



Application Correlation and Spectral Clustering

By : Fathimah Fadhilah Khoirunnisa

Tools : Python, Pandas, Scikit-Learn, Streamlit







Project Description

To analyze correlations among applicant data variables and perform clustering using Spectral Clustering.

Menu :

- Upload Data
 - Clustering Process
 - Cluster Analysis
 - Cluster Averages
 - Analysis of Prakerja Program Impact on Unemployment
 - Manual Data Input
- 
- 

Upload Data

Upload Data

Proses Analisis

Analisis Kluster

Rata-Rata Kluster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Upload Data

Silakan upload file untuk mulai proses.

Upload file data

Drag and drop file here
Limit 200MB per file • XLSX

Browse files

Data Pengangguran_Prakerja.xlsx 26.4KB

File berhasil diupload!

Tampilkan berapa data?

5

Jumlah data: 135

	No	Tahun	Wilayah	TPAK	Jumlah angkatan Kerja	Jumlah pengangguran terbuka	TPT	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Akasi
0	1	2020	Kabupaten Bogor	62.65	2733670	390731	14.29	262345	128386	67522	72817	75799	119828	39977
1	2	2020	Kabupaten Sukabumi	61.56	1146833	110098	9.6	76168	33930	27546	30293	29303	14473	5786
2	3	2020	Kabupaten Cianjur	69.26	1185595	131016	11.05	95492	35324	42686	30918	24505	24631	3342
3	4	2020	Kabupaten Bandung	62.2	1717376	147402	8.58	101373	46029	20621	39856	42656	28089	1618
4	5	2020	Kabupaten Garut	61.11	1156180	103607	8.95	74341	29266	27867	26971	33567	10192	2427

The file has been uploaded and is ready for analysis. You can also choose how many rows of data to display

Clustering Process

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Proses Analisis Data

Tabel Normalisasi (Min-Max)

	TPT	Jumlah pengangguran terbuka	Aktif	TPAK	Jumlah angkatan Kerja	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Akademik/Universitas	
0	1		1	0.6585	0.2831	0.9435	1	1	0.6369	1	1	1	0.399
1	0.6327		0.2742	0.204	0.2384	0.3955	0.2852	0.2561	0.259	0.4125	0.3808	0.1085	0.0567
2	0.7463		0.3283	0.1986	0.5539	0.4089	0.3594	0.2671	0.4022	0.4212	0.317	0.1944	0.0322
3	0.5529		0.3707	0.4885	0.2646	0.5925	0.382	0.3514	0.1936	0.5446	0.5587	0.2237	0.015
4	0.5818		0.2574	0.2616	0.22	0.3987	0.2782	0.2193	0.2621	0.3666	0.4376	0.0722	0.0231
5	0.4385		0.1635	0.183	0.5653	0.3257	0.1695	0.1564	0.2053	0.1757	0.1695	0.1076	0.0436
6	0.3242		0.0888	0.1183	0.642	0.234	0.0887	0.0948	0.1135	0.094	0.066	0.0728	0.0199
7	0.7596		0.1408	0.1122	0.2556	0.1796	0.1381	0.1516	0.1846	0.1511	0.1197	0.0882	0.0392
8	0.7831		0.3054	0.2376	0.3298	0.3656	0.3135	0.293	0.193	0.309	0.3517	0.3295	0.0164
9	0.3383		0.0898	0.1043	0.5703	0.2289	0.087	0.1012	0.0672	0.1013	0.1052	0.0734	0.0335

Normalisasi digunakan untuk menyamakan skala data antar fitur agar tidak ada fitur yang mendominasi. Di sini digunakan metode Min-Max Scaling untuk mengubah data ke rentang 0 hingga 1.

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Affinity Matrix (RBF Kernel)

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
0	1	0.018	0.0254	0.055	0.0154	0.0038	0.0011	0.003	0.0246	0.0012	0.0035	0.0068	0.0061	0.0019	0.0334	0.1616	0.0099	0.0003	0.0059	0.0006	0.0471	0.0006	0
1	0.018	1	0.8575	0.8002	0.9844	0.7488	0.5407	0.7675	0.9026	0.5809	0.7263	0.8146	0.8567	0.688	0.8768	0.437	0.8878	0.3123	0.7709	0.4851	0.6153	0.4968	0
2	0.0254	0.8575	1	0.6931	0.8117	0.7364	0.5212	0.6589	0.8743	0.527	0.718	0.8427	0.8016	0.5565	0.8695	0.4563	0.7508	0.3253	0.6511	0.332	0.5573	0.4084	0
3	0.055	0.8002	0.6931	1	0.8131	0.4985	0.3115	0.4174	0.7456	0.3366	0.4347	0.5553	0.5651	0.3513	0.7574	0.5672	0.6076	0.1526	0.4989	0.2115	0.7087	0.2182	0
4	0.0154	0.9844	0.8117	0.8131	1	0.7489	0.5453	0.7445	0.8651	0.5867	0.7036	0.7877	0.8221	0.6686	0.8229	0.4085	0.8725	0.3141	0.7487	0.4787	0.6282	0.4901	0
5	0.0038	0.7488	0.7364	0.4985	0.7489	1	0.9277	0.7939	0.7122	0.9315	0.9205	0.9551	0.9113	0.7423	0.6632	0.2484	0.6981	0.7125	0.6984	0.4989	0.4272	0.6698	0
6	0.0011	0.5407	0.5212	0.3115	0.5453	0.9277	1	0.6963	0.5085	0.9905	0.851	0.8276	0.7895	0.6832	0.4626	0.1362	0.5224	0.8942	0.554	0.4786	0.2693	0.6993	0
7	0.003	0.7675	0.6589	0.4174	0.7445	0.7939	0.6963	1	0.762	0.7383	0.8931	0.8373	0.8873	0.9792	0.6811	0.2228	0.9021	0.49	0.8866	0.8375	0.396	0.8768	0
8	0.0246	0.9026	0.8743	0.7456	0.8651	0.7122	0.5085	0.762	1	0.5421	0.7692	0.8377	0.8486	0.6921	0.9772	0.5558	0.8976	0.298	0.8202	0.4744	0.6792	0.5015	0
9	0.0012	0.5809	0.527	0.3366	0.5867	0.9315	0.9905	0.7383	0.5421	1	0.8747	0.8416	0.8252	0.7347	0.4957	0.1499	0.5672	0.8626	0.5998	0.533	0.2953	0.74	0

Matriks ini menunjukkan tingkat kedekatan atau kemiripan antar data berdasarkan fungsi kernel (di sini: RBF kernel). Nilai mendekati 1 artinya dua data sangat mirip.

Degree Matrix

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
0	4.2543	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	16.7235	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Laplacian Matrix

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	3.2543	-0.018	-0.0254	-0.055	-0.0154	-0.0038	-0.0011	-0.003	-0.0246	-0.0012	-0.0035	-0.0068	-0.0061	-0.0019	-0.0334	-0.1616	-0.0099	-0.0003	-0.0059	-0.0006	-0.0471
1	-0.018	85.7235	-0.8575	-0.8002	-0.9844	-0.7488	-0.5407	-0.7675	-0.9026	-0.5809	-0.7263	-0.8146	-0.8567	-0.688	-0.8768	-0.437	-0.8878	-0.3123	-0.7709	-0.4851	-0.6153
2	-0.0254	-0.8575	78.8521	-0.6931	-0.8117	-0.7364	-0.5212	-0.6589	-0.8743	-0.527	-0.718	-0.8427	-0.8016	-0.5565	-0.8695	-0.4563	-0.7508	-0.3253	-0.6511	-0.332	-0.5573
3	-0.055	-0.8002	-0.6931	61.7895	-0.8131	-0.4985	-0.3115	-0.4174	-0.7456	-0.3366	-0.4347	-0.5553	-0.5651	-0.3513	-0.7574	-0.5672	-0.6076	-0.1526	-0.4989	-0.2115	-0.7087
4	-0.0154	-0.9844	-0.8117	-0.8131	84.2743	-0.7489	-0.5453	-0.7445	-0.8651	-0.5867	-0.7036	-0.7877	-0.8221	-0.6686	-0.8229	-0.4085	-0.8725	-0.3141	-0.7487	-0.4787	-0.6282
5	-0.0038	-0.7488	-0.7364	-0.4985	-0.7489	93.9808	-0.9277	-0.7939	-0.7122	-0.9315	-0.9205	-0.9551	-0.9113	-0.7423	-0.6632	-0.2484	-0.6981	-0.7125	-0.6984	-0.4989	-0.4272
6	-0.0011	-0.5407	-0.5212	-0.3115	-0.5453	-0.9277	84.0787	-0.6963	-0.5085	-0.9905	-0.851	-0.8276	-0.7895	-0.6832	-0.4626	-0.1362	-0.5224	-0.8942	-0.554	-0.4786	-0.2693
7	-0.003	-0.7675	-0.6589	-0.4174	-0.7445	-0.7939	-0.6963	86.8181	-0.762	-0.7383	-0.8931	-0.8373	-0.8873	-0.9792	-0.6811	-0.2228	-0.9021	-0.49	-0.8866	-0.8375	-0.396
8	-0.0246	-0.9026	-0.8743	-0.7456	-0.8651	-0.7122	-0.5085	-0.762	82.618	-0.5421	-0.7692	-0.8377	-0.8486	-0.6921	-0.9772	-0.5558	-0.8976	-0.298	-0.8202	-0.4744	-0.6792
9	-0.0012	-0.5809	-0.527	-0.3366	-0.5867	-0.9315	-0.9905	-0.7383	-0.5421	87.1177	-0.8747	-0.8416	-0.8252	-0.7347	-0.4957	-0.1499	-0.5672	-0.8626	-0.5998	-0.533	-0.2953

Laplacian Matrix diperoleh dari Degree Matrix dikurangi Affinity Matrix. Matriks ini penting dalam algoritma Spectral Clustering karena digunakan untuk menemukan struktur kluster alami dalam data.

Eigenvectors (2D Embedding)

	Eigenvector 2	Eigenvector 3	
0		-0.5277	0.2453

Cluster Analysis

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Data Klasterisasi Lengkap

	Tahun	Wilayah	Cluster	TPAK	Jumlah angkatan Kerja	Jumlah pengangguran terbuka	TPT	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Ak
0	2020	Kabupaten Bogor	0	62.65	2733670	390731	14.29	262345	128386	67522	72817	75799	119828	
1	2020	Kabupaten Sukabumi	0	61.56	1146833	110098	9.6	76168	33930	27546	30293	29303	14473	
2	2020	Kabupaten Cianjur	0	69.26	1185595	131016	11.05	95492	35324	42686	30918	24505	24631	
3	2020	Kabupaten Bandung	0	62.2	1717376	147402	8.58	101373	46029	20621	39856	42656	28089	
4	2020	Kabupaten Garut	0	61.11	1156180	103607	8.95	74341	29266	27867	26971	33567	10192	
5	2020	Kabupaten Tasikmalaya	1	69.54	944706	67290	7.12	46022	21268	21857	13148	13434	14370	
6	2020	Kabupaten Ciamis	1	71.41	679192	38425	5.66	24972	13453	12144	7237	5661	10258	
7	2020	Kabupaten Kuningan	1	61.98	521677	58513	11.22	37853	20660	19666	11372	9690	12077	
8	2020	Kabupaten Cirebon	0	63.79	1060362	122145	11.52	83532	38613	20555	22801	27117	40591	
9	2020	Kabupaten Majalengka	1	69.66	664374	38798	5.84	24535	14263	7250	7763	8604	10327	

Pilih Tahun

2020

Pilih Klaster

0.0 x 1.0 x

Terapkan Filter

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Pilih Tahun

2020

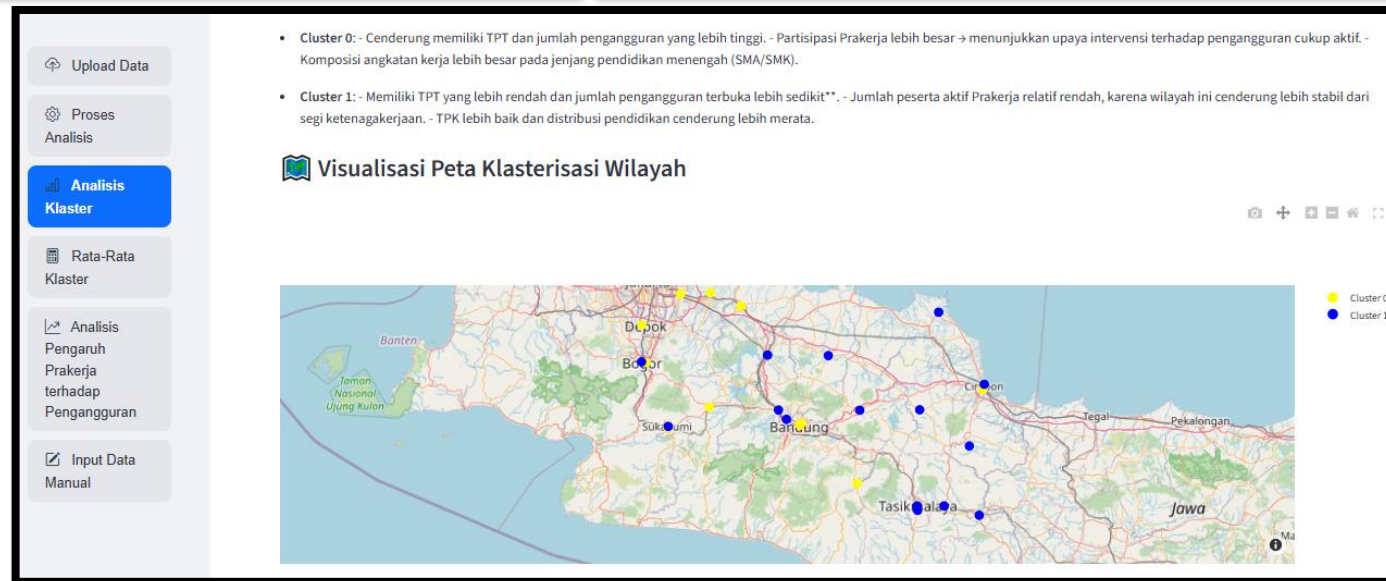
Pilih Klaster

0.0 x 1.0 x

Terapkan Filter

Data Klaster Terfilter

	Tahun	Wilayah	Cluster	TPAK	Jumlah angkatan Kerja	Jumlah pengangguran terbuka	TPT	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Ak
0	2020	Kabupaten Bogor	0	62.65	2733670	390731	14.29	262345	128386	67522	72817	75799	119828	
1	2020	Kabupaten Sukabumi	0	61.56	1146833	110098	9.6	76168	33930	27546	30293	29303	14473	
2	2020	Kabupaten Cianjur	0	69.26	1185595	131016	11.05	95492	35324	42686	30918	24505	24631	
3	2020	Kabupaten Bandung	0	62.2	1717376	147402	8.58	101373	46029	20621	39856	42656	28089	
4	2020	Kabupaten Garut	0	61.11	1156180	103607	8.95	74341	29266	27867	26971	33567	10192	
5	2020	Kabupaten Tasikmalaya	1	69.54	944706	67290	7.12	46022	21268	21857	13148	13434	14370	
6	2020	Kabupaten Ciamis	1	71.41	679192	38425	5.66	24972	13453	12144	7237	5661	10258	
7	2020	Kabupaten Kuningan	1	61.98	521677	58513	11.22	37853	20660	19666	11372	9690	12077	
8	2020	Kabupaten Cirebon	0	63.79	1060362	122145	11.52	83532	38613	20555	22801	27117	40591	
9	2020	Kabupaten Majalengka	1	69.66	664374	38798	5.84	24535	14263	7250	7763	8604	10327	



Cluster Averages



Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

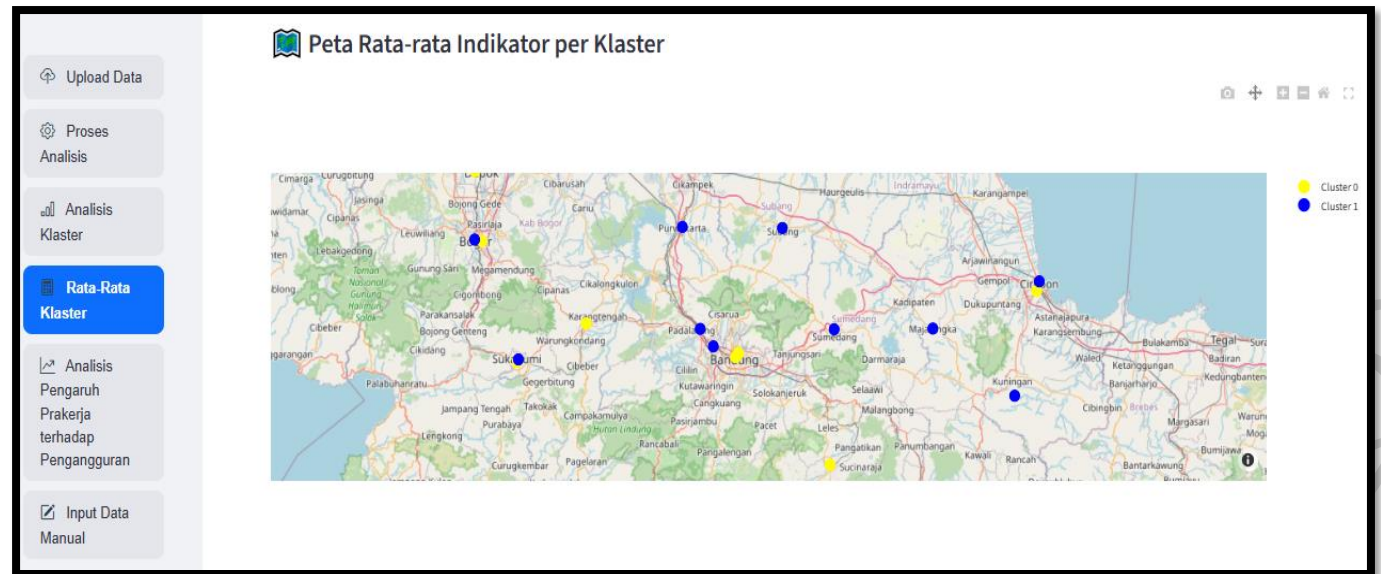
Input Data Manual

Rata-rata Indikator per Tahun, Wilayah, dan Klaster

	Wilayah	Cluster	TPT	Jumlah pengangguran terbuka	Aktif	TPAK	Jumlah angkatan Kerja	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Ak
0	Kabupaten Bandung	0	7.352	133896.6	48584.8	65.056	1828104.2	91868.6	42028	23021.4	18738.8	34105.6	25214.2	
1	Kabupaten Bandung Barat	1	9.668	80428.6	28050	64.068	851799.4	51998	28430.6	24087.6	16919	24815	15675.4	
2	Kabupaten Bekasi	0	9.926	180099.2	38380.8	65.144	1802671.2	118781	61318.2	23728.6	27866	44718	65484.6	
3	Kabupaten Bogor	0	10.592	296274.8	76039.8	63.894	2803206.6	204436.6	91838.2	67531.8	34099.2	56623.6	74906.4	
4	Kabupaten Ciamis	1	4.272	28684.8	13713.2	68.822	670835.8	20011	8673.8	18124.2	16873	14058.8	12884.8	
5	Kabupaten Cianjur	0	8.496	107075.4	30904.6	70.432	1278162.2	73742.6	33292.8	34929.4	15183.6	15498.6	23172	
6	Kabupaten Cirebon	0	8.88	100190.2	29624.8	65.848	1141543.6	52086.8	35166.8	26122	19640.4	23838.4	28063.8	
7	Kabupaten Garut	0	7.904	102214.4	40615.6	66.83	1303424.8	70554.6	31659.8	29729.8	23565	28420	16307.2	
8	Kabupaten Indramayu	1	7.342	70129.6	22451.2	68.812	956280	45824.4	24305.2	22678.2	19192.2	18106.2	19867.4	
9	Kabupaten Karawang	0	10.042	119923.2	25498.2	64.622	1199159.6	80903	39020.2	28625	19180.4	32386.4	36703.8	

Rata-rata Indikator per Klaster

	Cluster	TPT	Jumlah pengangguran terbuka	Aktif	TPAK	Jumlah angkatan Kerja	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Ak
0	0	8.9542	137208.9818	37580.7091	65.7675	1513606.0909	90828.3273	45192.6909	29165.3091	26385.6909	33945.5091	35560.3455	
1	1	7.6453	41635.6	12235.7	66.6341	558027.5375	26538.0875	13441.5875	16158.4064	13879.6875	12981.175	13954.5625	



Analysis of Prakerja Program Impact on Unemployment



Manual Data Input

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Input Data Manual untuk Cluster

TPT

14.29

-

+

Jumlah pengangguran terbuka

390731.00

-

+

Aktif

88469.00

-

+

TPAK

62.65

-

+

Jumlah angkatan Kerja

2733670.00

-

+

laki-laki

262345.00

-

+

perempuan

128386.00

-

+

SD ke bawah

67522.00

-

+

SMP

72817.00

-

+

SMA

75799.00

-

+

SMK

119828.00

-

+

Diploma I/II/III/Akademi/Universitas

39977.00

-

+

Tahun

2020

-

+

Wilayah

Kabupaten Bogor

▼

Klaster

✓ Data yang Anda input masuk ke Cluster 0.

Upload Data

Proses Analisis

Analisis Klaster

Rata-Rata Klaster

Analisis Pengaruh Prakerja terhadap Pengangguran

Input Data Manual

Detail Data yang Dimasukkan:

	TPT	Jumlah pengangguran terbuka	Aktif	TPAK	Jumlah angkatan Kerja	laki-laki	perempuan	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Akademi/Universitas
0	14.29	390731	88469	62.65	2733670	262345	128386	67522	72817	75799	119828	39977



Tahun: 2020 | Wilayah: Kabupaten Bogor

Korelasi Antar Fitur

	Aktif	TPT	SD ke bawah	SMP	SMA	SMK	Diploma I/II/III/Akademi/Universitas
Aktif	1.000	-0.338	-0.013	-0.237	-0.279	0.434	-0.333
TPT	-0.338	1.000	-0.328	0.247	0.239	-0.491	-0.290
SD ke bawah	-0.013	-0.328	1.000	-0.234	0.166	-0.254	-0.213
SMP	-0.237	0.247	-0.234	1.000	-0.623	0.322	0.406
SMA	-0.279	0.239	0.166	-0.623	1.000	-0.445	-0.372
SMK	0.434	-0.491	-0.254	0.322	-0.445	1.000	0.061
Diploma I/II/III	-0.333	-0.290	-0.213	0.406	-0.372	0.061	1.000

Penjelasan Korelasi:

- Nilai mendekati 1 atau -1 artinya hubungan kuat.
- Nilai mendekati 0 artinya hubungan lemah



Data preprocessing, correlation analysis, and Spectral Clustering using Python, Pandas, Scikit-Learn, Streamlit. Three main applicant clusters identified. Insights can help understand patterns in the dataset and inform decisions.

Contact & Links

Email : fathimahfadhilah17@gmail.com

Github : <https://github.com/fathimahfadhilah12/Application-and-Analysis-Correlation-and-Spectral-Clustering>

Linkedin : <https://www.linkedin.com/in/fathimah-fadhilah-khoirunnisa-547521214/>

