ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 10, Nomor 2, Tahun 2025, Halaman 325 - 334
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PEMODELAN ANGKA HARAPAN HIDUP DI INDONESIA DENGAN PENDEKATAN BAYESIAN: BAYESIAN ADAPTIVE SAMPLING DAN BAYESIAN MODEL AVERAGING DALAM SELEKSI VARIABEL

Fariz Budi Arafat[1*], Prajna Pramita Izati[2]
[1,2] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail**: farizbudia@lecturer.undip.ac.id

DOI: 10.14710/j.gauss.14.2.325-334

Article Info:
Received: 2025-06-04
Accepted: 2025-09-12
Available Online: 2025-10-07

Keywords:
regresi bayesian; bayesian adaptive sampling (BAS); g-prior; angka harapan hidup

Abstract: This study applies Bayesian Adaptive Sampling (BAS) and Bayesian Model Averaging (BMA) to model life expectancy in Indonesia, addressing model uncertainty and improving predictive accuracy. The analysis incorporates Zellner's g-prior, which enhances variable selection by balancing prior information and data-driven learning, ensuring more stable and reliable parameter estimation. Bayesian methods provide greater flexibility compared to classical regression, particularly in managing heterogeneous demographic data. The results identify poverty rate, healthcare professional ratio per 1,000 residents, percentage of infants receiving exclusive breastfeeding, and regional health expenditure as key determinants of life expectancy. The poverty rate negatively impacts life expectancy, whereas the other factors contribute positively, highlighting the importance of healthcare access, infant nutrition, and government investment in public health. The final model achieves an R^2 of 78.1%, indicating that these variables collectively explain a substantial proportion of life expectancy variability. By integrating Zellner's g-prior, Bayesian inference facilitates a robust selection of influential predictors, leading to more precise policy recommendations. The study suggests that enhancing healthcare distribution, promoting breastfeeding awareness, and optimizing health budget allocation can significantly improve life expectancy outcomes. Bayesian methods provide a powerful framework for demographic modeling by incorporating uncertainty and refining estimation accuracy.

1. PENDAHULUAN

Angka harapan hidup merupakan salah satu indikator utama dalam demografi dan kesehatan masyarakat. Secara umum, angka ini menunjukkan perkiraan rata-rata umur seseorang berdasarkan kondisi tertentu, seperti lokasi geografis, status sosial-ekonomi, serta akses terhadap layanan kesehatan. Indikator ini sering digunakan untuk menilai kualitas hidup dan tingkat kesejahteraan suatu populasi. Berdasarkan data dari Badan Pusat Statistik (BPS), rata-rata angka harapan hidup di Indonesia terus mengalami peningkatan sejak tahun 2014 sebesar 70,59 menjadi 73,20 di tahun 2023.

Angka harapan hidup dipengaruhi oleh berbagai faktor multidimensional, termasuk tingkat pendidikan, pendapatan, pola konsumsi makanan, kondisi lingkungan, serta kemajuan teknologi medis. Selain itu, prevalensi penyakit kronis, akses terhadap air bersih dan sanitasi, serta kualitas layanan kesehatan primer juga berperan penting dalam menentukan AHH suatu populasi. Negara dengan akses layanan kesehatan yang baik dan tingkat kemakmuran tinggi cenderung memiliki angka harapan hidup lebih tinggi dibandingkan negara yang mengalami keterbatasan sumber daya dan tingginya angka penyakit menular. Sebagai contoh, angka harapan hidup di Australia pada tahun 2023 mencapai 83,1 dan menduduki peringkat ke-4 dunia (Australian Bureau of Statistics, 2024).

Di Indonesia, variasi AHH antar provinsi menunjukkan adanya ketimpangan akses dan kualitas layanan publik, sehingga analisis faktor-faktor penentu AHH menjadi penting untuk mendukung kebijakan berbasis bukti.

Selain itu, angka harapan hidup merupakan indikator utama dalam pengukuran kualitas hidup dan pembangunan manusia. Di Indonesia, AHH digunakan sebagai salah satu komponen dalam Indeks Pembangunan Manusia (IPM), yang menjadi acuan dalam alokasi anggaran, evaluasi kinerja daerah, dan perencanaan pembangunan nasional. Oleh karena itu, pemahaman terhadap faktor-faktor yang memengaruhi AHH sangat penting untuk merancang intervensi kebijakan yang efektif, seperti program peningkatan layanan kesehatan, sistem asuransi sosial, dan strategi pembangunan ekonomi (Sitorus et al., 2024). Perubahan dalam angka harapan hidup dapat memberikan wawasan mengenai dampak perubahan sosial, ekonomi, serta kesehatan dalam suatu masyarakat.

Angka harapan hidup telah menjadi topik penelitian yang luas dalam berbagai disiplin ilmu, termasuk demografi, ekonomi, dan kesehatan masyarakat. Sejumlah studi telah mengeksplorasi faktor-faktor yang memengaruhi angka harapan hidup, mulai dari variabel sosio-ekonomi hingga perkembangan medis dan perubahan gaya hidup. Sugiantari dan Budiantara (2013) melakukan eksplorasi faktor-faktor yang memengaruhi angka harapan hidup di Jawa Timur dengan Regresi Spline. Adapun variabel yang memberikan pengaruh signifikan adalah angka kematian bayi, persentase bayi berusia 0-11 bulan yang diberi ASI selama 4-6 bulan, dan variabel persentase balita berusia 1-4 tahun yang mendapatkan imunisasi lengkap.

Maulana et al. (2024) juga melakukan penelitian terhadap angka harapan hidup di Provinsi Aceh. Diperoleh hasil bahwa pendapatan domestik regional bruto (PDRB) per kapita berpengaruh positif signifikan terhadap angka harapan hidup. Penelitian terhadap angka harapan hidup laki-laki dan perempuan secara terpisah dilakukan oleh Maryani dan Kristiana (2018). Hasil penelitian mengindikasikan bahwa faktor yang secara signifikan memengaruhi angka harapan hidup perempuan mencakup persentase penduduk yang pernah menjalani rawat inap, proporsi bayi di bawah dua tahun yang masih menerima ASI, serta rumah tangga yang memiliki akses terhadap fasilitas sanitasi untuk buang air besar. Sementara itu, angka harapan hidup laki-laki dipengaruhi oleh keberlanjutan pemberian ASI bagi bayi di bawah dua tahun, penggunaan fasilitas sanitasi dalam rumah tangga, serta pemanfaatan jaminan kesehatan untuk layanan rawat inap maupun rawat jalan.

Dalam konteks angka harapan hidup, metode Bayesian memberikan fleksibilitas dalam menangani data yang tidak selalu memenuhi asumsi distribusi normal atau memiliki jumlah sampel terbatas (Schmertmann, C.P. dan Gonzaga, M.R., 2018). Mahmudah et al. (2024) menggunakan pendekatan regresi logistik berbasis Bayesian dengan MCMC-Gibbs Sampling karena mampu menangani ketidakpastian data dan menghasilkan model yang lebih representatif untuk klasifikasi IPM di Jawa Timur. Dengan metode tersebut, diperoleh nilai *error* dari model sebesar 0,03. Penelitian Katianda et al. (2020) menggunakan regresi linier Bayesian dalam pemodelan kemiskinan di Kalimantan Timur karena pendekatan ini menyederhanakan proses estimasi model marginal dan menghasilkan pendugaan parameter yang lebih akurat dengan mempertimbangkan informasi *prior* selain data sampel. Metode ini dinilai lebih unggul dibandingkan pendekatan klasik, terutama dalam kondisi data yang terbatas.

Dalam pendekatan bayesian, penentuan distribusi prior merupakan komponen utama yang sangat menentukan kualitas hasil estimasi. Prior yang dipilih akan memengaruhi bentuk sebaran posterior, sehingga pemilihan prior yang tepat menjadi krusial, terutama dalam kondisi data yang terbatas atau heterogen. Dengan menggunakan distribusi prior yang

sesuai, seperti Zellner's g-prior, pendekatan ini mampu mengakomodasi pola heterogen dalam data serta meningkatkan ketepatan prediksi melalui penggabungan probabilitas model secara keseluruhan menggunakan Bayesian Model Averaging (BMA) (Clyde et al., 2011).

Selain itu, teknik seperti Bayesian Adaptive Sampling (BAS) memungkinkan seleksi variabel yang lebih optimal dalam konteks pemodelan angka harapan hidup, dengan mempertimbangkan berbagai kemungkinan model melalui pendekatan BMA, tanpa harus mengandalkan satu model terbaik secara deterministik (Clyde et al., 2011). Pendekatan ini memberikan keunggulan dalam menangkap ketidakpastian model dan menghasilkan estimasi yang lebih stabil. (Clyde et al., 2011). Dengan mempertimbangkan ketidakpastian dalam estimasi angka harapan hidup serta variasi antar kelompok, metode bayesian menjadi pendekatan yang tepat untuk menghasilkan inferensi yang lebih *robust* dan informatif dalam perencanaan kebijakan kesehatan dan evaluasi faktor-faktor sosial ekonomi yang memengaruhi umur harapan hidup suatu populasi.

Berdasarkan latar belakang tersebut, tujuan dari artikel ini adalah untuk mengidentifikasi faktor-faktor dominan yang memengaruhi angka harapan hidup di Indonesia dengan pendekatan metode Bayesian. Pendekatan ini memungkinkan pemilihan model yang optimal dari berbagai kombinasi variabel prediktor, serta menghasilkan estimasi yang mempertimbangkan ketidakpastian dalam pemilihan model. Dengan menggabungkan informasi dari berbagai model yang mungkin, metode Bayesian memberikan hasil yang lebih stabil dan dapat diandalkan dibandingkan pendekatan tunggal. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengambilan keputusan berbasis data untuk perencanaan kebijakan kesehatan dan pembangunan manusia.

2. TINJAUAN PUSTAKA

Dalam analisis regresi dengan banyak prediktor potensial, pemilihan model yang optimal menjadi tantangan utama. Bayesian Adaptive Sampling (BAS) merupakan metode yang dirancang untuk secara efisien menjelajahi ruang model yang luas, memungkinkan seleksi variabel yang optimal dan meminimalkan risiko *overfitting*. Pendekatan ini sangat berguna dalam penelitian angka harapan hidup, di mana faktor-faktor yang memengaruhi estimasi sering kali kompleks dan beragam.

Menurut Clyde et al. (2011), dalam konteks regresi linier dengan p prediktor yang potensial $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p$, model $\mathcal{M}_{\boldsymbol{\gamma}}$ direpresentasikan sebagai vektor biner $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_p)^T \in \{0,1\}^p \equiv \Gamma$, di mana setiap γ_j menunjukkan apakah prediktor $\boldsymbol{x}_j$ dimasukkan sebagai kolom dalam pada matrix $\boldsymbol{X}_{\boldsymbol{\gamma}}$ yang berukuran $n \times p_{\boldsymbol{\gamma}}$ dalam model $\mathcal{M}_{\boldsymbol{\gamma}}$. Model linier normal bersyarat pada $\mathcal{M}_{\boldsymbol{\gamma}}$ memiliki bentuk:

$$\boldsymbol{Y}|\alpha, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \phi, \mathcal{M}_{\gamma} \sim \mathcal{N}(\mathbf{1}_n\alpha + \boldsymbol{X}_{\boldsymbol{\gamma}}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{I}_n/\phi) \quad (1)$$

dengan $\boldsymbol{Y} = (Y_1, \dots, Y_n)'$, $\mathbf{1}_n$ menunjukkan vektor satu sebanyak n, α adalah konstanta, $\boldsymbol{\beta}_{\boldsymbol{\gamma}}$ merepresentasikan koefisien regresi dan ϕ adalah presisi (invers dari galat variansi) dengan $\boldsymbol{I}_n$ adalah matriks identitas $n \times n$.

Dalam pendekatan Bayesian, pemilihan model berdasarkan distribusi posterior dari model $\mathcal{M}_{\boldsymbol{\gamma}}$:

$$p(\mathcal{M}_{\boldsymbol{\gamma}}|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\mathcal{M}_{\boldsymbol{\gamma}})p(\mathcal{M}_{\boldsymbol{\gamma}})}{\sum_{\boldsymbol{\gamma}\in\Gamma} p(\boldsymbol{Y}|\mathcal{M}_{\boldsymbol{\gamma}})p(\mathcal{M}_{\boldsymbol{\gamma}})} \quad (2)$$

di mana $p(\boldsymbol{Y}|\mathcal{M}_{\boldsymbol{\gamma}}) = \int p(\boldsymbol{Y}|\boldsymbol{\theta}_{\boldsymbol{\gamma}}, \mathcal{M}_{\boldsymbol{\gamma}})p(\boldsymbol{\theta}_{\boldsymbol{\gamma}}|\mathcal{M}_{\boldsymbol{\gamma}})d\boldsymbol{\theta}_{\boldsymbol{\gamma}}$ adalah marginal likelihood dari $\mathcal{M}_{\boldsymbol{\gamma}}$ yang diperoleh dengan integral dari likelihood gabungan pada distribusi prior dari semua parameter $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = (\alpha, \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \phi)$ dan $p(\mathcal{M}_{\boldsymbol{\gamma}})$ adalah probabilitas prior dari model.

Ketidakpastian dalam pemilihan model merupakan tantangan yang perlu diperhitungkan untuk menghasilkan estimasi parameter yang lebih akurat dan robust. Bayesian Model Averaging (BMA) menawarkan pendekatan alternatif yang lebih sistematis dibandingkan metode tradisional yang hanya memilih satu model terbaik berdasarkan kriteria tertentu. Dengan BMA, inferensi dilakukan dengan mempertimbangkan distribusi posterior dari semua model yang mungkin, sehingga ketidakpastian dalam pemilihan model dapat terintegrasi dalam proses estimasi (Hoeting et al., 1999 dan Clyde dan George, 2004). Secara matematis, distribusi posterior untuk suatu kuantitas Δ dalam kerangka BMA didefinisikan sebagai:

$$p(\Delta|\boldsymbol{Y}) = \sum_{\boldsymbol{\gamma}\in\Gamma} p\left(\Delta\middle|\mathcal{M}_{\boldsymbol{\gamma}}, \boldsymbol{Y}\right) p\left(\mathcal{M}_{\boldsymbol{\gamma}}\middle|\boldsymbol{Y}\right) \tag{3}$$

dengan ekspektasi model rata-rata didefinisikan sebagai:

$$E[\Delta|\boldsymbol{Y}] = \sum_{\boldsymbol{\gamma}\in\Gamma} E\left(\Delta\middle|\mathcal{M}_{\boldsymbol{\gamma}}, \boldsymbol{Y}\right) p\left(\mathcal{M}_{\boldsymbol{\gamma}}\middle|\boldsymbol{Y}\right) \tag{4}$$

Dengan mempertimbangkan banyak model, Bayesian Model Averaging mengurangi risiko *overfitting* dan memungkinkan estimasi parameter yang lebih stabil, terutama dalam kasus di mana data memiliki dimensi tinggi atau jumlah sampel yang terbatas.

Dalam model regresi Bayesian, pemilihan distribusi prior memainkan peran penting dalam menentukan inferensi parameter. Salah satu prior yang banyak digunakan dalam konteks pemodelan regresi adalah Zellner's g-prior, yang dirancang untuk menangani ketidakpastian parameter regresi dengan mempertimbangkan informasi dari data yang diamati. Zellner's g-prior menawarkan pendekatan yang fleksibel dengan memasukkan kovarians dari prediktor ke dalam distribusi prior koefisien regresi. Zellner's g-prior memiliki persamaan (Zellner, 1986):

$$\begin{gathered} \boldsymbol{\beta}_{\boldsymbol{\gamma}}|\boldsymbol{\gamma}, \phi \sim \mathcal{N}_{p_{\boldsymbol{\gamma}}}\left(\mathbf{0}, g\left(\boldsymbol{X}_{\boldsymbol{\gamma}}'\boldsymbol{X}_{\boldsymbol{\gamma}}\right)^{-1}/\phi\right), \\ p(\alpha, \phi) \propto 1/\phi. \end{gathered} \tag{5}$$

Variansi prior dari parameter regresi ditentukan oleh nilai g yang dipilih oleh peneliti, dan nilai prior efektif dari jumlah sampel adalah n/g, di mana n adalah banyaknya sampel.

Berbagai pilihan nilai g telah diusulkan dalam literatur (Forte et al., 2018) untuk menyesuaikan tingkat pengaruh prior terhadap hasil estimasi. Zellner pertama kali mengusulkan penggunaan $g = n$, yang sesuai dengan ukuran sampel prior sebesar 1. Pendekatan ini dikenal sebagai Unit Information Prior (UIP) (Kass dan Wasserman, 1995), di mana prior memberikan jumlah informasi yang setara dengan satu observasi dalam sampel data.

Pilihan alternatif adalah $g = 1$, yang sesuai dengan ukuran sampel prior sebesar n (Van Zwet, 2019). Salah satu justifikasi untuk pendekatan ini adalah bahwa studi biasanya dirancang dengan ukuran sampel yang memiliki daya cukup untuk mendeteksi efek dengan ukuran tertentu. Oleh karena itu, ketika varians dalam distribusi prior disamakan dengan varians dalam distribusi sampling, kontribusi prior terhadap estimasi parameter menjadi seimbang. Keserupaan ini membantu menghindari dominasi salah satu sumber informasi, sehingga mengurangi potensi bias dan menghasilkan estimasi yang lebih stabil dan representatif terhadap data yang tersedia.

Pilihan menengah yang sering digunakan adalah $g = \sqrt{n}$ (Fernandez et al., 2001), yang memiliki ukuran sampel prior lebih moderat dibandingkan dua pendekatan sebelumnya. Pendekatan ini telah ditemukan bekerja dengan baik dalam konteks dimensi tinggi (Young et al., 2014), di mana jumlah variabel prediktor yang tersedia relatif besar dibandingkan

ukuran sampel. Dengan demikian, pemilihan g perlu disesuaikan dengan sifat data dan tujuan inferensi untuk memastikan stabilitas estimasi parameter dalam regresi Bayesian. Porwal dan Raftery (2022) telah menguji beberapa pilihan prior dan menyimpulkan bahwa g-prior dengan $g = \sqrt{n}$ memiliki performa terbaik.

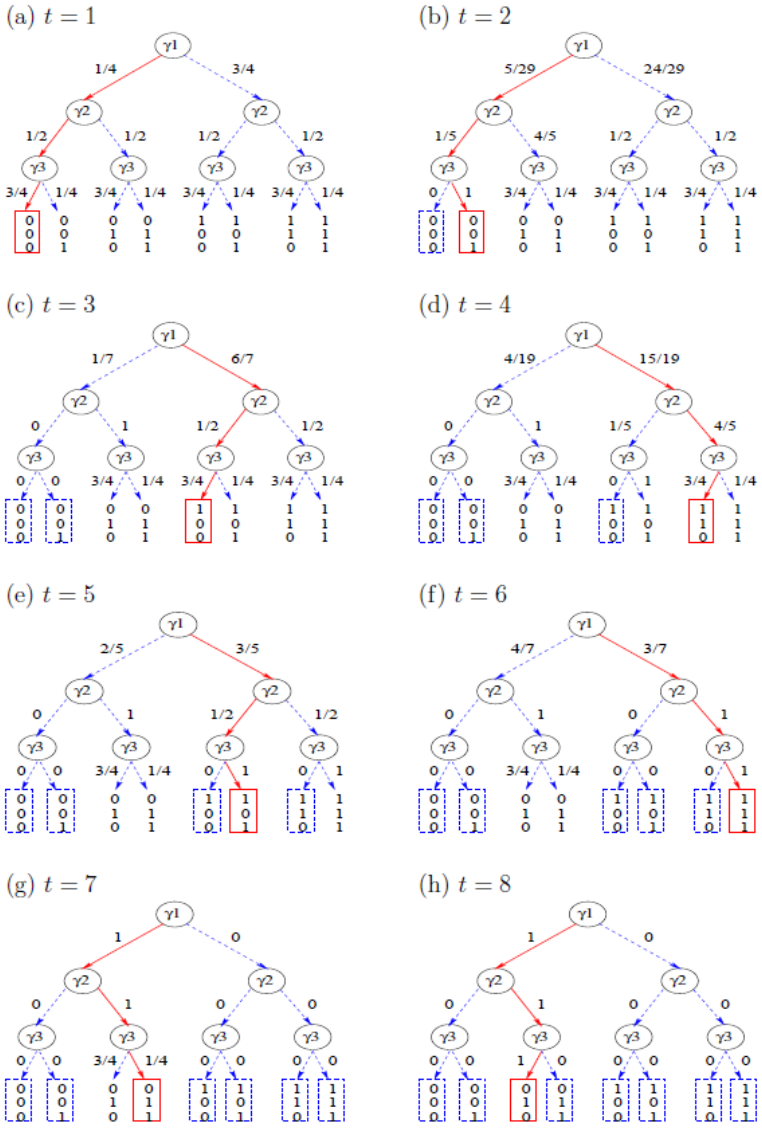
3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang diambil dari Badan Pusat Statistik (BPS). Variabel respon yang digunakan adalah angka harapan hidup (Y). Sedangkan variable prediktor terdiri atas gini rasio (X_1), persentase penduduk miskin (X_2), rasio jumlah tenaga kesehatan per 1000 penduduk (X_3), persentase bayi usia kurang dari 6 bulan yang mendapatkan asi eksklusif (X_4), produk domestik regional bruto (pdrb) per kapita dalam ribu rupiah (X_5), dan anggaran belanja daerah untuk kesehatan dalam milyar rupiah (X_6). Unit observasi dari data yang diambil adalah data 34 provinsi di Indonesia pada tahun 2023.

Berdasarkan dari Clyde et al. (2011), Bayesian Adaptive Sampling (BAS) adalah algoritma yang dirancang untuk melakukan pemilihan model secara efisien tanpa penggantian (*without replacement*), di mana probabilitas pemilihan model sebanding dengan suatu fungsi massa probabilitas yang memiliki konstanta normalisasi yang diketahui.

Dalam BAS, ruang model direpresentasikan sebagai pohon biner (*binary tree*), dengan γ_1 sebagai simpul utama (*root node*), diikuti oleh $\gamma_2, \dots, \gamma_p$. Setiap simpul j memiliki dua cabang yang mewakili kondisi $\gamma_j = 0$ (tidak dimasukkan dalam model) atau $\gamma_j = 1$ (dimaksudkan untuk dimasukkan dalam model). Dengan struktur ini, setiap model dalam ruang Γ direpresentasikan sebagai satu jalur unik di antara 2^p kemungkinan jalur dalam pohon biner.

Pendekatan ini memungkinkan perhitungan ulang probabilitas yang telah dinormalisasi tanpa perlu mencantumkan seluruh kemungkinan model 2^p secara eksplisit, serta memungkinkan proses sampling model langsung tanpa penggantian dari probabilitas yang telah dinormalisasi pada pohon. Ilustrasi BAS untuk kasus dengan $p = 3$ ditunjukkan oleh Gambar 1.



Gambar 1. Ilustrasi algoritma BAS untuk $p = 3$ dengan probabilitas inklusi yang dipetakan sepanjang cabang pohon. Garis solid dan persegi panjang menunjukkan model yang dipilih pada iterasi ke-t, sementara persegi panjang dengan garis putus-putus menunjukkan model yang telah dipilih pada iterasi sebelumnya (Clyde et al. 2011).

Setiap fungsi kepadatan probabilitas dapat ditulis sebagai rangkaian dari distribusi marginal dan bersyarat dengan bentuk

$$f(\boldsymbol{\gamma}) = \prod_{j=1}^{p} f(\gamma_j | \boldsymbol{\gamma}_{<j}) \tag{6}$$

di mana notasi $\boldsymbol{\gamma}_{<j}$ menunjukkan subset dar indicator inklusi $\{\gamma_k\}$ untuk $k < j$. Dapat juga dituliskan $\boldsymbol{\gamma}_{\geq j} = \{\gamma_k\}$ untuk $k \geq j$. Untuk $j = 1$, $f(\gamma_1 | \boldsymbol{\gamma}_{<1}) \equiv f(\gamma_1)$ adalah distribusi marginal dari γ_1. Karena γ_j adalah biner, persamaan (6) dapat dituliskan

$$f(\boldsymbol{\gamma}|\boldsymbol{\rho}) = \prod_{j=1}^{p} (\rho_{j|<j})^{\gamma_j} (1 - \rho_{j|<j})^{1-\gamma_j} \tag{7}$$

di mana $\rho_{j|<j} \equiv f(\gamma_j = 1|\boldsymbol{\gamma}_{<j})$ dan $\boldsymbol{\rho}$ adalah himpunan dari semua $\{\rho_{j|<j}\}$. Setelah melakukan sampling terhadap model dari persamaan (7), distribusi dari model yang lain memiliki persamaan yang sama dengan persamaan (7), namun dengan nilai $\boldsymbol{\rho}$ yang baru.

Langkah-langkah analisis yang dilakukan meliputi: (1) menghitung statistik deskriptif untuk seluruh variabel; (2) menetapkan distribusi *prior* untuk model; (3) menjalankan BAS untuk memperoleh model dengan probabilitas posterior tertinggi; (4) mengestimasi parameter dari model terpilih; dan (5) menyusun interpretasi hasil estimasi berdasarkan variabel yang masuk dalam model. Analisis data diimplementasikan dengan perangkat lunak R (4.5.0) dan paket BAS (1.7.5).

4. HASIL DAN PEMBAHASAN

Hasil analisis deskriptif terhadap angka harapan hidup dan faktor-faktor yang diduga memengaruhi di Indonesia yang terdiri atas rata-rata, standar deviasi, nilai minimum, dan nilai maksimum ditunjukkan oleh Tabel 1.

Tabel 1. Analisis Deskriptif

Variabel	Rata-rata	StandarDeviasi	Minimum	Maksimum
Y	73,20	1,94	68,22	75,88
X_1	0,34	0,05	0,25	0,45
X_2	10,09	5,18	4,25	26,03
X_3	0,70	0,38	0,07	2,13
X_4	70,41	7,67	55,11	82,45
X_5	81,948	61,907	23,078	322,615
X_6	6,57	6,33	1,49	29,19

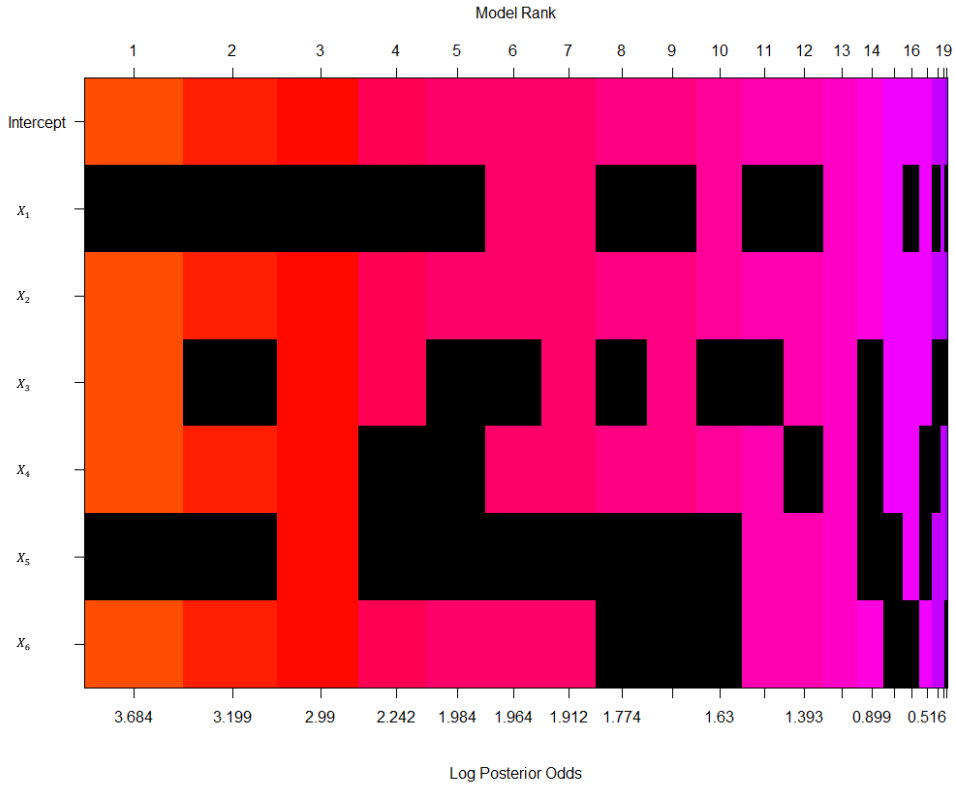
Berdasarkan Tabel 1, rata-rata angka harapan hidup di Indonesia mencapai 73,20 tahun, dengan Papua memiliki angka terendah sebesar 68,22 tahun, sedangkan DKI Jakarta menunjukkan angka tertinggi, yaitu 75,88 tahun.

Rata-rata gini rasio (X_1), yang mencerminkan tingkat ketimpangan ekonomi, tercatat sebesar 0,34, dengan nilai terendah 0,25 di Kepulauan Bangka Belitung dan tertinggi 0,45 di DI Yogyakarta. Sementara itu, rata-rata persentase penduduk miskin (X_2) berada di angka 10,09%, dengan Bali sebagai provinsi dengan tingkat kemiskinan terendah (4,25%) dan Papua sebagai provinsi dengan tingkat kemiskinan tertinggi (26,03%).

Rata-rata rasio jumlah tenaga kesehatan per 1.000 penduduk (X_3) tercatat 0,70, dengan distribusi yang sangat bervariasi—Papua memiliki rasio terendah (0,07), sementara DKI Jakarta memiliki rasio tertinggi (2,13). Proporsi bayi berusia kurang dari 6 bulan yang mendapatkan ASI eksklusif (X_4) rata-rata mencapai 70,41%, dengan Gorontalo sebagai provinsi dengan cakupan terendah (55,11%) dan Nusa Tenggara Barat dengan cakupan tertinggi (82,45%).

Selanjutnya, rata-rata Produk Domestik Regional Bruto (PDRB) per kapita (X_5) berada di angka Rp81.948, dengan nilai terendah Rp23.078 di Nusa Tenggara Timur dan tertinggi Rp322.615 di DKI Jakarta. Adapun anggaran belanja daerah untuk kesehatan (X_6) rata-rata sebesar Rp6,57 miliar, dengan alokasi terendah di Gorontalo (Rp1,49 miliar) dan tertinggi di Jawa Timur (Rp29,19 miliar).

Selanjutnya dilakukan pencarian model terbaik dengan BAS. Model prior yang digunakan adalah uniform, prior untuk parameter adalah g-prior dengan nilai $g = n$. Dengan banyaknya variable $p = 6$, maka jumlah model yang terbentuk ada sebanyak $2^6 = 64$ model. Hasil dari 20 model terbaik ditunjukkan oleh Gambar 2 dan nilai probabilitas posterior masing-masing prediktor ditunjukkan oleh Tabel 2.



Gambar 2. Dua puluh model terbaik dari BAS.

Tabel 1. Ringkasan dari 5 Model Terbaik

	$P(\beta \neq 0\|Y)$	Model 1[a]	Model 2[a]	Model 3[a]	Model 4[a]	Model 5[a]
Intercept	1,000	1	1	1	1	1
X_1	0,2094	0	0	0	0	0
X_2	0,9999	1	1	1	1	1
X_3	0,6057	1	0	1	1	0
X_4	0,8168	1	1	1	0	0
X_5	0,2489	0	0	1	0	0
X_6	0,8452	1	1	1	1	1
BayesFactor		1,000	0,6160	0,4997	0,2367	0,1827
PostProbs		0,2509	0,1545	0,1254	0,0594	0,0458
R^2		0,7810	0,7445	0,7968	0,7275	0,6878
logmarg		16,3464	15,8619	15,6528	14,9053	14,6465

[a] Nilai pada kolom Model 1 sampai Model 5 pada baris variabel menunjukkan apakah prediktor tersebut masuk ke dalam model (1 = Ya, 0 = Tidak).

Model terbaik dipilih berdasarkan nilai *log marginal likelihood* tertinggi, yang menunjukkan model dengan kesesuaian terbaik terhadap data. Berdasarkan Gambar 2, model terbaik terdiri dari variabel X_2, X_3, X_4, dan X_5. Selain itu, dapat terlihat bahwa variabel X_2 termasuk dalam 20 model terbaik, sedangkan X_6 muncul dalam 7 model terbaik, dan X_4 merupakan bagian dari 3 model terbaik. Probabilitas posterior $P(\beta \neq 0 \mid Y)$ yang ditampilkan pada Tabel 1 menunjukkan sejauh mana suatu prediktor berkontribusi terhadap pemodelan angka harapan hidup. Dari hasil tersebut, variabel X_2, X_4, dan X_6 memiliki probabilitas posterior yang tinggi (>0,8), mengindikasikan bahwa variabel-variabel ini memiliki kemungkinan besar untuk masuk dalam model final. Sementara itu, variabel X_3 juga memiliki probabilitas posterior yang relatif tinggi, yaitu 0,6057, sedangkan variabel X_1

dan X_5 menunjukkan probabilitas posterior yang lebih rendah, masing-masing 0,2094 dan 0,2489.

Selain itu, nilai R^2 yang diperoleh dari model terbaik adalah 78,1%, yang berarti bahwa variabel X_2, X_3, X_4, dan X_5 mampu menjelaskan 78,1% variabilitas variabel Y, menunjukkan tingkat prediktabilitas yang cukup baik dalam analisis ini.

Tabel 2. Estimasi Parameter

	$\hat{\beta}$	CI 95%[a]		Keterangan
		2.5%	97.5%	
Intercept	73,1976	72,8487	73,5636	Signifikan
X_1	0,7834	-2,3641	10,4867	Tidak Signifikan
X_2	-0,2307	-0,3226	-0,1418	Signifikan
X_3	0,7673	-0,0038	2,2516	Tidak Signifikan
X_4	0,0527	0,0000	0,1058	Signifikan
X_5	-0,0009	-0,0091	0,0009	Tidak Signifikan
X_6	0,0694	0,0000	0,1322	Signifikan

[a]Credibility Interval (Interval kredibilitas) 95%

Hasil estimasi parameter Bayesian pada Tabel 2 menunjukkan bahwa X_2 memiliki dampak negatif signifikan terhadap angka harapan hidup, sebagaimana ditunjukkan oleh interval kredibilitas yang tidak mencakup nol. Selain itu, X_4 dan X_6 memiliki pengaruh positif yang relevan, menunjukkan bahwa faktor-faktor ini berkontribusi terhadap peningkatan angka harapan hidup. Sebaliknya, variabel X_1, X_3, dan X_5 memiliki interval kredibilitas yang mencakup nol, sehingga pengaruhnya terhadap angka harapan hidup dalam model ini masih belum dapat dipastikan secara kuat. Meskipun demikian, variabel X_3 termasuk dalam model terbaik yang diperoleh.

5. KESIMPULAN

Hasil pemilihan model terbaik dengan Bayesian Adaptive Sampling (BAS) menunjukkan bahwa persentase penduduk miskin (X_2) memiliki dampak negatif terhadap angka harapan hidup. Temuan ini mengindikasikan bahwa tingginya tingkat kemiskinan berkontribusi pada rendahnya akses terhadap layanan kesehatan dan kondisi hidup yang lebih buruk. Sebaliknya, rasio jumlah tenaga kesehatan per 1000 penduduk (X_3), persentase bayi yang mendapatkan ASI eksklusif (X_4), serta anggaran belanja daerah untuk sektor kesehatan (X_6) memberikan pengaruh positif. Hal ini mencerminkan bahwa akses terhadap tenaga medis, pemberian gizi optimal sejak dini, dan dukungan kebijakan kesehatan merupakan faktor kunci dalam meningkatkan angka harapan hidup suatu populasi.

Berdasarkan hasil tersebut, rekomendasi kebijakan yang diajukan secara langsung merujuk pada variabel-variabel yang terbukti signifikan dalam model. Program pengentasan kemiskinan perlu dirancang secara terintegrasi dengan sektor kesehatan, seperti subsidi layanan kesehatan bagi masyarakat berpendapatan rendah, untuk mengatasi dampak negatif dari X_2. Peningkatan jumlah tenaga kesehatan (X_3) harus disertai dengan distribusi yang merata, terutama ke daerah dengan akses terbatas. Selain itu, edukasi dan promosi ASI eksklusif (X_4) harus diperkuat melalui kampanye kesehatan ibu dan anak, untuk memastikan bayi mendapatkan nutrisi optimal di awal kehidupan. Selain itu, alokasi anggaran kesehatan (X_6) harus diarahkan secara strategis agar investasi tersebut berdampak langsung terhadap peningkatan kesejahteraan dan kualitas hidup masyarakat.

DAFTAR PUSTAKA

Australian Bureau of Statistics 2021-2023, *Life expectancy*, ABS, viewed 3 May 2025, <https://www.abs.gov.au/statistics/people/population/life-expectancy/latest-release>.
Clyde, M.A., Ghosh, J. and Littman, M.L., 2011. Bayesian adaptive sampling for variable selection and model averaging. *Journal of Computational and Graphical Statistics*, *20*(1), pp.80-101.
Clyde, M. and George, E.I., 2004. Model uncertainty.
Fernandez, C., Ley, E. and Steel, M.F., 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*(2), pp.381-427.
Forte, A., Garcia-Donato, G. and Steel, M., 2018. Methods and tools for Bayesian variable selection and model averaging in normal linear regression. *International Statistical Review*, *86*(2), pp.237-258.
Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T., 1999. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and EI George, and a rejoinder by the authors. *Statistical science*, *14*(4), pp.382-417.
Kass, R.E. and Wasserman, L., 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the american statistical association*, *90*(431), pp.928-934.
Katianda, K. R., Goejantoro, R., & Satriya, A. M. A. 2020. Estimasi Parameter Model Regresi Linier dengan Pendekatan Bayes. *EKSPONENSIAL*, *11*(2), 127-132.
Mahmudah, N., Ningrum, I. K. and Dayanti, F. 2024. Implementasi Regresi Logistik Bayesian Pada Indeks Pembangunan Manusia. *Journal of Mathematics Education and Science*, *7*(1), pp. 85–91. doi: 10.32665/james.v7i1.2427.
Maryani, H., & Kristiana, L. 2018. Pemodelan angka harapan hidup (AHH) laki-laki dan perempuan di Indonesia tahun 2016. *Buletin Penelitian Sistem Kesehatan*, *21*(2), 71-81.
Maulana, M. A., Farlian, T., Handayani, M., & Juliansyah, R. 2024. PENGARUH PDRB PERKAPITA DAN PENGELUARAN PEMERINTAH TERHADAP ANGKA HARAPAN HIDUP DI PROVINSI ACEH. *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, *7*(3), 6769-6778.
Porwal, A. and Raftery, A.E., 2022. Comparing methods for statistical inference with model uncertainty. *Proceedings of the National Academy of Sciences*, *119*(16), p.e2120737119.
Schmertmann, C.P. and Gonzaga, M.R., 2018. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*, *55*(4), pp.1363–1388. https://doi.org/10.1007/s13524-018-0695-2
Sitorus, N., Yusrizal, Y. and Nasution, J. 2024. Peranan Program Jaminan Kesehatan Nasional (JKN) dalam mendorong Sustainable Development Goals (SDGs) di Indonesia. *Economic Reviews Journal*, *3*(1), pp. 45–60.
Sugiantari, A. P., & Budiantara, I. N. 2013. Analisis Faktor-faktor yang mempengaruhi angka harapan hidup di Jawa Timur menggunakan Regresi Semiparametrik Spline. *Jurnal sains dan Seni ITS*, *2*(1), D37-D41.
Van Zwet, E., 2019. A default prior for regression coefficients. *Statistical methods in medical research*, *28*(12), pp.3799-3807.
Young, W.C., Raftery, A.E. and Yeung, K.Y., 2014. Fast Bayesian inference for gene regulatory networks using ScanBMA. *BMC systems biology*, *8*, pp.1-11.
Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, North-Holland/Elsevier, pp. 233-243.