UJIAN AKHIR SEMESTER:
CLUSTERING RENTANG SUHU CUACA JAKARTA TAHUN 2021-2023 MENGGUNAKAN ALGORITMA K-MEAN DAN GAUSSIAN MIXTURE MODELS

UNIVERSITAS
BUDI LUHUR

Disusun oleh :

Muchamad Angga Dwi Wahyu 2112501339

UNIVERSITAS BUDI LUHUR
FAKULTAS TEKNOLOGI INFORMASI
JAKARTA
2022/2023

DAFTAR ISI

BAB I
PENDAHULUAN

1.1. Latar Belakang

Cuaca memainkan peran kunci dalam kehidupan sehari-hari dan dapat memengaruhi berbagai aspek kehidupan manusia, termasuk pertanian, transportasi, dan kesehatan. Pemahaman mendalam tentang pola cuaca sangat diperlukan untuk mengantisipasi dan mengelola dampaknya.

Jakarta, sebagai ibu kota Indonesia, mengalami variasi cuaca yang signifikan sepanjang tahun. Dari musim hujan hingga musim kemarau, pemahaman tentang perubahan cuaca dalam rentang waktu tertentu dapat memberikan wawasan tentang tren iklim dan perubahan yang mungkin terjadi.

Dalam menghadapi variasi cuaca yang mengalami perubahan yang signifikan, Penelitian ini memiliki dua tujuan utama: pertama, membentuk kelompok-kelompok suhu cuaca harian yang serupa menggunakan algoritma K-Means dan Gaussian Mixture Models (GMM); kedua, membandingkan kinerja kedua algoritma ini dengan menggunakan metrik evaluasi Silhouette Score. Hasil penelitian diharapkan dapat memberikan wawasan yang lebih baik tentang pola kondisi suhu cuaca di Jakarta serta menentukan algoritma yang paling optimal untuk tugas pengelompokan suhu cuaca harian.

1.2. Perumusan Masalah

Berdasarkan uraian latar belakang maka diuraikan perumusan masalah sebagai berikut :

1. Bagaimana membagi data suhu cuaca harian berdasarkan nilai suhu minimal dan maksimal ke dalam kelompok-kelompok yang serupa?
2. Manakah algoritma yang paling optimal untuk digunakan sebagai pembuat cluster untuk menyelesaikan permasalahan 1?

1.3. Tujuan

Berdasarkan uraian perumusan masalah maka diuraikan tujuan dari artikel ilmiah ini sebagai berikut :

1. Membuat kluster rentang kondisi suhu cuaca di jakarta berdasarkan nilai suhu cuaca minimal dan suhu cuaca maksimal jakarta tahun 2021-2023.
2. Mengetahui algoritma yang optimal antar K-Mean dan GMM untuk membuat cluster rentang suhu cuaca jakarta tahun 2021-2023 berdasarkan Silhouette Score.

BAB II
METODE PENELITIAN

2.1. Metode Yang Digunakan

Dalam penelitian ini akan digunakan dua metode machine learning yaitu metode K-Means dan Gaussian Mixture Models:

1. K-Means : Merupakan metode yang menggunakan pendekatan partisi untuk membentuk kelompok data. Cocok untuk data dengan kelompok yang jelas dan sederhana, memberikan solusi clustering dengan komputasi yang efisien.
2. Gaussian Mixture Models : Memiliki tujuan yang sama dengan K-Means tetapi menggunakan pendekatan probabilistik di mana setiap data dapat diberikan probabilitas untuk menjadi bagian dari setiap kelompok. Ideal untuk mengidentifikasi pola suhu cuaca yang mungkin bersifat kompleks dan tidak teratur.
3. Penilaian performa klastering, metode Silhouette akan dilibatkan untuk mengetahui algoritma yang optimal antar K-Mean dan GMM.

BAB III
INFORMASI DATASET

3.1. Sample Data Asli

name	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	precip	precipprob	precipcover	preciptype	snow	snowdepth	windgust	windspeed	winddir	sealevelpressure	cloudcover	visibility	solarradiation	solarenergy	uvindex	severerisk	sunrise	sunset	moonphase	conditions	description	icon	stations
Jakarta	07/04/2021	30.6	25	27.4	35.5	25	29.4	22.8	76.6	1.796	100	16.67	rain			53.6	30	270.8	1009.1	59.5	6.1	218.5	18.9	7		2021-04-07T04:07T1	0.83	artially ughout	rain	1909995	
Jakarta	08/04/2021	33	25	28.9	37.7	25	31.5	22.9	71.5	0	0	0				40.7	21.7	270	1007.9	49.2	6.1	276.6	24	9		2021-04-08T04:08T1	0.86	ially cic through	cloud	1909995	
Jakarta	09/04/2021	32	26	28.9	37.5	26	32.8	23.9	75.2	0	0	0					21.8	321.7	1008.6	49.2	5.8	216	18.6	7		2021-04-09T04:09T1	0.9	ially cic through	cloud	1909995	
Jakarta	10/04/2021	32	23.7	27.8	37.2	23.7	30.8	23.3	77.3	5.934	100	16.67	rain				27.7	322.3	1010.3	59.4	6.1	196.7	17.1	7		2021-04-10T04:10T1	0.93	artially with a c	rain	1909995	
Jakarta	11/04/2021	32.7	24.6	27.6	38.9	24.6	29.5	22.9	76.5	0.965	100	4.17	rain				26.1	289.1	1011	52.5	5.6	211.3	18.2	7		2021-04-11T04:11T1	0.97	artially ut the d	rain	1909995	
Jakarta	12/04/2021	33	24.3	28	38.4	24.3	29.9	22.9	75.2	26.529	100	8.33	rain				20.1	259.2	1010.9	55.2	5.8	135.4	11.7	5		2021-04-12T04:12T1	0	artially the day	rain	1909995	
Jakarta	13/04/2021	33	25	27.6	38.1	25	29.4	23.6	80.5	0.121	100	20.83	rain			35.3	16.6	334.7	1011	55.2	5.5	172.7	14.9	7		2021-04-13T04:13T1	0.04	artially with a c	rain	1909995	
Jakarta	14/04/2021	32	24.3	27	37.9	24.3	28.8	23.6	82.8	3.082	100	8.33	rain				22.3	259.6	1010.4	56.5	5.2	113.8	9.7	4		2021-04-14T04:14T1	0.07	artially ughout	rain	1909995	
Jakarta	15/04/2021	32	24	26.8	36.8	24	28.4	23.4	82.7	2.567	100	20.83	rain				19.5	275.5	1010.4	53.4	5.4	134.2	11.6	5		2021-04-15T04:15T1	0.1	artially ughout	rain	1909995	
Jakarta	16/04/2021	32	24.3	28.2	36.8	24.3	31.1	23.5	76.8	0	0	0					18	5.5	1010.4	53.8	5.8	222.4	19.3	8		2021-04-16T04:16T1	0.14	ially cic through	cloud	1674909	
Jakarta	17/04/2021	32	25.3	28.1	38.8	25.3	31.2	24.3	80.4	26.362	100	16.67	rain				14.8	338.6	1010.9	55.9	5	187.2	16.3	7		2021-04-17T04:17T1	0.17	artially ughout	rain	99,9674	
Jakarta	18/04/2021	32	25.3	28.3	37.3	25.3	31.3	24	78.3	0.055	100	8.33	rain				17.9	338.4	1011.5	58.3	5.3	248.8	21.3	9		2021-04-18T04:18T1	0.2	artially the day	rain	99,9674	
Jakarta	19/04/2021	32.7	25	28.6	37.7	25	31.6	23.5	74.9	0.036	100	4.17	rain				15.5	16.5	1010.4	49.6	5.3	149.9	13	6		2021-04-19T04:19T1	0.24	artially the day	rain	1909995	
Jakarta	20/04/2021	32.7	24	28.6	37.2	24	31.3	22.9	72.2	0.024	100	4.17	rain			31.7	25.4	49.2	1010.2	49.9	5.7	209.6	18.1	8		2021-04-20T04:20T1	0.25	artially the day	rain	99,9674	
Jakarta	21/04/2021	33.7	24.7	29	39	24.7	31.8	22.7	69.8	0	0	0					18.3	38.3	1010.4	39.3	5.7	296.9	25.5	10		2021-04-21T04:21T1	0.3	ially cic through	cloud	1909995	
Jakarta	22/04/2021	33	24.5	29.1	36.5	24.5	31.5	22.2	68	0	0	0					20.1	24.8	1010.2	42.1	5.8	291	25.2	10		2021-04-22T04:22T1	0.34	ially cic through	cloud	1673909	
Jakarta	23/04/2021	32.7	24.3	28.3	37.4	24.3	30.8	23	74.5	10.857	100	8.33	rain				15.3	1.6	1010.1	47.6	4.9	275	23.9	10		2021-04-23T04:23T1	0.37	artially the day	rain	1909995	
Jakarta	24/04/2021	32.7	24.3	28.8	37.3	24.3	31.6	22.6	70.2	0	0	0					20.4	336	1009.9	44.9	5.8	278.6	24	10		2021-04-24T04:24T1	0.4	ially cic through	cloud	1909995	
Jakarta	25/04/2021	33	25.3	29.3	38.9	25.3	33	23.2	70.2	0	0	0				40.7	21.8	44.8	1010.3	42.2	6.1	242.7	21	9		2021-04-25T04:25T1	0.44	ially cic through	cloud	1909995	
Jakarta	26/04/2021	34	26	29.7	38.8	26	33.8	23.7	71.4	0	0	0					27.3	77.6	1010	35.4	6.6	261.1	22.6	9		2021-04-26T04:26T1	0.47	ially cic through	cloud	1909995	
Jakarta	27/04/2021	34	25.3	29.4	40.2	25.3	33.3	23.5	71.7	0	0	0					23.9	93.5	1010.1	41.2	5.8	255.7	22	9		2021-04-27T04:27T1	0.5	ially cic through	cloud	1909995	
Jakarta	28/04/2021	32	24.3	27.7	38.1	24.3	30.5	24	80.8	60.883	100	25	rain				18.8	328.5	1011.5	58.7	5.5	63.3	5.5	3		2021-04-28T04:28T1	0.54	artially with a c	rain	1909995	
Jakarta	29/04/2021	33	24.3	28.4	38.3	24.3	31.2	23.6	76.6	0.03	100	4.17	rain			31.7	18	113.1	1010.3	48.5	5.5	209.5	18.1	7		2021-04-29T04:29T1	0.57	artially the day	rain	1909995	
Jakarta	30/04/2021	32	25	28.8	38.5	25	32.9	24.2	77.2	0.089	100	8.33	rain				21.8	41.2	1009.8	55.2	5.6	184	15.9	7		2021-04-30T04:30T1	0.6	artially the day	rain	1909995	
Jakarta	01/05/2021	33.7	26	29.5	39.6	26	34.4	24.6	76.2	0.027	100	12.5	rain				25.1	88.3	1009.7	49.5	5.8	133.8	11.5	5		2021-05-01T05:01T1	0.64	artially ughout	rain	1909995	
Jakarta	02/05/2021	33.1	26.3	29.5	40.6	26.3	34.6	24.6	75.5	0.001	100	4.17	rain				26.7	73	1009.3	45.9	6.3	162.2	14.2	6		2021-05-02T05:02T1	0.67	artially the day	rain	1909995	
Jakarta	03/05/2021	33	26.3	29.3	39.8	26.3	33.8	24.2	74.8	0	0	0					28.3	88.5	1007.5	49.9	6.4	134.1	11.7	5		2021-05-03T05:03T1	0.71	ially cic through	cloud	1909995	
Jakarta	04/05/2021	33	25.3	29.1	38.4	25.3	33.3	23.9	74.1	0.009	100	4.17	rain				14.2	107.4	1007.1	46.4	6.1	221.6	19.2	8		2021-05-04T05:04T1	0.75	artially ut the d	rain	99,WH	
Jakarta	05/05/2021	32.3	26	28.8	39.1	26	33.2	24.4	77.8	33.866	100	8.33	rain				14	33.3	1007.6	53.9	5.3	160.2	14	6		2021-05-05T05:05T1	0.78	artially the day	rain	1909995	
Jakarta	06/05/2021	32	24.6	27.5	38.3	24.6	29.9	24.1	82.8	6.334	100	12.5	rain				16.6	332.7	1009.1	61.7	5.7	80.2	7	3		2021-05-06T05:06T1	0.81	artially the day	rain	1909995	
Jakarta	07/05/2021	34.4	24.6	28.9	40.7	24.6	32.8	24.3	77.1	0.763	100	4.17	rain				20.6	33.3	1009.3	49.1	5.9	150.6	13.2	6		2021-05-07T05:07T1	0.84	artially the day	rain	1909995	
Jakarta	08/05/2021	32.4	25	28.9	40.5	25	33.8	25	79.7	0.263	100	12.5	rain				21.9	44.1	1008.7	55.8	5.8	153	13	5		2021-05-08T05:08T1	0.88	artially ughout	rain	1909995	
Jakarta	09/05/2021	33.4	24	28.6	39.8	24	32.4	24.3	78.8	46.897	100	12.5	rain				18.4	77.2	1008.4	59.4	6	86.7	7.4	4		2021-05-09T05:09T1	0.91	artially the day	rain	1909995	
Jakarta	10/05/2021	33	26	29.4	39.3	26	33.5	24.2	75	4.377	100	8.33	rain				17.8	80.2	1008	47.5	5.9	254.8	22.1	9		2021-05-10T05:10T1	0.95	artially the day	rain	1909995	
Jakarta	11/05/2021	34.8	26.3	30.1	41.3	26.3	34.9	24.1	71.5	0	0	0					21.1	62.5	1007.8	38.7	6.1	242.7	20.9	9		2021-05-11T05:11T1	0.98	ially cic through	cloud	1909995	
Jakarta	12/05/2021	33.3	26.3	29.8	40.5	26.3	35.1	24.8	75.1	0	0	0					23.9	97	1007.2	47.3	5.7	204	17.5	7		2021-05-12T05:12T1	0	ially cic through	cloud	1909995	
Jakarta	13/05/2021	33	26	29.7	40.2	26	34.7	24.3	73.5	0.055	100	4.17	rain				21.1	81.3	1006.9	42.3	5.6	155.7	13.5	6		2021-05-13T05:13T1	0.05	artially ut the d	rain	1909995	
Jakarta	14/05/2021	33.1	26	29.6	38.5	26	33.6	23.5	70.7	0	0	0					26.7	103.5	1007.2	49	6.1	158.6	13.8	6		2021-05-14T05:14T1	0.08	ially cic through	cloud	1909995	
Jakarta	15/05/2021	33.3	26	29.8	39.6	26	34.3	23.7	70.5	0.005	100	8.33	rain				22.3	79.2	1007.6	48.9	6	176.2	15.2	7		2021-05-15T05:15T1	0.12	artially with ra	rain	1909995	
Jakarta	16/05/2021	33.7	26.3	29.7	39.8	26.3	34.1	23.8	71.7	0.001	100	4.17	rain				15.8	77.5	1007.9	47	6.1	153.3	13.4	6		2021-05-16T05:16T1	0.15	artially the day	rain	1909995	
Jakarta	17/05/2021	32.7	25	29.4	38.8	25	33.9	24.2	74.6	26.256	100	4.17	rain				18.1	16.1	1008.4	55.7	5.7	111.2	9.9	5		2021-05-17T05:17T1	0.18	artially the day	rain	1909995	
Jakarta	18/05/2021	32.1	23.9	27.4	37.3	23.9	30	23.6	80.8	65.395	100	12.5	rain				16.6	83	1008.9	65.1	5.9	104.8	9.3	4		2021-05-18T05:18T1	0.22	artially the day	rain	1909995	
Jakarta	19/05/2021	33	26	29.1	39.9	26	33.5	24.3	76	0.2	100	4.17	rain				18	51.5	1008.8	50.8	5.9	144.5	12.4	5		2021-05-19T05:19T1	0.25	artially the day	rain	1909995	
Jakarta	20/05/2021	32.8	26	29	39.6	26	33	24.2	75.8	0.193	100	4.17	rain				24.9	77.8	1010.1	55	5.7	187.1	16	7		2021-05-20T05:20T1	0.25	artially the day	rain	1909995	
Jakarta	21/05/2021	33.4	25.3	28.6	40.2	25.3	32.2	24.2	78.1	0.07	100	4.17	rain				17.8	353.2	1010.3	52.4	5.5	151.1	13.1	5		2021-05-21T05:21T1	0.32	artially the day	rain	1909995	
Jakarta	22/05/2021	32	25.8	28.6	39.8	25.8	32.6	24.1	77.2	0.012	100	4.17	rain			33.5	17.3	304.4	1009.6	57.5	5.1	167	14.3	6		2021-05-22T05:22T1	0.35	artially the day	rain	1674909	
Jakarta	23/05/2021	32.1	25.3	28.3	39.5	25.3	32.1	24.3	79.5	0	0	0					19.9	19.5	1009.5	51.3	5	108.1	9.3	4		2021-05-23T05:23T1	0.39	ially cic through	cloud	1909995	
Jakarta	24/05/2021	32.7	25	28.7	40.2	25	33.1	24.8	80.1	8.836	100	8.33	rain				19.5	28.2	1010	62.5	4.8	198.8	17.1	7		2021-05-24T05:24T1	0.42	artially ughout	rain	1909995	
Jakarta	25/05/2021	33.4	25	28.4	40.3	25	31.4	23.9	78	7.082	100	8.33	rain				16.5	83.4	1009.9	49.9	6	157.7	13.6	6		2021-05-25T05:25T1	0.46	artially the day	rain	IH,9674	
Jakarta	26/05/2021	32.8	26	29.3	39.2	26	34.1	24.6	76.4	0	0	0					33.5	84.3	1009.3	40.7	6.2	230	19.9	8		2021-05-26T05:26T1	0.5	ially cic through	cloud	1909995	
Jakarta	27/05/2021	33.7	26.3	29.8	40.6	26.3	34.9	24.5	73.9	0.018	100	4.17	rain				27.2	85.7	1009.7	42.9	6.4	212.7	18.6	8		2021-05-27T05:27T1	0.52	artially ut the d	rain	1674909	
Jakarta	28/05/2021	33.8	25.3	29.4	40.8	25.3	34.2	24.5	76	0.073	100	4.17	rain				17	78.7	1010.5	42.5	6	217	18.7	8		2021-05-28T05:28T1	0.56	artially ut the d	rain	1909995	
Jakarta	29/05/2021	30	24	26.8	36.1	24	29.2	24.1	85.7	9.842	100	12.5	rain				24.1	264.6	1011.1	67.5	5.7	95.8	8.2	4		2021-05-29T05:29T1	0.59	artially ughout	rain	1909995	
Jakarta	30/05/2021	32.7	24	28.4	38.6	24	32	24.4	80.3	0.35	100	4.17	rain				17.8	39.3	1009	55.3	5.8	212.2	18.5	8		2021-05-30T05:30T1	0.62	artially ut the d	rain	1909995	
Jakarta	31/05/2021	32.7	26	29	39.8	26	33.5	24.6	78.1	0.058	100	12.5	rain				11.2	343.6	1008.3	48.1	5.8	162	14	6		2021-05-31T05:31T1	0.66	artially the day	rain	1909995	
Jakarta	01/06/2021	32.7	26	28.5	39.8	26	32.3	24.6	80.2	0.237	100	4.17	rain				16.2	32	1008.6	55.5	5.1	169.8	14.5	6		2021-06-01T06:01T1	0.69	artially the day	rain	1909995	
Jakarta	02/06/2021	32.7	25.3	28.5	39.7	25.3	32.4	24.5	79.5	0.461	100	12.5	rain				14	79.8	1009	52.4	5.6	218.8	18.9	8		2021-06-02T06:02T1	0.75	artially ughout	rain	1909995	
Jakarta	03/06/2021	33	25.3	29.1	41	25.3	34.2	25.1	79.6	7.72	100	8.33	rain				15.8	69.9	1009.3	50.5	5.6	188.7	16.3	7		2021-06-03T06:03T1	0.76	artially ughout	rain	1909995	
Jakarta	04/06/2021	33.7	26	29.4	40.6	26	35	25.3	79.5	0.017	100	8.33	rain			50	24.1	95.3	1009.7	59.7	5.5	161.4	14	6		2021-06-04T06:04T1	0.79	artially ughout	rain	1909995	
Jakarta	05/06/2021	32.7	25.3	29.1	40.2	25.3	33.8	24.8	78	0	0	0					14.6	28.5	1010.2	49.7	5.7	214.9	18.5	8		2021-06-05T06:05T1	0.83	ially cic through	cloud	1909995	
Jakarta	06/06/2021	32.7	25.6	28.8	38.9	25.6	32.2	24.4	78.7	31.182	100	12.5	rain				16.8	323.2	1010.5	59.5	5.2	216	18.7	7		2021-06-06T06:06T1	0.86	artially ughout	rain	1909995	
Jakarta	07/06/2021	32.7	24.3	28.5	39.4	24.3	31.8	23.9	77.5	0.007	100	8.33	rain				18.1	330.2	1010.7	47.3	5.8	165.3	14.4	7		2021-06-07T06:07T1	0.89	artially ughout	rain	1909995	
Jakarta	08/06/2021	32.2	25.3	29	39.8	25.3	33.5	24.3	76.2	0	0	0					15.8	286.4	1010.3	48.7	5.9	226.1	19.5	8		2021-06-08T06:08T1	0.93	ially cic through	cloud	1909995	
Jakarta	09/06/2021	32.1	25	28.4	39.4	25	32.2	24.3	79.3	1.929	100	4.17	rain				22.4	51.6	1009.5	48.5	5.8	213	18.4	7		2021-06-09T06:09T1	0.96	artially the day	rain	1909995	

3.2. Meta Data

Berikut adalah *meta data* dari dataset yang dipakai :

1. Sumber : https://weather.visualcrossing.com/VisualCrossingWebServices/rest/services/timeline
2. Author : visualcrossing
3. Ukuran File : 506 kb
4. Jumlah Baris : 1001
5. Jumlah Kolom : 33

3.3. Transformasi Data

Dari data set original yang akan digunakan hanya kolom “name”, “datetime”, “tempmax”, dam “tempmin” yang akan digunakan, disini bisa dilihat bahwa penggunaan format data dalam kolom “tempmax”, dan “tempmin” masih menggunakan titik yang merupakan format general/text, di tahap selanjutnya data cleaning akan diubah menjadi format number menggunakan koma(,). Berikut data yang sudah di transformasi :

name	datetime	tempmax	tempmin
Jakarta	07/04/2021	30.6	25
Jakarta	08/04/2021	33	25
Jakarta	09/04/2021	32	26
Jakarta	10/04/2021	32	23.7
Jakarta	11/04/2021	32.7	24.6
Jakarta	12/04/2021	33	24.3
Jakarta	13/04/2021	33	25
Jakarta	14/04/2021	32	24.3
Jakarta	15/04/2021	32	24
Jakarta	16/04/2021	32	24.3
Jakarta	17/04/2021	32	25.3
Jakarta	18/04/2021	32	25.3
Jakarta	19/04/2021	32.7	25
Jakarta	20/04/2021	32.7	24
Jakarta	21/04/2021	33.7	24.7
Jakarta	22/04/2021	33	24.5
Jakarta	23/04/2021	32.7	24.3
Jakarta	24/04/2021	32.7	24.3
Jakarta	25/04/2021	33	25.3
Jakarta	26/04/2021	34	26
Jakarta	27/04/2021	34	25.3
Jakarta	28/04/2021	32	24.3
Jakarta	29/04/2021	33	24.3

BAB IV
PERSIAPAN DATA

4.1. Data Cleaning (Menggunakan Python)

8.1 Modul Yang Digunakan :

- Pandas

8.2 Kode Program :

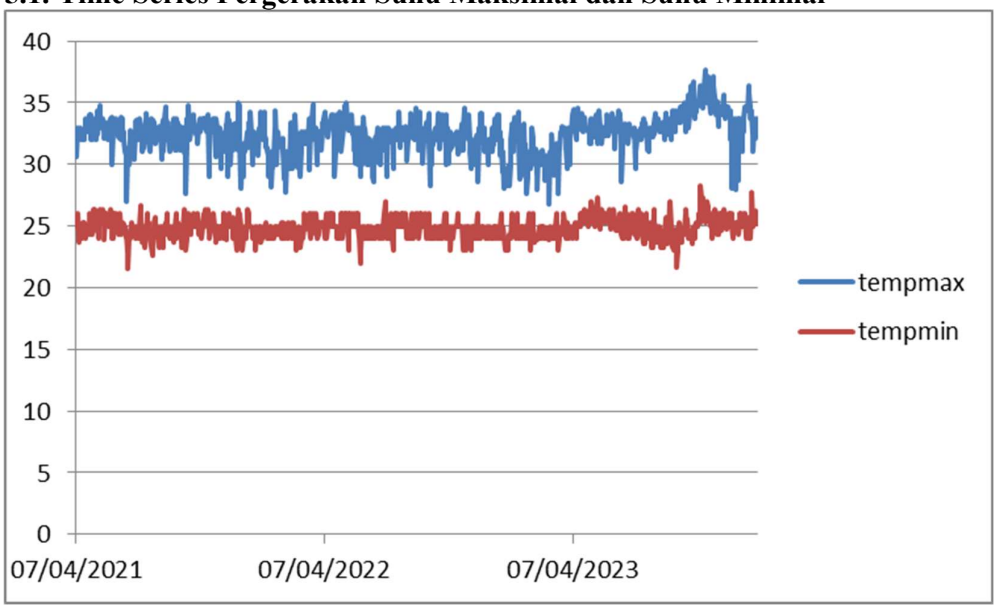
```
import pandas as pd

# Membaca file Excel
file_path = 'Data_Cuaca.xlsx'
data = pd.read_excel(file_path)

# 1. Mengambil kolom yang dibutuhkan
selected_columns = ['name', 'datetime', 'tempmax',
'tempmin']
data_selected = data[selected_columns]

# 2. Mengubah titik menjadi koma pada kolom tempmax dan
tempmin
data_selected['tempmax'] =
data_selected['tempmax'].str.replace(',',
'.').astype(float)
data_selected['tempmin'] =
data_selected['tempmin'].str.replace(',',
'.').astype(float)

# 3. Menghilangkan data duplikat
data_selected.drop_duplicates(inplace=True)

# 4. Menghilangkan baris dengan data kosong
data_selected.dropna(inplace=True)

# 5. Menghapus data outlier
data_selected = data_selected[(data_selected['tempmax']
<= 40) & (data_selected['tempmin'] >= 20)]

# Menyimpan data yang sudah diolah ke file baru
(Opsional)
output_file_path = 'Data_Cuaca_Bersih.xlsx'
data_selected.to_excel(output_file_path, index=False)

# Menampilkan data yang sudah diolah
print("Data Berhasil Dibersihkan")
```

8.3 Output :

name	datetime	tempmax	tempmin
Jakarta	2021-04-07 00:00:00	30,6	25
Jakarta	2021-04-08 00:00:00	33	25
Jakarta	2021-04-09 00:00:00	32	26
Jakarta	2021-04-10 00:00:00	32	23,7
Jakarta	2021-04-11 00:00:00	32,7	24,6
Jakarta	2021-04-12 00:00:00	33	24,3
Jakarta	2021-04-13 00:00:00	33	25
Jakarta	2021-04-14 00:00:00	32	24,3
Jakarta	2021-04-15 00:00:00	32	24
Jakarta	2021-04-16 00:00:00	32	24,3
Jakarta	2021-04-17 00:00:00	32	25,3
Jakarta	2021-04-18 00:00:00	32	25,3
Jakarta	2021-04-19 00:00:00	32,7	25
Jakarta	2021-04-20 00:00:00	32,7	24
Jakarta	2021-04-21 00:00:00	33,7	24,7
Jakarta	2021-04-22 00:00:00	33	24,5
Jakarta	2021-04-23 00:00:00	32,7	24,3
Jakarta	2021-04-24 00:00:00	32,7	24,3

BAB V
VISUALISASI DATA

5.1. Time Series Pergerakan Suhu Maksimal dan Suhu Minimal



BAB VI
PEMROSESAN DATA
MENGGUNAKAN PYTHON

6.1. Pemrosesan Metode K-Means : sekaligus uji performa Cluster menggunakan metode Silhouette

6.1.1. Modul Yang Digunakan :

- pandas
- matplotlib
- sklearn

6.1.2. Kode Program :

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt

# Membaca file Excel
file_path = "Data_Cuaca_Bersih.xlsx"
df = pd.read_excel(file_path)

# Memilih kolom tempmax dan tempmin
data = df[['tempmax', 'tempmin']]

# Scaling menggunakan Min-Max Scaling
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(data)

# Loop untuk mencoba nilai num_clusters dari 2 hingga 5
for num_clusters in range(2, 6):
    # Melakukan KMeans clustering
    kmeans = KMeans(n_clusters=num_clusters,
random_state=42)
    df['cluster'] = kmeans.fit_predict(data_scaled)

    # Menilai performa clustering menggunakan Silhouette
Score
    silhouette_avg = silhouette_score(data_scaled,
df['cluster'])
    print(f"Silhouette Score
(num_clusters={num_clusters}): {silhouette_avg}")

    # Menyimpan hasil clustering ke dalam file Excel
    output_file_path =
f"Data_Cuaca_Bersih_Clustered_{num_clusters}.xlsx"
    df.to_excel(output_file_path, index=False)
```

```
    # Menampilkan dan menyimpan scatter plot hasil
clustering
    plt.scatter(df['tempmax'], df['tempmin'],
c=df['cluster'], cmap='viridis')
    plt.title(f'Scatter Plot Hasil Clustering
(num_clusters={num_clusters})')
    plt.xlabel('TempMax')
    plt.ylabel('TempMin')
    plt.savefig(f'Scatter_Plot_Clustered_{num_clusters}.pn
g')
    plt.show()
```

6.2. Pemrosesan Metode Gaussian Mixture Models : sekaligus uji performa Cluster menggunakan metode Silhouette

6.2.1. Modul Yang Digunakan :

- pandas
- matplotlib
- statsmodels

6.2.2. Kode Program :

```
import pandas as pd
from sklearn.mixture import GaussianMixture
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt

# Membaca file Excel
file_path = "Data_Cuaca_Bersih.xlsx"
df = pd.read_excel(file_path)

# Memilih kolom tempmax dan tempmin
data = df[['tempmax', 'tempmin']]

# Scaling menggunakan Min-Max Scaling
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(data)

# Loop untuk mencoba nilai num_components dari 2 hingga 5
for num_components in range(2, 6):
    # Melakukan Gaussian Mixture Model clustering
    gmm = GaussianMixture(n_components=num_components,
random_state=42)
    df['cluster'] = gmm.fit_predict(data_scaled)

    # Menilai performa clustering menggunakan Silhouette
Score
```

```
    silhouette_avg = silhouette_score(data_scaled,
df['cluster'])
    print(f"Silhouette Score
(num_components={num_components}): {silhouette_avg}")

    # Menyimpan hasil clustering ke dalam file Excel
    output_file_path =
f"Data_Cuaca_Bersih_Clustered_GMM_{num_components}.xlsx"
    df.to_excel(output_file_path, index=False)

    # Menampilkan dan menyimpan scatter plot hasil
clustering
    plt.scatter(df['tempmax'], df['tempmin'],
c=df['cluster'], cmap='viridis')
    plt.title(f'Scatter Plot Hasil Clustering GMM
(num_components={num_components})')
    plt.xlabel('TempMax')
    plt.ylabel('TempMin')
    plt.savefig(f'Scatter_Plot_Clustered_GMM_{num_componen
ts}.png')
    plt.show()
```

BAB VII
HASIL

7.1. Hasil Metode K-Means : sekaligus uji performa Cluster menggunakan metode Silhouette

7.2.1. Output :

- Silhouette Score (num_clusters=2): 0.37363226954587125
- Silhouette Score (num_clusters=3): 0.3760170075621142
- Silhouette Score (num_clusters=4): 0.35835217804018804
- Silhouette Score (num_clusters=5): 0.3896178137614342

Berdasarkan hasil Silhouette Score yang telah diberikan, kita dapat menginterpretasikan kondisi rentang suhu cuaca sebagai berikut:

- **Variasi Ekstrem (num_clusters=5):**
  - Silhouette Score: 0.3896
  - Pada konfigurasi dengan lima kluster, terdapat kecenderungan variasi ekstrem dalam rentang suhu cuaca. Nilai Silhouette Score yang relatif tinggi (0.3896) menandakan bahwa pembentukan lima kluster memberikan hasil yang cukup baik, dan mungkin terdapat variasi suhu yang signifikan antar kluster.

- **Kecenderungan Cuaca yang Konstan (num_clusters=2, num_clusters=3):**
  - Silhouette Score (num_clusters=2): 0.3736
  - Silhouette Score (num_clusters=3): 0.3760
  - Pada konfigurasi dengan dua dan tiga kluster, terdapat kemungkinan adanya kecenderungan cuaca yang konstan. Meskipun nilai Silhouette Score sedikit lebih rendah dibandingkan dengan lima kluster, namun tetap menunjukkan adanya struktur yang baik dalam pembentukan kluster.

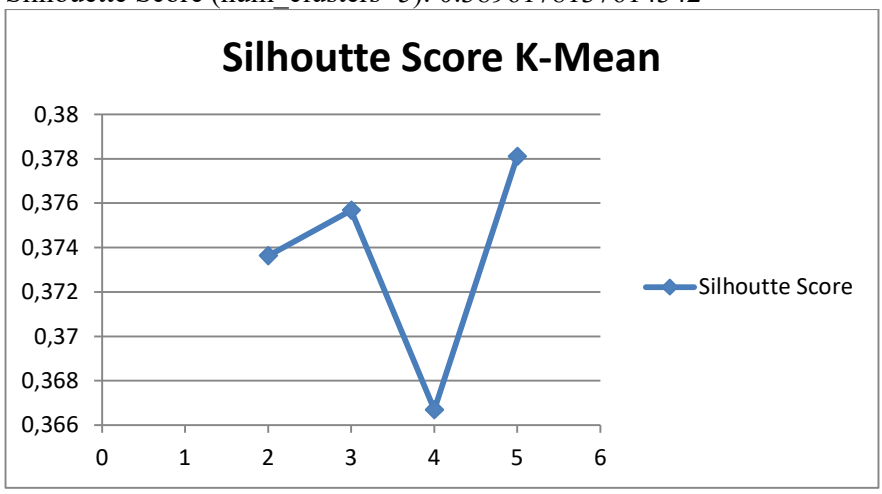
- **Analisis Tambahan (num_clusters=4):**
  - Silhouette Score: 0.3584
  - Konfigurasi dengan empat kluster memiliki nilai Silhouette Score yang lebih rendah, menunjukkan kemungkinan adanya beberapa kelompok suhu cuaca yang mungkin kurang terdefinisi dengan baik atau memiliki tumpang tindih.

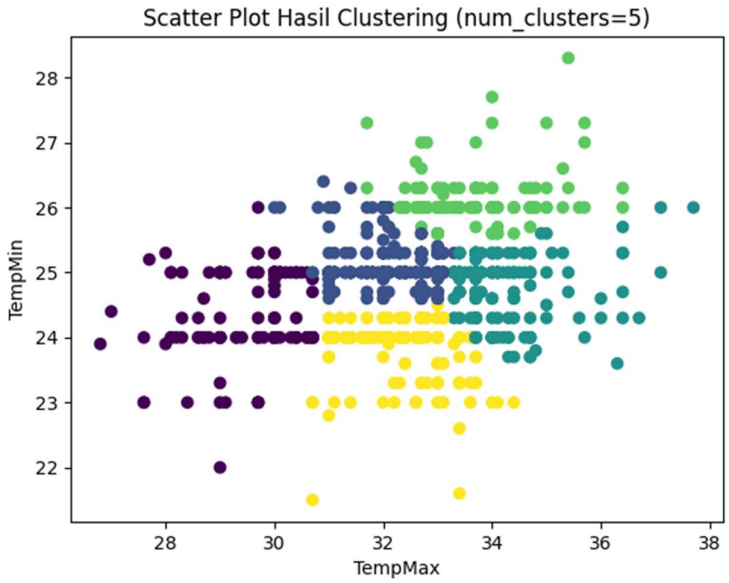
Berdasarkan hasil tersebut, konfigurasi dengan lima kluster mungkin lebih menggambarkan variasi ekstrem dalam rentang

suhu cuaca, sementara konfigurasi dengan dua atau tiga kluster mengindikasikan adanya kecenderungan cuaca yang konstan.

7.2.2. Grafik :

- Silhouette Score (num_clusters=5): 0.3896178137614342





Nilai Silhouette Score berkisar antara -1 hingga 1, dan semakin tinggi nilainya, semakin baik hasil clusteringnya. Dari hasil Uji performa Cluster K-Mean, bisa di lihat bahwa num_clusters=5 yang mendapatkan skor Silhouettle paling besar atau mendekati 1, menunjukkan bahwa pembentukan lima kluster memberikan hasil yang cukup baik. Nilai positif mengindikasikan bahwa objek-objek dalam kluster tersebut lebih serupa dengan anggota klusternya sendiri daripada dengan kluster lain.

7.2. Hasil Metode Gaussian Mixture Models : sekaligus uji performa Cluster menggunakan metode Silhouette

7.2.3. Output :

- SilhouetteScore (num_components=2): 0.32422138482991714
- SilhouetteScore (num_components=3): 0.32460686677351647
- Silhouett Score (num_components=4): 0.23162912616883433
- Silhouete Score (num_components=5): 0.16472131628154682

Berdasarkan hasil Silhouette Score, kita dapat menginterpretasikan kondisi rentang suhu cuaca sebagai berikut:

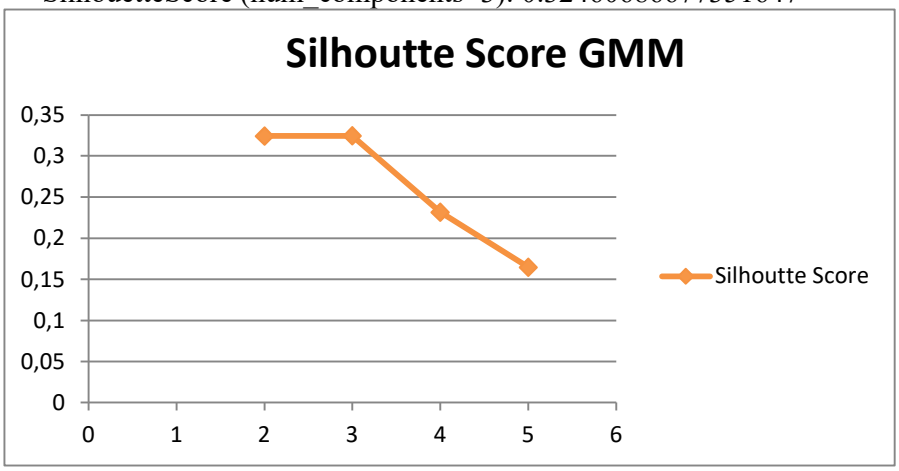
- **Variasi Ekstrem (num_components=3 dan num_components=2):**
  - Silhouette Score (num_components=3): 0.3246
  - Silhouette Score (num_components=2): 0.3242
  - Meskipun nilai Silhouette Score tidak sangat tinggi, konfigurasi dengan dua atau tiga komponen mungkin menunjukkan variasi ekstrem dalam rentang suhu cuaca. Keberadaan dua atau tiga komponen dapat mengindikasikan kecenderungan variasi suhu yang signifikan antar komponen.

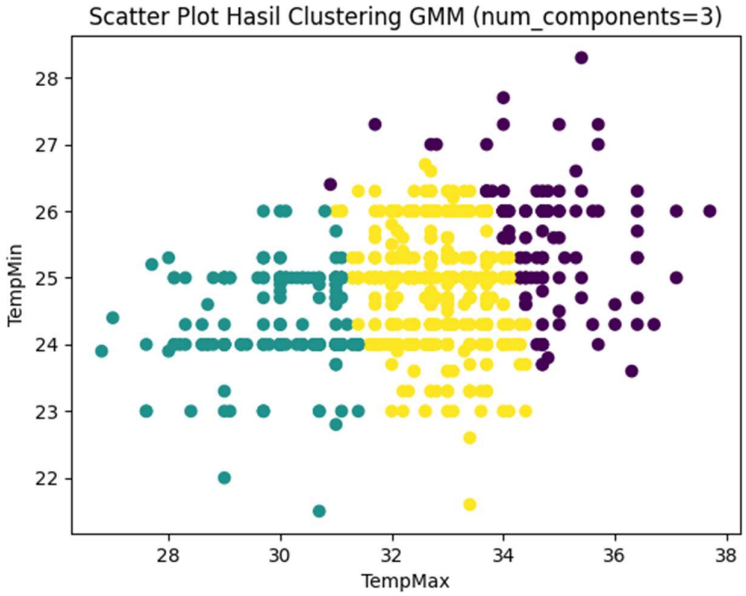
- **Kecenderungan Cuaca yang Konstan (num_components=4 dan num_components=5):**
  - Silhouette Score (num_components=4): 0.2316
  - Silhouette Score (num_components=5): 0.1647
  - Konfigurasi dengan empat dan lima komponen memiliki nilai Silhouette Score yang lebih rendah, menunjukkan kemungkinan adanya kecenderungan cuaca yang konstan. Nilai yang lebih rendah mungkin menandakan adanya struktur yang kurang terdefinisi dengan baik atau tumpang tindih antar komponen.

Berdasarkan hasil tersebut, konfigurasi dengan dua atau tiga komponen mungkin lebih menggambarkan variasi ekstrem dalam rentang suhu cuaca, sementara konfigurasi dengan empat atau lima komponen mengindikasikan adanya kecenderungan cuaca yang konstan.

7.2.4. Grafik :

- SilhouetteScore (num_components=3): 0.32460686677351647

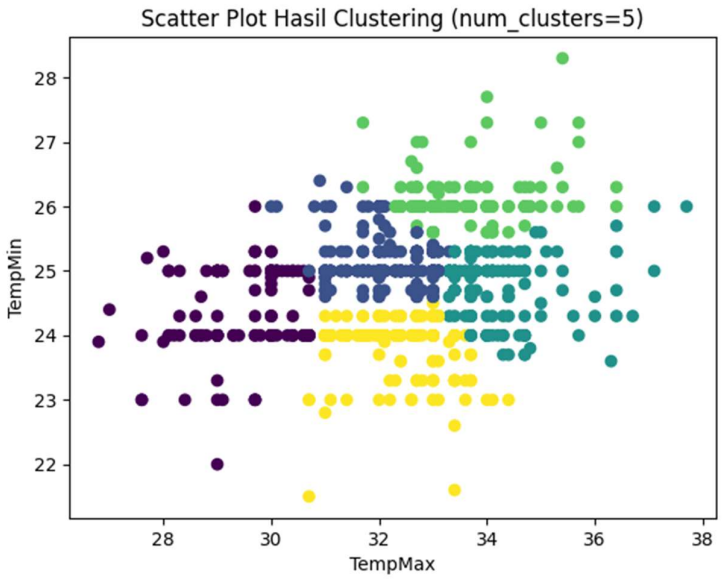




Dari hasil Uji performa Cluster Gaussian Mixture Models, bisa di lihat bahwa num_component=3 yang mendapatkan skor Silhouettle paling besar atau mendekati 1, menunjukkan bahwa pembentukan 3 kluster memberikan hasil yang cukup baik. Nilai positif mengindikasikan bahwa objek-objek dalam kluster tersebut lebih serupa dengan anggota klusternya sendiri daripada dengan kluster lain.

7.3. Perbandingan pemetaan data hasil metode K-Means dengan Gaussian Mixture Models

7.3.1. Grafik K-Means:



7.3.2. Grafik Gaussian Mixture Models:

Berdasarkan hasil Silhouette Score dan informasi kondisi rentang suhu cuaca, metode K-Means dengan 5 kluster dapat dianggap lebih optimal.

- **Lebih Tinggi Silhouette Score:** K-Means memiliki Silhouette Score yang lebih tinggi (0.3896) dibandingkan dengan Gaussian Mixture Models (GMM) (0.3246). Nilai Silhouette Score yang lebih tinggi menunjukkan kualitas klustering yang lebih baik.
- **Interpretasi Silhouette Score yang Lebih Tinggi:** Silhouette Score yang lebih tinggi pada metode K-Means menandakan bahwa objek-objek dalam kluster K-Means lebih serupa dengan anggota klusternya sendiri dan lebih berbeda dengan objek-objek dari kluster lain, memperlihatkan pembagian kelompok yang lebih baik.

BAB VIII
PENUTUP

8.1 Kesimpulan

8.1.1 Untuk membagi membagi data suhu cuaca harian berdasarkan nilai suhu minimal dan maksimal ke dalam kelompok-kelompok yang serupa dengan cara membuat kluster rentang kondisi suhu cuaca di jakarta berdasarkan nilai suhu cuaca minimal dan suhu cuaca maksimal jakarta tahun 2021-2023. berikut Langkah-langkahnya :

a. Pemahaman data nilai suhu cuaca minimal dan suhu cuaca maksimal jakarta tahun 2021-2023
b. Preprocessing terhadap data yang sudah disebutkan tadi
c. Melakukan visualisasi Time Series Pergerakan Suhu Maksimal dan Suhu Minimal
d. Kemudian, menerapkan algoritma K-Means dan GMM sekaligus uji performa Cluster menggunakan metode Silhouette per untuk mengelompokkan rentang kondisi suhu cuaca
e. Menginterpretasikan kondisi cuaca per cluster berdasarkan hasil cluster suhu cuaca minimal dan suhu cuaca maksimal

Berikut insterpretasi hasil rentang kondisi suhu cuaca di jakarta berdasarkan nilai suhu cuaca minimal dan suhu cuaca maksimal jakarta tahun 2021-2023:

a. Metode K-Means

- **Variasi Ekstrem (num_clusters=5):**

  Pada konfigurasi dengan lima kluster, terdapat kecenderungan variasi ekstrem dalam rentang suhu cuaca. Nilai Silhouette Score yang relatif tinggi (0.3896) menandakan bahwa pembentukan lima kluster memberikan hasil yang cukup baik, dan mungkin terdapat variasi suhu yang signifikan antar kluster.

- **Kecenderungan Cuaca yang Konstan (num_clusters=2, num_clusters=3):**

  Pada konfigurasi dengan dua dan tiga kluster, terdapat kemungkinan adanya kecenderungan cuaca yang konstan. Meskipun nilai Silhouette Score sedikit lebih rendah dibandingkan dengan lima kluster, namun tetap menunjukkan adanya struktur yang baik dalam pembentukan kluster.

- **Analisis Tambahan (num_clusters=4):**

  Konfigurasi dengan empat kluster memiliki nilai Silhouette Score yang lebih rendah, menunjukkan kemungkinan adanya beberapa kelompok suhu cuaca yang mungkin kurang terdefinisi dengan baik atau memiliki tumpang tindih.

b. Metode Gaussian Mixture Models

- **Variasi Ekstrem (num_components=3 dan num_components=2):**

  Meskipun nilai Silhouette Score tidak sangat tinggi, konfigurasi dengan dua atau tiga komponen mungkin menunjukkan variasi ekstrem dalam rentang suhu cuaca. Keberadaan dua atau tiga komponen dapat mengindikasikan kecenderungan variasi suhu yang signifikan antar komponen.

- **Kecenderungan Cuaca yang Konstan (num_components=4 dan num_components=5):**

  Konfigurasi dengan empat dan lima komponen memiliki nilai Silhouette Score yang lebih rendah, menunjukkan kemungkinan adanya kecenderungan cuaca yang konstan. Nilai yang lebih rendah mungkin menandakan adanya struktur yang kurang terdefinisi dengan baik atau tumpang tindih antar komponen.

8.1.2 Berdasarkan hasil Silhouette Score dan informasi kondisi rentang suhu cuaca, metode K-Means dengan 5 kluster dapat dianggap lebih optimal

a. **Lebih Tinggi Silhouette Score:** K-Means memiliki Silhouette Score yang lebih tinggi (0.3896) dibandingkan dengan Gaussian Mixture Models (GMM) (0.3246). Nilai Silhouette Score yang lebih tinggi menunjukkan kualitas klustering yang lebih baik.
b. **Interpretasi Silhouette Score yang Lebih Tinggi:** Silhouette Score yang lebih tinggi pada metode K-Means menandakan bahwa objek-objek dalam kluster K-Means lebih serupa dengan anggota klusternya sendiri dan lebih berbeda dengan objek-objek dari kluster lain, memperlihatkan pembagian kelompok yang lebih baik.