

Penerapan Normalisasi Data pada Angkatan Kerja Indonesia Bulan Februari 2025 Berdasarkan Kelompok Umur

Anastasya Jesica Sidauruk¹, Juan Sebastian Sirait², Sardo Sipayung³

^{1,2,3}Universitas Katolik Santo Thomas, Medan

Email: ¹anastasyasidauruk1877@gmail.com, ²juansebastian190405@gmail.com, ³pinsarsiphom@gmail.com

ABSTRAK

Normalisasi data adalah langkah awal yang krusial dalam proses penambangan data yang bertujuan untuk mereduksi perbedaan skala pada atribut numerik, sehingga analisis dapat dilakukan dengan lebih objektif dan tepat. Studi ini bertujuan untuk melaksanakan dan mengevaluasi teknik normalisasi data pada angkatan kerja Indonesia pada bulan Februari 2025 berdasarkan kategori usia. Data yang dipakai adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) melalui Survei Angkatan Kerja Nasional (SAKERNAS), yang mencakup atribut numerik seperti jumlah penduduk yang bekerja, jumlah penganggur, jumlah angkatan kerja, serta persentase penduduk yang bekerja. Metode normalisasi yang digunakan dalam studi ini terdiri dari *Min-Max Normalization*, *Z-Score Normalization*, dan *Decimal Scaling Normalization*. Proses penelitian mencakup pengumpulan data, pemilihan data dari periode Februari 2025, pembersihan data, penerapan teknik normalisasi, dan analisis hasil normalisasi tersebut. Temuan penelitian menunjukkan bahwa ketiga metode normalisasi berhasil meratakan skala nilai antar atribut yang sebelumnya menunjukkan perbedaan yang signifikan dalam rentang nilainya. Normalisasi Min-Max efektif dalam mengonversi data ke dalam rentang tertentu, Z-Score dapat mengetahui penyimpangan dari nilai rata-rata, sedangkan Decimal Scaling mempermudah perbandingan proporsional antara kategori usia. Secara empiris, penelitian ini mengkonfirmasi bahwa kelompok usia 25-44 tahun adalah yang paling dominan dalam struktur angkatan kerja Indonesia pada Februari 2025. Implementasi normalisasi data terbukti meningkatkan kualitas data dan menyokong analisis ketenagakerjaan yang lebih akurat.

Kata Kunci: Normalisasi Data, Penambangan Data, Angkatan Kerja, BPS Indonesia

ABSTRACT

Data normalization is a crucial initial step in the data mining process, aiming to reduce scale differences in numerical attributes, allowing for more objective and accurate analysis. This study aims to implement and evaluate data normalization techniques on the Indonesian workforce in February 2025 based on age category. The data used is secondary data obtained from the Central Bureau of Statistics (BPS) thru the National Labor Force Survey (SAKERNAS), which includes numerical attributes such as the number of employed people, the number of unemployed, the size of the labor force, and the percentage of the working population. The normalization methods used in this study consist of Min-Max Normalization, Z-Score Normalization, and Decimal Scaling Normalization. The research process includes data collection, selection of data from the period February 2025, data cleaning, application of normalization techniques, and analysis of the normalization results. The research findings indicate that all three normalization methods successfully leveled the value scales across attributes that previously showed significant differences in their value ranges. Min-Max normalization is effective in converting data to a specific range, Z-Score can identify deviations from the mean value, while Decimal Scaling facilitates proportional comparisons between age categories. Empirically, this study confirms that the 25-44 age group will be the most dominant in the structure of the Indonesian workforce in February 2025. Implementing data normalization has proven to improve data quality and support more accurate labor analysis.

Keywords: Data Normalization, Data Mining, Labor Force, BPS Indonesia

Penulis Korespondensi:

Anastasya Jesica Sidauruk

Email: Universitas Katolik Santo Thomas

Article Info

Diterima: 19 Januari 2026

Direvisi: 21 Januari 2026

Disetujui: 22 Januari 2026

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Data mining merupakan kegiatan mencari informasi dan pola yang berguna dari kumpulan data yang sangat besar. Kegiatan *data mining* meliputi pengumpulan data, ekstraksi data, analisis data, serta statistik data. Proses ini juga sering disebut sebagai penemuan pengetahuan, ekstraksi pengetahuan, analisis data/pola, pengambilan informasi, dan variasi lainnya [1].

Metode normalisasi data merupakan langkah untuk menyamakan beberapa variabel sehingga rentang nilainya serupa, tanpa ada yang terlalu ekstrem, baik dari sisi besar maupun kecil, agar analisis statistik dapat dilakukan dengan lebih efisien. Beberapa pendekatan dalam normalisasi data mencakup *Min-Max Normalization*, *Z-score Normalization*, dan *Decimal Scaling Normalization* [2]. Di antara metode-metode tersebut, salah satu yang sering diterapkan adalah Metode *Decimal Scaling*, yang mengubah data dengan cara normalisasi untuk menyelaraskan rentang nilai dari setiap atribut menggunakan skala tertentu, dengan cara memindahkan posisi desimal pada nilai-nilai data sesuai yang diinginkan [2]. Hal ini bertujuan untuk menciptakan data dengan kualitas yang lebih baik [3]. Penelitian terdahulu menunjukkan bahwa perbedaan dalam rentang nilai di setiap atribut selama proses transformasi data dapat menyebabkan atribut dengan nilai yang jauh lebih kecil tidak berfungsi dengan baik dibandingkan atribut lain. Oleh karena itu, data yang dinormalisasi menggunakan metode *Decimal Scaling* menunjukkan hasil yang secara signifikan lebih baik dibandingkan dengan hasil klasifikasi tanpa melakukan transformasi data [4].

Data angkatan kerja di Indonesia dikumpulkan oleh Badan Pusat Statistik (BPS) melalui sejumlah kegiatan statistik, termasuk survei. Survei yang khusus dibuat untuk mendapatkan informasi mengenai ketenagakerjaan adalah Survei Angkatan Kerja Nasional (SAKERNAS). BPS melaksanakan SAKERNAS secara terjadwal dua kali setahun, yaitu di bulan Februari dan Agustus, dengan jangkauan yang mencakup seluruh wilayah Indonesia. Dari hasil survei ini diperoleh data penting mengenai situasi ketenagakerjaan yang kemudian dipublikasikan secara resmi oleh BPS. Tujuan publikasi data tenaga kerja adalah untuk memberikan gambaran mengenai kondisi serta karakteristik tenaga kerja, termasuk jumlah angkatan kerja, individu yang bekerja, dan mereka yang menganggur. Selain itu, data tersebut juga menyajikan informasi tentang pola distribusi angkatan kerja berdasarkan karakteristik tertentu, salah satunya adalah usia. Informasi ini krusial untuk memahami struktur ketenagakerjaan pada periode waktu tertentu, termasuk bulan Februari 2025. Dalam pelaksanaan SAKERNAS, pertanyaan mengenai ketenagakerjaan difokuskan pada populasi yang berada dalam usia kerja, yakni individu yang berusia 15 tahun ke atas. Pada tahap pemrosesan data, individu usia kerja dikelompokkan berdasarkan karakteristik tertentu untuk menghasilkan informasi ketenagakerjaan yang sesuai [5]. Data yang telah diproses tersebut disajikan dalam format tabel dan memiliki berbagai atribut numerik dengan skala nilai yang bervariasi, sehingga memerlukan tahap awal pemrosesan data, salah satunya adalah dengan menerapkan normalisasi data sebelum digunakan dalam proses penambangan data [6].

Data angkatan kerja yang dirilis oleh Badan Pusat Statistik (BPS) disajikan dalam bentuk angka dengan berbagai atribut yang memiliki satuan dan rentang nilai yang berbeda-beda. Atribut seperti total tenaga kerja, jumlah orang yang bekerja, dan total pengangguran memiliki nilai skala yang cukup besar, sedangkan atribut lain yang berhubungan dengan karakteristik tertentu, seperti kelompok usia, memiliki nilai skala yang lebih kecil. Perbedaan nilai skala antara atribut tersebut dapat menyebabkan dominasi atribut dengan nilai yang lebih besar dalam analisis data, sehingga atribut dengan nilai lebih rendah menjadi kurang berpengaruh [7]. Situasi ini bisa membuat hasil analisis menjadi bias dan tidak mencerminkan karakteristik data secara keseluruhan jika tidak ada tahap praproses data yang dilakukan. Oleh karena itu, sangat penting untuk menerapkan teknik normalisasi data guna menyelaraskan skala nilai antar atribut, sehingga proses analisis data bisa dilakukan dengan lebih objektif dan tepat [8].

Penelitian ini bertujuan untuk mengimplementasikan metode normalisasi data dengan menggunakan *Min-Max Normalization* pada data angkatan kerja Indonesia pada bulan Februari 2025 berdasarkan kategori usia yang diambil dari Badan Pusat Statistik. Pelaksanaan normalisasi ini difokuskan pada penyeragaman skala nilai di antara atribut numerik sehingga data memiliki struktur yang lebih baik dan siap untuk digunakan dalam proses penggalian data [9]. Penelitian ini diharapkan mampu memberikan sumbangan berupa peningkatan mutu data statistik ketenagakerjaan dan menjadi acuan dalam penerapan teknik praproses data, terutama dalam hal normalisasi data, pada data resmi yang diterbitkan oleh BPS.

2. METODE PENELITIAN

Metode riset yang diterapkan dalam studi ini bertujuan untuk menjelaskan langkah-langkah dalam penerapan normalisasi data pada data tenaga kerja Indonesia bulan Februari 2025 menurut kategori usia. Riset ini menggunakan pendekatan kuantitatif dengan memanfaatkan data sekunder yang diambil dari Badan Pusat Statistik (BPS). Langkah-langkah penelitian dirancang dengan sistematis agar proses penelitian ini dapat dimengerti dan dipertanggungjawabkan dari segi ilmiah.

Dataset yang digunakan dalam studi ini adalah data sekunder yang diambil dari Badan Pusat Statistik (BPS) di Indonesia. Informasi tersebut merupakan data mengenai tenaga kerja Indonesia pada bulan Februari 2025 yang diorganisir berdasarkan kelompok usia. Data dari BLPS dipilih karena sifatnya yang resmi, dapat dipercaya, dan luas penggunaannya dalam penelitian akademik serta pengambilan keputusan publik [8].

2.1 Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis agar hasil yang diperoleh dapat dipertanggungjawabkan secara ilmiah. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Pengumpulan Data

Data angkatan kerja Indonesia bulan Februari 2025 dikumpulkan dari publikasi resmi Badan Pusat Statistik (BPS). Data yang dikumpulkan mencakup informasi angkatan kerja berdasarkan kelompok umur.

2. Seleksi Data

Pada tahap ini dilakukan pemilihan data yang relevan dengan tujuan penelitian, yaitu data angkatan kerja pada bulan Februari 2025. Data di luar periode tersebut tidak digunakan dalam proses analisis.

3. Pembersihan Data (*Data Cleaning*)

Tahap pembersihan data dilakukan untuk memastikan data bebas dari kesalahan, nilai kosong (*missing value*), dan inkonsistensi data. Proses ini bertujuan agar data yang digunakan siap untuk tahap normalisasi.

4. Normalisasi Data

Setelah data bersih, dilakukan proses normalisasi data menggunakan metode *Min-Max Normalization*. Normalisasi ini bertujuan untuk menyamakan rentang nilai antar atribut numerik agar tidak terjadi dominasi atribut tertentu dalam analisis.

5. Analisis Hasil Normalisasi

Data yang telah dinormalisasi kemudian dianalisis untuk melihat perubahan skala nilai serta kesiapan data untuk tahap penambangan data selanjutnya.

2.2 Metode Normalisasi Data

Penelitian ini menerapkan tiga teknik untuk normalisasi data, yaitu Normalisasi Min-Max, Normalisasi Z-Score, dan Normalisasi Skala Desimal. Pemanfaatan beberapa teknik normalisasi bertujuan untuk menganalisis karakteristik dari hasil normalisasi serta mengevaluasi kecocokan setiap teknik terhadap data tenaga kerja di Indonesia berdasarkan kategori usia.

2.2.1 Normalisasi Min-Max

Normalisasi Min-Max adalah teknik yang mengubah nilai data ke dalam *range* tertentu, biasanya di antara 0 hingga 1. Metode ini menjaga rasio jarak antar data, sehingga sering diterapkan pada fase pra-proses data [10].

Normalisasi min-max adalah pendekatan normalisasi yang memodifikasi angka dalam data secara linier, menjadikan nilai tertinggi 1 dan nilai terendah 0. Metode ini berguna untuk menurunkan nilai-nilai ekstrem pada fitur yang paling berpengaruh dalam proses pengklasifikasian. Rumus untuk normalisasi min-max dapat ditemukan dalam Rumus [11].

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

2.2.2 Z-Score Normalization

Z-Score Normalization merupakan metode normalisasi yang mengubah nilai data berdasarkan nilai rata-rata (*mean*) dan simpangan baku (standar deviasi). Metode ini menghasilkan data dengan rata-rata 0 dan standar deviasi 1.

Nilai Z-Score merupakan ukuran penyimpangan data dari nilai rata-ratanya (μ) yang diukur dalam satuan standar deviasinya (σ). Z-Score ini juga disebut dengan Nilai Standar atau Nilai baku. Persamaan (1) merupakan formula untuk menghitung nilai Z-Score dari setiap data yang diamati (x) [12]. Persamaan 1 merupakan formula untuk mengetahui Z-Score [13]:

$$Z = \frac{X - \mu}{\sigma}$$

2.2.3 Decimal Scaling Normalization

Salah satu cara untuk menormalkan data yang umum dipakai adalah dengan menggunakan metode skala desimal. Metode ini adalah teknik untuk mentransformasi data melalui normalisasi agar sebanding dalam jangkauan nilai setiap atribut dengan skala tertentu dengan cara memindahkan posisi desimal data ke arah yang diinginkan [14]. Hal ini bertujuan untuk memperoleh data yang lebih efektif [15]. Penelitian sebelumnya mengungkapkan bahwa perbedaan jangkauan nilai pada tiap atribut selama proses transformasi data bisa menyebabkan atribut dengan nilai yang jauh lebih kecil tidak berfungsi dengan baik, sehingga data yang dinormalisasi menggunakan metode skala desimal menunjukkan hasil yang lebih baik secara signifikan dibandingkan dengan klasifikasi yang tidak melakukan transformasi data [4].

Metode desimal Scaling berfungsi dengan memindahkan posisi titik desimal pada nilai-nilai data. Jumlah pemindahan titik desimal ditentukan oleh nilai absolut maksimum dari setiap fitur data atau variabel. Rumus untuk mentransformasi pengukuran desimal dengan j adalah nilai terkecil sehingga nilai $\max(|x'_i|) < 1$, seperti yang dijelaskan berikut ini [16]:

$$X' = \frac{X}{10^j}$$

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah statistik ketenagakerjaan Indonesia berdasarkan penduduk umum pada Februari 2025, yang diperoleh dari Badan Pusat Statistik (BPS). Kelompok usia yang dianalisis meliputi kelompok usia berikut: 15–24 tahun, 25–34 tahun, 35–44 tahun, 45–54 tahun, 55–64 tahun, dan 65 tahun. Atribut numerik yang digunakan adalah jumlah karyawan, jumlah pekerja, dan jumlah pengangguran.

Berdasarkan data yang dikumpulkan, terbukti bahwa jumlah pekerja dan karyawan sangat besar (jutaan orang), meskipun jumlah karyawan relatif kecil. Perbedaan skala ini menunjukkan bahwa data tidak memiliki keseragaman skala, sehingga proses normalisasi harus diselesaikan sebelum analisis lebih lanjut.

3.1. Hasil Penelitian

3.1.1 Data Asli Angkatan Kerja Bulan Februari 2025

Data asli mengindikasikan adanya perbedaan yang cukup signifikan antara kelompok usia. Kelompok usia yang produktif, terutama yang berusia antara 25 hingga 44 tahun, menunjukkan angka angkatan kerja dan penduduk yang bekerja paling tinggi dibandingkan dengan kelompok usia lainnya. Di sisi lain, kelompok yang berusia 65 tahun ke atas menunjukkan jumlah angkatan kerja dan pengangguran yang cenderung lebih rendah.

Perbedaan signifikan dalam nilai antara atribut dan kelompok usia ini dapat menghasilkan bias jika data dipakai langsung dalam analisis *data mining*. Untuk itu, langkah normalisasi data menjadi krusial guna menyelaraskan rentang nilai antar atribut numerik. Berikut merupakan datanya:

Tabel 1. Data Asli

IE-Kelompok Umur	Angkatan Kerja (AK) Menurut Kelompok Umur							
	Bekerja 2025		Jumlah Pengangguran 2025		Jumlah AK 2025		% Bekerja / AK 2025	
	Februari	Agustus	Februari	Agustus	Februari	Agustus	Februari	Agustus
15-24	18421906	18913205	3549928	3843567	21971834	22756772	83.84	83.11
25-34	33184087	33490603	1939099	1791256	35123186	35281859	94.48	94.92
35-44	33951539	33819741	684028	728703	34635567	34548444	98.03	97.89
45-54	30245618	30231916	528796	572637	30774414	30804553	98.28	98.14
55-64	19706443	19750464	416113	333664	20122556	20084128	97.93	98.34
65+	10261587	10336335	160343	191680	10421930	10528015	98.46	98.18
Total	145771180	146542264	7278307	7461507	153049487	154003771	95.24	95.15

Tabel 2. Data AK

IE-Kelompok Umur	Angkatan Kerja (AK) Menurut Kelompok Umur bulan Februari 2025			
	Bekerja	Jumlah Pengangguran	Jumlah AK	% Bekerja / AK
15-24	18421906	3549928	21971834	83.84
25-34	33184087	1939099	35123186	94.48
35-44	33951539	684028	34635567	98.03
45-54	30245618	528796	30774414	98.28
55-64	19706443	416113	20122556	97.93
65+	10261587	160343	10421930	98.46
Total	145771180	7278307	153049487	95.24

3.1.2 Hasil Normalisasi Min–Max

Hasil normalisasi menggunakan metode Min–Max menunjukkan bahwa data berhasil dipetakan ke dalam rentang nilai tertentu sesuai batas minimum dan maksimum. Kelompok umur 35–44 tahun dan 45–54 tahun memperoleh nilai tertinggi pada variabel bekerja dan angkatan kerja, menegaskan peran dominan kelompok usia produktif menengah dalam perekonomian.

Kelompok umur 15–24 tahun menunjukkan nilai negatif pada variabel % Bekerja/AK, yang menandakan rendahnya rasio bekerja dibandingkan angkatan kerja. Kondisi ini sejalan dengan fakta bahwa sebagian besar kelompok usia ini masih berada pada jenjang pendidikan formal. Metode Min–Max sangat efektif untuk visualisasi data dan pemodelan berbasis skala tetap, meskipun sensitif terhadap nilai ekstrem.

Tabel 3. Hasil Normalisasi min max

IE-Kelompok Umur	Bekerja	Jumlah Pengangguran	Jumlah AK	% Bekerja / AK
15-24	8160318.7	3389584.955	11549904	-11.59957749
25-34	22922500	1778755.955	24701256	-0.959577493
35-44	23689952	523684.9548	24213637	2.590422507
45-54	19984031	368452.9548	20352484	2.840422507

55-64	9444855.7	255769.9548	9700625.7	2.490422507
65+	-0.302242	-0.045167958	-0.296725	3.020422507

3.1.3 Hasil Normalisasi Desimal Scaling

Berdasarkan hasil yang diperoleh dari normalisasi dengan penerapan Metode Desimal Scaling, semua nilai data telah berhasil direformasi ke dalam rentang antara 0 sampai 1. Kelompok umur antara 25 hingga 44 tahun mencatat nilai tertinggi pada variabel tenaga kerja dan jumlah angkatan kerja, yang menunjukkan keberadaan dominan kelompok usia yang produktif dalam struktur kerja di tingkat nasional.

Sebaliknya, kategori usia 65 tahun ke atas menunjukkan nilai terendah pada hampir semua variabel, yang mencerminkan penurunan partisipasi dalam dunia kerja sejalan dengan bertambahnya usia. Persentase aktif bekerja relatif tinggi di hampir semua kelompok umur menandakan bahwa mayoritas angkatan kerja telah terserap dalam pasar tenaga kerja.

Penerapan Metode Desimal Scaling terbukti berhasil dalam menyederhanakan skala data tanpa menghilangkan perbandingan proporsional antar kelompok umur, sehingga analisis perbandingan awal menjadi lebih mudah dilakukan.

Tabel 4. Hasil Normalisasi Desimal Scaling

IE-Kelompok Umur	Bekerja	Jumlah Pengangguran	Jumlah AK	% Bekerja / AK
15-24	0.18421906	0.3549928	0.21971834	0.8384
25-34	0.33184087	0.1939099	0.35123186	0.9448
35-44	0.33951539	0.0684028	0.34635567	0.9803
45-54	0.30245618	0.0528796	0.30774414	0.9828
55-64	0.19706443	0.0416113	0.20122556	0.9793
65+	0.10261587	0.0160343	0.1042193	0.9846

3.1.4 Hasil Normalisasi Z-Score

Normalisasi Z-Score menghasilkan angka positif dan negatif yang mencerminkan seberapa jauh data berada dari rata-rata dalam hal satuan deviasi standar. Kelompok usia 35–44 tahun menunjukkan nilai Z-Score tertinggi dalam variabel pekerjaan dan jumlah angkatan kerja, yang menunjukkan bahwa kelompok ini memiliki posisi yang jauh di atas rata-rata nasional.

Di sisi lain, kelompok usia 15–24 tahun dan 65 tahun ke atas menunjukkan nilai Z-Score negatif, terutama dalam variabel pekerjaan, yang menunjukkan bahwa tingkat partisipasi kerja mereka berada di bawah rata-rata. Hal ini bisa dijelaskan oleh tingkat pendidikan pada usia muda dan faktor kesehatan serta pensiun pada usia lanjut.

Variabel % Bekerja/AK menunjukkan variasi Z-Score yang cukup signifikan, yang menunjukkan adanya perbedaan dalam tingkat efektivitas penyerapan tenaga kerja di antara kelompok umur. Metode Z-Score sangat berguna untuk mengenali kelompok umur yang menunjukkan perbedaan signifikan dari keadaan rata-rata nasional.

Tabel 5. Hasil Normalisasi Z-Score

IE-Kelompok Umur	Bekerja	Jumlah Pengangguran	Jumlah AK	% Bekerja / AK
15-24	-0.670660993	1.965739472	-0.398468055	-2.159640677
25-34	1.01500715	0.6107386	1.083370297	-0.131522689
35-44	1.102641178	-0.445004934	1.028427465	0.545152016
45-54	0.679468412	-0.575583347	0.593369183	0.592805164
55-64	-0.523981915	-0.670370295	-0.606836826	0.526090756
65+	-1.602473831	-0.885519496	-1.699862064	0.62711543
Rata-rata	24295196.67	1213051.167	25508247.83	95.17
Standard Deviasi	8757465.733	1188802.925	8875024.716	5.2462431

3.2. Pembahasan

Hasil penelitian menunjukkan bahwa setiap teknik normalisasi menghasilkan karakteristik data yang berbeda, sehingga memberikan implikasi yang bervariasi terhadap analisis lanjutan dan penerapan algoritma *Machine Learning*. Normalisasi Min–Max menghasilkan data dengan rentang nilai yang seragam, sehingga efektif digunakan pada algoritma yang sensitif terhadap skala, seperti *K-Means* dan Jaringan Syaraf Tiruan. Namun, kelemahan utama metode ini terletak pada sensitivitasnya terhadap nilai ekstrem (*outlier*). Pada data angkatan kerja, kelompok usia dengan jumlah tenaga kerja yang sangat besar berpotensi mempersempit rentang nilai kelompok lainnya, sehingga mengurangi representasi variasi data secara keseluruhan.

Sebaliknya, Normalisasi Z-Score lebih unggul dalam mendeteksi penyimpangan karena proses normalisasi didasarkan pada nilai rata-rata dan simpangan baku. Nilai positif dan negatif yang dihasilkan merepresentasikan tingkat deviasi suatu kelompok umur terhadap rata-rata nasional. Dalam konteks data angkatan kerja, metode ini efektif untuk mengidentifikasi kelompok usia dengan tingkat partisipasi kerja yang berada jauh di atas atau di bawah kondisi rata-rata, sehingga relevan untuk analisis anomali dan evaluasi ketimpangan pasar tenaga kerja. Penerapan Z-Score pada algoritma pengelompokan seperti *K-Means* memungkinkan pembentukan klaster yang lebih informatif karena jarak antar data mencerminkan deviasi statistik, bukan sekadar perbedaan skala.

Sementara itu, Normalisasi Skala Desimal mempertahankan proporsi data asli dengan menyederhanakan skala nilai tanpa menghilangkan hubungan relatif antar atribut. Metode ini relatif stabil terhadap nilai ekstrem dan sesuai digunakan pada tahap analisis eksploratif maupun sebagai praproses untuk algoritma klasifikasi yang tidak terlalu sensitif terhadap distribusi data.

Secara keseluruhan, temuan dari studi ini menegaskan bahwa pemilihan teknik normalisasi perlu diselaraskan dengan tujuan dari analisis serta algoritma *Machine Learning* yang diterapkan. Normalisasi yang sesuai dapat meningkatkan mutu pemodelan, mempercepat pelatihan model, dan menghasilkan analisis ketenagakerjaan yang lebih tepat dan dapat diandalkan.

4. KESIMPULAN

Penelitian ini bertujuan untuk menerapkan teknik normalisasi data pada data angkatan kerja Indonesia bulan Februari 2025 berdasarkan kelompok umur, sebagaimana telah diuraikan pada bagian pendahuluan. Berdasarkan hasil dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa tujuan penelitian berhasil dicapai. Hasil penerapan tiga metode normalisasi, yaitu Min–Max Normalization, Z-Score Normalization, dan Decimal Scaling Normalization, menunjukkan bahwa proses normalisasi mampu menyelaraskan skala nilai antar atribut numerik yang sebelumnya memiliki perbedaan rentang nilai yang cukup signifikan. Dengan demikian, potensi bias dalam analisis data akibat dominasi atribut tertentu dapat diminimalkan.

Secara empiris, hasil penelitian mengonfirmasi bahwa kelompok usia 25–44 tahun merupakan kelompok usia paling dominan dalam struktur angkatan kerja Indonesia pada Februari 2025, dengan tingkat partisipasi dan penyerapan tenaga kerja yang tinggi. Sementara itu, kelompok usia 15–24 tahun masih menunjukkan tingkat partisipasi kerja yang relatif rendah, dan kelompok usia 65 tahun ke atas mengalami penurunan aktivitas kerja yang bersifat struktural. Temuan ini konsisten dengan data dan publikasi resmi Badan Pusat Statistik (BPS). Dari sisi metodologis, setiap teknik normalisasi memiliki keunggulan masing-masing. Normalisasi Min–Max efektif untuk pemetaan data dalam rentang tertentu, Z-Score unggul dalam mendeteksi penyimpangan terhadap rata-rata, dan Decimal Scaling mempermudah analisis perbandingan awal tanpa menghilangkan proporsi data. Oleh karena itu, pemilihan metode normalisasi sebaiknya disesuaikan dengan kebutuhan analisis lanjutan. Sebagai prospek pengembangan, hasil penelitian ini dapat dijadikan dasar untuk penerapan teknik penambangan data lebih lanjut, seperti *clustering* atau klasifikasi, guna menggali pola ketenagakerjaan yang lebih kompleks. Selain itu, penelitian selanjutnya dapat memperluas cakupan data dengan menambahkan dimensi wilayah, tingkat pendidikan, atau periode waktu yang berbeda untuk memperoleh gambaran ketenagakerjaan yang lebih komprehensif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Republik Indonesia atas ketersediaan data ketenagakerjaan yang bersifat terbuka dan dapat diakses oleh publik, sehingga penelitian ini dapat dilaksanakan dengan baik. Ucapan terima kasih juga disampaikan kepada Universitas Katolik Santo Thomas atas dukungan akademik dan fasilitas yang diberikan selama proses penyusunan penelitian ini. Selain itu, penulis mengapresiasi berbagai pihak yang telah memberikan masukan dan dukungan, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat diselesaikan dan disajikan dalam bentuk artikel ilmiah.

REFERENSI

- [1] A. J. Wahidin and W. A. Syukrilla, *Data mining*.
- [2] P. P. Alloreng, A. Erna, and M. Bagussahrir, “Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit,” vol. 9, no. 3, pp. 178–191, 2024.
- [3] Y. D. Pratama and A. Salam, “Comparison of Data Normalization Techniques on KNN Classification Performance for Pima Indians Diabetes Dataset,” vol. 9, no. 3, 2025.
- [4] M. R. Kusnaldi, T. Gulo, and S. Aripin, “Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal,” vol. 3, no. 4, pp. 330–338, 2022, doi: 10.47065/josyc.v3i4.2112.
- [5] A. Transactions and M. Learning, “Evaluating the Impact of Data Normalization on Rice Classification Using Machine Learning Algorithms,” pp. 162–171, 2024.
- [6] K. Lima and P. Kota, “Angkatan kerja,” 2021.
- [7] H. S. Silva *et al.*, “The Impact of Feature Scaling In Machine Learning : Effects on Regression and Classification Tasks,” 2025, doi: 10.1109/ACCESS.2025.3635541.
- [8] B. P. Statistik, *No Title*. 2025. [Online]. Available: <https://www.bps.go.id/id>
- [9] J. Manoel *et al.*, “The Impact of Feature Scaling In Machine Learning : Effects on Regression and Classification Tasks,” vol. XX, no. X, 2025.
- [10] B. Practices, “Data Preprocessing and Feature Engineering for Data Mining :,” 2025.
- [11] J. Reynaldo, P. P. Adikara, and R. C. Wihandika, “Analisis Sentimen Mengenai Produk Toyota Avanza Menggunakan Metode Learning Vector Quantization Versi 3 (LVQ 3) dengan Seleksi Fitur Chi Square , Lexicon-Based Features serta Normalisasi Min-Max,” vol. 4, no. 3, pp. 830–839, 2020.
- [12] K. Penderita, D. Menggunakan, and M. L. D. A. N. Z-score, “Jurnal Teknologi Terpadu,” vol. 8, no. 2, pp. 94–99, 2022.

-
- [13] P. Studi and T. Informatika, "PENDEKATAN Z-SCORE DAN FUZZY DALAM PENGUJIAN AKURASI," vol. 3, no. 2, pp. 149–156, 2020.
 - [14] M. A. Othman and A. Samad, "When to Use Standardization and Normalization : Empirical Evidence From Machine Learning Models and XAI," *IEEE Access*, vol. 12, no. July, pp. 135300–135314, 2024, doi: 10.1109/ACCESS.2024.3462434.
 - [15] L. B. V. De Amorim, G. D. C. Cavalcanti, and R. M. O. Cruz, "The choice of scaling technique matters for classification performance," pp. 1–37, 2022.
 - [16] Y. Kim, M. Keun, N. Fu, J. Liu, J. Wang, and J. Srebric, "Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models," *Sustain. Cities Soc.*, vol. 118, no. June 2024, p. 105570, 2025, doi: 10.1016/j.scs.2024.105570.