

Implementasi *K-Medoids Clustering* Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Kepesertaan Aktif Jaminan Sosial Tenaga Kerja

Yuni Sulistyawati^{1,*}

¹ Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

*Corresponding author: 20611141@students.uii.ac.id



P-ISSN: 2986-4178
E-ISSN: 2988-4004

Riwayat Artikel

Dikirim: 04 April 2024
Direvisi: 25 Juni 2024
Diterima: 23 Oktober 2024

ABSTRAK

Berdasarkan data dari Portal Data Ketenagakerjaan RI, jumlah peserta BPJS Ketenagakerjaan di Indonesia hingga Desember 2022 mencapai 55,38 juta orang, baik aktif maupun non-aktif. Ini menunjukkan pentingnya BPJS Ketenagakerjaan dalam melindungi kesejahteraan tenaga kerja. Program jaminan sosial sudah diatur dalam undang-undang, namun banyak pekerja yang belum terdaftar sebagai peserta atau peserta non-aktif. Oleh karena itu, diperlukan analisis untuk mengidentifikasi pengelompokan peserta aktif berdasarkan karakteristiknya. Salah satu cara untuk melakukan analisis tersebut adalah dengan menggunakan metode *clustering* atau pengelompokan data. Tujuan dari penelitian ini adalah untuk menganalisis secara deskriptif data kepesertaan dan mengelompokkan provinsi di Indonesia berdasarkan jumlah kepesertaan aktif menggunakan metode *K-Medoids Clustering*. Analisis ini menghasilkan 3 *cluster*, *cluster* 1 terdiri dari 29 provinsi dengan tingkat kepesertaan terendah, *cluster* 2 terdiri dari provinsi DKI Jakarta, dengan tingkat kepesertaan tertinggi, dan *cluster* 3 terdiri dari 4 provinsi yaitu Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten dengan tingkat kepesertaan sedang.

Kata Kunci: BPU, *Clustering K-Medoids*, Jaminan sosial, Jasa Konstruksi, PU

ABSTRACT

Based on the data from the Indonesian Employment Data Portal, the number of BPJS Employment participants in Indonesia as of December 2022 reached 55.38 million people, both active and inactive participants. This shows the importance of BPJS Employment in protecting the welfare of workers. The social security programs is regulated by law, but many workers have not registered as participant or are non-active participants. Therefore, analysis is needed to identify the grouping of active participants based on their characteristics. One way to carry out this analysis is to use the clustering or data grouping method. The aim of this research is to descriptively analyze membership data and group provinces in Indonesia based on the number of active participants using the *K-Medoids Clustering* method. This analysis produces 3 clusters, cluster 1 consists of 29 provinces with the lowest level of participation, cluster 2 consists of the province of DKI Jakarta, with the highest level of participation, and cluster 3 consists of 4 provinces, namely West Java, Central Java, East Java, and Banten with medium participation level.

Keywords: Non-wage earners, K-Medoids Clustering, Social security, Construction services, Wage earners.

1. Pendahuluan

Jaminan sosial tenaga kerja merupakan suatu sistem perlindungan yang memberikan manfaat bagi tenaga kerja dan keluarganya dari hal yang tidak terduga akibat resiko yang disebabkan selama bekerja. Seorang tenaga kerja dalam melaksanakan tugas-tugasnya, kemungkinan akan dihadapkan pada berbagai resiko yang dapat membahayakan kesehatannya, keselamatannya, keberlangsungan karirnya, dan nyawanya. Resiko yang dihadapi dapat meliputi penyakit atau gangguan kesehatan yang terkait dengan pekerjaan, kecelakaan yang terjadi di tempat kerja, kecacatan yang terjadi akibat kecelakaan atau penyakit, kehilangan pekerjaan secara tiba-tiba, dan kematian yang dapat terjadi akibat berbagai penyebab, termasuk kecelakaan atau penyakit yang terkait pekerjaan. Berdasarkan berlakunya Undang-undang Nomor 24 tahun 2011 tentang Badan Penyelenggara Jaminan Sosial, pelaksanaan jaminan sosial bagi tenaga kerja telah dialihkan dari PT.Jamsostek kepada PT.BPJS Ketenagakerjaan. Menurut Wijaya (2022), BPJS Ketenagakerjaan memiliki tanggung jawab dalam menyelenggarakan jaminan Kecelakaan Kerja (JKK), Jaminan Hari Tua (JHT), Jaminan Pensiun (JP), dan Jaminan Kematian (JKM).

Jumlah peserta BPJS Ketenagakerjaan di Indonesia hingga Desember tahun 2022 tercatat sebanyak 55,38 juta orang yang telah terdaftar, baik itu peserta aktif maupun non aktif. Angka ini menunjukkan pentingnya peran BPJS Ketenagakerjaan dalam memberikan perlindungan dan kesejahteraan bagi para tenaga kerja di Indonesia. Program-program jaminan sosial tenaga kerja telah diatur dalam peraturan perundang-undangan, namun masih terdapat banyak tenaga kerja yang belum terdaftar sebagai peserta atau peserta yang sudah terdaftar namun tidak aktif. Oleh karena itu, diperlukan suatu analisis untuk mengidentifikasi pengelompokan peserta aktif berdasarkan karakteristiknya. Adapun kepesertaan aktif terdiri dari Penerima Upah (PU), Bukan Penerima Upah (BPU), dan Jasa Konstruksi[1]. Salah satu cara untuk melakukan analisis tersebut adalah dengan menggunakan metode *Clustering* atau pengelompokan data.

Pada penelitian ini, peneliti akan melakukan pengelompokan terhadap Kepesertaan aktif untuk melihat Provinsi mana yang berpartisipasi tinggi dalam berkontribusi terkait jumlah kepesertaan aktif jaminan sosial tenaga kerja di Indonesia. Selain itu, dengan melakukan analisis *Clustering*, diharapkan dapat ditemukan pola atau karakteristik khusus dari kepesertaan aktif sehingga dapat dilakukan upaya peningkatan kepesertaan dan manfaat dari program jaminan sosial tenaga kerja.

Tujuan dari penelitian ini yaitu mengelompokkan 34 Provinsi di Indonesia berdasarkan jumlah kepesertaan aktif jaminan sosial tenaga kerja menggunakan metode K-Medoids. Terdapat asumsi yang harus dipenuhi sebelum melakukan analisis *Cluster* yaitu uji multikolinieritas yang bertujuan untuk menghindari pengelompokan yang tidak akurat, sebaiknya variabel dalam model tidak memiliki korelasi yang signifikan satu sama lain. Penyelesaian analisis *Clustering* pada data kepesertaan aktif ini menggunakan *software Rstudio*. Implementasi K-Medoids *Clustering* diharapkan dapat mempermudah proses pengelompokan serta mempermudah rencana peningkatan kepesertaan program jaminan sosial tenaga kerja di Indonesia.

Dalam melakukan penelitian, memerlukan dukungan dari penelitian sebelumnya sebagai acuan atau titik awal dalam menjalankan penelitian. Dukungan dari hasil-hasil penelitian sebelumnya juga dapat membantu peneliti dalam menentukan metode penelitian yang tepat, membangun hipotesis yang solid, dan merancang studi yang relevan. Penelitian yang pertama yaitu penelitian yang dilakukan oleh (Nahdliyah et al., 2019) dengan studi kasus “Metode *K-Medoids Clustering* Dengan Validasi *Silhouette Index* Dan *C-Index*”. Data yang digunakan dalam penelitian ini yaitu data jumlah kriminalitas di Jawa Tengah pada tahun 2018. Variabel penelitian yang digunakan pada penelitian ini berjumlah 11 variabel yang terdiri dari jenis-jenis kriminalitas. Metode yang digunakan adalah *Non-hierarchical clustering* menggunakan algoritma *K-Medoids* dengan validasi *Silhouette Index* Dan *C-Index*. Hasil penelitian menunjukkan pengelompokan ke dalam 4 *cluster* [2].

Penelitian kedua yaitu penelitian oleh (Fialine et.al., 2021) dengan judul “Implementasi Metode *K-Medoids Clustering* untuk Pengelompokan provinsi di Indonesia Berdasarkan Indikator Pendidikan”. Penelitian ini menggunakan 8 variabel indikator Pendidikan yaitu APK, RLS, APM, AM, APS, APtS, persentase guru layak, dan TMS. Metode yang digunakan adalah *K-Medoids Clustering*. Hasil penelitian menunjukkan adanya 3 *cluster* dengan tingkat indikator pendidikan tinggi, sedang, dan rendah. [3]

2. Metodologi Penelitian

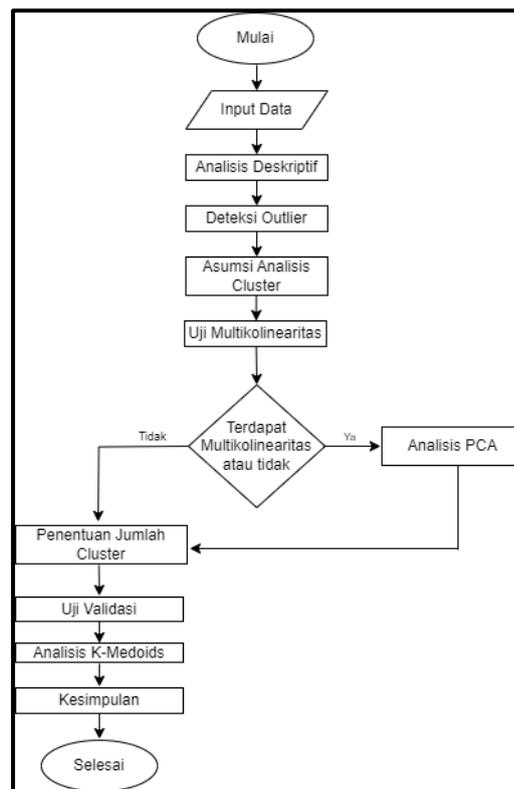
Data yang digunakan oleh peneliti dalam penelitian ini merupakan data sekunder mengenai kepesertaan aktif BPJS Ketenagakerjaan dengan periode sampai dengan bulan Desember 2022 yang diperoleh dari Portal Satu Data Ketenagakerjaan RI. Dalam penelitian ini, terdapat 3 variabel yang akan digunakan dalam analisis yaitu, Penerima Upah (PU), Bukan Penerima Upah (BPU), dan Jasa Kontruksi.

Tabel 1 Variabel yang dianalisis

| No | Nama Variabel | Definisi Variabel | Satuan |
|----|---------------------------|---|--------|
| 1 | Penerima Upah (PU) | Jumlah pekerja penerima upah yang terdaftar dalam kepesertaan aktif BPJS Ketenagakerjaan. | Jiwa |
| 2 | Bukan Penerima Upah (BPU) | Jumlah pekerja bukan penerima upah yang terdaftar dalam kepesertaan aktif BPJS Ketenagakerjaan. | Jiwa |
| 3 | Jasa Kontruksi | Jumlah pekerja jasa konstruksi yang terdaftar dalam kepesertaan aktif BPJS Ketenagakerjaan. | Jiwa |

Metode awal yang digunakan dalam penelitian ini adalah statistika deskriptif, yang bertujuan untuk memberikan gambaran yang sistematis dan objektif tentang data dengan menggunakan bentuk grafik maupun tabel [4]. Informasi yang disajikan dalam bentuk grafik dan tabel merupakan suatu hal yang penting dalam statistika deskriptif sehingga dapat membantu dalam memahami suatu permasalahan [5]. Grafik dan tabel membantu memvisualisasikan data sehingga mudah dipahami. Dengan menggunakan visualisasi grafik dan tabel, peneliti dapat mengidentifikasi hubungan-hubungan antar variabel serta menggambarkan karakteristik dari data [3]. Hasil dari analisis statistika deskriptif ini kemudian digunakan sebagai landasan untuk melanjutkan analisis pengelompokan *non-hierarki* yang bertujuan untuk mengelompokan provinsi di Indonesia berdasarkan jumlah kepesertaan aktif jaminan sosial tenaga kerja. Variabel yang digunakan dalam analisis ini meliputi, variabel Penerima Upah, Bukan Penerima Upah, dan Jasa Kontruksi. Data yang digunakan dalam penelitian ini adalah data pada tahun 2022 dan metode pengelompokan yang digunakan adalah metode *cluster K-Medoids*.

Dalam proses analisis, pengecekan *outlier* dilakukan menggunakan salah satu metode pengecekan *outlier* yaitu metode *Quan* (*Quantile-based method*). Metode tersebut merupakan metode yang bertujuan untuk menganalisis distribusi data berdasarkan kuantil dengan cara membagi data menjadi beberapa bagian yang sama besar, kemudian dengan metode *Quan* akan mengidentifikasi data yang menyimpang jauh dari data yang lain atau disebut dengan nilai ekstrim. Namun demikian, didalam analisis *K-Medoids* nantinya jika terdapat *outlier* pada data, data *outlier* tetap digunakan dalam analisis pembentukan *cluster* dikarenakan metode tersebut dirancang untuk mengatasi data yang mengandung *outlier*, metode *K-Medoids* memastikan bahwa hasil pengelompokan tetap representative terhadap keseluruhan data walaupun terdapat *outlier*. Penelitian ini direpresentasikan dalam bentuk *flow chart* yang menggambarkan tahapan penelitian sebagai berikut.



Gambar 1. Diagram Alir Penelitian

Dalam penelitian, peneliti menggunakan *software Rstudio* dan menerapkan metode *K-Medoids* untuk melakukan pengelompokan data. Proses pengelompokan dimulai dengan menginputkan data ke dalam *software Rstudio*. Selanjutnya, peneliti melakukan analisis deskriptif yang meliputi deteksi *outlier* didalamnya serta pengujian asumsi seperti uji multikolinieritas. Berikut ini merupakan langkah-langkah algoritma *K-Medoids* [2].

1. Menginputkan data kepesertaan aktif jaminan sosial tenaga kerja di Indonesia tahun 2022.
2. Melakukan analisis deskriptif pada data.
3. Melakukan pengecekan *outlier* untuk mengidentifikasi nilai-nilai *ekstrim* pada data menggunakan metode *Quan*.
4. Melakukan uji asumsi dalam analisis *cluster*, yaitu uji no-multikolinieritas.

Untuk mengetahui apakah asumsi multikolinieritas terpenuhi atau tidak dapat diperiksa dengan melihat nilai *VIF* (*Variance Inflation Factor*) [6]. Nilai *VIF* yang

kurang dari 10 menunjukkan bahwa tidak ada multikolinieritas dalam model. Berikut ini merupakan rumus VIF yang disajikan dalam Persamaan 5.

$$VIF = \frac{1}{1 - r^2} \tag{1}$$

$$r_{x_1x_2} = \frac{(n \sum x_1x_2 - (\sum x_1) \cdot (\sum x_2))}{\sqrt{(n \sum x_1^2 - (\sum x_1)^2) \cdot (n \sum x_2^2 - (\sum x_2)^2)}} \tag{2}$$

Keterangan:

VIF : Variance Inflation Factor

$r_{x_1x_2}$: Korelasi antar variabel x

n : Jumlah data yang digunakan

5. Menentukan jumlah *cluster* yang ingin dibentuk sebagai k menggunakan metode *Elbow* dan *Silhouette*. Metode *Elbow* menganalisis jumlah *cluster* dengan cara menghitung total *whitin-cluster sum of squares* (WCSS) dan memplot hasilnya. Saat jumlah *cluster* meningkat, WCSS akan menurun. Titik dimana penurunan plot secara drastis ini yang digunakan untuk menentukan jumlah *cluster*. Adapun menggunakan metode *Silhouette* dengan mengevaluasi seberapa baik data dikelompokkan dengan menghitung nilai yang berkisar antara -1 dan 1. Semakin tinggi nilainya semakin baik pengelompokkan *cluster*.
6. Membangkitkan k pusat *cluster* sebagai perwakilan (*medoid*) secara acak.
7. Menghitung jarak antara setiap objek *non-medoid* dengan *medoid* pada setiap *cluster*, kemudian menempatkan tiap objek *non-medoid* ke *medoid* terdekat dan hitung total jaraknya. Adapun rumus yang digunakan untuk menghitung jarak *Eulidean* disajikan pada Persamaan berikut.

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \tag{3}$$

Keterangan :

$d(x_i, x_j)$: Jarak antara objek ke-i dan objek ke-j

x_{ik} : Nilai objek ke-i pada variabel ke-k

x_{ij} : Nilai objek ke-i pada variabel ke-j

8. Memilih secara acak objek *non-medoid* dari setiap *cluster* sebagai kandidat *medoid* baru.
9. Menghitung jarak antara setiap objek *non-medoid* dengan kandidat *medoid* baru, kemudian menempatkan setiap objek *non-medoid* ke kandidat *medoid* terdekat dan hitung total jarak.

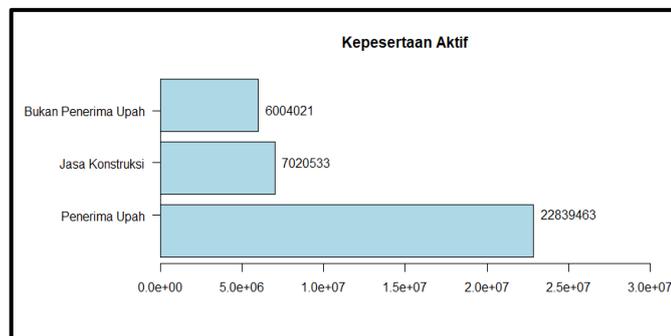
10. Menghitung selisih total jarak atau simpangan (S) dengan mengurangi total jarak pada kandidat *medoid* baru dengan total jarak pada *medoid* lama.
11. Jika $S < 0$, maka *medoid* baru diterima sebagai *medoid* yang baru. Jika $S > 0$, iterasi berhenti.
12. Mengulangi langkah 7 sampai 11 sehingga tidak ada perubahan *medoid* atau selisih total jarak > 0 .

Dengan melakukan langkah-langkah tersebut, algoritma *K-Medoids* akan mengoptimalkan pengelompokan objek ke dalam *cluster* dengan memperhatikan jarak antara objek dan *medoid*nya.

3. Hasil dan Pembahasan

3.1. Statistika Deskriptif

Statistika deskriptif digunakan untuk memberikan gambaran yang jelas dan terperinci tentang suatu fenomena dengan cara yang sistematis dan objektif dalam bentuk grafik untuk dianalisis dan ditafsirkan dengan penarikan kesimpulan. Dalam konteks penelitian ini, statistika deskriptif digunakan untuk memberikan gambaran umum atau ringkasan dari data jumlah kepesertaan Jaminan Sosial Tenaga Kerja di Indonesia berdasarkan Provinsi pada tahun 2022. Meskipun statistika deskriptif tidak menyediakan pengujian hipotesis atau inferensi statistik, namun tetap menjadi langkah awal yang penting dalam proses analisis data. Gambar berikut merupakan *Barchart* yang menggambarkan kepesertaan aktif jaminan sosial tenaga kerja di Indonesia.



Gambar 2. *Barchart* kepesertaan jaminan sosial tenaga kerja

Berdasarkan Gambar visualisasi *Barchart*, dapat disimpulkan bahwa kepesertaan aktif dalam jaminan sosial tenaga kerja di Indonesia terdiri dari 3. Pertama terdapat peserta penerima upah (PU) yang jumlahnya mencapai 22.839.463 orang. Peserta penerima upah merupakan yang terbanyak dari keseluruhan kepesertaan aktif. Diikuti oleh kepesertaan kategori jasa konstruksi yang memiliki jumlah peserta lebih sedikit dibandingkan peserta penerima upah, namun cukup signifikan dengan jumlah 7.020.533 orang. Sementara, peserta bukan penerima upah memiliki jumlah yang lebih rendah atau paling rendah yaitu sebanyak 6.004.021 orang. Dari data tersebut, dapat diketahui bahwa jumlah tenaga kerja yang terdaftar pada peserta penerima upah jauh lebih banyak dibandingkan dengan kategori lainnya. Peserta penerima upah memiliki kontribusi yang signifikan terhadap kepesertaan aktif jaminan sosial tenaga kerja di Indonesia. Selain dalam bentuk grafik, berikut adalah statistika deskriptif dalam ukuran pemusatan data dari data kepesertaan aktif jaminan sosial tenaga kerja yang dijelaskan secara numerik. Statistika deskriptif ini mencakup ukuran pemusatan data seperti nilai minimum, maksimum, dan nilai rata-rata setiap variabel kepesertaan aktif jaminan sosial tenaga kerja di Indonesia.

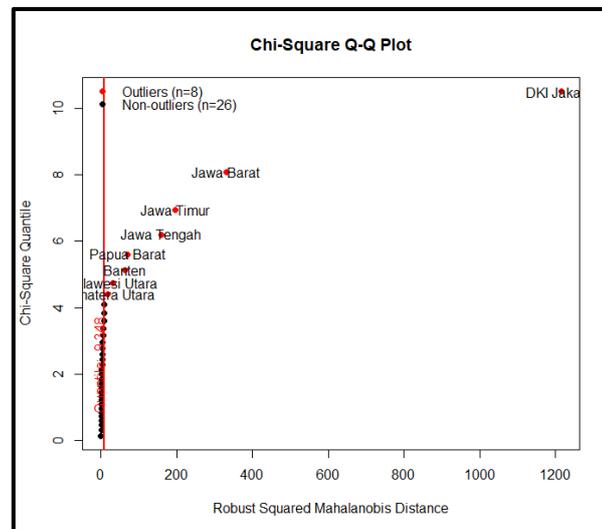
Tabel 2 Statistika Deskriptif

| Variabel | Minimum | Maximum | Mean |
|---------------------|---------|---------|--------|
| Penerima Upah | 63290 | 5059080 | 671749 |
| Bukan Penerima Upah | 18635 | 707363 | 176589 |
| Jasa Konstruksi | 20892 | 987602 | 206486 |

Berdasarkan tabel, menunjukkan hasil statistik deskriptif dari 3 variabel kepesertaan aktif jaminan sosial tenaga kerja di Indonesia. Tabel ini mencakup 3 jenis kepesertaan aktif yang meliputi Penerima Upah, Bukan Penerima Upah, dan Jasa Konstruksi. Statistika deskriptif yang ditampilkan meliputi nilai minimum, maksimum, dan nilai rata-rata (*mean*). Diketahui variabel Penerima Upah memiliki rata-rata peserta sebesar 671749 jiwa, dengan nilai maksimum 5059080 yang tercatat di provinsi DKI Jakarta. Sedangkan, nilai minimum sebesar 63290 peserta berada di Provinsi Sulawesi Barat. Untuk variabel bukan penerima upah, rata-rata pesertanya mencapai 176589 Jiwa, kemudian untuk nilai maksimum sebesar 707363 jiwa tercatat di Provinsi DKI Jakarta dan nilai minimum 18635 tercatat di Provinsi Maluku Utara. Terakhir rata-rata peserta jasa konstruksi adalah 206486 jiwa dengan nilai maksimum 987602 yaitu jumlah peserta yang berada di Provinsi Jawa Timur dan nilai minimum 20892 yang tercatat di Provinsi Kalimantan Utara.

3.2. Pengecekan Data Outlier

Sebelum melanjutkan analisis *clustering*, peneliti akan melakukan evaluasi terhadap data untuk mendeteksi adanya nilai yang berbeda jauh (*outlier*) pada Data Kepesertaan Aktif Jaminan Sosial Tenaga Kerja menggunakan metode *Quan*. Berikut ini disajikan hasil pemeriksaan *outlier* pada data.



Gambar 3. Deteksi Outlier pada data

Berdasarkan gambar dapat disimpulkan bahwa Data Kepesertaan Aktif Jaminan Sosial Tenaga Kerja di Indonesia pada tahun 2022 dari 34 Provinsi menunjukkan adanya 8 provinsi yang tergolong sebagai *outlier*. Provinsi-provinsi tersebut meliputi DKI Jakarta, Jawa Barat, Jawa Timur, Jawa Tengah, Papua Barat, Banten, Sulawesi Utara, dan Sumatera Utara. Sementara itu, 26 provinsi lainnya dapat dianggap sebagai data yang tidak termasuk dalam kategori *outlier*. Walaupun data tersebut mengandung data *outlier*, namun dalam metode *K-Medoids*, data *outlier* akan tetap dipertimbangkan dalam pembentukan *cluster* karena metode *K-Medoids* dirancang untuk mengatasi data yang mengandung *outlier* dengan cara yang lebih tahan terhadap pengaruhnya, salah satu faktornya yaitu jika kita bandingkan terhadap *K-Means* yang menggunakan rata-rata sebagai pusat *cluster*

sedangkan pada *K-Medoids* menggunakan *medoid* atau salah satu titik yang ada dalam *dataset* asli yang dapat meminimalkan jarak total terhadap titik-titik lain dalam *cluster*. Selain itu, karena *medoid* merupakan data asli jadi titik tersebut tidak begitu dipengaruhi oleh nilai ekstrem. Sehingga hasil analisis menggunakan *K-Medoids* tidak akan terlalu dipengaruhi oleh adanya *outlier*.

3.3. Uji Asumsi Cluster

3.3.1. Uji Multikolinieritas

Uji multikolinieritas digunakan untuk mengetahui apakah terdapat masalah multikolinieritas antar variabel independen. Untuk menguji multikolinieritas, peneliti menggunakan nilai VIF yang digunakan untuk melihat apakah terdapat hubungan atau korelasi antara variabel atau tidak. Dalam analisis multivariat, multikolinieritas antar variabel dapat menyebabkan pengaruh besar dalam menghasilkan solusi sehingga dapat mengganggu dalam proses analisis. Berikut ini merupakan nilai VIF dari 3 variabel yang digunakan dalam analisis.

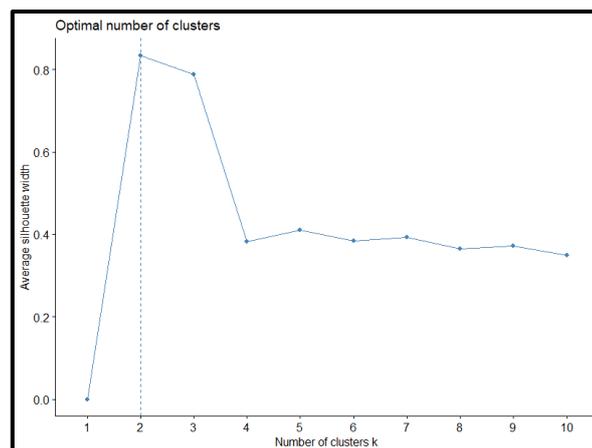
Tabel 3 Nilai VIF

| Variabel | Nilai VIF |
|---------------------|-----------|
| Penerima Upah | 6.715983 |
| Bukan Penerima Upah | 4.050455 |
| Jasa Konstruksi | 3.266368 |

Berdasarkan tabel menunjukkan bahwa nilai-nilai VIF yang diperoleh kurang dari 10, artinya tidak terjadi multikolinieritas sehingga asumsi *no-multikolinieritas* terpenuhi.

3.3.2. Penentuan Jumlah Cluster Optimum

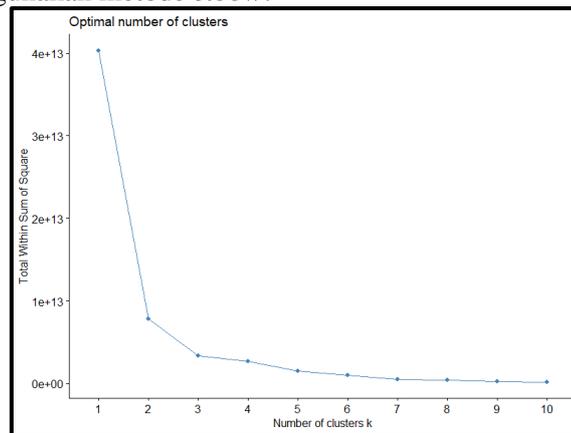
Untuk mengetahui bagaimana hasil pengclusteran atau pengelompokan jumlah kepesertaan Aktif jaminan tenaga kerja menurut provinsi pada tahun 2022 menggunakan metode *K-Medoids clustering*, sebelum melakukan analisis lebih lanjut, langkah awal yang dilakukan peneliti adalah menentukan jumlah *cluster* terbaik pada setiap metode. Terdapat beberapa cara untuk menentukan jumlah *cluster* terbaik, pada penelitian ini, peneliti menggunakan metode pendekatan *silhouette* dan *Elbow*.



Gambar 4. Plot *Silhouette*

Berdasarkan *Plot* metode *silhouette* pada Gambar 4 menunjukkan bahwa jumlah *cluster* terbaik yang disarankan untuk analisis *cluster* menggunakan 2 *cluster* ($K=2$), namun ada pilihan lain jika peneliti ingin menggunakan $K=3$ maka masih diperbolehkan dikarenakan dalam menentukan jumlah *cluster* metode *silhouette* ditunjukkan dengan nilai yang tertinggi atau dapat dikatakan bahwa nilai *silhouette* yang lebih tinggi secara umum menunjukkan partisi *cluster* yang lebih baik. Penggunaan metode *silhouette* didukung teknik atau metode evaluasi lainnya, seperti *elbow method* atau *gap statistics*. Maka dari

itu untuk memperkuat argument pemilihan jumlah *cluster* yang tepat, peneliti juga akan menganalisis menggunakan metode *elbow*.



Gambar 5. Hasil Plot WSS/ *Elbow method*

Berdasarkan hasil *Plot WSS* atau *elbow method* menunjukkan adanya titik patahan atau titik yang membentuk siku yang menandakan perubahan signifikan pada penurunan nilai *sum squared error* (SSE) saat jumlah *cluster* ditingkatkan. Titik patah terjadi pada nilai $k=3$. Pada analisis ini, peneliti memutuskan untuk menggunakan nilai K optimal = 3, karena terdapat siku atau patahan pada nilai k tersebut. Setelah nilai K optimal ditentukan, selanjutnya peneliti akan melakukan uji validasi indeks *silhouette* dan kemudian dilanjutkan dengan analisis *clustering K-Medoids* untuk mengelompokan data ke dalam *cluster* yang sesuai.

3.4. Uji Validasi Indeks *Silhouette*

Setelah menentukan jumlah *cluster* optimum, selanjutnya mengukur validitas dengan menggunakan validasi indeks *silhouette*. Berikut ini merupakan nilai koefisien *silhouette* (s_i) masing-masing objek hasil pengelompokan.

| Objek Ke- | Koefisien <i>Silhouette</i> (s_i) |
|-----------|---------------------------------------|
| 1 | 0.88537971 |
| 2 | 0.50413007 |
| 3 | 0.89621148 |
| 4 | 0.79978846 |
| 5 | 0.90944223 |
| 6 | 0.85302344 |
| 7 | 0.89132529 |
| 8 | 0.90741868 |
| 9 | 0.89161246 |
| 10 | 0.89666209 |
| 11 | 0.00000000 |
| 12 | 0.38980878 |
| 13 | 0.63637967 |
| 14 | 0.90137441 |
| 15 | 0.66088095 |
| 16 | -0.08756613 |
| 17 | 0.86783733 |
| 18 | 0.91181348 |
| 19 | 0.90643440 |
| 20 | 0.89269304 |
| 21 | 0.89736332 |

| Objek Ke- | Koefisien <i>Silhouette</i> (s_i) |
|-----------|---------------------------------------|
| 22 | 0.90041439 |
| 23 | 0.76795024 |
| 24 | 0.88963898 |
| 25 | 0.87493034 |
| 26 | 0.91162927 |
| 27 | 0.83317930 |
| 28 | 0.90170233 |
| 29 | 0.89255213 |
| 30 | 0.88742335 |
| 31 | 0.89391336 |
| 32 | 0.89194395 |
| 33 | 0.85979217 |
| 34 | 0.88522592 |

Berdasarkan tabel secara keseluruhan nilai koefisien *silhouette* (s_i) dari masing-masing objek lebih dari 0 sehingga dapat disimpulkan objek sudah berada dalam kelompok yang benar. Berdasarkan hasil perhitungan, terdapat satu objek yaitu objek ke-16 Provinsi Banten dengan nilai koefisien *silhouette* -0.08756613, ini menunjukkan bahwa objek tersebut memiliki tingkat kesesuaian yang lebih rendah dengan *clusternya* dibandingkan dengan *cluster* lainnya. Nilai koefisien *silhouette* yang negatif menunjukkan bahwa objek tersebut mungkin lebih cocok untuk ditempatkan dalam *cluster* yang berbeda. Namun dengan melihat konteks secara keseluruhan dan tidak hanya berfokus pada satu objek. Evaluasi peng-*clusteran* tidak hanya didasarkan pada satu objek atau nilai koefisien *silhouette* tunggal. Perlu dievaluasi secara keseluruhan terhadap kesesuaian *cluster*, termasuk melihat nilai koefisien *silhouette* objek lain dalam *cluster* yang sama. Objek ke 16 yang memiliki nilai koefisien *silhouette* negatif berada di *cluster* 3, mayoritas objek dalam *cluster* tersebut memiliki nilai koefisien *silhouette* yang tinggi dan hanya ada satu objek dengan nilai yang rendah, itu tidak selalu berarti peng-*clusteran* tidak sesuai. Objek tersebut merupakan *outlier* yang dapat dilihat pada pembahasan *outlier* pada data. Jadi dapat disimpulkan bahwa secara keseluruhan objek sudah berada dalam kelompok yang benar. Selanjutnya menghitung rentang nilai koefisien *silhouette* (SC). Dengan perhitungan persamaan 3.5 rentang nilai koefisien *silhouette* (SC) yang diperoleh yaitu sebesar 0.7883032, maka dapat disimpulkan bahwa struktur pengelompokan sudah sangat baik karena rentang nilai koefisien *silhouette* (SC) lebih besar dari 0.71

Analisis *cluster* menggunakan metode *K-Medoids* dengan menentukan K optimal sebanyak 3, maka didapatkan hasil pengelompokan dari data dan diperoleh hasil pusat *medoids* atau titik yang akan dijadikan *medoids* (perwakilan) adalah sebagai berikut.

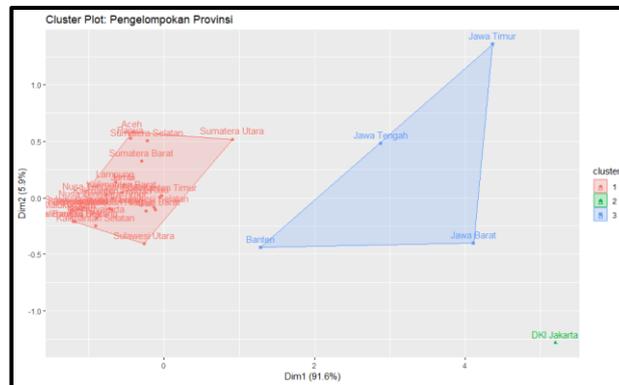
Tabel 5 Hasil Pusat Medoids

| Cluster | ID | Pusat Medoids | Bukan Penerima Upah | Penerima Upah | Jasa Konstruksi |
|---------|----|-----------------|---------------------|---------------|-----------------|
| 1 | 26 | Sulawesi Tengah | 254342 | 106047 | 97747 |
| 2 | 11 | DKI Jakarta | 5059080 | 707363 | 651237 |
| 3 | 13 | Jawa Tengah | 2279961 | 445815 | 656312 |

Dari hasil *output* Tabel 5 analisis *K-Medoids* disimpulkan bahwa objek yang dipilih sebagai *medoid* pada setiap *cluster* memiliki peran penting dalam menentukan pembentukan *cluster* yang akurat. Dalam hal ini, obyek ke-26, ke-11, dan ke-13 telah terpilih sebagai *medoid* untuk masing-masing *cluster*. Hal ini menunjukkan bahwa *medoid* pada setiap *cluster* didasarkan pada objek yang memiliki jarak terdekat dengan objek lain pada *cluster* tersebut. Perhitungan menggunakan jarak *Euclidean*, dapat memudahkan dalam menghitung jarak antar objek dan mempermudah dalam menentukan *medoid* pada setiap *cluster*.

3.5. Plot Hasil Cluster dengan K-Medoids

Analisis *cluster* menggunakan metode *K-Medoids* menghasilkan pengelompokan 34 Provinsi di Indonesia berdasarkan jumlah kepesertaan aktif jaminan sosial tenaga kerja, dengan menggunakan nilai $K=3$ atau jumlah *cluster* yaitu 3 (Rendah, Sedang, dan Tinggi) berdasarkan jenis kepesertaan aktifnya divisualisasikan dalam *Plot* berikut ini.



Gambar 6. Hasil plot *Cluster*

Berdasarkan Gambar 6 menunjukkan hasil visualisasi *Plot cluster K-Medoids*. *Plot* hasil *cluster* menunjukkan bahwa terdapat 3 *cluster* yang didapatkan berdasarkan hasil perhitungan similaritas antara kepesertaan di Indonesia. *Plot* hasil *cluster* terdapat 3 warna, yaitu warna merah menunjukkan hasil *cluster* 1, warna hijau menunjukkan hasil *cluster* 2, dan warna biru menunjukkan hasil *cluster* 3. Setiap warna memiliki ciri khas masing-masing. Pada gambar, sumbu x dan y memberikan keterangan bahwa hasil Dim 1 (91.6%) dan Dim 2 (5.9%), hal tersebut menunjukkan bahwa dua dimensi utama yang divisualisasikan dalam plot tersebut menjelaskan 97.5 % dari variabilitas data. Dim 1 atau dimensi 1 menjelaskan sebesar 91.6% data dari variabilitas yang berarti sebagian besar informasi dalam data ditangkap oleh dimensi 1 dan 5.9% pada dimensi 2 menjelaskan mengenai informasi tambahan namun dalam proporsi yang lebih kecil daripada dimensi 1. Kemudian, hasil anggota setiap *clusternya* akan dijabarkan per Provinsi dengan hasil sebagai berikut.

Tabel 6 Hasil *Cluster* Per-Provinsi

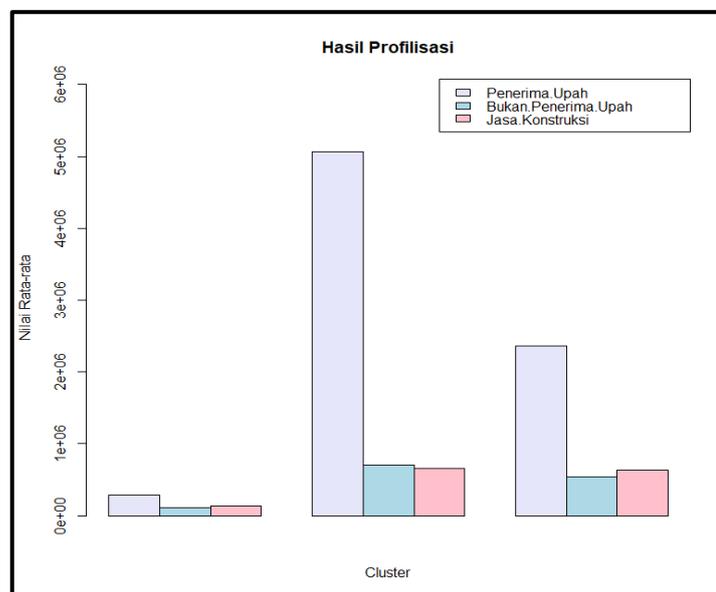
| <i>Cluster</i> | Jumlah Anggota | Kelompok <i>Cluster</i> |
|----------------|----------------|---|
| 1 | 29 | Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua |
| 2 | 1 | DKI Jakarta |
| 3 | 4 | Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten |

Berdasarkan tabel menunjukkan hasil pengelompokan provinsi di Indonesia berdasarkan kepesertaan aktif jaminan sosial tenaga kerja. Terlihat bahwa *cluster* 1 merupakan *cluster* yang memiliki anggota paling banyak yaitu 29, kemudian untuk *cluster* 2 memiliki anggota sebanyak 1 yaitu hanya provinsi DKI Jakarta, untuk *cluster* 3 memiliki anggota sebanyak 4 yaitu provinsi Jawa Barat, Jawa tengah, Jawa Timur, dan Banten. Kemudian berikut ini merupakan profilisasi kelompok dengan perhitungan rata-rata untuk masing-masing *cluster* dan mendapatkan hasil sebagai berikut :

Tabel 7

| Cluster | Bukan Penerima Upah | Penerima Upah | Jasa Konstruksi |
|---------|---------------------|---------------|-----------------|
| 1 | 287943. | 108678. | 131696. |
| 2 | 5059080 | 707363 | 651237 |
| 3 | 2357508. | 536247. | 637529. |

Berdasarkan Tabel 7 warna hijau menunjukkan tingkat kepesertaan aktif jaminan sosial tenaga kerja kategori tinggi, jumlah kepesertaan pada daerah tersebut sudah tinggi. Warna biru mengindikasikan tingkat kepesertaan aktif jaminan sosial tenaga kerja dengan kategori sedang yang berarti daerah-daerah dengan warna biru memiliki jumlah kepesertaan jaminan sosial yang sudah mencapai tingkat memadai atau baik. Terakhir, warna merah menunjukkan tingkat kepesertaan aktif jaminan sosial tenaga kerja yang rendah. Hal ini mengindikasikan bahwa daerah tersebut masih menghadapi tantangan dalam peningkatan jumlah kepesertaan jaminan sosial tenaga kerja. Kemudian untuk memudahkan melihat nilai rata-rata dari setiap variabel pada *cluster*-nya dapat menggunakan visualisasi *barplot* yang dapat dilihat pada gambar berikut.



Gambar 7. Hasil Profilisasi

Berdasarkan **Gambar 7.** menunjukkan visualisasi dalam bentuk *barchart* dari hasil profilisasi. Pada gambar terlihat bahwa pada setiap *cluster*, rata-rata tertinggi merupakan kepesertaan aktif penerima upah yang ditunjukkan dengan *bar* yang berwarna lavender. Pada *cluster* 1 dan 3 untuk rata-rata kepesertaan paling rendah ditunjukkan dengan kepesertaan bukan penerima upah yang diwakilkan oleh *bar* berwarna *lightblue*, sedangkan pada *cluster* 2 rata-rata terendah ada pada kepesertaan jasa konstruksi dengan *bar* yang berwarna pink. Dengan demikian, perlu adanya upaya yang lebih insentif dan strategi untuk meningkatkan kesadaran dan partisipasi masyarakat dalam program jaminan sosial tenaga kerja di daerah tersebut. Terdapat 3 kelompok atau *cluster* dengan karakteristik sebagai berikut.

1. *Cluster* 1 yang beranggotakan 29 provinsi memiliki karakteristik tingkat kepesertaan yang rendah pada semua variabel yaitu variabel penerima upah, bukan penerima upah, dan jasa konstruksi dibandingkan dengan *cluster* lainnya.
2. *Cluster* 2 yang beranggotakan 1 provinsi, yaitu DKI Jakarta memiliki karakteristik tingkat kepesertaan yang tinggi pada semua variabel dibandingkan dengan *cluster* lainnya.

3. *Cluster* 3 yang beranggotakan 4 provinsi, yaitu Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten memiliki karakteristik tingkat kepesertaan yang sedang pada semua variabel, lebih tinggi daripada *cluster* 1 dan lebih rendah dari *cluster* 2.

3.6. Visualisasi Dengan QGIS

Dalam rangka memperjelas dan memperdalam pemahaman terkait hasil dari profilisasi sebelumnya, akan disajikan visualisasi dengan bentuk Peta Indonesia menggunakan *software* QGIS. Visualisasi ini bertujuan untuk memberikan representasi grafis yang lebih jelas dan detail mengenai hasil *cluster*. Berikut ini merupakan tampilannya.



Gambar 8. Visualisasi hasil *cluster*

Berdasarkan Gambar 8. Menunjukkan visualisasi hasil *cluster* pada peta Indonesia, dapat disimpulkan bahwa provinsi-provinsi ditampilkan dengan menggunakan 3 warna yang berbeda, setiap warna menandakan adanya perbedaan karakteristik pada setiap daerah. Yang pertama warna hijau yang hanya dimiliki oleh provinsi DKI Jakarta, salah satu provinsi yang ada di Pulau Jawa (Terlihat lebih jelas pada gambar yang berada di pojok kiri bawah) yaitu provinsi yang masuk pada *cluster* 2 yang mengindikasikan bahwa pada provinsi tersebut, memiliki karakteristik tingkat kepesertaan yang tinggi. Warna biru menunjukkan provinsi-provinsi yang termasuk dalam *cluster* 2 dengan karakteristik tingkat kepesertaan yang sedang. Provinsi yang termasuk ke dalam *cluster* 2, keempatnya merupakan provinsi yang berada di Pulau Jawa yaitu, Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten. Terakhir untuk warna merah menunjukkan provinsi-provinsi yang termasuk dalam *cluster* 1 dengan karakteristik tingkat kepesertaan jaminan sosial tenaga kerja yang rendah.

4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, peneliti dapat mengambil kesimpulan bahwa:

1. Dari hasil analisis statistik deskriptif yang disajikan dalam bentuk visualisasi *barchart*, dapat disimpulkan bahwa provinsi DKI Jakarta merupakan provinsi dengan tingkat kepesertaan tertinggi dalam kategori penerima upah dan bukan penerima upah. Sementara itu, untuk kepesertaan jasa konstruksi, wilayah Jawa Timur menduduki peringkat tertinggi. Provinsi-provinsi dengan Tingkat kepesertaan terendah untuk penerima upah, bukan penerima upah, dan jasa konstruksi berturut-turut adalah wilayah Sulawesi Barat, Maluku Utara, dan wilayah Kalimantan Utara.
2. Dari analisis menggunakan metode *Cluster K-Medoids* didapatkan pengelompokan menjadi 3 *Cluster*, dengan *Cluster* 1 terdiri dari 29 provinsi, *Cluster* 2 hanya terdiri

dari 1 provinsi yaitu DKI Jakarta, dan *Cluster 3* terdiri dari 4 provinsi yaitu Jawa Barat, Jawa Tengah, Jawa Timur, dan Banten.

3. Berdasarkan hasil pengelompokan menjadi 3 *cluster*, dinyatakan bahwa tingkat kepesertaan aktif jaminan sosial tenaga kerja tertinggi terdapat pada *Cluster 2*, diikuti oleh *cluster 3* dengan tingkat kepesertaan aktif jaminan sosial tenaga kerja sedang, dan *Cluster 1* memiliki Tingkat kepesertaan terendah.

5. Saran

Adapun saran yang dapat peneliti sampaikan setelah dilakukannya penelitian ini adalah:

1. Upaya peningkatan kepesertaan di provinsi yang termasuk pada *cluster 1* diperlukan melalui sosialisasi dan pemberian informasi yang lebih aktif tentang program jaminan sosial tenaga kerja.
2. Pemerintah perlu mengevaluasi manfaat yang diterima peserta program jaminan sosial tenaga kerja dengan mengkaji ulang besaran manfaat dan melakukan penyesuaian untuk mengoptimalkan manfaat yang diterima peserta.

6. Daftar Pustaka

- [1] Kemenaker, *Permenaker_5_2021.pdf*. 2021. [Daring]. Tersedia pada: https://jdih.kemnaker.go.id/asset/data_puu/Permenaker_5_2021.pdf
- [2] M. A. Nahdliyah, T. Widiharih, dan A. Prahutama, "Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018)," *J. Gaussian*, vol. 8, no. 2, hal. 161–170, 2019.
- [3] D. A. Alodia, A. P. Fialine, D. Endriani, dan E. Widodo, "Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan," *Sepren*, vol. 2, no. 2, hal. 1–13, 2021, doi: 10.36655/sepren.v2i2.606.
- [4] S. E. Vivi Silvia, *Statistika Deskriptif*. Penerbit Andi, 2020.
- [5] H. Huwaida, *Statistika deskriptif*. Poliban Press, 2019.
- [6] S. Santoso, *Mahir Statistik Multivariat dengan SPSS*. Jakarta: PT ELex Media Komputindo, 2018.