



POLITEKNIK STATISTIKA STIS

Sidang Skripsi Prodi Komputasi Statistik T.A. 2022/2023

Pemanfaatan Sumber *Big Data* dalam Memperkirakan Konsumsi Listrik Per Kapita Rumah Tangga Menggunakan *Small Area Estimation* (Studi Kasus: Level Kecamatan di Provinsi Jawa Tengah 2021)

Raflizal Fikrar Odriansyah (221910812/4SD2)

Dosen Pembimbing : Prof. Setia Pramana, S.Si., M.Sc., Ph.D.

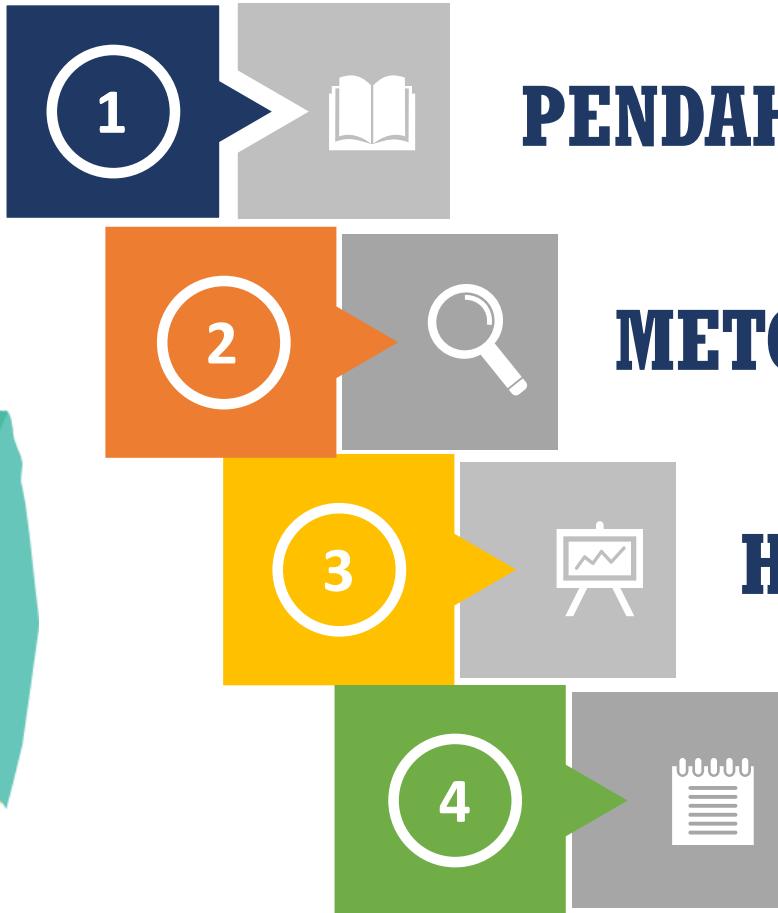
Dosen Penguji 1 : Dr. Eng. Arie Wahyu Wijayanto, S.S.T., M.T.

Dosen Penguji 2 : Farid Ridho, M.T.

Jakarta, 17 Juli 2023
R. 262, 14.30-15.30 WIB



Pokok Bahasan



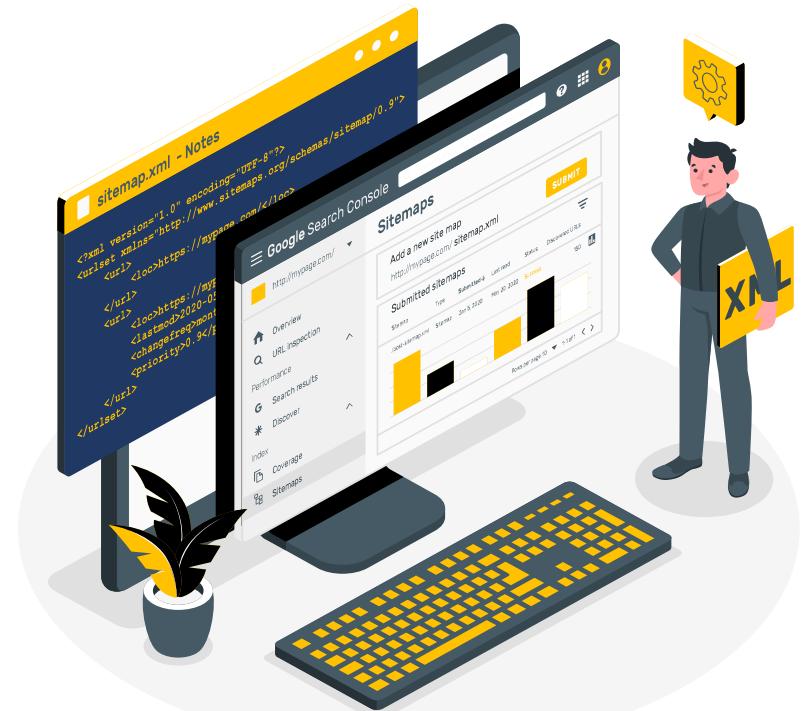
PENDAHULUAN

METODOLOGI

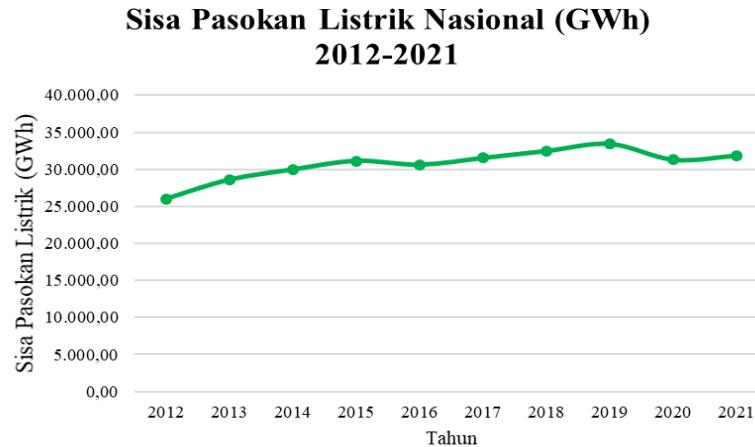
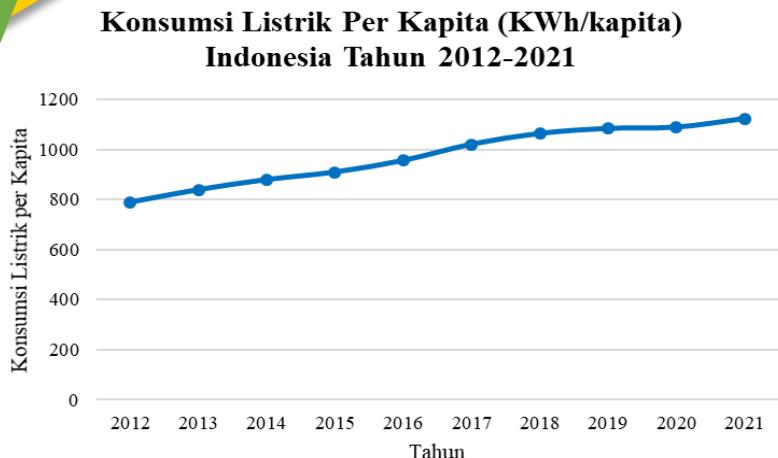
HASIL & PEMBAHASAN

PENUTUP

PENDAHULUAN



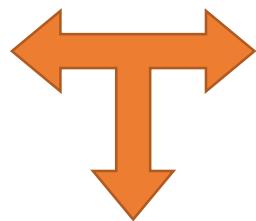
1. Latar Belakang Penelitian



“
Produksi energi listrik di **Jawa Tengah** hingga November 2021 masih mengalami **surplus** sehingga perlu meningkatkan **pemanfaatan sisa pasokan listrik** yang ada [1].
”

Sumber: Statistik Ketenagalistrikan Kementerian ESDM

Konsumsi listrik per kapita di Indonesia terus meningkat



Sumber: Statistik PLN 2021

Pasokan listrik PLN mengalami *oversupply* yang cukup besar (tidak efisien)



Urgensi terhadap perencanaan penyediaan energi listrik yang tepat dan efisien

Menimbulkan kerugian bagi PT. PLN dan negara

Diperlukannya data konsumsi listrik yang akurat dan presisi dengan metode lain



Referensi:

[1] M. Z. Arifin, “Produksi Energi Listrik di Jateng hingga November 2021 Alami Surplus,” Tribun Jateng, 2021. <https://pantura.tribunnews.com/2021/12/22/produksi-energi-listrikdi-jateng-hingga-november-2021-alami-surplus> (accessed Apr. 23, 2023).



1. Latar Belakang Penelitian



Pendugaan langsung (*direct estimation*) dengan data SUSENAS

Kelemahan



- Level penyajian estimasi hanya terbatas sampai tingkat kabupaten/kota
- Tingkat akurasi & presisi rendah untuk menghasilkan estimasi pada cakupan area lebih kecil
- Membutuhkan penambahan ukuran sampel yang lebih besar → biaya survei meningkat



Solusi yang diusulkan:

Penggunaan metode pendugaan tak langsung dengan *Small Area Estimation* yang memanfaatkan informasi tambahan dari sumber data lain, seperti sumber *big data* dan hasil pendataan potensi desa (Podes).

2. Research Gap & Opportunity



SAE

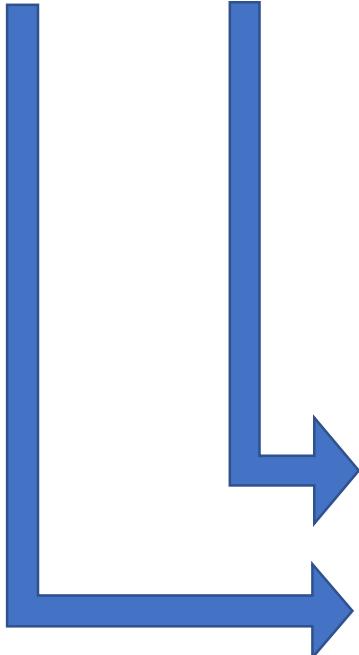


Research Gap

Integrasi pemanfaatan Sumber Big Data & Data Official Statistics dengan Small Area Estimation dalam memperkirakan konsumsi listrik per kapita level kecamatan

Memperhitungkan:

- a. Efek acak dan variabilitas dari unit area kecil
- b. Pengaruh spasial antarunit area
- c. Metode SAE berbasis GLMM (distribusi data variabel *interest* lebih adaptif)



Data *official statistics* (Pendataan Potensi Desa)

- ✓ Data penginderaan jauh (*remote sensing*)
- ✓ Data geospasial terbuka

Opportunity

- (a) Model SAE EBLUP-FH [2]
- (a)(b) Model SAE SEBLUP-FH [3]
- (a)(c) Model SAE HB [4]

Opportunity

Manfaatkan informasi tambahan yang tidak tercakup dalam survei

REFERENSI:

- [2] R. Gustiawan, "Pemanfaatan Big Data dan Deep Learning dalam Menduga Pengeluaran per Kapita di Tingkat Kecamatan Menggunakan Small Area Estimation (SAE)," pp. 1– 15, 2022.
- [3] A. Apriliansyah and I. Y. Wulansari, "Application of Spatial Empirical Best Linear Unbiased Prediction (SEBLUP) of Open Unemployment Rate on Sub-District Level Estimation in Banten Province," in Proceedings of The International Conference on Data Science and Official Statistics, 2021, vol. 2021, no. 1, pp. 905–913.
- [4] M. Mauliani, Maiyastri, and R. Diana, "Small Area Estimation dengan Pendekatan Hierarchical Bayes (HB) Lognormal," J. Mat. UNAND, vol. VII, no. 4, pp. 15–21, 2018.



3. Tujuan Penelitian

01

Melakukan pendugaan langsung (*direct estimation*) nilai rata-rata konsumsi listrik per kapita ruta level kecamatan di Provinsi Jawa Tengah dengan data SUSENAS Maret 2021

02

Melakukan pendugaan tak langsung (*indirect estimation*) nilai rata-rata konsumsi listrik per kapita ruta level kecamatan di Provinsi Jawa Tengah tahun 2021 menggunakan metode SAE dengan memanfaatkan informasi tambahan dari sumber *big data* dan hasil pendataan potensi desa (Podes).

03

Melakukan evaluasi hasil estimasi dari metode pendugaan langsung dan metode SAE untuk mendapatkan metode terbaik

04

Membuat pemetaan nilai estimasi rata-rata konsumsi listrik per kapita ruta level kecamatan berdasarkan metode terbaik



4. Batasan Penelitian



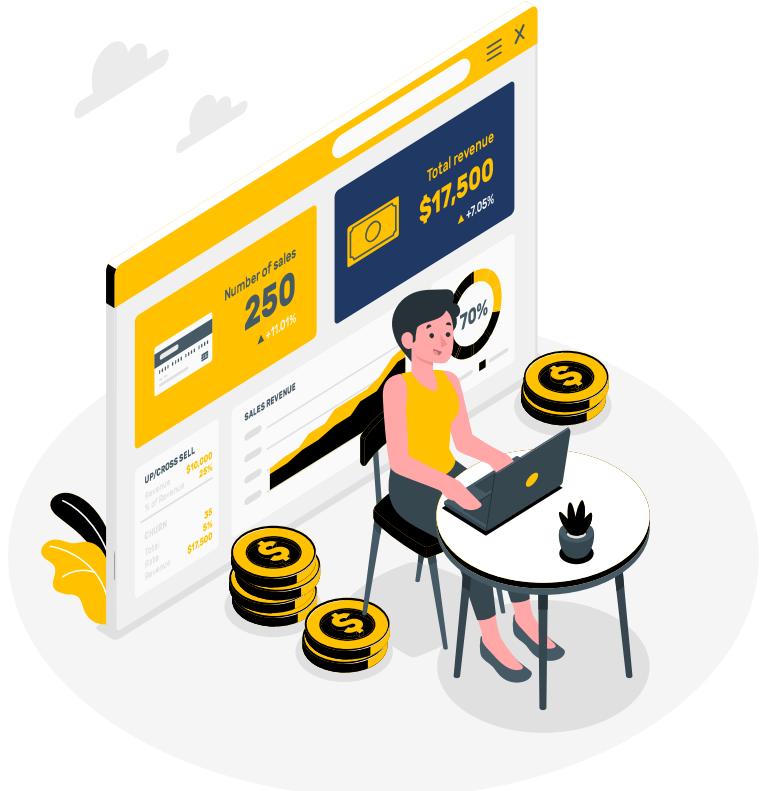
Estimasi mencakup wilayah sejumlah 572 dari 576 kecamatan di Provinsi Jawa Tengah sebagai unit analisis penelitian



Variabel yang diperoleh dari sumber *big data* bergantung pada resolusi spasial dan periode data yang digunakan



Variabel yang diperoleh dari sumber *big data* diasumsikan tidak mengandung *measurement error* (kesalahan pengukuran)



METODOLOGI

1. Variabel Penelitian

Variabel *Interest*: ESTIMASI RATA-RATA KONSUMSI LISTRIK PER KAPITA RUTA

- Nilai estimasi rata-rata konsumsi listrik per kapita ruta (KWh/Kapita) dihitung menggunakan data SUSENAS Maret dengan **metode pendugaan langsung** yang diformulasikan sbb:

$$\bar{y}_{kwh/kapita} = \frac{\sum_{i=1}^n \frac{y_{kwh_i}}{q_i} \times W_i}{\sum_{i=1}^n W_i}$$

$\bar{y}_{kwh/kapita}$: rata-rata konsumsi listrik per kapita ruta di suatu kecamatan

y_{kwh_i} : penggunaan listrik rumah tangga (KWh) selama sebulan

q_i : jumlah anggota rumah tangga pada rumah tangga ke- i

W_i : penimbang (bobot) individu pada rumah tangga ke- i

dengan $i = 1, \dots, n$ dimana n adalah ukuran sampel rumah tangga di setiap kecamatan.



1. Variabel Penelitian

Kandidat Variabel Penyerta: DATA PODES 2021

- ❑ Karena keterbatasan teori penelitian mengenai hubungan konsumsi listrik ruta dengan variabel dalam data PODES, maka digunakan **pendekatan kemiskinan dan ketimpangan dalam kaitannya dengan pemilihan kandidat peubah penyerta tersebut.**

Aspek	Kandidat Peubah Penyerta dari Data Podes
Infrastruktur	<ol style="list-style-type: none">1. Jumlah keluarga pengguna listrik PLN2. Jumlah tempat ibadah
Pendidikan	<ol style="list-style-type: none">1. Jumlah sarana pendidikan
Kesehatan	<ol style="list-style-type: none">1. Jumlah sarana kesehatan
Ekonomi	<ol style="list-style-type: none">1. Jumlah industri mikro dan kecil2. Jumlah sentra industri3. Jumlah sarana dan prasarana ekonomi

Sumber: [5]

- ❑ Selain itu, variabel **kepadatan penduduk (*population density*)** juga ditambahkan sebagai kandidat variabel penyerta yang diperoleh dari data Podes [6].

REFERENSI:

- [5] T. Siswanting, "Geoinformatika pada Kasus Area Kecil dan Penerapannya untuk Mendeteksi Kantong-Kantong Kemiskinan di Jember [Disertasi]," Bogor Inst. Pertan. Bogor, 2013.
- [6] I. F. U. Muzayannah, H. H. Lean, D. Hartono, K. D. Indraswari, and R. Partama, "Population density and energy consumption: A study in Indonesian provinces," *Heliyon*, vol. 8, no. 9, p. e10634, 2022, doi: 10.1016/j.heliyon.2022.e10634



1. Variabel Penelitian

Kandidat Peubah Penyerta: SUMBER *BIG DATA*

- ❑ Pemilihan kandidat peubah penyerta dari sumber *big data* didasarkan dari **teori penelitian** yang membahas tentang **isu lingkungan, infrastruktur, aktivitas masyarakat, dan electricity usage adaptation.**
- ❑ Berikut ini rincian kandidat peubah penyerta dari variabel sumber *big data*:

Kandidat Peubah Penyerta	Sumber Data	Resolusi Spasial (meter)	Periode Data	Referensi	Korelasi dengan Konsumsi Listrik Menurut Referensi
<i>Nighttime Light Intensity</i>	NOAA-VIIRS	463,83	2021	Shi et al. (2014), Ngestrini (2019)	Positif
<i>Built-Up Index</i>	Sentinel-2	10	2021	Faisal et al.(2016), Ouyang et al. (2016)	Positif
<i>Land Surface Temperature</i>	MODIS	1000	2021	Rajeshwari & Mani (2014), Ayuningtyas et al. (2015)	Positif
<i>Elevation</i>	SRTM-NASA DEM	30	2000	Luo et al. (2019), Katsoulakos & Kaliamvakos (2014)	Negatif
<i>Kepadatan Senyawa CO</i>	Sentinel-5P	1113,33	2021	Eom et al. (2020), T. Liu et al. (2018), Saberian et al. (2017)	Positif
<i>Konsentrasi Particulate Matter 2,5 (PM_{2,5})</i>	<i>Socioeconomic Data and Applications Center (SEDAC) NASA</i>	1111	2019		
<i>Global Aridity Index</i>	<i>Global Aridity and PET Database</i>	1000	1970-2000	Zomer et al. (2022), Eom et al. (2020)	Negatif
<i>Relative Wealth Index</i>	<i>Facebook's Data for Good</i>	2400	2021	Ferguson et al. (2000), Chi et al. (2022)	Positif



2. Metode Pengumpulan Data

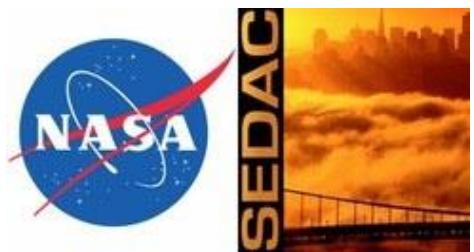


Raw data SUSENAS Konsumsi & Pengeluaran (KP) Maret 2021 dan data mikro PODES 2021 diperoleh dari Badan Pusat Statistik.



Google Earth Engine

Data *nighttime light intensity* (NTL), *land surface temperature* (LST), *build-up index* (BUI), *elevation*, dan kepadatan senyawa CO diekstrak dan diekspor dalam bentuk raster (GeoTIFF) menggunakan platform layanan Google Earth Engine.

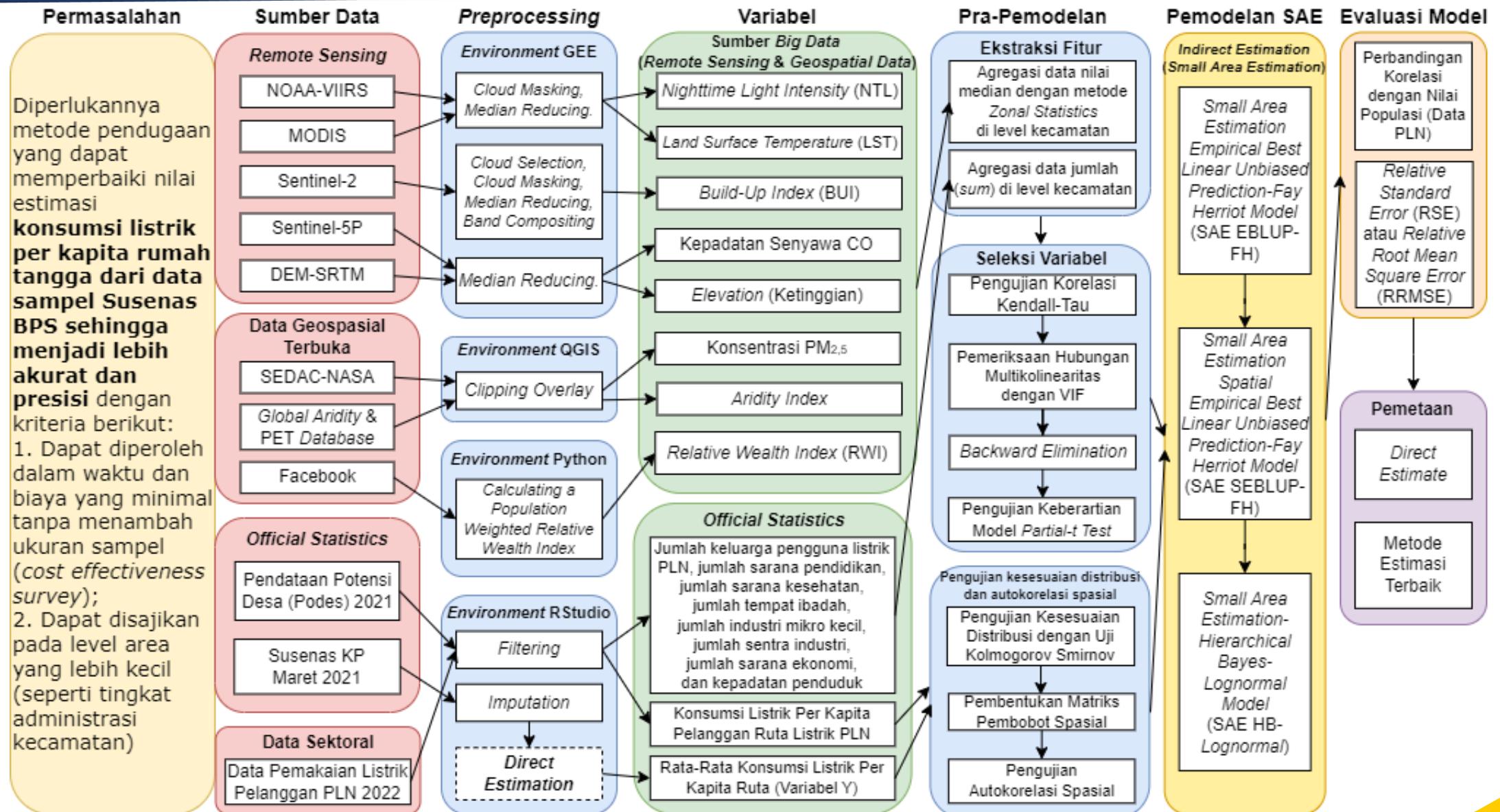


Awesome GEE
Community
Catalog

Data konsentrasi PM_{2,5} dan *global aridity index*, diunduh secara langsung dalam bentuk raster (GeoTIFF). Sedangkan data *relative wealth index* (RWI) diunduh langsung dalam bentuk CSV, kemudian dilakukan *adjustment* dengan data *high resolution population density*.



3. Kerangka Pikiran

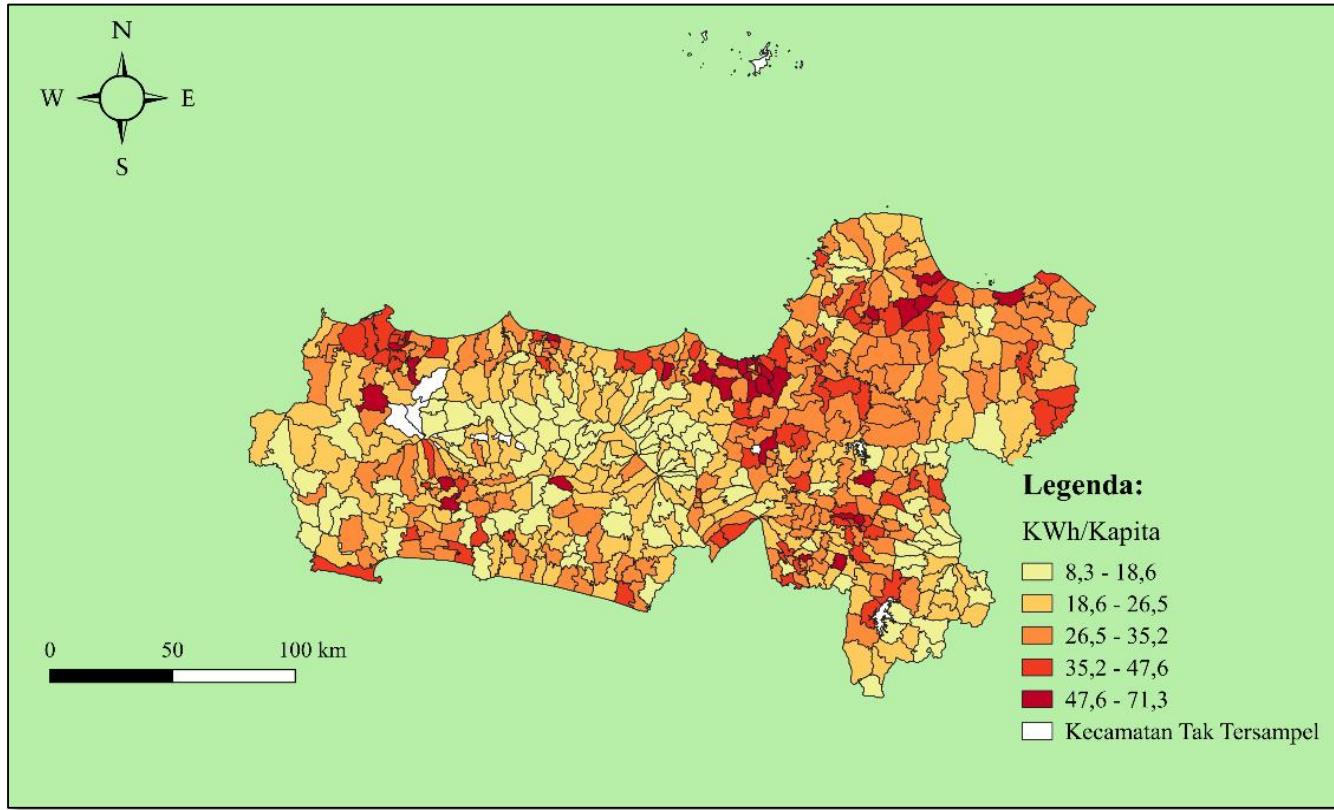


HASIL & PEMBAHASAN



Tujuan 1: Direct Estimation

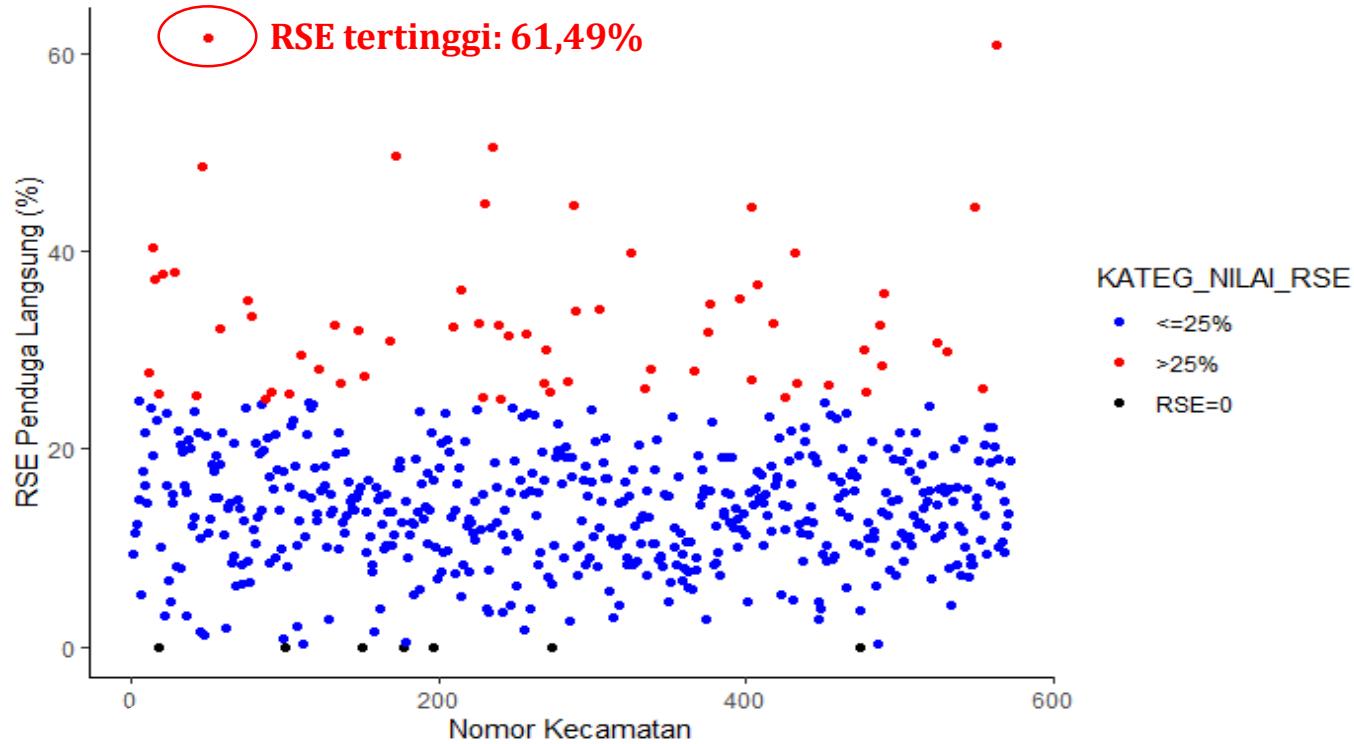
A. Hasil Pendugaan Langsung



Wilayah kecamatan di daerah **perkotaan** cenderung memberikan hasil estimasi rata-rata KWh/Kapita rumah tangga yang **relatif lebih tinggi** dibandingkan dengan wilayah kecamatan yang berada di **perdesaan** di mana banyak dijumpai di daerah kabupaten.

Tujuan 1: Direct Estimation

B. Relative Standard Error (RSE) Penduga Langsung

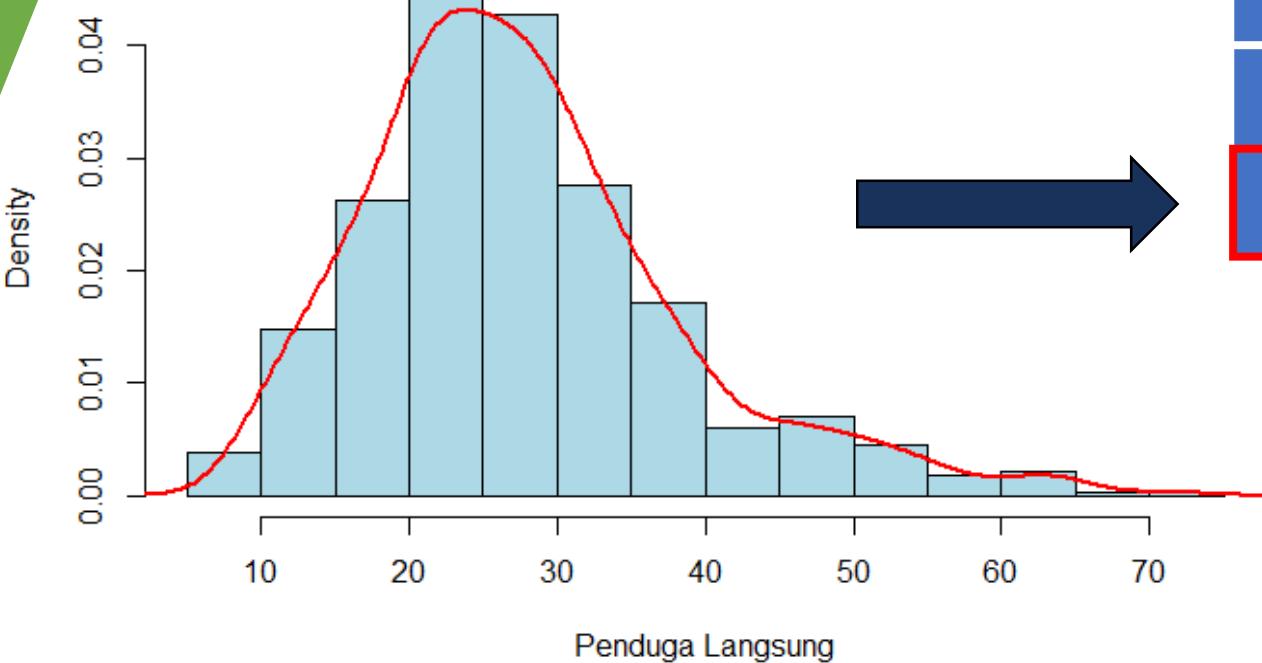


Dot plot nilai RSE dari penduga langsung

Tujuan 1: Direct Estimation

C. Pemeriksaan Distribusi Penduga Langsung

Distribusi Data Penduga Langsung



Hasil pengujian kesesuaian distribusi penduga langsung dengan Kolmogorov Smirnov

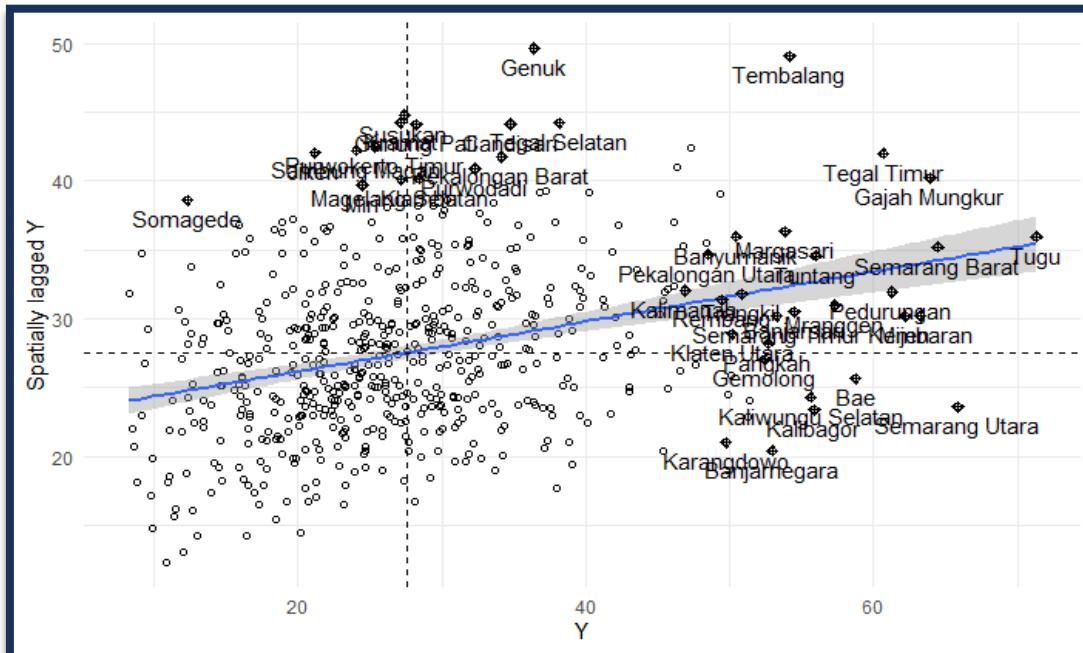
Statistik Uji (D)	P-value	Keterangan
0,0783	0,0018	Penduga langsung tidak berdistribusi normal
0,0416	0,2748*	Penduga langsung berdistribusi lognormal

Penduga langsung **berdistribusi lognormal** sehingga metode **SAE pendekatan Hierarchical Bayes (HB) Lognormal** dapat digunakan dalam menduga nilai rata-rata konsumsi listrik per kapita ruta level kecamatan.



Tujuan 1: Direct Estimation

D. Pengujian Autokorelasi Spasial Penduga Langsung



Moran's I scatter plot penduga langsung

tetangga ke- <i>j</i>	tetangga ke- <i>i</i>					Jumlah
	1	2	3	4	...	
1	0	0,333	0,333	0	...	0
2	0,200	0	0,200	0	...	0
3	0,167	0,167	0	0	0	1
4	0	0	0	0	0	1
...
572	0	0	0	0	...	0

Bentuk ilustrasi matriks pembobot spasial tipe *Queen* yang distandarisasi

Hasil pengujian autokorelasi spasial dengan *Moran's I test* pada penduga langsung

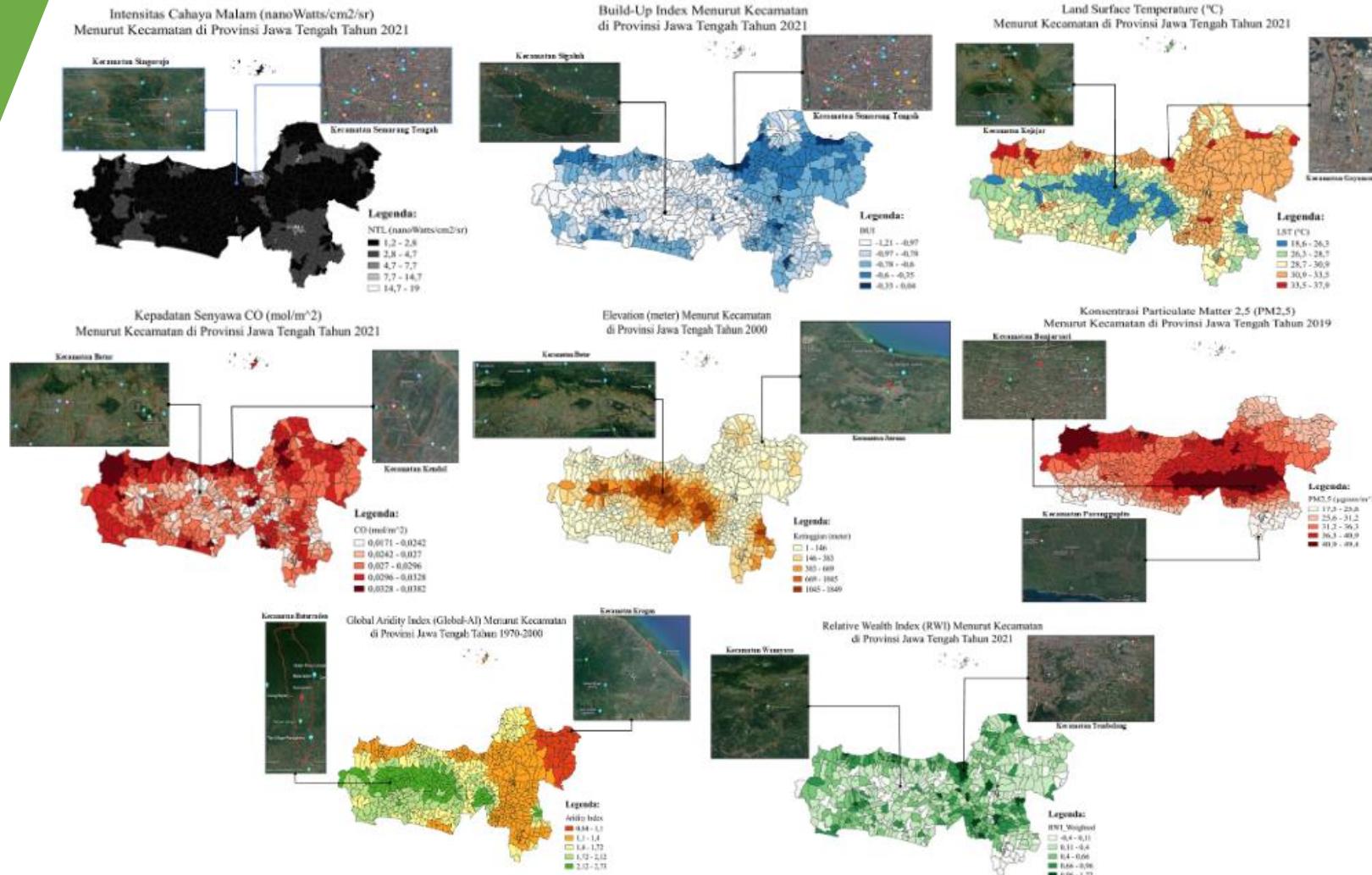
Model *Spatial EBLUP-FH* dapat digunakan

Indeks Moran	Z-statistics	P-value	Keterangan
0,1805	7,0841	$P < 0,0001$	Terdapat dependensi spasial antarwilayah



Tujuan 1: Direct Estimation

E. Analisis Deskriptif Kandidat Variabel Penyerta



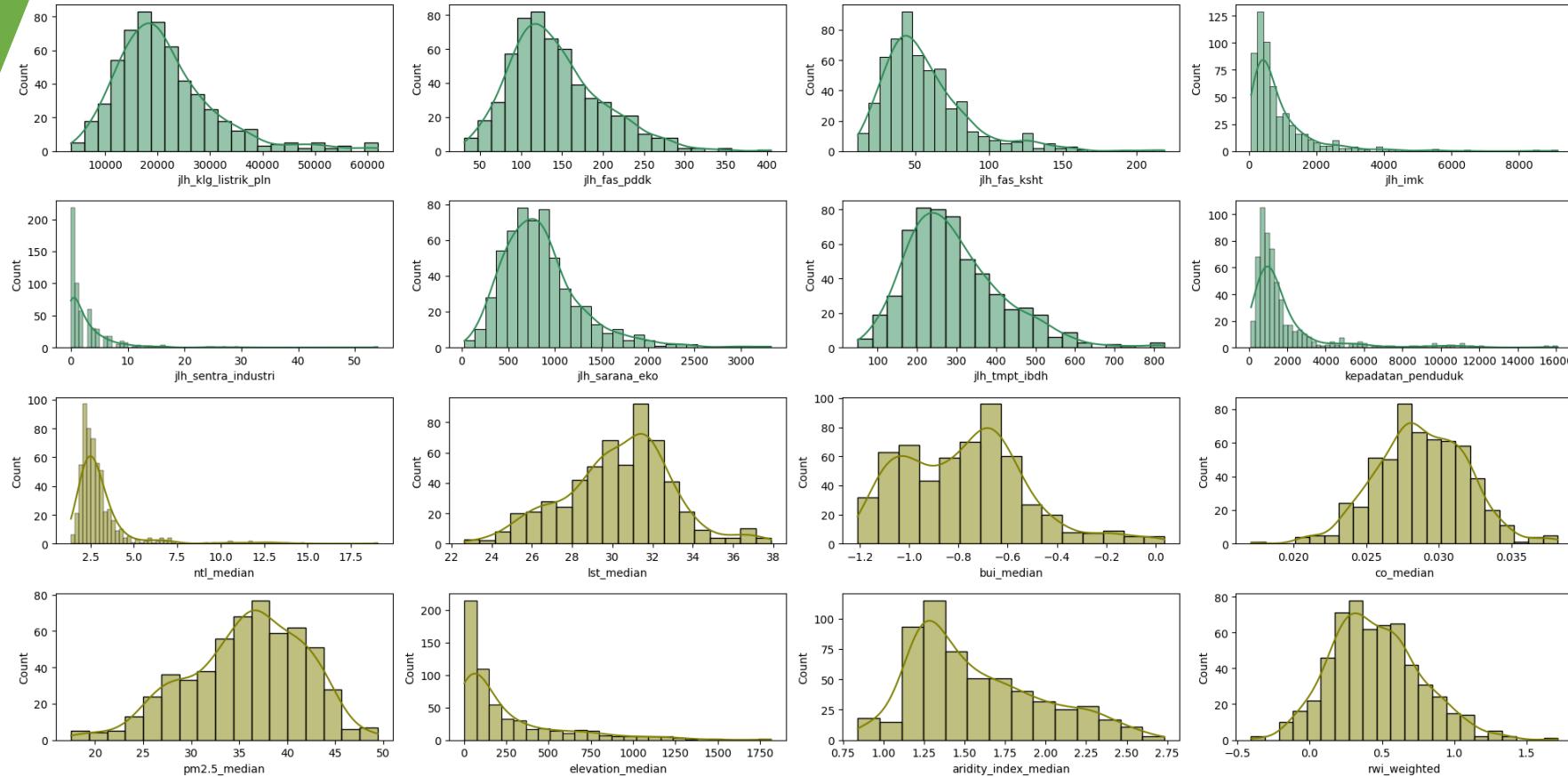
Peta tematik hasil ekstraksi sumber *big data* sebagai kandidat variabel penyerta

Secara garis besar, wilayah dengan NTL tinggi, BUI tinggi, LST tinggi, elevasi rendah kepadatan senyawa CO tinggi, konsentrasi PM2.5, indeks kekeringan tinggi, dan RWI tinggi berasosiasi dengan wilayah yang memiliki tingkat rata-rata konsumsi listrik ruta yang relatif tinggi, vice versa.



Tujuan 1: Direct Estimation

E. Analisis Deskriptif Kandidat Variabel Penyerta



Histogram data seluruh kandidat variabel penyerta
Keterangan: hijau (data Podes); coklat (sumber big data)

Sebaran data dari seluruh kandidat variabel penyerta secara umum tampak tidak simetris sehingga digunakan metode statistik nonparametrik pada uji korelasi yaitu dengan Rank Kendall-Tau



Tujuan 2: Prapemodelan SAE

A. Seleksi Variabel: Pengujian Korelasi Kendall-Tau

Variabel	Nama Kandidat Variabel Penyerta	Korelasi	P-value	Keterangan
X_1	Jumlah keluarga pengguna listrik PLN	0,2711	$P < 0,0001$	Signifikan
X_2	Jumlah sarana pendidikan	0,2014	$P < 0,0001$	Signifikan
X_3	Jumlah sarana kesehatan	0,3234	$P < 0,0001$	Signifikan
X_4	Jumlah tempat ibadah	0,0952	0,0007	Signifikan
X_5	Jumlah industri mikro dan kecil	-0,0241	0,3887	Tidak Signifikan
X_6	Jumlah sentra industri	-0,0527	0,0819	Tidak Signifikan
X_7	Jumlah sarana dan prasarana ekonomi	0,2757	$P < 0,0001$	Signifikan
X_8	Kepadatan penduduk	0,3215	$P < 0,0001$	Signifikan
X_9	<i>Nighttime Light Intensity</i>	0,2999	$P < 0,0001$	Signifikan
X_{10}	<i>Built-Up Index</i>	0,3931	$P < 0,0001$	Signifikan
X_{11}	<i>Land Surface Temperature</i>	0,4103	$P < 0,0001$	Signifikan
X_{12}	<i>Elevation</i>	-0,3244	$P < 0,0001$	Signifikan
X_{13}	Kepadatan Senyawa CO	0,2150	$P < 0,0001$	Signifikan
X_{14}	Konsentrasi Particulate Matter 2.5 ($PM_{2.5}$)	0,1357	$P < 0,0001$	Signifikan
X_{15}	<i>Global Aridity Index</i>	-0,2948	$P < 0,0001$	Signifikan
X_{16}	<i>Relative Wealth Index</i>	0,2828	$P < 0,0001$	Signifikan

Dari hasil uji korelasi Kendall-Tau diperoleh **14 variabel memiliki hubungan linear yang signifikan terhadap nilai penduga langsung dengan taraf signifikansi sebesar 0,05.**



Tujuan 2: Prapemodelan SAE

B. Seleksi Variabel: Pemeriksaan Multikolinearitas

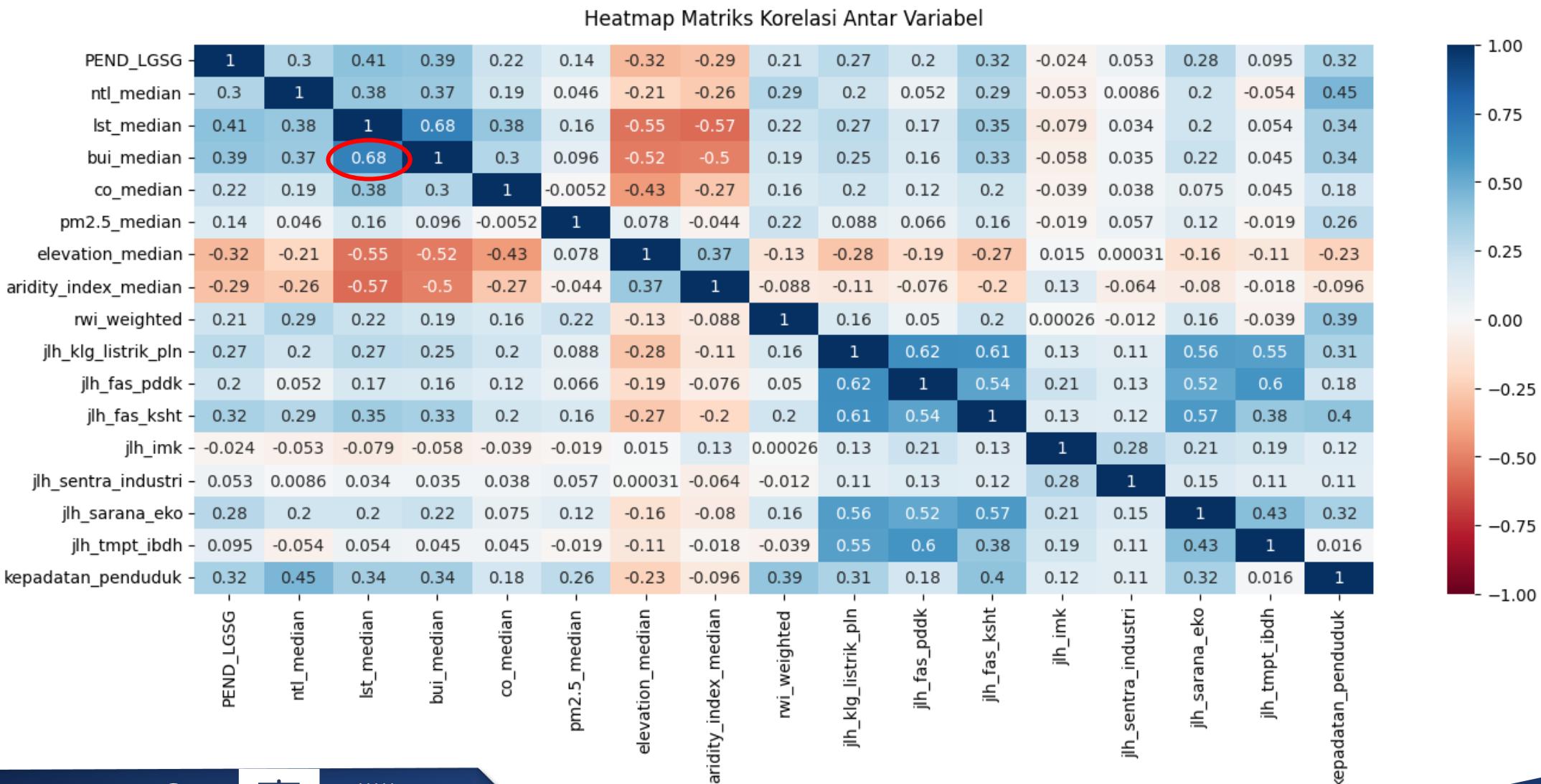
Variabel	Nama Kandidat Variabel Penyerta	Variance Inflation Factor (VIF)
X_1	Jumlah keluarga pengguna listrik PLN	5,816031
X_2	Jumlah sarana pendidikan	4,880905
X_3	Jumlah sarana kesehatan	5,096888
X_4	Jumlah tempat ibadah	4,323902
X_7	Jumlah sarana dan prasarana ekonomi	2,881605
X_8	Kepadatan penduduk	6,467060
X_9	<i>Nighttime Light Intensity</i>	5,631555
X_{10}	<i>Built-Up Index</i>	6,443919
X_{11}	<i>Land Surface Temperature</i>	12,699985
X_{12}	<i>Elevation</i>	4,359110
X_{13}	Kepadatan Senyawa CO	1,700263
X_{14}	Konsentrasi Particulate Matter 2.5 ($PM_{2.5}$)	1,646249
X_{15}	<i>Global Aridity Index</i>	3,670783
X_{16}	<i>Relative Wealth Index</i>	1,548146

Variabel LST memiliki nilai **VIF > 10** dan **berkorelasi kuat** dengan **variabel BUI** sebesar **0,68**. Oleh sebab itu, **salah satu** dari kedua variabel tersebut dieliminasi, yaitu **variabel LST**.



Tujuan 2: Prapemodelan SAE

B. Seleksi Variabel: Pemeriksaan Multikolinearitas



Tujuan 2: Prapemodelan SAE

C. Seleksi Variabel: *Backward Elimination*

Step	Daftar Kandidat Variabel Penyerta	AIC
1	$X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}$	2427,40
2	$X_1, X_2, X_3, X_4, X_7, X_8, X_9, X_{10}, X_{12}, X_{14}, X_{15}, X_{16}$	2425,49
3	$X_1, X_2, X_3, X_4, X_7, X_8, X_{10}, X_{12}, X_{14}, X_{15}, X_{16}$	2423,58
4	$X_1, X_2, X_4, X_7, X_8, X_{10}, X_{12}, X_{14}, X_{15}, X_{16}$	2421,67
5	$X_2, X_4, X_7, X_8, X_{10}, X_{12}, X_{14}, X_{15}, X_{16}$	2420,66
6	$X_2, X_4, X_7, X_8, X_{10}, X_{12}, X_{14}, X_{16}$	2419,59

Dari hasil eliminasi dengan metode *backward* diperoleh **8 variabel** yakni: $X_2, X_4, X_7, X_8, X_{10}, X_{12}, X_{14}$, dan X_{16} .



Tujuan 2: Prapemodelan SAE

D. Seleksi Variabel: Uji Keberartian Model dengan *Partial-t Test*

Variabel	Statistik <i>t</i>	P-value	Keterangan
X_2	1,695	0,091	Tidak Signifikan
X_4	-2,283	0,023*	Signifikan
X_7	4,457	$P < 0,001^*$	Signifikan
X_8	1,561	0,119	Tidak Signifikan
X_{10}	4,937	$P < 0,001^*$	Signifikan
X_{12}	-5,055	$P < 0,001^*$	Signifikan
X_{14}	2,094	0,037*	Signifikan
X_{16}	2,300	0,022*	Signifikan

Variabel X_2 dan X_8 tidak signifikan sehingga dieliminasi dan hanya terpilih **6 variabel** yang digunakan sebagai variabel penyerta pada pemodelan SAE.

Keterangan variabel:

Jumlah sarana pendidikan (X_2), jumlah tempat ibadah (X_4), jumlah sarana & prasarana ekonomi (X_7), kepadatan penduduk (X_8), *build-up index* (X_{10}), *elevation* (X_{12}), konsentrasi PM_{2,5} (X_{14}), dan *relative wealth index* (X_{16}).

***) variabel yang terpilih masuk pada pemodelan SAE.**



Tujuan 2: Hasil Pemodelan SAE

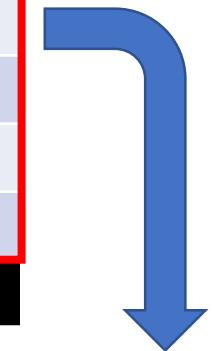
A. Model SAE EBLUP-FH

Hasil estimasi koefisien regresi dan koefisien ragam pengaruh acak area model EBLUP-FH

Variabel	Estimasi koefisien	Standard error	P-value
<i>Intercept</i> (b_0)	27,8024	2,4605	$1,3194 \times 10^{-29}$
Jumlah tempat ibadah (b_1)	-0,0067	0,0031	$2,764 \times 10^{-2}$
Jumlah sarana dan prasarana ekonomi (b_2)	0,0064	0,0009	$2,0953 \times 10^{-11}$
<i>Build-Up Index</i> (b_3)	11,7122	1,6722	$2,4879 \times 10^{-12}$
Konsentrasi PM _{2,5} (b_4)	0,1312	0,0547	$1,6503 \times 10^{-2}$
<i>Elevation</i> (b_5)	-0,0064	0,0012	$4,7032 \times 10^{-8}$
<i>Relative Wealth Index</i> (b_6)	1,8265	1,1152	$1,0146 \times 10^{-1}$
Ragam pengaruh acak area ($\hat{\sigma}_v^2$)	33,1056		

Hasil pengujian normalitas pengaruh acak area model SAE EBLUP-FH dengan Uji Kolmogorov Smirnov (KS)

Variabel Uji	Statistik Uji (D)	P-value	Keterangan
v_i	0,0398	0,3262	Pengaruh acak area berdistribusi normal



Variabel RWI tidak
signifikan pada taraf
signifikansi 0,05



Asumsi normalitas pengaruh acak area pada model EBLUP-FH
terpenuhi sesuai yang diungkapkan oleh Rao & Molina (2015)

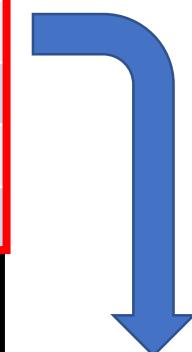


Tujuan 2: Hasil Pemodelan SAE

B. Model SAE SEBLUP-FH

Hasil estimasi koefisien regresi, koefisien ragam pengaruh acak area, dan koefisien autoregresif spasial model SEBLUP-FH

Variabel	Estimasi koefisien	Standard error	P-value
Intercept (b_0)	27,9392	2,5418	$4,1909 \times 10^{-28}$
Jumlah tempat ibadah (b_1)	-0,0074	0,0031	$1,5929 \times 10^{-2}$
Jumlah sarana dan prasarana ekonomi (b_2)	0,0064	0,0009	$2,4882 \times 10^{-11}$
Build-Up Index (b_3)	11,8871	1,6979	$2,5411 \times 10^{-12}$
Konsentrasi PM _{2,5} (b_4)	0,1375	0,0566	$1,5197 \times 10^{-2}$
Elevation (b_5)	-0,0062	0,0012	$2,3376 \times 10^{-7}$
Relative Wealth Index (b_6)	1,6733	1,1252	$1,3698 \times 10^{-1}$
Ragam pengaruh acak area ($\hat{\sigma}_v^2$)	32,9143		
Autoregresi spasial ($\hat{\rho}$)	0,1177		



Variabel RWI tidak

signifikan pada taraf

signifikansi 0,05

Hasil pengujian normalitas pengaruh acak area model SAE SEBLUP-FH dengan Uji Kolmogorov Smirnov (KS)

Variabel Uji	Statistik Uji (D)	P-value	Keterangan
v_i	0,0394	0,3380	Pengaruh acak area berdistribusi normal

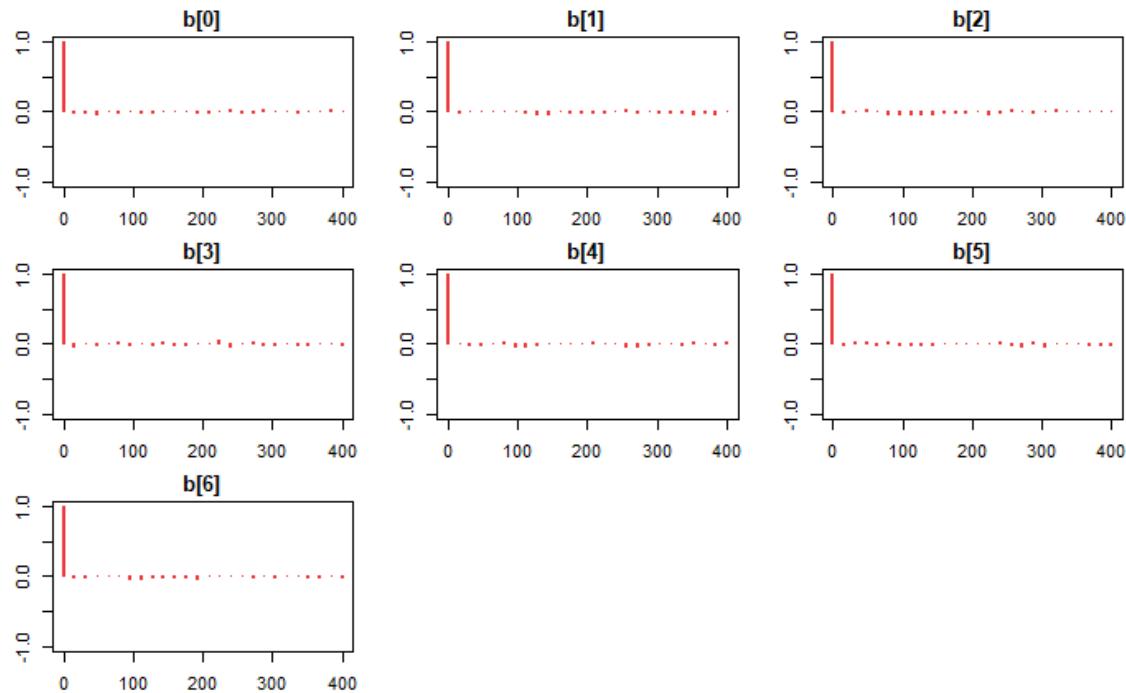


Asumsi normalitas pengaruh acak area pada model SEBLUP-FH terpenuhi sesuai yang diungkapkan oleh Rao & Molina (2015)

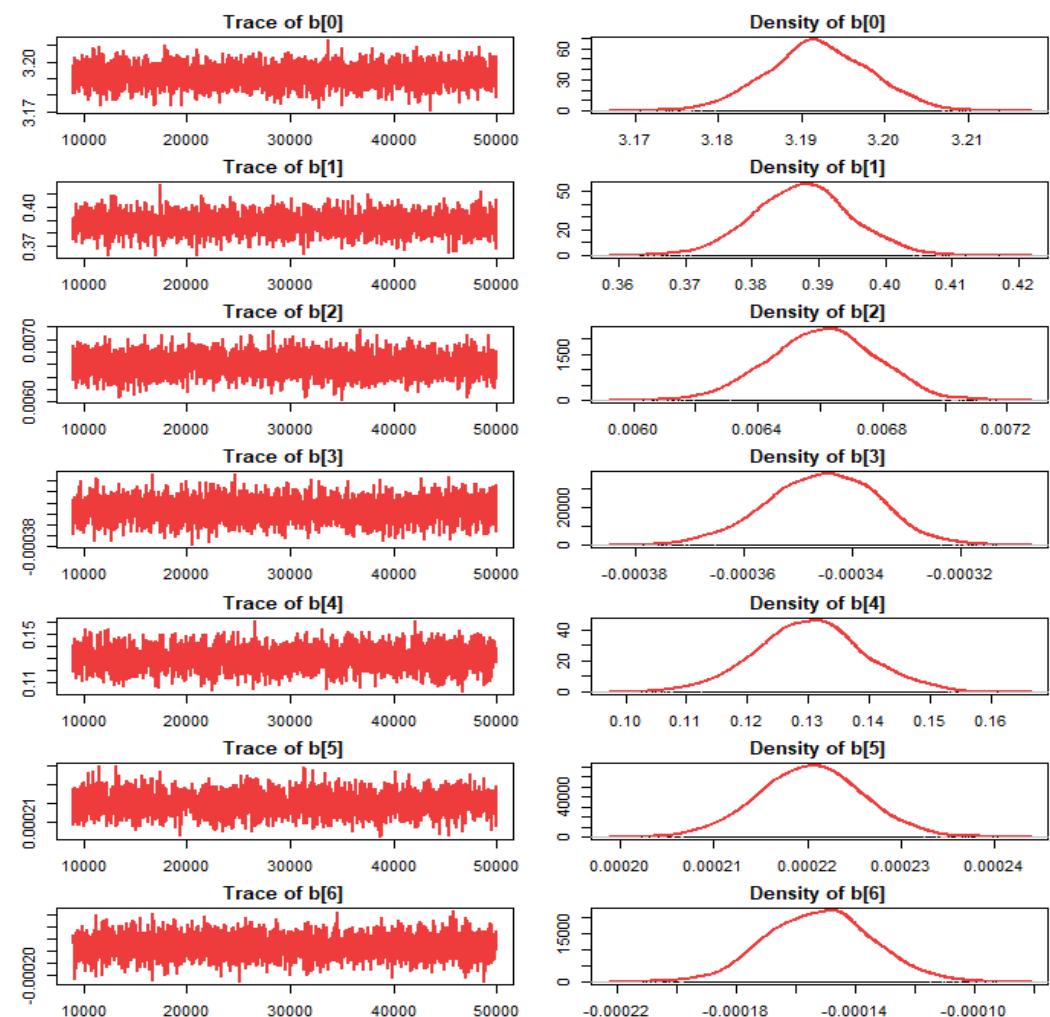
Tujuan 2: Hasil Pemodelan SAE

C. Model SAE HB Lognormal

- Penentuan prior β dan σ_v^2 diasumsikan bahwa $\beta_j \sim N_{iid}(\mu_{[\beta]_j}, \sigma_{[\beta]_j}^2)$ dan $v_i \sim N_{iid}(0, \tau_v)$ dengan nilai awal prior $\mu_{[\beta]_j} = 0$, $\sigma_{[\beta]_j}^2 = 1$, dan $\tau_v = 1$, di mana $\tau_v = \frac{1}{\sigma_v^2}$.
- Algoritma MCMC mencapai konvergen pada model percobaan ke-8 dengan spesifikasi: iterasi = 50.000 kali, burn in = 9.000 kali, thinning = 16 kali, dan update = 16 kali.



Autocorrelation plot dari parameter β model percobaan ke-8



Trace plot dan density plot dari parameter β model percobaan ke-8

Tujuan 2: Hasil Pemodelan SAE

C. Model SAE HB *Lognormal*

Hasil estimasi koefisien regresi dan koefisien ragam pengaruh acak area model SAE HB *Lognormal* dengan 95% *credible interval*

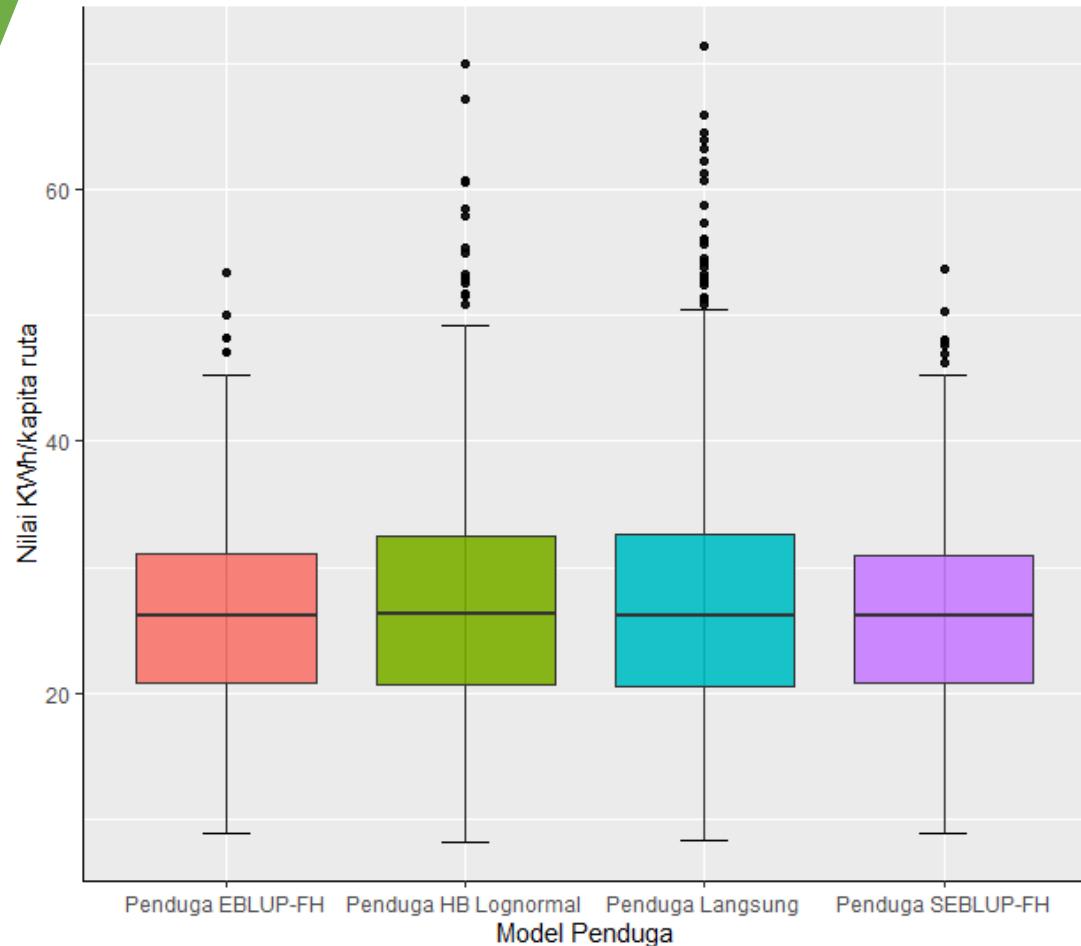
Variabel	Estimasi koefisien	Standard error	2,5%	97,5%
<i>Intercept</i> (b_0)	3,19199	$6,089 \times 10^{-3}$	3,18011	3,20364
Jumlah tempat ibadah (b_1)	-0,00040	$2,78 \times 10^{-5}$	-0,00045	-0,00034
Jumlah sarana dan prasarana ekonomi (b_2)	0,00016	$8,38 \times 10^{-6}$	0,00015	0,00018
<i>Build-Up Index</i> (b_3)	0,38755	$7,307 \times 10^{-3}$	0,37327	0,40186
Konsentrasi PM _{2,5} (b_4)	0,00661	$1,690 \times 10^{-4}$	0,00628	0,00692
<i>Elevation</i> (b_5)	-0,00035	$1,000 \times 10^{-5}$	-0,00037	-0,00033
<i>Relative Wealth Index</i> (b_6)	0,13026	$8,376 \times 10^{-3}$	0,11310	0,14806
Ragam pengaruh acak area ($\hat{\sigma}_v^2$)	0,06995			

Dapat dilihat bahwa 95% *credible interval* dari seluruh parameter tidak mengandung nilai nol sehingga menunjukkan **seluruh variabel signifikan** memengaruhi nilai *logaritma* dari penduga langsung.



Tujuan 2: Hasil Pemodelan SAE

D. Perbandingan Hasil Estimasi



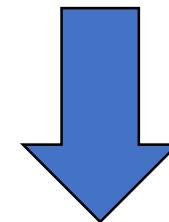
Box plot hasil estimasi antara empat model pendugaan

Nilai korelasi Kendall-Tau dan signifikansinya antara penduga langsung dengan penduga SAE

Penduga SAE	Korelasi Kendall-Tau	P-value	Keterangan
EBLUP-FH	0,8492	$P < 0,0001$	Signifikan
SEBLUP-FH	0,8502	$P < 0,0001$	Signifikan
HB Lognormal	0,9334	$P < 0,0001$	Signifikan

Model

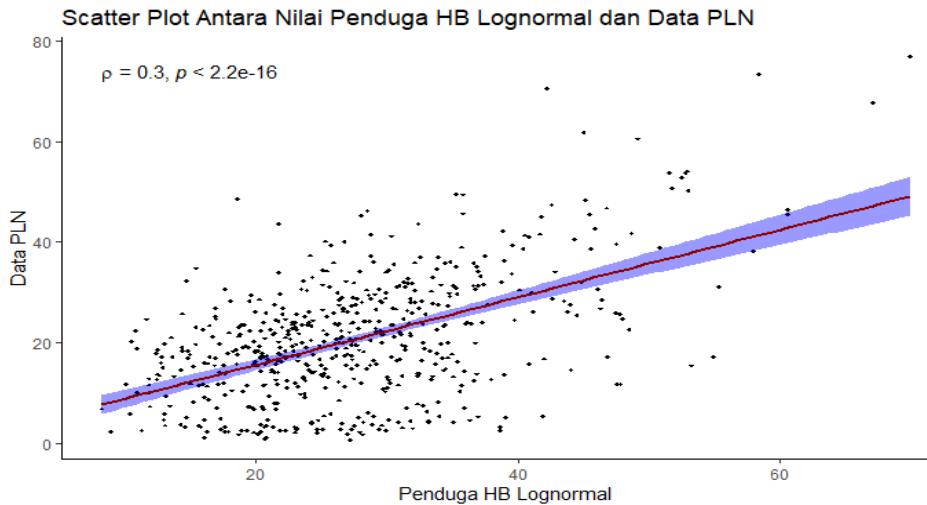
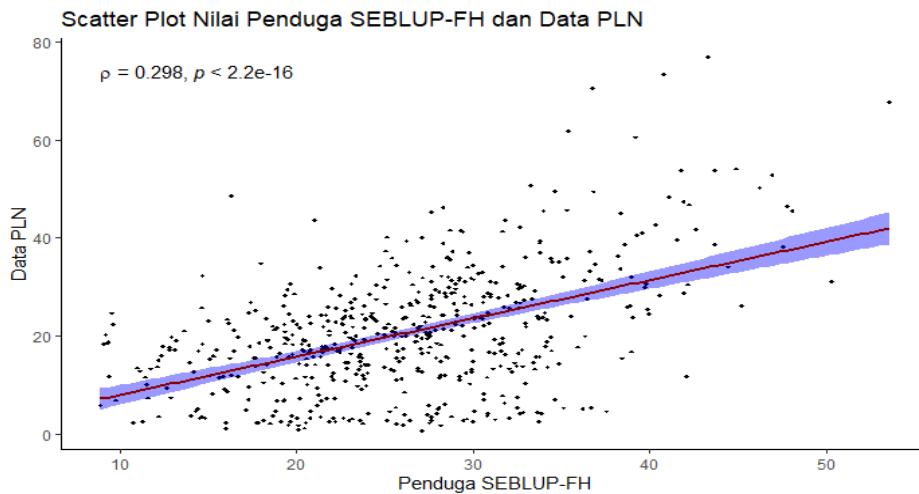
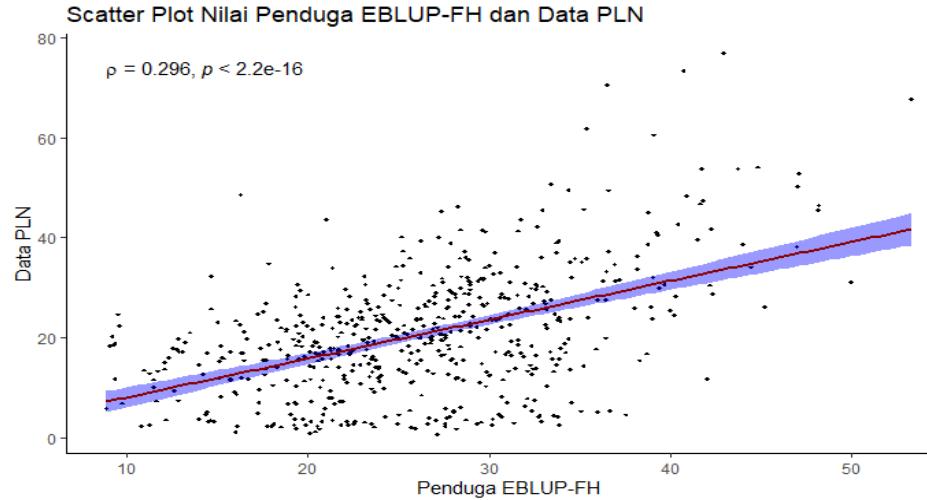
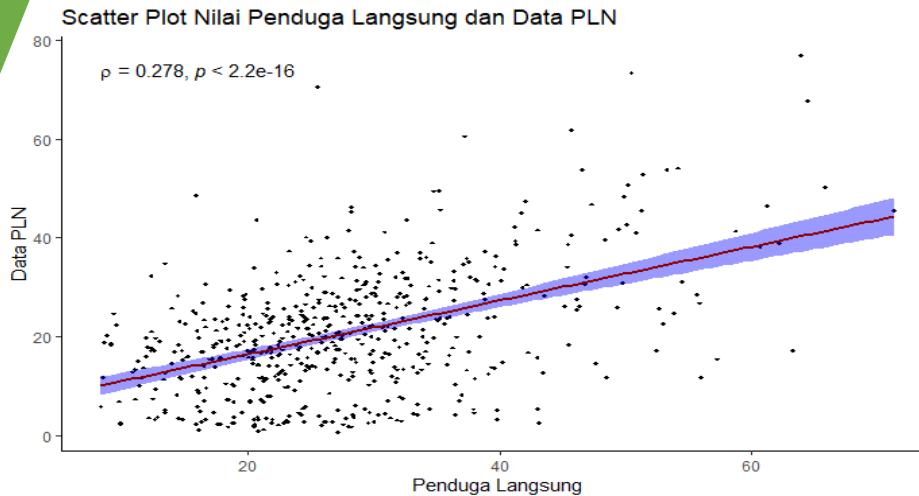
- Penduga EBLUP-FH
- Penduga HB Lognormal
- Penduga Langsung
- Penduga SEBLUP-FH



Penduga HB Lognormal secara statistik memiliki karakteristik nilai yang paling mendekati nilai penduga langsung

Tujuan 3: Evaluasi Hasil Model

A. Pemilihan Metode Terbaik: Perbandingan dengan data PLN

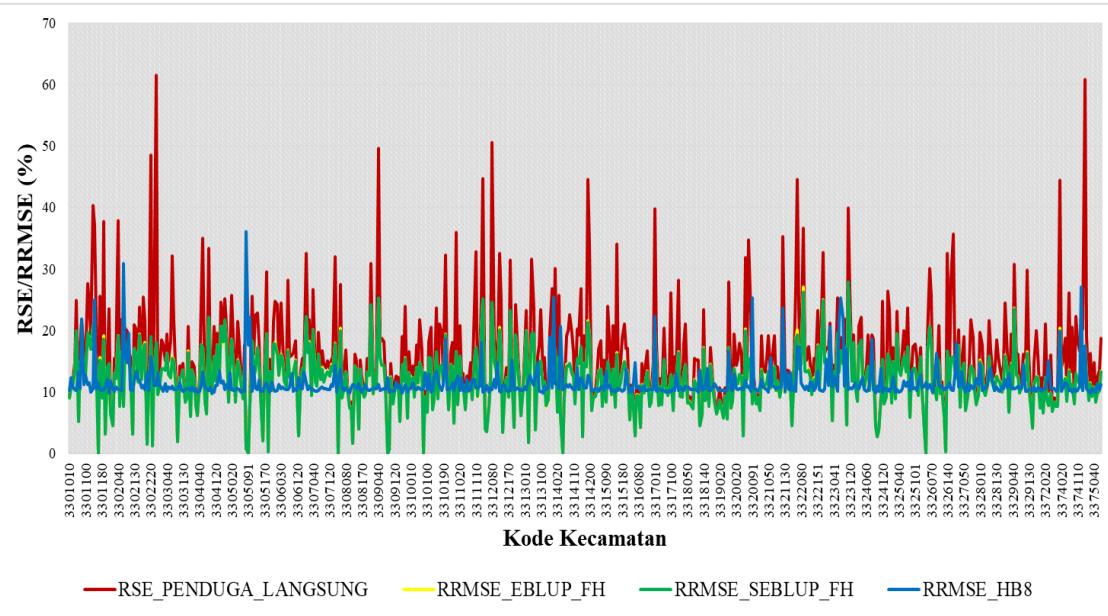


- Nilai korelasi Kendall-Tau:
- Pend.Lgsg & data PLN: 0,278
 - EBLUP-FH & data PLN: 0,296
 - SEBLUP-FH & data PLN: 0,298
 - HB Lognormal & data PLN: 0,3

Penduga HB Lognormal memiliki nilai korelasi Kendall-Tau paling tinggi terhadap data PLN

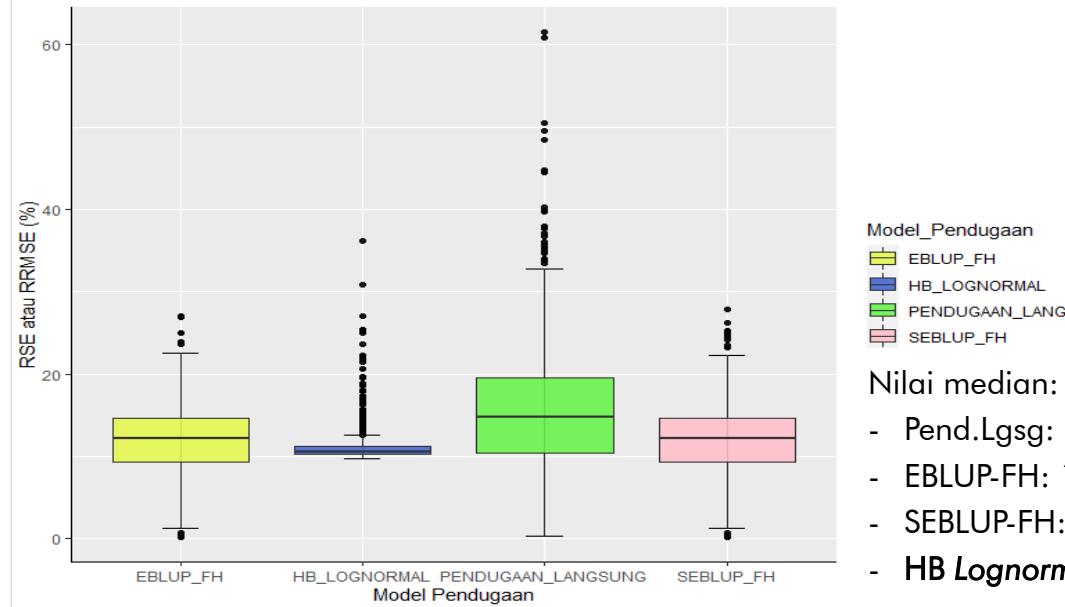
Tujuan 3: Evaluasi Hasil Model

B. Pemilihan Metode Terbaik: *Relative Standard Error (RSE)/Relative Root Mean Squared Error (RRMSE)*



Jumlah kecamatan menurut kategori nilai RSE/RRMSE dari empat penduga

Kategori nilai RSE atau RRMSE	Jumlah observasi (kecamatan)			
	Penduga langsung	Penduga EBLUP-FH	Penduga SEBLUP-FH	Penduga HB Lognormal
RSE = 0%	7	7	7	0
0% < RSE ≤ 25%	500	561	561	566
25% < RSE ≤ 50%	62	4	4	6
RSE > 50%	3	0	0	0



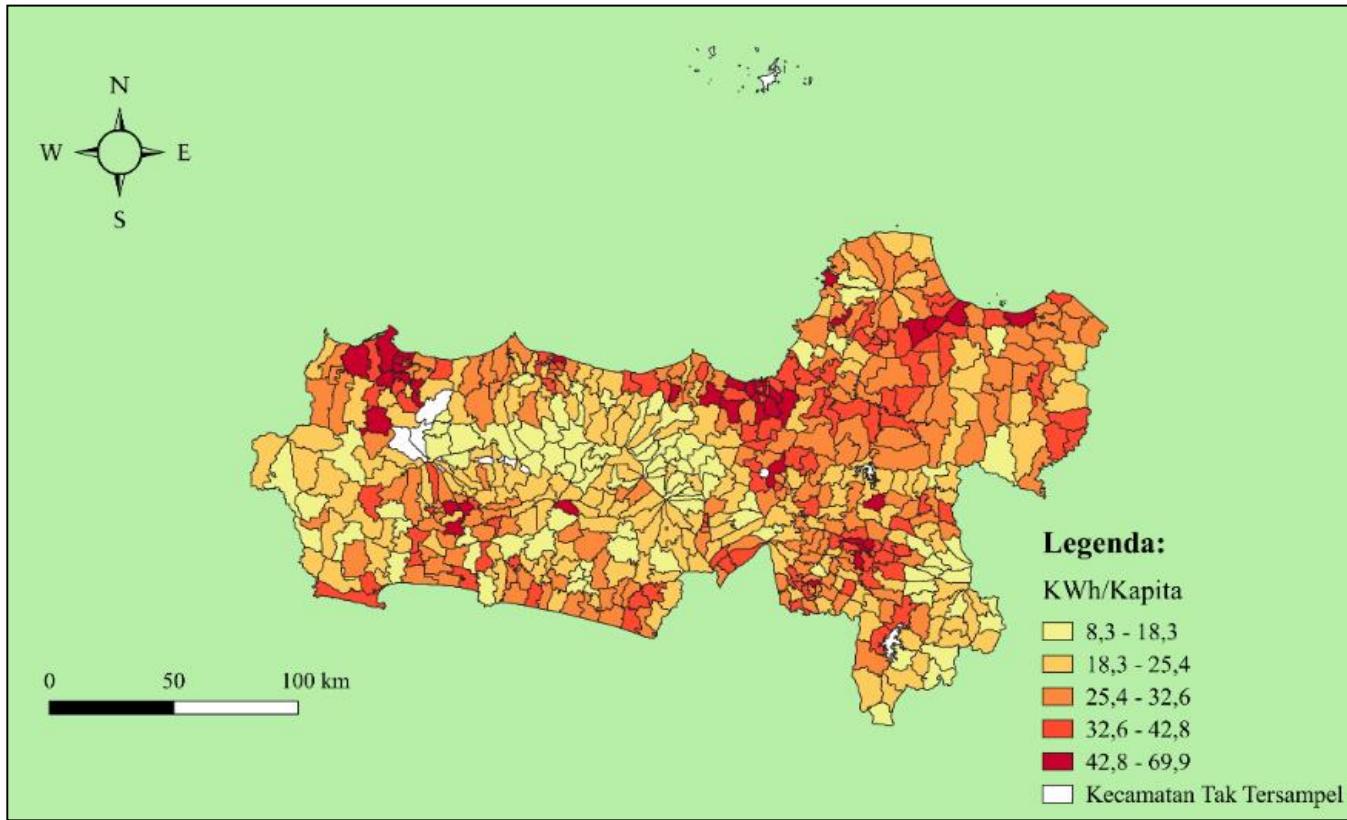
- Nilai median:
- Pend.Lgsg: 14,76%
 - EBLUP-FH: 12,18%
 - SEBLUP-FH: 12,17%
 - HB Lognormal: 10,55%

- Jumlah kecamatan dengan $RSE > 25\%$ dari penduga HB Lognormal masih lebih banyak dari penduga EBLUP-FH dan SEBLUP-FH.
- Nilai median RRMSE HB Lognormal $<$ RRMSE EBLUP-FH $<$ RRMSE SEBLUP-FH $<$ RSE pendugaan langsung.



Tujuan 4: Pemetaan Hasil

Pemetaan Hasil Estimasi Metode Terbaik: Model SAE HB *Lognormal*

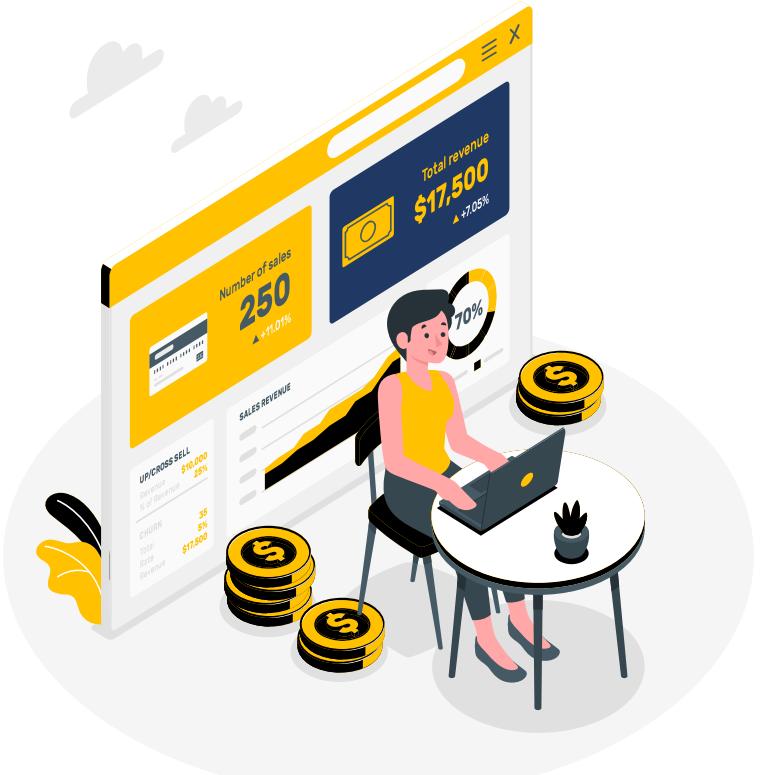


Peta tematik nilai penduga HB *Lognormal*

- Tingkat rata-rata konsumsi listrik per kapita ruta level kecamatan di Jawa Tengah tahun 2021 sebagian besar berada pada rentang antara 18,3 hingga 25,4 KWh/Kapita.
- Kecamatan Gajah Mungkur tertinggi dengan 69,91 KWh/Kapita.
- Kecamatan Petungkriono terendah dengan 8,26 KWh/Kapita.
- Wilayah kecamatan dengan tingkat konsumsi listrik ruta yang relatif tinggi sebagian besar berada di daerah perkotaan dengan tingkat kesejahteraan masyarakat yang juga relatif tinggi, vice versa.

Tautan Dashboard Informasi Hasil Skripsi: <https://raflizal17.github.io/dashboard-hasil-skripsi/>





PENUTUP

1. Kesimpulan Penelitian

1

Penduga langsung memiliki nilai median RSE sebesar 14,76% dengan karakteristik data berdistribusi *lognormal* dan terdapat dependensi spasial antarwilayah kecamatan.

2

Penduga HB *Lognormal* memiliki karakteristik nilai yang paling mendekati nilai penduga langsung dengan korelasi Kendall-Tau sebesar 0,9334.

3

Model SAE HB *Lognormal* dipilih menjadi metode terbaik dalam mengestimasi nilai rata-rata konsumsi listrik per kapita ruta level kecamatan berdasarkan hasil evaluasi.

4

Hasil pemetaan penduga HB *Lognormal* menunjukkan bahwa sebagian besar wilayah kecamatan dengan tingkat konsumsi listrik ruta yang relatif tinggi berada di daerah perkotaan dengan tingkat kesejahteraan masyarakat yang juga relatif tinggi, begitu juga sebaliknya.



2. Saran Penelitian

1

BPS dapat mengintegrasikan pemanfaatan sumber *big data* dan data Podes dengan penggunaan metode SAE HB *Lognormal* untuk memperbaiki hasil estimasi Susenas pada indikator yang memiliki tipe variabel berbentuk rill dan positif.

2

Bagi PT. PLN & Kementerian ESDM, hasil estimasi dari model HB *Lognormal* dapat dijadikan sebagai bahan informasi terkait wilayah kecamatan yang masih memerlukan dukungan infrastruktur kelistrikan sehingga pemanfaatan energi listrik dapat terserap dengan optimal.

3

Untuk penelitian selanjutnya, model SAE HB dengan memasukkan pengaruh spasial dapat diimplementasikan, terutama pada saat kondisi data variabel *interest* yang tidak berdistribusi normal.



POLITEKNIK STATISTIKA STIS

Sidang Skripsi Prodi Komputasi Statistik T.A. 2022/2023

TERIMA KASIH

Pemanfaatan Sumber *Big Data* dalam Memperkirakan
Konsumsi Listrik Per Kapita Rumah Tangga
Menggunakan *Small Area Estimation*
(Studi Kasus: Level Kecamatan di Provinsi Jawa Tengah 2021)

Raflizal Fikrar Odriansyah (221910812/4SD2)

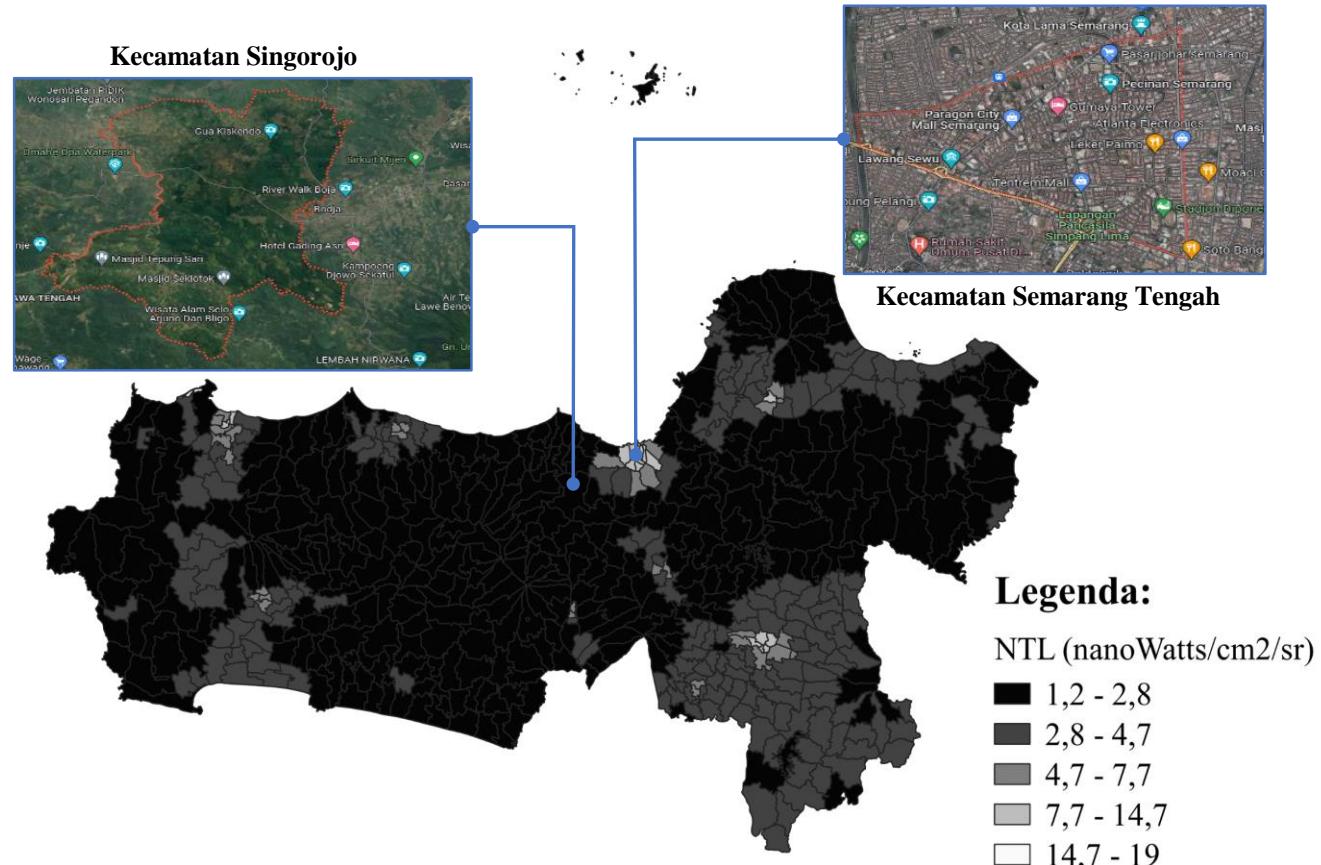


LAMPIRAN

Hasil Ekstraksi Fitur Sumber Big Data

A. *Nighttime Light Intensity*

Intensitas Cahaya Malam (nanoWatts/cm²/sr)
Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2021



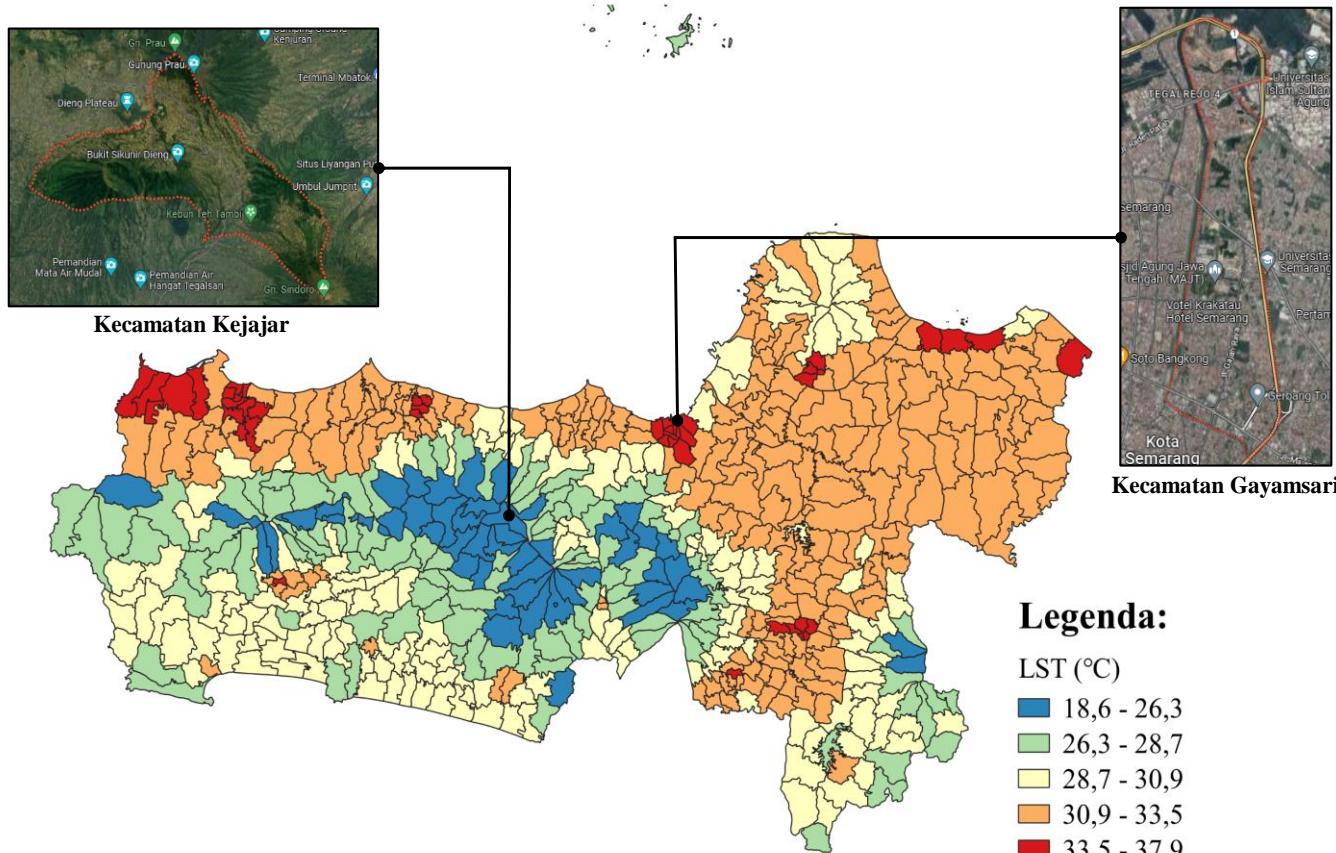
Nilai intensitas cahaya malam tertinggi sebesar 19,035 nanoWatts/cm²/sr berada di Kecamatan Semarang Tengah, Kota Semarang. Sedangkan nilai intensitas cahaya malam terendah sebesar 1,3725 nanoWatts/cm²/sr yang berada di Kecamatan Singorojo, Kabupaten Kendal.



B. Land Surface Temperature

Land Surface Temperature (°C)

Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2021

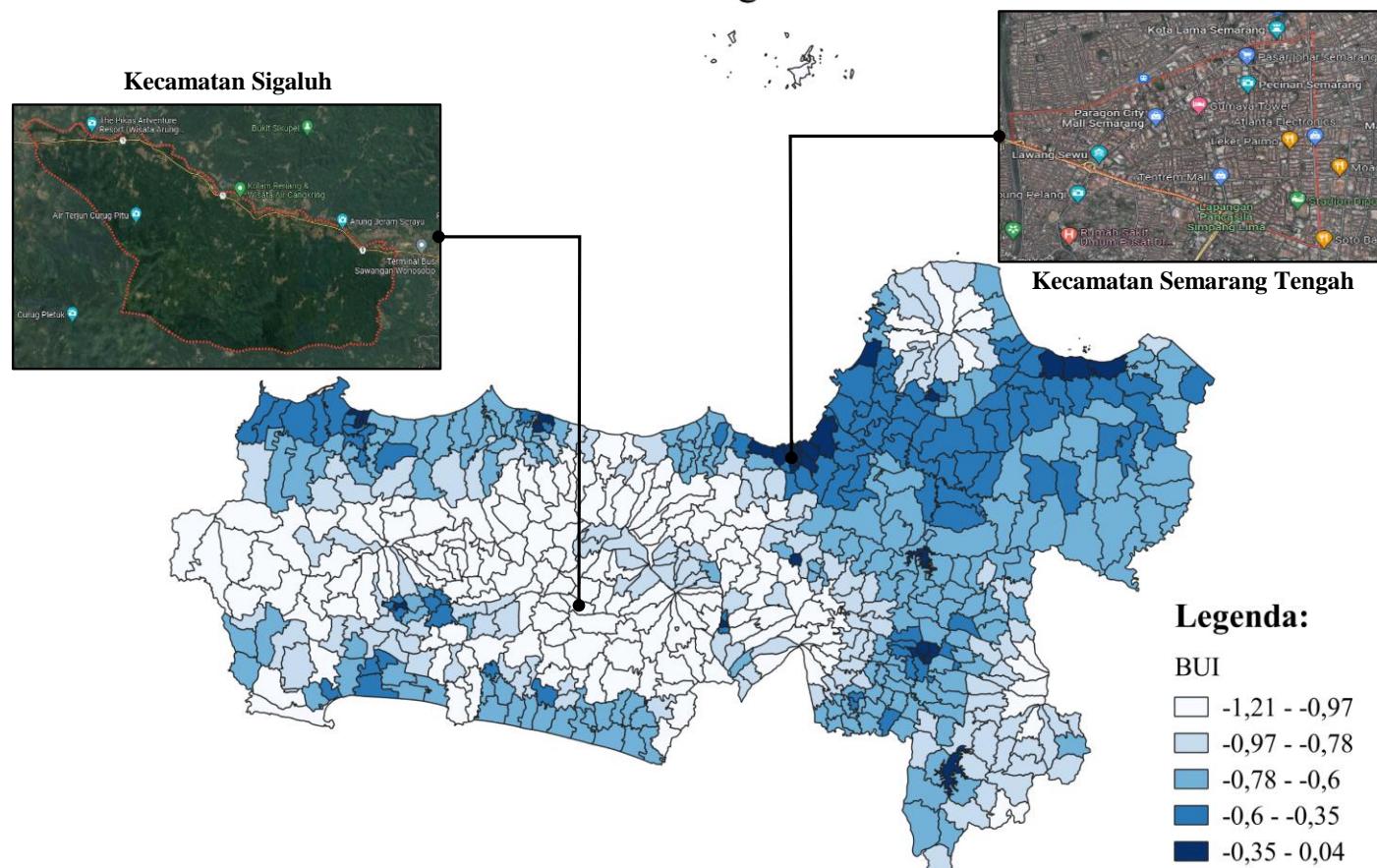


Nilai suhu permukaan tanah tertinggi yakni sebesar 37,88°C berada di Kecamatan Gayamsari, Kota Semarang. Sedangkan suhu permukaan tanah terendah yakni sebesar 22,64°C yang berada di Kecamatan Kejajar, Kabupaten Wonosobo.

Hasil Ekstraksi Fitur Sumber Big Data

C. Build-Up Index

Build-Up Index Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2021

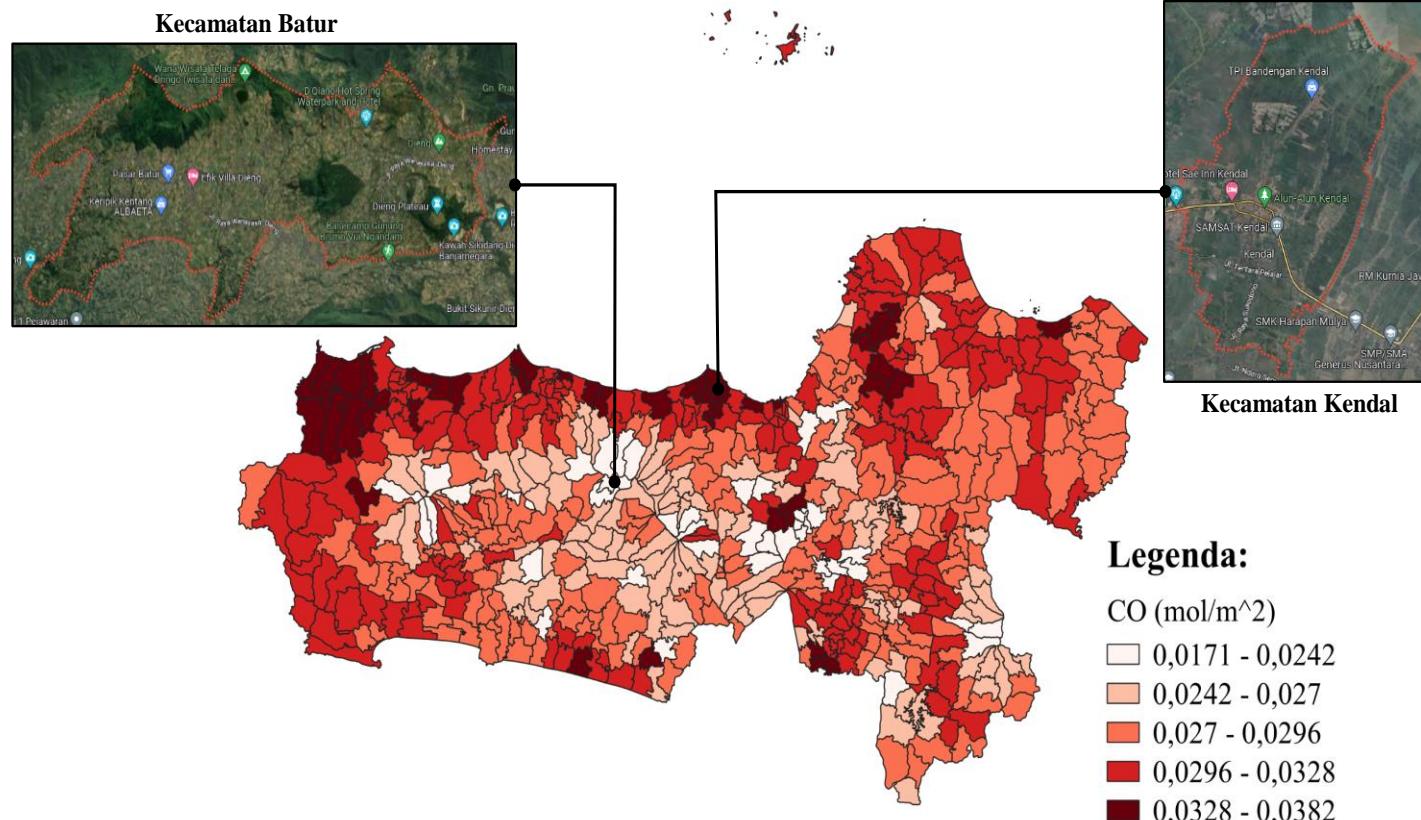


Nilai BUI tertinggi yakni sebesar 0,03596 berada di Kecamatan Semarang Tengah, Kota Semarang. Sedangkan nilai BUI terendah yakni sebesar -1,20879 yang berada di Kecamatan Sigaluh, Kabupaten Banjarnegara.

D. Kepadatan Senyawa CO

Kepadatan Senyawa CO (mol/m^2)

Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2021

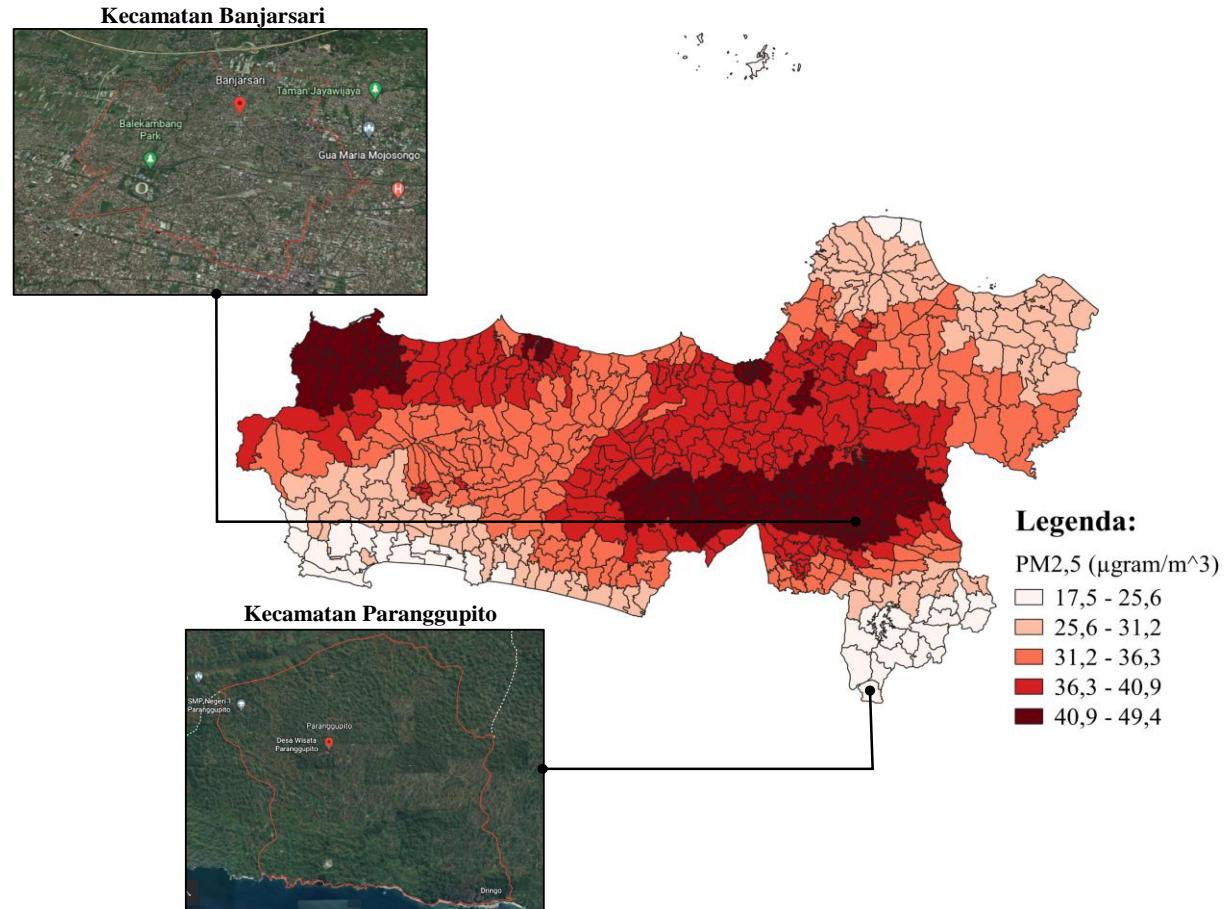


Nilai kepadatan senyawa CO tertinggi yakni sebesar $0,03821 \text{ mol}/\text{m}^2$ yang berada di Kecamatan Kendal, Kabupaten Kendal. Sedangkan kepadatan senyawa CO terendah yakni sebesar $0,01706 \text{ mol}/\text{m}^2$ yang berada di Kecamatan Batur, Kabupaten Banjarnegara.

Hasil Ekstraksi Fitur Sumber Big Data

E. Konsentrasi PM_{2.5}

Konsentrasi Particulate Matter 2,5 (PM_{2.5})
Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2019

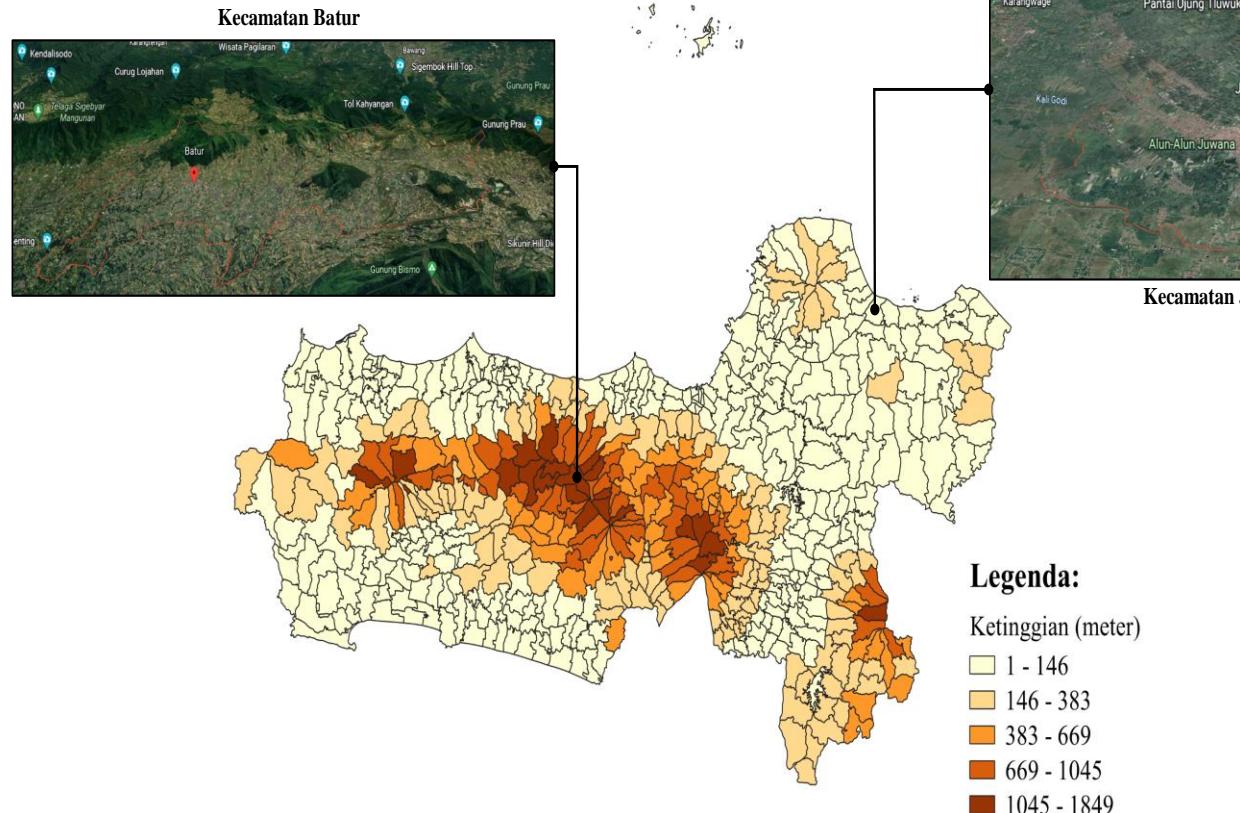


Nilai konsentrasi PM_{2.5} tertinggi yakni sebesar 49,4 $\mu\text{gram}/\text{m}^3$ yang berada di Kecamatan Banjarsari, Kota Surakarta. Sedangkan nilai konsentrasi PM_{2.5} terendah yakni sebesar 17,5 $\mu\text{gram}/\text{m}^3$ yang berada di Kecamatan Paranggupito, Kabupaten Wonogiri.

Hasil Ekstraksi Fitur Sumber Big Data

F. Elevation/Ketinggian

Elevation (meter) Menurut Kecamatan di Provinsi Jawa Tengah Tahun 2000



Nilai ketinggian terbesar yakni 1.816 meter yang berada di Kecamatan Batur., Kabupaten Banjarnegara Sedangkan nilai ketinggian terendah yaitu 1 meter yang berada di Kecamatan Juwana, Kabupaten Pati.

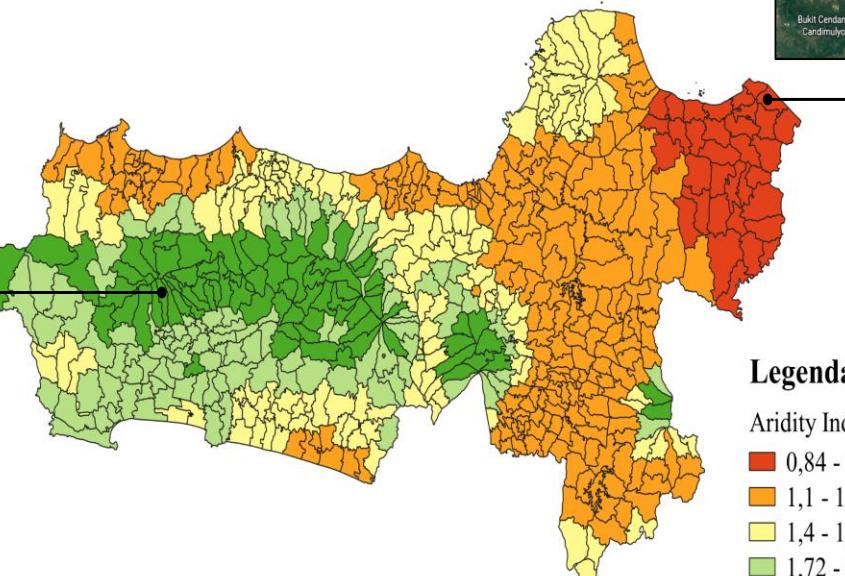
Hasil Ekstraksi Fitur Sumber Big Data

G. Aridity Index (Indeks Kekeringan)

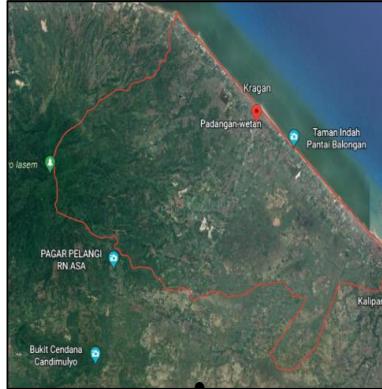
Kecamatan Baturraden



Global Aridity Index (Global-AI) Menurut Kecamatan di Provinsi Jawa Tengah Tahun 1970-2000



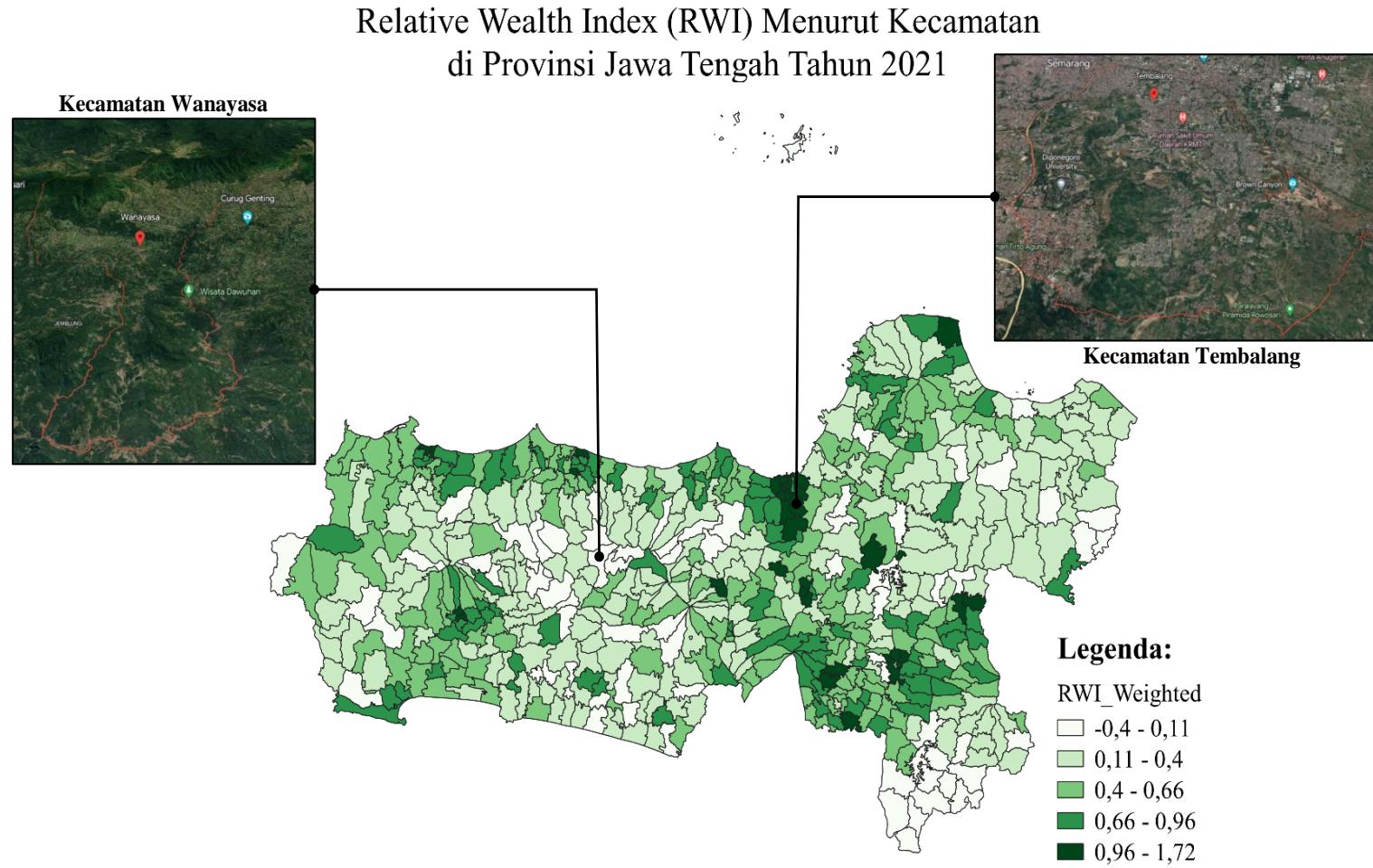
Kecamatan Kragan



Nilai *aridity index* tertinggi sebesar 2,7325 yang berada di Kecamatan Baturraden, Kabupaten Banyumas. Sedangkan nilai *aridity index* terendah sebesar 0,8404 yang berada di Kecamatan Kragan, Kabupaten Rembang.



H. Relative Wealth Index (Indeks Kekayaan Relatif)



Nilai RWI tertinggi sebesar 1,7157 yang berada di Kecamatan Tembalang, Kota Semarang. Sedangkan nilai RWI terendah sebesar -0,4047 yang berada di Kecamatan Wanayasa, Kabupaten Banjarnegara.

Hasil Ekstraksi Data PODES

Statistik Deskriptif Variabel PODES 2021

Variabel	Minimum	Maksimum	Rata-Rata	Median	Standar Deviasi
X_1	3516	61913	21474	19732	9729,44
X_2	31	405	142,6	133,5	56,164
X_3	11	219	56,86	50	29,499
X_4	50	826	294,6	273,5	122,274
X_5	60	9152	883,8	558,5	1009,352
X_6	0	54	2,57	1	4,287
X_7	28	3322	854	789,5	437,237
X_8	117	16094,02	1747,448	1121,47	2063,243

Keterangan:

Jumlah keluarga pengguna listrik PLN (X_1), jumlah sarana pendidikan (X_2), jumlah sarana kesehatan (X_3), jumlah tempat ibadah (X_4), jumlah industri mikro dan kecil (X_5), jumlah sentra industri (X_6), jumlah sarana dan prasarana ekonomi (X_7), dan kepadatan penduduk (X_8)



Sampling Scheme SUSENAS Maret 2021

Tahap/ Phase	Unit	Jumlah unit strata h		Metode penarikan sampel	Peluang pemilihan sampel	Fraksi sampling
		Populasi	Sampel			
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	Blok sensus	N_{kh}	n'_{kh}	PPS-with replacement	$\frac{M_{khi}}{M_{kh}}$	$n'_{kh} \frac{M_{khi}}{M_{kh}}$
		n'_{kh}	n_{kh}	Systematic	$\frac{1}{n'_{kh}}$	$\frac{n_{kh}}{n'_{kh}}$
2	Rumah tangga	M^{up}_{khi}	\bar{m}	Systematic	$\frac{1}{M^{up}_{khi}}$	$\frac{\bar{m}}{M^{up}_{khi}}$

Sumber: Metodologi sampling Susenas Maret 2021

Master *sampling frame*:

1. Kerangka sampel tahap pertama: daftar blok sensus biasa SP2020 dan daftar 40 persen blok sensus SP2020 yang sudah terdapat informasi kode strata.
2. Kerangka sampel tahap kedua: daftar rumah tangga hasil pendaftaran (*listing*) ruta di setiap blok sensus terpilih.

Prosedur penarikan sampel:

1. Tahap 1: Memilih 40 persen blok sensus populasi secara PPS dengan size jumlah KK hasil Rekap Daftar (RD) Penduduk SP2020 di setiap strata di kabupaten. Kemudian memilih sejumlah n blok sensus sesuai alokasi secara *systematic* di setiap strata urban/rural per kab/kota.
2. Tahap 2: Memilih 10 ruta hasil pendaftaran (*listing*) secara *systematic* dengan *implicit stratification* menurut pendidikan Kepala Rumah Tangga (KRT).



Teori Hubungan Variabel

- ❑ Penelitian dari Alfiany (2022) yang berjudul "*Hubungan Kausalitas antara Konsumsi Listrik, Pertumbuhan Ekonomi, Ketimpangan dan Kemiskinan di Indonesia Tahun 1971-2019*" memberikan hasil kesimpulan bahwa peningkatan atau penurunan konsumsi listrik memiliki hubungan kausalitas satu arah dan jangka panjang dengan peningkatan atau penurunan kemiskinan dan ketimpangan di Indonesia ➔ ***Dasar digunakannya pendekatan kemiskinan dan ketimpangan dalam pemilihan kandidat peubah penyerta dari data PODES***
- ❑ Penelitian dari Titin Siswantining (2013) meninjau permasalahan kemiskinan ke dalam empat aspek yaitu:

Aspek	Kandidat Peubah Penyerta dari Data Podes
Infrastruktur	1. Jumlah keluarga pengguna listrik PLN 2. Jumlah tempat ibadah
Pendidikan	1. Jumlah sarana pendidikan
Kesehatan	1. Jumlah sarana kesehatan
Ekonomi	1. Jumlah industri mikro dan kecil 2. Jumlah sentra industri 3. Jumlah sarana dan prasarana ekonomi



Teori Hubungan Variabel

Landasan teori pemilihan variabel berdasarkan penelitian terdahulu

Variabel	Referensi	Penjelasan
<i>Nighttime Light Intensity</i> (NTL)	Ngestrini (2019), Shi et al (2014)	Data intensitas cahaya malam (NTL) telah banyak digunakan untuk mempelajari aktivitas ekonomi di suatu wilayah (Ngestrini, 2019). Semakin tinggi intensitas cahaya malam di suatu wilayah, maka semakin tinggi tingkat elektrifikasi di wilayah tersebut. Tingkat elektrifikasi yang tinggi sejalan dengan meningkatnya kemampuan masyarakat dalam memenuhi kebutuhan listriknya (Shi et al, 2014). Oleh karena itu, intensitas cahaya malam yang tinggi di suatu wilayah sejalan dengan tingginya konsumsi listrik di wilayah tersebut.
<i>Build-Up Index</i> (BUI)	Ouyang et al. (2016), Faisal et al. (2016)	Nilai BUI yang tinggi dan positif menandakan bahwa suatu wilayah merupakan kawasan terbangun dengan tutupan lahan yang rendah (Ouyang et al., 2016). Salah satu faktor yang memengaruhi nilai BUI tinggi yaitu tingginya program pembangunan di suatu wilayah sehingga menaikkan aktivitas ekonomi penduduk yang dapat berdampak pada peningkatan kesejahteraan masyarakat (Faisal et al., 2016).



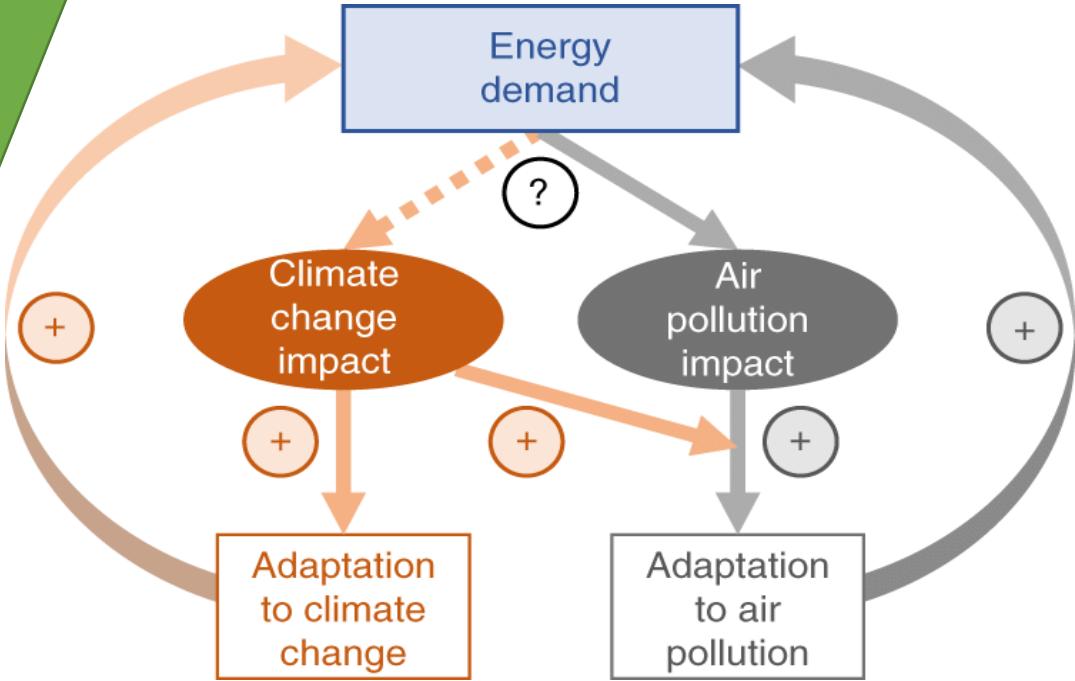
Teori Hubungan Variabel

Landasan teori pemilihan variabel berdasarkan penelitian terdahulu

Variabel	Referensi	Penjelasan
<i>Land Surface Temperature</i> (LST)	Rajeshwari & Mani (2014), Ayuningtyas et al. (2015)	LST dapat menangkap fenomena perubahan iklim global dimana seiring meningkatnya kandungan gas rumah kaca , maka nilai LST juga meningkat (Rajeshwari & Mani, 2014). Sumber gas rumah kaca dapat dihasilkan melalui emisi yang diperoleh dari penggunaan listrik (Ayuningtyas et al, 2015).
<i>Elevation</i> (Ketinggian)	Luo et al. (2019), Katsoulakos & Kaliampakos (2014)	Penggunaan DEM secara spesifik banyak diterapkan untuk berbagai aplikasi, salah satunya untuk memodelkan distribusi populasi penduduk dan pengaruhnya terhadap permintaan energi. Luo et al. (2019) menyatakan bahwa rata-rata elevasi berkorelasi negatif dengan kepadatan populasi di Tiongkok . Apabila kepadatan populasi diasumsikan berkorelasi positif dengan tingkat konsumsi listrik, maka semakin tinggi suatu wilayah akan semakin rendah tingkat konsumsi listrik rumah tangga di wilayah tersebut .



Teori Hubungan Variabel

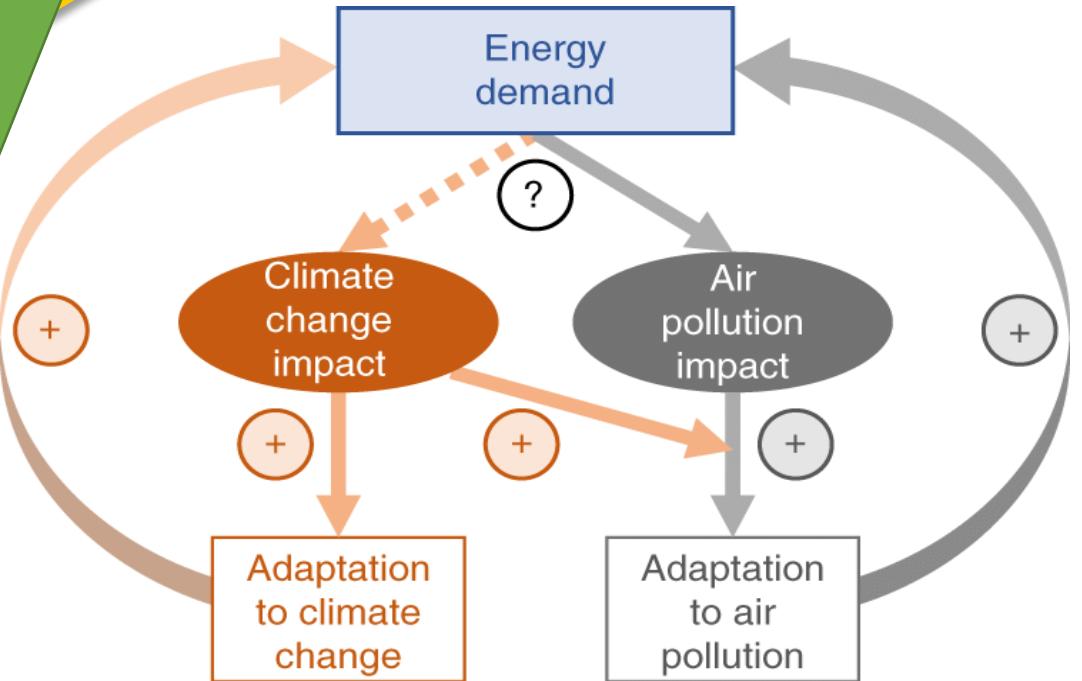


Siklus permintaan energi, dampak perubahan iklim dan polusi udara, dan hasil adaptasi intensif energi
Sumber: Jiyeon Eom et al. (2020)

Landasan teori pemilihan variabel berdasarkan penelitian terdahulu

Variabel	Referensi	Penjelasan
Polusi Udara: 1. Kepadatan Senyawa CO 2. Konsentrasi PM2.5	Eom et al. (2020), T. Liu et al. (2018), Saberian et al. (2017).	Polusi udara dapat memengaruhi perilaku penggunaan energi listrik rumah tangga berupa respons dan tindakan adaptasi yang membutuhkan penggunaan energi seperti menyalakan pengering mekanis untuk menghilangkan polutan di dalam ruangan (Liu et al, 2018). Polusi udara juga memicu perubahan gaya hidup , misalnya seseorang membatasi aktivitas di luar ruangan sehingga memerlukan penggunaan energi listrik lebih untuk melakukan kegiatan di dalam ruangan (Saberian et al, 2017)

Teori Hubungan Variabel



Siklus permintaan energi, dampak perubahan iklim dan polusi udara, dan hasil adaptasi intensif energi

Sumber: Jiyeon Eom et al. (2020)

Landasan teori pemilihan variabel berdasarkan penelitian terdahulu

Variabel	Referensi	Penjelasan
<i>Global Aridity Index (Global-AI)</i>	Zomer et al. (2022), Eom et al. (2020).	Data Global-AI dapat mendukung studi yang berkontribusi pada pembangunan berkelanjutan, pengentasan kemiskinan, dan adaptasi terhadap perubahan iklim (Zomer & Trabucco, 2022). Dampak perubahan iklim dapat menciptakan perilaku baru sebagai adaptasi dari perubahan iklim dan polusi udara sehingga memengaruhi penggunaan energi listrik bagi rumah tangga (Jiyeon Eom et al, 2020).

Skema klasifikasi iklim umum *aridity index value*

Sumber: UNEP (1997)

Aridity Index Value	Climate Class
< 0,03	Hyper-Arid
0,03 - 0,2	Arid
0,2 - 0,5	Semi-Arid
0,5 - 0,65	Dry sub-humid
> 0,65	Humid



Teori Hubungan Variabel

Landasan teori pemilihan variabel berdasarkan penelitian terdahulu

Variabel	Referensi	Penjelasan
<i>Relative Wealth Index</i> (RWI)	Ferguson et al. (2000), Chi et al. (2022).	Terdapat korelasi kuat antara penciptaan kekayaan (<i>wealth creation</i>) terhadap penggunaan listrik di negara-negara maju dan beberapa negara berkembang. Di Indonesia, korelasi antara penggunaan listrik per kapita dengan penciptaan kekayaan sebesar 0,986. Penciptaan kekayaan mengacu pada data GDPPP/kapita tahun 1971-1995 (Ferguson R. et al, 2000). RWI mengindikasikan tingkat kekayaan dari suatu lokasi yang relatif dengan lokasi lainnya dalam satu negara yang sama (Guanhua Chi et al, 2022).

*Note: GDPPP perkapita adalah nilai GDP perkapita dalam USD (selanjutnya disebut GDP perkapita saja) yang telah dikonversi menggunakan bobot *purchasing power parity* (PPP).



Metode Analisis Data

Model SAE *Empirical Best Linear Unbiased Prediction Fay Herriot* (EBLUP-FH)

Penduga EBLUP-FH dirumuskan sebagai berikut:

$$\hat{\theta}_i^{EBLUP} = \mathbf{x}_i^T \hat{\beta} + \hat{\gamma}_i (\hat{\theta}_i - \mathbf{x}_i^T \hat{\beta}); \text{ dimana } \hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \psi_i} \text{ dan}$$
$$\hat{\beta} = \sum_{i=1}^m (\mathbf{x}_i (\hat{\sigma}_v^2 z_i^2 + \psi_i)^{-1} \mathbf{x}_i^T)^{-1} \sum_{i=1}^m \mathbf{x}_i (\hat{\sigma}_v^2 z_i^2 + \psi_i)^{-1} \hat{\theta}_i$$

MSE EBLUP-FH dirumuskan sebagai berikut:

$$MSE(\hat{\theta}_i^{EBLUP}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2); g_{1i}(\hat{\sigma}_v^2) = \frac{\hat{\sigma}_v^2 \psi_i}{\hat{\sigma}_v^2 + \psi_i} = \hat{\gamma}_i \psi_i$$
$$g_{2i}(\hat{\sigma}_v^2) = (1 - \hat{\gamma}_i)^2 \mathbf{x}_i^T \left[\sum_{i=1}^m \frac{\mathbf{x}_i \mathbf{x}_i^T}{\hat{\sigma}_v^2 + \psi_i} \right]^{-1} \mathbf{x}_i; g_{3i}(\hat{\sigma}_v^2) = \psi_i^2 \mathbf{x}_i^4 (\hat{\sigma}_v^2 z_i^2 + \psi_i)^{-3} \bar{V}(\sigma_v^2)$$

Keterangan:

\mathbf{x}_i : vektor *covariate* unit area ke- i

$\hat{\beta}$: vektor koefisien regresi

$\hat{\sigma}_v^2$: penduga ragam pengaruh acak area

ψ_i : ragam galat acak unit area ke- i

z_i : konstanta positif yang diketahui

$\hat{\theta}_i$: penduga langsung unit area ke- i

$\bar{V}(\sigma_v^2)$: varians asimtot dari $\hat{\sigma}_v^2$

dengan $\bar{V}(\sigma_v^2) = 2m^{-2} \sum_{i=1}^m (\hat{\sigma}_v^2 + \psi_i)^2$ dan

m : jumlah observasi unit area



Metode Analisis Data

Model SAE *Spatial Empirical Best Linear Unbiased Prediction Fay Herriot* (SEBLUP-FH)

Penduga SEBLUP-FH dirumuskan sebagai berikut:

$$\begin{aligned}\hat{\theta}_i^{SEBLUP}(\hat{\sigma}_v^2, \hat{\rho}) &= \mathbf{x}_i^T \hat{\beta} + \mathbf{b}_i^T \{ \hat{\sigma}_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \} \mathbf{Z}^T \\ &\times \{ \text{diag}(\psi_i) + \mathbf{Z} \hat{\sigma}_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{Z}^T \}^{-1} (\hat{\theta} - X \hat{\beta})\end{aligned}$$

MSE SEBLUP-FH dirumuskan sebagai berikut:

$$MSE(\hat{\theta}_i^{SEBLUP}, (\hat{\sigma}_v^2, \hat{\rho})) \approx g_{1i}(\hat{\sigma}_v^2, \hat{\rho}) + g_{2i}(\hat{\sigma}_v^2, \hat{\rho}) + 2g_{3i}(\hat{\sigma}_v^2, \hat{\rho});$$

dimana:

$$g_{1i}(\hat{\sigma}_v^2, \hat{\rho}) = \mathbf{b}_i^T \{ \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \} \mathbf{Z}^T \times \{ \text{diag}(\psi_i) + \mathbf{Z} \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{Z}^T \}^{-1} \mathbf{Z} \sigma_v^2 \times [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{b}_i;$$

$$g_{2i}(\hat{\sigma}_v^2, \hat{\rho}) = (\mathbf{x}_i - \mathbf{b}_i^T \{ \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \} \mathbf{Z}^T \times \{ \text{diag}(\psi_i) + \mathbf{Z} \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{Z}^T \}^{-1} \mathbf{X}) \times (\mathbf{X}^T \{ \text{diag}(\psi_i) + \mathbf{Z} \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{Z}^T \}^{-1} \mathbf{X})^{-1} \times (\mathbf{x}_i - \mathbf{b}_i^T \{ \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \} \mathbf{Z}^T \times \{ \text{diag}(\psi_i) + \mathbf{Z} \sigma_v^2 [(\mathbf{I} - \hat{\rho} \mathbf{W})(\mathbf{I} - \hat{\rho} \mathbf{W})^T]^{-1} \mathbf{Z}^T \}^{-1} \mathbf{X})^T; \text{ dan}$$

$$\begin{aligned}g_{3i}(\hat{\sigma}_v^2, \hat{\rho}) &= \text{tr} \left\{ \begin{bmatrix} \mathbf{b}_i^T (\mathbf{C}^{-1} \mathbf{Z}^T \mathbf{V}^{-1} + \sigma_u^2 \mathbf{C}^{-1} \mathbf{Z}^T (-\mathbf{V}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^T \mathbf{V}^{-1})) \\ \mathbf{b}_i^T \mathbf{A} \mathbf{Z}^T \mathbf{V}^{-1} + \sigma_u^2 \mathbf{C}^{-1} \mathbf{Z}^T (-\mathbf{V}^{-1} \mathbf{Z} \mathbf{A} \mathbf{Z}^T \mathbf{V}^{-1}) \end{bmatrix}_{\mathbf{V}} \right. \\ &\quad \left. \times \begin{bmatrix} \mathbf{b}_i^T (\mathbf{C}^{-1} \mathbf{Z}^T \mathbf{V}^{-1} + \sigma_u^2 \mathbf{C}^{-1} \mathbf{Z}^T (-\mathbf{V}^{-1} \mathbf{Z} \mathbf{C}^{-1} \mathbf{Z}^T \mathbf{V}^{-1})) \\ \mathbf{b}_i^T \mathbf{A} \mathbf{Z}^T \mathbf{V}^{-1} + \sigma_u^2 \mathbf{C}^{-1} \mathbf{Z}^T (-\mathbf{V}^{-1} \mathbf{Z} \mathbf{A} \mathbf{Z}^T \mathbf{V}^{-1}) \end{bmatrix}^T \bar{V}(\hat{\sigma}_v^2, \hat{\rho}) \right\}\end{aligned}$$

Keterangan:

\mathbf{x}_i : vektor covariate unit area ke- i

$\hat{\beta}$: vektor koefisien regresi

\mathbf{b}_i^T : vektor berukuran $1 \times n$

$\hat{\sigma}_v^2$: penduga ragam pengaruh acak area

ψ_i : ragam galat acak unit area ke- i

\mathbf{W} : matriks pembobot spasial

$\hat{\rho}$: koefisien spatial autoregressive

\mathbf{Z} : matriks desain variabel prediktor

$\hat{\theta}$: vektor penduga langsung

\mathbf{V} : matriks kovarians dari $\hat{\theta}$

$\bar{V}(\hat{\sigma}_v^2, \hat{\rho})$: varians asimtot dari $\hat{\sigma}_v^2$ dan $\hat{\rho}$



Metode Analisis Data

Model SAE *Hierarchical Bayes* (HB)

Rumus besaran posterior rata-rata dan standar deviasi:

$$\hat{E}(G(\boldsymbol{\theta})|y) = \frac{1}{T'} \sum_{t=1}^{T'} G(\boldsymbol{\theta}^{(t)}) ; \text{ dimana } T' = \frac{T - B}{L}$$

$$\widehat{SD}(G(\boldsymbol{\theta})|y) = \sqrt{\frac{1}{T' - 1} \sum_{t=1}^{T'} [G(\boldsymbol{\theta}^{(t)}) - \hat{E}(G(\boldsymbol{\theta})|y)]^2}$$

MSE dari penduga HB dirumuskan sebagai berikut:

$$MSE(G(\boldsymbol{\theta})|y) = (\widehat{SD}(G(\boldsymbol{\theta})|y))^2$$

RRMSE dari penduga HB dirumuskan sebagai berikut:

$$RRMSE(\hat{\theta}_i^{HB}) = \frac{\sqrt{(V[G(\boldsymbol{\theta})|y])}}{\phi_i^{HB}}$$

Keterangan:

$G(\boldsymbol{\theta}^{(t)})$: fungsi parameter $\boldsymbol{\theta}$ yang diduga pada sampel ke- t

T' : jumlah sampel akhir yang dipertahankan

T : total iterasi yang dilalui pada algoritma MCMC

B : *burn in*

L : *thinning interval* atau *sampling lag*

$V[G(\boldsymbol{\theta})|y]$: varians posterior

ϕ^{HB} : penduga HB

dimana $\phi^{HB} = E[h(\boldsymbol{\theta})|y]$: rata-rata posterior dan

$V[G(\boldsymbol{\theta})|y] = MSE(G(\boldsymbol{\theta})|y)$



SAE HB LOGNORMAL

i. Sampling model

$$\hat{\theta}_i | \theta_i \sim^{id} lognormal(\mu_i, \psi_i) \quad i = 1, 2, \dots, m$$

di mana $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta} + \nu_i$ dengan $\nu_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\nu^2)$ dan ψ_i adalah ragam dari distribusi *lognormal* dengan $\psi_i = \frac{1}{\tau_\theta}$

$\tau_\theta \sim Gamma(a, b)$ sehingga $E(\hat{\theta}_i) = \theta_i = \exp\left\{\mu_i + \frac{\psi_i}{2}\right\}$ dengan nilai a dan b ditentukan secara tetap (*fixed*).

ii. Linking model

$$\log(\theta_i) | \boldsymbol{\beta}, \sigma_\nu^2 \sim^{id} N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_\nu^2) \quad i = 1, 2, \dots, m$$

dengan $\boldsymbol{\beta}$ merupakan vektor koefisien regresi (pengaruh tetap) dan σ_ν^2 merupakan ragam pengaruh acak area, di mana $\nu_i \sim N(0, \sigma_\nu^2)$. Inferensi HB pada parameter θ_i mengasumsikan nilai *flat prior* untuk $\boldsymbol{\beta}$ dan σ_ν^2 . Maka,

distribusi *prior* untuk masing-masing parameter dapat dituliskan yaitu $\beta_j \sim N_{iid}(\mu_{[\beta]_j}, \sigma_{[\beta]_j}^2)$ dan

$\sigma_\nu^2 \sim InversGamma(c_1, c_2)$ dengan $\tau_\nu = \frac{1}{\sigma_\nu^2}$ sehingga $\tau_\nu \sim Gamma(c_1, \frac{1}{c_2})$, dimana $\boldsymbol{\beta}$ dan σ_ν^2 saling independen dan

nilai dari $\mu_{[\beta]_j}, \sigma_{[\beta]_j}^2, c_1, c_2$, dan τ_ν ditentukan secara tetap (*fixed*).



Fungsi *Likelihood*

$$f(\boldsymbol{\theta}|\boldsymbol{\beta}, \tau_v) = \prod_{i=1}^n f(\theta_i|\boldsymbol{\beta}, \tau_v)$$

$$f(\boldsymbol{\theta}|\boldsymbol{\beta}, \tau_v) = \prod_{i=1}^n \left(\frac{\left(\tau_\theta^{1/2} \exp \left[-\frac{\tau_\theta}{2} (\ln(\hat{\theta}_i) - \mu_i)^2 \right] \right)}{\theta_i (2\pi)^{1/2}} \times \frac{\tau_v^{1/2} \exp \left[-\frac{\tau_v}{2} (\mu_i - x_{ij}\beta_j)^2 \right]}{(2\pi)^{1/2}} \right)$$

$$f(\boldsymbol{\theta}|\boldsymbol{\beta}, \tau_v) = \frac{(\tau_\theta)^{n/2} \exp \left[-\frac{\tau_\theta}{2} \sum_{i=1}^n (\ln(\hat{\theta}_i) - \mu_i)^2 \right]}{(2\pi)^{n/2} \prod_{i=1}^n \theta_i} \times \frac{\tau_v^{n/2} \exp \left[-\frac{\tau_v}{2} \sum_{i=1}^n (\mu_i - \sum_{j=1}^n x_{ij}\beta_j)^2 \right]}{(2\pi)^{n/2}}$$

Fungsi *Prior*

$$f(\boldsymbol{\beta}) = \frac{\tau_{[\beta]_j} \exp\left[-\frac{\tau_{[\beta]_j}}{2} (\beta_j - \mu_{[\beta]_j})^2\right]}{(2\pi)^{\frac{1}{2}}}$$

$$f(\tau_\theta) = \tau_\theta^{(a-1)} \exp[-b \times \tau_\theta]; a \sim Gamma(1,1); b \sim Gamma(1,1)$$

$$f(\tau_v) = \tau_v^{(c_1-1)} \exp\left[-\frac{1}{c_2} \times \tau_v\right]; c_1 \sim Gamma(1,1); c_2 \sim Gamma(1,1)$$

SAE HB LOGNORMAL

Fungsi Proporsional Distribusi Posterior

$$f(\boldsymbol{\beta}, \tau_v | \theta_i) \propto \frac{(\tau_\theta)^{n/2} \exp\left[-\frac{\tau_\theta}{2} \sum_{i=1}^n (\ln(\hat{\theta}_i) - \mu_i)^2\right]}{\prod_{i=1}^n \theta_i} \times \frac{\tau_v^{n/2} \exp\left[-\frac{\tau_v}{2} \sum_{i=1}^n (\mu_i - \sum_{j=1}^n x_{ij}\beta_j)^2\right]}{(2\pi)^n} \times \\ \frac{\tau_{[\beta]_j} \exp\left[-\frac{\tau_{[\beta]_j}}{2} (\beta_j - \mu_{[\beta]_j})^2\right]}{(2\pi)^{1/2}} \times \tau_\theta^{(a-1)} \exp[-b \times \tau_\theta] \times \tau_v^{(c_1-1)} \exp\left[-\frac{1}{c_2} \times \tau_v\right]$$

dimana ditentukan nilai awal *prior* dari $\tau_v = 1, \mu_{[\beta]_j} = 0, \tau_{[\beta]_j} = 1, a \sim Gamma(1,1), b \sim Gamma(1,1), c_1 \sim Gamma(1,1), c_2 \sim Gamma(1,1)$.

- Penentuan *prior* ini didasarkan pada ketiadaan informasi awal dari distribusi parameter.
- Proses penyelesaian distribusi *posterior marginal* dari setiap parameter mengandung integral berdimensi tinggi dan kompleks sehingga digunakan pendekatan numerik dengan metode MCMC dan *Gibbs Sampling*.
- Metode tersebut dilakukan dengan mengambil sampel secara berulang dari bentuk distribusi *full conditional posterior* dari setiap parameter.
- Setelah menjalankan beberapa percobaan, algoritma telah konvergen pada model percobaan ke-8 dengan spesifikasi iterasi sebanyak 50.000 kali, *burn in* sebanyak 9.000 kali, *thin* sebanyak 16 kali, dan *update* sebanyak 16 kali.



Potongan Kode Program

```
model {  
    mu.b = mu.b.value # 0  
    tau.b = tau.b.value # 1  
    tau.aa = tau.ab = tau.ba = tau.bb = 1  
    tau.ua = tau.ub = 1  
    a.var = 1  
  
    for (i in 1:n1) {  
        y_sampled[i] ~ dlnorm(mu.ln[i],tau[i])  
        mu.ln[i] <- b[1] + sum(b[2:nvar]*x_sampled[i,]) + u[i]  
        u[i] ~ dnorm(0,tau.u)  
        mu[i] <- exp(mu.ln[i])  
        tau[i] ~ dgamma(tau.a, tau.tb)  
    }  
  
    # prior  
    for (k in 1:nvar) {  
        b[k] ~ dnorm(mu.b[k],tau.b[k])  
    }  
    tau.a ~ dgamma(tau.aa, tau.ab)  
    tau.tb ~ dgamma(tau.ba, tau.bb)  
    tau.u ~ dgamma(tau.ua, tau.ub)  
    a.var <- 1 / tau.u  
}
```



Preprocessing Remote Sensing

1. Cloud Masking

Teknik ini digunakan untuk menghilangkan awan yang menghalangi citra satelit. *Cloud masking* pada GEE dilakukan dengan menggunakan beberapa algoritma, seperti Fmask atau QA band, untuk mendeteksi lokasi awan pada citra. Setelah lokasi awan dideteksi, citra satelit akan diolah dengan cara menghilangkan piksel-piksel yang terpengaruh oleh awan.

```
// Create cloud masking function
var masking = function CloudMask(image) {
  var QA60 = image.select(['QA60']);
  var clouds = QA60.bitwiseAnd(1<<10).or(QA60.bitwiseAnd(1<<11)); // this gives us cloudy pixels
  return image.updateMask(clouds.not()); // remove the clouds from image
};
```

2. Cloud Selection

Teknik ini digunakan untuk memilih citra satelit yang bebas dari awan. *Cloud selection* dilakukan dengan cara membandingkan beberapa citra satelit pada waktu yang berbeda, dan memilih citra satelit yang paling sedikit mengandung awan.

```
// Import and filtering data S2, cloud selection, and cloud masking
var S2_masked = ee.ImageCollection('COPERNICUS/S2_SR')
  .filterDate("2021-01-01", "2021-12-31")
  .filterBounds(jateng)
  .filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 20)) // cloud selection
  .map(masking); // cloud masking
```

Persentase piksel berawan yang kurang dari 20% digunakan sebagai kriteria untuk memilih citra-citra yang relatif bebas dari awan.



Preprocessing Remote Sensing

3. Median Reducing atau Median Compositing

Teknik ini digunakan untuk mengurangi *noise* pada citra satelit dengan menggunakan median dari beberapa citra satelit yang diambil pada waktu yang berbeda atau pada seluruh piksel citra yang diperoleh dari beberapa waktu yang berbeda. Median reducing pada GEE dapat dilakukan dengan menggunakan fungsi *ImageCollection.median()* atau *ee.Reducer.median()*.

```
// Reducing an ImageCollection  
var S2_masked_median = S2_masked.median();
```

ATAU

```
// Create a reducer that will compute the  
median of the inputs  
var co_median =  
co.reduce(ee.Reducer.median()).clip(jateng);
```

4. Band Compositing

Teknik ini digunakan untuk menggabungkan beberapa band citra satelit menjadi satu citra yang lebih informatif. Pada GEE, teknik ini dapat dilakukan dengan menggunakan fungsi *Image.select* untuk memilih beberapa band citra satelit, dan fungsi *Image.addBands* untuk menggabungkan beberapa band tersebut.

```
// Compute BUI  
var BUI = S2_masked.expression(  
    "((SWIR - NIR) / (SWIR + NIR)) - ((NIR-  
RED) / (NIR+RED)) ",  
    {  
        RED : S2_masked.select("B4"),  
        NIR : S2_masked.select("B8"),  
        SWIR : S2_masked.select("B11")  
    });  
  
// Visualizing  
Map.addLayer(BUI.clip(jawatengah), {min:-1, max:1}, "BUI");
```



Preprocessing RWI (weighted-pop)

