

Final SSO SSF UNS 2020

Tim Arah

SSF118308





Contents

- I. Problem Statement
- II. Workflow Penyelesaian Masalah

I. Problem Statement

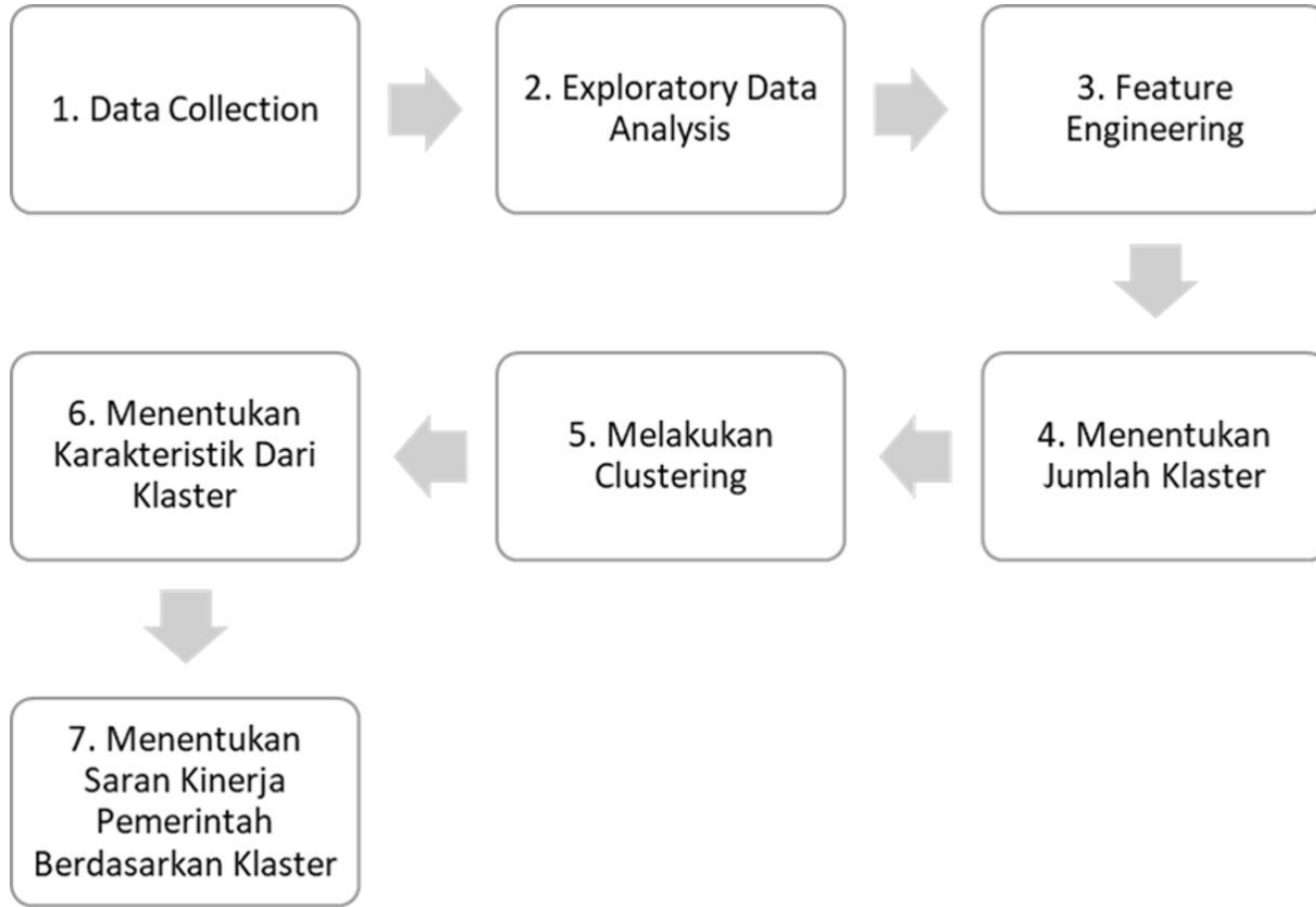
Masalah

Mempercepat **penanganan COVID-19** yang lebih terarah untuk semua provinsi di Indonesia

Tujuan

1. Menentukan **kelompok/cluster** provinsi di Indonesia berdasarkan data
2. Menentukan **saran kinerja pemerintah** dalam penanganan COVID-19 yang lebih efektif dan efisien **berdasarkan karakteristik dari setiap klaster**

II. Workflow Penyelesaian Masalah



1. Data Collection

Sumber Data :

Naskah Final SSO SSF UNS 2020 yang bersumber dari situs resmi BPS dan kawal covid-19 diakses pada tanggal 17 Oktober 2020 16:00 WIB

2. Exploratory Data Analysis

	count	mean	std	min	25%	50%	75%	max
X1	34.0	10522.411765	17805.485760	499.00	1589.750	4052.00	11193.500	93356.00
X2	34.0	8282.117647	14885.514349	368.00	1394.500	3215.50	8111.500	77969.00
X3	34.0	126.500000	198.614023	0.00	24.500	48.50	123.250	974.00
X4	34.0	119.058824	206.392674	0.00	17.250	41.00	117.000	1106.00
X5	34.0	365.617647	701.378081	5.00	43.750	131.00	376.000	3529.00
X6	34.0	2.470588	4.560936	0.00	0.000	0.50	2.000	21.00
X7	34.0	0.758529	0.122503	0.38	0.685	0.79	0.855	0.94
X8	34.0	0.030294	0.016420	0.01	0.020	0.03	0.040	0.07
X9	34.0	2.646765	2.700350	0.00	0.710	2.20	3.390	13.13
X10	34.0	25715.529412	78780.344321	0.00	0.000	0.00	5453.500	385352.00
X11	33.0	19995.333333	63543.060652	0.00	0.000	0.00	3983.000	284304.00
X12	34.0	5430.235294	16902.228211	0.00	0.000	0.00	2551.000	96844.00
X13	34.0	294.500000	833.206521	0.00	0.000	0.00	113.000	4519.00
X14	34.0	4542.235294	16030.010348	0.00	0.000	0.00	1346.000	92325.00
X15	34.0	461.029412	885.797038	0.00	0.000	0.00	369.250	3781.00
X16	34.0	742.029412	2708.764311	9.00	54.500	103.50	261.250	15900.00
X17	34.0	457414.529412	123845.255516	0.00	395236.250	454847.00	523044.750	732570.00

3. Feature Engineering

Ditemukan feature yang bermasalah yaitu

X10 : Kumulatif total kasus OTG (jiwa)

X11 : Total kasus OTG selesai (jiwa)

X12 : Kumulatif total kasus ODP (jiwa)

X13 : Total kasus ODP dalam proses (jiwa)

X14 : Total kasus ODP selesai (jiwa)


X15 : Kumulatif total kasus PDP (jiwa)

X17 : Garis Kemiskinan (Rupiah/kapita/bulan)

Pearson Correlation of features

X1	1	1	0.88	0.95	0.81	0.88	0.15	0.17	-0.16	0.82	0.87	-0.0047	0.0041	0.015	0.049	0.36	-0.054
X2	1	1	0.85	0.94	0.83	0.89	0.2	0.19	-0.17	0.84	0.89	-0.036	-0.025	-0.018	0.021	0.39	-0.092
X3	0.88	0.85	1	0.95	0.58	0.76	-0.11	0.014	-0.066	0.58	0.65	0.14	0.11	0.16	0.22	0.078	0.14
X4	0.95	0.94	0.95	1	0.66	0.79	0.049	0.047	-0.1	0.68	0.74	-0.014	-0.026	0.0026	0.044	0.16	0.077
X5	0.81	0.83	0.58	0.66	1	0.79	0.23	0.49	-0.19	0.94	0.95	-0.033	-0.019	-0.014	0.1	0.78	-0.4
X6	0.88	0.89	0.76	0.79	0.79	1	0.1	0.23	-0.076	0.77	0.82	0.22	0.23	0.24	0.25	0.38	-0.18
X7	0.15	0.2	-0.11	0.049	0.23	0.1	1	0.3	-0.13	0.2	0.19	-0.29	-0.3	-0.32	-0.16	0.2	-0.16
X8	0.17	0.19	0.014	0.047	0.49	0.23	0.3	1	-0.29	0.32	0.29	-0.15	-0.11	-0.12	-0.021	0.42	-0.37
X9	-0.16	-0.17	-0.066	-0.1	-0.19	-0.076	-0.13	-0.29	1	-0.13	-0.12	0.5	0.31	0.38	0.21	-0.045	-0.049
X10	0.82	0.84	0.58	0.68	0.94	0.77	0.2	0.32	-0.13	1	0.99	-0.09	-0.099	-0.077	-0.081	0.79	-0.42
X11	0.87	0.89	0.65	0.74	0.95	0.82	0.19	0.29	-0.12	0.99	1	-0.096	-0.11	-0.084	-0.16	0.73	-0.39
X12	-0.0047	-0.036	0.14	-0.014	-0.033	0.22	-0.29	-0.15	0.5	-0.09	-0.096	1	0.93	0.98	0.57	-0.056	-0.099
X13	0.0041	-0.025	0.11	-0.026	-0.019	0.23	-0.3	-0.11	0.31	-0.099	-0.11	0.93	1	0.94	0.6	-0.068	-0.12
X14	0.015	-0.018	0.16	0.0026	-0.014	0.24	-0.32	-0.12	0.38	-0.077	-0.084	0.98	0.94	1	0.56	-0.062	-0.084
X15	0.049	0.021	0.22	0.044	0.1	0.25	-0.16	-0.021	0.21	-0.081	-0.16	0.57	0.6	0.56	1	-0.096	0.12
X16	0.36	0.39	0.078	0.16	0.78	0.38	0.2	0.42	-0.045	0.79	0.73	-0.056	-0.068	-0.062	-0.096	1	-0.68
X17	-0.054	-0.092	0.14	0.077	-0.4	-0.18	-0.16	-0.37	-0.049	-0.42	-0.39	-0.099	-0.12	-0.084	0.12	-0.68	1
X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	

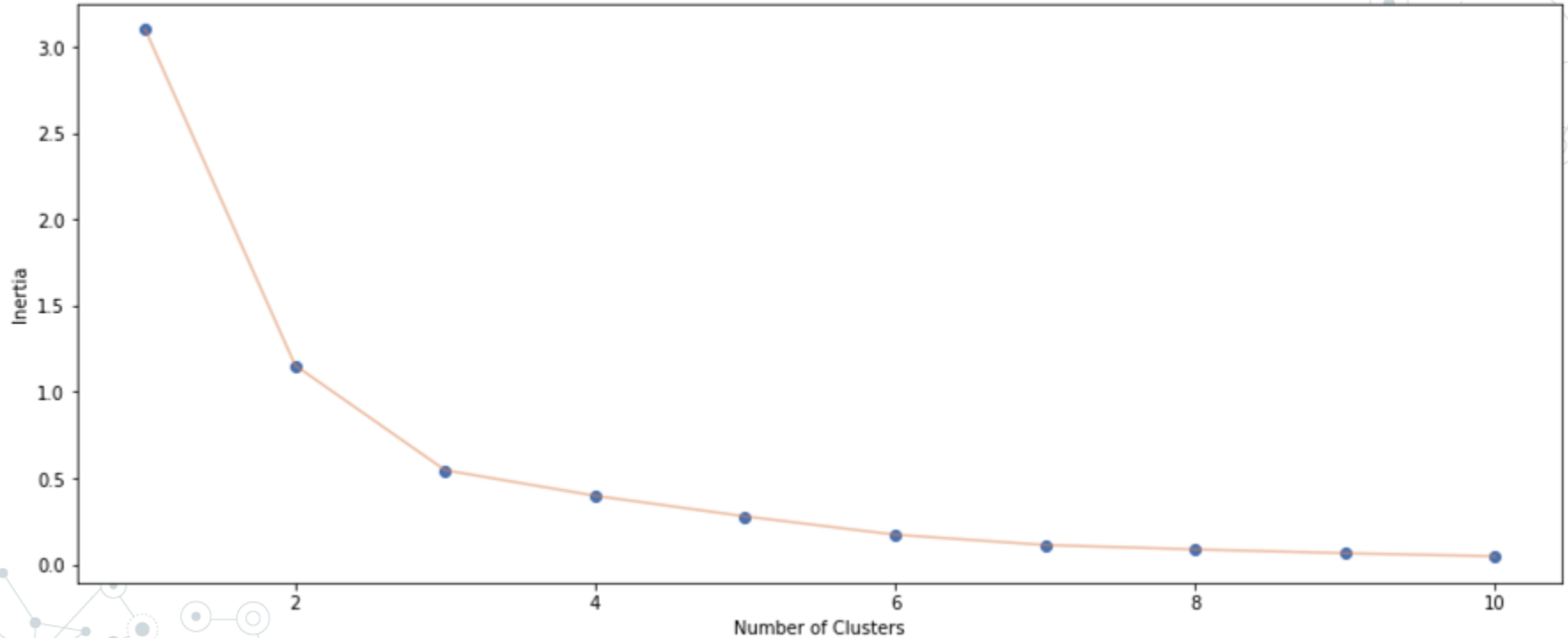




Maka, feature yang disubstitusi missing value-nya adalah feature X17. Feature X10, X11, X12, X13, X14, X15 tidak digunakan dalam pemodelan selanjutnya.

4. Menentukan Jumlah Kluster

Metode Elbow



Metode Silhouette

For n_clusters=2, The Silhouette Coefficient is 0.530659664518871
For n_clusters=3, The Silhouette Coefficient is 0.5429858339033454
For n_clusters=4, The Silhouette Coefficient is 0.4872891042142783
For n_clusters=5, The Silhouette Coefficient is 0.522886616218323
For n_clusters=6, The Silhouette Coefficient is 0.5352079262375548
For n_clusters=7, The Silhouette Coefficient is 0.5013648263382846
For n_clusters=8, The Silhouette Coefficient is 0.483223290661232
For n_clusters=9, The Silhouette Coefficient is 0.49951372221459195
For n_clusters=10, The Silhouette Coefficient is 0.48302786299627115

Berdasarkan 2 metode di atas, jumlah klaster yang dipilih yaitu 3 klaster

5. Melakukan Clustering

K-Means Clustering

- Klaster 0 : Jawa Barat, Jawa Tengah, Sulawesi Tenggara, Sulawesi Selatan, Nusa Tenggara Timur
- Klaster 1 : DI Yogyakarta, Jambi, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kepulauan Riau, Nusa Tenggara Barat, Sumatera Selatan, Sulawesi Tengah, Riau, Maluku Utara, Maluku, Papua Barat
- Klaster 2 : Aceh, Bali, Banten, Bangka Belitung, Bengkulu, DKI Jakarta, Jawa Timur, Kalimantan Utara, Sumatera Barat, Sulawesi Utara, Sumatera Utara, Lampung, Papua, Sulawesi Barat, Gorontalo

6. Menentukan Karakteristik dari Klaster

Klaster	Statistika Deskriptif	X1=Kumulatif total terinfeksi COVID-19 (jiwa)	X2=Kumulatif total sembuh dari COVID-19 (jiwa)	X5=Kumulatif total meninggal akibat terinfeksi COVID-19 (Jiwa)	X7=Recovery Index
Klaster 0	Minimum	584	406	7	0,63
	Maximum	30.043	22.980	1.579	0,86
	Average	16.145	12.131	532	0,73
Klaster 1	Minimum	668	368	9	0,38
	Maximum	11.503	9.984	451	0,90
	Average	4.809,071	3.646,143	160,929	0,75
Klaster 2	Minimum	499	430	5	0,54
	Maximum	93.356	77.969	3.529	0,94
	Average	13.980,667	11.326,067	501,2	0,776

7. Menentukan Saran Kinerja Pemerintah Berdasarkan Klaster

Klaster 0 :

Rata-rata total kumulatif terinfeksi paling tinggi, rata-rata total kumulatif kematian paling tinggi dan indeks kesembuhan paling rendah, maka **kebijakan yang diambil haruslah ketat** di berbagai aspek.

Walaupun rata-rata total kumulatif sembuh tinggi, namun Index Recovery-nya masih rendah. Artinya, tetap perlu kebijakan yang ketat untuk menghindari risiko pasien terinfeksi tidak sembuh.

Klaster 1 :

Rata-rata total kumulatif kesembuhan COVID-19 paling rendah, laju kesembuhan sedang, tingkat kematian tinggi dan minimum total kumulatif meninggal paling tinggi maka kebijakan yang diambil fokus terhadap **kebijakan yang berhubungan dengan kesehatan**. Hal ini dilakukan untuk meningkatkan jumlah kesembuhan dan menekan tingkat kematian akibat COVID-19.

Klaster 2 :

Dapat dilihat bahwa rata-rata tingkat recovery klaster ini paling tinggi dibandingkan klaster lain dan total kumulatif terinfeksi COVID-19 dalam tingkat sedang. Maka kebijakan yang diberikan dapat berupa **rekomen-dasi-rekomen-dasi** yang tidak terlalu ketat seperti klaster 0.



Terimakasih

“Data are just summaries of thousands of stories - tell a few of those stories to make the data meaningful.”

-Chip and Dan Heath

