

Klasterisasi Kecamatan DKI Jakarta berdasarkan Tingkat Pertumbuhan Angka Kesembuhan

David Imago Dei Gloriawan

10 April 2021

Pendahuluan

rumusan masalah, hipotesis, batasan masalah

Rumusan Masalah

- Apakah **tingkat pertumbuhan angka kesembuhan** covid-19 di setiap kecamatan DKI Jakarta **mempunyai pola tertentu** sehingga membentuk kelompok-kelompok kecamatan?

Hipotesis

- Terdapat **pola tertentu pada tingkat pertumbuhan angka kesembuan** sehingga **membentuk kelompok-kelompok kecamatan** dengan karakteristik tersendiri

Batasan Masalah

- **Pemrograman** : Bahasa R – RStudio
- **Dataset** : Angka kesembuhan setiap kecamatan per hari
- **Dataset** : Jangka waktu 1 Januari 2021 – 8 April 2021 (98 file)
- **Output** : Kelompok-kelompok kecamatan berdasarkan tingkat pertumbuhan angka kesembuhan

Terminologi & Catatan Penting

Terminologi

- **Tingkat pertumbuhan angka kesembuhan** : Kenaikan/penurunan angka kesembuhan covid-19 suatu kecamatan dari hari ke hari.

Catatan

- Dataset terdiri dari 98 file, dimana masing-masing file merupakan data harian dari 1 Januari 2021 – 8 April 2021 (98 hari).

Pra Proses

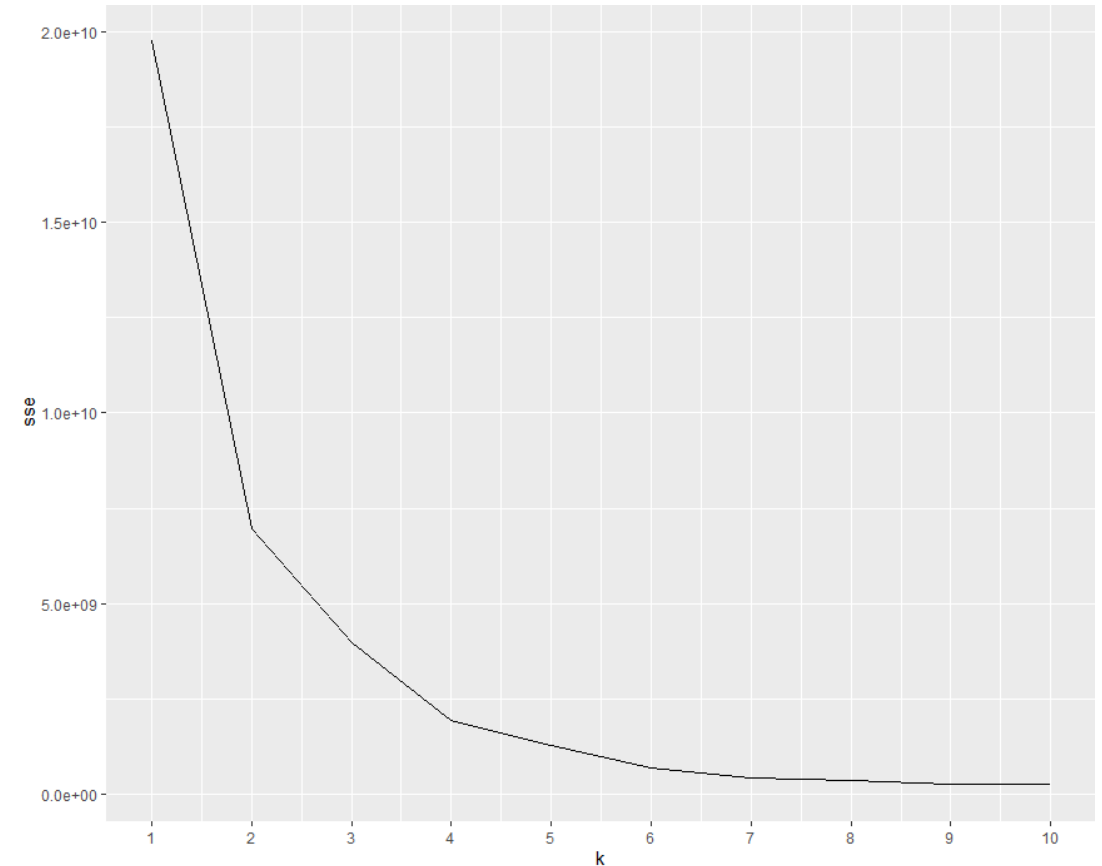
Pra Proses

- Import 98 file ke dalam 1 file list.
- Eliminasi kolom dan rename kolom untuk 98 dataset.
- Filter baris hanya untuk provinsi DKI Jakarta.
- Pengecekan jumlah baris, masing-masing data harus memiliki jumlah baris sebanyak 44 (sesuai jumlah kecamatan DKI Jakarta).
- Pengecekan missing value.
- Menggabungkan (join) 98 data menjadi 1 dataset.
- Eliminasi kolom kecamatan, untuk menjadi data siap mining.

Mining – k-means

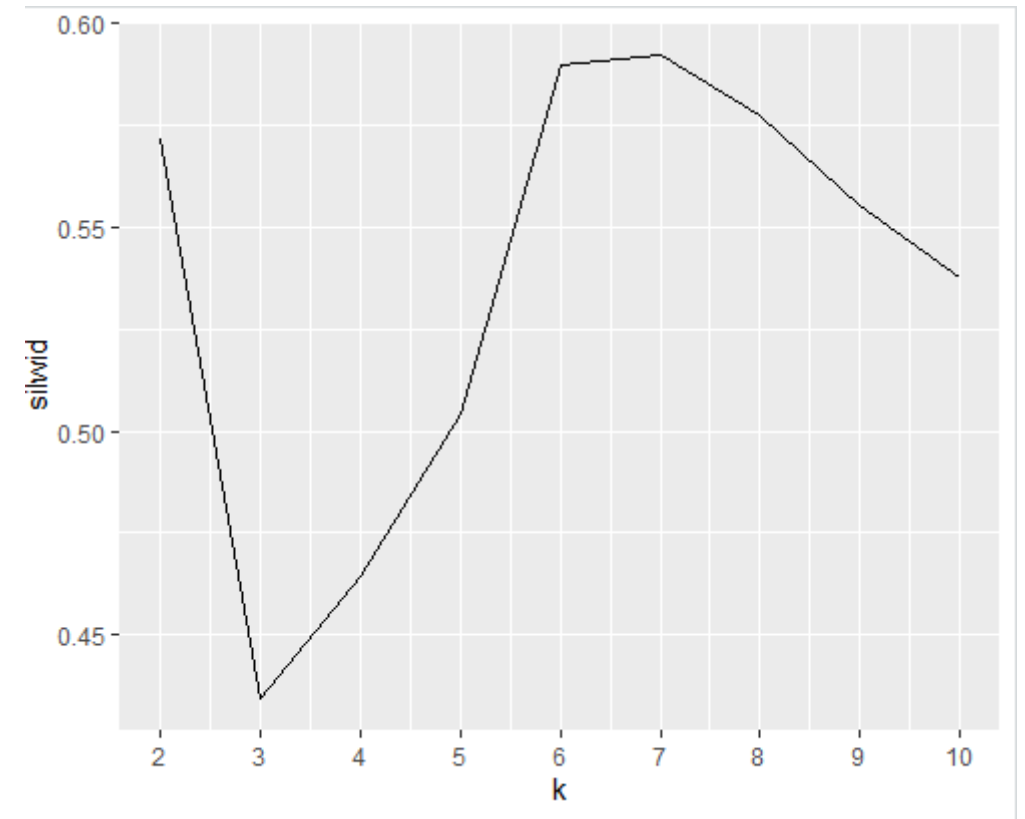
Visualisasi **SSE** berdasarkan nilai **k**

- Nilai **k** ditentukan berdasarkan "elbow".
- Semakin kecil nilai SSE-nya, semakin baik.
- Nilai **k** terbaik ada pada 2,3, atau 4.
- Dipilih **k** → **4** dengan SSE paling kecil.



Visualisasi **Avg Silhouette Width** berdasarkan k

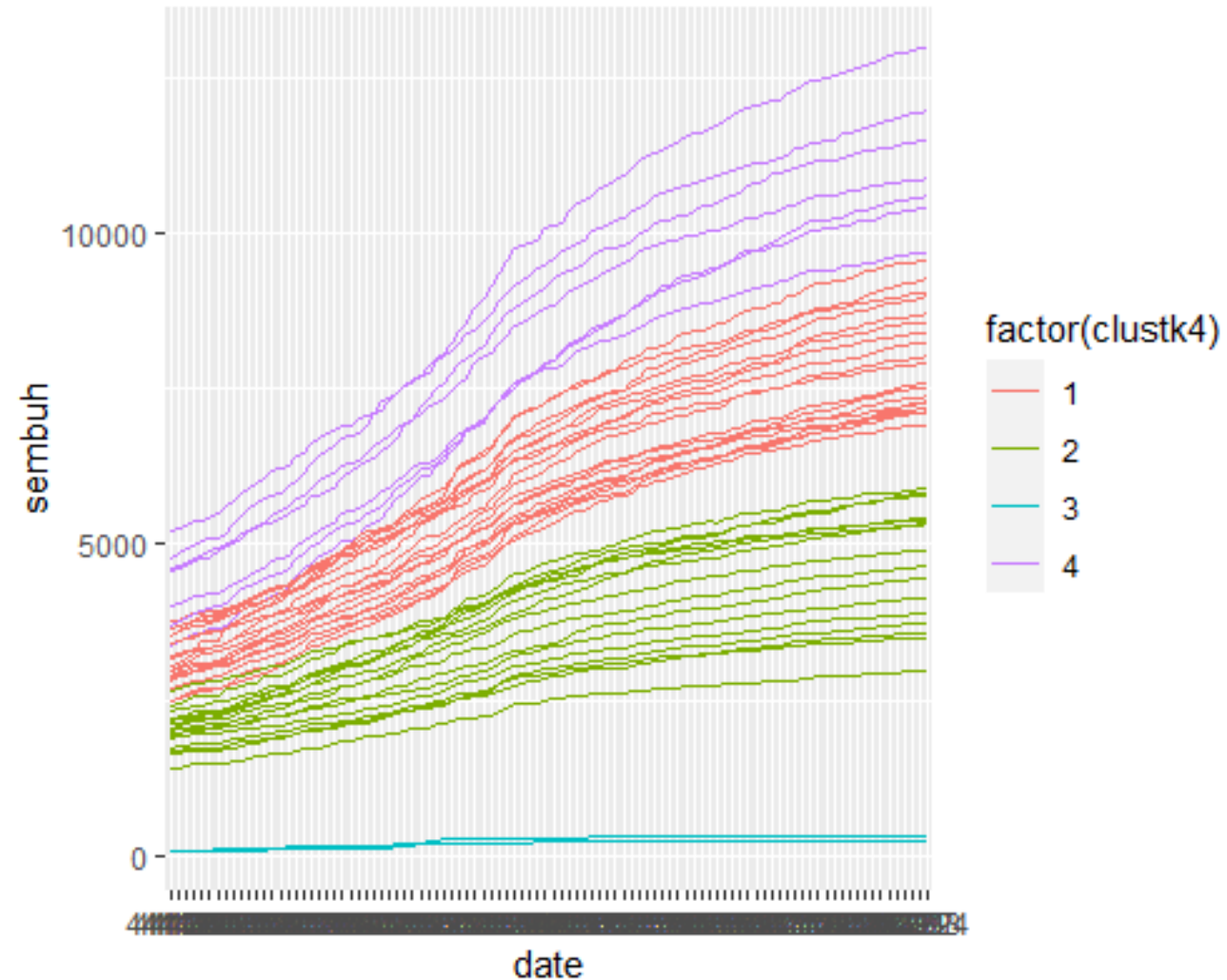
- Nilai k ditentukan berdasarkan nilai silhouette width tertinggi.
- Untuk $k = 7$ memiliki nilai yang lebih tinggi daripada $k = 6$, namun semakin tinggi k semakin tinggi resiko model menjadi *overfit* (*high variance*).
- Dipilih **$k \rightarrow 6$** .



K-means dengan $k = 4$

- Persebaran data pada masing-masing cluster

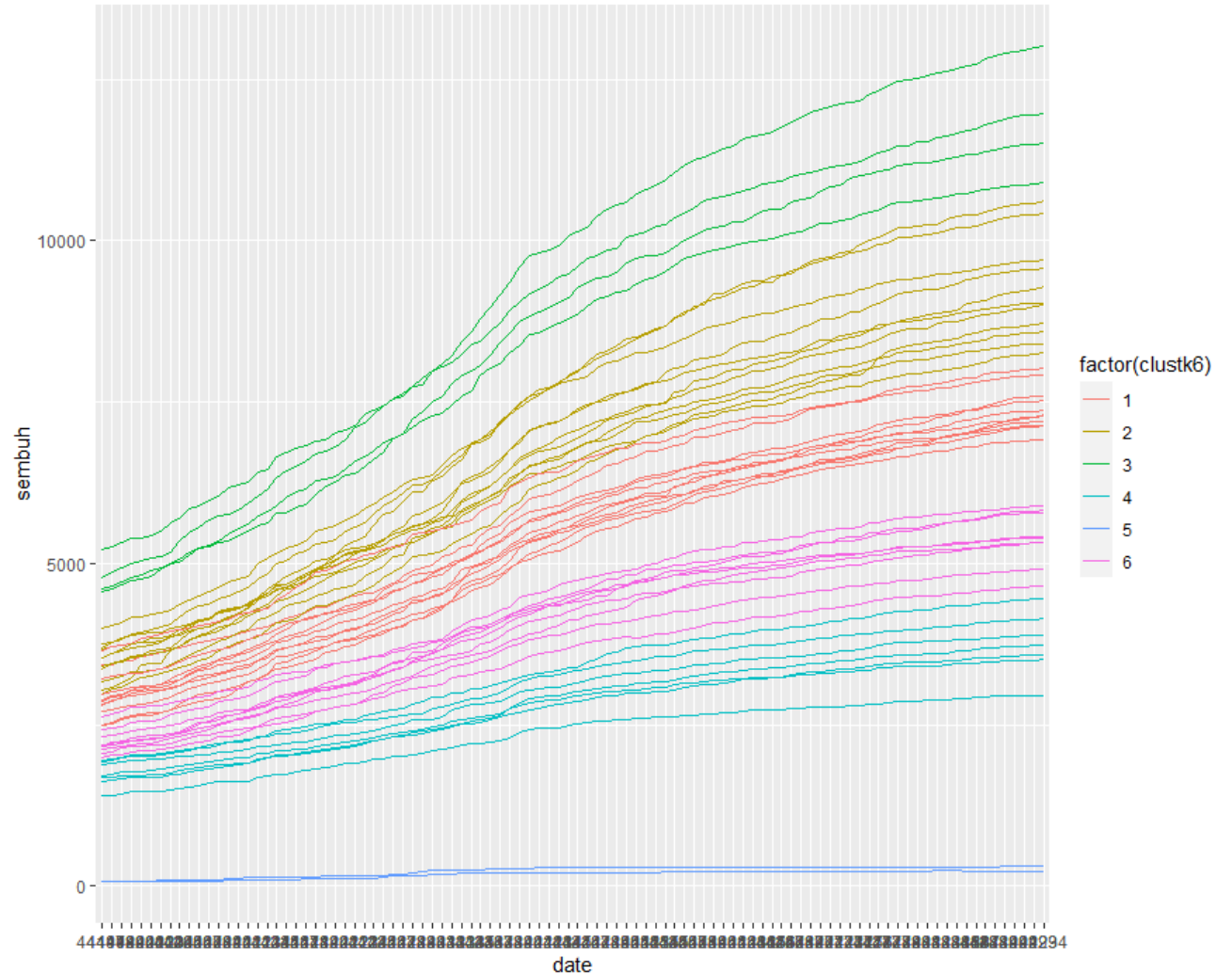
	Cluster	Jumlah Kecamatan
	1	19
	2	16
	3	2
	4	7



K-means dengan $k = 6$

- Persebaran data pada masing-masing cluster

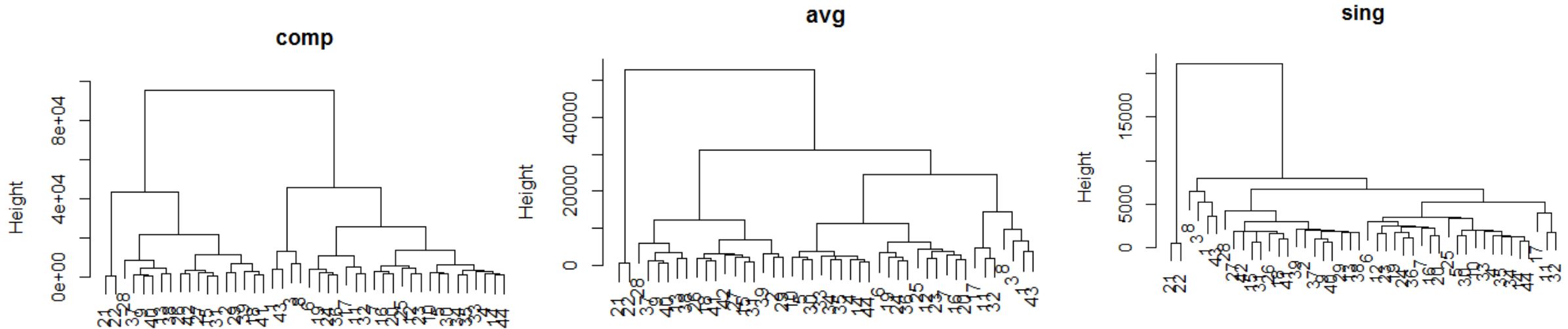
	Cluster	Jumlah Kecamatan
	1	11
	2	11
	3	4
	4	7
	5	2
	6	9



Mining – Hierarchical (Aglo)

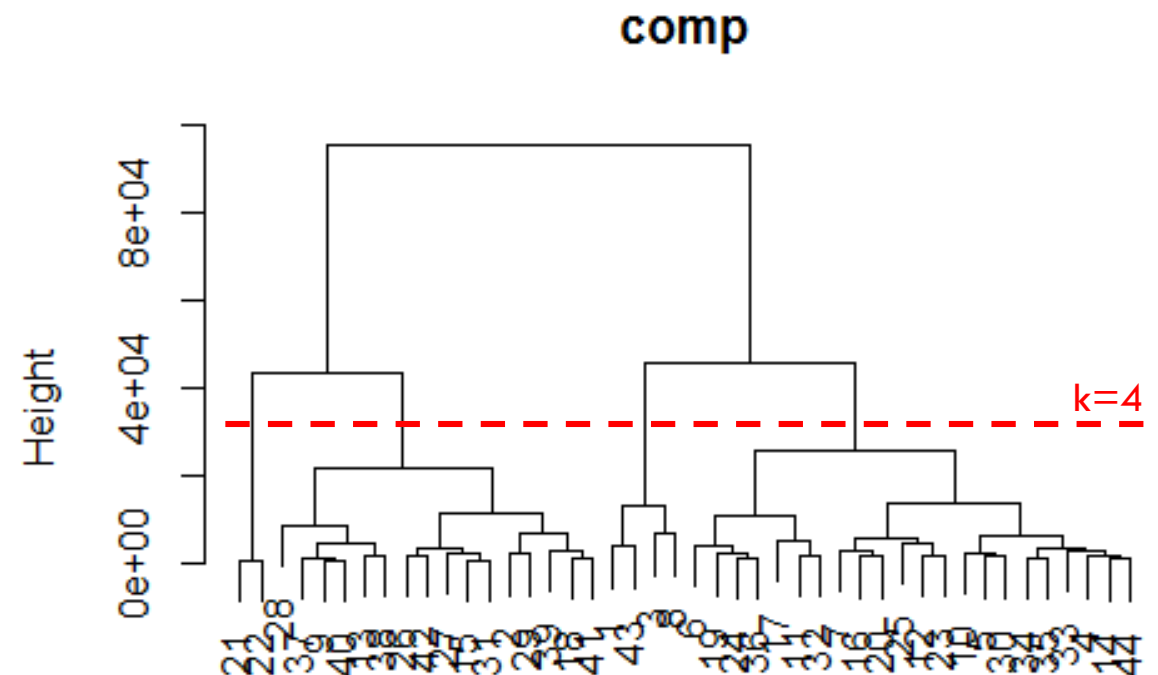
Pemilihan Linkage Method

- Dalam kasus ini, dipilih dendrogram yang paling seimbang / simetris agar persebaran data pada masing-masing kluster merata.
- Dipilih metode **complete**.



Penentuan nilai k

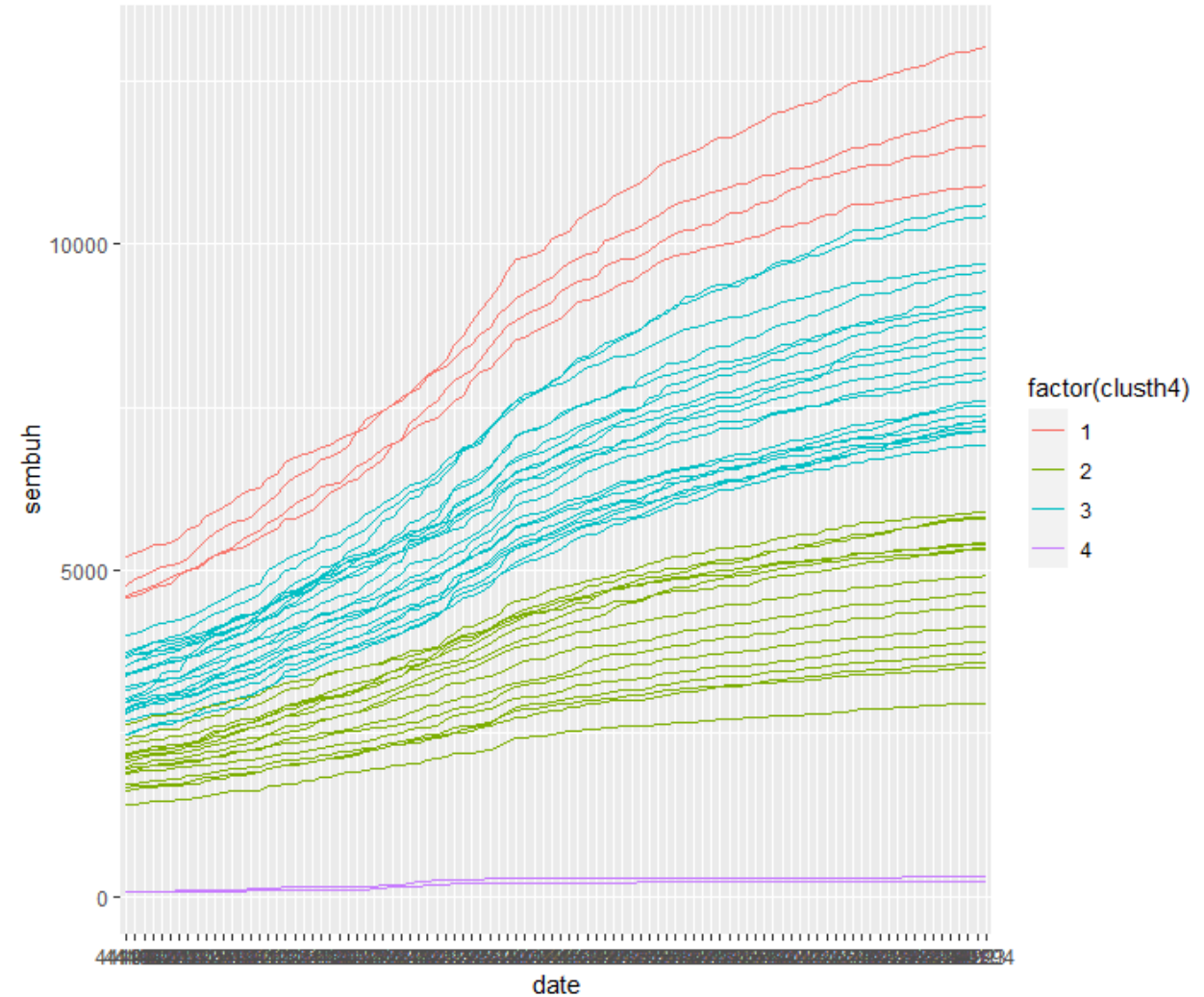
- Dipilih **k** → **4**, dimana memiliki persebaran data yang cukup merata antar cluster dan simetris.



Hierarchical dengan $k = 4$

- Persebaran data pada masing-masing cluster

	Cluster	Jumlah Kecamatan
	1	4
	2	16
	3	22
	4	2

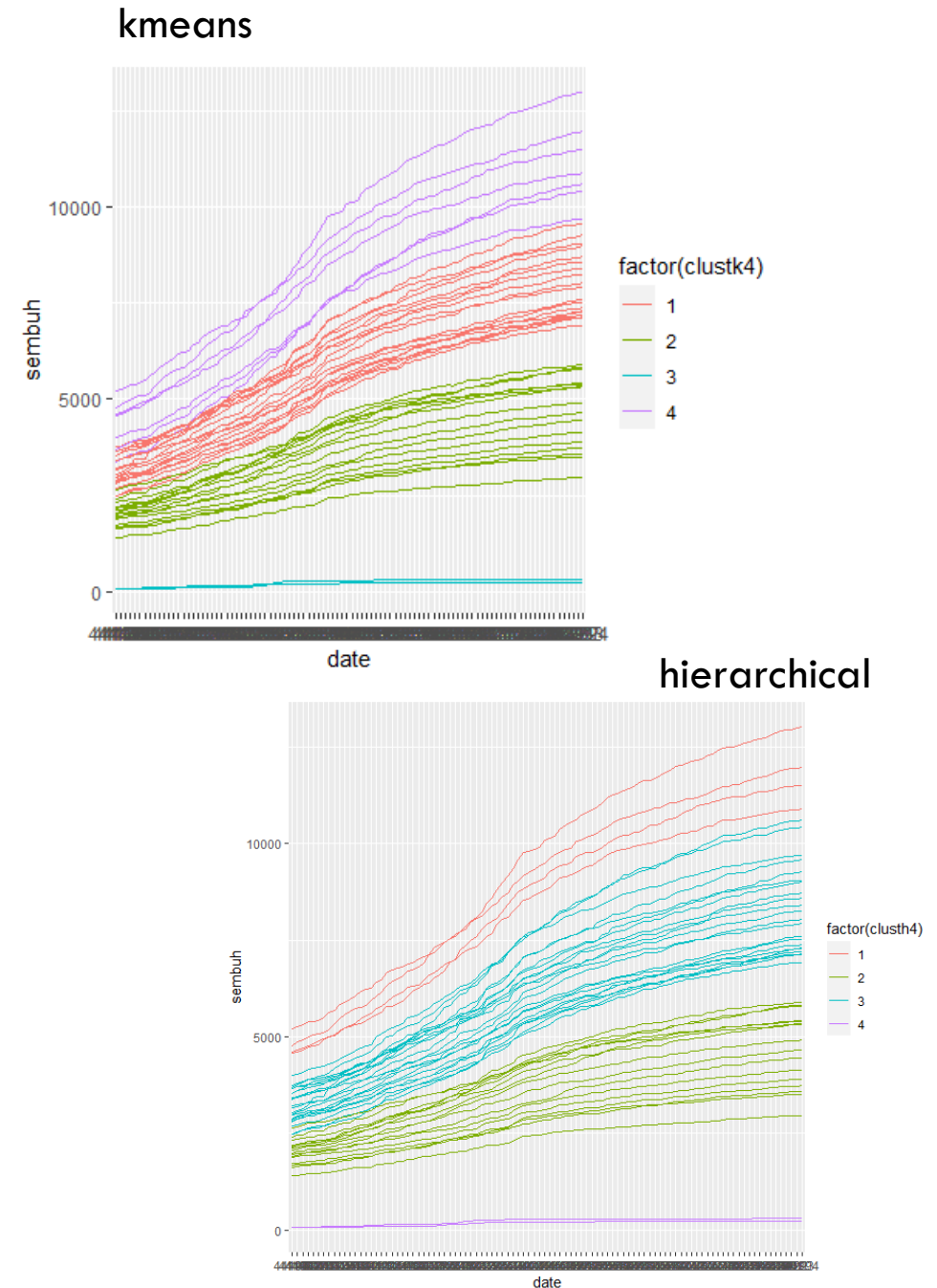


Pemilihan Metode Terbaik

Motode Terpilih

K-means dengan $k=4$

- Dipilih mining dengan $k = 4$,
 - karena baik pada kmeans maupun hierarchical menunjukkan bahwa nilai $k = 4$ adalah baik,
 - memiliki persebaran data paling merata dan grafik mudah diinterpretasikan,
 - dinilai cukup untuk merepresentasikan perbedaan antar kluster.
- Dipilih **kmeans** daripada hierarchical,
 - karena penyebaran data pada kmeans lebih merata,
 - berdasarkan grafik, kmeans menghasilkan grafik yang perbedaan antar cluster-nya lebih **didasarkan pada tingkat pertumbuhan (slope)**. Sementara pada hierarchical **didasarkan pada titik awal (intercept)**.

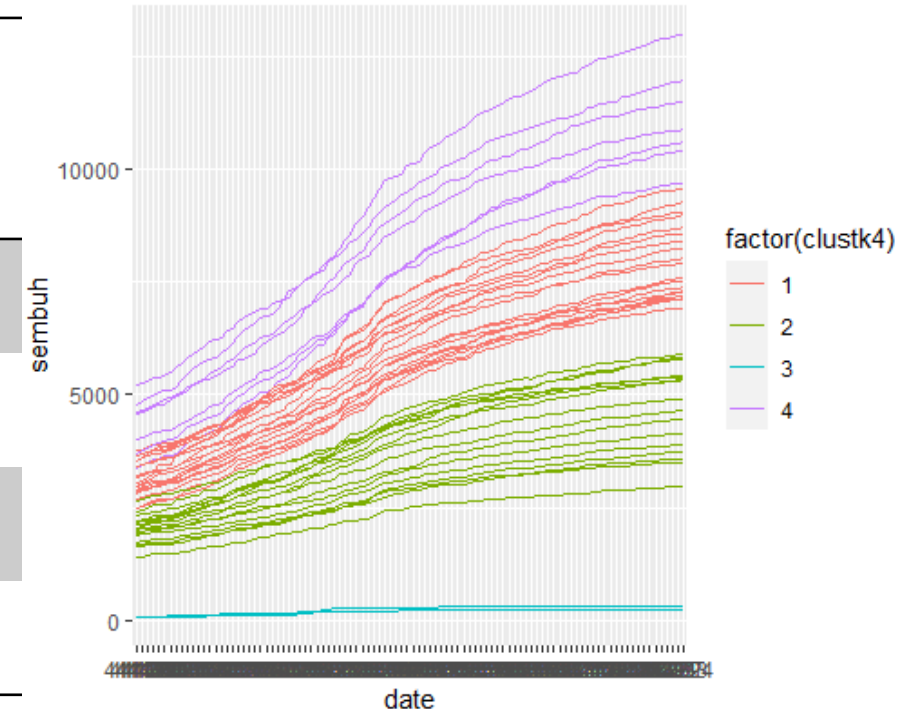


Analisis dan Pembahasan

Analisis

Cluster 4 merupakan kelompok kecamatan dengan angka pertumbuhan kesembuhan tertinggi

	Cluster	Jml Kecamatan	Angka Awal Kesembuhan (1 Jan) (Intercept)	Angka Pertumbuhan Kesembuhan per Hari (Slope)	Ratio Jumlah Kesembuhan dari Total Kasus Positif per Hari
	1	19	3154.7	55.3	0.99
	2	16	2036.8	30.2	0.99
	3	2	92.7	2.27	0.99
	4	7	4370.9	78.3	1.00



- Berdasarkan hasil clustering, dapat disimpulkan bahwa perbedaan mendasar pada masing-masing cluster adalah berdasarkan **angka pertumbuhan kesembuhan per hari**.
- Dimana baik tingkat pertumbuhan kesembuhan maupun ratio kesembuhan dari kasus positif **tidak menunjukkan perbedaan yang signifikan**.
- Cluster 3 merupakan kelompok kecamatan dengan **angka pertumbuhan terendah**, namun hal tersebut juga disebabkan karena kasus positif di cluster 3 juga rendah (berdasarkan ratio kesembuhan dr kasus positif).

Analisis

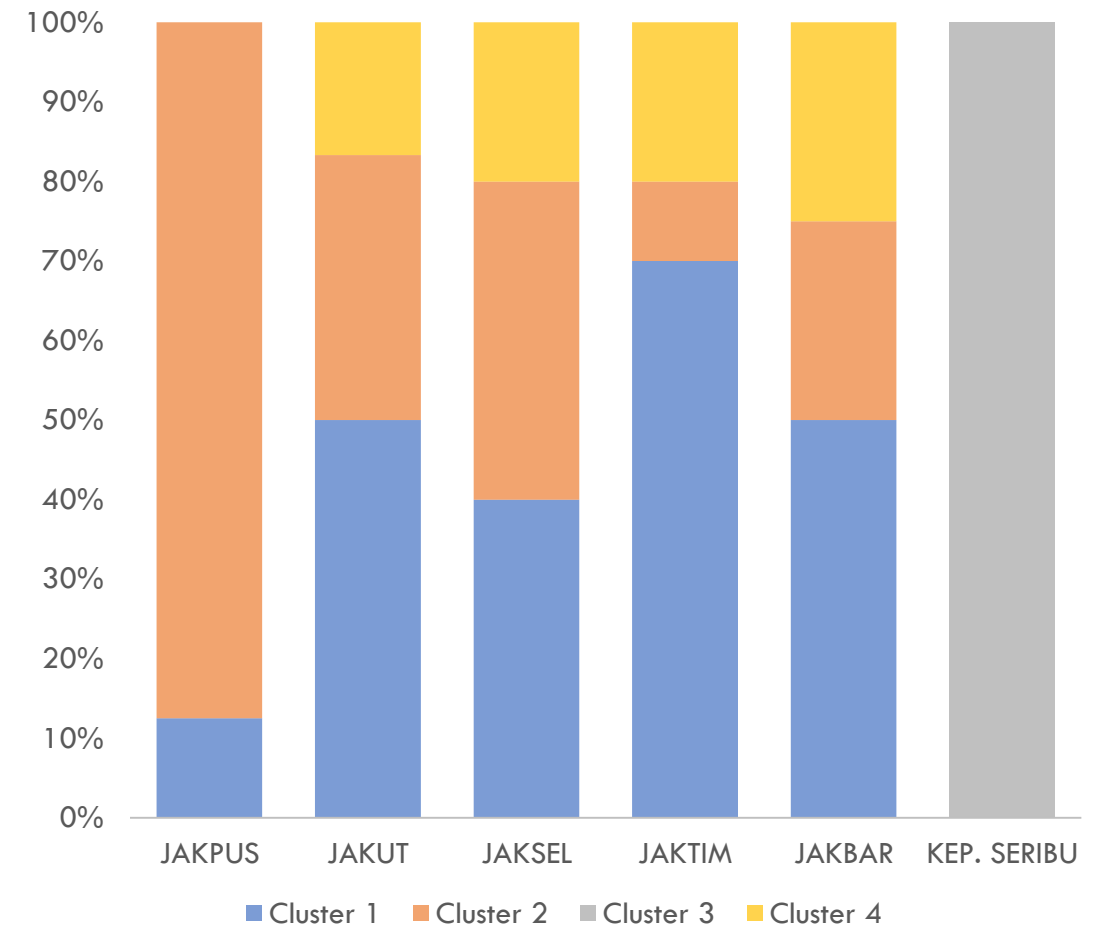
Cluster 2 dan 3 merupakan kelompok kecamatan yang masih bisa meraih pertumbuhan lebih tinggi sehingga perlu melakukan upaya tambahan

Cluster	Jml Kecamatan	Angka Pertumbuhan Kesembuhan per Hari	Tingkat Pertumbuhan	Perlakukan
1	19	55.3	Tinggi	Cluster 1 sudah memiliki angka pertumbuhan kesembuhan yang tinggi namun akan semakin baik jika dapat dioptimalkan hingga mencapai angka optimal (78.3) yang diraih cluster 4.
2	16	30.2	Sedang	Cluster 2 masih memiliki gap yang cukup besar dibanding cluster 1 dan 4. Sehingga dirasa perlu adanya upaya yang lebih serius lagi dari cluster 2.
3	2	2.27	Rendah	Meskipun angka kesembuhan cluster 3 rendah namun masih sebanding dengan jumlah kasus positif yang sedikit. Tetapi cluster 3 masih memiliki potensi untuk meraih angka yang lebih tinggi.
4	7	78.3	Tinggi	Kebijakan yang sudah dilakukan di cluster 4 dapat dijadikan contoh untuk cluster lain.

Analisis

Cluster 4 (cluster terbaik) tersebar di 4 kota, sementara cluster 3 (cluster terendah) seluruhnya ada di Kepulauan Seribu

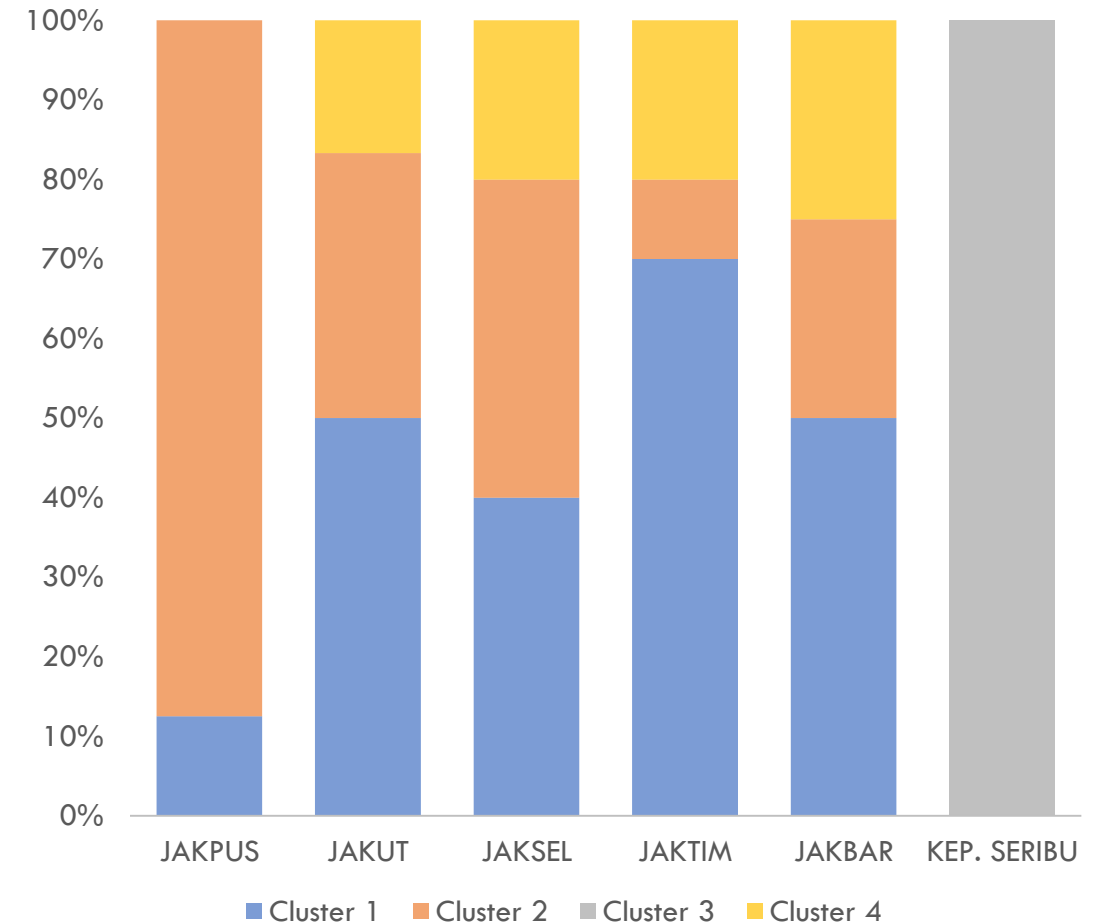
Kota	Jumlah Kecamatan	Cluster			
		1	2	3	4
Jakarta Pusat	8	1	7	0	0
Jakarta Barat	8	4	2	0	2
Jakarta Timur	10	7	1	0	2
Jakarta Selatan	10	4	4	0	2
Jakarta Utara	6	3	2	0	1
Kepulauan Seribu	2	0	0	2	0



Analisis

Berdasarkan angka pertumbuhan kesembuhan, kota Jakarta Timur merupakan kota dengan pertumbuhan terbaik

Kota	Angka Awal Kesembuhan (1 Jan) (Intercept)	Angka Pertumbuhan Kesembuhan per Hari (Slope)
Jakarta Pusat	2140,7	29,0
Jakarta Barat	3189,9	50,8
Jakarta Timur	3441,2	63,4
Jakarta Selatan	2775,3	52,1
Jakarta Utara	3085,9	48,0
Kepulauan Seribu	92,7	2,27



Kesimpulan & Rekomendasi

Kesimpulan

Simpulan

- Terdapat **pola tertentu** pada angka kesembuhan, sehingga **membentuk kelompok-kelompok kecamatan** dengan karakteristik tersendiri.

Saran

- Sangat dianjurkan untuk dilakukan **analisis lanjutan** mengenai faktor-faktor lain.
 - sangat memungkinkan adanya korelasi kuat dengan faktor lain seperti populasi penduduk, kepadatan penduduk, kondisi ekonomi, dan tingkat kesehatan setiap wilayah (kota / kecamatan).

Rekomendasi

- Pada wilayah-wilayah dengan tingkat kesembuhan rendah sebaiknya **diterapkan peraturan PSBB** (atau yang serupa) dengan **lebih ketat lagi**.
- Pada wilayah-wilayah dengan tingkat kesembuhan rendah sebaiknya **lebih digalakkan program vaksinasi**.
- Jika pada analisis lanjutan mengenai korelasi hasil mining ini dengan faktor lain benar (terdapat korelasi kuat), maka kebijakan dapat dibuat berdasarkan hasil analisis tersebut.
 - Contoh, ada korelasi kuat antara kondisi ekonomi dengan tingkat kesembuhan. Maka usulan rekomendasinya seperti pemberian subsidi sembako / bantuan dana yang lebih digalakkan pada daerah tertentu, harapannya hal ini dapat mengurangi mobilisasi warga.

Terima Kasih