

Pengelompokan Kabupaten/Kota di Kawasan Timur Indonesia Tahun 2021 Berdasarkan Indikator Sosial Ekonomi

Grouping of Regencies/Municipalities in Eastern Indonesia in 2021 Based on Socio-Economic Indicators

¹Annisa Firnanda*, ²Arie Wahyu Wijayanto

^{1,2}Politeknik Statistika STIS

Jalan Otto Iskandardinata No. 64C, Jatinegara, Jakarta, Indonesia

*e-mail: 211911031@stis.ac.id

(received: 14 Januari 2023, revised: 18 Maret 2023, accepted: 25 Maret 2023)

Abstrak

Perbedaan kondisi sosial, ekonomi, demografi dan sumber daya pada masing-masing daerah mampu menyebabkan ketimpangan, sehingga perlu mendorong pembangunan ekonomi sesuai kemampuan yang dimiliki oleh daerah tersebut. Pentingnya melihat indikator sosial ekonomi yang menjadi target pembangunan, khususnya di Kawasan Timur Indonesia (KTI). Penelitian ini menggunakan dua metode non hierarki, yaitu *K-Means* dan *K-Medoids*. Dalam penelitian ini dilakukan *Principal Component Analysis* (PCA) menghasilkan 3 faktor. Penentuan jumlah *cluster* menggunakan validitas internal dan stabilitas menunjukkan bahwa metode *K-Means* dengan jumlah *cluster* sebanyak 2 menghasilkan *cluster* yang paling optimal. *Cluster* 1 terdiri dari 152 kabupaten/kota, sedangkan *cluster* 2 terdiri dari 80 kabupaten/kota. *Cluster* 1 memiliki karakteristik sarana dan prasarana serta ekonomi di atas rata-rata, sementara kualitas manusia nya masih di bawah rata-rata.

Kata kunci: Sosial Ekonomi, *K-Means*, *K-Medoids*, Uji Validitas

Abstract

Differences in social, economic, demographic and resource conditions in each region can cause inequality, so it is necessary to encourage economic development according to the capabilities of the region. The importance of looking at socio-economic indicators that are development targets, especially in Eastern Indonesia (KTI). This study uses two non-hierarchical methods, namely *K-Means* and *K-Medoids*. In this study, *Principal Component Analysis* (PCA) was carried out to produce 3 factors. Determining the number of clusters using internal validity and stability shows that the *K-Means* method with a number of clusters of 2 produces the most optimal clusters. *Cluster* 1 consists of 152 regencies/municipalities, while *cluster* 2 consists of 80 regencies/municipalities. *Cluster* 1 has above average infrastructure and economic characteristics, while its human quality is still below average.

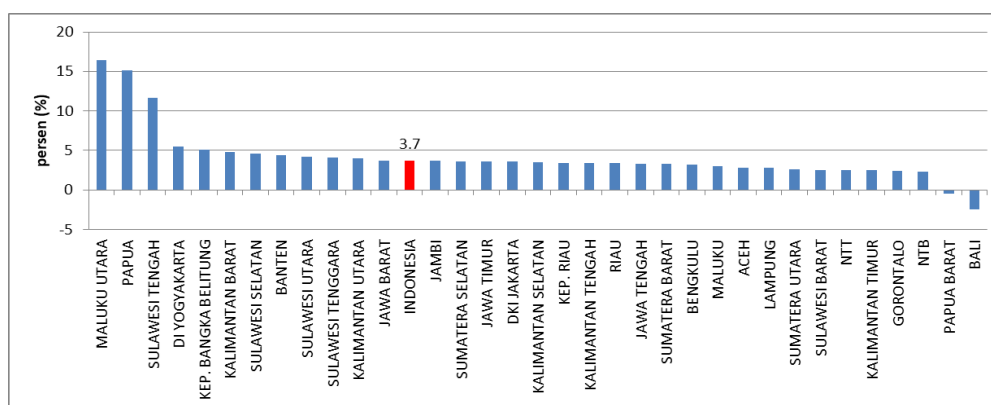
Keywords: Socioeconomics, *K-Means*, *K-Medoids*, Validity Test

1 Pendahuluan

Dalam mendukung perekonomian suatu negara, dibutuhkan pembangunan ekonomi yang mampu menyediakan infrastruktur yang memadai, meningkatkan pendidikan masyarakatnya, tersedianya kesempatan kerja, serta meningkatkan pendapatan guna terciptanya kesejahteraan masyarakat. Adanya ketidakmerataan atau kesenjangan distribusi pendapatan yang berdampak pada kemiskinan merupakan salah satu masalah yang sering dihadapi oleh kebanyakan negara-negara berkembang, termasuk Indonesia [1]. Perbedaan kondisi sosial, ekonomi, demografi dan sumber daya pada masing-masing daerah mampu menyebabkan ketimpangan, sehingga perlu mendorong pembangunan ekonomi sesuai kemampuan yang dimiliki oleh daerah tersebut [2]. Potensi daerah Indonesia berbeda baik secara geografis maupun sumber daya alamnya, sehingga daerah yang berpotensi tinggi lebih maju dibandingkan daerah lainnya [3].

Secara makro, ketimpangan pembangunan di Indonesia dipengaruhi oleh kesenjangan alokasi sumber daya, baik sumber daya alam maupun manusia, teknologi, fisik serta kapital [2]. Ketimpangan tersebut tidak dapat dihilangkan, namun dapat dikurangi agar terjadi keselarasan dalam pertumbuhannya [4]. Ketimpangan atau kesenjangan pendapatan yang semakin tinggi mengisyaratkan bahwa pertumbuhan ekonomi yang tinggi hanya dinikmati golongan dengan pendapatan tinggi dari total pendapatan [1]. Ukuran merata atau tidaknya distribusi pendapatan antar penduduk dapat dilihat dari angka *gini ratio*. Angka *gini ratio* yang tinggi artinya tingkat ketimpangan di suatu daerah juga semakin tinggi, begitu pula sebaliknya.

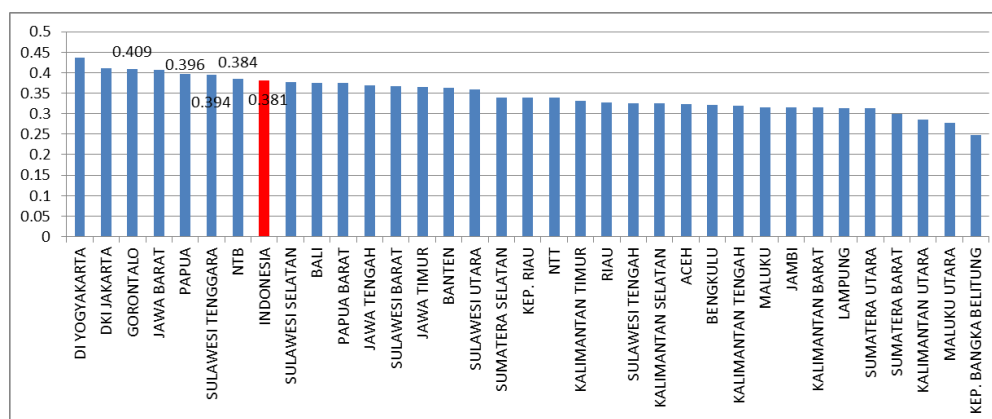
Berdasarkan tiga pilar utama tujuan *Sustainable Development Goals* (SDGs), meliputi *Human Development*, *Social Economic Development*, dan *Environmental Development*, kemiskinan menjadi isu yang perlu diperhatikan [5]. Meningkatkan kesejahteraan masyarakat merupakan tujuan dari pembangunan ekonomi itu sendiri, sehingga perlu adanya pertumbuhan ekonomi yang meningkat dan distribusi pendapatan yang merata. Pertumbuhan ekonomi, struktur ekonomi dan ketimpangan pembangunan yang semakin kecil merupakan indikator keberhasilan dalam pembangunan daerah [6].



Sumber: BPS (2021), diolah

Gambar 1. Pertumbuhan Ekonomi Provinsi di Indonesia tahun 2021

Pertumbuhan ekonomi dapat diartikan sebagai peningkatan kemampuan perekonomian suatu negara untuk menghasilkan barang dan jasa [7]. Jika dilihat pada Gambar 1, wilayah-wilayah dengan pertumbuhan ekonomi di atas rata-rata nasional, secara umum berada pada Kawasan Timur Indonesia (KTI). Namun, pertumbuhan ekonomi tanpa pemerataan ekonomi akan memperlebar kesenjangan antar kelompok masyarakat. Meski memiliki angka pertumbuhan ekonomi yang tinggi, namun beberapa provinsi di KTI memiliki angka *gini ratio* diatas rata-rata nasional yang terlihat pada Gambar 2, yakni Gorontalo, Papua, Sulawesi Tenggara dan NTB masing-masing 0.409, 0.396, 0.394, 0.384 dan 0.381.



Sumber: BPS (2021), diolah

Gambar 2. Gini Ratio Provinsi di Indonesia tahun 2021

Sebagai roda pertumbuhan ekonomi, infrastruktur memegang peranan penting dalam pembangunan daerah. Hal ini karena pembangunan ekonomi memerlukan sarana dan prasarana yang memadai [8]. Tak hanya itu, dalam mendukung pembangunan ekonomi, modal manusia (*human capital*) berperan penting. Adanya pemerataan dalam penyelenggaraan pendidikan merupakan salah satu target SDGs. Pendidikan yang merata, dapat meningkatkan kesempatan memperoleh pekerjaan dan penghasilan yang mampu meningkatkan kesejahteraan masyarakat [9].

Berdasarkan kawasan di Indonesia, dari tahun 2015-2019 Rata-rata Lama Sekolah (RLS) di KTI selalu lebih rendah dari RLS di Kawasan Barat Indonesia (KBI). Hal ini mengindikasikan bahwa adanya ketimpangan pencapaian pendidikan di KTI. Rendahnya RLS KTI diikuti dengan rendahnya pendapatan perkapita dan persentase penduduk miskin yang tinggi di wilayah KTI [10]. Namun, berdasarkan penelitian yang dilakukan oleh Kapriaji dan Siswidiyanto [11], rasio ketersediaan SMA/MA di KTI lebih baik dibandingkan wilayah KBI. Hal ini berarti jumlah SMA/MA yang ada di KTI lebih banyak dibandingkan di KBI. Sementara itu, jumlah murid per SMA/SMK/MA Negeri lebih besar di KBI. Dapat disimpulkan jika di KTI sudah ada sekolah yang menampung, namun kekurangan murid yang ingin bersekolah.

Dalam hal ini, perlu perhatian pemerintah dalam meningkatkan kesejahteraan masyarakat di KTI dengan adanya pengelompokan kabupaten/kota berdasarkan kondisi sosial ekonomi sehingga dapat mendukung pemerintah dalam pengambilan kebijakan pembangunan dan mewujudkan tujuan SDGs. Oleh karena itu, dalam penelitian ini akan dilakukan perbandingan performa metode *clustering* yaitu *K-Means* dan *K-Medoids* dalam mengelompokkan kabupaten/kota di KTI tahun 2021 berdasarkan indikator sosial ekonomi dengan menggunakan validitas internal serta stabilitas.

2 Tinjauan Literatur

Penelitian dengan menggunakan analisis *cluster* pernah dilakukan oleh Tirayo [5] untuk mengelompokkan kabupaten/kota di Provinsi Sulawesi Tengah menurut Indikator Sosial Ekonomi menggunakan *K-Medoids Clustering*. Dalam penelitiannya, menghasilkan 2 *cluster* dimana *cluster* pertama merupakan wilayah yang memiliki rata-rata tingkat kemiskinan rendah, Indeks Pembangunan Manusia (IPM) tinggi, serta TPT yang tinggi, begitupun sebaliknya untuk *cluster* ke-2.

Selain itu, Herviany et al. [12] membandingkan algoritma *K-Means* dan *K-Medoids* dalam mengelompokkan daerah rawan tanah longsor di Provinsi Jawa Barat. Diperoleh hasil pada data kejadian tanah longsor di Provinsi Jawa Barat, metode *K-Means* lebih optimal dibandingkan *K-Medoids* dengan jumlah *cluster* sebanyak 6. Namun, dalam penelitian Luthfi dan Wijayanto [13] yang mengelompokkan Indeks Pembangunan Manusia (IPM), metode *K-Medoids* lebih unggul jika dibandingkan dengan metode *K-Means* maupun Hierarki. Hal ini terlihat dari ukuran validitas dari nilai indeks *Dunn*, *Davies-Bouldin* dan *Calinski-Harabaz* serta nilai rasio simpangan baku dalam kelompok yang minimum dan antar kelompok yang maksimum.

Penelitian lain yang membandingkan algoritma *K-Means* dan *K-Medoids* dilakukan oleh Faisal et al. [14] dalam mengelompokkan suhu di Provinsi Riau. Dalam penelitiannya menggunakan validitas DBI dalam membandingkan kedua metode dimana nilai DBI yang semakin kecil menunjukkan bahwa semakin baik *cluster* yang dihasilkan. Didapatkan bahwa algoritma *K-Means* memiliki nilai DBI terbaik, yaitu 0.2 dengan jumlah *cluster* sebanyak 6 dan 10 iterasi, sedangkan algoritma *K-Medoids* memperoleh nilai terbaik sebesar 0.279 dengan jumlah *cluster* sebanyak 8 dan 100 iterasi.

Pada penelitian yang membandingkan algoritma *K-Means* dan *K-Medoids* lainnya juga dilakukan untuk mengelompokkan wilayah sebaran cacat pada anak yang menghasilkan bahwa *K-Medoids* lebih baik dalam melakukan pengelompokan pada data sebaran anak cacat dibandingkan algoritma *K-Means* karena berdasarkan nilai validitas indeks *Silhouette*, *K-Medoids* lebih besar yaitu 0.5009 [15].

Penelitian dengan membandingkan metode membandingkan metode *clustering* juga dilakukan oleh Sikana dan Wijayanto [16] yang mengelompokkan IPM kabupaten/kota di Indonesia tahun 2019 dengan membandingkan metode *partitioning* (non hierarki) dan hierarki. Diperoleh hasil bahwa metode *K-Means* dengan jumlah *cluster* optimum sebanyak 6 dengan menggunakan indeks *Silhouette*, *Calinski-Harabaz*, dan *Davies-Bouldin*.

Berdasarkan uraian diatas, penelitian ini akan membandingkan metode *K-Means* dan *K-Medoids* dalam mengelompokkan kabupaten/kota di KTI tahun 2021 berdasarkan indikator sosial ekonomi. Kedua algoritma tersebut akan dibandingkan performanya dengan cara uji validitas internal dan

stabilitas guna menemukan algoritma dengan *cluster* terbaik. Penelitian ini diharapkan mampu menjadi dasar bagi pemangku kebijakan dalam pengimplementasian kebijakan tepat sasaran. Penelitian ini lebih mencakup keseluruhan kabupaten/kota di KTI, bukan hanya satu provinsi saja sehingga lebih terperinci dengan harapan mampu menggambarkan keadaan sosial ekonomi di KTI yang membandingkan algoritma terbaik terlebih dahulu.

3 Metode Penelitian

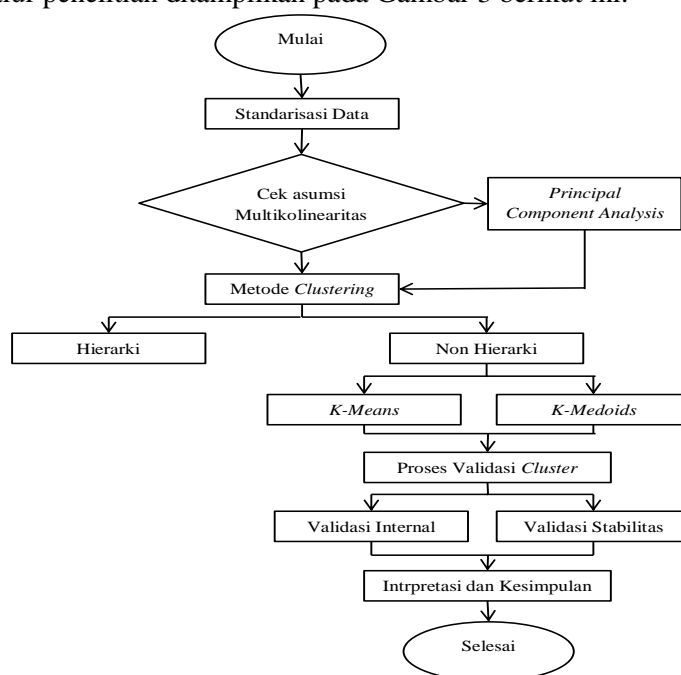
Penelitian ini merupakan penelitian kuantitatif dengan melakukan perbandingan dua algoritma *clustering* yaitu *K-Means* dan *K-Medoids* dengan menggunakan ukuran validitas internal maupun stabilitas. Data yang digunakan pada penelitian ini diambil dari tabel dinamis pada website Badan Pusat Statistika (BPS) dan Publikasi Provinsi dalam Angka tahun 2021. Unit analisis dalam Unit analisis dalam penelitian ini meliputi 232 kabupaten/kota di 17 provinsi yang tergabung dalam Kawasan Timur Indonesia (KTI).

Adapun provinsi-provinsi yang tergabung dalam KTI meliputi Provinsi Maluku, Maluku Utara, Sulawesi Utara, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Tenggara, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kalimantan Timur, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Gorontalo, Papua dan Papua Barat. Dataset, *R code* dan hasil *clustering* dalam penelitian dapat diakses melalui <https://github.com/annisafirnanda15/Paper-Annisa-Firnanda-Sistemasi.git>. Indikator-indikator dalam penelitian ini terlihat pada tabel 1 berikut.

Tabel 1. Indikator-Indikator Penelitian

Nama Variabel	Keterangan Variabel	Tipe Data
PE	Pertumbuhan Ekonomi	Numerik
TPAK	Tingkat Partisipasi Angkatan Kerja	Numerik
RASIOPUSKES	Rasio Puskesmas	Numerik
RASIOSMA	Rasio SMA	Numerik
AHH	Angka Harapan Hidup	Numerik
HLS	Harapan Lama Sekolah	Numerik
RLS	Rata-rata Lama Sekolah	Numerik
PENG	Pengeluaran Pemerintah	Numerik

Adapun diagram alur penelitian ditampilkan pada Gambar 3 berikut ini.



Gambar 3. Diagram Alur Penelitian

1) Analisis Faktor

Analisis faktor mencakup analisis komponen utama dan analisis faktor umum. Analisis faktor merupakan pendekatan statistik yang digunakan untuk menganalisis "interaksi" dengan jumlah variabel yang besar dan menginformasikan dimensi mana yang mendasari variabel tersebut [17]. Untuk mereduksi jumlah variabel yang terlalu banyak dan berkorelasi tinggi, digunakanlah analisis faktor dengan mengelompokkan variabel-variabel baru yang jumlahnya lebih sedikit, namun masih mengandung sebagian besar informasi dari variabel aslinya [18]. Berikut merupakan tahapan analisis faktor secara umum menurut Supranto [18]:

1. Pemeriksaan matriks korelasi

Sebelum melakukan analisis faktor, digunakan *Bartlett's test of sphericity* untuk menguji ada atau tidak korelasi antar variabel. Hasil yang signifikan menunjukkan bahwa matriks korelasi yang dihasilkan tidak sama dengan matriks identitas ditunjukkan jika $\chi_{hitung}^2 > \chi_{\alpha; \frac{1}{2}p(p-1)}^2$ atau $p - value < \alpha$ dengan hipotesis sebagai berikut.

$$H_0: \rho = I_{p \times p} \text{ (matriks korelasi adalah matriks identitas)}$$

$$H_1: \rho \neq I_{p \times p} \text{ (matriks korelasi bukan matriks identitas)}$$

Adapun statistik ujinya sebagai berikut.

$$\chi_{hitung}^2 = - \left[n - \frac{1}{6}(2p + 5) \right] \ln |R| \sim \chi_{\alpha; \frac{1}{2}p(p-1)}^2 \quad (1)$$

Dengan keterangan: n adalah jumlah amatan, p adalah jumlah variabel/indikator dan R adalah matriks korelasi

Selain itu, jika dilihat dari matriks korelasi, adanya indikasi multikolinearitas ditunjukkan pada nilai korelasi yang tinggi antar variabel yaitu di atas 0.8. Selanjutnya digunakan *Kaiser-Meyer-Olkin (KMO)* yang merupakan ukuran kelayakan suatu data untuk dilakukan analisis faktor. KMO juga dilakukan untuk melihat kecukupan sampel yang memilih rentang dari 0 hingga 1 [19]. Jika nilai KMO lebih besar dari 0.5, analisis faktor dapat digunakan.

Hipotesis:

H_0 : data yang digunakan layak untuk analisis faktor

H_1 : data yang digunakan tidak layak untuk analisis faktor

Statistik uji dalam uji KMO:

$$KMO = \frac{\sum_j \sum_{j \neq k} r_{jk}^2}{\sum_j \sum_{j \neq k} r_{jk}^2 + \sum_j \sum_{j \neq k} q_{jk}^2} \quad (2)$$

Dengan keterangan: r_{jk}^2 adalah koefisien korelasi antara variabel j dan k dan q_{ij}^2 adalah koefisien korelasi parsial antara variabel j dan k

2. Ekstraksi Faktor

Apabila data sudah layak untuk dianalisis faktor, lakukan pemilihan metode untuk mengekstraksi faktor. Langkah ini dilakukan untuk mereduksi beberapa variabel menjadi suatu himpunan variabel baru yang dinamakan faktor. Salah satu metode yang umum digunakan yaitu *Principal Component Analysis (PCA)* yang dapat membentuk faktor sesedikit mungkin dengan mempertimbangkan jumlah ragam maksimal dari data [20].

Banyak faktor dapat ditentukan dengan beberapa kriteria seperti apriori yang merupakan banyak faktor berdasarkan penelitian sebelumnya, nilai eigen yang merupakan faktor dengan nilai eigen lebih dari satu akan dipilih, dan *scree plot* merupakan melihat titik dimana kurva mulai melandai dengan nilai eigen pada sumbu y dan banyak faktor pada sumbu x .

3. Rotasi Faktor

Proses memutar sumbu faktor dari kedudukan semula sehingga variabel asal memiliki korelasi tinggi terhadap faktor tertentu saja disebut rotasi faktor. Rotasi tersebut dilakukan jika faktor dari hasil analisis masih sulit untuk diinterpretasikan.

4. Pembentukan skor faktor

Skor faktor dapat digunakan sebagai pengganti dari variabel asli pada analisis lanjutan untuk analisis multivariate.

2) Analisis Cluster

Analisis *cluster* dilakukan dengan tujuan untuk mengelompokkan individu maupun objek-objek ke dalam suatu kelompok dengan karakteristik yang relatif sama, sehingga dapat mengetahui perbedaan karakteristik antar kelompok satu dengan lainnya secara rinci. Anggota dalam kelompok bersifat homogen, sedangkan antarkelompok bersifat heterogen [19].

Umumnya, ada dua metode yang digunakan dalam analisis *cluster*, yaitu metode hierarki dan non hierarki [19]. Metode hierarki biasanya digunakan dalam mengelompokkan pengamatan secara terstruktur berdasarkan kemiripan sifat serta banyaknya kelompok yang diinginkan belum diketahui. Hal ini bisa melalui penggabungan atau *agglomerative* dan pemisahan atau *divisive*. Sementara itu, metode non hierarki digunakan untuk mengelompokkan pengamatan secara tidak teratur dan banyak kelompok yang diinginkan telah ditentukan terlebih dahulu.

3) Metode K-Means

Salah satu metode non hierarki yang sering digunakan yaitu metode *K-Means clustering* yang bertujuan untuk mengelompokkan objek sedemikian rupa sehingga jarak antar objek ke pusat kelompok (*centroid*) menjadi minimum. Algoritma *K-Means* memiliki kemampuan dalam mengelompokkan jumlah data yang besar dengan waktu yang cepat dan efisien. Algoritma *K-Means* dengan jumlah *cluster* yang berbeda lebih sering digunakan karena kesederhanaan dan kompleksitas komputasi yang rendah. Algoritma *K-Means* sensitif terhadap *outlier* karena benda-benda tersebut berada jauh dari pusat data secara umum, sehingga ketika *K-Means* digunakan, maka dapat mengubah nilai rata-rata dari *cluster* [13]. Secara umum, algoritma *clustering* dengan *K-Means* sebagai berikut [21].

1. Tentukan banyak *cluster* yang ingin dibentuk.
2. Data dialokasikan ke dalam *cluster* secara random.
3. Hitung jarak setiap data ke masing-masing *centroid* dengan rumus korelasi antar dua objek, seperti *Euclidean Distance*. Rumusnya sebagai berikut.

$$d(x, y) = \sqrt{(x_i - s_i)(y_i - t_i)} \quad (3)$$

Dengan keterangan:

d : *Euclidean Distance*
 x, y : titik koordinat observasi
 s, t : titik koordinat *centroid*
 i : jumlah observasi

4. Hitung pusat *cluster* baru dengan anggota *cluster* sekarang. Pusat *cluster* merupakan rata-rata dalam sebuah *cluster* dengan rumus berikut [22].

$$V_{ij} = \frac{1}{N} \sum_{k=0}^{N_i} X_{kj} \quad (4)$$

Dengan keterangan:

V_{ij} : Centroid rata-rata *cluster* ke-i untuk variabel ke-j
 N_i : jumlah anggota *cluster* ke-i
 i, k : Indeks dari variabel
 j : Indeks dari variabel
 X_{kj} : Nilai dari data ke-k variabel ke-j untuk *cluster* tersebut

5. Ulangi langkah ketiga dan keempat hingga *centroid* bernilai optimal.

4) Metode K-Medoids

K-Medoids berusaha meminimumkan *sum of dissimilarities* antar data di sebuah *cluster* dan memilih sebuah data di dalam *cluster* sebagai pusat (*medoids*) [13]. Metode *K-Medoids* lebih kuat daripada *K-Means* pada data yang terdapat *noise* dan *outlier*. *K-Medoids clustering* atau biasa dikenal sebagai *Partitioning Around Medoids* (PAM) merupakan algoritma yang lebih tahan terhadap *outlier* jika dibandingkan dengan algoritma *K-Means*. Algoritma *K-Medoids* dimulai dengan memilih titik acak sebagai pengganti centroid yang representatif. Kemudian algoritma *K-Medoids* dilakukan secara *iteratif* hingga diperoleh objek sebenarnya yang mewakili kelompok. Algoritma *clustering* dengan *K-Medoids* secara umum adalah sebagai berikut [23].

1. Tentukan jumlah *cluster* sebanyak k dan pusat *cluster* secara acak.
2. Alokasikan setiap observasi ke dalam *cluster* dengan *Euclidean Distance* terdekat.
3. Pilih observasi secara acak pada tiap *cluster* sebagai kandidat *medoids* yang baru.
4. Hitung jarak dari setiap observasi pada tiap-tiap *cluster* dengan kandidat *medoids* yang baru.
5. Hitung total simpangan (S) yang merupakan selisih dari total *distance* baru dan total *distance* lama. Apabila $S < 0$, maka tukar observasi tersebut untuk membentuk sekumpulan k observasi baru sebagai *medoids*.
6. Ulangi langkah 2 sampai 5 hingga tidak terjadi perubahan *medoids*.

5) Validitas *Cluster*

Validitas hasil analisis *cluster* dilakukan untuk mendapatkan partisi terbanyak. Sesuai dengan data. Kegagalan validasi *cluster* akan memengaruhi hasil analisis [24]. Dalam penelitian ini, validitas *cluster* yang digunakan, yaitu validitas internal dan validitas stabilitas dengan penjelasan sebagai berikut.

1. Indeks *Connectivity*

Nilai Indeks *Connectivity* berkisar antara 0 sampai tak hingga. Semakin kecil nilai dari indeks *Connectivity*, maka akan semakin baik *cluster* yang terbentuk. Konektivitasnya adalah sebagai berikut.

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L x_{i,nni(j)} \quad (4)$$

Dengan keterangan: L merupakan parameter yang menentukan jumlah tetangga yang berkontribusi pada ukuran konektivitas. Koefisien konektivitas memiliki nilai antara nol dan satu dan nilainya harus diminimalkan.

2. Indeks *Silhouette* (SI)

Indeks *Silhouette* digunakan untuk melakukan evaluasi pengelompokan pada setiap *cluster*. SI memperoleh jumlah *cluster* optimal dengan menghitung selisih antara jarak rata-rata di dalam *cluster* dengan jarak minimum antar *cluster* [25].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

Dengan keterangan:

$a(i)$: jarak rata-rata antara i dan seluruh pengamatan lainnya pada *cluster* yang sama

$b(i)$: jarak rata-rata antara i dengan pengamatan pada *cluster* terdekat

Jika nilai SI mendekati positif 1, maka dapat diartikan bahwa bahwa suatu titik data telah berada pada kluster yang tepat..

3. Indeks *Dunn*

Indeks *Dunn* merupakan salah satu metode validitas *cluster* yang membandingkan jarak terkecil antara pengamatan yang tidak sama antar dua *cluster* ke jarak *intracluster* terbesar dengan nilai indeks yang semakin besar menunjukkan jumlah *cluster* terbaik. Indeks *Dunn* adalah perbandingan jarak terkecil antara pengamatan pada *cluster* yang berbeda dengan jarak terbesar pada setiap *cluster* data [26].

$$C = \frac{d_{min}}{d_{max}} \quad (6)$$

Dengan keterangan:

d_{min} : jarak terkecil antara observasi pada *cluster* yang berbeda

d_{max} : jarak terbesar pada masing-masing *cluster* data.

Indeks *Dunn* rentan terhadap *outlier* dan *noise* (Dunn, 1973 dalam Luthfi & Wijayanto) [13].

4. Indeks *Davies-Bouldin* (DBI)

Untuk menghitung rata-rata nilai setiap titik yang ada di himpunan data digunakanlah Indeks *Davies-Bouldin* (DBI). Pada setiap titik, perhitungan nilainya merupakan jumlah *compactness* yang dibagi dengan jarak antara kedua titik pusat *cluster* sebagai pemisah. Semakin kecil nilai DB, semakin baik *cluster* yang dihasilkan (Davies, 1979 dalam Luthfi & Wijayanto) [13].

5. Validitas Stabilitas

Pada validitas stabilitas, cara yang dilakukan yaitu dengan membandingkan hasil dari analisis *cluster* pada penghapusan setiap kolom data satu per satu. Pada validitas ini, nilai yang digunakan adalah *average proportion of non-overlap* (APN), *average distance* (AD), *average*

distance between means (ADM), dan *figure of merit* (FOM). Apabila keempat nilai tersebut semakin kecil, maka semakin baik *cluster* yang diperoleh [24].

4 Hasil dan Pembahasan

1) Deskripsi Statistik

Tabel 2. Deskripsi Statistik

Indikator	Min	Max	Mean	Median
PE	-2.2900	162.0000	4.2012	3.2300
TPAK	2.1200	97.9300	63.0194	68.8900
RASIOPUSKES	0.0045	0.3000	0.0733	0.0604
RASIOSMA	0.0041	0.2899	0.0733	0.0579
AHH	55.4300	74.7600	68.2299	68.4250
HLS	3.8700	16.8900	12.7509	12.8350
RLS	1.4200	12.5100	8.1394	8.1550
PENG	82.0	16765.0	8743.6422	8834.5

Sumber: BPS (2021), diolah

Tabel 2 menunjukkan deskripsi dari delapan indikator penelitian yang digunakan, mulai dari nilai minimum, nilai maksimum, rata-rata dan median dari tiap indikator. Adanya perbedaan satuan dari tiap indikator sehingga perlu dilakukan standarisasi untuk semua indikator penelitian.

2) Uji Bartlett, KMO dan Asumsi Multikolinearitas

Sebelum dilakukan pengelompokan tiap kabupaten/kota di Kawasan Timur Indonesia (KTI), perlu adanya pengecekan asumsi bahwa data tidak terdapat multikolinearitas dan cukup secara statistik.

Tabel 3. Nilai Uji Bartlett dan KMO

	Uji	Nilai
Uji Bartlett	Approx.Chi-Square	763.903
	<i>df</i>	28
	<i>sig</i>	2.22e-16
Kaiser-Meyer-Olkin (KMO)		0.6697

Sumber: BPS (2021), diolah

Dari tabel III di atas, diperoleh *p-value* uji *bartlett* sebesar 2.22e-16 yang kurang dari tingkat signifikansi $\alpha = 0.05$, sehingga cukup bukti untuk menyatakan bahwa indikator-indikator yang digunakan saling berkorelasi (multikolinearitas). Didapatkan pula nilai KMO sebesar 0.6697 yang lebih dari 0.5, sehingga dapat disimpulkan bahwa dapat cukup untuk dianalisis menggunakan *Principal Component Analysis* (PCA).

Tabel 4. Matriks Korelasi

Indikator	PE	TPAK	RASIO PUSKES	RASIO SMA
PE	1	-0.0155	0.0163	0.0006
TPAK	-0.0155	1	-0.3307	-0.3140
RASIOPUSKES	0.0163	-0.3307	1	0.8414
RASIOSMA	0.0006	-0.3140	0.8414	1
AHH	-0.0147	-0.1504	0.0848	0.1868
HLS	0.0434	-0.2185	0.1849	0.3397
RLS	0.0892	-0.3142	0.1453	0.2825
PENG	-0.0008	-0.1508	0.0874	0.1741
	AHH	HLS	RLS	PENG
PE	-0.0147	0.0434	0.0892	-0.0008
TPAK	-0.1504	-0.2185	-0.3142	-0.1508
RASIOPUSKES	0.0848	0.1849	0.1453	0.0874

RASIOSMA	0.1868	0.3397	0.2825	0.1741
AHH	1	0.4146	0.4855	0.4383
HLS	0.4146	1	0.8117	0.4222
RLS	0.4855	0.8117	1	0.5199
PENG	0.4383	0.4222	0.5199	1

Sumber: BPS (2021), diolah

Pada matriks korelasi, terdapat korelasi yang tinggi antar indikator yang ditunjukkan pada tabel 4. Jika terdapat multikolinieritas, analisis faktor dapat digunakan.

3) Analisis Faktor

Tabel 5. Nilai Eigen Hasil Ekstraksi Faktor

Dimensi	eigenvalue	variance. percent	cumulative. variance. percent
Dim.1	3.0369	37.9618	37.9618
Dim.2	1.6202	20.2519	58.2137
Dim.3	1.0125	12.6568	70.8705
Dim.4	0.7942	9.9276	80.7981
Dim.5	0.6674	8.3423	89.1403
Dim.6	0.5610	7.0129	96.1533
Dim.7	0.1672	2.0904	98.2437
Dim.8	0.1405	1.7563	100

Sumber: BPS (2021), diolah

Untuk mereduksi indikator, digunakan metode PCA. Kriteria nilai eigen merupakan cara menentukan banyak faktor berdasarkan sumbangan ragam faktor terhadap ragam keseluruhan. Penentuan jumlah faktor dilihat dari nilai eigen yang lebih dari 1 seperti pada tabel 5. Diperoleh bahwa ketiga faktor tersebut cukup untuk menjelaskan keragaman sebesar 70.87 persen.

Tahapan selanjutnya adalah menentukan komponen dari tiap faktor yang terbentuk. Proses penentuan indikator menggunakan nilai *loading factor*. Metode rotasi faktor dapat digunakan jika proses ekstraksi faktor sulit diinterpretasikan. Berikut adalah nilai *loading factor* yang terbentuk setelah dirotasi.

Tabel 6. Nilai Loading Faktor Hasil Ekstraksi Faktor

Indikator	PC1	PC2	PC3
PE	0.0326	0.0266	0.9773
TPAK	-0.2823	0.2245	-0.0391
RASIOPUSKES	0.3061	-0.6113	-0.0089
RASIOSMA	0.3747	-0.5209	-0.0336
AHH	0.3508	0.2881	-0.1456
HLS	0.4562	0.2226	0.0479
RLS	0.4755	0.2811	0.0842
PENG	0.3575	0.3038	-0.1073

Sumber: BPS (2021), diolah

Sehingga, indikator penyusunan setiap faktor dapat diuraikan sebagai berikut.

Tabel 7. Indikator Penyusun Tiap Faktor

Faktor	Indikator	Pencirian
1	TPAK, AHH, HLS, RLS, dan PENG	Kualitas Manusia
2	RASIOPUSKES, RASIOSMA	Sarana dan Prasarana
3	PE	Ekonomi

Dari tabel 7 diatas, indikator TPAK, AHH, HLS, RLS, dan PENG dinamakan sebagai faktor kualitas manusia. Sementara itu, untuk indikator rasio puskesmas dan rasio sma dinamakan sebagai faktor sarana dan prasarana, serta indikator pertumbuhan ekonomi dinamakan sebagai faktor ekonomi.

4) Validitas Model

Tabel 8. Validitas Internal

Metode	\sum Klaster	2	3	4	5
K-Means	Connectivity	2.9290	9.7056	24.7802	34.1722
	Dunn	0.9788	0.0464	0.0487	0.0495
	Silhouette	0.8198	0.4607	0.4635	0.4175
	Davies-Bouldin	1.2032	0.8141	0.8193	0.9999
K-Medoids	Connectivity	27.7794	41.5175	47.0381	56.7448
	Dunn	0.0105	0.0176	0.0168	0.0106
	Silhouette	0.3525	0.3601	0.4509	0.3133
	Davies-Bouldin	1.2380	1.0341	0.7466	0.9634

Sumber: Data Pengolahan

Uji validitas *cluster* terlebih dahulu dilakukan untuk mengetahui jumlah *cluster* terbaik dan juga model terbaik yang akan digunakan dalam penelitian sebelum melakukan analisis *cluster*. Pada penelitian ini menggunakan menggunakan indeks *Connectivity*, *Silhouette*, *Dunn*, dan *Davies-Bouldin*. Nilai indeks *Silhouette* dan *Dunn* yang semakin mendekati 1, maka semakin baik jumlah *cluster* tersebut. Sementara itu, semakin kecil nilai *Connectivity* dan *Davies-Bouldin*, maka semakin baik *cluster* yang didapatkan. Pada tabel 8 dapat dilihat nilai validitas internal pada penelitian ini.

Tabel 9. Nilai Optimal Validitas Internal

Indeks	Score	Method	Clusters
Connectivity	2.9290	K-Means	2
Dunn	0.9788	K-Means	2
Silhouette	0.8198	K-Means	2
Davies-Bouldin	0.7466	K-Medoids	4

Sumber: Data Pengolahan

Berdasarkan tabel 9, dari keempat uji validitas terlihat bahwa yang paling optimal adalah metode *K-Means* dengan jumlah *cluster* sebanyak 2 yang memiliki nilai Indeks *Connectivity* sebesar 2.929, nilai Indeks *Dunn* sebesar 0.979 dan nilai Indeks *Silhouette* sebesar 0.820. Namun, nilai indeks *Davies Bouldin* metode *K-Medoids* yang paling kecil yaitu sebesar 0.747 dengan 4 *cluster*. Sehingga, metode yang digunakan untuk mengelompokkan kabupaten/kota di KTI tahun 2021 berdasarkan indikator sosial ekonomi adalah metode *K-Means*.

Tabel 10. Validitas Stabilitas

Metode	\sum Klaster	2	3	4	5
K-Means	APN	0.0618	0.192	0.2838	0.4186
	AD	2.6085	2.3403	2.1175	2.0626
	ADM	0.2666	0.6858	0.9263	1.0339
	FOM	1.3429	1.3414	1.3381	1.3047
K-Medoids	APN	0.203	0.2583	0.3395	0.3012
	AD	2.3681	2.1346	1.9365	1.8086
	ADM	0.7015	0.7309	0.7475	0.742
	FOM	1.3358	1.3239	1.2991	1.3002

Sumber: Data Pengolahan

Tabel 11. Nilai Optimal Validitas Stabilitas

	Score	Method	Clusters
APN	0.0618	<i>K-Means</i>	2
AD	1.8086	<i>K-Medoids</i>	5
ADM	0.2666	<i>K-Means</i>	2
FOM	1.2991	<i>K-Medoids</i>	4

Sumber: Data Pengolahan

Selanjutnya, uji validitas stabilitas dilakukan dengan menggunakan nilai APN, AD, ADM dan FOM. Nilai validitas stabilitas dapat dilihat pada tabel 11. Berdasarkan tabel 11, nilai APN dan ADM menunjukkan metode *K-Means* dengan jumlah *cluster* 2 adalah metode terbaik. Sedangkan nilai AD dan FOM yang menunjukkan metode *K-Medoids* dengan jumlah *cluster* masing-masing sebanyak 5 dan 4 merupakan metode yang terbaik.

5) Hasil Pengelompokan

Dari hasil uji validitas internal dan stabilitas, menunjukkan bahwa metode terbaik yang digunakan dalam penelitian ini adalah metode *K-Means* dengan jumlah *cluster* sebanyak 2. *Cluster* yang terbentuk oleh metode *K-Means*, dapat diketahui karakteristik dari setiap *cluster* berdasarkan indikator-indikator Sosial Ekonomi.

Tabel 12. Rata-rata Tiap Cluster

Faktor	Cluster 1	Cluster 2
Faktor Kualitas Manusia	-0.8971	1.7046
Faktor Sarana dan Prasarana	0.1935	-0.3676
Faktor Ekonomi	0.0474	-0.0901

Sumber: Data Pengolahan

Pada tabel 8, nilai rata-rata faktor diatas telah distandarisasi sebelumnya, sehingga nilai yang menjadi penentu baik buruknya suatu faktor adalah angka nol. Artinya, apabila faktor bernilai positif atau berada di atas rata-rata, maka faktor tersebut dapat dikatakan baik pada suatu *cluster*. Sementara itu, untuk rata-rata suatu faktor bernilai negatif atau berada di bawah rata-rata, maka indikator tersebut dikatakan kurang atau tidak baik untuk suatu *cluster*.

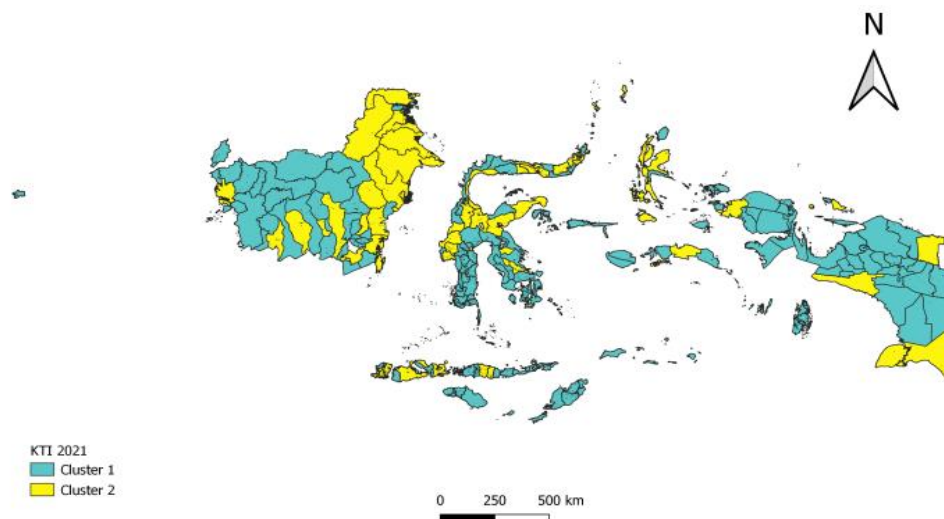
1. Cluster 1

Cluster ini terdiri dari 152 kabupaten/kota yang memiliki karakteristik sarana dan prasarana serta ekonomi di atas rata-rata, sedangkan kualitas manusia masih di bawah rata-rata. Faktor yang paling dominan pada *cluster* ini yaitu faktor sarana dan prasarana. Hal ini menandakan bahwa kondisi sosial ekonomi di *cluster* 1 didukung oleh sarana dan prasarana yang memadai. Begitupun dengan faktor ekonomi yang lebih unggul dibandingkan *cluster* 2. Namun, seiring dengan perkembangan tersebut, pemerintah perlu memperhatikan faktor kualitas manusia pada wilayah tersebut.

2. Cluster 2

Cluster ini terdiri dari 80 kabupaten/kota yang memiliki kualitas manusia di atas rata-rata, sedangkan sarana dan prasarana serta ekonominya masih di bawah rata-rata. Dengan kata lain, bahwa sumber daya manusia di *cluster* ini sudah memadai. Hanya saja kondisi sarana dan prasarana serta ekonomi perlu ditingkatkan. Seperti adanya pemenuhan fasilitas sekolah dan puskesmas yang dapat menampung murid untuk sekolah serta mengontrol kesehatan masyarakatnya.

Pengelompokan kabupaten/kota di Kawasan Timur Indonesia berdasarkan indikator sosial ekonomi tahun 2021 dengan *K-Means* berdasarkan *cluster* dapat dilihat pada lampiran yang ada pada bab metodologi. Selain itu, peta sebaran *cluster* dapat dilihat pada gambar 4 berikut ini.



Gambar 4. Peta Sebaran Cluster

Sumber: Data Pengolahan

5 Kesimpulan

Berdasarkan hasil analisis faktor, terdapat 3 faktor yang dapat menggambarkan keadaan sosial ekonomi di Kawasan Timur Indonesia (KTI), yaitu faktor kualitas manusia, sarana dan prasarana, serta ekonomi. Metode terbaik untuk pengelompokan kabupaten/kota di KTI berdasarkan indikator sosial ekonomi tahun 2021 adalah metode *K-Means* dengan jumlah *cluster* sebanyak 2. *Cluster 1* terdiri dari 152 kabupaten/kota yang memiliki karakteristik sarana dan prasarana serta ekonomi di atas rata-rata, sedangkan kualitas manusia masih di bawah rata-rata. Sementara itu, *cluster 2* terdiri dari 80 kabupaten/kota yang memiliki karakteristik berkebalikan dari *cluster 1*.

Adapun saran untuk pemerintah, yaitu agar dapat melihat kondisi sosial ekonomi di kabupaten-kabupaten yang ada di KTI sesuai dengan faktor-faktor yang terbentuk. Terkhusus pada kabupaten/kota yang masuk ke dalam *cluster 2* yang masih memiliki faktor sarana dan prasarana serta ekonomi yang masih rendah (di bawah rata-rata) dengan cara meningkatkan jumlah puskesmas dan SMA yang dapat menampung masyarakatnya serta pertumbuhan ekonomi di wilayah tersebut yang perlu ditingkatkan dengan melihat faktor-faktor yang memengaruhi pertumbuhan ekonomi di setiap wilayah. Hal ini disebabkan karena kondisi sosial ekonomi serta potensi di tiap wilayah berbeda, sehingga perlu penanganan yang berbeda pula untuk menuju pemerataan.

Saran untuk penelitian selanjutnya yaitu dapat mencoba pengelompokan dengan metode hierarki baik dengan *single linkage*, *complete linkage*, serta *average linkage* dan membandingkannya dengan *K-Means*, *K-Medoids* atau bahkan *Fuzzy C-Means*.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada dosen pembimbing, yaitu Dr. Eng. Arie Wahyu Wijayanto, S.S.T., M.T. yang telah memberikan ilmu dan bimbingannya selama penulisan dalam penelitian ini.

Referensi

- [1] Amaluddin, B. R. A. Hanoeboen, and A. Ramly, "Determinan sosial ekonomi kesenjangan distribusi pendapatan penduduk di Kawasan Timur Indonesia," *J. Paradig. Ekon.*, vol. 15, no. 1, pp. 39–50, 2020, doi: 10.22437/paradigma.v15i1.9218.
- [2] R. Raharti, T. Laras, and O. Oktavianti, "Model Ketimpangan Pembangunan Ekonomi di Indonesia," *J. Samudra Ekon. dan Bisnis*, vol. 12, no. 2, pp. 257–270, 2021, doi: 10.33059/jseb.v12i2.2422.
- [3] C. A. Kamaruddin and S. Alam, "Analisis Potensi Sektor Unggulan dan Pemetaan Kemiskinan

<http://sistemasi.ftik.unisi.ac.id>

- Masyarakat di Wilayah Maminasata Sulawesi Selatan,” *J. Ad’ministrare*, vol. 5, no. 2, p. 85, 2019, doi: 10.26858/ja.v5i2.7886.
- [4] Y. Putri, S. Amar, and H. Aimon, “Analisis Faktor-Faktor Yang Mempengaruhi Pertumbuhan Ekonomi Dan Ketimpangan Pendapatan Di Indonesia,” *J. Kaji. Ekon.*, vol. 3, no. 6, p. 102918, 2015.
- [5] E. R. Tirayo, “Pengelompokan Kabupaten / Kota di Sulawesi Tengah Menurut Indikator Sosial Ekonomi Menggunakan K-Medoids *Clustering*,” vol. 2019, no. Sns Viii, 2019.
- [6] S. A. R. Khan, C. Jian, Y. Zhang, H. Golpîra, A. Kumar, and A. Sharif, “Environmental, social and economic growth indicators spur logistics performance: From the perspective of South Asian Association for Regional Cooperation countries,” *J. Clean. Prod.*, vol. 214, pp. 1011–1023, 2019, doi: 10.1016/j.jclepro.2018.12.322.
- [7] A. Munandar, “Analisis Regresi Data Panel pada Pertumbuhan Ekonomi di Negara-Negara Asia,” *Ekon. Glob. Masa Kini*, vol. 8, no. 1, pp. 59–67, 2017, [Online]. Available: <http://ejournal.uigm.ac.id/index.php/EGMK/article/view/246>.
- [8] T. Sukwika, “Peran Pembangunan Infrastruktur terhadap Ketimpangan Ekonomi Antarwilayah di Indonesia,” *J. Wil. dan Lingkung.*, vol. 6, no. 2, p. 115, 2018, doi: 10.14710/jwl.6.2.115-130.
- [9] A. Rahayu, “Ketimpangan Pendidikan di Indonesia (Education Inequality in Indonesia),” *Journal Ekonomi & Studi Pembangunan*, vol. 6, no. 1. pp. 21–40, 2005.
- [10] A. Nurmutiazifah and A. F. Yuniasih, “Penerapan Model Regresi Data Panel : Determinan Ketimpangan Capaian Pendidikan Di Kawasan Timur Indonesia (Kti) 2015-2019,” *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 1294–1304, 2021, doi: 10.34123/semnasoffstat.v2020i1.705.
- [11] M. N. Kapriaji and S. Siswidiyanto, “Desentralisasi dan Kapasitas Pemerintah dalam Pembangunan Sumber Daya Manusia di Indonesia,” *J. Ilm. Adm. Publik*, vol. 8, no. 2, pp. 241–255, 2022, doi: 10.21776/ub.jiap.2022.008.02.14.
- [12] M. Herviany, S. P. Delima, T. Nurhidayah, and Kasini, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor di Provinsi Jawa Barat,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, [Online]. Available: <https://journal.irpi.or.id/index.php/malcom/article/view/60>.
- [13] E. Luthfi and A. W. Wijayanto, “Analisis perbandingan metode hirearchical , k-means , dan k-medoids *clustering* dalam pengelompokan indeks pembangunan manusia Indonesia Comparative analysis of hirearchical , k-means , and k-medoids *clustering* and methods in grouping Indonesia ’ s human,” *Inovasi*, vol. 17, no. 4, pp. 770–782, 2021, [Online]. Available: <https://journal.feb.unmul.ac.id/index.php/INOVASI/article/download/10106/1437>.
- [14] F. Faisal, L. A. Giopani, M. Fitriah, Z. C. Dwyne, and S. Syahidatul, “Comparison of K-Means and K-Medoids Algorithms for Temperature Grouping in Riau Province Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Suhu di Provinsi Riau,” vol. 2, no. 2, pp. 128–134, 2022.
- [15] D. Marlina, N. F. Putri, A. Fernando, and A. Ramadhan, “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak,” vol. 4, no. 2, pp. 64–71, 2018.
- [16] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical *Clustering*,” *J. Ilmu Komput.*, vol. 14, no. 2, p. 66, 2021, doi: 10.24843/jik.2021.v14.i02.p01.
- [17] N. Afira and A. W. Wijayanto, “Analisis *Cluster* dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019,” *Komputika J. Sist. Komput.*, vol. 10, no. 2, pp. 101–109, 2021, doi: 10.34010/komputika.v10i2.4317.
- [18] J. Supranto, *Analisis Multivariat: Arti & Interpretasi*. Jakarta: Rineka Cipta, 2004.
- [19] R. E. Hair, J F; Black, W C; Babin, B J; Anderson, *Multivariate Data Analysis, Eighth Edition*. Hampshire: Cengage Learning, 2019.
- [20] D. W. Johnson, R A; Wichern, *Applied Multivariate Statistical Analysis: Pearson New International Edition, 6th Edition*. Essex: Pearson Education Limited, 2013.
- [21] S. Handi Kurniawan Sohdianata, “Penerapan Algoritma K-Means Untuk Menetapkan Kelompok Mutu Karet,” *Progresif*, vol. 7, no. 2, pp. 2–9, 2011, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>

- <http://ojs.stmik-banjarbaru.ac.id/index.php/progresif/article/viewFile/60/60>.
- [22] H. D. Tampubolon, S. Suhada, M. Safii, S. Solikhun, and D. Suhendro, "Penerapan Algoritma K-Means dan K-Medoids *Clustering* untuk Mengelompokkan Tindak Kriminalitas Berdasarkan Provinsi," *J. Ilmu Komput. dan Teknol.*, vol. 2, no. 2, pp. 6–12, 2021, doi: 10.35960/ikomti.v2i2.703.
- [23] G. Dwilestari, Mulyawan, Martanto, and I. Ali, "Analisis *Clustering* menggunakan K-Medoid pada Data Penduduk Miskin Indonesia," *JURSIMA J. Sist. Inf. dan Manaj.*, vol. 9, no. 3, pp. 282–290, 2021.
- [24] M. A. Nahdliyah, T. Widiharih, and A. Prahutama, "METODE k-MEDOIDS *CLUSTERING* DENGAN VALIDITAS SILHOUETTE INDEX DAN C-INDEX (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018)," *J. Gaussian*, vol. 8, no. 2, pp. 161–170, 2019, doi: 10.14710/j.gauss.v8i2.26640.
- [25] X. Wang and Y. Xu, "An improved index for *clustering* validation based on Silhouette index and Calinski-Harabasz index," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.
- [26] N. I. Febianto and N. Palasara, "Analisa *Clustering* K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, pp. 130–140, 2019, doi: 10.32736/sisfokom.v8i2.653.