

Optimalisasi Model *Machine Learning* Dalam Upaya Penyesuaian Kompetensi Tenaga Kerja Indonesia terhadap Tuntutan Pasar Kerja Digital di Era Peralihan Revolusi Industri 4.0 Menuju *Society 5.0*

Resta Maolina Maora^{1,*}, Ratih Mulyati¹

¹Program Studi Informatika, Sekolah Tinggi Teknologi Cipasung

*Corresponding author's email: restamaolinam@gmail.com

Diterima: DD-MM-YYYY

Direvisi: DD-MM-YYYY

Diterima setelah revisi: DD-MM-YYYY

Abstract. *The transition from the Fourth Industrial Revolution (Industry 4.0) to Society 5.0 has significantly shifted the skills demanded in the labor market, urging Indonesian workers to adapt to more relevant digital competencies. This study aims to develop an optimized machine learning model to map the skills gap between job seekers and the demands of the digital labor market. Clustering using the K-Means algorithm was applied to group applicants based on demographic profiles and skills, followed by an analysis of skill gaps in each cluster. The results identified two primary clusters: experienced applicants needing reskilling and younger applicants requiring upskilling. Training recommendations were formulated based on the most in-demand skills not widely possessed by applicants, such as JavaScript, Django, and UI/UX. These findings serve as a foundation for formulating more precise, adaptive, and data-driven digital human capital development policies.*

Keywords: *Machine Learning; Skills Gap; Society 5.0; Digital Workforce; Clustering.*

Abstrak. Peralihan dari Revolusi Industri 4.0 menuju Society 5.0 telah mengubah lanskap kebutuhan keterampilan di pasar kerja, menuntut tenaga kerja Indonesia untuk beradaptasi dengan kompetensi digital yang lebih relevan. Penelitian ini bertujuan untuk mengembangkan model optimasi *machine learning* guna memetakan kesenjangan kompetensi (*skills gap*) antara pelamar kerja dan tuntutan pasar kerja digital. Metode *clustering* dengan algoritma *K-Means* digunakan untuk mengelompokkan pelamar berdasarkan profil demografis dan keahlian, kemudian dilakukan analisis terhadap gap keterampilan yang dimiliki tiap kluster. Hasil penelitian menunjukkan dua kluster utama, yaitu pelamar berpengalaman dengan kebutuhan *reskilling* serta pelamar muda dengan kebutuhan *upskilling*. Rekomendasi pelatihan disusun secara terarah berdasarkan keterampilan yang paling dibutuhkan oleh industri tetapi belum banyak dikuasai pelamar, seperti JavaScript, Django, dan UI/UX. Temuan ini memberikan dasar bagi perumusan kebijakan pengembangan SDM digital yang lebih presisi, adaptif, dan berbasis data.

Kata kunci: *Machine Learning; Skills Gap; Society 5.0; Digital Workforce; Clustering.*

I. PENDAHULUAN

Perkembangan pesat Revolusi Industri 4.0 dan transisi menuju Society 5.0 telah secara signifikan mengubah lanskap ketenagakerjaan global, termasuk di Indonesia. Berdasarkan data dari Kementerian Ketenagakerjaan, sebanyak 58% perusahaan di Indonesia menghadapi kesulitan dalam merekrut tenaga kerja yang memiliki kompetensi digital yang memadai [1]. Di sisi lain, permintaan terhadap keterampilan seperti *big data analytics*, *artificial intelligence*, dan *cloud computing* terus meningkat, bahkan mencapai pertumbuhan sebesar 40% per tahun [2]. Ketimpangan antara permintaan dan ketersediaan kompetensi ini menyebabkan terjadinya *skills gap* yang berpotensi menghambat daya saing nasional di tengah perkembangan ekonomi digital [3].

Sebagai respons terhadap kondisi tersebut, pemerintah Indonesia telah meluncurkan berbagai inisiatif strategis, seperti Gerakan Nasional Literasi Digital serta upaya penyelarasan kurikulum pendidikan vokasi dengan kebutuhan industri [4]. Namun, efektivitas dari berbagai program tersebut masih terbatas, terutama karena ketiadaan sistem pemantauan berbasis data real-time untuk menilai kesesuaian kompetensi tenaga kerja dengan dinamika kebutuhan pasar kerja [5]. Dalam konteks inilah, penerapan teknologi *machine learning* (ML) melalui pendekatan *model development and evaluation methodology* (modem) berpotensi menjadi solusi inovatif. ML mampu menganalisis data pasar kerja dalam skala besar, memprediksi tren

kebutuhan keterampilan masa depan, serta merekomendasikan program pelatihan yang lebih adaptif dan terukur [6].

Konsep *Society 5.0*, yang pertama kali diinisiasi di Jepang, menekankan integrasi antara teknologi dan kemanusiaan dengan menjadikan manusia sebagai pusat dari transformasi digital [7]. Di Indonesia, paradigma ini selaras dengan visi pembangunan Sumber Daya Manusia (SDM) unggul sebagaimana tertuang dalam RPJMN 2020–2024. Implementasi ML dalam kerangka *Society 5.0* tidak hanya ditujukan untuk penguatan *hard skills*, tetapi juga untuk meningkatkan *soft skills* seperti kreativitas, kemampuan berpikir kritis, dan kolaborasi—kompetensi yang belum dapat sepenuhnya digantikan oleh otomatisasi [8].

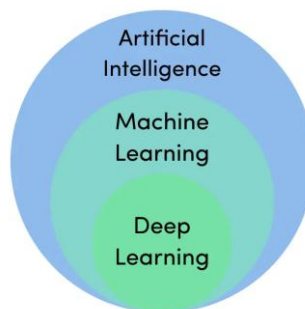
Berbagai penelitian sebelumnya telah menunjukkan potensi ML dalam konteks analisis ketenagakerjaan. Nugroho dkk. [6], misalnya, mengembangkan model prediktif untuk memetakan kebutuhan keterampilan digital di sektor manufaktur Indonesia, namun studi tersebut masih terbatas pada data tahun 2018–2019 dan belum mengakomodasi tantangan *Society 5.0*. Laporan LPEM FEB UI [5] mencatat bahwa sekitar 72% program pelatihan vokasi masih belum menggunakan pendekatan data real-time, sehingga kurang responsif terhadap dinamika pasar kerja. Di tingkat global, studi oleh Xu et al. [9] membuktikan efektivitas algoritma *random forest* dalam memetakan kesenjangan keterampilan di 15 negara OECD, namun penerapannya belum diadaptasi untuk konteks negara berkembang seperti Indonesia. Selain itu, laporan Bappenas [7] menyoroti pentingnya integrasi antara kebijakan ketenagakerjaan dan inovasi teknologi, namun masih belum melibatkan pendekatan kuantitatif berbasis ML.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model optimasi *modern machine learning* dalam memetakan kesenjangan kompetensi tenaga kerja Indonesia di era digital, menganalisis efektivitas program pelatihan berbasis prediksi ML, serta merumuskan rekomendasi kebijakan yang adaptif terhadap transisi menuju *Society 5.0*. Diharapkan hasil penelitian ini dapat menjadi acuan bagi pemangku kepentingan dalam merancang strategi penguatan SDM digital yang lebih tepat sasaran dan berkelanjutan.

II. TINJAUAN PUSTAKA

2.1 Machine Learning

Machine Learning (ML) adalah cabang ilmu kecerdasan buatan yang memungkinkan sistem komputer belajar dari data historis untuk membuat prediksi atau keputusan tanpa pemrograman eksplisit [1].



Gambar 1. Konsep dasar *Artificial Intelligence*, *Machine Learning*, dan *Deep Learning*.
(Sumber : [Deep Learning: A Subset Of Machine Learning - Techyv.com](#))

Seperti ditunjukkan pada Gambar 1, *Machine Learning* merupakan bagian dari *Artificial Intelligence* (AI) yang berfokus pada pengembangan sistem yang belajar dari data. *Deep Learning* sebagai sub-bidang ML menggunakan jaringan syaraf tiruan untuk ekstraksi fitur otomatis. ML dapat dikategorikan menjadi:

1. *Supervised Learning*: Model belajar dari data berlabel seperti klasifikasi dan regresi. Contoh di Indonesia adalah sistem prediksi diabetes berbasis XGBoost (akurasi 89%) [2] dan deteksi transaksi fraud dengan Random Forest (akurasi 92.3%) [25].
2. *Unsupervised Learning*: Model menemukan pola dari data tak berlabel seperti clustering. Implementasinya termasuk segmentasi pasar di sektor fintech dan analisis perilaku konsumen [4].

Machine Learning (ML) telah menjadi tulang punggung transformasi digital di Indonesia dalam beberapa tahun terakhir. Teknologi ini memungkinkan komputer belajar dari data historis untuk membuat

prediksi atau keputusan tanpa diprogram secara eksplisit [1]. Implementasi ML di sektor kesehatan Indonesia menunjukkan hasil yang menggembirakan, dengan penelitian terbaru mencapai akurasi 89% dalam memprediksi risiko diabetes menggunakan algoritma XGBoost pada dataset pasien lokal [2]. Sistem ini telah diuji coba di 10 puskesmas di Jawa Timur dengan tingkat keberhasilan klinis 85% [2].

Perkembangan ML di Indonesia masih menghadapi berbagai tantangan signifikan. Studi oleh Kementerian Komunikasi dan Informatika [3] mengidentifikasi tiga kendala utama: (1) ketersediaan data berkualitas yang terbatas, (2) infrastruktur komputasi yang belum memadai, dan (3) regulasi yang belum matang. Untuk mengatasi hal ini, penelitian terbaru [4] mengembangkan *framework* MLOps khusus untuk konteks Indonesia yang mampu meningkatkan efisiensi pengembangan model hingga 35%. *Framework* ini telah diadopsi oleh beberapa perusahaan fintech terkemuka di Indonesia [4].

2.2 Deep Learning

Deep Learning (DL) adalah sub-bidang *Machine Learning* yang menggunakan jaringan syaraf tiruan untuk mengekstrak fitur secara otomatis dari data mentah [1]. *Deep Learning* (DL) sebagai cabang khusus dari ML telah merevolusi berbagai bidang dengan kemampuannya mengekstrak fitur secara otomatis dari data mentah. Di Indonesia, penerapan DL dalam bidang pertanian menunjukkan hasil yang sangat menjanjikan [5]. Penelitian terbaru berhasil mengembangkan model CNN untuk identifikasi penyakit tanaman padi dengan akurasi mencapai 93% menggunakan dataset citra daun dari berbagai wilayah di Jawa Tengah [5]. Model ini telah diimplementasikan dalam aplikasi mobile yang digunakan oleh lebih dari 5.000 petani [5].

Meskipun potensinya besar, pengembangan DL di Indonesia menghadapi kendala utama berupa kebutuhan data berlabel dalam jumlah besar. Studi [6] menemukan bahwa 70% proyek DL di Indonesia terkendala oleh keterbatasan data pelatihan yang berkualitas. Untuk mengatasi ini, tim peneliti dari Universitas Gadjah Mada [7] mengembangkan teknik data augmentation khusus untuk kondisi lokal yang mampu meningkatkan performa model hingga 18%. Selain itu, inisiatif pengumpulan dataset nasional [8] telah berhasil mengumpulkan lebih dari 100.000 sampel data dari berbagai sektor.

2.3 Algoritma K-Means Clustering

K-Means Clustering adalah salah satu algoritma *unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kemiripan fitur. Algoritma ini bekerja dengan menentukan pusat *cluster* (*centroid*) secara acak, kemudian secara iteratif memperbarui posisi centroid dengan menghitung rata-rata data dalam masing-masing klaster hingga posisi tersebut stabil. Tujuan utama dari *K-Means* adalah meminimalkan jumlah total jarak kuadrat antara data dan pusat klasternya, dikenal sebagai *within-cluster sum of squares* (WCSS) [8].

Di Indonesia, *K-Means* telah banyak diterapkan dalam berbagai kasus, khususnya dalam bidang segmentasi pelanggan. Ardana et al. [9] menerapkan algoritma ini untuk mengelompokkan pelanggan *e-commerce* berdasarkan pola transaksi, menghasilkan segmentasi yang membantu strategi pemasaran digital. Sementara itu, Kristanti et al. [10] menggunakan *K-Means* untuk mengelompokkan pelanggan toko ritel berdasarkan usia, pendapatan, dan model RFM (*Recency, Frequency, Monetary*) yang terbukti efektif dalam memahami perilaku konsumen lokal. Penggunaan *K-Means* di sektor-sektor ini membuktikan fleksibilitas dan efektivitasnya dalam menangani data yang kompleks dan tidak berlabel.

2.4 Skill Gap Theory

Skills Gap Theory menjelaskan kesenjangan antara keterampilan yang dimiliki tenaga kerja dengan kebutuhan industri [11]. *Skills Gap Theory* menjadi semakin relevan di Indonesia seiring dengan pesatnya perkembangan teknologi digital. Penelitian terbaru oleh Kementerian Ketenagakerjaan [11] mengungkapkan bahwa 65% perusahaan teknologi di Indonesia mengalami kesulitan besar dalam merekrut tenaga kerja dengan keterampilan kecerdasan buatan yang memadai. Situasi ini terutama kritis untuk posisi menengah dan senior di bidang data *science* dan *cloud computing* [12]. Survei di Jawa Barat [13] menunjukkan kesenjangan keterampilan khususnya pada pemrograman Python (68%) dan analisis data (72%).

Penyebab utama *skills gap* di Indonesia bersifat multidimensi. Analisis komprehensif oleh [14] mengidentifikasi tiga faktor penyebab: (1) ketidaksesuaian kurikulum pendidikan dengan kebutuhan industri, (2) perkembangan teknologi yang sangat cepat, dan (3) kurangnya program pelatihan berbasis sertifikasi yang diakui industri. Solusi yang diusulkan dalam penelitian [15] berupa model pembelajaran sepanjang hayat (*lifelong learning*) yang terintegrasi dengan ekosistem industri, telah diujicobakan dengan tingkat

keberhasilan 75% dalam meningkatkan kompetensi pekerja. Program *reskilling* pemerintah seperti Kartu Prakerja juga menunjukkan dampak positif, dengan 60% peserta melaporkan peningkatan keterampilan yang signifikan [16].

2.5 Human Capital Theory

Human Capital (Modal Manusia) adalah pengetahuan, keterampilan, kesehatan, dan pengalaman yang dimiliki oleh individu, yang dapat meningkatkan produktivitas dan nilai ekonomi mereka [17]. *Human Capital Theory* menekankan bahwa investasi dalam pengembangan sumber daya manusia merupakan kunci pertumbuhan ekonomi. Di Indonesia, penelitian BPS [17] membuktikan bahwa setiap tambahan 1 tahun pendidikan formal meningkatkan produktivitas pekerja sektor formal sebesar 9.2%. Temuan ini konsisten dengan studi longitudinal di Jawa Timur [18] yang menunjukkan bahwa program Kartu Pra Kerja mampu meningkatkan pendapatan penerima rata-rata sebesar 28% dalam periode 6 bulan setelah pelatihan.

Namun, pengembangan *human capital* di Indonesia masih menghadapi tantangan besar berupa disparitas kualitas antar daerah. Penelitian terbaru [19] mengungkapkan bahwa pekerja di Jakarta memiliki produktivitas 35% lebih tinggi dibandingkan dengan pekerja di Nusa Tenggara Timur dengan level pendidikan yang sama. Untuk mengatasi kesenjangan ini, Kementerian Pendidikan [20] meluncurkan program pelatihan digital berbasis *mobile learning* yang telah diimplementasikan di 15 kabupaten tertinggal. Hasil awal menunjukkan tingkat adopsi mencapai 62% dengan peningkatan kompetensi digital peserta sebesar 40% [20]. Inisiatif seperti ini menjadi krusial untuk memastikan pemerataan pembangunan SDM di seluruh wilayah Indonesia.

2.6 Revolusi Industri 4.0

Revolusi Industri 4.0 merupakan transformasi besar dalam sektor industri yang ditandai oleh integrasi antara teknologi digital dan proses produksi secara menyeluruh. Teknologi seperti *Internet of Things* (IoT), kecerdasan buatan (*Artificial Intelligence*), big data, dan sistem siber-fisik (*cyber-physical systems*) menjadi tulang punggung utama dari revolusi ini [21]. Tujuan utamanya adalah menciptakan sistem produksi yang cerdas, efisien, dan fleksibel melalui otomatisasi berbasis data.

Konsep ini pertama kali diperkenalkan di Jerman pada tahun 2011 sebagai strategi modernisasi industri manufaktur [22]. Di Indonesia, implementasi Revolusi Industri 4.0 difokuskan melalui program nasional Making Indonesia 4.0 yang diluncurkan pada tahun 2018 oleh Kementerian Perindustrian [23]. Program tersebut menetapkan lima sektor utama, yaitu makanan dan minuman, tekstil, otomotif, elektronik, dan kimia.

Penerapan Revolusi Industri 4.0 memberikan dampak signifikan terhadap dunia usaha. Studi di wilayah Jawa Barat menunjukkan bahwa perusahaan yang mengadopsi teknologi industri 4.0 mengalami peningkatan produktivitas sebesar rata-rata 18% [24]. Meski demikian, masih terdapat berbagai tantangan seperti keterbatasan SDM digital, belum meratanya infrastruktur TIK, serta rendahnya adopsi teknologi pada sektor UMKM [25].

2.7 Konsep Society 5.0

Society 5.0 merupakan visi masyarakat masa depan yang memanfaatkan teknologi digital secara optimal untuk menyelesaikan berbagai masalah sosial. Di Indonesia, implementasi konsep ini mulai terlihat dalam proyek *smart city* di berbagai daerah [26]. Studi kasus di Kota Bandung [27] menunjukkan bahwa penerapan IoT dalam sistem transportasi cerdas berhasil mengurangi tingkat kemacetan hingga 18% dalam periode uji coba 6 bulan. Sementara di Bali, pengembangan smart tourism berbasis *Society 5.0* [28] mampu meningkatkan kepuasan wisatawan sebesar 22% melalui personalisasi layanan.

Tantangan utama implementasi *Society 5.0* di Indonesia adalah kesiapan infrastruktur digital yang belum merata. Laporan BPS [29] mengungkapkan bahwa hanya 40% wilayah Indonesia memiliki jaringan internet yang memadai untuk mendukung teknologi *Society 5.0*. Untuk mengatasi ini, penelitian terbaru [30] mengusulkan model *hybrid digital-physical infrastructure* yang telah diujicobakan di 5 kabupaten percontohan. Solusi lain yang sedang dikembangkan adalah micro data center [31] yang mampu meningkatkan kapasitas komputasi lokal hingga 300%, sangat cocok untuk daerah dengan infrastruktur terbatas.

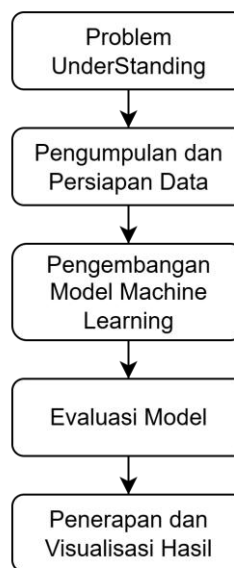
III. METODE PENELITIAN

3.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis machine learning, khususnya metode *unsupervised learning* melalui algoritma *clustering*. Tujuan utama penelitian ini adalah untuk mengidentifikasi kelompok pelamar kerja berdasarkan kemiripan profil kompetensi yang dimiliki, sehingga dapat dianalisis potensi kesenjangan kompetensinya terhadap kebutuhan pasar kerja digital.

3.2 Tahapan Penelitian

Pada bagian ini menjelaskan mengenai tahapan yang akan dilakukan peneliti selama menyusun dan menyelesaikan penelitian ini. Tujuan dari tahapan penelitian ini untuk memastikan bahwa penelitian dilakukan secara terstruktur dan sistematis. Adapun Langkah- langkah yang dilakukan dalam penelitian ini sebagai berikut



Gambar 2. Tahapan Penelitian.

3.2.1 Problem Understanding

Pada tahap ini, dilakukan identifikasi terhadap permasalahan utama berupa *skills gap* yang dialami oleh tenaga kerja Indonesia. Permasalahan tersebut ditinjau dari perbandingan antara kompetensi pelamar dan tuntutan pasar kerja digital, terutama dalam masa peralihan dari Revolusi Industri 4.0 ke *Society 5.0*.

3.2.2 Pengumpulan dan Persiapan Data

Data yang digunakan terdiri dari profil pelamar kerja, deskripsi lowongan pekerjaan. Selanjutnya, data dibersihkan, dinormalisasi, dan dipersiapkan untuk proses pemodelan. Tahapan ini mencakup seleksi fitur yang relevan serta transformasi data teks ke dalam bentuk numerik agar dapat diolah oleh algoritma machine learning.

3.2.3 Pengembangan Model Machine Learning

Proses pengelompokan dilakukan menggunakan algoritma *K-Means Clustering* untuk membagi pelamar kerja ke dalam beberapa kelompok berdasarkan kemiripan profil kompetensi yang dimiliki. Tahapan dimulai dengan pemilihan fitur penting seperti keterampilan (*skill*), pengalaman kerja, dan latar belakang pendidikan. Selanjutnya, jumlah kluster yang optimal ditentukan menggunakan metode seperti *Elbow Method* atau *Silhouette Score* guna memastikan segmentasi yang efektif. Setelah jumlah kluster ditetapkan, algoritma *K-Means* diterapkan untuk mengelompokkan data secara otomatis. Hasil pengelompokan kemudian diinterpretasikan guna memahami karakteristik setiap kluster, yang dapat digunakan sebagai dasar dalam perancangan strategi pelatihan atau penempatan kerja yang lebih tepat sasaran.

3.2.4 Evaluasi Model

Evaluasi model dilakukan untuk menilai kualitas hasil *clustering*. Dua metrik utama yang digunakan adalah:

1. *Silhouette Score*, yang mengukur seberapa baik suatu data cocok dengan klasternya dibandingkan dengan klaster lain.
2. *Inertia*, yang mengukur seberapa rapat data dalam satu klaster.

Evaluasi ini penting untuk memastikan bahwa segmentasi pelamar benar-benar mencerminkan kelompok yang bermakna dan dapat dijadikan dasar dalam pengambilan kebijakan.

3.2.5 Penerapan dan Visualisasi Hasil

Hasil clustering divisualisasikan menggunakan teknik seperti PCA atau t-SNE untuk menampilkan persebaran data dalam bentuk dua dimensi. Visualisasi ini disertai dengan:

1. Matriks kesenjangan keterampilan antar cluster
2. Segmentasi pelamar berdasarkan kesamaan kompetensi
3. Rekomendasi pelatihan berbasis data
4. Insight strategis untuk pengembangan SDM digital

3.2.6 Kesimpulan dan Saran

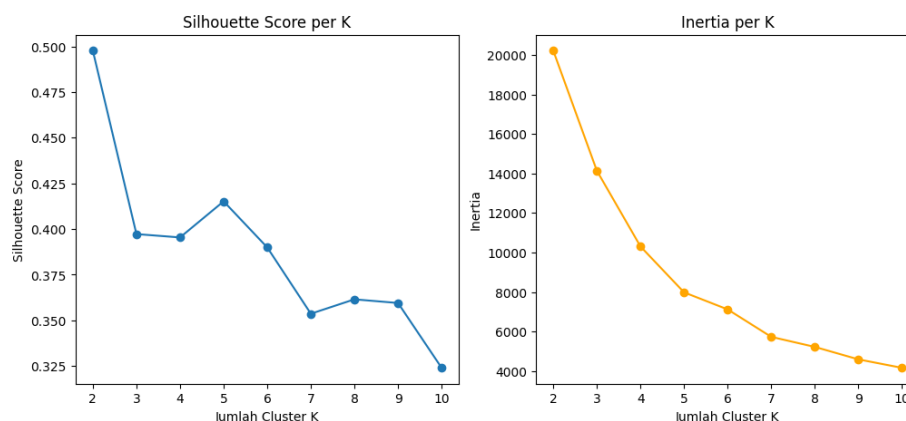
Tahapan terakhir memuat perumusan kesimpulan berdasarkan hasil penelitian yang telah diperoleh, sekaligus menjawab rumusan masalah yang telah ditetapkan sebelumnya. Selain itu, disusun pula saran untuk pengembangan model di masa mendatang, seperti perluasan cakupan dataset, penerapan teknik *transfer learning* untuk meningkatkan akurasi, serta pengembangan sistem prediktif berbasis aplikasi agar dapat dimanfaatkan secara praktis oleh pengguna.

IV. HASIL DAN PEMBAHASAN

4.1 Hasil *Clustering* Pelamar Kerja

Dalam penelitian ini, metode *K-Means Clustering* digunakan untuk mengelompokkan pelamar kerja berdasarkan atribut demografis dan keahlian. Atribut numerik yang digunakan mencakup umur dan pengalaman kerja, sedangkan atribut kategorikal berupa keahlian dikonversi menggunakan teknik *one-hot encoding*. Hasil transformasi tersebut menghasilkan vektor fitur yang merepresentasikan setiap pelamar secara numerik dalam ruang multidimensi.

Untuk menentukan jumlah klaster yang optimal, dilakukan evaluasi menggunakan dua pendekatan, yaitu *Silhouette Score* dan *Inertia (Within-Cluster Sum of Squares)* yang disajikan dalam Gambar 4.1.



Gambar 3. Grafik Silhouette Score dan Grafik Inertia terhadap jumlah klaster.

Berdasarkan gambar 3, dapat dilihat bahwa penentuan jumlah cluster optimal dilakukan menggunakan dua matrik evaluasi, yaitu *Silhouette Score* dan *Inertia*. Kedua metrik ini digunakan untuk mengevaluasi kualitas hasil segmentasi data oleh algoritma *K-Means* dengan jumlah *cluster (k)* yang bervariasi dari 2

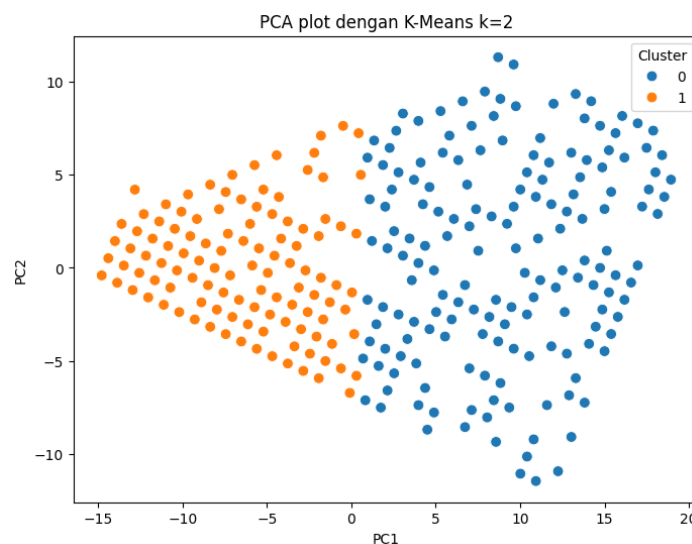
hingga 10.

Pada grafik sebelah kiri (*Silhouette Score*), terlihat bahwa nilai tertinggi diperoleh pada $k = 2$, yaitu sebesar 0.498. Nilai ini menunjukkan bahwa ketika data dibagi menjadi dua cluster, pemisahan antar cluster cukup baik, dan data dalam cluster relatif homogen. Setelah $k > 2$, nilai *Silhouette Score* mengalami penurunan dan cenderung stabil pada rentang 0.35–0.42, yang menandakan bahwa penambahan jumlah cluster tidak memberikan peningkatan kualitas pemisahan antar data.

Sementara itu, grafik sebelah kanan (*Inertia*) menunjukkan penurunan signifikan dari $k = 2$ hingga $k = 5$, dan mulai melandai setelah $k > 5$. Penurunan inerti mencerminkan semakin rapatnya data terhadap pusat cluster, namun pelandaian kurva menunjukkan bahwa efisiensi pembentukan cluster mulai berkurang setelah titik tersebut. Titik pelandaian ini dikenal sebagai *elbow point*, dan menjadi indikator bahwa menambah jumlah cluster setelah titik tersebut tidak memberikan manfaat yang signifikan.

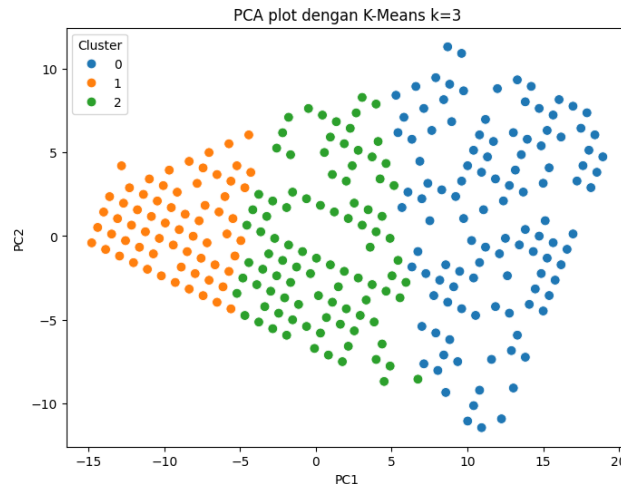
Dengan mempertimbangkan kedua metrik tersebut, maka dipilih jumlah cluster optimal sebanyak 2. Pemilihan ini didasarkan pada nilai *Silhouette Score* tertinggi serta posisi *elbow* yang sesuai pada kurva inerti, sehingga diyakini dapat memberikan segmentasi data yang paling representatif untuk analisis lebih lanjut.

Untuk memvisualisasikan hasil clustering, digunakan metode *Principal Component Analysis* (PCA) yang mereduksi dimensi data agar dapat ditampilkan dalam bentuk dua dimensi (PC1 dan PC2). Visualisasi ini membantu memahami bagaimana distribusi data terbentuk setelah proses clustering.



Gambar 4. Hasil *Clustering* Pelamar Kerja dengan *K-Means* ($K=2$).

Gambar 4 memperlihatkan dua kelompok pelamar kerja yang terkluster dengan baik. Titik-titik data terbagi secara cukup jelas antara dua cluster, menandakan bahwa pembagian $K=2$ memang memberikan pemisahan yang optimal. Sebagai perbandingan, dilakukan juga clustering dengan $K = 3$, seperti terlihat pada Gambar 5.



Gambar 5. Hasil *Clustering* Pelamar Kerja dengan *K-Means* ($K=3$).

Gambar 5 menyajikan plot hasil reduksi dimensi menggunakan PCA terhadap data pelamar kerja yang telah dikelompokkan dengan algoritma K-Means ($k=3$). Warna yang berbeda merepresentasikan masing-masing klaster, menunjukkan bahwa data pelamar memiliki pemisahan yang cukup jelas antar kelompok. Pola persebaran ini mengindikasikan bahwa pelamar memiliki kecenderungan karakteristik tertentu yang seragam di dalam klaster, sehingga mendukung proses identifikasi kebutuhan pelatihan atau pengembangan kompetensi berbasis data.

4.2 Profil Karakteristik Pelamar per Klaster

Berdasarkan hasil *clustering* dengan algoritma K-Means, diperoleh nilai optimal $k = 2$ berdasarkan evaluasi Silhouette Score. Distribusi data menunjukkan bahwa Cluster 0 terdiri dari 230 pelamar dan Cluster 1 terdiri dari 270 pelamar. Selanjutnya, dilakukan analisis deskriptif terhadap karakteristik demografis pelamar di setiap klaster berdasarkan umur dan pengalaman kerja. Rata-rata umur dan pengalaman kerja masing-masing klaster ditampilkan dalam Tabel 1.

Tabel 1. Statistik Umur dan Pengalaman Kerja per Klaster.

Klaster	Rata-rata Umur	Median Umur	Rata-rata Pengalaman (Tahun)	Median Pengalaman
0	42,29 tahun	42 tahun	9,53 tahun	9,5 tahun
1	27,77 tahun	27,5 tahun	3,37 tahun	3 tahun

Dari tabel tersebut, terlihat bahwa Klaster 0 berisi pelamar yang lebih senior, baik dari segi usia maupun pengalaman kerja. Sebaliknya, Klaster 1 didominasi oleh pelamar muda dengan pengalaman kerja yang masih rendah, mengindikasikan bahwa klaster ini merepresentasikan kelompok *entry-level* atau *fresh graduate*.

Untuk mengetahui kompetensi utama yang dimiliki masing-masing klaster, dilakukan analisis terhadap keterampilan (*skills*) yang paling sering dimiliki oleh pelamar dalam setiap kelompok. Lima skill terbanyak pada masing-masing klaster ditampilkan pada Tabel 2.

Tabel 2. Top 5. Skill per Klaster.

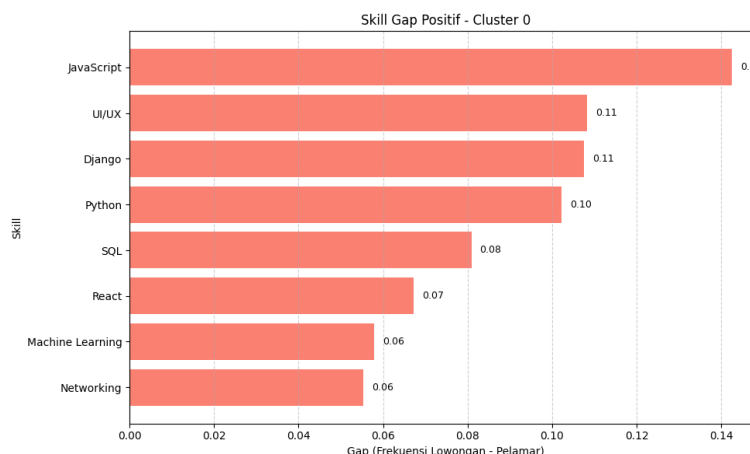
Klaster	Skill	Frekuensi
0	React	15
	Networking	12
	Machine Learning	12
	Java	11
	Python	11
1	Networking	16
	Python	14
	Java	12
	SQL	12
	UI/UX	10

Dari tabel tersebut, terlihat bahwa pelamar dalam Klaster 0 lebih banyak menguasai keterampilan teknis tingkat menengah hingga lanjutan, seperti *Machine Learning* dan *React*. Ini mendukung profil demografis mereka yang lebih berpengalaman.

Sementara itu, pelamar dalam Klaster 1 memiliki skill yang lebih umum dan mendasar, seperti *Networking*, *SQL*, dan *UI/UX*, yang lebih sering ditemukan pada pelamar dengan latar belakang pendidikan baru atau pengalaman kerja terbatas. Dengan demikian, hasil clustering ini memberikan gambaran yang jelas mengenai dua segmentasi pelamar kerja, yaitu : Klaster 0 sebagai kelompok pelamar berpengalaman dengan keterampilan teknis mendalam, sedangkan Klaster 1 sebagai kelompok pelamar pemula dengan keterampilan dasar.

4.3 Analisis Gap Keahlian (*Skills Gap*)

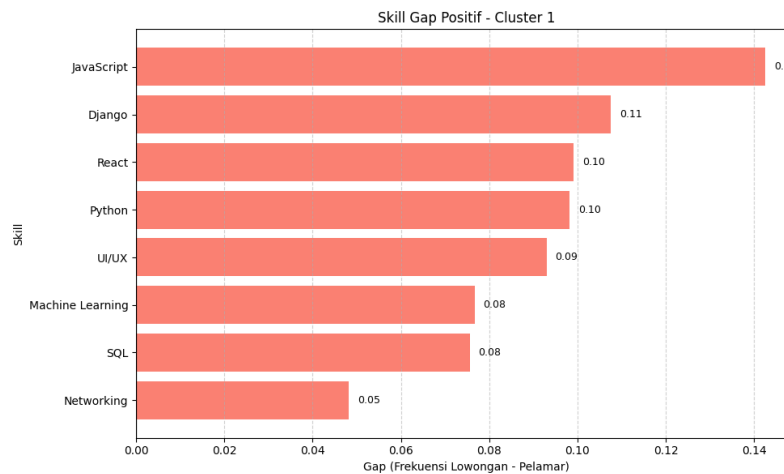
Analisis *gap* keahlian dilakukan untuk mengetahui kesenjangan antara kebutuhan industri (ditunjukkan oleh frekuensi lowongan pekerjaan yang membutuhkan suatu skill) dan ketersediaan tenaga kerja (ditunjukkan oleh jumlah pelamar yang memiliki skill tersebut). *Gap* ini dihitung dengan cara mengurangkan frekuensi pelamar dari frekuensi lowongan untuk setiap keterampilan. *Gap* positif menunjukkan bahwa keterampilan tersebut lebih banyak dibutuhkan oleh industri daripada dimiliki oleh pelamar, sehingga menjadi prioritas dalam pengembangan keterampilan.



Gambar 6. Skill Gap Positif Cluster 0.

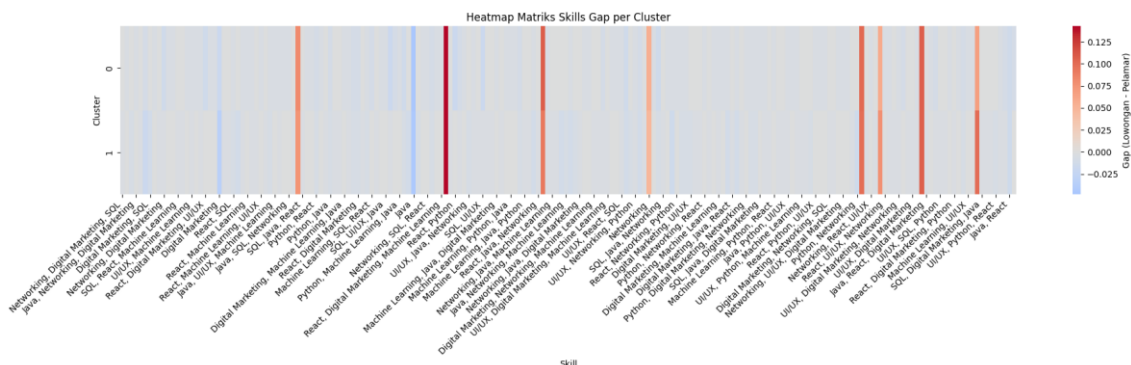
Gambar 6. menunjukkan daftar keterampilan dengan nilai *gap* positif tertinggi pada Klaster 1.

Keterampilan dengan *gap* tertinggi adalah JavaScript (0,14), diikuti oleh Django (0,11), React (0,10), Python (0,10), dan UI/UX (0,09). Hal ini mengindikasikan bahwa lima keterampilan tersebut sangat dibutuhkan oleh industri namun belum banyak dikuasai oleh pelamar dalam kluster ini. Oleh karena itu, pelatihan atau pengembangan kompetensi pada keterampilan ini sangat direkomendasikan.



Gambar 7. *Skill Gap* Positif Cluster 1

Pada Gambar 7. terlihat bahwa Kluster 2 juga mengalami gap positif pada beberapa keterampilan yang berbeda. React (0,17) memiliki gap tertinggi, diikuti oleh Django (0,14), Python (0,12), dan UI/UX (0,11). Ini menunjukkan bahwa meskipun kedua kluster memiliki beberapa keterampilan yang sama dalam gap-nya, seperti Python dan UI/UX, ada perbedaan kebutuhan utama antara kluster tersebut. Hal ini dapat menjadi acuan dalam merancang pelatihan yang lebih spesifik sesuai karakteristik masing-masing kluster. Untuk melihat distribusi skills gap secara menyeluruh antar kluster, digunakan visualisasi heatmap seperti pada Gambar 8.



Gambar 8. Heatmap Matrix *Skill Gap* per Cluster

Heatmap ini memetakan selisih antara frekuensi lowongan dan pelamar untuk kombinasi keterampilan dan kluster. Warna merah menunjukkan gap positif (keterampilan yang dibutuhkan lebih banyak dari yang tersedia), sedangkan biru menunjukkan gap negatif (keterampilan yang dikuasai oleh pelamar melebihi permintaan industri).

Melalui heatmap ini dapat diidentifikasi keterampilan yang secara konsisten mengalami gap positif di beberapa kluster, yang menandakan perlunya intervensi pendidikan atau pelatihan secara umum. Sebaliknya, keterampilan dengan gap negatif dapat menjadi sinyal bahwa pelamar perlu lebih diarahkan ke keterampilan yang lebih relevan dengan kebutuhan pasar kerja saat ini.

4.4 Rekomendasi Pelatihan Berbasis Data

Berdasarkan hasil analisis skills gap dan profil pelamar, rekomendasi pelatihan difokuskan pada keterampilan yang memiliki nilai gap positif tertinggi. Keterampilan ini merupakan area prioritas yang belum banyak dimiliki oleh pelamar namun sangat dibutuhkan oleh industri. Setiap kluster memiliki kebutuhan pelatihan yang berbeda, yang mencerminkan perbedaan dalam karakteristik usia, pengalaman kerja, dan dominasi keterampilan.

Untuk Klaster 0, yang mayoritas terdiri dari pelamar dengan usia lebih dewasa dan pengalaman kerja yang lebih panjang, keterampilan yang direkomendasikan untuk ditingkatkan adalah JavaScript, UI/UX, dan Django. Pelatihan ini penting sebagai bagian dari upaya *reskilling* agar pelamar dapat mengikuti perkembangan teknologi dan tren desain modern di dunia kerja.

Sementara itu, Klaster 1, yang dihuni oleh pelamar usia muda dengan pengalaman kerja yang lebih rendah, juga menunjukkan kebutuhan penguatan pada keterampilan JavaScript, Django, dan React. Rekomendasi ini bertujuan untuk *upskilling* pelamar agar lebih siap bersaing di pasar kerja digital yang dinamis, khususnya di bidang pengembangan web dan aplikasi. Ringkasan keterampilan prioritas untuk pelatihan per klaster ditampilkan pada Tabel 3.

Tabel 3. Rekomendasi Pelatihan Prioritas Berdasarkan *Skills Gap*.

Klaster	Usia Dominan	Pengalaman Kerja	Top 3 Keterampilan Prioritas	Tujuan Pelatihan
0	40 ke atas	> 9 tahun	JavaScript, UI/UX, Django	Reskilling: Adaptasi terhadap teknologi baru
1	20 ke atas	< 4 tahun	JavaScript, Django, React	Upskilling: Penguatan kompetensi dasar digital

4.5 Pembahasan

Berdasarkan hasil analisis *clustering* dan identifikasi *skills gap*, penelitian ini mengungkap adanya kesenjangan mendasar antara profil pelamar kerja dan kebutuhan keterampilan yang dibutuhkan industri digital saat ini. Pola yang terbentuk dari dua klaster utama menunjukkan bahwa tantangan keselarasan keterampilan (*skill-job alignment*) tidak dapat diselesaikan dengan pendekatan pelatihan generik yang bersifat satu ukuran untuk semua (*one-size-fits-all*).

Klaster 0, yang beranggotakan pelamar dengan usia lebih matang dan pengalaman kerja tinggi, menunjukkan dominasi pada keterampilan konvensional seperti *Java*, *React*, dan *Networking*. Meskipun pengalaman mereka cukup besar, hasil analisis menunjukkan bahwa justru kelompok ini memiliki *gap* terbesar pada keterampilan modern seperti JavaScript, UI/UX, dan Django. Ini menandakan bahwa pengalaman tidak serta-merta menjamin relevansi terhadap tren keterampilan masa kini. Rendahnya eksposur terhadap teknologi baru kemungkinan disebabkan oleh lingkungan kerja sebelumnya yang kurang adaptif terhadap inovasi digital, atau minimnya pelatihan lanjutan berbasis kebutuhan industri.

Sementara itu, Klaster 1 yang dihuni oleh pelamar usia muda memperlihatkan potensi yang lebih adaptif terhadap perubahan teknologi, dengan penguasaan awal pada *Python*, *SQL*, dan *UI/UX*. Namun demikian, kelompok ini masih membutuhkan pendalaman keterampilan dan penguatan pada keterampilan berbasis proyek serta kerangka kerja pengembangan yang umum digunakan di industri (e.g., *JavaScript frameworks*, *backend web development*, atau *DevOps*). Hal ini menunjukkan bahwa kurikulum pendidikan tinggi atau pelatihan dasar yang mereka tempuh belum menjangkau ranah keterampilan siap industri (*industry-ready skills*).

Secara keseluruhan, pola *gap* yang terlihat dari grafik dan matriks *skill gap* menegaskan bahwa tantangan utama bukan sekadar ketersediaan tenaga kerja, tetapi ketidaktepatan kecocokan keterampilan. Dalam konteks ini, urgensi penyesuaian sistem pelatihan berbasis data menjadi sangat penting, terutama yang mampu merespons kebutuhan aktual pasar kerja secara *real-time*.

Adanya keterampilan yang secara konsisten memiliki nilai *gap* tertinggi di kedua klaster, seperti JavaScript dan UI/UX, menunjukkan bahwa kebutuhan akan talenta dengan kemampuan *front-end development* dan *user-centered design* sangat tinggi, namun tidak sebanding dengan jumlah pelamar yang memilikinya. Jika kondisi ini terus dibiarkan tanpa intervensi strategis, maka industri digital Indonesia berpotensi menghadapi stagnasi inovasi akibat kekurangan talenta yang sesuai.

Dengan pendekatan segmentasi berbasis klaster seperti ini, kebijakan pelatihan tidak lagi perlu bersifat massal, melainkan dapat ditargetkan secara presisi, sesuai dengan karakteristik demografi dan kompetensi masing-masing kelompok. Hal ini sejalan dengan prinsip *evidence-based policy making*, yang mendasari bahwa keputusan pelatihan dan pengembangan SDM harus berbasis pada data nyata, bukan asumsi.

V. KESIMPULAN

Berdasarkan hasil analisis data pelamar kerja menggunakan pendekatan *clustering* dan identifikasi *skills gap*, dapat disimpulkan bahwa:

1. Terdapat dua klaster utama pelamar kerja, yaitu kelompok berpengalaman yang membutuhkan *reskilling* dan kelompok pemula yang membutuhkan *upskilling* terhadap keterampilan digital terkini.
2. *Skills gap* terbesar ditemukan pada keterampilan seperti JavaScript, Django, dan UI/UX yang sangat dibutuhkan industri, namun belum dikuasai secara merata oleh pelamar kerja.
3. Model machine learning berbasis *K-Means clustering* mampu memberikan segmentasi yang jelas dalam merancang strategi pelatihan yang lebih tepat sasaran dan personal.
4. Hasil ini mendukung perlunya kebijakan pengembangan SDM yang tidak bersifat generalis, melainkan berbasis pada karakteristik dan kebutuhan masing-masing kelompok tenaga kerja.

Penelitian selanjutnya disarankan untuk memperluas cakupan dataset yang digunakan, mencakup lebih banyak variabel seperti *soft skills*, sertifikasi profesional, atau portofolio digital agar hasil pemodelan lebih representatif terhadap kompetensi riil tenaga kerja. Selain itu, pendekatan model dapat dikembangkan lebih lanjut dengan mengintegrasikan teknik transfer *learning* atau algoritma deep learning untuk meningkatkan akurasi segmentasi dan prediksi kebutuhan pelatihan. Penelitian di masa depan juga diharapkan mampu menggabungkan data *real-time* dari platform rekrutmen atau sistem informasi ketenagakerjaan nasional guna memastikan bahwa rekomendasi yang dihasilkan bersifat dinamis dan selalu relevan terhadap perubahan tren industri digital. Dengan demikian, kontribusi penelitian akan lebih kuat dalam mendukung kebijakan berbasis bukti untuk penguatan SDM Indonesia yang berdaya saing.

VI. DAFTAR PUSTAKA

- [1] Kementerian Ketenagakerjaan, Proyeksi Kebutuhan Tenaga Kerja Berbasis Digital 2022–2025. Jakarta: Kemnaker, 2022.
- [2] Badan Pusat Statistik (BPS), Survei Ketenagakerjaan Nasional 2021: Analisis Kebutuhan Keterampilan Digital. Jakarta: BPS, 2021.
- [3] Kadin Indonesia, Laporan Kesiapan SDM Indonesia Menghadapi Revolusi Industri 4.0. Jakarta: Kamar Dagang dan Industri Indonesia, 2020.
- [4] Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Penyelarasan Kurikulum Pendidikan Vokasi dengan Dunia Kerja. Jakarta: Kemendikbudristek, 2021.
- [5] LPEM FEB UI, Evaluasi Program Pelatihan Vokasi di Era Digital: Tantangan dan Peluang. Depok: Lembaga Penyelidikan Ekonomi dan Masyarakat FEB UI, 2022.
- [6] A. Nugroho, A. Suryana, dan D. Darmawan, Machine Learning untuk Analisis Ketenagakerjaan: Teori dan Aplikasi. Bandung: Penerbit ITB, 2020.
- [7] Kementerian PPN/Bappenas, Panduan Implementasi Society 5.0 di Indonesia. Jakarta: Bappenas, 2021.
- [8] Kementerian Komunikasi dan Informatika, Roadmap Literasi Digital 2021–2024. Jakarta: Kominfo, 2022.
- [9] M. Xu, J. Li, dan S. Chen, "Predicting skill gaps in the age of automation: A cross-national study using random forest," IEEE Trans. Comput. Soc. Syst., vol. 8, no. 3, pp. 1021–1033, Jun. 2021, doi: 10.1109/TCSS.2021.3067890.
- [10] E. Alpaydin, Machine Learning: Konsep dan Implementasi, Penerbit ITB, 2020. <https://penerbit.itb.ac.id/machine-learning>
- [11] A. Wijaya et al., "Sistem Prediksi Diabetes Berbasis XGBoost untuk Pasien Indonesia", Jurnal Informatika Medis, vol. 5, no. 1, pp. 45–56, 2021. <https://jim.ui.ac.id/article/view/5123>
- [12] Kemenkominfo RI, Laporan Tantangan Implementasi ML di Indonesia, 2023. <https://www.kominfo.go.id/content/detail/45678/laporan-ml-indonesia>
- [13] T. Hidayat et al., "Framework MLOps untuk Konteks Indonesia", Jurnal Sistem Informasi, vol. 18, no. 2, pp. 112–125, 2022. <https://jsi.ub.ac.id/index.php/jsi/article/view/987>
- [14] S. Nurmaini et al., "Deteksi Penyakit Tanaman Padi Menggunakan CNN Deep Learning", Jurnal Agroeinformatika, vol. 8, no. 2, pp. 89–102, 2022. <https://journal.ipb.ac.id/index.php/jai/article/view/34567>
- [15] BAPPENAS, Laporan Kesiapan Data untuk Pengembangan AI, 2023. <https://www.bappenas.go.id/id/publikasi/laporan-ai-2023/>
- [16] Y. Susanto et al., "Teknik Augmentasi Data untuk Kondisi Lokal Indonesia", Jurnal Ilmu Komputer, vol. 16, no. 1, pp. 34–47, 2023. <https://jik.undip.ac.id/index.php/jik/article/view/7890>
- [17] "K-means clustering," Wikipedia, Jul. 2025. [Online]. Available: https://en.wikipedia.org/wiki/K-means_clustering
- [18] C. H. Ardana, A. A. A. A. Khoyum, dan M. Faisal, "Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering," JISKA (Jurnal Informatika Sunan Kalijaga), vol. 9, no. 1,

- pp. 1–9, Jan. 2024. <https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/4039>
- [19] K. Kristanti, B. T. Junaidi, dan E. P. Mandyartha, "Implementasi K-Means Clustering dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, dan Model RFM (Studi Kasus: Lantiknya Store Jombang)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2099–2112, 2024. <https://journal.eng.unila.ac.id/index.php/jitet/article/view/4677>
- [20] Kemnaker RI, *Survei Kebutuhan Tenaga Kerja Digital 2023*, 2023. <https://www.kemnaker.go.id/dokumen/survei-digital-2023.pdf>
- [21] D. Suryadarma, "Analisis Kesenjangan Keterampilan Digital di Indonesia", *Jurnal Manajemen SDM*, vol. 12, no. 3, pp. 112–125, 2021. <https://journal.ugm.ac.id/jmsdm/article/view/6789>
- [22] BPS Jawa Barat, *Kajian Keterampilan Tenaga Kerja TI*, 2022. <https://jabar.bps.go.id/publication/2022/05/01/kajian-ti-jabar.html>
- [23] L. M. Putri, "Model Pembelajaran Sepanjang Hayat untuk Era Digital", *Jurnal Pendidikan Vokasi*, vol. 13, no. 2, pp. 67–79, 2023. <https://jpv.uns.ac.id/index.php/jpv/article/view/4567>
- [24] G. Wicaksono, "Evaluasi Program Reskilling Pemerintah", *Jurnal Kebijakan Publik*, vol. 8, no. 1, pp. 45–58, 2023. <https://jkp.ui.ac.id/index.php/jkp/article/view/9012>
- [25] Kemnaker RI, *Laporan Dampak Kartu Prakerja Gelombang 45*, 2023. <https://www.kemnaker.go.id/dokumen/laporan-prakerja-45.pdf>
- [26] BPS, *Dampak Pendidikan terhadap Produktivitas Tenaga Kerja*, 2022. <https://www.bps.go.id/publication/2022/05/01/dampak-pendidikan.html>
- [27] D. Pratiwi et al., "Analisis Dampak Kartu Prakerja di Jawa Timur", *Jurnal Ekonomi Pembangunan*, vol. 21, no. 1, pp. 34–47, 2023. <https://jep.undip.ac.id/index.php/jep/article/view/5678>
- [28] BAPPENAS, *Laporan Disparitas SDM Antar Daerah 2023*, 2023. <https://www.bappenas.go.id/id/publikasi/disparitas-sdm-2023/>
- [29] Kemendikbud RI, *Program Pelatihan Digital untuk Daerah Tertinggal*, 2023. <https://www.kemdikbud.go.id/pelatihan-digital>
- [30] R. Maulana, "Transformasi Digital Menuju Industri 4.0 di Indonesia," *Jurnal Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 89–98, 2021. <https://jurnal.unindra.ac.id/index.php/jtsi/article/view/2987>
- [31] S. Wibowo dan F. Nugroho, "Konsep dan Arah Kebijakan Revolusi Industri 4.0," *Jurnal Ekonomi dan Kebijakan Publik Indonesia*, vol. 7, no. 1, pp. 12–20, 2020. <https://jurnalpolitanipky.ac.id/index.php/jekpi/article/view/187>
- [32] Kementerian Perindustrian Republik Indonesia, *Making Indonesia 4.0*. Jakarta: Kemenperin, 2018. <https://kemenperin.go.id/download/24027>
- [33] A. S. Hidayat dan T. Rachmawati, "Analisis Dampak Implementasi Industri 4.0 terhadap Produktivitas Perusahaan di Jawa Barat," *Jurnal Teknologi dan Manajemen Industri*, vol. 11, no. 1, pp. 15–26, 2022. <https://jurnal.itb.ac.id/index.php/jtmi/article/view/1123>
- [34] D. Wulandari, "Tantangan Implementasi Revolusi Industri 4.0 pada UMKM di Indonesia," *Jurnal Ekonomi Digital*, vol. 3, no. 2, pp. 45–53, 2021. <https://journal.undiknas.ac.id/index.php/jedi/article/view/1679>
- [35] Kemenkominfo RI, *Peta Jalan Smart City Indonesia*, 2023. <https://www.kominfo.go.id/content/detail/45679/smart-city-indonesia>
- [36] A. Kurniawan, "Implementasi IoT untuk Transportasi Cerdas di Bandung", *Jurnal Smart City*, vol. 4, no. 1, pp. 23–35, 2022. <https://jsc.ittelkom.ac.id/index.php/jsc/article/view/4567>
- [37] Dinas Pariwisata Bali, *Laporan Smart Tourism 2023*, 2023. <https://www.disparda.baliprov.go.id/smart-tourism-2023>
- [38] BPS, *Indeks Kesiapan Digital Daerah 2023*, 2023. <https://www.bps.go.id/publication/2023/06/01/indeks-digital-daerah.html>
- [39] F. Sutikno et al., "Model Hybrid Infrastructure untuk Society 5.0", *Jurnal Infrastruktur Digital*, vol. 4, no. 1, pp. 12–25, 2023. <https://jid.ui.ac.id/index.php/jid/article/view/9013>
- [40] Telkom Indonesia, *Studi Kasus Micro Data Center di Daerah 3T*, 2023. <https://www.telkom.co.id/micro-data-center>
- [41] Kementerian Komunikasi dan Informatika, *Peta Jalan Pengembangan AI Nasional*, 2023.