



**ANALISIS FAKTOR PENENTU DAN POTENSI AKSELERASI
PEMBANGUNAN MANUSIA DI PULAU JAWA DAN KALIMANTAN:
PENDEKATAN *CLUSTERING* DAN REGRESI LOGISTIK BAYESIAN**

SKRIPSI

**untuk memenuhi persyaratan dalam menyelesaikan program sarjana
Strata-1 Statistika**

**Oleh
ADE IRAWAN
NIM. 2211017110008**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**

SKRIPSI

Analisis Faktor Penentu dan Potensi Akselerasi Pembangunan Manusia di Pulau Jawa dan Kalimantan: Pendekatan *Clustering* dan Regresi Logistik Bayesian

Oleh:
Ade Irawan
2211017110008

Telah dipertahankan pada hari Senin, tanggal 29-06-2026 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

Penguji I



Dewi Sri Susanti, S.Si., M.Si
NIP. 197305161999032002

Pembimbing II



Maisarah, S.Pd.I., M.Pd
NIP. 19890713201801213001

Penguji II



Rifqi Aulya Rahman, S.Mat., M.Si.
NIP. 199708072024061003

Banjarbaru, 15 Juli 2026
Mengetahui,
Ketua Jurusan/Koordinator PS Statistika
FMIPA/ILM



Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 15 Juli 2026



Ade Irawan
NIM. 2211017110008

ABSTRAK

Analisis Faktor Penentu dan Potensi Akselerasi Pembangunan Manusia di Pulau Jawa dan Kalimantan: Pendekatan *Clustering* dan Regresi Logistik Bayesian (Oleh: Ade Irawan; Pembimbing: Dewi Anggraini dan Maisarah, 2026; 64 halaman)

Pembangunan manusia merupakan indikator penting dalam menilai kualitas hidup masyarakat melalui aspek kesehatan, pendidikan, dan ekonomi. Penelitian ini bertujuan mengelompokkan kabupaten/kota berdasarkan karakteristik pembangunan manusia, mengidentifikasi faktor yang memengaruhi peluang suatu wilayah masuk ke kelompok pembangunan manusia relatif tinggi, serta mengevaluasi potensi akselerasi pembangunan melalui simulasi intervensi. Data yang digunakan adalah komponen penyusun Indeks Pembangunan Manusia (IPM) tahun 2024 pada 175 kabupaten/kota di Pulau Jawa dan Kalimantan, yaitu Umur Harapan Hidup (UHH), Harapan Lama Sekolah (HLS), Rata-rata Lama Sekolah (RLS), dan Pengeluaran per Kapita. Analisis dilakukan menggunakan *Hierarchical Clustering* dengan *Ward's Method* dan *Euclidean Distance*, kemudian dilanjutkan dengan Regresi Logistik Bayesian. Hasil penelitian menunjukkan terbentuk dua kelompok utama, yaitu Status Pembangunan Manusia Relatif (SPMR) tinggi sebanyak 53 kabupaten/kota dan SPMR rendah sebanyak 122 kabupaten/kota. Hasil regresi menunjukkan bahwa UHH, RLS, dan Pengeluaran per Kapita berpengaruh positif signifikan terhadap peluang masuk ke kelompok SPMR tinggi, sedangkan HLS tidak signifikan. Variabel RLS merupakan faktor yang paling dominan. Model memiliki akurasi sebesar 99,43% dan *Brier Score* sebesar 0,0083. Simulasi intervensi menunjukkan bahwa peningkatan RLS berpotensi meningkatkan peluang beberapa daerah untuk masuk ke kelompok SPMR tinggi, sehingga peningkatan pendidikan dapat menjadi strategi penting dalam mendorong pembangunan manusia.

Kata Kunci: *Hierarchical Clustering*, Regresi Logistik Bayesian, Pembangunan Manusia, Rata-rata Lama Sekolah, *Ward's Method*.

ABSTRACT

Determinant Factors and Potential Acceleration of Human Development in Java and Kalimantan: A Clustering and Bayesian Logistic Regression Approach (By: Ade Irawan; Supervisors: Dewi Anggraini and Maisarah, 2026; 64 pages)

Human development is an important indicator for assessing people's quality of life through health, education, and economic dimensions. This study aims to classify regencies/cities based on their human development characteristics, identify factors influencing the probability of a region belonging to the high human development group, and evaluate the potential acceleration of human development through intervention simulations. The data used consist of the 2024 Human Development Index (HDI) components for 175 regencies/cities in Java and Kalimantan, namely Life Expectancy (LE), Expected Years of Schooling (EYS), Mean Years of Schooling (MYS), and Expenditure per Capita. The analysis was conducted using Hierarchical Clustering with Ward's Method and Euclidean Distance, followed by Bayesian Logistic Regression. The results revealed two main groups, namely High Relative Human Development Status (RHDS) consisting of 53 regencies/cities and Low RHDS consisting of 122 regencies/cities. The regression results showed that LE, MYS, and Expenditure per Capita had a significant positive effect on the probability of belonging to the high RHDS group, while EYS was not statistically significant. MYS was identified as the most dominant factor. The model achieved an accuracy of 99.43% and a Brier Score of 0.0083. Intervention simulations indicated that increasing MYS has the potential to raise the probability of several regions moving into the high RHDS group. Therefore, improving educational attainment can be considered an important strategy for promoting human development.

Keywords: Hierarchical Clustering, Bayesian Logistic Regression, Human Development, Mean Years of Schooling, Ward's Method.

PRAKATA

Puji syukur ke hadirat Allah Swt. atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Analisis Faktor Penentu dan Potensi Akselerasi Pembangunan Manusia di Pulau Jawa dan Kalimantan: Pendekatan *Clustering* dan Regresi Logistik Bayesian”. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Dalam proses penyusunan Tugas Akhir ini, penulis memperoleh banyak bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua dan keluarga yang senantiasa memberikan doa, dukungan finansial dan emosional, dan motivasi.
2. Ibu Dewi Anggraini, S.Si., M.App.Sci., Ph.d dan Ibu Maisarah, S.Pd.I., M.Pd selaku dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan selama penelitian dan penyusunan Tugas Akhir.
3. Seluruh dosen penguji atas kritik dan saran yang diberikan untuk penyempurnaan Tugas Akhir ini.
4. Seluruh dosen dan staf Program Studi Statistika FMIPA Universitas Lambung Mangkurat atas ilmu, bantuan, dan dukungan selama masa perkuliahan.
5. Teman-teman serta semua pihak yang telah membantu dan memberikan semangat selama proses penyusunan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga Tugas Akhir ini dapat memberikan manfaat dan menjadi referensi bagi pembaca.

Banjarbaru, 15 Juli 2026

Ade Irawan

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN	ii
ABSTRAK.....	iii
<i>ABSTRACT</i>	iv
PRAKATA	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	i
DAFTAR ISTILAH, SINGKATAN, DAN LAMBANG	i
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Penelitian Terdahulu.....	5
2.2 Kajian Teori.....	8
2.2.1 Indeks Pembangunan Manusia dan Status Pembangunan Manusia Relatif (SPMR).....	8
2.2.2 Statistika Deskriptif	10
2.2.3 Standarisasi <i>Z-score</i>	11
2.2.4 <i>Euclidean Distance</i>	12
2.2.5 <i>Hierarchical Clustering</i>	13
2.2.6 Metode <i>Linkage</i>	13
2.2.7 <i>Ward's Method</i>	14
2.2.8 Dendrogram.....	15
2.2.9 Penentuan Jumlah <i>Cluster</i> (Pemotongan Dendrogram)	16
2.2.10 <i>Elbow Method</i>	17
2.2.11 Regresi Linear Sederhana.....	18
2.2.12 Regresi Logistik	18
2.2.13 Regresi Logistik Bayesian.....	19
2.2.14 <i>Markov Chain Monte Carlo</i> (MCMC)	21
2.2.15 Diagnostik Konvergensi (<i>Trace Plot</i>)	23
2.2.16 <i>Confusion Matrix</i>	24
2.2.17 <i>Brier Score</i>	25
2.2.18 <i>Credible Interval</i>	26
BAB III METODE PENELITIAN	28
3.1 Sumber Data	28
3.2 Variabel Penelitian.....	28
3.3 Prosedur Penelitian	29
BAB IV HASIL DAN PEMBAHASAN	32

4.1	Pengumpulan dan Persiapan Data Penelitian	32
4.1.1	Pengumpulan Data	32
4.1.2	Statistika Deskriptif	32
4.1.3	Persiapan Data	33
4.1.3	Standarisasi Data	34
4.2	Analisis Kluster (<i>Hierarchical Clustering</i>).....	35
4.2.1	Perhitungan Jarak Menggunakan <i>Euclidean Distance</i>	36
4.2.2	Pengelompokan Menggunakan <i>Hierarchical Clustering</i> dengan <i>Ward's Method</i>	37
4.2.3	Penentuan Jumlah <i>Cluster</i> Melalui Dendrogram.....	38
4.2.4	<i>Profiling Cluster</i> dan Penetapan Status Pembangunan Manusia Relatif (SPMR).....	41
4.3	Estimasi Model Regresi Logistik Bayesien.....	42
4.4	Diagnostik Model dan Uji Signifikansi	46
4.4.1	Uji Konvergensi MCMC Melalui <i>Trace Plot</i>	46
4.4.2	Uji Signifikansi Parameter Menggunakan 95% <i>Credible Interval</i>	47
4.5	Evaluasi Performa Model dan Prediksi	48
4.5.1	Prediksi Probabilitas Masuk ke Dalam <i>Cluster</i> Tinggi.....	48
4.5.2	Evaluasi Klasifikasi Menggunakan <i>Confusion Matrix</i> dan Akurasi Model 48	
4.5.3	Evaluasi Probabilistik Menggunakan <i>Brier Score</i>	49
4.6	Simulasi Intervensi dan Analisis Sensitivitas.....	50
4.6.1	Penyusunan Skenario Intervensi	51
4.6.2	Perhitungan <i>Delta Probability</i>	51
4.7	Interpretasi Hasil	53
4.7.1	Pola Pengelompokan Kabupaten/Kota Berdasarkan Komponen Pembangunan Manusia	53
4.7.2	Pengaruh Masing-Masing Komponen Pembangunan Manusia terhadap Peluang Masuk Cluster Pembangunan Tinggi.....	55
4.7.3	Dampak Simulasi Peningkatan Rata-Rata Lama Sekolah terhadap Peluang Akselerasi Pembangunan Daerah	57
BAB V	PENUTUP	60
5.1	Kesimpulan.....	60
5.2	Saran	61
	Daftar Pustaka.....	62
	Lampiran.....	65
	RIWAYAT HIDUP	79

DAFTAR GAMBAR

Gambar 1.1 IPM Indonesia 2020-2024	1
Gambar 1.2 Rata-rata IPM 5 Pulau Besar Indonesia.....	2
Gambar 2.1 Ilustrasi Perbandingan Metode <i>Linkage</i>	14
Gambar 2.2 Ilustrasi Dendrogram	16
Gambar 2.3 Ilustrasi Pemotongan Dendrogram	17
Gambar 2.4 Ilustrasi <i>Trace Plot</i> yang Konvergen.....	23
Gambar 3.1 Diagram Alir Penelitian.....	31
Gambar 4.1 Grafik <i>Elbow Method</i>	39
Gambar 4.2 Dendrogram Hasil <i>Hierarchical Clustering</i>	40
Gambar 4.3 Dendrogram Hasil Pengelompokan	40
Gambar 4.4 Trace Plot Parameter Model Regresi Logistik Bayesian	46
Gambar 4.5 Dendrogram <i>Hierarchical Clustering</i> Metode Ward	54
Gambar 4.6 <i>Forest Plot</i> Area Distribusi Posterior dan 95% <i>Credible Interval</i>	56

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	5
Tabel 3.1 Variabel Peneltian.....	28
Tabel 4.1 Data yang telah dikumpulkan	32
Tabel 4.2 Statistika Deskriptif Variabel Penelitian	33
Tabel 4.3 Data yang telah distandarisasi.....	35
Tabel 4.4 Hasil Perhitungan euclidean distance.....	36
Tabel 4.5 Standarisasi Kota Pontianak dan Singkawang.....	36
Tabel 4.6 Penggabungan Cluster Menggunakan <i>Ward's Method</i>	37
Tabel 4.7 Nilai WSS untuk jumlah Cluster	38
Tabel 4. 8 Karakteristik Rata-rata Indikator IPM per Cluster	41
Tabel 4.9 Hasil Estimasi Parameter Regresi Logistik Bayesian.....	45
Tabel 4.10 95% Credible Interval Parameter Regresi Logistik Bayesian.....	47
Tabel 4.11 <i>Confusion Matrix</i>	49
Tabel 4.12 Hasil Simulasi Intervensi Peningkatan RLS	52
Tabel 4.13 Hasil Pengelompokan Kabupaten/Kota Berdasarkan Cluster.....	54
Tabel 4.14 Distribusi Cluster Berdasarkan Pulau	54
Tabel 4.15 Estimasi Posterior Regresi Logistik Bayesian.....	56
Tabel 4.16 Kabupaten/Kota Potensial Setelah Simulasi Peningkatan RLS.....	57

DAFTAR LAMPIRAN

Lampiran 1. Data yang telah dikumpulkan.....	65
Lampiran 2. Data yang telah distandarisasi	67
Lampiran 3. Hasil Analisis Lengkap.....	69
Lampiran 4. Syntax R untuk Seluruh Prosedur Analisis.....	73

PRODI STATISTIKA

DAFTAR ISTILAH, SINGKATAN, DAN LAMBANG

<i>Warm-up</i>	Iterasi awal MCMC yang digunakan untuk proses penyesuaian algoritma sebelum pengambilan sampel posterior
MLE	<i>Maximum Likelihood Estimation</i>
MCMC	<i>Markov Chain Monte Carlo</i>
x_{ij}	Nilai observasi daerah ke-i pada variabel ke-j
μ_j	Rata-rata (mean) populasi dari variabel ke-j
σ_j	Standar deviasi populasi dari variabel ke-j
Z_{ij}	Nilai <i>Z-score</i> daerah ke-i pada variabel ke-j
$d(X, Y)$	Jarak Euclidean antara objek X dan Y
$\Delta(A, B)$	Perubahan variansi saat <i>cluster</i> A dan B digabung
n_A, n_B	jumlah anggota <i>cluster</i> A dan B
$\bar{x}_A, \bar{x}_B$	<i>centroid</i> (rata-rata) <i>cluster</i> A dan B
$\ \bar{x}_A - \bar{x}_B\ ^2$	<i>squared Euclidean distance</i> antar <i>centroid</i>
y_i	Variabel dependen (variabel respon) untuk pengamatan ke- <i>i</i>
x_i	Variabel independen (variabel penjelas) untuk pengamatan ke- <i>i</i>
β_0	<i>Intercept</i> (titik potong sumbu y saat $x=0$)
β_i	Koefisien regresi atau <i>slope</i> (kemiringan garis)
ϵ	<i>Error</i> atau sisaan (perbedaan antara nilai pengamatan dengan nilai prediksi model)
p_i	Peluang kejadian pada pengamatan ke- <i>i</i>
$L(\beta)$	Fungsi <i>likelihood</i>
$\hat{\beta}$	Nilai estimasi titik (rata-rata) untuk parameter
S	Jumlah total sampel yang diambil dari distribusi posterior (setelah masa <i>warm-up</i>)
$\beta^{(s)}$	Nilai sampel parameter pada iterasi ke- <i>s</i>