

Pemetaan Prioritas Investasi terhadap Provinsi di Indonesia untuk Program Makan Bergizi Gratis dengan Pendekatan Klasterisasi Berbasis Graf

¹Stefano Budi ²Gary Chen ³Jason Alexander Wijaya ⁴Marcellino Wilson Rusli
 Dept. of Computer Science, Dept. of Computer Science, Dept. of Computer Science, Dept. of Computer Science,
 BINUS University, BINUS University, BINUS University, BINUS University,
 Semarang, Indonesia Semarang, Indonesia Semarang, Indonesia Semarang, Indonesia
stefano.budi@binus.ac.id gary.chen@binus.ac.id jason.wijaya013@binus.ac.id marcellino.rusli@binus.ac.id

ABSTRAK

Pada tahun 2024, program Makan Bergizi Gratis (MBG) yang diprakarsai oleh Bapak Presiden Prabowo mulai dilaksanakan. Program MBG memiliki banyak tantangan, seperti perbedaan kondisi pasokan pangan di berbagai provinsi yang mengakibatkan kesenjangan alokasi dana. Penelitian ini dilakukan untuk mengembangkan pemetaan prioritas investasi terhadap 38 provinsi di Indonesia untuk mengatasi perbedaan kondisi pangan tersebut. Dataset yang digunakan adalah data demografis, ekonomi, dan potensi pertanian. Model yang digunakan untuk melakukan pemetaan prioritas investasi adalah Spectral Clustering yang menggunakan algoritma Laplacian untuk mempertimbangkan kedekatan antar-provinsi secara graf. Kesimpulan yang dihasilkan oleh pola yang dideteksi model clustering adalah bahwa Indonesia bagian timur harus diinvestasikan secepat mungkin oleh karena krisis pangan dan potensi perekonomian yang bagus.
Kata kunci: Klasterisasi, Ketahanan Pangan, Kesenjangan Regional

ABSTRACT

In 2024, the Makan Bergizi Gratis (MBG) program which was founded by President Prabowo officially started. The MBG program had a lot of obstacles, such as the difference of food supply condition in various provinces, which could later cause problems relating to fund allocation. This study is therefore conducted to develop an investment priority mapping for 38 provinces In Indonesia to address the differences in food conditions. The dataset used in this study includes demographic, economic, and agricultural potential data. The model used to map the investment priority is Spectral Clustering which uses the Laplacian algorithm to also consider the distance between a province to its neighbour graphically. The conclusion drawn from the clustering pattern indicates that the eastern part of Indonesia needs immediate investment to address food insecurity and market failure.
Keywords: Clustering, Food Security, Regional Disparity

BAB I PENDAHULUAN

Program Makan Bergizi Gratis (MBG) menjadi program dengan prioritas tertinggi oleh Presiden Prabowo. Program MBG menjadi cara pemerintah mengatasi problematika kekurangan gizi yang dialami oleh anak-anak. Pemerintah menitikberatkan program MBG karena dapat nantinya meningkatkan pertumbuhan kemampuan kognitif anak (Qomarullah et al., 2025).

Pemerintah berulang kali menekankan bahwa program MBG bukanlah program bantuan sosial, namun menjadi langkah awal membentuk kualitas SDM dalam jangka panjang. Siklus gizi buruk, sebagaimana dikutip dari UNICEF, telah mengakibatkan 7 juta anak balita mengalami stunting, 2 juta anak kurus, dan 1 dari 4 remaja mengalami anemia. Dampak dari siklus gizi buruk adalah kebodohan, karena anak yang lapar akan susah berkonsentrasi dan menyerap ilmu, sedangkan kebodohan itu sendiri dapat mengakibatkan kemiskinan.

Program MBG memiliki banyak tantangan sosio-ekonomi. Mengenai kemiskinan, berdasarkan data BPS (2023), terdapat ketimpangan kondisi ekonomi di provinsi bagian timur dan di Jawa. Provinsi-provinsi di Indonesia bagian timur memiliki tingkat kemiskinan sebesar 26%, sedangkan provinsi-provinsi di Jawa memiliki tingkat kemiskinan sebesar 9-10%. Mengenai stunting, kemkes.go.id (2023) menyatakan bahwa NTT memiliki tingkat stunting tertinggi di Indonesia, yakni 35,3%, sedangkan Bali hanya berkisar di angka 8%.

Program MBG juga memiliki banyak tantangan geografis, yakni utamanya logistik dan Sumber Daya Alam (SDA) lokal. Disparitas biaya distribusi bahan pangan menjadi salah satu penyebab utama perbedaan harga pangan antar provinsi. Data Indeks Kemahalan Konstruksi (IKK) tahun 2023 dari BPS menunjukkan bahwa Provinsi Papua Pegunungan memiliki IKK sebesar 179,86, sementara Provinsi Jawa Tengah hanya 98,99. Ketimpangan IKK antara kedua provinsi tersebut menandakan biaya konstruksi dan logistic di Papua Pegunungan jauh lebih mahal dibandingkan di Jawa Tengah. Ketersediaan bahan pangan lokal seperti sayur-sayuran, ikan, dan telur sangat berlimpah di suatu provinsi, namun terbatas di provinsi lain. Ketimpangan ketersediaan bahan pangan dapat diamati pada Jawa Timur yang menjadi lumbung pangan nasional melalui produksi ikan dan beras, sedangkan NTT seringkali mengalami kekurangan dalam komoditas tertentu dan bergantung pada pasokan antar provinsi.

Untuk mengatasi problematika tersebut, peneliti berusaha mengembangkan pemetaan prioritas investasi untuk seluruh provinsi di Indonesia. Peneliti mempertimbangkan beberapa faktor yang berpengaruh, seperti faktor geografis, demografis, ekonomi, pertanian, dan cuaca. Peneliti memilih model Graph-Based Clustering yang dapat turut mempertimbangkan faktor kedekatan antar provinsi

untuk menarik kesimpulan. Peneliti juga menghadapi tantangan dan keterbatasan, seperti

- 1) Data hanya berasal dari tahun 2022 dan 2023 karena keterbatasan data.
- 2) Data mengenai provinsi baru di Papua (Papua Barat Daya, Papua Tengah, Papua Selatan, Papua Pegunungan) tidak sepenuhnya akurat karena baru didirikan di awal tahun 2023.

BAB II TINJAUAN PUSTAKA

A. Makan Bergizi Gratis demi Keberlanjutan Pendidikan Indonesia

Program MBG ditujukan untuk meningkatkan kualitas SDM Indonesia. Program ini memiliki target awal untuk mencapai 3 juta penerima (Kemenko PMK, 2025). Implementasi program ini dirancang untuk memenuhi 30-35% kebutuhan gizi harian anak sekolah, sesuai standar yang ditetapkan Badan Gizi Nasional. Menurut United Nations' World Food Programme (2021), negara Afrika berhasil membuktikan bahwa program yang menyerupai MBG dapat memperlancar usaha petani lokal dan mendorong ekonomi pedesaan. Badan Gizi Nasional (2025) juga menyatakan bahwa program MBG dirancang untuk juga melibatkan Usaha Mikro, Kecil, dan Menengah (UMKM) dalam penyediaan pangan. Strategi ini berhasil mendorong perekonomian sekaligus pemenuhan gizi rakyat.

B. Potensi Produksi Pertanian di Indonesia

Sektor pertanian berkontribusi besar terhadap perekonomian nasional, yakni sekitar 12,53% dari total Produk Domestik Bruto (PDB) Indonesia di tahun 2023 (World Bank, 2023). Pentingnya sektor pertanian didukung oleh potensi Indonesia sebagai produsen berbagai produk pertanian tropis melalui banyaknya lahan subur yang dimiliki (Indonesia Investments, 2025). Besarnya produksi pertanian berhasil membawa Indonesia mencapai swasembada pangan pada level nasional di tahun 2025, walaupun dalam penerapannya, distribusi pencapaian ini tidak merata ke seluruh wilayah Indonesia. BPS (2025) menunjukkan bahwa terdapat tren penurunan luas lahan pertanian tanpa peningkatan efektivitas produksi. Divergensi ini menimbulkan kekhawatiran mengenai keberlanjutan capaian swasembada pangan nasional. Sektor pertanian Indonesia memiliki banyak tantangan. *Pertama*, keterbatasan akses modal investasi di daerah yang memiliki potensi pertanian tinggi. *Kedua*, kesenjangan teknologi antar daerah. *Ketiga*, infrastruktur distribusi dan logistic yang belum optimal. Ketiga faktor ini menyebabkan rendahnya produktivitas pertanian di berbagai daerah di Indonesia.

C. Peran Investasi dalam Pengembangan Sektor Pertanian

Masru'ah dan Soejoto (2013) membuktikan melalui penelitiannya, bahwa investasi ke dalam sektor pertanian memiliki hubungan linear terhadap produktivitas pertanian. Abdullah dan Rahman (2017) turut mendukung temuan ini melalui penelitiannya yang menunjukkan bahwa setiap 1% dari peningkatan investasi terhadap sektor pertanian dapat meningkatkan hasil pertanian sebesar 0,73%. Statizsta (2023) menunjukkan bahwa *Foreign Direct Investment* (FDI) terhadap sektor pertanian, kehutanan, dan perikanan Indonesia pada tahun 2023 mencapai 2 miliar USD. Angka ini menunjukkan kepercayaan investor asing

terhadap sektor pertanian di Indonesia, walaupun distribusi investasi masih terpusat pada wilayah-wilayah tertentu. Siregar dan Wahyuni (2019) melalui studinya menekankan bahwa alokasi investasi terhadap daerah-daerah dengan produktivitas pertanian rendah dapat menjadi katalis peningkatan hasil produksi di daerah tersebut. Ketiga hasil penelitian membuktikan bahwa strategi menghimpun investasi ke dalam sektor pertanian sejalan dengan konsep pembangunan berkelanjutan untuk pemeratakan pembangunan ekonomi di seluruh wilayah Indonesia. Dalam konteks program MBG, banyaknya investasi sektor pertanian ke dalam provinsi yang memiliki potensi dapat nantinya menurunkan harga pangan, sehingga program MBG dapat berjalan lebih merata.

BAB III TAHAP PELAKSANAAN

A. Pengumpulan Data (*Data-gathering*)

Data yang diambil adalah data dari tahun 2022 (34 provinsi) dan 2023 (38 provinsi). Detail dari dataset tersebut adalah:

Kategori	No.	Nama Fitur	Nama kolom	Satuan
Demografi	1	Luas wilayah	<i>area of land</i>	km ²
	2	Jumlah populasi	<i>population</i>	orang
	3	Jumlah petani	<i>farmer amount</i>	orang
	4	Kesejahteraan petani	<i>farmers tottrade</i>	(index)
	5	Jumlah kelompok petani	<i>poktan_amount</i>	kelompok
	6	Jumlah gabungan kelompok petani	<i>gapoktan_amount</i>	kelompok
Ekonomi	1	Unit investasi asing	<i>fdi units</i>	unit
	2	Jumlah investasi asing	<i>fdi_rupiahs</i>	rupiah
	3	Unit investasi lokal	<i>ldi units</i>	unit
	4	Jumlah investasi lokal	<i>ldi_rupiahs</i>	rupiah
	5	Produk Domestik Regional Bruto (PDRB)	<i>pdrb</i>	rupiah/orang
Potensi Pertanian	1	Luas sawah	<i>wetlands area</i>	Ha (hectare)
	2	Produksi daging ayam ras pedaging	<i>broiler_production</i>	Ton
	3	Produksi telur ayam ras petelur	<i>layer_production</i>	Ton
	4	Jumlah ayam pedaging	<i>broiler_amount</i>	ekor
	5	Jumlah ayam petelur	<i>layer_amount</i>	ekor
	6	Produksi padi	<i>paddy_production</i>	Ton
	7	Produksi kedelai	<i>soybean_production</i>	Ton
	8	Imbal hasil padi	<i>paddy_yield</i>	Qu/Ha
	9	Imbal hasil kedelai	<i>soybean_yield</i>	Qu/Ha
Iklim	1	Temperatur	<i>temperature</i>	°C
	2	Kelembapan	<i>humidity</i>	%
	3	Curah hujan	<i>rainfall</i>	mm
Harga komoditas pangan	1	Harga bawang merah	<i>avg_Bawang Merah price</i>	rupiah
	2	Harga bawang putih	<i>avg_Bawang Putih Bonggol price</i>	rupiah
	3	Harga beras medium	<i>avg_Beras Medium price</i>	rupiah
	4	Harga beras premium	<i>avg_Beras Premium price</i>	rupiah

	5	Harga cabai merah keriting	<i>avg_Cabai Merah Keriting_price</i>	rupiah
	6	Harga cabai rawit merah	<i>avg_Cabai Rawit Merah_price</i>	rupiah
	7	Harga daging ayam ras pedaging	<i>avg_Daging Ayam Ras_price</i>	rupiah
	8	Harga telur ayam ras petelur	<i>avg_Telur Ayam Ras_price</i>	rupiah
	9	Harga daging sapi murni	<i>avg_Daging Sapi Murni_price</i>	rupiah
	10	Harga gula	<i>avg_Gula Konsumsi_price</i>	rupiah
	11	Harga minyak goreng curah	<i>avg_Minyak Goreng Curah_price</i>	rupiah
	12	Harga minyak goreng kemasan sederhana	<i>avg_Minyak Goreng Kemasan Sederhana_price</i>	rupiah
	13	Harga tepung terigu curah	<i>avg_Tepung Terigu (Curah)_price</i>	rupiah

B. Penggantian Data Kosong (*imputation*)

Terdapat beberapa kekosongan dalam dataset tahun 2022 dan 2023. Model yang dipilih (*clustering*) tidak dapat digunakan pada dataset yang mengandung NaN (data kosong), sehingga diperlukan adanya penggantian data kosong atau *imputation*. Teknik yang digunakan utamanya adalah penggantian dengan mean untuk data yang memiliki distribusi normal, dan median untuk data yang tidak memiliki distribusi normal (*skewed*). Berikut data-data yang kosong pada tahun 2022 dan cara mengatasinya.

Provinsi	Fitur yang kosong	Nama kolom yang kosong	Cara mengatasi
DKI Jakarta	Produksi daging ayam ras pedaging	<i>broiler_production</i>	Diganti dengan mean dari Banten, D.I. Yogyakarta, dan Bali.
	Produksi telur ayam ras petelur	<i>layer_production</i>	
	Produksi kedelai	<i>soybean_production</i>	
	Jumlah ayam ras pedaging	<i>broiler_amount</i>	
	Jumlah ayam ras petelur	<i>layer_amount</i>	
	Imbal hasil kedelai	<i>soybean_yield</i>	
Kalimantan Utara	Produksi kedelai	<i>soybean_production</i>	Diganti dengan rata-rata dari Kalimantan Barat,
	Imbal hasil kedelai	<i>soybean_yield</i>	

	Temperatur	<i>temperature</i>	Kalimantan Timur, Kalimantan Selatan, Kalimantan Tengah
	Kelembapan	<i>humidity</i>	
	Curah hujan	<i>rainfall</i>	
Papua Barat Daya	Semua kolom kosong	-	Seluruh provinsi dihapus seluruhnya
Papua Selatan	Semua kolom kosong	-	
Papua Tengah	Semua kolom kosong	-	
Papua Pegunungan	Semua kolom kosong	-	

Berikut beberapa kekosongan dalam dataset tahun 2023 dan cara mengatasinya.

Provinsi	Fitur yang kosong	Nama kolom yang kosong	Cara mengatasi
DKI Jakarta	Produksi daging ayam ras pedaging	<i>broiler_production</i>	Diganti dengan <i>median</i> dari Banten, D.I. Yogyakarta, Bali, Jawa Barat.
	Produksi telur ayam ras petelur	<i>layer_production</i>	
	Produksi kedelai	<i>soybean_production</i>	
	Jumlah ayam ras pedaging	<i>broiler_amount</i>	
	Jumlah ayam ras petelur	<i>layer_amount</i>	
	Imbal hasil kedelai	<i>soybean_yield</i>	
Bangka Belitung	Produksi kedelai	<i>soybean_production</i>	Diganti dengan <i>median</i> dari semua provinsi lain di Sumatera
	Imbal hasil kedelai	<i>soybean_yield</i>	
Papua Barat Daya, Papua Barat	Jumlah petani	<i>farmer_amount</i>	Menggunakan rasio <i>farmer_amount</i> dibagi dengan <i>population</i> dari Papua Barat (lama), kemudian mengaplikasikan rasio tersebut dengan <i>population</i> provinsi yang hendak di-impute.
	Kesejahteraan petani	<i>farmers_totrade</i>	Copy dari <i>farmers_totrade</i>

			yang dimiliki Papua Barat (lama).
	Jumlah kelompok petani	<i>poktan_amount</i>	Menggunakan rasio $\frac{farmer_amount}{poktan_amount}$, milik Papua Barat (lama), kemudian mengaplikasikan rasio tersebut ke <i>farmer_amount</i> provinsi yang hendak di-impute.
	Jumlah gabungan kelompok petani	<i>gapoktan_amount</i>	Menggunakan rasio $\frac{farmer_amount}{gapoktan_amount}$ milik Papua Barat (lama), kemudian mengaplikasikan rasio tersebut ke <i>farmer_amount</i> provinsi yang hendak di-impute.
	Produksi kedelai	<i>soybean_production</i>	Menggunakan rasio $\frac{soybean_production}{wetland_area}$ milik Papua Barat (lama), kemudian mengaplikasikan rasio tersebut ke <i>wetland_area</i> provinsi yang hendak di-impute.
	Imbal hasil kedelai	<i>soybean_yield</i>	Copy dari <i>soybean_yield</i> yang dimiliki Papua Barat (lama).
Papua Tengah, Papua Selatan, Papua, Papua Pegunungan	Jumlah petani	<i>farmer_amount</i>	Menggunakan rasio $\frac{farmer_amount}{population}$ dibagi dengan <i>population</i> dari Papua (lama), kemudian mengaplikasikan rasio tersebut dengan <i>population</i> provinsi yang hendak di-impute.

	Kesejahteraan petani	<i>farmers_totrade</i>	Copy dari <i>farmers_totrade</i> yang dimiliki Papua (lama)
	Jumlah kelompok petani	<i>poktan_amount</i>	Menggunakan rasio $\text{farmer_amount} / \text{poktan_amount}$, milik Papua (lama), kemudian mengaplikasikan rasio tersebut ke <i>farmer_amount</i> provinsi yang hendak di-impute.
	Jumlah gabungan kelompok petani	<i>gapoktan_amount</i>	Menggunakan rasio $\text{farmer_amount} / \text{gapoktan_amount}$ milik Papua (lama), kemudian mengaplikasikan rasio tersebut ke <i>farmer_amount</i> provinsi yang hendak di-impute.
	Produksi kedelai	<i>soybean_production</i>	Menggunakan rasio $\text{soybean_production} / \text{wetland_area}$, milik Papua (lama) kemudian mengaplikasikan rasio tersebut ke <i>wetland_area</i> provinsi yang hendak di-impute.
	Imbal hasil kedelai	<i>soybean_yield</i>	Copy dari <i>soybean_yield</i> yang dimiliki Papua (lama).

C. Rekayasa Fitur Baru (*Feature Engineering*)

Fitur baru yang ditambahkan adalah sebagai berikut:

a) Imbal hasil daging ayam ras pedaging (*broiler_yield*)

Mencerminkan seberapa produktif ayam ras pedaging di setiap provinsi.

$$broiler_{yield} = \frac{broiler_{production}}{broiler_{amount}}$$

b) Imbal hasil telur ayam ras petelur (*layer_yield*)

Mencerminkan seberapa produktif ayam ras petelur di setiap provinsi. Merupakan hasil pembagian antara produksi telur ayam ras petelur dengan jumlah ayam ras petelur

$$layer_{yield} = \frac{layer_{yield}}{layer_{yield}}$$

c) Selisih imbal hasil daging ayam ras pedaging (*broiler_yield_gap*)

Mencerminkan seberapa jauh selisih produktivitas sebuah provinsi terhadap provinsi yang paling produktif. Menghasilkan angka potensi yang masih belum dapat dicapai.

$$broiler_{yield_gap} = 1 - \frac{broiler_{yield}}{broiler_{yield_max}}$$

d) Selisih imbal hasil telur ayam ras petelur (*layer_yield_gap*)

Mencerminkan seberapa jauh selisih produktivitas sebuah provinsi terhadap provinsi yang paling produktif. Menghasilkan angka potensi yang masih belum dapat dicapai.

$$layer_{yield_gap} = 1 - \frac{layer_{yield}}{layer_{yield_max}}$$

e) Tingkat keteraturan petani (*farmer_organization_rate*)

Mencerminkan seberapa teratur petani, melalui seberapa banyak organisasi petani yang berdiri di tengah banyaknya jumlah petani.

$$farmer_{organization_rate} = \frac{poktan_{amount} + gapoktan_{amount}}{farmer_{amount}}$$

f) Produk Domestik Regional Bruto perkapita (*pdrb_per_capita*)

Mencerminkan total produksi barang dan jasa (*goods and services*) per rakyat di sebuah provinsi.

$$pdrb_{per_capita} = \frac{pdrb}{population}$$

g) Dominasi investasi lokal (*ldi_dominance*)

Mencerminkan seberapa besar total investasi yang diterima di sebuah provinsi adalah investasi lokal dan bukan asing.

$$ldi_{dominance} = \frac{ldi_{rupiahs}}{(fdi_{rupiahs} + ldi_{rupiahs})}$$

h) Tingkat ketahanan pangan (*msg_basket_ssr*)

Mencerminkan angka ketahanan pangan, karena membandingkan konsumsi dan produksi sebuah komoditas di sebuah provinsi.

$$msg_basket_ssr = \frac{1}{4} \times \sum_{i \in \{paddy, soy, layer, broiler\}} \frac{produksi_i}{populasi_{provinsi} \times konsumsi_i}$$

Dengan penjelasan konsumsi setiap komoditas sebagai berikut:

Tahun	Komoditas	Konsumsi (kg/kapita/tahun)
2022	Padi	93,791
	Kedelai	8,388
2023	Daging ayam ras pedaging	17,459
	Telur ayam ras petelur	6,969

i) Rasio curah hujan setiap satuan luas (*rainfall_per_area*)

Mencerminkan seberapa basah sebuah provinsi.

$$rainfall_per_area = \frac{rainfall}{area_of_land}$$

j) Rasio jumlah petani terhadap total populasi (*farmer_over_population*)

Mencerminkan seberapa banyak dari total populasi di sebuah provinsi adalah seorang petani.

$$farmer_over_population = \frac{farmer_amount}{population}$$

k) Kepadatan penduduk (*population_density*)

Mencerminkan kepadatan penduduk (berapa penduduk yang ada tiap satuan luas)

$$population_density = \frac{population}{area_of_land}$$

l) Selisih curah hujan optimal (*rainfall_optimal_gap*)

Mencerminkan seberapa jauh curah hujan sebuah provinsi dari curah hujan optimal. Angka negatif berarti tanaman cenderung tergenang, sedangkan angka positif berarti tanaman cenderung kering.

$$rainfall_{optimal_gap} = \frac{(200 - rainfall)}{200}$$

m) Selisih suhu optimal (*temperature_optimal_gap*)

Mencerminkan seberapa jauh temperatur sebuah provinsi dari curah hujan optimal. Angka negatif berarti tanaman cenderung kering (terlalu panas), sedangkan angka positif berarti tanaman cenderung dingin.

$$temperature_{optimal_gap} = \frac{(27.5 - temperature)}{27.5}$$

n) Rata-rata harga komoditas (*commodity_basket_index*)

Mencerminkan rata-rata harga komoditas dari 13 komoditas utama di Indonesia, sesuai dengan fitur yang dipilih dari dataset utuh.

$$commodity_{basket_index} = \sum_{j=1}^n w_j \times NKomoditas_j$$

$$NKomoditas_j = \frac{Komoditas_j - \min(Komoditas_j)}{\max(Komoditas_j) - \min(Komoditas_j)}$$

o) Rasio harga terhadap tingkat ketahanan pangan (*price_to_ssr*)

Mencerminkan harga yang harus dibayar untuk kemandirian pangan sebuah provinsi.

$$price_to_ssr = \frac{commodity_{basket_index}}{msg_basket_ssr}$$

D. Normalisasi Data

Variabel-variabel yang sudah dibentuk ditransformasi sesuai dengan algoritma yang paling cocok untuk persebaran data variable tersebut. Transformasi logaritmik (\log_{1p}) digunakan untuk mengatasi persebaran yang kurang merata (*skewed*), seperti *population*, *pdrb_per_capita*, *price_to_ssr*. Transformasi StandardScaler digunakan untuk mengatasi persebaran yang mendekati distribusi

normal dan tidak memiliki banyak *outlier*, seperti *population_density*, *area_of_land*, *wetlands_area*, *farmer_over_population*, *temperature*, *rainfall_per_area*, *rainfall_optimal_gap*, *temperature_optimal_gap*, dan *commodity_basket_index*. MinMaxScaler digunakan untuk variabel yang memiliki nilai minimum dan maksimum yang jelas, seperti persentase. RobustScaler digunakan apabila ada kolom yang memiliki banyak *outlier*.

E. Pemilihan Fitur (*Feature Selection*)

Setelah melakukan normalisasi, masih terdapat beberapa fitur yang distribusinya belum normal, sehingga penghapusan untuk data tersebut diperlukan. Fitur-fitur yang akhirnya dipilih untuk dipakai meliputi:

<i>province</i>	<i>farmer_organization_rate</i>	<i>temperature</i>
<i>population</i>	<i>farmers_totrade</i>	<i>humidity</i>
<i>area_of_land</i>	<i>ldi_dominance</i>	<i>rainfall_per_area</i>
<i>population_density</i>	<i>msg_basket_ssr</i>	<i>rainfall_optimal_gap</i>
<i>wetlands_area</i>	<i>soybean_yield_gap</i>	<i>temperature_optimal_gap</i>
<i>pdrb_per_capita</i>	<i>layer_yield_gap</i>	<i>commodity_basket_index</i>
<i>farmer_over_population</i> ,	<i>paddy_yield_gap</i>	<i>price_to_ssr</i>

F. Pemilihan dan Pembentukan Model *Machine Learning*

Jenis model *machine learning* yang dipilih adalah *unsupervised learning*, karena tidak ada jawaban pasti mengenai prioritas investasi sebuah provinsi. *Unsupervised learning* dipilih agar model *machine learning* dapat secara mandiri mendeteksi pola dan mengembalikan kesimpulan berdasarkan pola tersebut. Model *unsupervised learning* yang dipilih adalah klasterisasi atau *clustering*. *Clustering* mempelajari semua fitur yang ada di dataset dan mengelompokkan *datapoint* yang memiliki kemiripan karakteristik menjadi satu *cluster* atau kelompok.

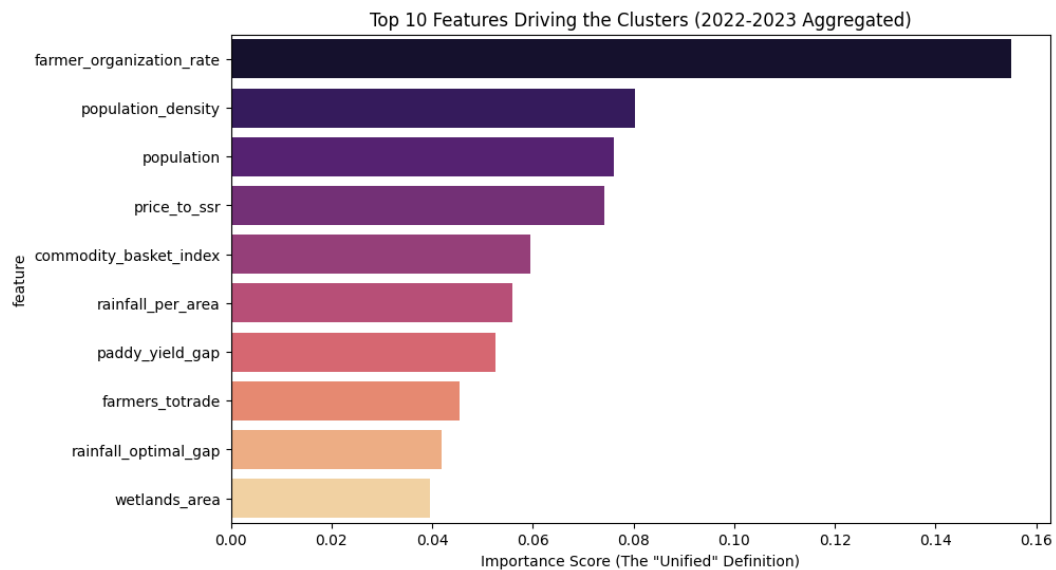
Metode *clustering* yang dipilih adalah *graph-based clustering*. Model *graph-based clustering* membagi data menjadi 3 jenis cluster dengan tingkat urgensi investasi yang berbeda. Setiap golongan memiliki tingkatan urgensi investasi dari pemerintah yang berbeda-beda. Pendekatan Graph-Based Clustering dipilih karena provinsi yang berdekatan seharusnya memiliki prioritas investasi yang mirip, karena fasilitas investasi di provinsi yang berdekatan saling mendukung satu sama lain, dan juga karena provinsi yang berdekatan cenderung memiliki karakteristik yang sama.

Seluruh dataset dengan fitur-fitur yang sudah dipilih pada fase *feature selection* beserta data koordinat geografis 38 provinsi di Indonesia dimasukkan ke dalam model *clustering*. Fase *clustering* dilakukan dua kali, yang pertama untuk dataset di tahun 2022, dan yang kedua untuk dataset di tahun 2023. *Clustering* menghasilkan dua set label *cluster* (untuk tahun 2022 dan tahun 2023) yang belum

tentu memiliki arti yang sama di antara kedua tahun, sehingga peneliti melakukan *alignment* dengan algoritma *centroid distance* yang menghitung jarak *euclidean* antara dua *centroids*. *Alignment* ini dilakukan agar *cluster* n pada tahun 2022 memiliki makna yang sama dengan *cluster* n pada tahun 2023.

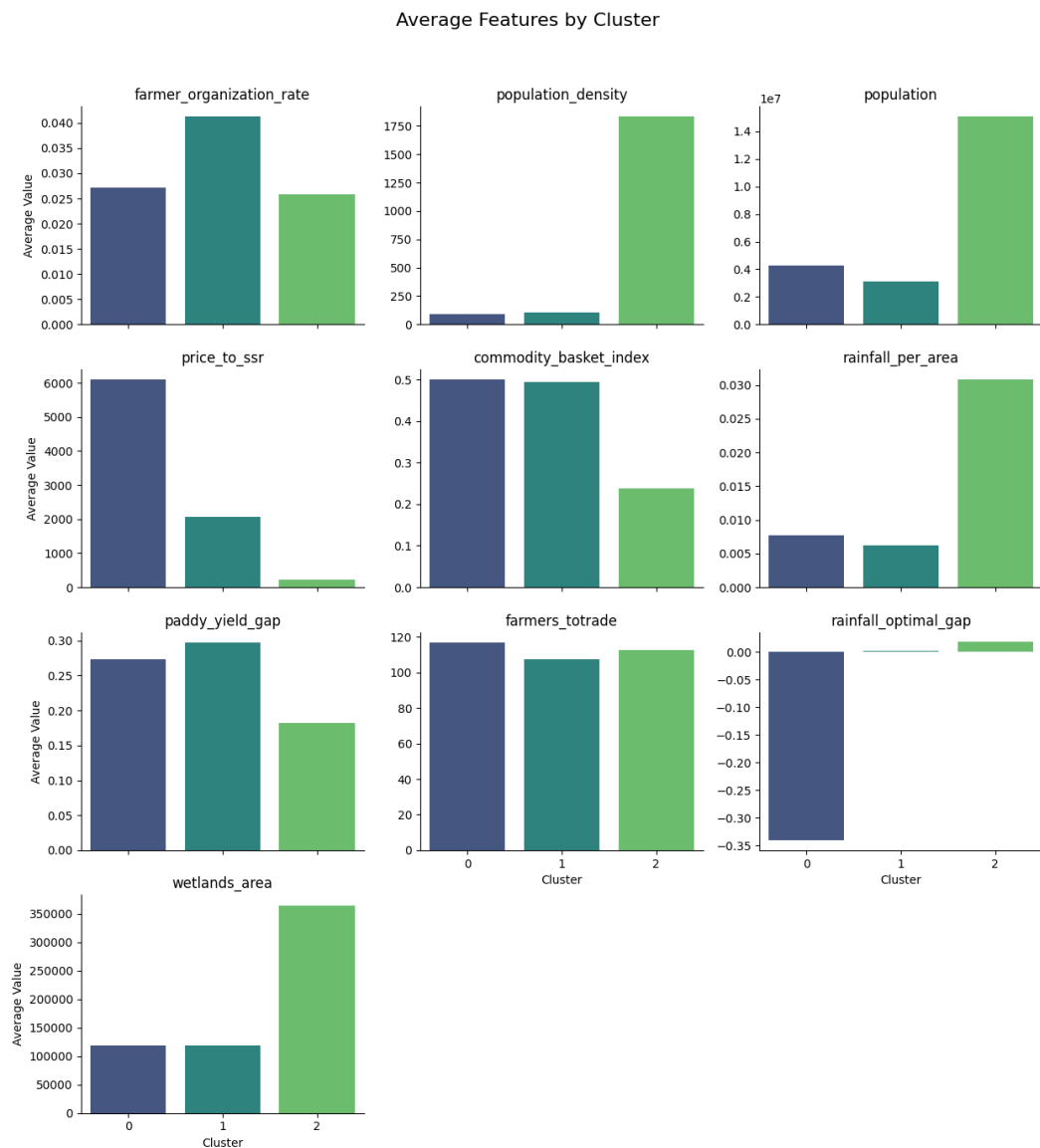
G. Interpretasi

Klasterisasi menghasilkan label numerik bagi masing-masing provinsi. Label numerik ini nantinya akan menjadi acuan tingkat prioritas investasi terhadap provinsi tersebut, namun sebelum arti dari masing-masing label *cluster* diketahui, diperlukan adanya analisis karakteristik masing-masing *cluster*. Teknik yang dipilih oleh peneliti untuk mendapatkan interpretasi *cluster* adalah menggunakan *feature importance* dengan XGBoost. Peneliti melakukan *training* model XGBoost dengan dataset utuh sebagai fitur dan label *cluster* sebagai fitur target yang harus diprediksi, kemudian melakukan ekstraksi *feature importance* untuk mengenali fitur mana saja yang membedakan antara data satu dengan yang lain. Peneliti dapat membedakan prioritas investasi masing-masing *cluster* dengan cara meneliti karakteristik fitur-fitur yang dideteksi oleh *feature importance*. Berikut 10 fitur paling penting yang dideteksi oleh model XGBoost.



Gambar 1. 10 Fitur Paling Penting dari XGBoost

Menggunakan sepuluh fitur terpenting tersebut, peneliti dapat memahami karakteristik ketiga *cluster* yang ada melalui agregasi data berdasarkan *cluster*. Berikut detail agregasi data ketiga *cluster*.



Gambar 2. Data Agregat per *Cluster* Berdasarkan 10 Fitur Paling Penting

Cluster 0 memiliki tingkat organisasi petani (*farmer_organization_rate*) di tengah-tengah ketiga provinsi, namun memiliki harga pangan yang sangat tinggi, dilihat dari *price_to_ssr* dan *commodity_basket_index* yang sangat tinggi dibandingkan *cluster* lain. *Cluster 0* juga memiliki luas sawah yang belum terlalu banyak, dan memiliki kesenjangan efektivitas lahan (*paddy_yield_gap*) yang sangat tinggi dari *cluster lainnya*. Berdasarkan data tersebut, dapat disimpulkan bahwa *cluster 0* merupakan *cluster* yang membutuhkan investasi darurat.

Cluster 1 memiliki tingkat organisasi petani (*farmer_organization_rate*) paling tinggi, dan harga pangan yang tidak terlalu tinggi. Walau begitu, *cluster 1* memiliki kesenjangan efektivitas lahan (*paddy_yield_gap*) yang paling tinggi dibandingkan *cluster* lain. Berdasarkan karakteristik ini, *cluster 1* membutuhkan investasi, namun tidak sedarurat *cluster 0*.

Cluster 2 memiliki lahan sawah (*wetlands_area*) paling banyak dibandingkan *cluster* lain, sehingga menjelaskan harga pangan (*price_to_ssr* dan *commodity_basket_index*) yang paling murah dibandingkan *cluster* lain. Faktor terakhir adalah *cluster 2* memiliki kesenjangan efektivitas (*paddy_yield_gap*) yang rendah. Dari karakteristik ini, *cluster 2* tidak membutuhkan investasi lagi.

H. Evaluasi Algoritma

Algoritma *SpectralClustering* mengandalkan *domain knowledge* dalam evaluasi hasilnya. Berdasarkan hasil agregat data dari ketiga *cluster* yang ada, ditemukan bahwa hasil *cluster* dan interpretasinya memang logis. Evaluasi matematis yang biasanya digunakan untuk *clustering* meliputi *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan skor *Calinski-Harabasz*, namun ketiganya tidak mempertimbangkan kedekatan graf.

Peneliti mengevaluasi hasil *clustering* menggunakan skor *Calinski-Harabasz*, karena metrik ini lebih tangguh atau *robust* untuk *SpectralClustering*. Skor yang didapatkan oleh *Calinski-Harabasz* adalah 3,94, dengan *benchmark* angka acak sebesar 1,01 dilakukan sebanyak 1000 kali. Skor ini menandakan bahwa *SpectralClustering* memang berhasil mengelompokkan data, karena nilainya lebih besar daripada sekedar menebak-nebak *cluster* secara asal. Walau begitu, ketika dibandingkan dengan nilai K-means sebesar 10.48, *SpectralClustering* kalah. Hal ini disebabkan karena *Calinski-Harabasz* mengevaluasi kelompok data berupa *blob*, sehingga sangat sesuai dengan algoritma K-means. Hal lain yang perlu diingat adalah *SpectralClustering* mempertimbangkan kedekatan graf, sedangkan K-means tidak. Dengan mempertimbangkan skor *Calinski-Harabasz* yang lebih tinggi sebesar tiga kali lipat dibandingkan nilai acak beserta data agregat yang masuk akal, dapat disimpulkan bahwa *SpectralClustering* berhasil mendeteksi pola yang ada di dalam dataset.

BAB IV HASIL YANG DICAPAI DAN POTENSI KHUSUS

A. Hasil

Model *graph-based clustering* yang digunakan oleh peneliti memberikan hasil *cluster* sebagai berikut:

Label <i>cluster</i>	No	Provinsi
1	1	Nanggroe Aceh Darussalam
	2	Sumatera Utara
	3	Sumatera Barat
	4	Sumatera Selatan
	5	Bengkulu
	6	Riau
	7	Kepulauan Riau
	8	Jambi
	9	Lampung
	10	Bangka Belitung
	11	Banten
	12	DKI Jakarta
	13	Daerah Istimewa Yogyakarta
	14	Jawa Tengah
	15	Jawa Barat
	16	Jawa Timur
	17	Kalimantan Barat
2	1	Kalimantan Utara
	2	Kalimantan Timur
	3	Kalimantan Tengah
	4	Kalimantan Selatan
	5	Sulawesi Utara
	6	Sulawesi Tengah
	7	Sulawesi Tenggara
	8	Sulawesi Barat
	9	Sulawesi Selatan
	10	Gorontalo
	11	Maluku
	12	Maluku Utara
	13	Nusa Tenggara Barat
	14	Nusa Tenggara Timur
	15	Bali
	16	Papua Barat
	17	Papua Barat Daya
3	1	Papua Tengah
	2	Papua Selatan
	3	Papua
	4	Papua Pegunungan

B. Potensi Khusus

Berdasarkan hasil *cluster* yang didapatkan dari *graph-based clustering*, diketahui bahwa *cluster* 1 memiliki provinsi-provinsi yang tergolong sangat stabil, mencakup provinsi-provinsi di Jawa yang cenderung sudah lebih stabil dalam hal kondisi pangan, *cluster* 2 memiliki provinsi-provinsi yang tergolong sudah stabil, namun masih bisa lebih baik, dan *cluster* 3 memiliki provinsi-provinsi yang tergolong belum stabil dalam hal kondisi pangan. Berdasarkan karakteristik dari masing-masing provinsi, dapat disimpulkan bahwa *cluster* 1 merupakan kumpulan provinsi yang sudah stabil (tidak perlu investasi lebih lagi), *cluster* 2 merupakan kumpulan provinsi yang kurang stabil (perlu investasi untuk *capital expenditure* atau CAPEX), dan *cluster* 3 merupakan kumpulan provinsi yang sama sekali belum stabil (perlu investasi untuk *operational expenditure* atau OPEX)

BAB V DOKUMENTASI

A. *Human-Led Work*

Dalam mengerjakan proyek ini, peneliti sendiri yang memilih dan mencari dataset yang dibutuhkan, memilih model *machine learning* yang sesuai, dan menginterpretasikan masing-masing arti dari label *cluster*. Peneliti memilih model *clustering* sebagai bagian dalam *unsupervised learning* karena dataset yang ada tidak memiliki jawaban asli atau label. Peneliti memilih dataset berupa data ekonomi, demografi, geografi, potensi pertanian, serta iklim karena kelima data tersebut berhubungan dengan prioritas investasi. Peneliti mengunduh dataset dari BPS dan Badan Pangan Nasional. Peneliti menginterpretasikan masing-masing arti dari label *cluster* dengan bantuan model *machine learning* XGBoost (*feature importance*), namun tetap menggunakan *domain knowledge* untuk menginterpretasikan agregat data setiap *clusternya*. Dalam tahap *preprocessing*, peneliti sendiri yang menentukan cara mengimputasi nilai yang kosong, mentransformasi dataset yang persebarannya tidak merata. Dalam tahap *feature engineering*, peneliti sendiri yang membuat fitur-fitur baru yang lebih representative.

B. *AI-Assisted Tasks*

Dalam mengerjakan proyek ini, AI membantu dalam menghasilkan kesimpulan berupa pengelompokan dataset dengan karakteristik mirip menjadi 1 *cluster* atau grup. AI juga membantu dalam mengartikan arti dari masing-masing *cluster*, melalui *feature importance* dalam fase pelatihan atau *train* XGBoost.

C. *AI Tool Usage*

Dalam pengerjaan *code* python untuk membuat model clustering, peneliti menggunakan bantuan AI, spesifiknya model LLM dalam beberapa hal, seperti dalam *code* untuk membuat dan melatih model *clustering*, *code* untuk menyelaraskan hasil *clustering* antara tahun 2022 dan 2023, *code* untuk membuat dan melatih model XGBoost untuk *feature importance*, dan ide pembuatan beberapa fitur dalam tahap *feature engineering*, seperti fitur *price_to_ssr*.

D. *Reflections on AI Limitations*

AI masih kurang paham dalam menentukan fitur-fitur penting yang perlu dibuat dalam tahap *feature engineer*, menentukan cara imputasi nilai kosong, menentukan cara mengatasi data *outlier*, menentukan cara mengatasi persebaran dataset yang tidak rata (*imbalanced*).

Manusia dan AI bukanlah dua pihak yang saling bertentangan. Manusia seharusnya bekerjasama dengan AI untuk menciptakan solusi nyata bagi permasalahan dunia. Dalam penerapannya, AI dapat mengerjakan banyak hal teknis, namun manusia yang menjadi percikan yang menyalakan inovasi. AI dapat membuat *code* yang benar dan efisien, namun manusia yang mengarahkan bagaimana cara kerja dan alur dari algoritma yang dibutuhkan. Permasalahan yang diangkat oleh proyek AOL ini adalah permasalahan yang melibatkan kondisi demografis, geografis, ekonomi, pertanian, dan iklim. Penelitian menjadi bukti bahwa manusia sendiri atau AI sendiri tidak cukup dalam menghadapi masalah kompleks di kehidupan nyata, namun perlu adanya kolaborasi antara manusia maupun AI.

E. Ethical and Safety Review

Diperlukan kehati-hatian dalam memahami atau menginterpretasikan kesimpulan yang diambil oleh model *machine learning*, karena keputusan yang diambil oleh pihak yang berwenang dapat memengaruhi hajat hidup orang banyak. Mengambil keputusan yang memengaruhi hidup dan mati orang banyak, seluruhnya berdasarkan kesimpulan sebuah model matematika dan statistika bukanlah hal yang bijak. Diperlukan adanya kolaborasi antara pihak yang berwenang dan para ahli dalam mengatasi permasalahan ini maupun dalam mengeksekusi solusi yang dihasilkan.

BAB VI PENUTUP

Program MBG yang telah diinisiasi oleh Presiden Prabowo memiliki tujuan yang baik, namun memiliki banyak tantangan dalam penerapannya, seperti disparitas harga pangan di setiap provinsi. Disparitas ini terbukti dapat diselesaikan melalui penambahan jumlah investasi kepada provinsi yang memiliki ketergantungan pangan terhadap provinsi lain. Peneliti mengembangkan pemetaan prioritas investasi terhadap 38 provinsi di Indonesia dalam rangka mendukung program MBG pemerintah.

Peneliti menggunakan model *machine learning* berupa *unsupervised learning*, spesifiknya *graph-based clustering* yang dapat mendeteksi pola tersembunyi dan mengelompokkan data yang memiliki karakteristik mirip berdasarkan pola tersebut. Model *clustering* yang digunakan peneliti menghasilkan kesimpulan bahwa provinsi-provinsi di sebelah timur Indonesia, seperti Papua Tengah, Papua Selatan, Papua Pegunungan, dan Papua Tenggara membutuhkan investasi darurat untuk *operational expenditure*.

DAFTAR PUSTAKA

- [1] Qomarrullah, R., Suratni, Wulandari, L., & Sawir, M. (2024). THE LONG-TERM IMPACT OF THE FREE NUTRITIOUS MEAL PROGRAM ON HEALTH AND EDUCATIONAL SUSTAINABILITY. Indonesian Journal of Intellectual Publication.
- [2] Fatimah, S., Rasyid, A., & Arwakon, H. O. (2024). Kebijakan Makan Bergizi Gratis di Indonesia Timur: Tantangan, Implementasi, dan Solusi untuk Ketahanan Pangan. Journal of Governance and Policy Innovation, 4(1), 14-21.
- [3] Sondakh, J., Rembang, J. H. W., & Syahyuti, N. (2021). KARAKTERISTIK, POTENSI GENERASI MILENIAL DAN PERSPEKTIF PENGEMBANGAN PERTANIAN PRESISI DI INDONESIA. Forum Penelitian Agro Ekonomi, 38(2), 155. <https://doi.org/10.21082/fae.v38n2.2020.155-166>
- [4] Wijaya, A. R., & Susandi, A. (2018). KONSEP FORECAST BASEDFINANCING UNTUK PERTANIAN PRESISI DI INDONESIA (Vol. 17).
- [5] Dewi, E. (2018). ANALISIS KEBIJAKAN SWASEMBADA BERAS DALAM UPAYA PENINGKATAN KETAHANAN PANGAN. In Jurnal Agribisnis Fakultas Pertanian Unita-Oktober.
- [6] Fatimah, S., Rasyid, A., & Arwakon, H. O. (2024). Kebijakan Makan Bergizi Gratis di Indonesia Timur: Tantangan, Implementasi, dan Solusi untuk Ketahanan Pangan. Journal of Governance and Policy Innovation, 4(1), 14-21.
- [7] Kurniawati, S. (2020). Prosiding Seminar Akademik Tahunan Ilmu Ekonomi dan Studi Pembangunan.
- [8] Qomarrullah, R., Suratni, Wulandari, L., & Sawir, M. (2024). THE LONG-TERM IMPACT OF THE FREE NUTRITIOUS MEAL PROGRAM ON HEALTH AND EDUCATIONAL SUSTAINABILITY. Indonesian Journal of Intellectual Publication.
- [9] Sondakh, J., Rembang, J. H. W., & Syahyuti, N. (2021). KARAKTERISTIK, POTENSI GENERASI MILENIAL DAN PERSPEKTIF PENGEMBANGAN PERTANIAN PRESISI DI INDONESIA. Forum Penelitian Agro Ekonomi, 38(2), 155. <https://doi.org/10.21082/fae.v38n2.2020.155-166>

- [10] Wijaya, A. R., & Susandi, A. (2018). KONSEP FORECAST BASEDFINANCING UNTUK PERTANIAN PRESISI DI INDONESIA (Vol. 17).
- [11] Yunitasari, D. (2025). JURNAL MANAJEMEN EKONOMI DAN BISNIS. JURNAL MANAJEMEN EKONOMI DAN BISNIS, 1(2), 30–42.
- [12] Puspa, A. W. (2023). 10 Provinsi dengan Persentase Penduduk Miskin Tertinggi di Indonesia. Bisnis.com.
- [13] Wikipedia. (2025). Health in Indonesia. Wikipedia, The Free Encyclopedia.