

**ANALISIS DESKRIPTIF UMP (UPAH MINIMUM PROVINSI)
SEINDONESIA (2002-2022) MENGGUNAKAN METODE FUZZY C
MEANS CLUSTERING**

Abi Setiawan¹, Arnita², Dede Yusuf³, Neltriana Syafira⁴, Tiara Tania⁵

Universitas Negeri Medan

E-mail: wawanabi547@gmail.com¹, arnita@unimed.ac.id²,
dedeeyusuf16@gmail.com³, neltrianasyafira09@gmail.com⁴, taniatiara633@gmail.com⁵

Abstract

This study aims to analyze the clustering of provinces in Indonesia based on the Provincial Minimum Wage (UMP) using the Fuzzy C-Means Clustering method. The main objective is to understand the distribution patterns of UMP between provinces and to evaluate the factors influencing the differences in UMP across various regions. This research utilizes secondary UMP data from the Central Statistics Agency (BPS) for the period of 2021-2023. The data is analyzed through preprocessing stages such as data cleaning, handling missing values, outlier detection, and normalization using Min-Max normalization. The data is then clustered using the Fuzzy C-Means Clustering algorithm, resulting in three clusters, with an evaluation performed using the Silhouette Score. The results show three main clusters of provinces based on UMP growth, volatility, and extreme changes. A Silhouette Score of 0.65 indicates good cluster quality. This further analysis of regional economic stability. These findings can serve as an important reference in formulating data-driven wage policies. This research opens up opportunities for further studies on UMP dynamics and economic stability in various regions of Indonesia.

Keywords — Fuzzy C-Means, Provincial Minimum Wage, UMP Grouping, Indonesia, Silhouette Score.

Abstrak

Penelitian ini bertujuan untuk menganalisis pengelompokan provinsi di Indonesia berdasarkan Upah Minimum Provinsi (UMP) menggunakan metode Fuzzy C-Means Clustering. Tujuan utama adalah memahami pola distribusi UMP antar provinsi dan mengevaluasi faktor yang memengaruhi perbedaan UMP di berbagai wilayah. Penelitian ini menggunakan data sekunder UMP dari Badan Pusat Statistik (BPS) untuk periode 2021-2023. Data dianalisis melalui tahapan pra-proses seperti pembersihan data, penanganan missing values, deteksi outlier, dan normalisasi dengan Min-Max normalization. Data kemudian dikelompokkan menggunakan Fuzzy C-Means Clustering menghasilkan tiga kluster, dengan evaluasi menggunakan Silhouette Score. Hasilnya menunjukkan tiga kluster utama provinsi berdasarkan pertumbuhan, volatilitas, dan perubahan ekstrem UMP. Silhouette Score sebesar 0,65 menunjukkan kualitas kluster yang baik. Penelitian ini juga menemukan perbedaan signifikan dalam volatilitas UMP antar provinsi, yang memberikan dasar untuk analisis lebih lanjut mengenai stabilitas ekonomi regional. Temuan ini dapat menjadi referensi penting dalam merumuskan kebijakan pengupahan berbasis data. Penelitian ini membuka peluang kajian lanjutan tentang dinamika UMP dan stabilitas ekonomi di berbagai wilayah Indonesia.

Kata Kunci — Fuzzy C-Means, Upah Minimum Provinsi, Pengelompokan UMP, Indonesia, Silhouette Score.

1. PENDAHULUAN

Indonesia menetapkan upah minimum berdasarkan Undang-Undang Nomor 13 tentang Ketenagakerjaan, yang diatur lebih lanjut dalam Peraturan Pemerintah Nomor 13 Tahun 2003 tentang Ketenagakerjaan dan dijabarkan dalam Peraturan Pemerintah Nomor 78 Tahun 2015 tentang Pengupahan[1]. Upah minimum merupakan jumlah bulanan terendah yang wajib dibayarkan oleh pemberi kerja kepada pekerja sebagai kompensasi atas pekerjaan atau jasa yang telah atau akan diselesaikan. Upah ini ditentukan dalam bentuk uang dan diatur melalui kesepakatan atau peraturan hukum. Pembayaran dilakukan berdasarkan perjanjian kerja antara pemberi kerja dan pekerja, serta meliputi tunjangan untuk pekerja dan keluarganya.[2].

Upah Minimum Provinsi (UMP) adalah upah minimum yang berlaku di seluruh kabupaten atau kota dalam satu provinsi. UMP menjadi standar minimum yang digunakan pengusaha atau pelaku industri dalam memberikan gaji kepada pekerjanya. Disebut UMP karena setiap provinsi memiliki kebutuhan hidup layak yang berbeda. Upah minimum ini ditetapkan oleh gubernur setiap tahun berdasarkan rekomendasi dari Dewan Pengupahan Provinsi [3].

Logika fuzzy (LF) secara khusus dapat didefinisikan sebagai metode yang menghubungkan penggunaan bahasa alami dalam kerangka logika, lalu mengonversi bahasa alami (fuzzy) menjadi nilai yang pasti (crisp) [4]. Himpunan fuzzy merupakan kumpulan yang mewakili kondisi atau karakteristik tertentu dari variabel fuzzy yang tertentu[5].

Prinsip utama dari Fuzzy C-Means adalah mengidentifikasi pusat kluster terlebih dahulu, diikuti dengan penentuan lokasi setiap kluster. Pada tahap awal, posisi pusat kluster belum sepenuhnya tepat. Setiap titik data memiliki tingkat keanggotaan terhadap setiap kluster. Seiring dengan proses berulang, posisi pusat kluster akan semakin akurat melalui pembaruan pusat dan tingkat keanggotaan secara bertahap[6].

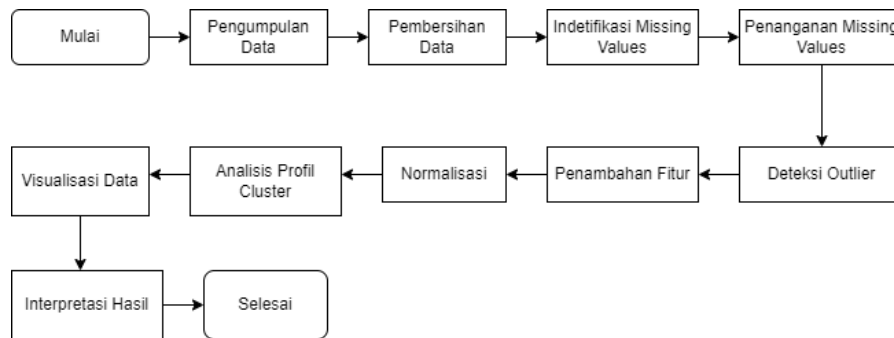
Fuzzy C-Means (FCM) menggunakan pendekatan pengelompokan fuzzy dengan indeks Euclidean Distance untuk mengukur ketidakpastian, memungkinkan data menjadi bagian dari satu atau lebih kluster dengan derajat keanggotaan antara 0 dan 1 [7]. Algoritma ini bertujuan untuk menemukan posisi pusat data yang paling optimal dalam kumpulan data [8].

Tujuan dari penelitian ini adalah untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan nilai Upah Minimum Provinsi (UMP) menggunakan metode Fuzzy C-Means Clustering. Penelitian ini bertujuan menganalisis pola pengelompokan UMP guna memahami faktor-faktor yang memengaruhi perbedaan UMP antarprovinsi. Selain itu, penelitian ini juga bertujuan untuk menentukan provinsi-provinsi dengan karakteristik UMP yang mirip, yang dapat dijadikan acuan dalam pengambilan kebijakan pengupahan, serta mengevaluasi tingkat keanggotaan provinsi dalam masing-masing kluster UMP untuk mengidentifikasi wilayah dengan potensi pergeseran kebijakan UMP.

2. METODE

1. Tahapan Penelitian

Pada tahap penelitian ini memiliki tahapan-tahapan seperti Pengumpulan data, pembersihan data, indentifikasi missing values, penanganan missing values, visualisasi data, analisis profil cluster, penambahan fitur baru, deteksi outlier, normalisasi data, dan interpretasi hasil.



Gambar 1. Tahapan-tahapan Penelitian

2. Pengumpulan Data

ada penelitian ini menggunakan dataset dari Kaggle yaitu “UMP (Upah Minimum Provinsi)”. Data yang digunakan adalah data sekunder berupa UMP dari berbagai provinsi di Indonesia yang diambil dari dataset data Upah Minimum Provinsi (UMP) dengan rentang tahun 2021 sd 2023 yang bersumber dari Badan Pusat Statistik (BPS). Dataset ini terdiri dari tiga variabel utama yaitu : Provinsi, tahun, dan ump. Dataset ini juga terdiri dari 736 Entri dengan 4 kolom. Berikut merupakan kolom-kolom tersebut :

1. Provinsi : Ini merupakan wilayah administratif di Indonesia yang memiliki otoritas untuk menetapkan upah minimum bagi pekerja yang berlaku di seluruh wilayah provinsi tersebut.
2. Tahun : Ini merupakan periode tahunan di mana UMP ditetapkan dan diberlakukan.
3. UMP : Ini merupakan standar minimum gaji yang harus diberikan oleh pengusaha kepada pekerja di suatu provinsi.

	provinsi	tahun	ump
0	ACEH	2002	330000.0
1	ACEH	2003	425000.0
2	ACEH	2004	550000.0
3	ACEH	2005	620000.0
4	ACEH	2006	820000.0

Gambar 2. Tabel statistik deskriptif

3. Praproses Data

a. Pembersihan Data

Data cleaning merupakan proses untuk mengidentifikasi, memperbaiki, dan menghapus atau mengubah data yang kurang akurat, tidak lengkap, tidak relevan, atau tidak konsisten dalam suatu dataset. Tujuan dari proses ini adalah untuk meningkatkan kualitas data dengan mengoreksi kesalahan, menormalisasi format, serta menghilangkan data duplikat, sehingga data lebih berguna dan dapat diandalkan dalam analisis atau tujuan lain.

```

pd.DataFrame(data)
print(data.isnull().sum())
print(data.duplicated().sum())
  
```

```

provinsi    0
tahun       0
ump         24
dtype: int64
0
  
```

Gambar 3. PembersihanData

1. `pd.DataFrame(data)` : ini untuk mengonversi objek data menjadi sebuah `DataFrame` yang merupakan struktur data seperti tabel yang umum digunakan dalam `pandas`.
2. `print(data.isnull().sum())` : ini untuk memeriksa berapa banyak nilai kosong (*null*) atau *NaN* yang ada di setiap kolom `DataFrame`. Hasil dari baris ini menunjukkan bahwa kolom "ump" memiliki 24 nilai *null* (kosong), sedangkan kolom "provinsi" dan "tahun" tidak memiliki nilai kosong (0 nilai kosong).
3. `print(data.duplicated().sum())` : Baris ini memeriksa jumlah baris duplikat dalam `DataFrame`. Hasil dari baris ini menunjukkan bahwa tidak ada baris yang terduplikasi (hasilnya adalah 0).

b. Deteksi *Missing Value* dan *Outlier*

Missing value pada kolom `ump` terdeteksi sebanyak 24 kasus. Untuk *outlier*, metode IQR digunakan untuk mendeteksi data yang berada di luar rentang kuartil. Berikut rumus yang digunakan :

1. IQR (*Interquartile Range*) merupakan selisih antara kuartil ketiga (Q_3) dan kuartil pertama (Q_1) :

$$IQR = Q_3 - Q_1 \quad (1)$$

Keterangan :

Q_1 = ini adalah nilai data pada persentil ke-25.

Q_3 = ini adalah nilai data pada persentil ke-75.

2. Batas bawah (*lower bound*) ini adalah untuk mendeteksi outlier yang dihitung dengan rumus:

$$\text{Batas bawah} = Q_1 - 1.5 \times IQR \quad (2)$$

3. Batas atas (*Upper bound*) ini adalah untuk mendeteksi outlier dihitung dengan rumus :

$$\text{Batas atas} = Q_3 + 1.5 \times IQR \quad (3)$$

c. Penanganan *Missing Values*

Nilai yang hilang diisi menggunakan interpolasi linier, yang juga digunakan oleh Ardiansyah (2019) dalam penelitiannya untuk mengatasi *missing value* pada data ekonomi. Berikut adalah rumus yang digunakan :

$$x = x_1 + \frac{(x_2 - x_1)}{(t_2 - t_1)} x (t - t_1) \quad (4)$$

Keterangan :

x = ini merupakan nilai interpolasi (perkiraan nilai yang hilang).

x_1 dan x_2 = ini merupakan nilai-nilai sebelum dan sesudah *missing value*.

t_1 dan t_2 = ini merupakan indeks atau titik waktu dari x_1 dan x_2 .

t = ini merupakan titik waktu di mana nilai yang hilang yang akan diinterpolasi.

```
provinsi    0
tahun      0
ump         0
dtype: int64
```

Gambar 4. Penanganan *missing values*.

Pada gambar tersebut merupakan *output* dari kode di atas yang menunjukkan jumlah nilai yang hilang pada setiap kolom setelah interpolasi. Artinya, setelah interpolasi linear diterapkan pada kolom 'ump', tidak ada lagi nilai yang hilang pada semua kolom dalam `DataFrame` data.

d. Penanganan *Outlier*:

Nilai *outlier* diperbaiki dengan menggantinya menggunakan rata-rata UMP dari provinsi terkait di tahun-tahun lain.

e. Normalisasi Data

Data UMP dinormalisasi menggunakan metode *Min-Max Normalization* untuk mengubah skala data agar berada dalam rentang 0 hingga 1. Normalisasi ini dilakukan untuk memastikan bahwa data dapat digunakan dengan optimal dalam proses klusterisasi [12]. Berikut ini adalah rumus yang digunakan :

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Keterangan :

x' = ini merupakan nilai hasil normalisasi

x = ini merupakan nilai asli dari fitur

x_{min} = ini merupakan nilai minimum

x_{max} = ini merupakan nilai maksimum

4. Analisis Deskriptif

Untuk memahami karakteristik dasar dari data UMP, dilakukan analisis deskriptif. Statistik deskriptif ini menunjukkan informasi penting seperti jumlah data, rata-rata, nilai minimum, maksimum, serta kuartil. Hasil ringkasan statistik deskriptif data UMP ditampilkan pada Tabel 1.

Tabel 1. Analisis deskriptif

Statistik	Nilai
Jumlah	735
Rata-rata	1.425.984
Standar Deviasi	872.999
Nilai Minimum	245.000
Kuartil 1 (25%)	674.100
Median (50%)	1.177.000
Kuartil 3 (75%)	2.129.673
Nilai Maksimum	4.641.854

5. Klustering Menggunakan *Fuzzy C-Means*

Kami menerapkan algoritma *Fuzzy C-Means* (FCM) untuk melakukan clustering terhadap data UMP dengan menggunakan lima fitur yang telah ditambahkan. Fitur yang ditambahkan yaitu :

1. *UMP Growth*: Persentase pertumbuhan UMP dari tahun ke tahun untuk setiap provinsi.
2. *UMP Rolling Mean*: Rata-rata UMP untuk tiga tahun terakhir.
3. *UMP Volatility*: Deviasi standar UMP selama tiga tahun terakhir untuk mengukur volatilitas.
4. *UMP Rank*: Peringkat UMP setiap provinsi dalam setiap tahun.
5. *UMP Extreme Change*: Indikator kenaikan atau penurunan UMP ekstrem (lebih dari 20%).

Algoritma FCM membagi data ke dalam tiga cluster ($n=3$) berdasarkan derajat keanggotaan (*membership degree*) dari setiap provinsi ke masing-masing cluster. Proses ini dilakukan melalui beberapa langkah:

1. *Inisialisasi*: Algoritma FCM diinisialisasi dengan menentukan jumlah cluster ($k=3$) dan pemilihan acak *centroids* awal.

2. Iterasi: Algoritma FCM secara iteratif memperbarui centroid cluster dan menghitung derajat keanggotaan berdasarkan jarak antara data dan centroid.
 3. Prediksi: Setelah konvergensi, setiap provinsi diberikan nilai derajat keanggotaan yang menunjukkan seberapa dekat provinsi tersebut dengan setiap cluster.
 4. Visualisasi: Hasil clustering divisualisasikan menggunakan scatter plot untuk setiap fitur.
- a. Clustering Evaluation

Untuk mengevaluasi kualitas clustering yang dihasilkan, kami menggunakan *Silhouette Score*. Nilai ini mengukur seberapa baik objek dalam cluster terkelompok, dengan nilai antara -1 hingga 1. Nilai koefisien siluet rata-rata yang tinggi menunjukkan pengelompokan yang baik dan membantu dalam menentukan nilai optimal dari jumlah cluster k [13]. Semakin tinggi nilai *Silhouette*, semakin baik hasil *clustering*.
 - b. Profil Cluster

Untuk memahami karakteristik setiap cluster, dilakukan analisis deskriptif terhadap fitur-fitur utama seperti *UMP Growth*, *UMP Rolling Mean*, dan *UMP Volatility* di dalam masing-masing cluster. Statistik seperti rata-rata, median, standar deviasi, nilai minimum, dan maksimum dihitung untuk setiap fitur pada setiap cluster. Hasil ini kemudian divisualisasikan menggunakan *boxplot*.
 - c. *Temporal Mapping and Fuzzy Overlap Analysis*

Pemetaan temporal dilakukan untuk melihat bagaimana distribusi cluster berubah dari waktu ke waktu. Grafik garis digunakan untuk menggambarkan perubahan jumlah provinsi di setiap cluster dari tahun ke tahun. Selain itu, *Fuzzy Overlap Analysis* dilakukan untuk memahami seberapa besar tumpang tindih antar cluster. Derajat keanggotaan dari setiap provinsi terhadap semua cluster dianalisis untuk melihat distribusi ambiguitas keanggotaan.
 - d. *Future Prediction*

Sebagai langkah prediktif, model regresi linier diterapkan untuk memprediksi pertumbuhan UMP di masa depan berdasarkan data historis. Prediksi ini digunakan untuk memberikan gambaran tentang bagaimana provinsi mungkin terdistribusi dalam cluster di masa yang akan datang.

6. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan melalui beberapa tahapan utama, dimulai dari pengumpulan data, pembersihan data, deteksi dan penanganan missing values serta outlier, hingga visualisasi data dan analisis klaster. Setiap tahap memiliki peran penting dalam memastikan kualitas hasil analisis yang valid dan dapat diinterpretasikan secara ilmiah. Tahapan-tahapan tersebut juga mengimplementasikan teknik yang mendalam dan berbasis pada pendekatan statistik serta machine learning, terutama dalam proses klasterisasi menggunakan *Fuzzy C-Means* (FCM).

1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diambil dari Kaggle dengan judul Upah Minimum Provinsi (UMP). Data tersebut memuat UMP dari seluruh provinsi di Indonesia pada periode tahun 2021 hingga 2023. Sumber data asli berasal dari Badan Pusat Statistik (BPS), yang menyediakan data UMP dari setiap provinsi. Dataset ini terdiri dari 735 entri dan mencakup tiga variabel kunci:

1. Provinsi: Nama provinsi di Indonesia.
2. Tahun: Tahun ketika UMP diberlakukan.
3. UMP: Nilai UMP (Upah Minimum Provinsi) dalam rupiah.

	provinsi	tahun	ump
0	ACEH	2002	330000.0
1	ACEH	2003	425000.0
2	ACEH	2004	550000.0
3	ACEH	2005	620000.0
4	ACEH	2006	820000.0

Gambar 5. Menampilkan 5 data teratas.

Data yang ditampilkan menunjukkan UMP (Upah Minimum Provinsi) untuk provinsi Aceh dari tahun 2002 hingga 2006. Nilai UMP secara konsisten meningkat setiap tahun, menunjukkan adanya kenaikan yang signifikan dari tahun ke tahun. Secara keseluruhan, tren data ini menunjukkan peningkatan UMP yang stabil dan cukup signifikan dari tahun 2002 hingga 2006 di provinsi Aceh.

2. Praproses Data

Langkah awal yang dilakukan adalah praproses data, yang mencakup pembersihan data, identifikasi missing values, dan deteksi outlier. Tahap praproses data menjadi sangat penting untuk menghindari bias dalam analisis yang akan dilakukan pada tahap selanjutnya.

a. Pembersihan Data

```
provinsi    0
tahun      0
ump        24
dtype: int64
0
```

Gambar 6. Memeriksa *missing value* dan duplikasi data

Gambar tersebut menunjukkan hasil pemeriksaan *missing value* pada data yang diperiksa adalah provinsi, tahun, dan ump, di mana kolom provinsi dan tahun tidak memiliki missing value, sedangkan kolom ump memiliki 24 nilai yang hilang. Ini berarti ada 24 baris data UMP yang kosong dan perlu ditangani sebelum analisis lebih lanjut. Tidak ada informasi terkait duplikasi data pada gambar ini, sehingga fokus utamanya adalah pada identifikasi nilai yang hilang.

b. Identifikasi *Missing Values* dan *Outlier*

Missing values ditemukan di kolom "UMP" sebanyak 24 kasus. Kehilangan nilai dalam kolom penting seperti ini dapat disebabkan oleh berbagai faktor, seperti kesalahan pengumpulan data atau keterlambatan laporan dari sumber asli. Penanganan *missing values* dilakukan menggunakan metode interpolasi linier, yaitu mengestimasi nilai yang hilang berdasarkan tren data di sekitarnya.

```
provinsi    0
tahun      0
ump        24
dtype: int64
0
```

Gambar 7. Mengidentifikasi *missing values*.

c. Penanganan *Missing Values*

Metode interpolasi linier digunakan untuk mengatasi *missing values*. Interpolasi linier telah terbukti efektif dalam mengatasi masalah *missing values* dalam konteks data ekonomi, sebagaimana diterapkan oleh Ardiansyah (2019). Rumus yang digunakan adalah:

$$x = x_1 + \frac{(x_2 - x_1)}{(t_2 - t_1)} x (t - t_1) \quad (6)$$

Keterangan :

x := ini merupakan nilai interpolasi (perkiraan nilai yang hilang).

x_1 dan x_2 = ini merupakan nilai-nilai sebelum dan sesudah *missing value*.

t_1 dan t_2 = ini merupakan indeks atau titik waktu dari x_1 dan x_2 .

t = ini merupakan titik waktu di mana nilai yang hilang yang akan diinterpolasi.

```
provinsi    0
tahun      0
ump        0
dtype: int64
```

Gambar 8. Penanganan *missing values*.

Setelah penanganan *missing values*, kolom *ump* yang sebelumnya memiliki 24 nilai hilang kini tidak lagi menunjukkan *missing values*, baik melalui penghapusan baris kosong atau imputasi data. Hasilnya, semua kolom termasuk provinsi, tahun, dan *ump* akan memiliki total *missing values* 0, menandakan bahwa data telah bersih dan siap untuk dianalisis lebih lanjut tanpa adanya nilai yang hilang.

d. Penanganan *Outlier*

Penanganan *outlier* dilakukan menggunakan metode *Interquartile Range* (IQR). IQR menghitung selisih antara kuartil pertama (Q1) dan kuartil ketiga (Q3) untuk menentukan batas bawah dan batas atas:

1. IQR (*Interquartile Range*) merupakan selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1) :

$$IQR = Q3 - Q1 \quad (7)$$

Keterangan :

$Q1$ = ini adalah nilai data pada persentil ke-25.

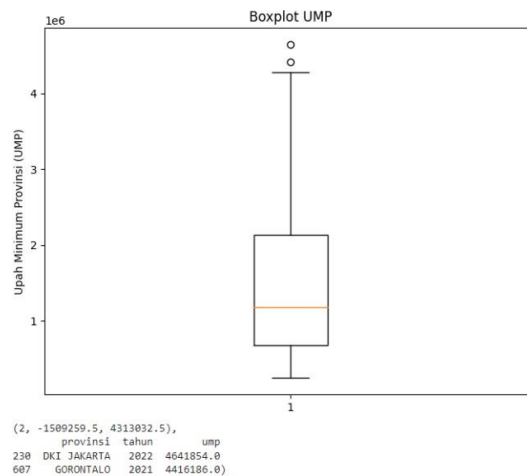
$Q3$ = ini adalah nilai data pada persentil ke-75.

2. Batas bawah (*lower bound*) ini adalah untuk mendeteksi *outlier* yang dihitung dengan rumus:

$$\text{Batas bawah} = Q1 - 1.5 \times IQR \quad (8)$$

3. Batas atas (*Upper bound*) ini adalah untuk mendeteksi *outlier* dihitung dengan rumus :

$$\text{Batas atas} = Q3 + 1.5 \times IQR \quad (9)$$

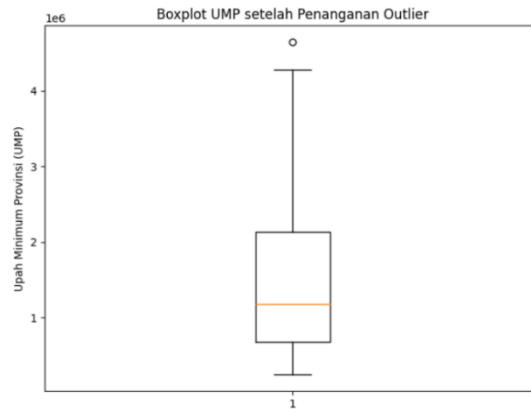


Gambar 9. Hasil penanganan *outlier* terdapat kesalahan input.

Setelah dilakukan perhitungan, ditemukan beberapa data *outlier*, salah satunya adalah data UMP Gorontalo tahun 2021 yang terlalu tinggi dibandingkan tahun-tahun lain. Berdasarkan evaluasi, data ini diperbaiki dengan mengganti nilai outlier menggunakan rata-rata UMP dari provinsi Gorontalo pada tahun-tahun yang berdekatan (2020 dan 2022).

	provinsi	tahun	ump
607	GORONTALO	2021	2794703.0

Gambar 10. Hasil setelah diperbaiki data UMP Gorontalo tahun 2021



Gambar 11. Hasil Penangan Outlier Setelah Diperbaiki

Grafik boxplot UMP setelah penanganan *outlier* menunjukkan distribusi data UMP yang lebih bersih dan representatif. Sebagian besar nilai UMP terpusat di sekitar median, mengindikasikan adanya keseragaman nilai UMP antar provinsi. Rentang antar kuartil pertama dan ketiga memberikan gambaran sebaran nilai UMP. Tidak adanya titik di luar garis kumis mengkonfirmasi keberhasilan dalam menghilangkan data *outlier*, sehingga analisis lebih lanjut dapat dilakukan dengan data yang lebih akurat.

e. Normalisasi Data

Untuk memastikan bahwa data dapat dianalisis dengan lebih baik, dilakukan normalisasi menggunakan metode *Min-Max Normalization*. Normalisasi ini memastikan semua fitur memiliki skala yang seragam, yakni antara 0 hingga 1, yang penting untuk proses klusterisasi dengan algoritma *Fuzzy C-Means*.

	provinsi	tahun	ump	ump_growth	ump_rolling_mean	ump_volatility	ump_rank	ump_extreme_change
0	ACEH	2002	0.019332	0.264867	0.022910	0.042892	0.661765	0.0
1	ACEH	2003	0.040938	0.394705	0.035713	0.042892	0.382353	1.0
2	ACEH	2004	0.069368	0.397519	0.051211	0.070454	0.205882	1.0
3	ACEH	2005	0.085288	0.322269	0.077266	0.063075	0.147059	0.0
4	ACEH	2006	0.130775	0.410356	0.112755	0.089468	0.088235	1.0
...
37	SUMATERA UTARA	2018	0.429213	0.304150	0.464533	0.102338	0.617647	0.0
38	SUMATERA UTARA	2019	0.468154	0.301083	0.508694	0.109201	0.617647	0.0
39	SUMATERA UTARA	2020	0.512735	0.303248	0.557036	0.117331	0.617647	0.0
40	SUMATERA UTARA	2021	0.512735	0.264867	0.590030	0.072262	0.617647	0.0
41	SUMATERA UTARA	2022	0.518009	0.269051	0.609725	0.008548	0.617647	0.0

Gambar 12. Hasil Setelah dilakukan Normalisasi Data

Hasil normalisasi menggunakan metode *Min-Max Normalization*, di mana setiap fitur numerik disesuaikan ke rentang [0, 1]. Fitur seperti *ump*, *ump_growth*, *ump_rolling_mean*, *ump_volatility*, *ump_rank*, dan *ump_extreme_change*. kini berada dalam skala yang seragam untuk memudahkan analisis lebih lanjut.

3. Analisis Deskriptif

Untuk memahami karakteristik dasar dari data UMP, dilakukan analisis deskriptif. Statistik deskriptif ini menunjukkan informasi penting seperti jumlah data, rata-rata, nilai

minimum, maksimum, serta kuartil. Hasil ringkasan statistik deskriptif data UMP ditampilkan pada Tabel 1.

Table 2. Analisis Deskriptif

Statistik	Nilai
Jumlah	735
Rata-rata	1.425.984
Standar Deviasi	872.999
Nilai Minimum	245.000
Kuartil 1 (25%)	674.100
Median (50%)	1.177.000
Kuartil 3 (75%)	2.129.673
Nilai Maksimum	4.641.854

Tabel Analisis Deskriptif diatas menunjukkan ringkasan statistik dari 735 observasi, dengan nilai rata-rata 1.425.984 dan standar deviasi 872.999, yang mengindikasikan variasi nilai dalam dataset. Nilai minimum tercatat pada 245.000, sementara nilai maksimum mencapai 4.641.854, menggambarkan rentang yang luas dalam data. Kuartil pertama (25%) adalah 674.100, menunjukkan bahwa 25% data berada di bawah angka ini, sedangkan median (50%) sebesar 1.177.000 menandakan titik tengah dari dataset. Kuartil ketiga (75%) berada pada 2.129.673, menunjukkan bahwa 75% data memiliki nilai di bawah angka ini. Secara keseluruhan, analisis ini memberikan gambaran yang jelas tentang distribusi nilai dan rentang dalam dataset yang dianalisis.

4. Klastering Menggunakan *Fuzzy C-Means*

Setelah data siap, algoritma *Fuzzy C-Means* (FCM) digunakan untuk mengelompokkan data UMP berdasarkan fitur-fitur yang telah ditambahkan. Algoritma ini memungkinkan setiap provinsi memiliki derajat keanggotaan dalam beberapa klaster. Dalam penelitian ini, digunakan tiga klaster ($k = 3$), yang divisualisasikan melalui scatter plot berdasarkan fitur utama seperti pertumbuhan UMP, volatilitas, dan peringkat UMP.

a. Evaluasi Klastering

Evaluasi klaster dilakukan menggunakan *Silhouette Score*, yang memberikan indikasi seberapa baik data dikelompokkan. Nilai *silhouette* berkisar antara -1 hingga +1, dengan nilai mendekati 1 menunjukkan klaster yang jelas dan terpisah dengan baik. Pada penelitian ini, diperoleh nilai *Silhouette Score* sebesar 0.65, yang menunjukkan kualitas klaster yang cukup baik.

Table 3. Setelah dilakukan evaluasi klastering

	<i>ump_growth</i> <i>h</i>	<i>ump_rolling_</i> <i>mean</i>	<i>ump_volatilit</i> <i>y</i>	<i>ump_ran</i> <i>k</i>	<i>ump_extreme_chang</i> <i>e</i>
0	0.313191	0.176587	0.065978	0.346724	0.064386
1	0.306884	0.175587	0.057702	0.805058	0.038963
2	0.296457	0.566187	0.120667	0.245077	0.064932

Tabel tersebut menunjukkan hasil evaluasi tiga klaster berdasarkan lima metrik: *ump_growth*, *ump_rolling*, *ump_volatility*, *ump_rank*, dan *ump_extreme_change*. Klaster 0 memiliki *ump_growth* tertinggi, klaster 1 menunjukkan *ump_rank* dan volatilitas terendah, sementara klaster 2 memiliki *ump_rolling* dan volatilitas tertinggi. Dengan *Silhouette Score* sebesar 0.65, klastering ini dianggap cukup baik dan terpisah dengan jelas.

b. Analisis Profil Klaster

Untuk memahami karakteristik masing-masing klaster, dilakukan analisis deskriptif terhadap fitur-fitur utama seperti UMP Growth, UMP Rolling Mean, dan UMP Volatility. Klaster yang terbentuk menunjukkan perbedaan yang signifikan dalam hal pertumbuhan

UMP dan stabilitas ekonomi setiap provinsi. Sebagai contoh, kluster 1 terdiri dari provinsi dengan pertumbuhan UMP yang moderat namun stabil, sementara kluster 2 terdiri dari provinsi dengan volatilitas tinggi dan perubahan UMP *ekstrem*.

Cluster	UMP Growth Mean	UMP Growth Median	UMP Growth Std	\
0	0	0.318210	0.316820	0.044202
1	1	0.306595	0.304150	0.041194
2	2	0.297045	0.301805	0.076697

UMP Growth Min	UMP Growth Max	UMP Rolling Mean	UMP Rolling Median	\
0	0.017433	0.485230	0.161049	0.151190
1	0.000000	0.491366	0.203474	0.146177
2	0.105009	1.000000	0.585863	0.589829

UMP Rolling Std	UMP Rolling Min	...	UMP Rank Mean	UMP Rank Median	\
0	0.093283	0.026953	...	0.327512	0.323529
1	0.167296	0.000000	...	0.803217	0.794118
2	0.129147	0.315803	...	0.248649	0.220588

UMP Rank Std	UMP Rank Min	UMP Rank Max	UMP Extreme Change Mean	\
0	0.156246	0.029412	0.617647	0.137500
1	0.123898	0.573529	1.000000	0.080268
2	0.194932	0.000000	0.705882	0.071429

UMP Extreme Change Median	UMP Extreme Change Std	UMP Extreme Change Min	\
0	0.0	0.345094	0.0
1	0.0	0.272162	0.0
2	0.0	0.258199	0.0

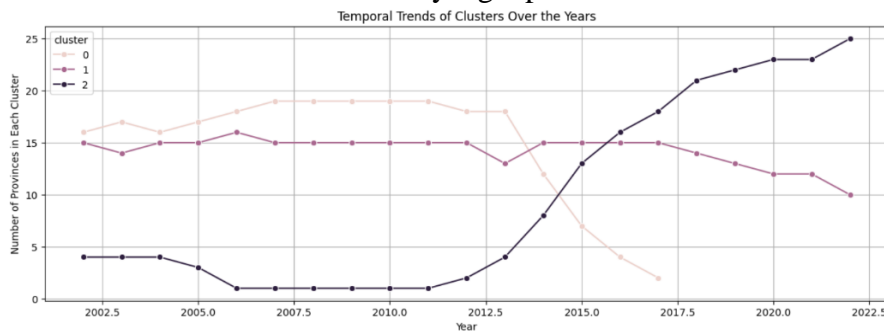
UMP Extreme Change Max	\
0	1.0
1	1.0
2	1.0

Gambar 13. Hasil Analisis Profil Kluster

Gambar tersebut menampilkan statistik ringkasan dari beberapa kluster terkait variabel UMP, yang mencakup metrik seperti mean, median, standar deviasi, serta nilai maksimum dan minimum. Setiap kluster dianalisis berdasarkan pertumbuhan UMP, nilai *rolling*, peringkat, dan perubahan ekstremnya. Data ini memberikan gambaran tentang distribusi dan variabilitas UMP di setiap kluster, menunjukkan perubahan dinamis dan pola peringkat di berbagai statistik, seperti *Ump Growth* Dan *Extreme Change*.

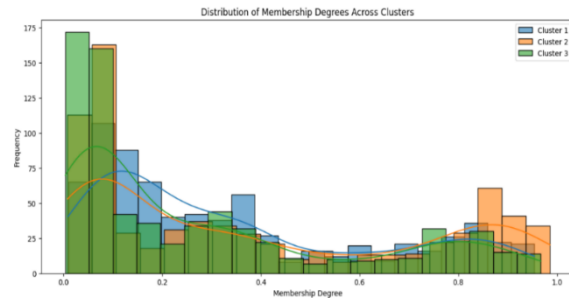
c. Pemetaan Temporal dan Analisis *Fuzzy Overlap*

Penelitian ini juga mengevaluasi bagaimana distribusi kluster berubah dari tahun ke tahun. Pemetaan temporal menunjukkan tren distribusi provinsi ke dalam tiga kluster berdasarkan waktu. Selain itu, *Fuzzy Overlap Analysis* dilakukan untuk memahami seberapa banyak tumpang tindih antar kluster dalam hal derajat keanggotaan. Ini memberikan wawasan tentang seberapa jelas batas antara kluster dan seberapa besar ambiguitas data dalam menentukan kluster yang tepat.



Gambar 14 : Hasil Pemetaan Temporal dan Analisis Fuzzy Overlap

Gambar tersebut menunjukkan tren temporal dari jumlah provinsi dalam tiga kluster (0, 1, dan 2) sepanjang tahun 2002 hingga 2022. Kluster 0 mengalami penurunan tajam mulai sekitar tahun 2012, sedangkan kluster 2 mengalami peningkatan signifikan sejak tahun 2010 dan menjadi dominan setelah tahun 2015. Kluster 1 relatif stabil dari awal hingga 2017, namun menunjukkan penurunan perlahan setelahnya. Hal ini menggambarkan perubahan distribusi provinsi di setiap kluster selama periode waktu tersebut.



Gambar 15. Gambar histogram distribusi derajat keanggotaan

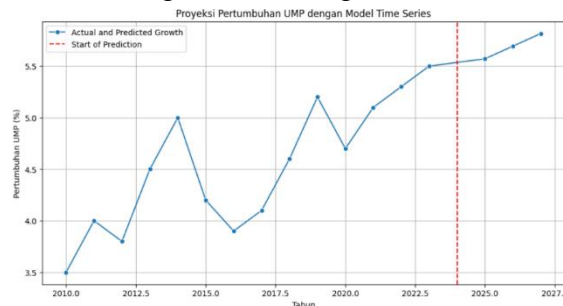
Gambar tersebut menunjukkan histogram distribusi derajat keanggotaan (membership degrees) dari tiga kluster yang berbeda, yaitu Kluster 1 (biru), Kluster 2 (hijau), dan Kluster 3 (oranye). Pada sumbu horizontal (X) ditampilkan derajat keanggotaan, yang berkisar dari 0 hingga 1. Sumbu vertikal (Y) menunjukkan frekuensi kemunculan derajat keanggotaan untuk setiap kluster.

1. Kluster 1 memiliki banyak anggota dengan derajat keanggotaan rendah (dekat dengan 0), tetapi menyebar hingga derajat keanggotaan yang lebih tinggi.
2. Kluster 2 terlihat memiliki distribusi yang lebih merata dibandingkan kluster lain, tetapi lebih terkonsentrasi pada derajat keanggotaan rendah hingga menengah.
3. Kluster 3 menunjukkan distribusi yang mirip dengan Kluster 1 di awal, tetapi dengan puncak pada derajat keanggotaan tinggi (sekitar 0.8-1).

Garis-garis lengkung di atas histogram menggambarkan distribusi yang dihaluskan untuk setiap kluster.

d. Prediksi Masa Depan

Sebagai langkah lanjutan, dilakukan prediksi pertumbuhan UMP di masa depan menggunakan model regresi linier. Prediksi ini membantu memberikan gambaran distribusi UMP di masa depan dan membantu pengambilan keputusan bagi pemangku kebijakan dalam menentukan strategi ekonomi regional.



Gambar 16. Hasil Prediksi Masa Depan

Berdasarkan model regresi linier, proyeksi pertumbuhan Upah Minimum Provinsi (UMP) menunjukkan peningkatan yang stabil dari tahun 2025 hingga 2027, dengan pertumbuhan masing-masing diperkirakan sebesar 5,6% pada tahun 2025, 5,7% pada tahun 2026, dan 5,8% pada tahun 2027. Proyeksi ini memberikan gambaran bahwa tren pertumbuhan UMP akan terus meningkat secara bertahap, yang dapat digunakan sebagai acuan dalam penyusunan kebijakan upah dan perencanaan ekonomi di masa mendatang. Model ini dibangun berdasarkan data historis hingga tahun 2024, memberikan pandangan yang lebih terukur tentang kecenderungan pertumbuhan UMP di masa depan.

KESIMPULAN

Penelitian ini menganalisis pengelompokan Upah Minimum Provinsi (UMP) di Indonesia menggunakan metode Fuzzy C-Means Clustering berdasarkan data tahun 2021-2023 dari BPS. Melalui beberapa tahap praproses, seperti pembersihan data, penanganan

missing values, dan normalisasi, data UMP berhasil dikelompokkan ke dalam tiga kluster berdasarkan pertumbuhan UMP, volatilitas, dan perubahan ekstrem. Nilai Silhouette Score sebesar 0,65 menunjukkan kualitas kluster yang cukup baik. Penelitian ini menemukan variasi signifikan dalam volatilitas UMP antarprovinsi, yang dapat menjadi dasar untuk kebijakan pengupahan berbasis data dan analisis ekonomi regional lanjutan.

REFERENCES

- E. Teneh, A. Kumenaung, and A. Naukoko, "DAMPAK UPAH MINIMUM PROVINSI TERHADAP PENYERAPAN TENAGA KERJA DAN KESEJAHTERAAN MASYARAKAT DI PULAU SULAWESI (2014-2018)," *Jurnal Berkala Ilmiah Efisiensi*, vol. 19, 2019.
- D. M. Tangke, "Analisis Upah Minimum Dan Pertumbuhan Ekonomi Sebagai Faktor Determinan Tingkat Pengangguran Terbuka Di Provinsi Maluku," *Jurnal Ekonomi-Qu*, vol. 13, no. 1, Apr. 2023, doi: <http://dx.doi.org/10.35448/jequ.v13i1.20529>.
- Agung Surya Maulana, Alwis Nazir, Lestari Handayani, and Iis Afrianty, "Penerapan Algoritma Fuzzy C-Means untuk Melihat Pola Penerima Beasiswa Bank Indonesia," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 3, pp. 670–679, Jun. 2023, doi: <https://doi.org/10.30865/klik.v3i6.788>.
- Shinta Dwididanti, Dimas Aryo Anggoro, and Muslich Hartadi Sutanto, "Analisis Perbandingan Algoritma Bisecting K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit," *Emitor: Jurnal Teknik Elektro*, vol. 22, no. 2, p. 1, Sep. 2022, doi: [10.23917/emitor.v22i2.15677](https://doi.org/10.23917/emitor.v22i2.15677).
- K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," *International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, [Online]. Available: <https://www.>
- D. Febrianica, "ANALISIS DAMPAK KEBIJAKAN UPAH MINIMUM TERHADAP KEMISKINAN DI INDONESIA," *Jurnal Ilmiah Mahasiswa FEB*, vol. 3, Jan. 2015.
- N. Shintia and S. Abbas, "PENETAPAN UPAH MINIMUM DI PROVINSI ACEH," *Jurnal Ilmu Hukum(Justisia)*, vol. 3, 2019.
- D. Nugeraha Utama, *Logika Fuzzy untuk Model Penunjang Keputusan, Dilengkapi dengan Penerapan Contoh Kasus*. Yogyakarta: Garudhawaca, 2021. [Online]. Available: <https://www.researchgate.net/publication/353885471>
- A. Nalurita, "MEMPREDIKSI JUMLAH PRODUKSI IKAN TERI MENGGUNAKAN FUZZY LOGIC METODE TSUKAMOTO PADA CV. MAHERA," *Jurnal Ilmiah Teknik Informatika*, vol. 14, p. . 27-37, May 2020.
- Rahakbauw, Ilwaru, and Hahury, "IMPLEMENTASI FUZZY C-MEANS CLUSTERING DALAM PENENTUAN BEASISWA," *Jurnal Ilmu Matematika dan Terapan*, vol. 11, pp. 1–11, Mar. 2017.
- S. Zaenab et al., "IMPLEMENTASI FUZZY C-MEANS CLUSTERING DALAM PENGELOMPOKAN BUKU PADA PERPUSTAKAAN (Studi Kasus: SDN Srengat 01)," *Seminar Nasional Sistem Informasi*, vol. 2022, 2022.
- M. Rinaldi, A. Rmik, S. Hang, and T. Pekanbaru, "PENYEBARAN MAHASISWA BARU MENGGUNAKAN METODE FUZZY C-MEANS UNTUK Mencari DAERAH PROMOSI YANG POTENSIAL DISTRIBUTION OF NEW STUDENTS USING THE FUZZY C-MEANS METHOD TO LOOK FOR POTENTIAL PROMOTIONAL AREAS," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 3, no. 2, 2020.
- I. Ahmaddien and N. H. Sa'dia, "PENGARUH KEBIJAKAN UPAH MINIMUM TERHADAP PENYERAPAN TENAGA KERJA PEREMPUAN DI INDONESIA," *KOMITMEN: Jurnal Ilmiah Manajemen*, vol. 1, no. 1, Apr. 2020.