

Segmentasi Mahasiswa Berdasarkan Kesiapan Karir menggunakan Algoritma K-Means dan Visualisasi Interaktif di Telkom University Surabaya

1st Ananda Taqhsya Dwiyana
Sains Data
Telkom University Surabaya
Surabaya, Indonesia

nandatqhsy@student.telkomuniversity.ac.id

2nd Regita Putri Permata, S.Stat., M.Stat
Sains Data
Telkom University Surabaya
Surabaya, Indonesia

regitapermata@telkomuniversity.ac.id

3rd Rifdatun Ni'mah, S.Si., M.Si
Sains Data
Telkom University Surabaya
Surabaya, Indonesia

rifdatun@telkomuniversity.ac.id

Abstrak — Tingginya angka keraguan mahasiswa semester akhir terhadap motivasi dan kompetensi kerja mereka menunjukkan pentingnya evaluasi terhadap kesiapan karir mahasiswa. Pra-survei yang dilakukan di Telkom University Surabaya mengungkap bahwa 78% mahasiswa merasa tidak yakin terhadap motivasi internal mereka, dan 83% meragukan kemampuan mereka untuk bersaing di dunia kerja. Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengelompokkan mahasiswa berdasarkan tingkat kesiapan karir menggunakan algoritma K-Means, serta menyajikan hasilnya dalam bentuk dashboard interaktif. Lima faktor utama yang dianalisis meliputi motivasi, kematangan pribadi, kematangan sosial, sikap kerja, dan kompetensi kerja. Data dikumpulkan melalui kuesioner skala Likert dan dianalisis secara langsung menggunakan algoritma K-Means untuk membentuk kelompok mahasiswa dengan karakteristik kesiapan karir yang serupa. Setelah kluster terbentuk, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA) guna memvisualisasikan hasil kluster dalam ruang dua dimensi. Validasi jumlah kluster optimal dilakukan menggunakan metode Elbow dan Silhouette Score. Penelitian ini menghasilkan tiga kluster utama yaitu kluster Siap Kerja, kluster Menuju Siap Kerja, dan kluster Butuh Pembinaan. Visualisasi interaktif melalui Looker Studio membantu dalam memahami karakteristik tiap kluster secara lebih dinamis. Hasil penelitian ini mendukung pengambilan keputusan berbasis data oleh Career Development Center (CDC) dalam merancang program pengembangan karir yang lebih tepat sasaran.

Kata kunci— Kesiapan karir, K-Means, segmentasi mahasiswa, PCA visualisasi, dashboard interaktif

I. PENDAHULUAN

Kesiapan karir merupakan aspek penting yang menentukan kesuksesan mahasiswa dalam memasuki dunia kerja yang kompetitif dan dinamis. Namun, hasil pra-survei terhadap mahasiswa semester akhir di Telkom University Surabaya menunjukkan bahwa 78% responden masih

meragukan motivasi internal mereka, dan 83% merasa tidak yakin terhadap kompetensi kerja yang dimiliki. Temuan ini mengindikasikan perlunya strategi pengembangan karir yang lebih tepat sasaran dan berbasis data. Dalam dunia kerja yang menuntut soft skills, pengambilan keputusan yang tepat, serta kemampuan teknis dan interpersonal yang kuat, kesiapan karir tidak lagi cukup diukur melalui pencapaian akademik semata, melainkan juga oleh kematangan pribadi, sosial, dan profesional mahasiswa.

Berbagai penelitian sebelumnya telah menyoroti pentingnya faktor-faktor seperti motivasi, kematangan pribadi dan sosial, sikap kerja, serta kompetensi kerja sebagai indikator kesiapan karir [1], [2]. Namun, penelitian-penelitian tersebut sebagian besar masih berfokus pada validasi instrumen atau penerapan algoritma dalam konteks non-karir, seperti pengelompokan topik atau nilai akademik. Integrasi antara teknik segmentasi berbasis machine learning dan visualisasi interaktif dalam konteks kesiapan karir mahasiswa masih jarang dilakukan. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menggabungkan pendekatan algoritmik dan representasi visual yang informatif untuk mendukung pengambilan keputusan strategis di lingkungan pendidikan tinggi.

Berdasarkan latar belakang dan studi terkini, penelitian ini mengangkat dua permasalahan utama. Pertama, bagaimana membangun model segmentasi kesiapan karir mahasiswa menggunakan algoritma K-Means. Kedua, bagaimana menyajikan hasil segmentasi tersebut secara informatif melalui dashboard interaktif yang dapat dimanfaatkan oleh pemangku kepentingan seperti Career Development Center (CDC) Telkom University Surabaya.

Tujuan dari penelitian ini adalah untuk membangun model segmentasi berbasis algoritma K-Means yang mampu

mengelompokkan mahasiswa berdasarkan kesiapan karir mereka, serta mengembangkan dashboard visualisasi interaktif menggunakan Looker Studio. Dengan demikian, diharapkan hasil penelitian ini dapat digunakan sebagai landasan dalam merancang program pengembangan karir yang lebih personal, efisien, dan relevan dengan kebutuhan masing-masing klaster mahasiswa.

II. KAJIAN TEORI

A. Kesiapan Karir Mahasiswa

Kesiapan karir merupakan kemampuan individu dalam mempersiapkan diri menghadapi dunia kerja, baik dari aspek teknis maupun psikologis. Kesiapan karir mencakup kecakapan dalam mengambil keputusan, mengelola diri, dan merancang masa depan profesional secara mandiri[3]. Mahasiswa yang memiliki kesiapan karir tinggi cenderung mampu merespons perubahan dan tuntutan dunia kerja dengan lebih baik. Kesiapan ini tidak hanya ditentukan oleh kemampuan akademik, tetapi juga oleh kematangan dalam berpikir, berinteraksi, dan bertindak secara profesional [4]. Oleh karena itu, perguruan tinggi memiliki tanggung jawab untuk membekali mahasiswa dengan keterampilan yang relevan agar transisi ke dunia kerja dapat berjalan lancar.

B. Faktor-Faktor Kesiapan Karir

Kesiapan karir mahasiswa dipengaruhi oleh sejumlah faktor yang saling berkaitan. Lima faktor utama yang menjadi fokus dalam penelitian ini, yaitu motivasi, kematangan pribadi, kematangan sosial, sikap kerja, dan kompetensi kerja[1]. Motivasi karir mencerminkan dorongan internal untuk mencapai tujuan pekerjaan, sedangkan kematangan pribadi berkaitan dengan pengendalian emosi dan kemampuan membuat keputusan secara mandiri. Kematangan sosial mengacu pada keterampilan berkomunikasi dan bekerja sama dalam tim, yang sangat penting di lingkungan profesional. Sikap kerja mencakup etos kerja, kedisiplinan, dan tanggung jawab, sementara kompetensi kerja mengarah pada keterampilan teknis maupun non-teknis yang dibutuhkan dalam dunia kerja. Kelima faktor ini menjadi kerangka penting dalam mengukur kesiapan mahasiswa untuk menghadapi dunia profesional.

C. Algoritma K-Means

K-Means merupakan salah satu algoritma dalam data mining yang digunakan untuk melakukan pengelompokan (clustering) data berdasarkan kesamaan karakteristik. Dalam penelitian ini, K-Means digunakan untuk mengelompokkan mahasiswa ke dalam beberapa klaster berdasarkan tingkat kesiapan karir mereka. Algoritma ini bekerja dengan cara menginisialisasi sejumlah pusat klaster, lalu menghitung jarak setiap data ke pusat tersebut menggunakan rumus Euclidean, dan memperbarui posisi pusat hingga stabil. K-Means sangat efisien dalam mengelompokkan data besar dan dapat membantu institusi pendidikan dalam mengenali pola kesiapan karir mahasiswa[5]. Penerapan algoritma ini memungkinkan Career Development Center (CDC) mengidentifikasi kelompok mahasiswa yang memerlukan intervensi lebih lanjut dalam pengembangan karir.

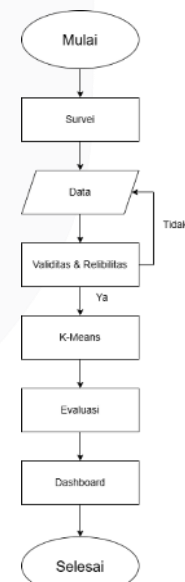
D. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah teknik reduksi dimensi yang digunakan untuk menyederhanakan struktur data yang kompleks tanpa kehilangan informasi penting. Dalam penelitian ini, PCA diterapkan setelah proses klasterisasi untuk membantu memvisualisasikan hasil pengelompokan dalam bentuk scatter plot dua dimensi. PCA bekerja dengan cara mentransformasikan data ke dalam komponen-komponen utama yang mewakili varian terbesar. PCA dapat meningkatkan interpretabilitas data dan sangat berguna dalam mendukung analisis visual, terutama untuk melihat distribusi klaster secara lebih jelas dan intuitif[6]. Melalui visualisasi ini, proses evaluasi hasil klaster menjadi lebih informatif dan mudah dipahami oleh pemangku kepentingan.

E. Visualisasi Interaktif dan Dashboard

Visualisasi data merupakan komponen penting dalam menyampaikan hasil analisis secara efektif. Dalam konteks penelitian ini, visualisasi interaktif digunakan untuk mempermudah eksplorasi dan interpretasi hasil segmentasi kesiapan karir mahasiswa. Looker Studio dipilih sebagai platform visualisasi karena kemampuannya dalam menyajikan data secara real-time, interaktif, dan mudah diakses. Dashboard interaktif memungkinkan pengguna untuk melakukan filter, membandingkan klaster, dan mengidentifikasi tren berdasarkan faktor-faktor seperti jenis kelamin, usia, dan program studi [7]. Dengan demikian, visualisasi interaktif tidak hanya memperjelas hasil analisis, tetapi juga mendukung pengambilan keputusan berbasis data yang lebih tepat sasaran di tingkat institusional.

III. METODE



GAMBAR 1
(FLOWCHART)

A. Desain dan Waktu Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksploratif-deskriptif, bertujuan untuk mengelompokkan mahasiswa berdasarkan kesiapan karir serta menyajikannya dalam visualisasi interaktif. Penelitian dilaksanakan pada bulan April hingga Mei 2025 di lingkungan Telkom University Surabaya. Fokus utama

penelitian adalah untuk membangun segmentasi kesiapan karir mahasiswa menggunakan algoritma K-Means Clustering dan memvisualisasikan hasilnya melalui dashboard interaktif guna mendukung pengambilan keputusan di Career Development Center (CDC).

B. Sumber dan Teknik Pengumpulan Data

Data primer diperoleh melalui penyebaran kuesioner daring menggunakan Google Form. Responden yang ditargetkan adalah mahasiswa aktif semester 7 ke atas dari berbagai program studi. Kuesioner mengukur lima dimensi kesiapan karir: motivasi, kematangan pribadi, kematangan sosial, sikap kerja, dan kompetensi kerja. Masing-masing dimensi terdiri dari beberapa indikator dengan total 30 item pernyataan, diukur menggunakan skala Likert 1–5. Jumlah sampel ditentukan menggunakan rumus Slovin dengan margin of error 5%:

$$n = \frac{N}{1 + N \cdot e^2} \quad (1)$$

di mana n adalah ukuran sampel, N adalah populasi mahasiswa, dan e adalah margin of error.

C. Uji Validitas dan Reliabilitas

Instrumen penelitian diuji validitas dan reliabilitasnya sebelum digunakan. Uji validitas dilakukan menggunakan korelasi Pearson Product Moment, dengan kriteria valid jika nilai korelasi lebih besar dari r tabel. Rumus korelasi Pearson dituliskan sebagai berikut:

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{[n \sum X^2 - (\sum X)^2][n \sum Y^2 - (\sum Y)^2]}} \quad (2)$$

di mana r adalah koefisien validitas, X adalah skor item, Y adalah skor total, dan n adalah jumlah responden. Reliabilitas instrumen diuji menggunakan Cronbach's Alpha, dengan interpretasi reliabel jika nilai $\alpha \geq 0.7$. Rumusnya adalah:

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum s_i^2}{s_t^2} \right) \quad (3)$$

dengan α adalah koefisien reliabilitas, n adalah jumlah item, s_i^2 adalah variansi tiap item, dan s_t^2 adalah variansi total.

D. Prosedur Pengolahan Data

Data yang valid dan reliabel dianalisis menggunakan algoritma K-Means Clustering, yaitu metode *unsupervised learning* yang membagi data ke dalam k kelompok berdasarkan kesamaan karakteristik. Pengelompokan didasarkan pada jarak Euclidean antara data dan pusat kluster, dengan rumus:

$$D(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (4)$$

di mana $D(i, j)$ adalah jarak antara data ke- i dan ke- j , dan x_{ik} adalah nilai pada dimensi ke- k . Untuk menentukan jumlah kluster optimal, digunakan dua metode:

1. Elbow Method, yang mengamati titik infleksi pada grafik *Within-Cluster Sum of Squares (WCSS)* di hitung dengan rumus:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} ||x - \mu_i||^2 \quad (5)$$

2. Silhouette Score, yaitu indeks evaluasi kluster yang dihitung dengan rumus:

$$S = \frac{b - a}{\max(a, b)} \quad (6)$$

dengan a adalah rata-rata jarak antar data dalam satu kluster, dan b adalah jarak rata-rata ke kluster terdekat.

E. Visualisasi Hasil Segmentasi

Setelah proses klusterisasi selesai, dilakukan reduksi dimensi menggunakan *Principal Component Analysis (PCA)* untuk memetakan data ke dalam dua dimensi visual. PCA digunakan agar hasil kluster dapat divisualisasikan dalam scatter plot dengan representasi yang lebih mudah dipahami. Proyeksi data ke komponen utama dihitung menggunakan rumus:

$$Y = Z \cdot V_k \quad (7)$$

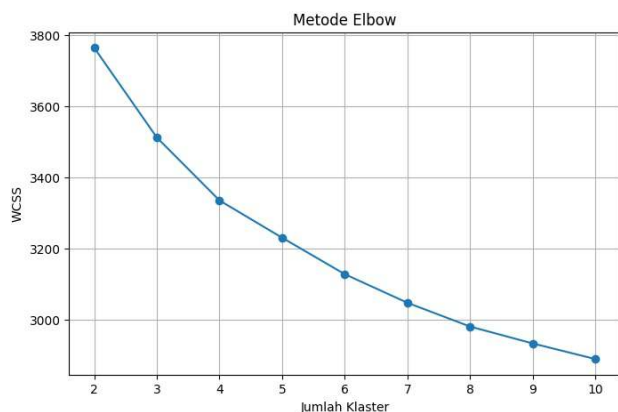
di mana Y adalah data hasil transformasi, Z adalah data yang telah dinormalisasi, dan V_k adalah vektor eigen dari komponen utama. Hasil akhir divisualisasikan ke dalam dashboard interaktif menggunakan Looker Studio, yang menyediakan filter berdasarkan variabel demografis (jenis kelamin, usia, program studi) serta kluster. Visualisasi ini berfungsi sebagai alat bantu utama dalam menganalisis tren kesiapan karir mahasiswa dan merancang program intervensi yang lebih tepat sasaran oleh CDC.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan segmentasi mahasiswa Telkom University Surabaya berdasarkan kesiapan karir menggunakan algoritma *K-Means Clustering*. Data kesiapan karir diperoleh dari kuesioner yang mencakup lima faktor utama: motivasi, kematangan pribadi, kematangan sosial, sikap kerja, dan kompetensi kerja. Analisis dilakukan terhadap 244 responden mahasiswa semester akhir.

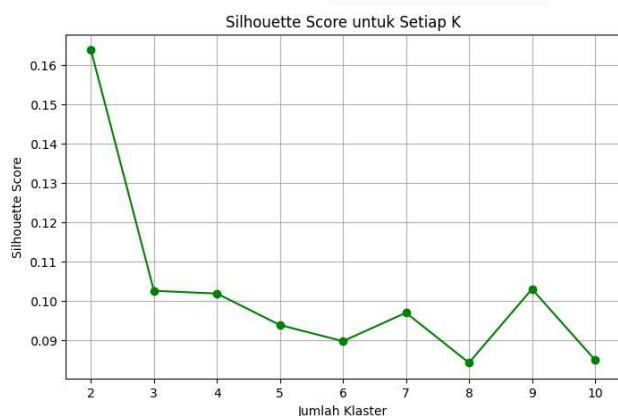
A. Penentuan Jumlah Kluster

Penentuan jumlah kluster optimal dilakukan menggunakan dua metode, yaitu Elbow Method dan Silhouette Score. Berdasarkan Elbow Method, titik siku yang signifikan terlihat pada $k = 3$, menandakan bahwa segmentasi paling optimal terjadi ketika data dibagi menjadi tiga kluster.



GAMBAR 2
(METODE ELBOW)

Berdasarkan Gambar 3, grafik metode elbow menunjukkan bahwa penurunan nilai *Within-Cluster Sum of Squares* (WCSS) paling signifikan terjadi antara kluster ke-2 dan ke-3. Setelah kluster ke-3, penurunan WCSS mulai melandai, menandakan bahwa penambahan kluster selanjutnya hanya memberikan sedikit kontribusi terhadap pengurangan variasi dalam kluster. Titik siku tajam pada kluster ke-3 diidentifikasi sebagai jumlah kluster optimal. Pemilihan tiga kluster mencerminkan keseimbangan antara representasi struktur data yang akurat dan kesederhanaan model. Jumlah ini dianggap ideal untuk menghindari over-segmentasi maupun under-representasi, serta menjadi dasar dalam proses analisis segmentasi kesiapan karir mahasiswa.



GAMBAR 3
(SILHOUETTE SCORE)

Berdasarkan Gambar 3, nilai Silhouette Score tertinggi diperoleh pada saat $k = 2$, menunjukkan kohesi kluster yang kuat dan pemisahan yang jelas. Namun, dua kluster dinilai terlalu sederhana untuk merepresentasikan kompleksitas kesiapan karir mahasiswa, karena cenderung menghasilkan pembagian biner yang kurang informatif. Meskipun nilai skor menurun pada $k = 3$, penurunan tersebut justru mencerminkan pembelahan struktural yang penting dalam data. Setelah $k = 3$, penurunan Silhouette Score berlangsung lebih landai, menunjukkan stabilitas model. Oleh karena itu, pemilihan tiga kluster dianggap sebagai titik optimal yang seimbang antara akurasi segmentasi dan kemudahan interpretasi. Struktur tiga kluster memungkinkan kategorisasi kesiapan karir yang lebih representatif, yakni “siap kerja”, “menuju siap kerja”, dan “butuh pembinaan”.

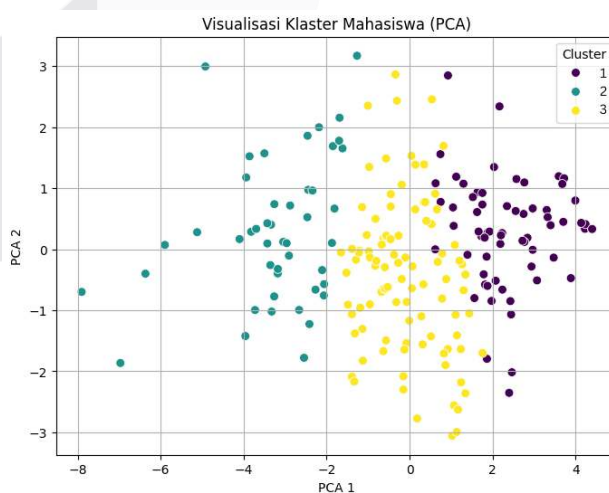
B. Karakteristik Kluster Kesiapan Karir

Setelah pembentukan kluster menggunakan algoritma *K-Means*, diperoleh tiga kelompok kesiapan karir mahasiswa, yaitu Kluster 1 (Mahasiswa Siap Kerja), Kluster 2 (Mahasiswa Butuh Pembinaan), dan Kluster 3 (Mahasiswa Menuju Siap Kerja). Analisis deskriptif dilakukan untuk mengetahui jumlah mahasiswa, proporsi, serta rata-rata skor kesiapan karir pada tiap kluster. Hasilnya disajikan pada Tabel 1 berikut:

TABEL 1
(HASIL KLASERISASI)

Kluster	Jumlah Mahasiswa	Persentase	Rata-rata Skor
Kluster 1 Mahasiswa Siap Kerja	76 mahasiswa	31%	4,63
Kluster 2 Mahasiswa Butuh Pembinaan	54 mahasiswa	22%	3,62
Kluster 3 Mahasiswa Menuju Siap Kerja	114 mahasiswa	47%	4,19

Dari tabel tersebut terlihat bahwa mayoritas mahasiswa berada dalam Kluster 3 (Menuju Siap Kerja), menunjukkan bahwa hampir separuh populasi mahasiswa sedang berada dalam tahap transisi menuju kesiapan penuh untuk memasuki dunia kerja. Kluster 1 memiliki rata-rata skor tertinggi sebesar 4,63, menandakan bahwa mahasiswa dalam kelompok ini telah menunjukkan kesiapan karir yang matang dari berbagai aspek. Sementara itu, Kluster 2 menunjukkan skor paling rendah (3,62), yang mengindikasikan perlunya intervensi dan pembinaan lanjutan untuk meningkatkan kesiapan mereka. Distribusi ini memperkuat pentingnya segmentasi dalam merancang strategi pengembangan karir yang lebih personal dan tepat sasaran.



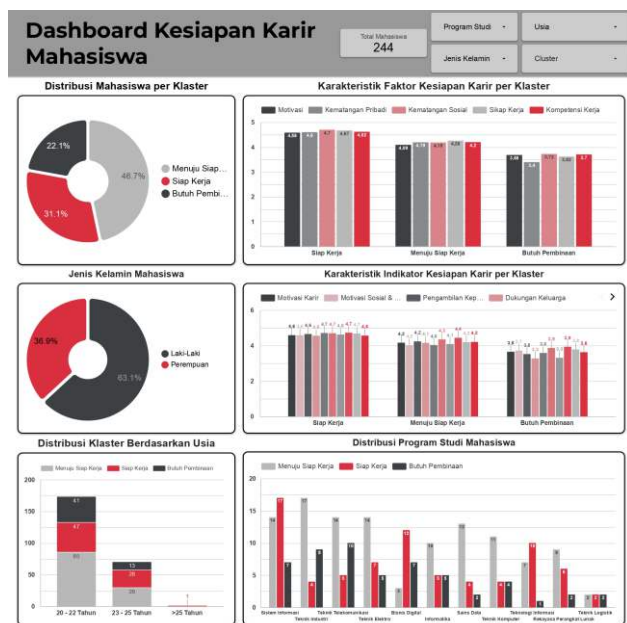
GAMBAR 4
(VISUALISASI PCA DUA DIMENSI TERHADAP HASIL KLASERISASI KESIAPAN KARIR)

Gambar 5 menyajikan hasil visualisasi data menggunakan teknik Principal Component Analysis (PCA), yang mereduksi 30 item pernyataan dalam kuesioner menjadi dua

komponen utama: PCA 1 dan PCA 2. Transformasi ini menyederhanakan data berdimensi tinggi ke dalam dua dimensi tanpa kehilangan informasi utama, sehingga setiap mahasiswa dapat direpresentasikan sebagai titik dalam bidang dua dimensi. Scatter plot hasil PCA menunjukkan tiga kluster hasil segmentasi, dengan Kluster 1 cenderung berada di sisi kanan (kesiapan karir tinggi), Kluster 2 di sisi kiri (kesiapan rendah), dan Kluster 3 tersebar di tengah (kesiapan bervariasi). Visualisasi ini mendukung validitas pemisahan antar kluster dan memperlihatkan struktur karakteristik kesiapan karir yang khas. Meskipun tidak digunakan untuk interpretasi faktor, PCA berperan penting dalam menyederhanakan data untuk proses klusterisasi yang lebih efektif dan dapat divisualisasikan secara jelas.

C. Visualisasi Interaktif dan Dashboard

Untuk meningkatkan keterbacaan hasil penelitian dan memudahkan proses pengambilan keputusan, hasil segmentasi divisualisasikan menggunakan dashboard interaktif berbasis Looker Studio. Dashboard ini menampilkan distribusi mahasiswa per kluster, persebaran berdasarkan jenis kelamin, usia, serta karakteristik indikator kesiapan karir.



GAMBAR 5
(DASHBOARD INTERAKTIF)

Pengguna dapat mengeksplorasi data melalui filter yang tersedia, misalnya memilih kluster tertentu, membandingkan program studi, atau menganalisis kesiapan berdasarkan usia dan gender. Visualisasi ini memberikan fleksibilitas kepada CDC untuk merancang intervensi spesifik. Sebagai contoh, mahasiswa dalam kluster “Butuh Pembinaan” dapat difokuskan pada pelatihan peningkatan motivasi dan soft skill.

V. KESIMPULAN

Penelitian ini berhasil membangun model segmentasi kesiapan karir mahasiswa Telkom University Surabaya menggunakan algoritma K-Means. Berdasarkan lima faktor utama motivasi, kematangan pribadi, kematangan sosial, sikap kerja, dan kompetensi kerja terbentuk tiga kluster utama: Siap Kerja, Menuju Siap Kerja, dan Butuh Pembinaan. Validasi menggunakan *Elbow Method* dan Silhouette Score menunjukkan bahwa tiga kluster merupakan jumlah optimal secara struktural dan interpretatif. Hasil klusterisasi divisualisasikan dengan PCA dalam dua dimensi untuk memperjelas pemisahan antar kelompok, dan disajikan melalui dashboard interaktif berbasis Looker Studio. Dashboard ini memfasilitasi pemahaman mendalam terhadap distribusi mahasiswa dan mendukung *Career Development Center (CDC)* dalam merancang program pengembangan karir yang lebih tepat sasaran dan berbasis data.

REFERENSI

- [1] A. Prianto, W. Winardi, and U. K. N. Qomariyah, “The Effect of the Implementation of Teaching Factory and Its Learning Involvement toward Work Readiness of Vocational School Graduates,” *International Journal of Instruction*, 2021, [Online].
- [2] R. Saputra and K. T. Kustina, “Minat Mahasiswa Akuntansi Untuk Mengikuti Pendidikan Profesi Penilai Ditinjau Dari Motivasi Sosial, Motivasi Karir Dan Motivasi Ekonomi,” *Jurnal Ilmiah Akuntansi dan Bisnis*, 2019, [Online].
- [3] N. C. Gysbers, “Career-Ready Students: A Goal of Comprehensive School Counseling Programs,” *Career Dev Q*, vol. 61, no. 3, pp. 283–288, 2013, doi: <https://doi.org/10.1002/j.2161-0045.2013.00057.x>.
- [4] M. Moore and J. Thaller, “Career readiness: preparing social work students for entry into the workforce,” *Front Educ (Lausanne)*, vol. 8, Nov. 2023, doi: 10.3389/feduc.2023.1280581.
- [5] D. Chi, “Research on the Application of K-Means Clustering Algorithm in Student Achievement,” *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pp. 435–438, 2021, [Online].
- [6] J. Colyar, K. Michael, R. Maciejewski, and L. Tate, “Improving Educational Standards Using Visualization Dashboards for Decision Making,” in *2022 IEEE International Symposium on Technology and Society (ISTAS)*, IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ISTAS55053.2022.10227101.
- [7] A. Bellanov and L. Nurhayati, “K-Means Clustering Analysis Untuk Menentukan Strategi Promosi Kampus,” *Jurnal Teknik Industri: Jurnal Hasil Penelitian dan Karya Ilmiah dalam Bidang Teknik Industri*, 2023, [Online].