26.

Furqon, M. T., & Widodo, A. W. (2017). Implementasi Metode Improved K-Means untuk Mengelompokkan Titik Panas Bumi. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *1*(11), 1270-1276.

Gu, Z. (2022). Complex heatmap visualization. *Imeta*, *1*(3), e43.

Hunt, E. B. (2014). *Artificial intelligence*. Academic Press.

Husna, F. I., & Tranggono, T. (2024). Implementasi Data Analytic Dalam Upaya Peningkatan Penjualan Properti Sebesar 10% Di NYC Amerika Serikat. *Venus: Jurnal Publikasi Rumpun Ilmu Teknik*, *2*(1), 134-144.

Khairunnas, M. A., Jamaludin, A., & Adam, R. I. (2023). Pengaruh Pendapatan Orang Tua terhadap Hasil Belajar Siswa Menggunakan Algoritma K-Means Clustering. *Jurnal Pendidikan Tambusai*, *7*(3), 31434-31444. https://doi.org/10.31004/jptam.v7i3.12130.

Khakim, E. N. R., Hermawan, A., & Avianto, D. (2023). Implementasi correlation matrix pada klasifikasi dataset wine. *JIKO (Jurnal Informatika dan Komputer)*, *7*(1), 158-166. http://dx.doi.org/10.26798/jiko.v7i1.771.

Manalu, D. A., & Gunadi, G. (2022). Implementasi Metode Data Mining K-Means Clustering Terhadap Data Pembayaran Transaksi Menggunakan Bahasa Pemrograman Python Pada Cv Digital Dimensi. *Infotech: Journal of Technology Information*, *8*(1), 43-54. https://doi.org/10.37365/jti.v8i1.131.

Mayasari, S. N., & Nugraha, J. (2023). Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *3*(2), 317-329. https://doi.org/10.24002/konstelasi.v3i2.7200.

Rizquina, A. Z. (2023). *Perbandingan Penjualan Produk Halal Labeled dan Non-labeled pada E-commerce Tokopedia Indonesia* (Doctoral dissertation, Universitas Islam Indonesia).

Sari, N. N., Anisah, T. T., & Fitriani, R. (2024). Implementasi Machine Learning Untuk Prediksi Harga Laptop Menggunakan Algoritma Regresi Linear Berganda. *Jurnal Manajemen Informatika (JAMIKA)*, *14*(2), 162-177. https://doi.org/10.34010/jamika.v14i2.1292.

Sarimole, F. M., & Hakim, L. (2024). Klasifikasi Barang Menggunakan Metode Clustering K-Means Dalam Penentuan Prediksi Stok Barang. *Jurnal Sains dan Teknologi*, *5*(3), 846-854. https://doi.org/10.55338/saintek.v5i3.2709.

Sembiring, M. A., Agus, R. T. A., & Sibuea, M. F. L. (2021). Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD). *Journal of Science and Social Research*, *4*(3), 336-341. https://doi.org/10.54314/jssr.v4i3.712.

Sihombing, P. R., Suryadiningrat, S., Sunarjo, D. A., & Yuda, Y. P. A. C. (2022). Identifikasi data outlier (pencilan) dan kenormalan data pada data univariat serta alternatif penyelesaiannya. *Jurnal Ekonomi Dan Statistik Indonesia*, *2*(3), 307-316.

Singgalen, Y. A. (2023). Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting, Targeting, Positioning) Layanan Akomodasi Hotel, Homestay, dan Resort. *Jurnal Media Informatika Budidarma*, *7*(4), 1980-1993.

Wahyudi, T., & Silfia, T. (2022). Implementation of Data Mining Using K-Means Clustering Method to Determine Sales Strategy In S&R Baby Store. *Journal of Applied Engineering and Technological Science (JAETS)*, *4*(1), 93-103. https://doi.org/10.37385/jaets.v4i1.913.

Winarta, A., & Kurniawan, W. J. (2021). Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python. *JTIK (Jurnal Teknik Informatika Kaputama)*, *5*(1), 113-119.