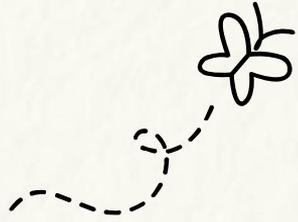
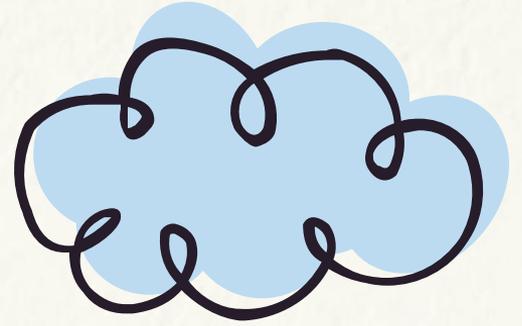
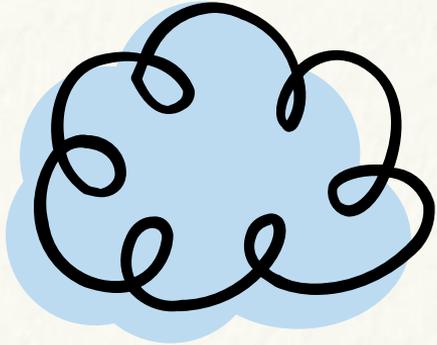


ANALISIS CLUSTER MENGGUNAKAN
METODE K-MEANS PADA
KABUPATEN/KOTA DI PROVINSI
JAWA TENGAH BERDASARKAN
INDIKATOR INDEKS PEMBANGUNAN
MANUSIA TAHUN 2021



01

Latar Belakang





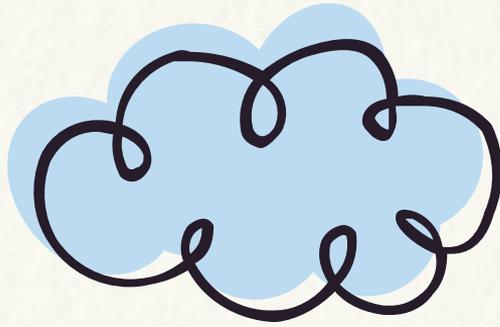
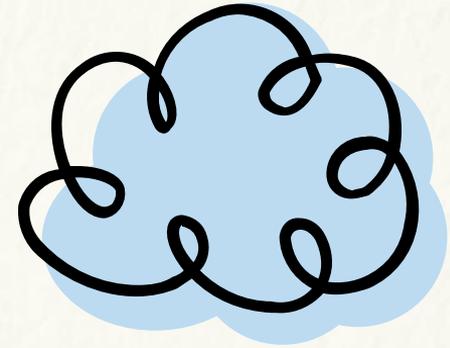
Latar Belakang



Indeks Pembangunan Manusia (IPM) atau *Human Development Index* (HDI) merupakan salah satu ukuran kualitas yang dapat digunakan untuk mengetahui sejauh mana kualitas pembangunan manusia yang telah berhasil dicapai. IPM menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan lain sebagainya.

Sebagai alat ukur untuk membangun kualitas hidup masyarakat yang lebih baik diperlukan nilai IPM yang meningkat setiap tahunnya secara progresif. Untuk mencapai hal tersebut, perlu dilakukan upaya-upaya untuk menstabilkan dan meningkatkan nilai IPM yang berbasis pada data dengan mengidentifikasi karakteristik dari masing-masing kabupaten/kota di Jawa Tengah berdasarkan indikator yang mempengaruhi IPM. Lalu, kita akan mendapatkan kluster atau kelompok berdasarkan wilayah yang memiliki karakteristik indikator peningkatan IPM yang sama.





02

Teori Pendukung

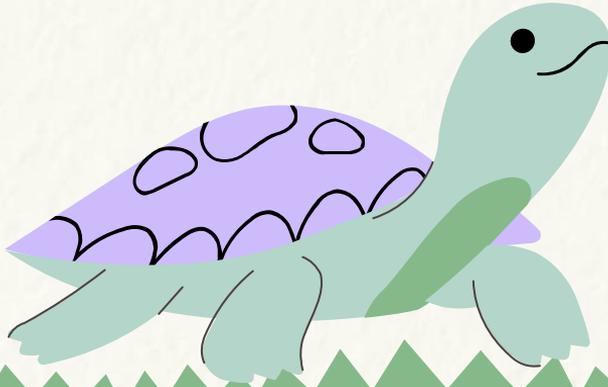
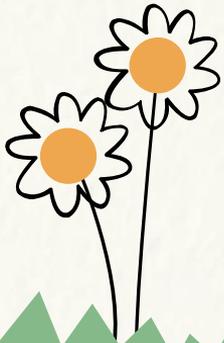
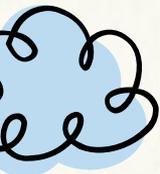


Table of contents



2.1

IPM

2.2

UHH

2.3

HLS

2.4

KPT

2.5

TPAK

2.6

Clustering

2.7

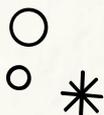
Metode Elbow

2.8

Metode Silhouette

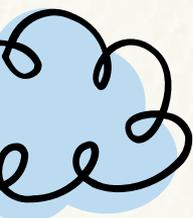
2.9

K-Means

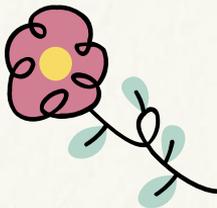




2.1 Indeks Pembangunan Manusia



Indeks pembangunan manusia (IPM) atau *Human Development Index* (HDI) adalah pengukuran perbandingan dari harapan hidup, melek huruf pendidikan, dan standar hidup untuk semua negara seluruh dunia. IPM menjadi indikator utama dalam mengukur keberhasilan pembangunan.

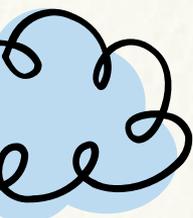


2.2 Usia Harapan Hidup

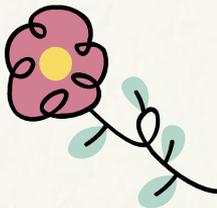
Usia Harapan Hidup saat Lahir (UHH) dapat diartikan sebagai rata-rata dari perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir. UHH dipengaruhi oleh jumlah kematian dan kelahiran bayi. Apabila jumlah kematian bayi jumlahnya cukup besar, maka akan mempengaruhi nilai UHH secara negatif atau menurun.



2.3 Harapan Lama Sekolah



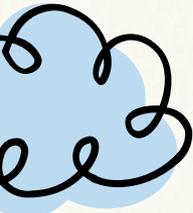
Angka Harapan Lama Sekolah (HLS) diartikan sebagai lamanya masa sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Angka HLS menunjukkan peluang anak usia 7 tahun ke atas untuk mengenyam pendidikan formal pada waktu tertentu.



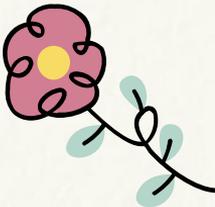


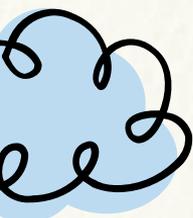
2.4

Pengeluaran Perkapita



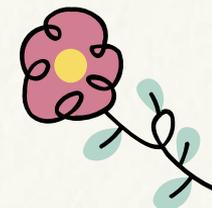
Pengeluaran per kapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga yang telah disesuaikan dengan paritas daya beli. Pengeluaran rumah tangga dibedakan menurut kelompok makanan dan bukan makanan. Semakin tinggi pendapatan, semakin tinggi pengeluaran bukan makanan.





2.5 Tingkat Pengangguran Angkatan Kerja

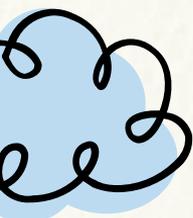
Tingkat Pengangguran Terbuka adalah persentase jumlah pengangguran terhadap jumlah angkatan kerja. Semakin tinggi nilai indikator TPAK, semakin banyak persediaan tenaga kerja yang tidak termanfaatkan. Indikator ini dapat memberikan sinyal tentang kinerja pasar kerja dan berlangsungnya kondisi ekonomi tertentu. TPAK berfungsi untuk mengindikasikan besarnya persentase angkatan kerja yang termasuk dalam pengangguran



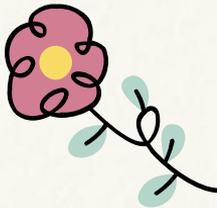


2.6

Clustering



Clustering merupakan proses mengelompokkan kumpulan object data ke dalam satu atau lebih kelompok sehingga data yang terhimpun pada suatu kelompok memiliki tingkat kesamaan (*similarity*) yang tinggi. Tujuan *clustering* adalah untuk mengidentifikasi suatu kelompok data dari populasi data untuk menghasilkan sifat-sifat dari data itu sendiri.



2.7 Metode Elbow

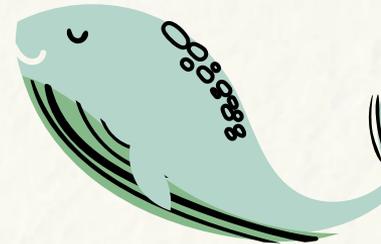
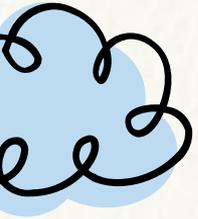
Metode Elbow merupakan metode yang digunakan untuk menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai cluster dan kemudian menambah nilai cluster tersebut untuk dijadikan model data dalam penentuan cluster terbaik.

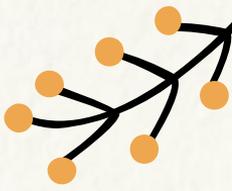
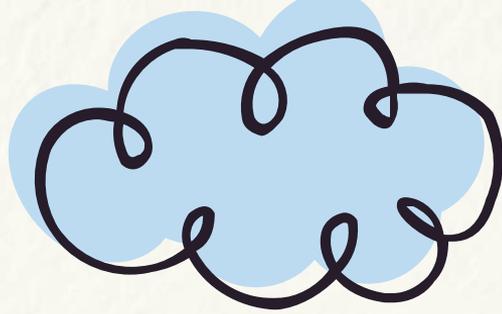
2.8 Metode Silhouette

Metode Silhouette Coefficient dapat digunakan untuk melihat kualitas serta kekuatan kluster dan seberapa baik objek berada pada cluster. Metode ini gabungan dari dua metode yaitu *cohesion* dan *separation*. Metode *cohesion* dapat mengetahui relasi antar objek di dalam *cluster*. Sedangkan, metode *separation* dapat mengetahui seberapa jauh atau terpisah satu *cluster* dengan *cluster* yang lain.

2.9 K-Means

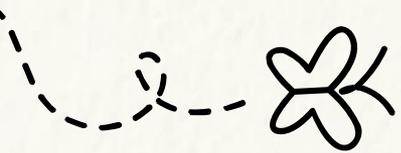
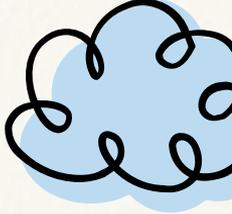
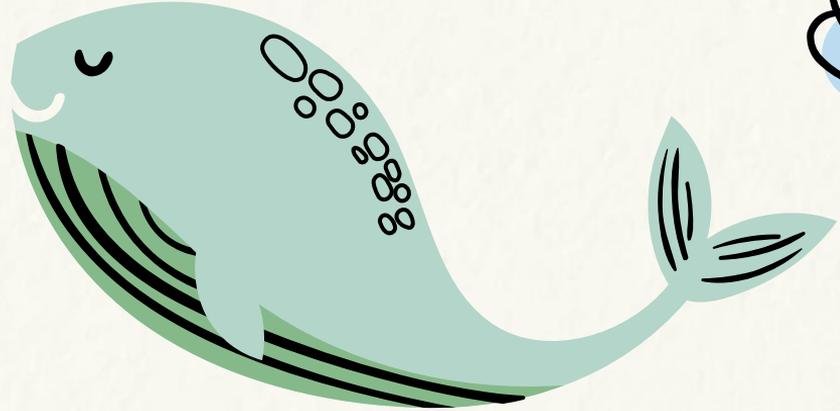
K-Means adalah suatu metode pengelompokan data non-hirarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok lain.





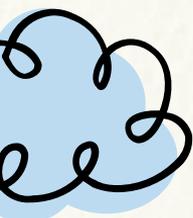
03

Metode Penelitian





Sumber Data : BPS Jawa Tengah



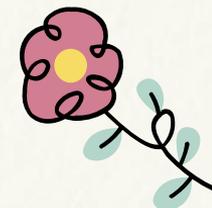
Variabel yang digunakan yaitu :

UHH : Usia Harapan Hidup

HLS : Harapan Lama Sekolah

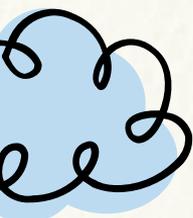
KPT : Pengeluaran per Kapita

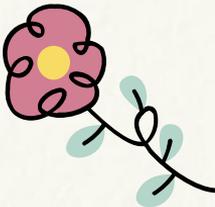
TPAK : Tingkat Pengangguran Angkatan Kerja



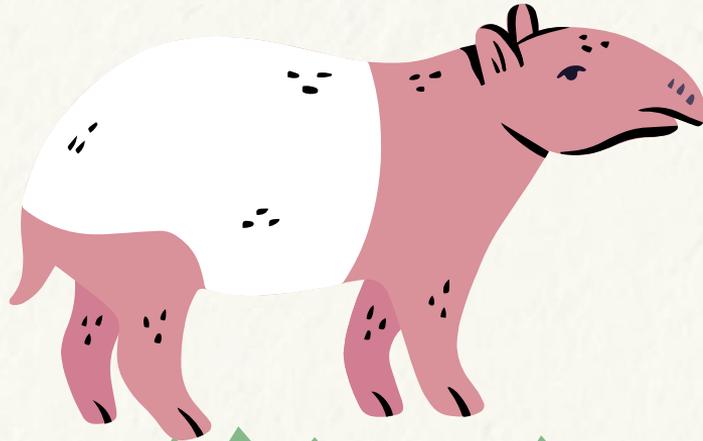
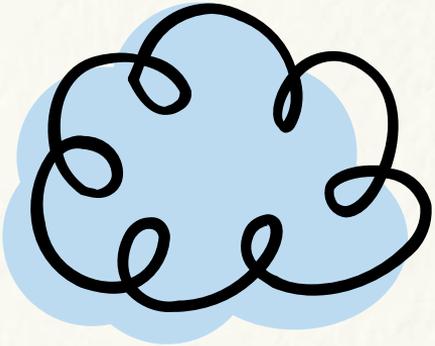


Langkah Penelitian

- 
1. Input variabel yang akan dianalisis, diantaranya: UHH, HLS, KPT, dan TPAK.
 2. Melakukan analisis deskriptif untuk mengetahui gambaran umum setiap variabel.
 3. Melakukan uji asumsi non multikolinearitas, dan uji Kaiser Meyer Olkin untuk mengetahui apakah sampel dapat mewakili populasi dari suatu data.
 4. Melakukan standarisasi data
 5. Menentukan jumlah cluster menggunakan metode Elbow dan Silhoutte.
 6. Melakukan interpretasi dari jumlah cluster yang terbentuk.



04 Hasil dan Pembahasan



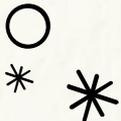
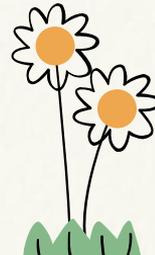


Analisis Deskriptif

Indikator	N	Min	Max	Mean	Std. Dev
Usia Harapan Hidup (UHH)	35	69.54	77.73	75.021	1.829
Harapan Lama Sekolah (HLS)	35	11.63	15.53	12.973	0.928
Pengeluaran per Kapita (KPT)	35	8573	15843	11139	1772.996
Tingkat Pengangguran Angkatan Kerja (TPAK)	35	62.91	75.79	70.076	3.422

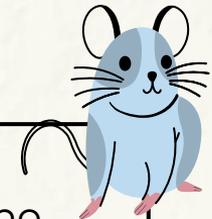


Dari tabel dapat diketahui bahwa data perlu dilakukan standarisasi terlebih dahulu sebelum dilakukan clustering.





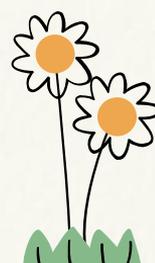
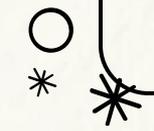
Uji Asumsi Sampel Mewakili Populasi



Asumsi pertama yang harus dipenuhi yaitu sampel mewakili populasi, yang dapat dilakukan menggunakan uji Kaiser Meyer Olkin (KMO). Sampel dikatakan telah mewakili populasi apabila nilai keseluruhan KMO $> 0,5$.

Tabel 2 Uji Keyser Meyer Olkin
Kaiser-Meyer-Olkin factor adequacy
Overall MSA= 0,55

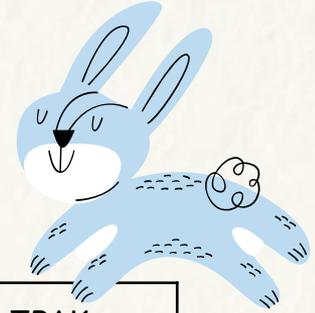
diperoleh nilai KMO sebesar 0,55. Hal tersebut berarti sampel telah mewakili populasi atau memenuhi uji asumsi.



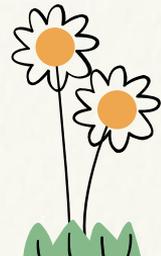


Standardisasi

Satuan yang berbeda -> STANDARDISASI

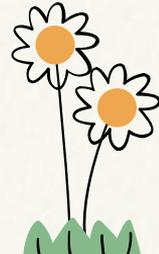
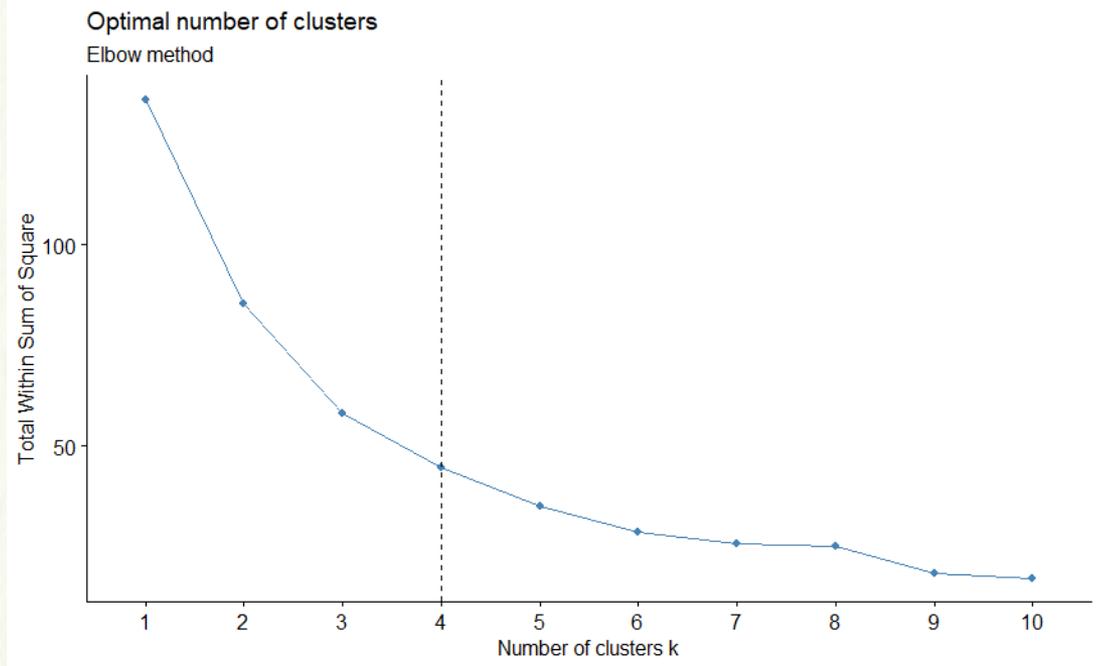


Indikator	UHH	HLS	KPT	TPAK
Kab. Cilacap	-0.60765358	-0.36973353	-0.3413431	-2.0939644
Kab. Banyumas	-0.66229869	0.06090453	0.2294421	-1.4628700
Kab. Purbalingga	-0.98470482	-1.04798848	-0.6244797	-0.1626988
Kab. Banjarnegara	-0.40000218	-1.44632868	-0.9769903	-0.0925772
Kab. Kebumen	-0.79891145	0.40541497	-1.1907528	-0.6769238



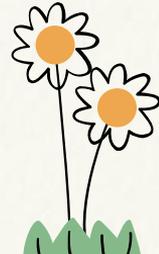
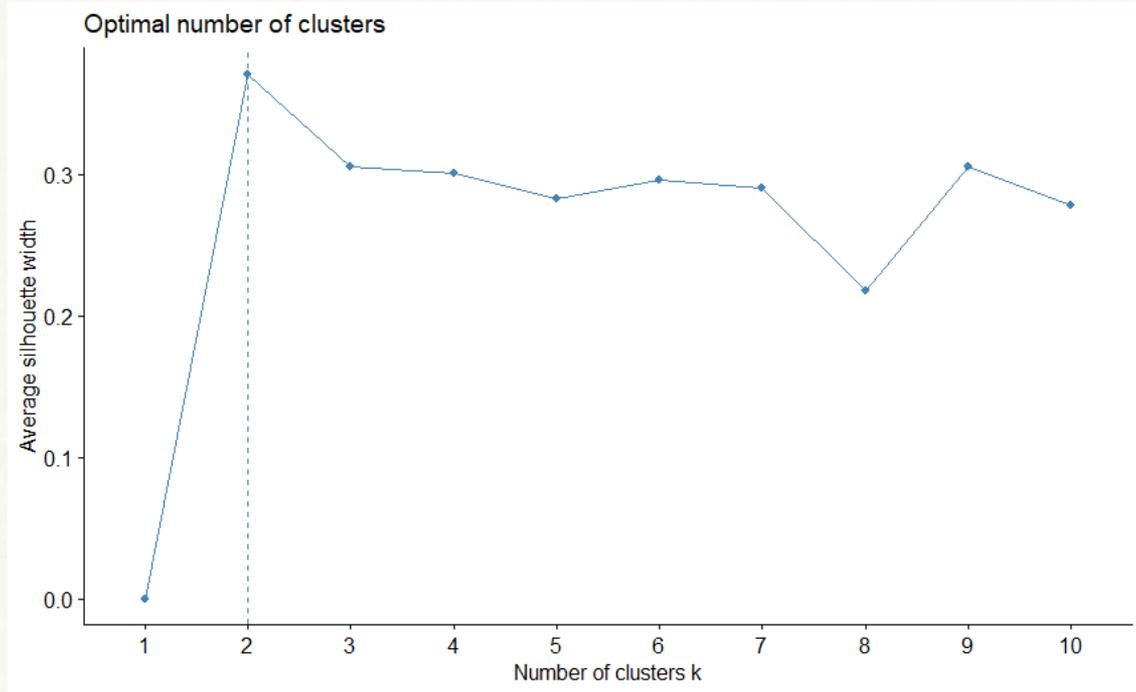
Penentuan Jumlah Cluster Optimal

Metode Elbow



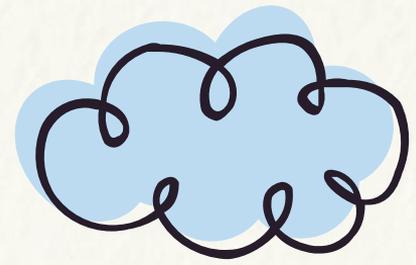
Penentuan Jumlah Cluster Optimal

Metode Silhouette

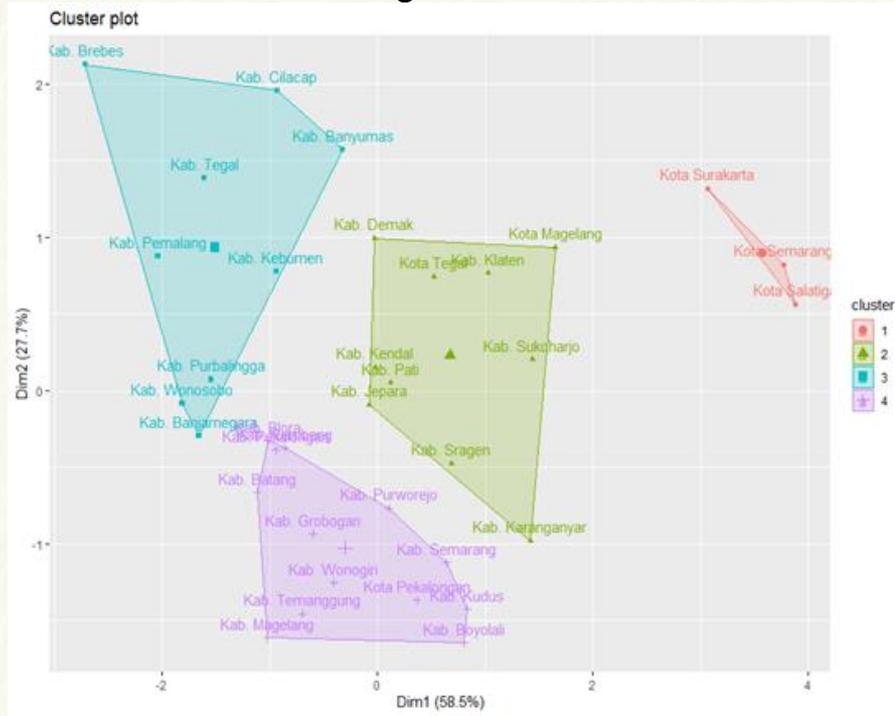




K-Means Clustering

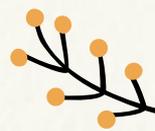


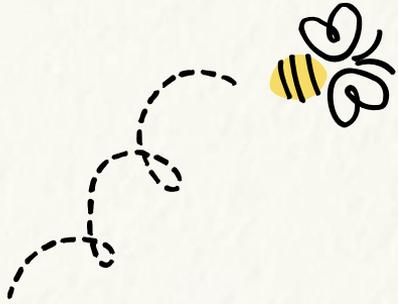
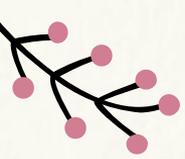
K = 4



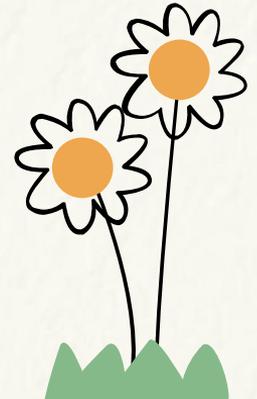
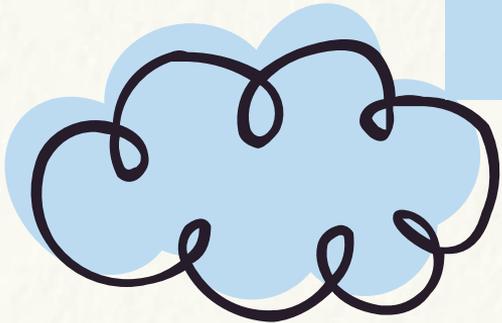
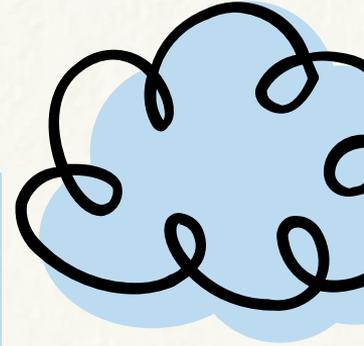
- Cluster 1 = 3 Kab/Kota
- Cluster 2 = 10 Kab/Kota
- Cluster 3 = 9 Kab/Kota
- Cluster 4 = 13 Kab/Kota

($between_SS / total_SS = 67.3\%$)

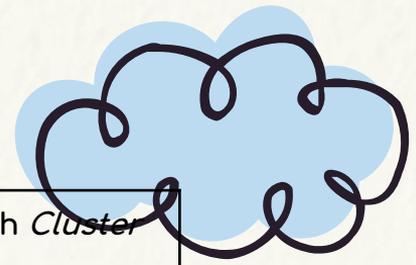




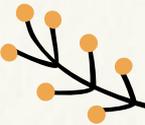
Berdasarkan pengklasteran untuk kedua nilai k , saat $k = 4$ memiliki nilai **within cluster sum of square by cluster** yang lebih besar yaitu dengan tingkat kebaikan sebesar 67.3 %. Sehingga dapat disimpulkan banyaknya cluster yang optimal pada saat ini adalah **4 cluster.**



Anggota Cluster

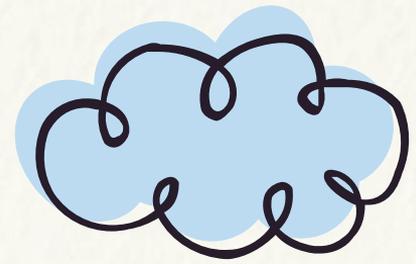


Cluster	Kabupaten/ Kota	Jumlah Cluster
1	Kota Surakarta, Kota Salatiga, Kota Semarang	3
2	Kab. Klaten, Kab. Sukoharjo, Kab. Karanganyar, Kab. Sragen, Kab. Pati, Kab. Jepara, Kab. Demak, Kab. Kendal, Kota Magelang, Kota Tegal	10
3	Kab. Cilacap, Kab. Banyumas, Kab. Purbalingga, Kab. Banjarnegara, Kab. Kebumen, Kab. Wonosobo, Kab. Pemalang, Kab. Tegal, Kab. Brebes	9
4	Kab. Purworejo, Kab. Magelang, Kab. Boyolali, Kab. Wonogiri, Kab. Grobogan, Kab. Blora, Kab. Rembang, Kab. Kudus, Kab. Semarang, Kab. Temanggung, Kab. Batang, Kab. Pekalongan, Kota Pekalongan	13



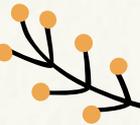


Rerata Masing-masing Cluster



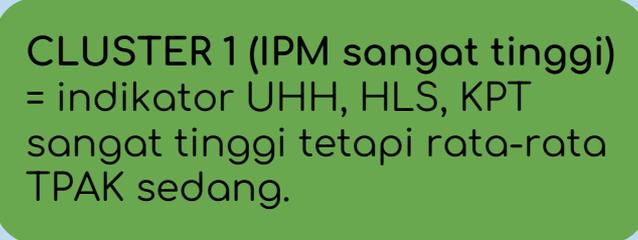
Cluster	UHH	HLS	KPT	TPAK
1	77.46000	15.27667	15393.000	68.88667
2	76.15200	13.29600	11602.300	69.05800
3	72.83000	12.36667	9970.222	66.87778
4	75.08077	12.61385	10610.615	73.35000

Kuning : Sangat Tinggi
Hijau : Tinggi
Ungu : Sedang
Biru : Rendah

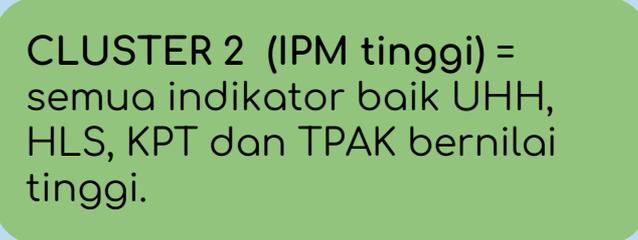




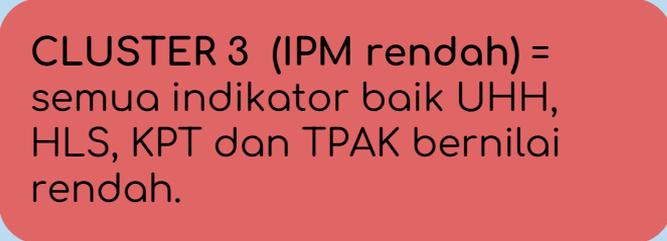
Kesimpulan



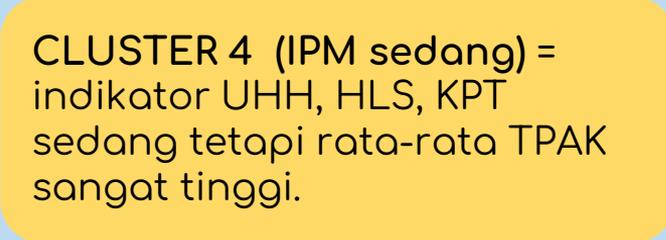
CLUSTER 1 (IPM sangat tinggi) = indikator UHH, HLS, KPT sangat tinggi tetapi rata-rata TPAK sedang.



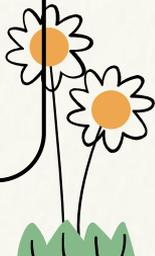
CLUSTER 2 (IPM tinggi) = semua indikator baik UHH, HLS, KPT dan TPAK bernilai tinggi.



CLUSTER 3 (IPM rendah) = semua indikator baik UHH, HLS, KPT dan TPAK bernilai rendah.



CLUSTER 4 (IPM sedang) = indikator UHH, HLS, KPT sedang tetapi rata-rata TPAK sangat tinggi.





Daftar Pustaka



Badan Pusat Statistik. (2015). *Tingkat Pengangguran Angkatan Kerja (TPAK)*. Diambil kembali dari Sistem Informasi Rujukan Statistik: <https://sirusa.bps.go.id/sirusa/index.php/indikator/44>

Badan Pusat Statistik. (2017). *Angka Harapan Hidup saat Lahir*. Diambil kembali dari Sistem Informasi Rujukan Statistik: <https://sirusa.bps.go.id/sirusa/index.php/indikator/1060>

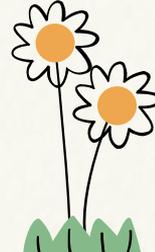
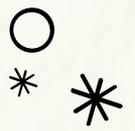
Badan Pusat Statistik. (2017). *Harapan Lama Sekolah*. Diambil kembali dari Sistem Informasi Rujukan Statistik: <https://sirusa.bps.go.id/sirusa/index.php/indikator/1016>

Badan Pusat Statistik. (2017). *Pengeluaran per Kapita*. Diambil kembali dari Sistem Informasi Rujukan Statistik: <https://sirusa.bps.go.id/sirusa/index.php/indikator/197>

Badan Pusat Statistik. (2020). *Indeks Pembangunan Manusia*. Jakarta: Badan Pusat Statistik.

Badan Pusat Statistik. (2020, February 17). Retrieved from BPS provinsi Jawa Tengah: <https://jateng.bps.go.id/pressrelease/2020/02/17/1222/>

Bholowalia, Purnima & Kumar, Arvind, (2014). EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. *International Journal of Computer Application*, 9(105), 17-24





Daftar Pustaka



Du. (2010). *Clustering: A neural network approach*. Neural Networks, 89-107.

Hardjanto, Imam. *Teori Pembangunan*. Malang: UB Press. 2013

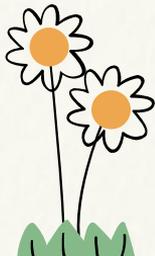
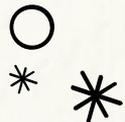
Larose, D., & Larose, C. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons.

Madhulatha, T.S., 2012. An Overview On Clustering Methods. *IOSR Journal of Engineering*, 2(4), 719-725

Metisen M.B., Sari L.H., (2015). Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2), 110-118.

Wira, B., Budianto, A. E., & Wiguna, A. S., (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. Rainstek: *Jurnal Terapan Sains & Teknologi*, 1(3), 53-68.

Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678.



Thanks!

