

Pengelompokkan Kabupaten/Kota di Provinsi Kalimantan Selatan Berdasarkan IPM dan Kemiskinan Tahun 2022 Menggunakan *K-Means* dan *K-Medoids*

M. Difa Farady^{1*}, Akbar Fikriawan², Nur Abdillah Al Fatih³, Akhmad Aufar Romeo Bakhtiar⁴, and Joycelin Gracelda Resi Gaya⁵

^{1,2,3,4,5}Departemen Statistika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

*Corresponding author: difarady1713@gmail.com

ABSTRAK – Penelitian ini mengevaluasi kondisi kemiskinan dan Indeks Pembangunan Manusia (IPM) di Kabupaten/Kota Kalimantan Selatan pada tahun 2022 dengan menggunakan pendekatan *K-Means* dan *K-Medoids*. Hasil analisis menunjukkan bahwa kondisi kemiskinan relatif stabil dan konsisten antar daerah, dengan rendahnya varians Gini Ratio, P0, P1, dan P2. Sebaliknya, tingginya varians IPM dan Garis Kemiskinan menandakan ketidakmerataan pembangunan antar kabupaten/kota. Temuan ini menggambarkan hubungan negatif antara IPM dan indikator kemiskinan, menunjukkan bahwa peningkatan IPM berkorelasi dengan penurunan tingkat kemiskinan. Khususnya, Kota Banjarbaru dan Banjarmasin menonjol dengan nilai IPM yang jauh lebih tinggi, mencerminkan sumber daya manusia yang lebih baik. Analisis *clustering* dengan *K-Means* dan *K-Medoids* menghasilkan tiga dan dua kluster, masing-masing, yang dapat memisahkan daerah berdasarkan IPM dan tingkat kemiskinan. Terlihat bahwa daerah dengan tingkat kemiskinan dan IPM rendah cenderung berdekatan, terutama di bagian utara Kalimantan Selatan. Beberapa kabupaten, seperti Tabalong dan Kotabaru, memiliki Gini Ratio yang lebih tinggi karena menjadi penghasil batubara. Penelitian ini memberikan wawasan tentang distribusi kemiskinan dan pembangunan di Kalimantan Selatan serta relevansi pendekatan *clustering* dalam pemetaan kondisi regional.

Kata kunci – *Clustering*, IPM, Kemiskinan.

I. PENDAHULUAN

A. Latar Belakang

Kemiskinan merupakan kondisi dimana seseorang atau sekelompok orang tidak mampu memenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Menurut Arsyad (2010) kemiskinan secara garis besar dapat digolongkan menjadi dua aspek, yaitu aspek primer dan aspek sekunder. Aspek primer adalah aspek yang berupa miskin aset (harta), organisasi sosial politik, pengetahuan, dan keterampilan. Sementara itu, aspek sekunder adalah aspek yang berupa miskin terhadap jaringan sosial, sumber-sumber keuangan dan informasi. Untuk mengukur kemiskinan konsep yang dipakai BPS adalah kemampuan memenuhi kebutuhan dasar (*basic needs approach*), dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Jadi Penduduk Miskin adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan dibawah Garis Kemiskinan (GK).

Garis Kemiskinan sendiri merupakan representasi dari rupiah yang diperlukan atau harga yang dibayarkan agar penduduk dapat hidup layak secara minimum yang mencakup pemenuhan kebutuhan minimum makanan (setara dengan 2.100 kilokalori per kapita per hari) dan non makanan *essential*. Garis Kemiskinan yang digunakan oleh BPS terdiri dari dua komponen, yaitu Garis Kemiskinan Makanan (GKM) yang terdiri atas 52 jenis komoditi dan Garis Kemiskinan Non Makanan (GKNM) yang terdiri dari 51 jenis komoditi untuk perkotaan, dan 47 jenis komoditi untuk perdesaan. Garis Kemiskinan (GK) merupakan penjumlahan dari GKM dan GKNM. Didalam Indeks Garis Kemiskinan ini terdapat juga variable Head Count Index (HCI-P0) atau presentase penduduk miskin, yang merupakan persentase penduduk yang berada dibawah Garis Kemiskinan (GK), kemudian ada Indeks Kedalaman Kemiskinan (Poverty Gap Index-P1), merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan dan juga terdapat Indeks Keparahan Kemiskinan (Poverty Severity Index/P2) yang memberikan informasi mengenai gambaran penyebaran pengeluaran di antara penduduk miskin. Di sisi lain tingkat kemiskinan dapat dipengaruhi oleh Gini Ratio. Menurut Barber (2008) dalam Annim et al. (2012) memandang hubungan antara kesenjangan dan kemiskinan sebagai hubungan yang pragmatis, yaitu bahwa kesenjangan menyebabkan kemiskinan semakin parah.

Adapun Indeks Pembangunan Manusia (IPM) yang merupakan alat untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui

pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan sehat, pengetahuan, dan kehidupan yang layak. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak. Komponen ini terbagi atas Angka Harapan Hidup, Angka Melek Huruf, Rata-rata lama sekolah dan Pengeluaran Riil per Kapita yang disesuaikan.

Berdasarkan data BPS, proyeksi jumlah penduduk Indonesia tahun 2022 adalah sebanyak 275,77 juta jiwa. Dari 34 Provinsi yang ada di Indonesia, Provinsi Kalimantan Selatan memiliki total penduduk sebanyak 4.182.100 dan merupakan top 18 provinsi yang memiliki penduduk terbanyak diantara 34 provinsi lainnya. Banyaknya jumlah penduduk di Provinsi Kalimantan ini tentunya akan menyebabkan permasalahan-permasalahan di bidang sosial seperti kemiskinan. Seperti yang diketahui jumlah penduduk miskin di provinsi Kalimantan Selatan 2022 adalah sebanyak 195.702 ribu jiwa. Oleh karena itu, penelitian tentang kemiskinan sangat penting dilakukan dengan tujuan untuk mengetahui kondisi kemiskinan dan IPM kabupaten/kota di Provinsi Kalimantan Selatan, dengan tujuan agar dapat diupayakan bagaimana cara menurunkan kemiskinan di Provinsi Kalimantan Selatan ditinjau dengan beberapa variabel, seperti kemiskinan seperti angka Garis Kemiskinan, Gini Ratio, Tingkat Kedalaman Kemiskinan(P1), Tingkat Keparahan Kemiskinan(P2), Indeks Pembangunan Manusia dan Presentase Penduduk Miskin(P0).

B. Rumusan Masalah

Adapun rumusan masalah yang diangkat dalam penelitian ini sebagai berikut.

1. Bagaimana gambaran kondisi IPM dan kemiskinan di provinsi Kalimantan Selatan tahun 2022?
2. Bagaimana pengelompokan kabupaten/kota di provinsi Kalimantan Selatan berdasarkan IPM dan kemiskinan 2022?
3. Bagaimana perbedaan karakteristik kelompok yang terbentuk?

C. Tujuan

Berdasarkan rumusan masalah, berikut merupakan tujuan penelitian.

1. Mendeskripsikan kondisi IPM dan kemiskinan di provinsi Kalimantan Selatan tahun 2022.
2. Mengelompokkan kabupaten/kota di provinsi Kalimantan Selatan berdasarkan IPM dan kemiskinan tahun 2022.
3. Menganalisis perbedaan karakteristik kelompok yang terbentuk.

II. TINJAUAN PUSTAKA

A. Statistika Deskriptif

Statistika deskriptif adalah ilmu statistika yang mempelajari tentang pengumpulan, penyederhanaan, pengolahan, analisis, dan penyajian data yang bertujuan untuk memaparkan informasi dari hasil suatu penelitian berdasarkan data populasi [1]. Kegiatan statistika deskriptif meliputi mencari rata-rata (mean), modus, median, ragam atau standar deviasi tanpa disertai tindak lanjut (penyimpulan/pengujian suatu hipotesis) [2]. Berikut ini adalah informasi-informasi yang dapat diperoleh dari metode statistika deskriptif pada penelitian ini.

1) Mean

Mean atau rata-rata merupakan ukuran pemusatan data didapatkan dari hasil pembagian jumlah keseluruhan nilai penelitian dengan banyaknya penelitian yang dilakukan. Secara matematis, mean dapat dihitung dengan menggunakan rumus pada Persamaan (1).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Dimana \bar{x} adalah rata-rata, x_i adalah data ke- i , dan n adalah banyaknya data.

2) Varians

Varians adalah ukuran penyebaran data yang mengukur rata-rata jarak kuadrat semua titik penelitian terhadap titik pusatnya (rata-rata). Secara matematis, varians dapat dihitung dengan menggunakan rumus pada Persamaan (2).

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (2)$$

Dimana s^2 adalah varians, \bar{x} adalah rata-rata, x_i adalah data ke- i , dan n adalah banyaknya data.

3) Standar Deviasi

Standar deviasi merupakan statistik yang digunakan untuk melihat bagaimana persebaran suatu data dalam sampel dan melihat seberapa dekat titik data individu ke rata-rata nilai sampel. Secara matematis, standar deviasi dapat dihitung dengan menggunakan rumus pada Persamaan (3).

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3)$$

4) Nilai Minimum

Nilai minimum merupakan ukuran penyebaran data yang menunjukkan nilai terkecil dari suatu kumpulan data.

5) Nilai Maksimum

Nilai maksimum merupakan ukuran penyebaran data yang menunjukkan nilai terbesar dari suatu kumpulan data.

B. Analisis Kluster

Analisis Kluster atau analisis kelompok merupakan suatu metode analisis untuk mengelompokkan objek-objek pengamatan menjadi beberapa kelompok sehingga akan diperoleh kelompok dimana objek-objek dalam satu kelompok mempunyai banyak persamaan sedangkan dengan anggota kelompok yang lain mempunyai banyak perbedaan. Prosedur pada analisis kluster terdapat dua metode yakni metode hirarki dan non hirarki, pada penelitian ini akan digunakan metode analisis kluster non hirarki [3].

Analisis kluster dengan menggunakan metode non hirarki adalah metode pengelompokan dengan menentukan banyak kelompok terlebih dahulu, kemudian baru menentukan anggota tiap kelompok tersebut. Ukuran kesamaan merupakan hal yang paling penting pada analisis kluster. Salah satu ukuran kesamaan yang dapat digunakan adalah jarak euclidian dimana formulanya ditunjukkan pada persamaan sebagai berikut [3].

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (4)$$

1) *K-Means*

Pada pengelompokan dengan *K-Means*, pusat kelompok diwakili oleh mean (rata-rata) dari seluruh data anggota kelompok. Algoritma ini meminimalkan jumlah kuadrat jarak antara setiap data dan pusat kelompoknya. Metode ini dimulai dengan proses penentuan jumlah kluster terlebih dahulu. Metode non hirarki bertujuan mengelompokkan n obyek kedalam k kelompok ($k < n$) metode ini yang digunakan adalah *K-means*. Terdapat tiga pendekatan yang digunakan untuk menempatkan masing-masing observasi pada satu kluster yaitu *Sequential Threshold*, *Parallel Threshold* dan *Optimization*. Selanjutnya algoritma untuk *K-means*, dimana merupakan metode data *clustering* non hirarki yang digunakan untuk mengelompokkan data yang ada ke dalam bentuk satu atau lebih *cluster*. Metode ini mengelompokkan data kedalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu cluster dan memaksimalkan variasi antar cluster [4].

2) *Partitioning Around Medoids (PAM)*

Pada pengelompokan *K-Medoids*, pusat kelompok diwakili oleh salah satu data anggota kelompok (*medoid*), yang merupakan titik yang memiliki jarak total minimum ke semua titik lain dalam kelompok. PAM adalah implementasi spesifik dari *K-Medoids*. Jika berbicara tentang PAM dalam konteks *clustering*, itu sama dengan berbicara tentang *K-Medoids*. Algoritma PAM dikembangkan oleh Leonard Kaufman dan J. Rousseeuw, dan algoritma ini mirip dengan *K-Means*, terutama keduanya algoritma partitional dengan kata lain, keduanya memecah dataset menjadi kelompok-kelompok, dan keduanya bekerja secara acak, lalu proses perhitungan cost dilakukan kembali. Apabila total cost yang dihasilkan lebih kecil dari cost setiap objek dengan *medoid* lama, maka objek baru tersebut dapat menjadi *medoids* baru. Iterasi terakhir sampai tidak terdapat perubahan cost terhadap cost yang dihasilkan oleh *medoid* baru. Berikut tahapan algoritma PAM [5].

- Secara acak, pilih k objek pada sekumpulan n objek sebagai *medoid*.
- Ulangi.
- Tempatkan objek *non-medoid* ke dalam *cluster* yang paling dekat dengan *medoid* menggunakan metode perhitungan jarak *Euclidean Distance*.
- Secara acak, pilih O_{random} : sebuah objek *non-medoid*.
- Hitung total biaya, S, dari pertukaran *medoid* O_j dengan O_{random} .
- Jika $S < 0$, maka tukar O_j dengan O_{random} untuk membentuk sekumpulan k objek baru sebagai *medoid*.
- Ulangi proses 3-6 hingga tidak ada perubahan.

C. Silhouette Coefficient

Metode *Silhouette Coefficient* adalah metode yang digunakan untuk mengevaluasi atau mengukur kualitas kluster yang terbentuk [6]. Metode ini mengkombinasikan metode cohesi dan separation. Nilai *Silhouette Coefficient* berkisar antara -1 sampai 1, jika nilai *Silhouette Coefficient* mendekati 1 berarti pengelompokan pada kluster semakin baik. Sebaliknya jika mendekati -1 berarti pengelompokan pada kluster semakin buruk.

Berikut ini merupakan tahapan – tahapan atau algoritma dalam perhitungan *silhouette coefficient*.

- 1) Menghitung rata – rata jarak antara suatu data dengan data lain pada suatu *cluster* seperti pada Persamaan 4 dengan cara memisahkan i dengan seluruh data lain yang berada pada satu *cluster*.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (4)$$

Dimana $a(i)$ adalah perbedaan rata – rata objek (i) ke seluruh objek lain pada *cluster* A, $d(I, j)$ merupakan jarak antara data I dan data j , dan A merupakan *cluster*.

- 2) Menghitung rata – rata jarak dari suatu data dengan seluruh data di *cluster* lain yang tersaji pada Persamaan 5 berikut.

$$d(i, B) = \frac{1}{|B|} \sum_{j \in C} d(i, j) \quad (5)$$

Dimana $d(i, B)$ adalah perbedaan rata – rata objek i ke seluruh objek lain pada *cluster* B dan B merupakan *cluster* baru selain *cluster* A.

- 3) Menghitung $d(i, B)$ untuk seluruh B, dan diambil nilai terkecil yang tersaji pada Persamaan 6 berikut.

$$b(i) = \min_{B \neq A} d(i, B) \quad (6)$$

Dimana $b(i)$ adalah *cluster* tetangga dari objek (i) yang mencapai nilai minimum, sehingga *cluster* C yang mencapai minimum yaitu ($d(i, C) = b(i)$) disebut tetangga dari objek (i) dan merupakan *cluster* terbaik kedua untuk objek (i).

- 4) Menghitung nilai *silhouette coefficient* untuk setiap i sebagaimana yang ada pada Persamaan 7 berikut.

$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \quad (7)$$

Dimana $s(i)$ merupakan nilai dari *silhouette coefficient*, dan untuk interpretasi dari hasil *cislhouette coefficient* tersaji dalam Tabel 1 berikut.

Tabel 1 Interpretasi Nilai *Silhouette Coefficient*

Nilai <i>Silhouette Coefficient</i>	Interpretasi
0,71 – 1,00	<i>Cluster</i> yang sangat baik (kuat)
0,51 – 0,70	<i>Cluster</i> yang baik
0,26 – 0,50	<i>Cluster</i> yang lemah
$\leq 0,25$	Tidak dapat dikatakan sebagai <i>cluster</i>

D. Analysis of Varians (ANOVA)

Secara umum, *analysis of varians* (ANOVA) menguji dua varians (atau ragam) berdasarkan hipotesis nol bahwa kedua varians itu sama. Varians pertama adalah varians antarcontoh (among samples) dan varians kedua adalah varians di dalam masing-masing contoh (within samples). Adanya ide semacam ini, analisis varians dengan dua contoh akan memberikan hasil yang sama dengan uji-t untuk dua rerata (*mean*). Berikut komponen dan persamaan yang digunakan pada ANOVA [3].

Tabel 2 Persamaan ANOVA

Sumber	Derajat Bebas	Jumlah Kuadrat	Rata-Rata Kuadrat	Fhitung
Perlakuan	$g-1$	$\sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2$	$\frac{JKP}{dbp}$	$\frac{KTP}{KTE}$
Error	$\sum_{l=1}^g n_l - g$	$\sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2$	$\frac{JKE}{dbe}$	
Total	$\sum_{l=1}^g n_l - 1$	$\sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2$		

Dimana:

g : Banyaknya ulangan

- $$\sum_{l=1}^g n_l$$
- : Banyaknya perlakuan
-
- x_{ij}
- : Data pada baris ke-l, kolom ke-j
-
- dbp
- : Derajat bebas perlakuan
-
- dbe
- : Derajat bebas error
-
- JKP
- : Jumlah kuadrat perlakuan
-
- JKE
- : Jumlah kuadrat error
-
- KTP
- : Rata-rata kuadrat perlakuan
-
- KTE
- : Rata-rata kuadrat error

E. Garis Kemiskinan

Garis Kemiskinan (GK) merupakan representasi dari jumlah rupiah minimum yang dibutuhkan untuk memenuhi kebutuhan pokok minimum makanan yang setara dengan 2.100 kilokalori per kapita per hari dan kebutuhan pokok bukan makanan [7]. Menurut BPS, Garis Kemiskinan (GK) merupakan penjumlahan dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). Penduduk yang memiliki rata-rata pengeluaran perkapita per bulan dibawah Garis Kemiskinan dikategorikan sebagai penduduk miskin.

Garis Kemiskinan Makanan (GKM) merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2100 kilokalori perkapita perhari. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buah-buahan, minyak dan lemak, dll). Sedangkan Garis Kemiskinan Non Makanan (GKNM) adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Paket komoditi kebutuhan dasar non makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan [8].

F. Persentase Penduduk Miskin

Persentase penduduk miskin adalah persentase penduduk yang berada dibawah garis kemiskinan dengan rumus perhitungan sebagai berikut [8].

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha} \quad (8)$$

Dimana :

$\alpha = 0$

z = garis kemiskinan.

y_i = Rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($i=1, 2, 3, \dots, q$), $y_i < z$

q = Banyaknya penduduk yang berada di bawah garis kemiskinan.

n = jumlah penduduk.

G. Indeks Kedalaman Kemiskinan

Indeks Kedalaman Kemiskinan (*Poverty Gap Index-P1*), merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan. Semakin tinggi nilai indeks, semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan. Dengan rumus perhitungan sebagai berikut [8].

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha} \quad (9)$$

Dimana :

$\alpha = 1$

z = garis kemiskinan.

y_i = Rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($i=1, 2, 3, \dots, q$), $y_i < z$

q = Banyaknya penduduk yang berada di bawah garis kemiskinan.

n = jumlah penduduk.

H. Indeks Keparahan Kemiskinan

Indeks Keparahan Kemiskinan (*Proverty Severity Index-P2*) memberikan gambaran mengenai penyebaran pengeluaran diantara penduduk miskin. Semakin tinggi nilai indeks, semakin tinggi ketimpangan pengeluaran diantara penduduk miskin [8].

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[\frac{z - y_i}{z} \right]^{\alpha} \quad (10)$$

Dimana :

$\alpha = 2$

z = garis kemiskinan.

y_i = Rata-rata pengeluaran per kapita sebulan penduduk yang berada dibawah garis kemiskinan ($i=1, 2, 3, \dots, q$), $y_i < z$
 q = Banyaknya penduduk yang berada di bawah garis kemiskinan.
 n = jumlah penduduk.

I. Gini Ratio

Gini ratio merupakan suatu ukuran pemerataan yang dihitung dengan membandingkan luas antara diagonal dan kurva Lorenz dibagi dengan luas segitiga di bawah diagonal. Rasio gini bernilai antara 0 dan 1. Nilai 1 menunjukkan *complete inequality* atau *perfectly inequal*, di mana seluruh penduduk menempati satu lokasi di suatu negara dan tidak ada penduduk di lokasi lainnya. Nilai 0 menunjukkan *perfectly equal*, yaitu penduduk terdistribusikan sempurna di seluruh wilayah suatu negara. Jadi, semakin besar nilai rasio konsentrasi gini, semakin tidak merata antara distribusi penduduk dan jumlah lokasi [8].

$$GR = 1 - \sum f_i[Y_i + Y_{i-1}] \quad (11)$$

Dimana :

f_i = Jumlah persen (%) penerima pendapatan kelas ke- i

Y_i = Jumlah kumulatif (%) pendapatan pada kelas ke- i

J. Indeks Pembangunan Manusia

Mengutip isi Human Development Report (HDR) pertama tahun 1990, pembangunan manusia adalah suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Diantara banyak pilihan tersebut, pilihan yang terpenting adalah untuk berumur panjang dan sehat, untuk berilmu pengetahuan, dan untuk mempunyai akses terhadap sumber daya yang dibutuhkan agar dapat hidup secara layak. Badan Pusat Statistik menjelaskan bahwa Indeks Pembangunan Manusia (IPM) mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup. Sebagai ukuran kualitas hidup, IPM dibangun melalui pendekatan tiga dimensi dasar. Dimensi tersebut mencakup umur panjang dan sehat; pengetahuan, dan kehidupan yang layak. Ketiga dimensi tersebut memiliki pengertian sangat luas karena terkait banyak faktor. Untuk mengukur dimensi kesehatan, digunakan angka harapan hidup waktu lahir. Selanjutnya untuk mengukur dimensi pengetahuan digunakan gabungan indikator angka melek huruf dan rata-rata lama sekolah. Adapun untuk mengukur dimensi hidup layak digunakan indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari rata-rata besarnya pengeluaran per kapita sebagai pendekatan pendapatan yang mewakili capaian pembangunan untuk hidup layak.

III. METODOLOGI

A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari website Badan Pusat Statistika Kalimantan Selatan (<https://kalsel.bps.go.id/>) dan data diambil pada 9 Desember 2023.

B. Variabel Penelitian

Penelitian ini mengumpulkan dan mengolah variabel pada analisis ini sebagai berikut.

Tabel 3 Variabel Penelitian

Variabel	Satuan	Skala	Notasi
Garis Kemiskinan	Rupiah	Rasio	X_1
Gini Ratio	Persentase	Rasio	X_2
Persentase Penduduk Miskin (P0)	Indeks	Rasio	X_3
Indeks Kedalaman Kemiskinan (P1)	Indeks	Rasio	X_4
Indeks Keparahan Kemiskinan (P2)	Indeks	Rasio	X_5
Indeks Pembangunan Manusia	Indeks	Rasio	X_6

C. Struktur Data

Struktur data yang digunakan pada penelitian ini adalah sebagai berikut.

Tabel 4 Variabel Penelitian

No	X_1	X_2	X_3	X_4	X_5	X_6
1	X_{11}	X_{21}	X_{31}	X_{41}	X_{51}	X_{61}
2	X_{12}	X_{22}	X_{32}	X_{42}	X_{52}	X_{62}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N	X_{1N}	X_{2N}	X_{3N}	X_{4N}	X_{5N}	X_{6N}

D. Langkah Penelitian

Berikut adalah Langkah-langkah pada penelitian ini.

1. Merumuskan permasalahan

2. Mengumpulkan data
3. Melakukan analisis statistika deskriptif
4. Melakukan pengecekan outlier
5. Menentukan *cluster* optimal
6. Melakukan analisis cluster
7. Pengecekan asumsi normal multivariat
8. Analisis MANOVA dan ANOVA
9. Visualisasi *cluster* menggunakan *map chart*
10. Menarik kesimpulan dan saran

IV. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

Untuk mengetahui karakteristik suatu data, dapat dilakukan dengan menghitung nilai rata-rata (*mean*), standar deviasi, varians, nilai minimum, dan nilai maksimum. Berikut hasil perhitungan yang dilakukan.

Tabel 5 Karakteristik Data

Variabel	N	Mean	Varians	Sd	Min	Maks
Garis Kemiskinan (Ribu Rupiah)	14	531.97	6424018.49	80.15	375.71	692.83
Gini Ratio	14	0.296	0.0011	0.033	0.228	0.341
P0 (Persentase Penduduk Miskin)	14	4.686	1.046	1.023	2.790	6.490
P1 (Kedalaman Kemiskinan)	14	0.549	0.0424	0.206	0.168	0.863
P2 (Keparahan Kemiskinan)	14	0.108	0.0026	0.051	0.015	0.178
IPM (Indeks Pembangunan Manusia)	14	71.403	12.688	3.562	66.840	79.680

Berdasarkan Tabel 5, dapat disimpulkan bahwa kondisi kemiskinan di Kalimantan Selatan pada tahun 2022 relatif stabil dan konsisten antar daerah. Indikator kemiskinan, seperti Gini Ratio, P0, P1, dan P2, menunjukkan nilai varians yang rendah dan rentang nilai yang relatif kecil, mencerminkan konsistensi tingkat kemiskinan di sepanjang daerah-daerah tersebut. Namun, perbedaan yang signifikan antar daerah terlihat pada Garis Kemiskinan dan Indeks Pembangunan Manusia (IPM). Varians yang lebih besar pada kedua variabel ini menunjukkan bahwa pembangunan dan kesejahteraan penduduk di antara kabupaten/kota di Kalimantan Selatan bersifat kurang merata.

Garis Kemiskinan, dengan rata-rata sebesar 531.97, memiliki variasi yang cukup besar dan rentang nilai antara 375.71 hingga 692.83 menandakan perbedaan yang signifikan dalam tingkat kemiskinan di antara daerah tersebut. Sementara itu, Indeks Pembangunan Manusia (IPM) menunjukkan rata-rata sebesar 71.403, dengan variasi yang cukup besar (12.688) dan rentang nilai antara 66.840 hingga 79.680. Ini mencerminkan ketidakmerataan dalam pembangunan manusia antar kabupaten/kota.

B. Korelasi

Dilakukan perhitungan dan visualisasi korelasi dari keenam variabel sebagai berikut.



Gambar 1 Correlation Plot

Berdasarkan plot gambar 1 di atas, diketahui bahwa indikator P0, P1, dan P2 memiliki korelasi yang tinggi mengindikasikan bahwa jika persentase penduduk naik maka P1 dan P2 cenderung mengikuti naik dan begitu juga sebaliknya. Hal yang sama dapat dilihat pada variabel IPM, Gini Ratio, dan Garis Kemiskinan di mana jika IPM

mengalami kenaikan maka Gini Ratio dan Garis Kemiskinan cenderung naik. Informasi di atas memberi tahu bahwa jika persentase penduduk miskin naik maka kedalaman dan keparahan seringkali juga mengalami kenaikan, dan jika IPM mengalami kenaikan maka penduduk relatif lebih sejahtera yang menyebabkan naiknya garis kemiskinan dan gini ratio.

C. Outlier

Analisis *outlier* dilakukan sebelum melakukan analisis cluster untuk menghindari hal – hal yang tidak diinginkan. *Outlier* diidentifikasi dengan batas bawah $Q_1 - (IQR \times 1.5)$ dan batas atas $Q_3 + (IQR \times 1.5)$. Hasil dari perhitungan dan *outlier* yang teridentifikasi sebagai berikut.

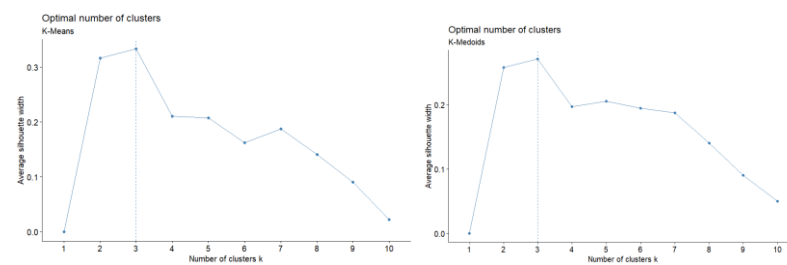
Tabel 6 Outlier

Variabel	Batas	Outlier
Garis Kemiskinan (Ribu Rupiah)	Atas : 688.62	692.83 (Kota Banjarbaru)
	Bawah : 370.00	
Gini Ratio	Atas : 0.409	Tidak ada.
	Bawah : 0.189	
P0 (Persentase Penduduk Miskin)	Atas : 7.611	Tidak ada.
	Bawah : 2.141	
P1 (Kedalaman Kemiskinan)	Atas : 1.224	Tidak ada.
	Bawah : -0.151	
P2 (Keparahan Kemiskinan)	Atas : 0.267	Tidak ada.
	Bawah : -0.054	
IPM (Indeks Pembangunan Manusia)	Atas : 74.97	77.97 (Kota Banjarmasin) 79.68 (Kota Banjarbaru)
	Bawah : 66.59	

Dari perhitungan tabel 6 di atas, teridentifikasi 2 kota yang memiliki nilai *outlier* yaitu Kota Banjarbaru dan Banjarmasin, oleh karena itu kami mengeluarkan variabel tersebut dari analisis selanjutnya.

D. Cluster Optimal

Dalam melakukan algoritma *K-Means* dan *K-Medoids*, diperlukan *k* optimal yaitu jumlah cluster dengan similaritas tertinggi. Perhitungan similaritas dilakukan menggunakan koefisien *silhouette* kemudian hasil yang didapat divisualisasikan sebagai berikut.

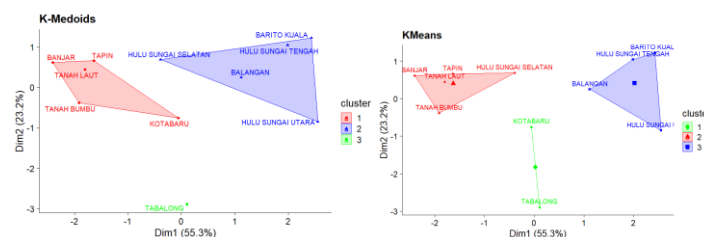


Gambar 2 Average Silhouette Width Plot

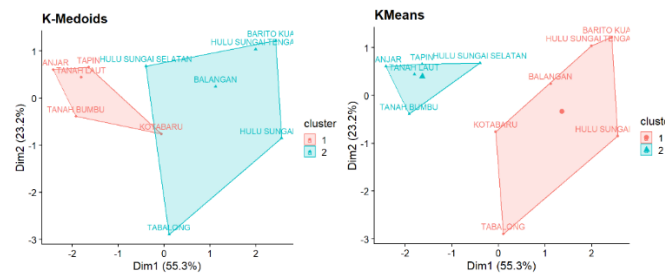
Pada Gambar 2 di atas diketahui bahwa kedua metode menghasilkan jumlah *cluster* optimal yang sama yaitu tiga *cluster* sehingga *k* optimal yaitu tiga.

E. Analisis Cluster

Dilakukan clusterisasi dengan tiga cluster dan hasil dari tiga cluster yang terbentuk seperti gambar berikut.

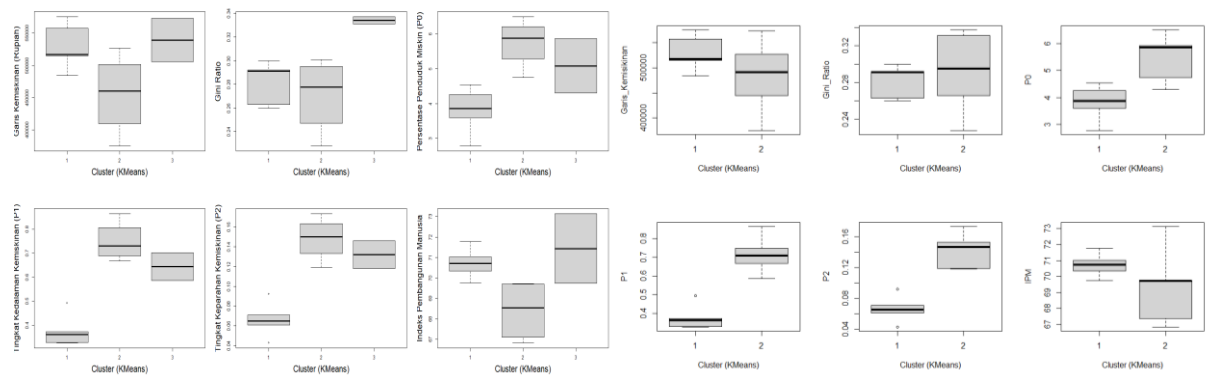


Gambar 3 Cluster Plot dengan 3 Kelompok



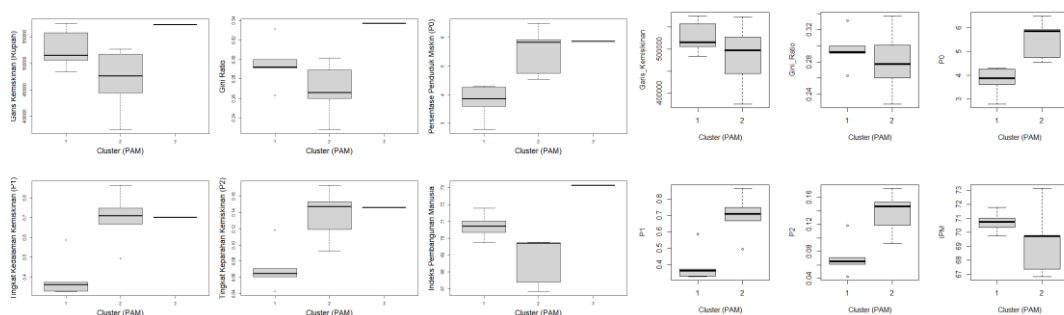
Gambar 4 Cluster Plot dengan 2 Kelompok

Gambar 3 di atas menunjukkan bahwa dengan 3 cluster tidak adanya *overlap* antar cluster. Apabila digunakan 2 cluster dengan K-Medoids, maka terjadi overlap antar cluster. Pada penelitian ini juga digunakan clustering berbasis hirarki dimana dengan metode Complete linkage 2 cluster memiliki anggota cluster yang sama dengan k-means 2 cluster. Untuk Single linkage tidak terlalu baik dalam membuat cluster sedangkan ward juga memberikan hasil yang mirip, oleh karena itu pada penelitian ini berfokus pada k-means dan k-medoid karena memiliki performa yang mirip dengan beberapa clustering hirarki. Selanjutnya dilakukan visualisasi *boxplot* untuk mengetahui perbedaan nilai antara tiap clusternya sebagai berikut.



Gambar 5 Boxplot Cluster K-Means

Dapat dilihat berdasarkan boxplot dengan 3 cluster kmeans mampu mengelompokkan variabel dengan lebih baik daripada dengan 2 variabel. Dapat dilihat dimana awalnya belum bisa membedakan garis kemiskinan dan gini ratio dengan 2 cluster, dapat dibedakan dengan 3 cluster. Dapat dilihat berdasarkan boxplot k-means 3 cluster, cluster ke-2 merupakan daerah yang paling mengkhawatirkan karena memiliki nilai Tingkat kemiskinan yang tinggi dengan nilai ipm yang juga rendah. Dapat dilihat juga gini ratio cluster ke-2 yang rendah menunjukkan Masyarakat kabupaten di cluster ke-2 masih memiliki pendapatan yang rendah. Oleh karena itu dipilih k-means 3 cluster untuk dianalisis lebih lanjut menggunakan manova.



Gambar 6 Boxplot Cluster K-Medoids

Berdasarkan gambar 5 dapat dilihat k-medoid dengan 3 cluster tidak terlalu berbeda dengan 3 cluster, dimana kabupaten tabalong yang ada di cluster ke-2 menjadi cluster sendiri. Oleh karena itu digunakan 2 cluster karena ada cluster yang memiliki satu anggota saja apabila dibuat 3 cluster. Dapat dilihat cluster 2 memiliki ciri nilai kemiskinan yang tinggi dan ipm yang rendah. Kemudian dengan k-medoid juga masih belum bisa mengelompokkan data berdasarkan gini ratio. Setelah diperoleh kelompok-kelompok yang terbentuk pada setiap metode, baik menggunakan k-means maupun k-medoids, maka analisis dilanjutkan dengan menganalisis perbedaan antar kelompok menggunakan Manova dan Anova.

F. MANOVA

a. Asumsi Normal Multivariat

Sebelum melakukan Manova dan anova dilakukan terlebih dahulu pengujian asumsi normal multivariat dan normal univariat. Berikut adalah hasil pengujiannya.

Tabel 7 Asumsi Normalitas

Variabel	Distribusi	p-value
Multivariat	Normal Multivariat	0.23
Garis Kemiskinan	Normal	0.42
Gini Ratio	Normal	0.75
P0	Normal	0.44
P1	Normal	0.25
P2	Normal	0.42
IPM	Normal	0.33

Dapat dilihat berdasarkan hasil pengujian normalitas, data berdistribusi normal multivariat dan normal univariat untuk setiap variabel. Oleh karena itu bisa dilanjutkan dengan menggunakan Manova dan Anova.

b. Analisis Manova dan Anova

Berikut ini adalah hasil dari Manova dan Anova berdasarkan cluster k-means dan k-medoid menggunakan alpha 5%.

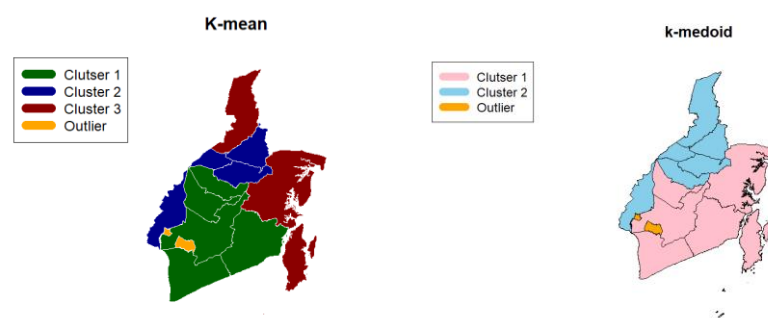
Tabel 8 Pengujian Manova dan Anova

Variabel	K-means	K-medoid
Multivariat	Signifikan	Signifikan
Garis Kemiskinan	Tidak Signifikan	Tidak Signifikan
Gini Ratio	Signifikan	Signifikan
P0	Signifikan	Signifikan
P1	Signifikan	Signifikan
P2	Signifikan	Signifikan
IPM	Signifikan	Signifikan

Dapat dilihat berdasarkan tabel diatasn k-means dan k-medoid memiliki performa yang serupa dimana mampu mengelompokkan data berdasarkan Gini Ratio, P0, P1, P2 dan IPM. Sedangkan untuk garis kemiskinan memiliki rata-rata yang sama. akan tetapi dapat kita lihat dengan boxplot sebelumnya baik k-means dan k-medoid adanya cluster yang memiliki garis berbeda dengan cluster lain. Dapat disimpulkan baik k-mean dan k-medoids memiliki performa yang serupa.

G. MAP Chart

Berdasarkan cluster yang dibentuk bisa dibuat peta kabupaten/kota di provinsi Kalimantan Selatan. berikut adalah hasil petanya.



Gambar 7 Peta Cluster K-means dan K-medoid

Dapat dilihat apabila persebaran berdasarkan peta cluster kmeans dan k-medoid daerah yang berdekatan cenderung memiliki kemiripan berdasarkan nilai IPM, dan berbagai indikator kemiskinan seperti garis kemiskinan, P0, P1, P2. Adanya outlier yaitu kota banjarماسin dan banjarbaru dikarenakan tinggi nilai IPM menunjukan pembangunan yang tidak merata. Pada cluster 1 di k-means didapatkan 2 daerah yang memiliki nilai gini ratio yang lebih tinggi dibandingkan dengan yang lain dikarenakan daerah tersebut merupakan daerah pertambangan. Pada

cluster 2 k-means dan k-medoids merupakan daerah yang harus diperhatikan lebih lanjut karena memiliki nilai IPM dan Garis kemiskinan yang rendah dan P0, P1 dan P2 yang tinggi. Hal ini juga bisa disebabkan letak daerah yang berada di utara dan tanah gambut yang membuat pembangunan infrastruktur lebih sulit.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil analisis didapatkan kesimpulan sebagai berikut ini.

1. Kondisi kemiskinan di Kalimantan Selatan relatif stabil dan konsisten antar kabupaten/kota dengan rendahnya varians Gini Ratio, P0, P1, dan P2. Tingginya varians IPM dan Garis Kemiskinan menandakan kurang meratanya pembangunan antar kabupaten/kota. Selain itu terdapat hubungan negatif antara IPM dan indikator kemiskinan seperti P0, P1 dan P2 yang menandakan bahwa jika IPM mengalami kenaikan maka kemiskinan di suatu daerah cenderung turun.
2. Kota Banjarbaru dan Banjarmasin memiliki nilai IPM yang jauh lebih tinggi dibandingkan dengan Kabupaten yang lain di provinsi Kalimantan Selatan, ini menunjukan kota di Kalimantan Selatan memiliki SDM yang lebih baik dari pada daerah yang lain. Berdasarkan clustering k-means dan k-medoid didapatkan cluster sebanyak 3 dan 2 dimana cenderung mampu mengelompokkan daerah berdasarkan nilai IPM, Garis Kemiskinan, Proporsi kemiskinan, keparahan kemiskinan dan kedalaman kemiskinan.
3. Daerah yang memiliki Proporsi kemiskinan, keparahan kemiskinan dan kedalaman kemiskinan tinggi dan IPM yang rendah cenderung berdekatan dan ada di bagian Utara Kalimantan Selatan. Kabupaten Tabalong dan Kotabaru memiliki gini ratio yang lebih tinggi dibandingkan dengan daerah yang lain ini disebabkan daerah tersebut merupakan penghasil Batubara.

B. Saran

Berikut ini saran berdasarkan hasil analisis yang didapatkan.

1. Untuk penelitian selanjutnya apabila ingin melakukan analisis faktor-faktor yang mempengaruhi IPM dan Kemiskinan bisa digunakan regresi spasial karena berdasarkan clustering adanya faktor wilayah yang mempengaruhi nilai IPM dan Kemiskinan
2. Mungkin bisa digunakan metode clustering lain seperti *density based clustering* (dbscan), *gaussian mixture model* (gmm), mean shift untuk dibandingkan juga performanya dengan metode k-means dan k-medoid.

DAFTAR PUSTAKA

- [1] L. Saptono, "Statistika Deskriptif untuk Pendidikan Ekonomi," Sanata Dharma University Press, Yogyakarta, 2022.
- [2] Y. Nalim and T. Salafudin, "Statistika Deskriptif," STAIN Pekalongan Press, Pekalongan, 2012.
- [3] R. A. Johnson and D. W. Winchurn, *Applied Multivariate Statistical Analysis* (Edisi 6 ed.), New Jersey: Prentice Hall International Inc, 2007.
- [4] A. C. Rencher, *Method of Multivariate Analysis* (Second Edition ed.), United States of Amerika: John Wiley & Sons, Inc, 2002.
- [5] J. Han and M. Kamber, *Data Mining: Concepts and Techniques* (Second Edition), New York: Morgan Kaufmann, 2006.
- [6] D. F. Azuri, Zulhanif and R. S. Pontoh, "Pengelompokan Kabupaten/Kota Di Pulau Jawa Berdasarkan Pembangunan Manusia Berbasis Fender Menggunakan Bisecting K-Means," *Prosiding Seminar Nasional MIPA*, pp. 78-83, 2016.
- [7] Rabiah, S. (2018). Kabupaten Indragiri Hilir Dalam Angka 2018. Kabupaten Indragiri Hilir: BPS Kabupaten Indragiri Hilir.
- [8] BPS. (2023). <https://jatim.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html>, BPS Jawa Timur.

LAMPIRAN

1. Lampiran Data

No	Kabupaten.Kota	Garis_Kemisikinan	Gini_Ratio	P0	P1	P2	IPM
1	BALANGAN	476190	0.266	5.83	0.667	0.119	69.73
2	BANJAR	514991	0.3	2.79	0.329	0.043	70.72
3	BANJARBARU	692830	0.341	4.17	0.168	0.015	79.68
4	BANJARMASIN	651383	0.33	4.74	0.723	0.178	77.97
5	BARITO KUALA	375707	0.289	4.75	0.711	0.173	67.37

6	HULU SUNGAI SELATAN	516857	0.26	4.54	0.493	0.092	69.76
7	HULU SUNGAI TENGAH	444331	0.228	5.92	0.749	0.147	69.7
8	HULU SUNGAI UTARA	526919	0.301	6.49	0.863	0.153	66.84
9	KOTABARU	505601	0.331	4.3	0.586	0.118	69.74
10	TABALONG	573018	0.337	5.87	0.702	0.146	73.13
11	TANAH BUMBU	557500	0.292	4.26	0.362	0.065	71.79
12	TANAH LAUT	575061	0.263	3.86	0.374	0.061	70.35
13	TAPIN	484113	0.291	3.6	0.328	0.071	71.02

2. Lampiran Syntax

```
library(ggpubr)
library(cluster)
library(factoextra)
library(ggcorrplot)
library(tidyverse)
setwd('C:/Users/ASUS/OneDrive - Institut Teknologi Sepuluh Nopember/Dokumen/Semester 7/Statof/eas')

dfs <- readxl::read_xlsx("data.xlsx") %>% as.data.frame();dfs[-1] <- lapply(dfs[-1], as.numeric) %>% as.data.frame()
dfs
rownames(dfs) <- dfs[[1]];dfs <- dfs[-1]
colnames(dfs)
dfs<-dfs[-1,]
cor_dfs <- cor(dfs);ggcorrplot(cor_dfs, lab = TRUE, type = "lower")
dim(df)
outliers <- function(x, probs = c(0.25, 0.75), labels = NULL) {
  if(length(probs) != 2) stop("Probability length must be 2.")
  Q1 <- quantile(x, probs=min(probs))
  Q3 <- quantile(x, probs=max(probs))
  iqr = Q3-Q1
  upper_limit = Q3 + (iqr*1.5)
  lower_limit = Q1 - (iqr*1.5)
  temp_outliers <- x > upper_limit | x < lower_limit
  cat("Upper Limit: ", round(upper_limit, 3), "\n", paste(max(probs) * 100, "%", sep = ""), "\nLower Limit: ", round(lower_limit, 3),
  "\n", paste(min(probs) * 100, "%", sep = ""), fill = TRUE)
  if(length(which(temp_outliers)) != 0){
    for(i in which(temp_outliers)) cat("Outlier: ", x[i], "\n", labels[i], fill = TRUE)
  } else cat("Tidak ada outlier.\n")
  return(!temp_outliers)
};for(cname in colnames(dfs)){
  cat("\n", cname, "\n")
  outliers(dfs[[cname]], probs = c(0.25, 0.75), labels = rownames(dfs[cname]))
  rm(cname)
}
dfs <- dfs[outliers(dfs$IPM),]

dfs1<-dfs

fviz_nbclust(dfs %>% scale(), kmeans,method = 'silhouette')
fviz_nbclust(dfs %>% scale(), pam)

c1 <- pam(dfs1 %>% scale(), k = 3)
c2 <- kmeans(dfs1 %>% scale(), centers = 2)
c11<-c1
c22<-c2

fviz_cluster(c1, main = "K-Medoids", ggtheme = cowplot::theme_cowplot(font_size = 16))
fviz_cluster(c2, data = dfs1, main = "KMeans", ggtheme = cowplot::theme_cowplot(font_size = 16))
dfs["PAM"] <- c1$clustering %>% as.factor(); dfs["KMeans"] <- c2$cluster %>% as.factor();rm(c1);rm(c2)
dfs
tname <- colnames(dfs);colnames(dfs) <- make.names(colnames(dfs));par(mfrow = c(2,3));
for(cname in make.names(colnames(dfs[1:6]))){
```

```

idx <- which(colnames(dfs) == cname)
boxplot(paste0(cname %>% as.symbol(), "~PAM") %>% as.formula(), data = dfs,
        ylab = tcname[idx], xlab = "Cluster (PAM)")
# cat(deparse(cname), fill = TRUE)
rm(cname)
};for(cname in make.names(colnames(dfs[1:6]))){
  idx <- which(colnames(dfs) == cname)
  boxplot(paste0(cname %>% as.symbol(), "~KMeans") %>% as.formula(), data = dfs,
          ylab = tcname[idx], xlab = "Cluster (KMeans)")
  # cat(deparse(cname), fill = TRUE)
  rm(cname);rm(idx)
};colnames(dfs) <- tcname;rm(tcname)
dfs
MVN::mvn(dfs[1:6], mvnTest = "royston", univariateTest = "SW", desc = FALSE)
m1 <- manova(dfs[1:6] %>% as.matrix() ~ PAM, data = dfs);summary(m1);summary.aov(m1)
m2 <- manova(dfs[1:6] %>% as.matrix() ~ KMeans, data = dfs);summary(m2);summary.aov(m2)
library(biotools)

library(car)

```