



PENERAPAN METODE *ENSEMBLE CLUSTERING* PADA PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TIMUR BERDASARKAN INDIKATOR PEMBANGUNAN EKONOMI

Syifa Nafi'ah Priansyah*, Universitas Negeri Yogyakarta
Dhoriva Urwatul Wutsqa, Universitas Negeri Yogyakarta
*e-mail: syifanafiah.2019@student.uny.ac.id

Abstrak. Pembangunan Ekonomi Regional merupakan pembangunan dimana pemerintah daerah dan warga negara bekerja sama untuk mengelola sumber daya yang ada. Pembangunan ekonomi pada Provinsi Jawa Timur mengalami ketimpangan yang terus berlanjut dikarenakan kesenjangan yang lebar antara kabupaten dan kota tertinggi dan terendah dalam hal PDRB per kapita. Berdasarkan fakta-fakta ini, pertumbuhan ekonomi di Provinsi Jawa Timur harus didistribusikan secara merata. Tujuan dari penelitian ini adalah menggunakan pendekatan *Ensemble clustering* untuk mengkategorikan dan mengidentifikasi fitur kabupaten dan kota Provinsi Jawa Timur menurut indikator pembangunan ekonomi. Penelitian ini memanfaatkan data sekunder yang dikumpulkan oleh Badan Pusat Statistik. Ada total tiga belas variabel yang terdiri dari sembilan numerik dan tiga kategori. Dimana pendekatan K-Means digunakan untuk data numerik dan metode K-modes untuk data kategorikal. Analisis yang dilakukan oleh pendekatan *Ensemble clustering* memanfaatkan hasil dari kedua metode tersebut. Penelitian ini menggunakan penggunaan *Davies Bouldin Index* (DBI) dan *Pseudo F-Statistics* untuk tujuan validasi. Hasil penelitian pengelompokan menggunakan metode *Ensemble clustering* didapatkan sebanyak 3 *cluster* dimana dilakukan pemeringkatan berdasarkan tingkatan. Dimana tingkat 3 memiliki 12 objek dengan tingkat pembangunan ekonomi tergolong rendah atau tingkat 3, dikarenakan memiliki nilai PDRB terendah dan angka kemiskinan yang tinggi. Sedangkan pada tingkat 1 terdapat 14 objek yang tergolong kedalam tingkat pembangunan ekonomi maju dimana pada *cluster* ini memiliki nilai PDRB tertinggi dan angka kemiskinan terendah. Selanjutnya tingkat yang terakhir yaitu tingkat 2 yang memiliki 12 objek dimana tingkat pembangunan ekonomi pada *cluster* ini tergolong menengah.

Kata kunci: *Pembangunan Ekonomi, Ensemble Clustering, Davies Bouldin Index, Pseudo F-Statistics.*

PENDAHULUAN

Pembangunan ekonomi bertujuan mengalihkan kegiatan ekonomi dari sektor primer ke sektor industri dan jasa, meningkatkan kesempatan kerja, memperkuat hubungan ekonomi regional, dan secara umum meningkatkan taraf hidup masyarakat (Setiarso, 2014). Pada saat yang sama, pembangunan ekonomi regional memerlukan pengelolaan sumber daya secara kooperatif dan pembentukan kemitraan antara sektor swasta dengan masyarakat dan pemerintah daerah untuk mendorong pertumbuhan ekonomi regional dan lapangan kerja.

Ketika merencanakan pertumbuhan suatu wilayah penting untuk mempertimbangkan batasan wilayah dan potensi wilayah tersebut. Arsyad (2010) menyatakan bahwa ada tiga hal yang dapat dilihat sebagai tanda bahwa pertumbuhan ekonomi telah berhasil. Kesejahteraan ekonomi dan pendapatan per kapita merupakan metrik moneter. Ketiga, ada indikator campuran yang menggabungkan indikator dasar Susenas dan Indeks Pembangunan Manusia (IPM). Indikator keempat dan terakhir adalah indikator nonmoneter yang terdiri dari indikator sosial dan indeks kualitas hidup. Tujuan pembangunan ekonomi ini adalah untuk meningkatkan PDB suatu wilayah atau negara melampaui apa yang diharapkan dari pertumbuhan penduduk saja. Berdasarkan penelitian terdahulu, tingkat ketimpangan di Provinsi Jawa Timur masih cukup tinggi karena masih tingginya kesenjangan antara kabupaten/kota dengan PDRB per kapita tertinggi dan terendah (Huda & Santoso, 2014). Diperlukan kebijakan untuk mengatasi masalah ketimpangan dan ketimpangan di Jawa Timur karena masih terdapat banyak sekali perbedaan antar kabupaten/kota. Para pengambil kebijakan di Provinsi Jawa Timur menginginkan adanya peta regional kabupaten/kota di provinsi tersebut berdasarkan indikator pembangunan ekonomi; analisis kluster menawarkan salah satu solusi untuk tantangan ini.

Dalam hal analisis data numerik, metodologi kluster K-Means merupakan salah satu teknik non-hierarkis yang paling populer dan banyak digunakan. Dengan menggunakan kluster K-Means, kumpulan data dipartisi menjadi K kelompok yang berbeda dan tidak tumpang tindih. Menentukan jumlah total kluster, atau K, merupakan langkah awal dalam melakukan analisis K-Means (Muhima et al., 2021). Tujuan penelitian ini adalah untuk mengklasifikasi kabupaten dan kota di Jawa Timur menurut metrik pembangunan ekonomi tahun 2019.

Variabel yang digunakan adalah nilai indikator pertumbuhan ekonomi di kabupaten dan kota di Provinsi Jawa Timur. Tujuan utama penelitian ini adalah mengoptimalkan indikator pembangunan ekonomi di Jawa Timur dalam upaya memperbaiki pertumbuhan ekonomi yang belum merata di wilayah tersebut. Para peneliti di Provinsi Jawa Timur berharap bahwa temuan penelitian mereka dapat memberikan kredibilitas pada strategi pemerintah untuk menyeimbangkan pertumbuhan ekonomi di seluruh kabupaten dan kota di provinsi tersebut.

METODE

Buku berjudul "Provinsi Jawa Timur Dalam Angka Tahun 2022" dan data sekunder yang dikumpulkan dari Badan Pusat Statistik Provinsi Jawa Timur (<https://jatim.bps.go.id/>) menjadi sumber informasi utama yang digunakan dalam penelitian ini. Data Survei Sosial Ekonomi Nasional (SUSENAS) 2019 yang digunakan di sini mencakup semua kabupaten dan kota di Jawa Timur dan memuat berbagai metrik pembangunan ekonomi. Format data yang digunakan adalah numerik dan kategoris. Informasi tentang kabupaten dan kota di Provinsi Jawa Timur yang akan berdampak pada pertumbuhan ekonomi daerah pada tahun 2019 merupakan bagian dari data yang dievaluasi dalam penelitian ini (Arsyad, 2010). Tabel 1 menunjukkan dua belas variabel yang membentuk indikator, yang meliputi sembilan variabel numerik dan tiga kategori kategoris. Indikator-indikator ini meliputi indikator moneter, non-moneter, dan campuran, yang semuanya telah dibahas pada bab sebelumnya.

Tabel 1. Variabel Penelitian

Variabel Numerik	
Variabel	Deskripsi
X_1	Produk Domestik Regional Bruto Atas Dasar Harga Berlaku
X_2	Indeks Pembangunan Manusia

X_3	Angka Partisipasi Sekolah berusia 7-12 tahun
X_4	Angka Partisipasi Sekolah berusia 13-15 tahun
X_5	Angka Partisipasi Sekolah berusia 16-18 tahun
X_6	Tingkat Pengangguran Terbuka
X_7	Tingkat Partisipasi Angkatan Kerja
X_8	Persentase Kemiskinan

Variabel Kategori

Variabel	Keterangan	Kategori
X_9	Tingkat Kriminalitas	0 : Rendah 1 : Sedang 2 : Tinggi
X_{10}	Sumber Air Minum Utama	0 : Air Hujan 1 : Danau/ Sungai/ Kolam 2 : Mata Air 3 : Sumur 4 : Leding/Pipa 5 : Air Kemasan
X_{11}	Sumber Penerangan Listrik Utama	0 : Bukan Listrik 1 : Listrik Non-PLN 2 : Listrik PLN

Pada penelitian ini menggunakan analisis *cluster* dengan algoritma *clustering ensemble*, dimana analisis *cluster* untuk variabel kategorik menggunakan metode *K-Means*, sedangkan variabel kategorik menggunakan *K-modes*. Peneliti menggunakan RStudio untuk pemrograman dan Microsoft Excel 2019 untuk pengolahan data. Variabel penelitian dianalisis secara deskriptif terkait dengan deskripsi data. Selain itu, teknik *clustering ensemble* digunakan untuk analisis data, dengan metode *K-Means* diterapkan pada variabel numerik dan metode *K-modes* digunakan untuk variabel kategori. Selain itu, setiap teknik menjalani validasi klaster menggunakan *Pseudo F-Statistics* dan *Davies Bouldin Index (DBI)* untuk memastikan ukuran klaster yang optimal. Untuk informasi lebih lanjut, berikut adalah langkah-langkah yang terlibat dalam analisis data saat menggunakan metode *clustering ensemble*:

1. Data berdasarkan indikator pertumbuhan ekonomi harus dijelaskan menggunakan statistik deskriptif yang diterapkan pada kategori data numerik dan kategori.
2. Pisahkan data numerik dan kategori.
3. Untuk data numerik dilakukan standarisasi data terlebih dahulu sebelum dilakukan analisis *clustering*. Standarisasi data yang diperlukan dalam penelitian ini menggunakan *Z score*.

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (1)$$

keterangan :

S_j : standar deviasi variabel ke- j

n : banyaknya objek penelitian

X_{ij} : objek pada pengamatan ke- i pada variabel ke- j

\bar{X}_j : rata-rata variabel ke- j

4. Melakukan analisis pengelompokan data numerik menggunakan metode *K-Means*.

- a) Inisialisasi K (jumlah *cluster*) dan *centroid* awal (K data yang digunakan sebagai *centroid* awal).
- b) Mengalokasikan setiap data ke *centroid* yang terdekat menggunakan salah satu metode pengukuran jarak salah satunya jarak *Euclidean*.
- c) Perhitungkan kembali *centroid* berdasarkan data yang termasuk kedalam *cluster*-nya masing-masing.
- d) Lakukan kembali langkah (2) dan (3) sampai mencapai keadaan konvergen, di mana tidak ada perpindahan data ke *cluster* lain.

Dalam menentukan *centroid* untuk memperoleh titik *centroid* (C), diperoleh dengan cara mengambil nilai rata-rata semua atribut yang dimiliki oleh setiap data dalam setiap kelompok yang didefinisikan sebagai berikut:

$$c_{kj} = \frac{1}{nk} \sum_{i=1}^{nk} x_{ij} \quad (2)$$

keterangan :

Dalam menghitung jarak *Euclidean* antara objek i dan pusat *cluster* ke- k

c_{kj} : *centroid*/rata-rata *cluster* ke- k untuk variabel ke- j

nk : banyaknya pengamatan/atribut pada *cluster* ke- k

i : indeks objek

k : indeks *cluster*

j : indeks dari variabel

x_{ij} : objek pada pengamatan ke- i untuk variabel ke- j
 digunakan persamaan berikut ini:

$$d(i, k) = \sqrt{\sum_{j=1}^p (x_{ij} - c_{kj})^2} \quad (3)$$

keterangan :

$d(i, k)$: jarak *Euclidean cluster* ke- k objek ke- i

k : indeks *cluster*

j : Indeks dari variabel

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

c_{kj} : *centroid*/rata-rata *cluster* ke- k untuk variabel ke- j

5. Menerapkan teknik pengelompokan dalam proses analisis menggunakan metode *K-modes* untuk data kategorikal.

- a) Menetapkan banyaknya *cluster* dan *centroid* awal untuk setiap *cluster*.
- b) Tentukan seberapa jauh jarak *centroid* dari setiap item, lalu tempatkan item tersebut dalam klaster dengan *centroid* terdekat. Ulangi proses ini hingga semua objek dikelompokkan. Jika atribut atau objek memiliki nilai jarak yang sama, maka objek masuk kedalam *cluster* terkecil. Jarak objek dengan *centroid* terdekat tersebut menggunakan jarak *simple matching* pada persamaan berikut.

$$d(i, k) = \sum_{j=1}^p \delta(x_{ij}, x_{kj}) \quad (4)$$

dengan :

$$\delta(i, k) = \begin{cases} 0, & x_{ij} = x_{kj} \\ 1, & x_{ij} \neq x_{kj} \end{cases} \quad (5)$$

keterangan :

- $d(i, k)$: jarak *Euclidean* antara objek i dan pusat *cluster* ke- k pada variabel ke- j
 j : banyaknya variabel untuk $j = 1, 2, \dots, p$
 x_{ij} : nilai objek ke- i pada variabel ke- j
 x_{kj} : nilai pusat *cluster* ke- k pada variabel ke- j

- c) Setelah semua objek telah dialokasikan pada setiap *cluster*, uji ulang ketidaksamaan objek terhadap *centroid* saat ini, Setiap kali terjadi pembaruan, pastikan untuk memperbarui pusat klaster setiap kelompok dengan nilai yang paling sering muncul atau yang merupakan modus setiap kelompok.
 - d) Ulangi langkah (3) sampai tidak ada objek yang berpindah kelompok atau iterasi berhenti.
6. Menyatukan hasil pengelompokan data numerik dan kategorik.
 7. Melakukan *clustering ensemble* untuk mengelompokkan hasil, karena hasil ini bersifat kategoris, sehingga digunakan metode *clustering* untuk jenis data tersebut.
 8. Evaluasi hasil proses *fuzzy clustering ensemble* menggunakan indeks ukuran validasi. Dalam penelitian ini digunakan 2 validasi indeks:
 - a) *Davies Bouldin Index* (DBI)

Pendekatan penilaian internal mencakup indeks validitas Davies Bouldin (DB) sebagai salah satu alat evaluasi klasternya; pendekatan ini menggunakan nilai separasi dan kohesi untuk mengevaluasi klaster (Alith, 2015). Ukuran klaster yang optimal ditentukan oleh nilai DB yang lebih kecil. (Davies & Bouldin, 1979) Berikut ini adalah metode untuk menghitung indeks validitas Davies Bouldin (DBI):

1. Menetapkan nilai *centroid* masing-masing *cluster* untuk diuji.
2. Melakukan perhitungan nilai S_k atau Seberapa jauh titik data dari titik pusat klaster, yang digunakan untuk menemukan derajat kekompakan dalam klaster ke- k .

$$S_k = \frac{1}{n_k} \sum_{i=1}^{n_k} d(i, k) \quad (6)$$

dengan,

$$d(i, k) = \sqrt{\sum_{j=1}^p (x_{ik} - c_{kj})^2} \quad (7)$$

keterangan :

- S_k : rata-rata jumlah nilai jarak objek terhadap *centroid* pada *cluster* ke- k
 K : banyaknya *cluster*
 n_k : banyaknya objek pada *cluster* ke- k
 $d(i, k)$: jarak *Euclidean* objek ke- i terhadap *centroid* *cluster* ke- k untuk $i = 1, 2, \dots, n$

- x_{ik} : objek ke- i pada cluster ke- k
 c_{kj} : *centroid cluster* ke- k variabel ke- j

3. Melakukan perhitungan nilai $d(c_k, c_{k'})$ atau jarak antar *centroid* Digunakan untuk tujuan menilai jarak antar kelompok.

$$d(c_k, c_{k'}) = \sqrt{\sum_{j=1}^p (c_{kj} - c_{k'j})^2} \quad (8)$$

keterangan :

$d(c_k, c_{k'})$: nilai jarak antar *centroid* pada *cluster* ke- k dan *cluster* ke- k' dimana $k \neq k'$

c_{kj} : *centroid cluster* ke- k pada variabel ke- j

$c_{k'j}$: *centroid cluster* ke- k' pada variabel ke- j

4. Setelah mendapat nilai kohesi dan separasi, langkah berikutnya adalah menghitung rasio ($R_{k,k'}$). *Cluster* Gugus yang memiliki kekompakan terbatas dan pemisahan signifikan dianggap sangat baik.

$$R_{k,k'} = \frac{S_k + S_{k'}}{d(c_k, c_{k'})} \quad (9)$$

5. Jumlah klaster, atau Indeks Davies Bouldin (DBI), dihitung menggunakan nilai rasio dari persamaan (2.21).

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{k \neq k'} R_{k,k'} \quad (10)$$

b) *Pseudo F-Statistics*

Jumlah kelompok ideal ditunjukkan oleh skor maksimum pada statistik pseudo F ini menyiratkan bahwa terdapat tingkat homogenitas yang tinggi dalam kelompok tetapi tingkat heterogenitas yang tinggi di antara kelompok. Untuk mendapatkan nilai statistik pseudo F, persamaan berikut digunakan.

$$Pseudo F - Statistics = \frac{\left(\frac{R^2}{i-1}\right)}{\left(\frac{1-R^2}{n-1}\right)} \quad (11)$$

dengan:

$$R^2 = \frac{(SST - SSW)}{SST} \quad (12)$$

$$SST = \sum_{i=1}^n \sum_{k=1}^c \sum_{j=1}^J (x_{ik}^j - \bar{x}^j)^2 \quad (13)$$

$$SSW = \sum_{i=1}^n \sum_{k=1}^c \sum_{j=1}^J (x_{ik}^j - \bar{x}_k^j)^2 \quad (14)$$

keterangan :

- SST : total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan
- SSW : total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya
- n : banyaknya objek
- c : banyaknya kelompok
- J : banyaknya variabel
- x_{ik}^j : objek ke- i pada kelompok ke- k dan variabel ke- j
- \bar{x}^j : rata-rata seluruh objek pada variabel ke- j
- \bar{x}_k^j : rata-rata objek pada kelompok ke- k dan variabel ke- j

HASIL DAN PEMBAHASAN

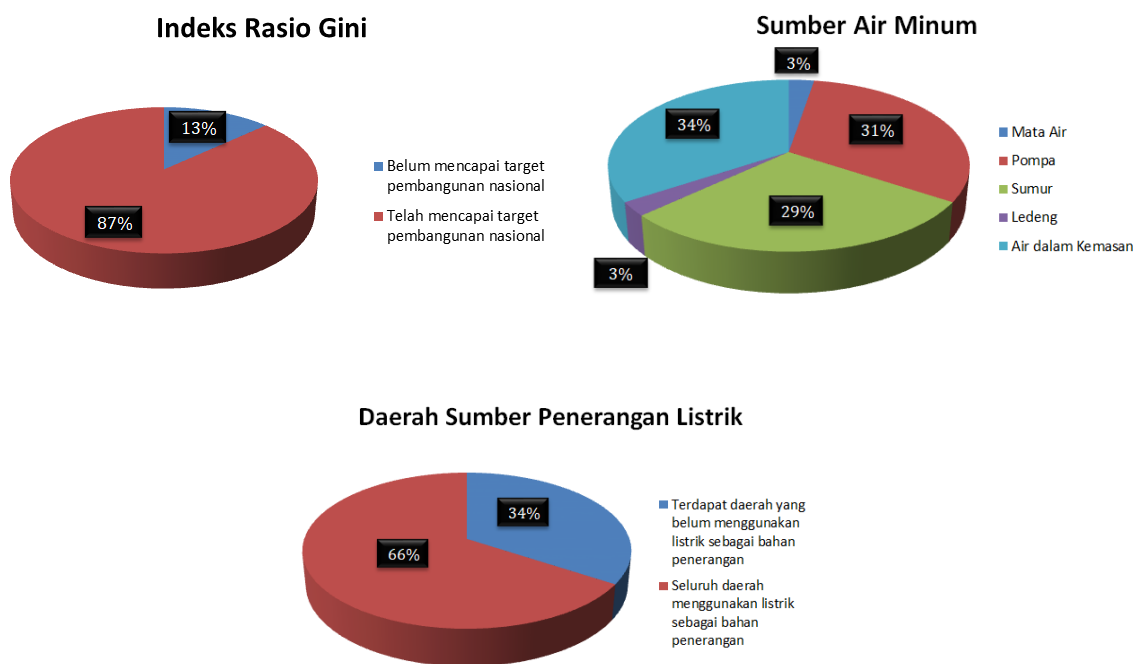
Analisis Deskriptif

Analisis ini akan memberikan ringkasan deskriptif dari data penelitian yang mencakup nilai rata-rata, minimum, dan maksimum. Tabel 2 akan memberikan informasi tentang data numerik dan analisis deskriptif.

Tabel 2. Statistika Deskriptif pada Variabel Numerik

Variabel	Mean	Min	Max
X_1	62476	6792	580489
X_2	71,69	61,94	82,22
X_3	99,66	98,84	100
X_4	97,40	91,96	99,64
X_5	75,21	50,18	95,83
X_6	3,65	0,91	5,88
X_7	69,65	63,44	79,62
X_8	10,30	3,81	20,71
X_9	835,68	108	3377

Indikator pembangunan ekonomi kabupaten dan kota di Jawa Timur dipelajari menggunakan statistik numerik dan deskriptif untuk lebih memahami fitur data kategorikal. Tiga variabel membentuk data kategorikal: indeks rasio Gini, sumber utama air minum, dan wilayah dengan penerangan listrik. Analisis deskriptif tersebut disajikan pada diagram lingkaran berikut ini.



Gambar 1. Diagram Lingkaran Variabel Kategorik

Sebaran variabel kategorik kabupaten/kota di Provinsi Jawa Timur dapat dilihat pada Gambar 1. Berdasarkan indeks gini ratio, terlihat bahwa sebagian besar Provinsi Jawa Timur telah memenuhi atau melampaui target pembangunan nasional. Sumber air minum di Provinsi Jawa Timur mayoritas menggunakan air dalam kemasan dan sebagian besar sudah menggunakan listrik sebagai sumber penerangan.

Analisis Kluster Metode *K-Means*

Data numerik yang distandarkan dianalisis menggunakan analisis kluster dengan menggunakan pendekatan *K-Means*. Validasi kluster diperlukan untuk menemukan ukuran kluster yang ideal. Pendekatan Davies Bouldin Index (DBI) dan Pseudo *F-Statistics* digunakan untuk validasi dalam penelitian ini yang menggunakan metode *K-Means*. Pada Tabel 3 berikut ini disajikan hasil perhitungan kalster pada data numerik menggunakan kedua validasi kluster tersebut.

Tabel 3. Hasil Perhitungan Validasi *Cluster* Data Numerik

Jumlah <i>Cluster</i>	DBI	<i>Pseudo F-Statistics</i>
2	2,5177	2,0566
3	1,5083	34,4191
4	3,7662	22,1625
5	15,6062	16,3505

Pada Tabel 3 dapat dilihat bahwa jumlah *cluster* sebanyak 3 merupakan jumlah *cluster* optimum. *Cluster* optimum didapatkan saat nilai DBI mencapai nilai minimum dan nilai *Pseudo F-Statistics* mencapai nilai maksimum. Hasil analisis *cluster* optimum data numerik menggunakan metode *K-Means* sebanyak 3 *cluster* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengelompokan 3 *Cluster* Metode *K-Means*

<i>Cluster</i>	Wilayah Kabupaten/Kota
1	Kabupaten Ponorogo, Kabupaten Tulungagung, Kabupaten Blitar, Kabupaten Kediri, Kabupaten Malang, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Pasuruan, Kabupaten Mojokerto, Kabupaten Jombang, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Ngawi, Kabupaten Bojonegoro, Kabupaten Lamongan, Kabupaten Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Batu.
2	Kabupaten Sidoarjo, Kota Surabaya.
3	Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Lumajang, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Tuban, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep.

Analisis Kluster Metode *K-Modes*

Analisis pengelompokan pada data kategorikal dilakukan dengan menggunakan pendekatan *K-modes*. Dalam penelitian ini, tiga variabel kategorikal akan digunakan. Tabel 1 menampilkan faktor-faktor tersebut. Mirip dengan pendekatan *K-Means*, metode *K-modes* menggunakan validasi kluster untuk memilih jumlah kluster yang ideal. Davies Bouldin Index (DBI) dan Pseudo *F-Statistics* adalah alat yang digunakan untuk validasi kluster. Tabel 5 menunjukkan hasil perhitungan validasi kluster (*K-modes*) menggunakan data kategorikal.

Tabel 5. Hasil Perhitungan Validasi *Cluster* Data Kategorik

Jumlah Cluster	DBI	<i>Pseudo F-Statistics</i>
2	1,5817	12,5365
3	2,0992	10,1986
4	1,8880	24,7463
5	1,4793	26,8138

Pada Tabel 5 dapat dilihat bahwa jumlah *cluster* sebanyak 5 merupakan jumlah *cluster* optimum. Hal tersebut dikarenakan nilai DBI minimum dan nilai *Pseudo F-Statistics* maksimum didapatkan pada *cluster* 5. Hasil analisis *cluster* optimum data kategorikal menggunakan metode *K-modes* sebanyak 5 *cluster* dapat dilihat pada Tabel 6.

Tabel 6. Hasil Akhir Pengelompokan Data Kategorik

<i>Cluster</i>	Wilayah Kabupaten/Kota
1	Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Malang
2	Kabupaten Tulungagung, Kabupaten Kediri, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Pasuruan, Kabupaten Jombang, Kabupaten Magetan, Kota Kediri, Kota Batu

<i>Cluster</i>	<i>Wilayah Kabupaten/Kota</i>
3	Kabupaten Ponorogo, Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Probolinggo, Kabupaten Nganjuk, Kabupaten Madiun, Kabupaten Bojonegoro, Kabupaten Bangkalan
4	Kabupaten Blitar, Kabupaten Lumajang, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep, Kota Blitar
5	Kabupaten Sidoarjo, Kabupaten Mojokerto, Kabupaten Ngawi, Kabupaten Tuban, Kabupaten Lamongan, Kabupaten Gresik, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya

Analisis Klaster Metode *Ensemble*

Pada analisis sebelumnya, data numerik dan data kategorik telah dianalisis menggunakan metode *K-Means* dan *K-modes*, selanjutnya dilakukan perhitungan analisis *clustering* yang terakhir yaitu *Ensemble clustering*. Dalam metode *Ensemble clustering* data yang digunakan merupakan hasil dari metode *K-Means* dan *K-modes*. Dikarenakan data yang digunakan merupakan hasil analisis *cluster* yang bertipe kategorik, maka metode *Ensemble clustering* menggunakan perhitungan yang sama dengan metode *K-modes*. Dalam menemukan jumlah *cluster* optimal metode *Ensemble clustering* memiliki perhitungan validasi *cluster* yang sama dengan metode *K-Means* dan *K-modes*. Validasi *Cluster* yang digunakan melibatkan metode *Davies Bouldin Index* (DBI) dan *Pseudo F-Statistics*. Validasi *cluster* pada data campuran dengan menggunakan metode *Ensemble clustering* memiliki hasil perhitungan yang dapat ditemukan di Tabel 7.

Tabel 7. Hasil Perhitungan Validasi Cluster Data Campuran

<i>Jumlah Cluster</i>	<i>DBI</i>	<i>Pseudo F-Statistics</i>
2	3,2651	2,9694
3	1,0174	22,7974
4	7,7005	2,0948
5	1,2932	10,5994

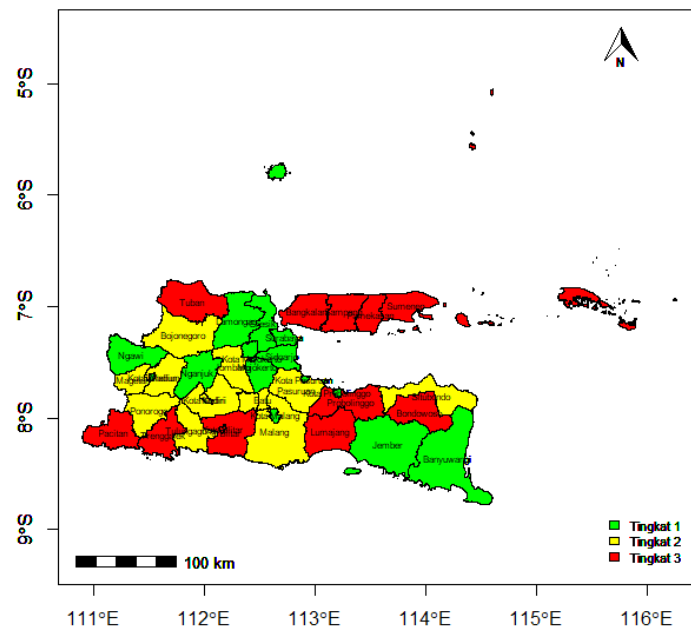
Pada Tabel 7 dapat dilihat bahwa jumlah *cluster* sebanyak 3 merupakan jumlah *cluster* optimum. Hal tersebut dikarenakan nilai DBI minimum dan nilai *Pseudo F-Statistics* maksimum didapatkan pada *cluster* 3. Hasil analisis *cluster* optimum data campuran menggunakan metode *Ensemble clustering* sebanyak 3 *cluster* dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengelompokan 3 Cluster Metode Ensemble

Cluster	Wilayah Kabupaten/Kota
1	Kabupaten Pacitan, Kabupaten Trenggalek, Kabupaten Blitar, Kabupaten Lumajang, Kabupaten Bondowoso, Kabupaten Probolinggo, Kabupaten Tuban, Kabupaten Bangkalan, Kabupaten Sampang, Kabupaten Pamekasan, Kabupaten Sumenep, Kota Blitar
2	Kabupaten Jember, Kabupaten Banyuwangi, Kabupaten Sidoarjo, Kabupaten Mojokerto, Kabupaten Nganjuk, Kabupaten Ngawi, Kabupaten Lamongan, Kabupaten Gresik, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya
3	Kabupaten Ponorogo, Kabupaten Tulungagung, Kabupaten Kediri, Kabupaten Malang, Kabupaten Situbondo, Kabupaten Pasuruan, Kabupaten Jombang, Kabupaten Madiun, Kabupaten Magetan, Kabupaten Bojonegoro, Kota Kediri, Kota Batu

Pemetaan Hasil Analisis Clustering Metode Ensemble

Pada analisis *clustering* menggunakan metode *Ensemble* yang sudah dilakukan pada perhitungan sebelumnya, didapatkan hasil yaitu sebanyak 3 *cluster*. Tujuan dari perhitungan dan validasi klaster adalah untuk mendapatkan hasil klaster terbaik. Berdasarkan data, terlihat bahwa klaster 1 terdiri dari 12 kabupaten/kota, klaster 2 terdiri dari 14 kabupaten/kota, dan klaster 3 terdiri dari 12 kabupaten/kota. Hasil pemerinkatan menunjukkan bahwa klaster 2 memiliki infrastruktur ekonomi paling maju, klaster 3 memiliki infrastruktur ekonomi terbaik kedua, dan klaster 1 memiliki infrastruktur ekonomi terburuk. Tingkat perkembangan ekonomi menunjukkan sejauh mana suatu tempat telah berkembang secara ekonomi. Tingkat yang lebih rendah menunjukkan tingkat yang lebih maju. Gambar 2 adalah peta kabupaten/kota di Jawa Timur yang disusun berdasarkan tingkat perkembangannya.



Gambar 2. Peta Persebaran Tingkat Pembangunan Ekonomi Provinsi Jawa Timur

Sebagian besar kabupaten/kota di Provinsi Jawa Timur berada pada level 2 atau level 3, seperti terlihat pada Gambar 2. Tabel 9 menunjukkan gambaran umum indikator pembangunan ekonomi, sekaligus kajian karakteristik pada masing-masing level daerah.

Tabel 9. Karakteristik Pembangunan Ekonomi pada Setiap *Cluster*

Tingkat	Cluster	Jumlah Kabupaten/Kota	Karakteristik Pembangunan Ekonomi	Status Pembangunan
1	2	14	Pada <i>cluster</i> ini memiliki wilayah dengan nilai PDRB, IPM, angka partisipasi sekolah pada usia SD dan SMA tertinggi, dan angka kemiskinan terendah, mayoritas sumber air minum yang digunakan pada wilayah ini adalah air dalam kemasan.	Maju
2	3	12	Wilayah pada <i>cluster</i> ini hampir seluruhnya memiliki nilai variabel sedang, hanya pada variabel angka partisipasi sekolah usia SMP memiliki nilai tertinggi daripada <i>cluster</i> yang lain.	Menengah
3	1	12	Wilayah pada <i>cluster</i> ini memiliki nilai PDRB, IPM, angka partisipasi sekolah terendah dan angka kemiskinan yang tertinggi, dimana sumber air minum mayoritas menggunakan sumur. Akan tetapi, wilayah pada <i>cluster</i> ini memiliki tingkat kriminalitas terendah.	Rendah

Klaster 2 yang meliputi empat belas kabupaten/kota di Provinsi Jawa Timur memiliki nilai PDRB terbesar dan tingkat kemiskinan terendah, sebagaimana terlihat pada Tabel 9. Oleh karena itu, pembangunan ekonomi tingkat 1 tergolong maju. Selain itu, 12 kabupaten/kota di Provinsi Jawa Timur masuk dalam klaster 3 yang merupakan tingkat 2 untuk indikator pembangunan ekonomi yang tergolong sedang. Sementara itu, klaster 1 yang meliputi 12 kabupaten/kota di Provinsi Jawa Timur memiliki nilai PDRB yang relatif rendah dan tingkat kemiskinan yang relatif tinggi, yang menunjukkan derajat pembangunan ekonomi yang relatif rendah pada tingkat 3.

Lincoln Arsyad (2010) berpendapat bahwa pembangunan ekonomi adalah peningkatan Produk Nasional Bruto (PNB) per tahun dan mempertahankan peningkatan tersebut hingga berlangsung dalam jangka panjang dengan memperhatikan perbaikan sistem kelembagaan di segala bidang (misalnya ekonomi, politik, hukum, sosial, dan budaya). Hasil pengelompokan dan analisis karakteristik tersebut sejalan dengan teori tersebut. Sebagian besar kabupaten dan kota di Provinsi Jawa Timur berada pada level perkembangan ekonomi antara bawah dan menengah, berdasarkan hasil pengelompokan metode klaster Ensemble.

SIMPULAN

Penelitian ini menggunakan obyek sebanyak 38 kabupaten dan kota di Provinsi Jawa

Timur. Data sekunder yang digunakan dalam penelitian ini adalah data yang dikumpulkan oleh Badan Pusat Statistik tahun 2019. Total variabel yang digunakan adalah tiga belas variabel, yaitu sembilan variabel numerik dan tiga variabel kategori. Langkah awal dalam pembuatan kluster menggunakan teknik ensemble clustering adalah menghitung jumlah kluster yang terbentuk, yaitu 2, 3, 4, dan 5. Dari keempat hasil pembentukan cluster dihitung nilai validitas masing-masing cluster menggunakan *Davies Bouldin Index* (DBI) dan *Pseudo F-Statistic* untuk menentukan cluster terbaik yaitu diperoleh sebanyak 3 cluster. Hasil penelitian pengelompokan menggunakan metode *ensemble clustering* didapatkan sebanyak 3 cluster dimana dilakukan pemeringkatan berdasarkan tingkatan pembangunan ekonomi. Cluster 3 memiliki 12 objek dengan tingkat pembangunan ekonomi tergolong rendah atau tingkat 3, dikarenakan memiliki nilai PDRB terendah dan angka kemiskinan yang tinggi. Sedangkan pada cluster 2 terdapat 14 objek yang tergolong ke dalam tingkat pembangunan ekonomi maju dimana pada cluster ini memiliki nilai PDRB tertinggi dan angka kemiskinan terendah. Selanjutnya cluster yang terakhir yaitu cluster 3 memiliki 12 objek dimana tingkat pembangunan ekonomi pada cluster ini tergolong menengah. Hasil pembentukan cluster ditampilkan dalam bentuk peta tingkat pembangunan ekonomi pada kabupaten/kota di Provinsi Jawa Timur.

UCAPAN TERIMA KASIH

Terimakasih kepada koor Prodi Statistika dan seluruh Dosen Statistika yang telah memberikan ilmu hingga terselesainya artikel ini.

DAFTAR PUSTAKA

- Alith, M. F. (2015). *Klasterisasi proses seleksi pemain menggunakan algoritma k-means (study kasus: tim hockey Kabupaten Kendal)*. Skripsi: Universitas Dian Nuswantoto.
- Alqurashi, T., & Wang, W. (2019). Metode ansambel pengelompokan. *Jurnal Internasional Pembelajaran Mesin dan Sibernetika*, 1227-1246.
- Arsyad, L. (2010). *Ekonomi pembangunan (5th ed)*. Yogyakarta: BP STIE YKP.
- Badan Pusat Statistik Provinsi Jawa Timur. (2019). *Tingkat Partisipasi Angkatan Kerja (TPAK) (Persen), 2016-2018*. Retrieved Mei 11, 2023, from BPS Provinsi Jawa Timur: <https://jatim.bps.go.id/indicator/6/277/1/tingkat-partisipasi-angkatan-kerja-tpak-.html>
- Badan Pusat Statistik Provinsi Jawa Timur. (2020). *Tingkat Pengangguran Terbuka (TPT) Provinsi Jawa Timur (Persen), 2017-2019*. Retrieved Mei 11, 2023, from BPS Provinsi Jawa Timur: <https://jatim.bps.go.id/indicator/6/54/2/tingkat-pengangguran-terbuka-tpt-provinsi-jawa-timur.html>
- Badan Pusat Statistik Provinsi Jawa Timur. (2022). *Provinsi Jawa Timur dalam angka 2022*. Surabaya: Badan Pusat Statistik.
- Huda, M., & Santoso, E. B. (2014). Pengembangan daya saing daerah kabupaten/kota di Propinsi Jawa Timur berdasarkan potensi daerahnya. *Jurnal Teknik Pomits*, 81-86.
- Muhima, R. R., Kurniawan, M., Wardhana, S. R., Yudhana, A., Sunardi, Rahmawati, W. M., et al. (2022). *Kupas tuntas algoritma clustering: konsep, perhitungan manual, dan program*. Yogyakarta: ANDI.
- Pemerintah Daerah. (2015). *Undang-undang (UU) Nomor 9 Tahun 2015 tentang Perubahan Kedua Atas Undang-Undang Nomor 23 Tahun 2014 tentang Pemerintahan Daerah*. Jakarta: LN.2015/NO.58, TLN NO.5679, LL SETNEG.

- Suhaeni, C., Kurnia, A., & Ristiyanti. (2018). Perbandingan hasil pengelompokan menggunakan analisis cluster berhirarki, k-means cluster, dan cluster ensemble (studi kasus data indikator pelayanan kesehatan ibu hamil). *Jurnal Media Infotama*, 31-38.
- Tayal, D. K., Jain, A., Arora, S., Agarwal, S., Gupta, T., & Tyagi, N. (2015). Crime detection and criminal identification in India using data mining techniques. *AI & Soc*, 117-127.
- Todaro, M. P., & Smith, S. C. (2011). *Pembangunan ekonomi (9th ed)*. Jakarta: Erlangga.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S (4th ed)*. New York: Springer.
- Walesiak, M., & Dudek, A. (2020). The choice of variable normalization method in cluster analysis. *Education Excellence and innovation management: a 2025 vision to sustain economic development during global challenges* (pp. 325-340). Seville, Spain: International Business Information Management Association (IBIMA).
- Weih, C., Ligges, U., Luebke, K., & Raabe, N. (2005). klaR analyzing German business cycles. In D. Baier, R. Decker, & L. Schmidt-Thieme, *Data analysis and decision support* (pp. 335-343). Berlin: Springer-Verlag.