



Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia Menggunakan Algoritma GUIDE

Syifa Maulia¹, Yasmin Erika Faridhan^{2*}, dan Maya Widyastiti³

^{1,2,3}Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Pakuan

*Correspondence author: yasminef@unpak.ac.id

ABSTRAK

Indeks Pembangunan Manusia (IPM) merupakan indikator penting untuk mengukur kualitas pembangunan manusia di suatu wilayah. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia kabupaten/kota di Indonesia menggunakan algoritma *Generalized, Unbiased, Interaction Detection and Estimation* (GUIDE). Data yang digunakan mencakup peubah sosial dan ekonomi yang berkaitan dengan pendidikan, kesehatan, dan kondisi ekonomi. Algoritma GUIDE diterapkan untuk mengidentifikasi peubah prediktor yang berpengaruh signifikan serta pola interaksi antarpeubah tanpa bias dalam pemilihan peubah. Hasil penelitian menunjukkan bahwa beberapa faktor utama, seperti angka harapan hidup, rataan lama sekolah, dan pengeluaran per kapita, memiliki peran dominan dalam menentukan nilai IPM di berbagai daerah. Model pohon regresi yang dihasilkan mampu memberikan interpretasi yang jelas mengenai hubungan antara peubah penjelas dan IPM. Penelitian ini menunjukkan bahwa algoritma GUIDE efektif digunakan dalam pemodelan IPM dan dapat menjadi alternatif metode analisis dalam kajian pembangunan daerah di Indonesia.

© 2026 Kantor Jurnal dan Publikasi UPI

ABSTRACT

Human Development Index (HDI) is an important indicator to measure the quality of human development across regions. This study aims to analyze factors influencing the Human Development Index of regencies/cities in Indonesia using the Generalized, Unbiased, Interaction Detection and Estimation (GUIDE) algorithm. The data used consists of socio-economic variables related to education, health, and economic conditions. The GUIDE algorithm is applied to identify significant predictor variables and interaction patterns that affect HDI values without bias in variable selection. The results show that several key factors, such as life expectancy, mean years of schooling, and expenditure per capita, play a dominant role in determining HDI across regions. The resulting regression tree provides an interpretable structure that illustrates the relationship between explanatory variables and HDI. This study demonstrates that the GUIDE algorithm is effective in modeling HDI data and can be used as an alternative analytical method for regional development analysis in Indonesia.

© 2026 Kantor Jurnal dan Publikasi UPI

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima 9 April 2026
Direvisi 29 Mei 2026
Disetujui 30 Mei 2026
Tersedia online 31 Mei 2026
Dipublikasikan 31 Mei 2026

Kata Kunci:

GUIDE,
Indeks Pembangunan Manusia (IPM),
Kesejahteraan masyarakat,
Pohon regresi.

Keywords:

GUIDE,
Human Development Index,
Regression tree,
Social welfare.

1. PENDAHULUAN

Indeks pembangunan manusia (IPM) merupakan indikator penting untuk mengukur keberhasilan pembangunan manusia di suatu daerah. Indeks Pembangunan Manusia (IPM) mencakup tiga dimensi utama, yaitu umur panjang dan hidup sehat, pendidikan, serta standar hidup layak (Alviani, 2021). Peningkatan nilai IPM Indonesia pada tahun 2024 mencerminkan adanya kemajuan dalam pembangunan manusia. Meskipun demikian, ketimpangan pembangunan antar wilayah masih menjadi tantangan. Wilayah dengan Pendapatan Asli Daerah (PAD) yang tinggi, seperti DKI Jakarta, cenderung memiliki nilai Indeks Pembangunan Manusia (IPM) lebih tinggi dibandingkan daerah yang masih bergantung pada dana dari pemerintah pusat (Mu'minah & Tjenreng, 2025). Oleh karena itu, pengidentifikasian faktor-faktor utama yang berpengaruh terhadap Indeks Pembangunan Manusia (IPM) penting dilakukan agar kebijakan pembangunan dapat dirancang dengan tepat sasaran di seluruh wilayah.

Salah satu metode yang dapat mengidentifikasi faktor-faktor yang mempengaruhi Indeks Pembangunan Manusia (IPM) dengan baik adalah metode berbasis pohon, di antaranya algoritma GUIDE (*Generalized, Unbiased, Interaction Detection and Estimation*). GUIDE bersifat nonparametrik, mampu mendeteksi interaksi lokal antar peubah saat pemilihan *split*, serta mengurangi bias pemilihan peubah yang umum terjadi pada algoritma pohon regresi klasik. Pohon regresi GUIDE menerapkan teknik *ensemble* dan *bagging*, serta mampu membangun empat jenis model, yaitu model konstan, linier berganda, linier bertatar (*stepwise*), dan polinomial (Jena & Dehuri, 2020).

Beberapa penelitian sebelumnya menunjukkan keunggulan algoritma GUIDE dalam membangun model regresi yang akurat dan menghindari bias pemilihan peubah. Loh *et al.*, (2019b) membandingkan 13 algoritma pohon dan hutan (*forest*) regresi dalam menganalisis contoh data tak lengkap dari survei, dan menyimpulkan bahwa baik pohon maupun hutan regresi GUIDE memiliki bias dan RMSE (*Root Mean Square Error*) yang lebih kecil daripada algoritma berbasis CART yang berbias terhadap peubah yang dapat di-*split* lebih banyak. Liu *et al.* (2019) menyimpulkan bahwa performa pohon regresi GUIDE lebih baik daripada tiga algoritma berbasis CART lainnya dalam mengidentifikasi subgrup pada kasus pengobatan presisi (*precision medicine*). Loh *et al.* (2020) mendemonstrasikan bahwa pohon dan hutan regresi GUIDE tidak menggunakan imputasi untuk data hilang yang kompleks. Namun demikian, GUIDE dapat digunakan untuk *impute*/mengisi data hilang untuk penggunaan lainnya, sebagai alternatif untuk AMELIA dan MICE. Untuk skema imputasi ini, GUIDE menghasilkan performa terbaik, diikuti oleh imputasi MICE dan imputasi mean/modus.

Di Indonesia, Isnani *et al.* (2019) menunjukkan bahwa pohon regresi GUIDE mampu menggambarkan struktur data secara jelas serta mengidentifikasi peubah kinerja program gizi yang signifikan pada prevalensi balita dengan gizi kurang. Sayangnya belum banyak penelitian menggunakan data di Indonesia yang menerapkan algoritma ini, sehingga diharapkan penelitian ini dapat mengisi rumpang yang ada.

Penelitian ini bertujuan untuk mengidentifikasi pola hubungan antara peubah bebas dan nilai Indeks Pembangunan Manusia (IPM) pada tingkat kabupaten/kota di Indonesia tahun 2024 menggunakan metode pohon regresi dengan algoritma GUIDE, menganalisis faktor-faktor yang paling berpengaruh terhadap IPM, serta membangun model prediksi IPM yang dihasilkan melalui metode pohon regresi dengan algoritma GUIDE. Penelitian seperti ini belum pernah dilakukan di Indonesia. Diharapkan model yang dihasilkan dari penelitian ini representatif untuk pengidentifikasian faktor-faktor utama yang mempengaruhi IPM, sehingga perencanaan pembangunan pemerintah pusat dan daerah tepat sasaran.

2. METODE

2.1. Algoritma GUIDE

Algoritma GUIDE (*Generalized, Unbiased, Interaction, Detection, and Estimation*) adalah metode pembelajaran mesin yang digunakan untuk membangun model pohon klasifikasi dan regresi dengan cara membagi pohon secara rekursif. Keunggulan GUIDE terletak pada kemampuannya untuk memilih peubah *split* (pembagi) secara tak bias, yang berarti tidak ada kecenderungan terhadap peubah yang memiliki lebih banyak nilai atau lebih mudah dibagi (Loh *et al.*, 2019b). Berikut ini adalah tahapan lengkap pembentukan pohon regresi algoritma GUIDE model konstan.

1. *Inisialisasi Model*

Inisialisasi model dilakukan dengan membentuk model dasar berupa model konstan sebagai titik awal proses partisi dan dasar perhitungan residu untuk mendeteksi kebutuhan pemisahan (*split*). Pada setiap simpul terminal, nilai prediksi ditentukan sebagai rata-rata peubah tak bebas (Y) dari data yang berada dalam simpul tersebut, yang sesuai untuk menangani hubungan nonlinier serta menghasilkan model yang mudah diinterpretasikan (Loh *et al.*, 2019b). Nilai prediksi awal dinyatakan sebagai rata-rata nilai Y , sebagaimana dirumuskan dalam persamaan (1) berikut:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (1)$$

Persamaan (1) menunjukkan bahwa nilai prediksi awal (\hat{Y}) dihitung dengan membagi jumlah nilai aktual peubah tak bebas ke-1 (Y_1) dengan total banyaknya data (n) yang tersedia dalam penelitian.

2. *Seleksi Peubah dengan Uji Kai-kuadrat*

Uji kai-kuadrat menurut Loh *et al.* (2019a) digunakan untuk pendeteksi peubah signifikan dalam pembelahan simpul sekaligus meminimalkan bias dengan mendeteksi non-linearitas melalui uji kurvatur dan ketergantungan antar peubah melalui uji interaksi.

a. *Uji kurvatur (Main Effect)*

Uji kurvatur dengan metode Kai-kuadrat berfungsi untuk mendeteksi adanya hubungan non-linear antara peubah bebas dan peubah tak bebas. Berikut langkah-langkah pengujian kurvatur dengan uji Kai-kuadrat.

1) Perhitungan sisaan awal

Adapun persamaan perhitungan sisaan awal sebagai persamaan (2) berikut:

$$e_i = Y_i - \hat{Y} \quad (2)$$

Persamaan (2) ditunjukkan dengan e_i adalah sisaan dari observasi ke- i , Y_i adalah nilai aktual peubah tak bebas ke- i , dan \hat{Y} adalah rata-rata keseluruhan dari peubah tak bebas atau nilai prediksi.

2) Tabel Kontingensi

Tabel kontingensi menyajikan frekuensi pengelompokan sisaan dan peubah yang dibagi menjadi kategori rendah serta tinggi berdasarkan nilai median. Tabel kontingensi 2 x 2 untuk uji kurvatur secara umum dapat dilihat pada Tabel 1.

Tabel 1. Tabel Kontingensi 2×2 Uji Kurvatur

		Kelompok Peubah Menurut Median		Total
		Rendah	Tinggi	
Kelompok Sisaan	+	O_{+1}	O_{+2}	b_1
	-	O_{-1}	O_{-2}	b_2
Total		a_1	a_2	N

(Sumber: Satria et al., 2023)

Keterangan:

O_{+j} : Frekuensi amatan pada kelompok sisaan positif (+) dan kelompok peubah ke- j

O_{-j} : Frekuensi amatan pada kelompok sisaan negatif (-) dan kelompok peubah ke- j

a_i : Total frekuensi kolom ke- j ($j = 1,2,3,4$)

b_i : Total frekuensi baris ke- i ($i = 1$ untuk positif, 2 untuk negatif)

N : Total keseluruhan frekuensi.

3) Statistik Kai-kuadrat

Perhitungan statistik uji Kai-kuadrat dapat dilakukan dengan formula sebagai persamaan (3) berikut:

$$\chi^2 = \sum_{i=1}^i \sum_{j=1}^j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

Pada persamaan (3) nilai statistik kai-kuadrat dinyatakan dengan χ^2 , indeks untuk baris ($1 =$ sisaan positif, $2 =$ sisaan negatif) dinyatakan dengan i , indeks untuk kolom ($1 =$ rendah, $2 =$ tinggi) dinyatakan dengan j , frekuensi observasi di baris ke- i kolom ke- j dinyatakan dengan O_{ij} , dan frekuensi harapan di baris ke- i kolom ke- j dinyatakan dengan E_{ij} , dengan perhitungan frekuensi harapan sebagai persamaan (4) berikut:

$$E_{ij} = \frac{(\text{total baris ke } - i) \times (\text{total kolom ke } - j)}{n} \quad (4),$$

n dinyatakan sebagai total amatan dalam simpul dengan derajat kebebasan untuk tabel $(i - 1)(j - 1)$.

4) Kriteria keputusan

Pengujian ini bertujuan untuk deteksi apakah sisaan menunjukkan pola non-*linear* terhadap suatu peubah bebas. Berikut analisis hipotesis uji kurvatur:

H_0 : Tidak ada pola non-*linear*

H_1 : Ada pola non-*linear* (signifikan).

Kriteria tolak H_0 jika memenuhi persamaan (5) berikut,

$$\chi^2 > \chi_{\alpha,df}^2 \text{ atau } p\text{-value} < \alpha, (\alpha = 0,05) \quad (5)$$

Persamaan (5) dengan χ^2 adalah nilai statistik kai-kuadrat, $\chi^2_{\alpha,df}$ adalah nilai kai-kuadrat tabel, α adalah tingkat signifikansi, dan df adalah derajat bebas $(i-1)(j-1)$.

b. Uji Interaksi

Uji interaksi dilakukan setelah uji kurvatur untuk mendeteksi apakah pengaruh suatu peubah terhadap peubah respon dipengaruhi oleh peubah bebas lain. Berikut langkah-langkah pengujian interaksi peubah dengan kai-kuadrat.

1) Sisaan awal

Perhitungan dapat dihitung menggunakan persamaan (2).

2) Tabel kontingensi

Tabel kontingensi menyajikan frekuensi pengelompokan sisaan dan peubah yang dibagi menjadi kategori rendah serta tinggi berdasarkan kombinasi antara nilai median dua peubah. Tabel kontingensi 2 x 4 untuk uji interaksi secara umum dapat dilihat pada Tabel 2.

Tabel 2. Tabel Kontingensi 2 × 4 Uji Interaksi

		Kelompok Kombinasi Median				Total
		k_1	k_2	k_3	k_4	
Kelompok Sisaan	+	O_{+1}	O_{+2}	O_{+3}	O_{+4}	b_1
	-	O_{-1}	O_{-2}	O_{-3}	O_{-4}	b_2
Total		a_1	a_2	a_3	a_3	N

(Sumber: Satria *et al.*, 2023)

Keterangan:

O_{+j} = Frekuensi amatan pada kelompok sisaan positif (+) dan kelompok peubah ke- j

O_{-j} = Frekuensi amatan pada kelompok sisaan positif (-) dan kelompok peubah ke- j

a_i = Total frekuensi kolom ke- j ($j = 1,2,3,4$)

b_i = Total frekuensi baris ke- i ($i = 1$ untuk positif, 2 untuk negatif)

N = Total keseluruhan frekuensi.

3) Statistik Kai-kuadrat

Perhitungan statistik uji kai-kuadrat uji interaksi dapat dilakukan dengan formula sebagai berikut:

$$\chi^2 = \sum_{i=1}^i \sum_{j=1}^j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

Pada persamaan (6) χ^2 adalah nilai statistik kai-kuadrat, i adalah indeks untuk baris (1 = sisaan positif, 2 = sisaan negatif), j adalah indeks untuk kolom (kelompok median), O_{ij} adalah frekuensi observasi (jumlah data aktual) di baris ke- i kolom ke- j , dan E_{ij} adalah frekuensi harapan di baris ke- i kolom ke- j .

4) Kriteria keputusan

Uji interaksi berasal dari interaksi nilai X_i & X_j . Berikut analisis hipotesis uji interaksi peubah:

H_0 : Tidak ada interaksi antara X_i & X_j .

H_1 : Ada interaksi.

Kriteria yang digunakan yaitu perbandingan nilai kai-kuadrat hitung (χ^2) dengan kai-kuadrat tabel ($\chi_{\alpha,df}^2$) maka:

Tolak H_0 jika,

$$\chi^2 > \chi_{\alpha,df}^2 \quad \text{atau} \quad p - \text{value} < \alpha, (\alpha = 0,05) \quad (7)$$

Persamaan (7) dengan χ^2 adalah nilai statistik kai-kuadrat, $\chi_{\alpha,df}^2$ adalah nilai kai-kuadrat tabel, α adalah tingkat signifikansi, dan df adalah derajat bebas $(i - 1)(j - 1)$.

c. Pemilihan peubah signifikan

Jika uji kai-kuadrat signifikan berasal dari uji kurvatur, peubah X akan dipilih langsung untuk menjadi peubah *split*. Sebaliknya, jika uji kai-kuadrat signifikan berasal dari uji interaksi, maka dua peubah (X_i, X_j) akan melalui proses pemilihan kembali dengan jumlah kuadrat galat (JKG) atau *Sum of Squared Errors* (SSE). Berikut langkah pemilihan peubah jika hasil uji kai-kuadrat signifikan berasal dari uji interaksi.

- 1) Pembagian data menjadi dua kelompok berdasarkan rata-rata X_i . Pembagian kelompok digambarkan dengan $X_i \leq \text{mean}(X_i)$ & $X_i > \text{mean}(X_i)$.
- 2) Perhitungan SSE dilakukan untuk kedua kelompok.
- 3) Pengulangan langkah (i) & (ii) untuk peubah X_j .
- 4) Total SSE dari kedua peubah (X_i dan X_j) dibandingkan. Peubah dengan total SSE lebih kecil terpilih menjadi peubah *split*.

3. Pemilihan Titik Split

GUIDE menentukan titik *split* yang optimal setelah peubah *split* telah ditentukan dengan uji kai-kuadrat. Berikut pendekatan utama penentuan titik *split* untuk peubah numerik (Loh et al., 2019c):

a. *Greedy Search* (G)

Metode *greedy search* pada GUIDE mengevaluasi seluruh titik potensial untuk menentukan titik *split* terbaik yang akurat bagi hubungan non-linear.

b. *Median split* (M)

Pembagian peubah *split* pada median X_j dan tidak perlu dihitung nilai SSE-nya. Pendekatan ini cepat dan cocok untuk data dengan distribusi tidak normal/distribusi miring (*skewed*).

4. Pembentukan Pohon Secara Rekursif

Setiap simpul baru dari pemisah sebelumnya dievaluasi kembali untuk menentukan apakah *split* lanjutan masih bisa dilakukan. Evaluasi ini dilakukan dengan pengulangan prosedur seleksi peubah dan titik *split*, seperti yang dikerjakan pada simpul akar. Berikut langkah-langkah pembentukan pohon rekursif (Loh et al., 2019a).

- 1) Nilai rata-rata untuk dugaan (prediksi) dihitung ulang berdasarkan data yang terdapat pada simpul pembelahan.
- 2) Nilai sisaan (residual) kemudian diperoleh dari selisih antara nilai IPM aktual dan rata-rata baru di simpul tersebut.
- 3) Uji kurvatur (*main effect*) dilakukan terhadap semua peubah bebas untuk mencari pengaruh terhadap sisaan pada simpul tersebut.
- 4) Jika terdapat peubah yang signifikan, maka pencarian titik *split* optimal dilakukan oleh algoritma dengan pendekatan *greedy search*.
- 5) Jika tidak ada *main effect* yang signifikan, maka uji interaksi dipertimbangkan oleh algoritma (jika syarat terpenuhi).
- 6) Proses diulang secara terus-menerus sehingga tidak ada lagi *split* yang layak dilakukan (tidak ada peubah yang signifikan).

5. Pemangkasan (*Pruning*)

Tujuan pemangkasan (*pruning*) pada pohon regresi GUIDE adalah penyederhanaan pohon serta pengurangan kompleksitas model tanpa adanya penurunan akurasi prediksi. Proses *pruning* pada algoritma GUIDE terdapat empat tahapan utama:

- 1) Pembentukan pohon maksimal
- 2) Perhitungan biaya kompleksitas
- 3) Pemangkasan bertahap
- 4) Validasi silang

6. Evaluasi Model

Evaluasi yang digunakan adalah *Mean Squared Error* (MSE) dan koefisien determinasi (R^2).

7. Hasil Akhir Pohon Regresi

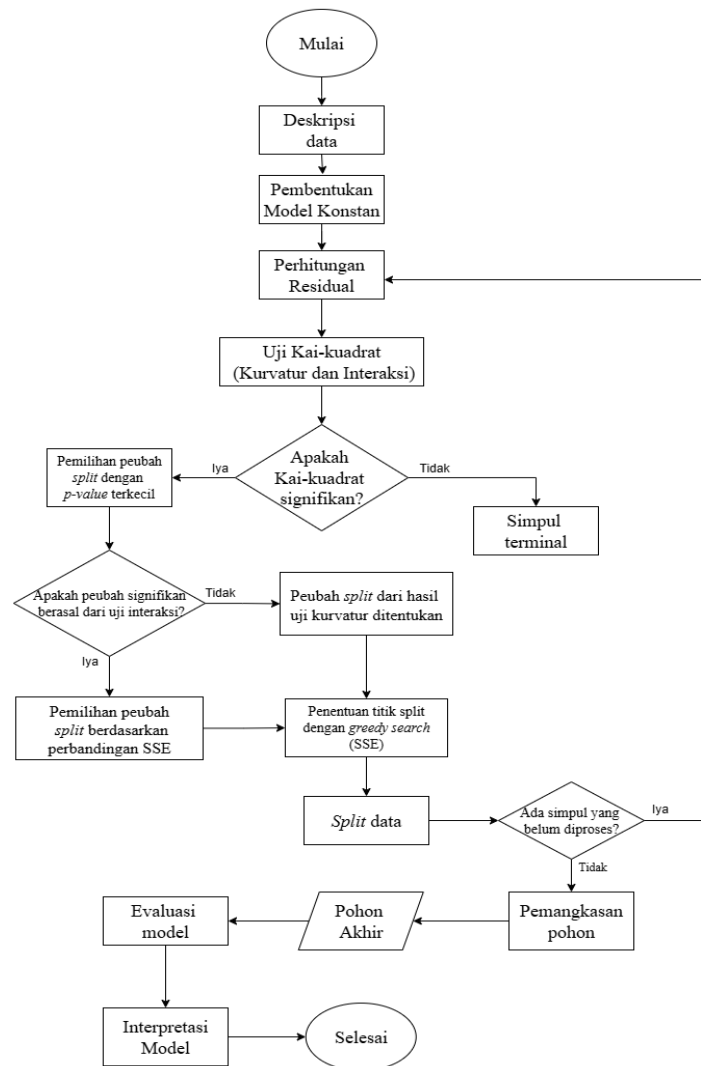
Algoritma GUIDE menghasilkan struktur pohon yang ringkas namun tetap mempertahankan *split-split* penting.

2.2 Tahapan Analisis

Pohon regresi GUIDE dibentuk secara bertahap, dimulai dari simpul akar. Pemilihan peubah *split* terbaik dan pemilihan titik *split* terbaik. Proses berlanjut hingga terbentuk simpul terminal, lalu dilakukan pemangkasan bertahap. Terakhir, model dievaluasi dan diinterpretasikan. Tahapan penelitian ini dilakukan dengan bantuan *software* GUIDE ver. 44.0. Tahapan pembentukan pohon regresi dengan algoritma GUIDE disajikan pada Gambar 1.

Diagram alir pada Gambar 1 menunjukkan tahapan penelitian dalam pemodelan Indeks Pembangunan Manusia (IPM) menggunakan pohon regresi dengan algoritma GUIDE. Penelitian diawali dengan tahap persiapan data. Data tersebut selanjutnya digunakan sebagai dasar dalam proses pembentukan model.

Tahap berikutnya adalah pembentukan model awal pada simpul akar berupa model konstan. Berdasarkan model awal tersebut, dilakukan perhitungan nilai sisaan untuk setiap amatan yang kemudian digunakan sebagai dasar dalam proses seleksi peubah. Seleksi peubah dilakukan melalui uji Kai-kuadrat, yang meliputi uji kurvatur untuk mendeteksi pengaruh utama peubah serta uji interaksi untuk mengidentifikasi adanya interaksi antarpeubah. Peubah atau pasangan peubah yang menunjukkan signifikansi statistik dipertimbangkan sebagai kandidat peubah pemisah (*split*).



Gambar 1. Diagram Alir Tahapan Analisis.

Apabila peubah yang signifikan berasal dari uji interaksi, maka penentuan peubah *split* dilakukan dengan membandingkan nilai *Sum of Squared Error* (SSE), sedangkan apabila berasal dari uji kurvatur, peubah dengan nilai statistik Kai-kuadrat terbesar dipilih secara langsung sebagai peubah *split*. Setelah peubah *split* ditentukan, tahap selanjutnya adalah penentuan titik *split* optimal menggunakan pendekatan *greedy search* dengan kriteria minimisasi SSE. Data kemudian dibagi menjadi dua subset berdasarkan peubah dan titik *split* yang terpilih.

Proses pembentukan simpul dan pemisahan data dilakukan secara rekursif hingga memenuhi kriteria pemberhentian, sehingga diperoleh pohon regresi maksimal. Untuk menghindari terjadinya *overfitting*, dilakukan tahap pemangkasan pohon (*pruning*) menggunakan pendekatan validasi silang. Proses pemangkasan melibatkan perhitungan biaya kompleksitas (*cost-complexity*), pemangkasan bertahap secara *bottom-up*, serta pemilihan pohon optimal menggunakan aturan satu galat baku (*1-SE rule*).

Tahap akhir penelitian adalah evaluasi dan interpretasi model pohon regresi final. Evaluasi kinerja model dilakukan menggunakan *Mean Squared Error* (MSE) dan koefisien determinasi (R^2), sedangkan interpretasi dilakukan dengan menganalisis struktur pohon, titik *split*, nilai prediksi pada setiap simpul, serta pola interaksi antar peubah yang mempengaruhi IPM.

3. HASIL DAN PEMBAHASAN

3.1 Data Penelitian

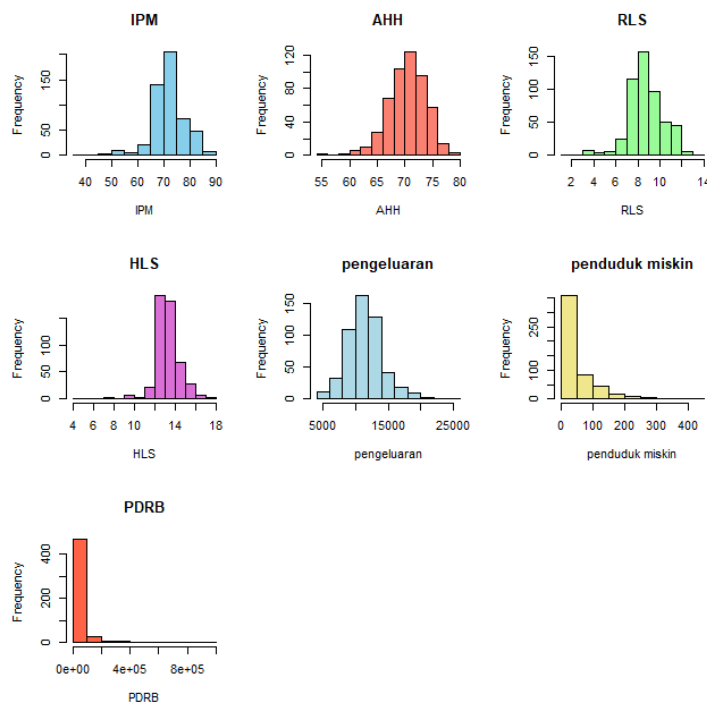
Penelitian ini menggunakan data sekunder Indeks Pembangunan Manusia (IPM) tahun 2024 dari Badan Pusat Statistik (BPS) yang mencakup 514 amatan, dengan rincian satu peubah respon dan enam peubah bebas. Adapun peubah dari penelitian disajikan pada Tabel 3.

Tabel 3. Peubah Penelitian

Peubah	Keterangan	Satuan
Y	Indeks Pembangunan Manusia	Nilai
X_1	Angka Harapan Hidup	Rataan Tahun
X_2	Rataan Lama Sekolah	Rataan Tahun
X_3	Harapan Lama Sekolah	Rataan Tahun
X_4	Pengeluaran per kapita	Ribu Rupiah
X_5	Jumlah Penduduk Miskin	Jiwa
X_6	PDRB	Milyar rupiah

3.2. Pembahasan

Eksplorasi data adalah langkah pertama yang dilakukan untuk mengetahui karakteristik umum dari masing-masing peubah.



Gambar 2. Histogram Semua Peubah.

Berdasarkan Gambar 2, penyebaran peubah menunjukkan beberapa pola penyebaran yang berbeda, terutama pada data yang memiliki kemiringan (*skewed*). Model konstan dipilih berdasarkan bentuk sebaran yang tidak selalu normal, karena dalam prosesnya dilakukan dengan perhitungan rata-ran di setiap simpulnya. Hal ini memungkinkan pola data untuk tetap tertangkap tanpa terpengaruh oleh sebaran data yang tidak normal (*skewed*).

Pembentukan model awal pada simpul akar dengan model konstan dinyatakan dengan

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{514} \sum_{i=1}^{514} Y_i = 71,9189$$

Berdasarkan perhitungan di atas, telah terbentuk simpul akar dengan nilai rata-rata IPM sebesar 71,9189 dari 514 kabupaten/kota di Indonesia. Nilai ini selanjutnya akan digunakan sebagai nilai penduga (\hat{Y}) atau prediksi pada simpul akar, serta untuk perhitungan sisaan. Sehingga didapatkan perhitungan sisaan pada simpul akar yang ditampilkan pada Tabel 4.

Tabel 4. Sisaan Peubah Y

No	IPM	\hat{Y}	$e_i = Y_i - \hat{Y}$
1	69,01	71,9189	-2,9089
2	70,75	71,9189	-1,1689
3	69,1	71,9189	-2,8189
⋮	⋮	⋮	⋮
512	51,85	71,9189	-20,0689
513	51,98	71,9189	-19,9389
514	49,36	71,9189	-22,5589

Pola sisaan pada Tabel 4 akan dianalisis pada tahap berikutnya untuk penentuan peubah yang paling berpengaruh terhadap IPM.

Tahapan pemilihan peubah *split* ini ditentukan dari peubah bebas paling berpengaruh terhadap sisaan. Dua uji akan dilakukan terhadap semua peubah bebas, yaitu uji kurvatur (*main effect*) dan uji interaksi antar peubah.

a. Uji kurvatur

Uji kurvatur dilakukan terhadap semua peubah pada simpul dengan uji kai-kuadrat. Ilustrasi perhitungan menggunakan peubah Pengeluaran per Kapita. Tabel kontingensi untuk peubah ini dapat dilihat pada Tabel 5.

Tabel 5. Tabel Kontingensi 2x2 Uji Kurvatur Pengeluaran per Kapita pada Simpul Akar

		Pengeluaran per kapita		Total
		Rendah	Tinggi	
Kelompok Sisaan	+	41	199	240
	-	230	44	274
Total		271	243	514

Frekuensi harapan dihitung berdasarkan Tabel 5. Perhitungan untuk sel kelompok sisaan (+) dan pengeluaran per kapita rendah adalah sebagai berikut:

$$E_{11} = \frac{240 \times 271}{514} = 126,54$$

Prosedur yang sama diterapkan untuk seluruh kombinasi sel kelompok sisaan dan kelompok pengeluaran per kapita lainnya. Hasil lengkap dari perhitungan frekuensi harapan disajikan pada Tabel 6.

Tabel 6. Hasil Frekuensi harapan Uji Interaksi Peubah Pengeluaran per Kapita

E_{ij}	Rendah	Tinggi
+	126,54	113,46
-	144,46	129,54

Tahapan setelah memperoleh frekuensi harapan (E_{ij}) pada masing-masing sel, langkah berikutnya adalah perhitungan nilai statistik Kai-kuadrat (χ^2). Sehingga nilai (χ^2) yang diperoleh:

$$\chi^2 = \sum_{i=1}^i \sum_{j=1}^j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 227,1$$

$$\text{sehingga } \chi^2 = 227,1 > \chi^2_{\alpha,df} = 3,8415$$

Sebagai hasilnya, hipotesis nol ditolak, yang berarti terdapat hubungan yang signifikan antar kelompok pengeluaran per kapita dan kelompok sisaan. Tabel 7 menampilkan nilai statistik kai-kuadrat pada simpul akar pada semua peubah.

Tabel 7. Nilai Statistik Kai-kuadrat pada Simpul Akar

Rank	Peubah	χ^2	$\chi^2_{0,05,1}$	Keterangan
1	Pengeluaran per kapita	227,1	3,841	Signifikan
2	Rataan Lama Sekolah (RLS)	174,7	3,841	Signifikan
3	Angka Harapan Hidup (AHH)	160,4	3,841	Signifikan
4	Harapan Lama Sekolah (HLS)	116,0	3,841	Signifikan
5	PDRB	74,28	3,841	Signifikan
6	Jumlah Penduduk Miskin	0,27	3,841	Tidak Signifikan

Berdasarkan Tabel 7, peubah dengan nilai statistik kai-kuadrat (χ^2) yang paling tinggi dinyatakan sebagai peubah paling berpengaruh terhadap sisaan. Hal ini menjadi indikasi bahwa pengaruh terbesar peubah bebas terhadap sisaan pada simpul akar diberikan oleh pengeluaran per kapita.

b. Uji interaksi

Sebagai ilustrasi, dilakukan pengujian interaksi terhadap peubah pengeluaran per kapita dan Rataan Lama Sekolah (RLS) terhadap sisaan pada simpul akar. Kedua peubah dikategorikan menjadi dua kelompok berdasarkan median dengan pengeluaran memiliki batas 11.305 dan RLS memiliki batas 8,62. Kombinasi yang dihasilkan dari kedua peubah adalah empat kelompok gabungan, yaitu:

$$k_1 = \text{pengeluaran} \leq 11.305 \ \& \ \text{RLS} \leq 8,62$$

$$k_2 = \text{pengeluaran} \leq 11.305 \ \& \ \text{RLS} > 8,62$$

$$k_3 = \text{pengeluaran} > 11.305 \ \& \ \text{RLS} \leq 8,62$$

$$k_4 = \text{pengeluaran} > 11.305 \ \& \ \text{RLS} > 8,62$$

Tabel 8. Hasil Tabel Kontingensi 2×4 Uji Interaksi

		Pengeluaran per kapita & RLS				Total
		k_1	k_2	k_3	k_4	
Kelompok sisaan	+	10	39	75	116	240
	-	120	88	54	12	274
Total		130	127	129	128	514

Tabel 8 menampilkan tabel kontingensi 2×4 untuk uji interaksi. Sebagai ilustrasi, perhitungan untuk sel kelompok sisaan (+) dan kombinasi k_1 adalah sebagai berikut:

$$E_{11} = \frac{240 \times 130}{514} = 60,70$$

Prosedur yang sama diterapkan untuk seluruh kombinasi sel kelompok sisaan dan kombinasi pengeluaran dan RLS lainnya. Hasil lengkap dari perhitungan frekuensi harapan disajikan pada Tabel 9.

Tabel 9. Hasil Frekuensi harapan Uji Interaksi

E_{ij}	k_1	k_2	k_3	k_4
+	60,70	59,30	60,23	59,77
-	69,30	67,95	69,02	68,48

Tahapan setelah memperoleh frekuensi harapan (E_{ij}) pada masing-masing sel, adalah perhitungan nilai statistik Kai-kuadrat (χ^2). Maka nilai (χ^2) yang diperoleh:

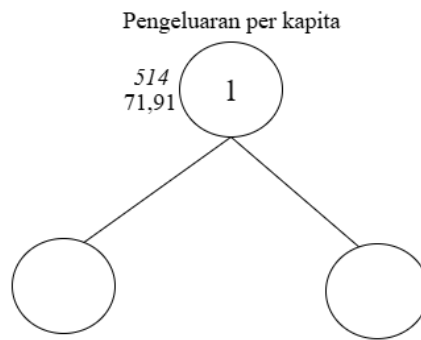
$$\chi^2 = \sum_{i=1}^i \sum_{j=1}^j \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 198,62$$

$$\text{sehingga } \chi^2 = 198,62 > \chi^2_{\alpha,df} = 7,8127$$

Sebagai hasilnya, hipotesis nol ditolak, artinya interaksi pengeluaran per kapita dan RLS secara statistik berhubungan dengan kelompok sisaan, sehingga layak digunakan sebagai peubah *split* pada simpul akar pembentukan pohon regresi GUIDE.

Peubah Pengeluaran per Kapita terpilih menjadi peubah *split* pada simpul akar dikarenakan pada uji kurvatur peubah tersebut sudah signifikan, sehingga tidak perlu menggunakan keputusan dari uji interaksi. Ilustrasi pembentukan peubah *split* pada simpul akar ditunjukkan pada Gambar 3. Angka 514 merupakan banyaknya amatan (n) yaitu kota/kabupaten di seluruh Indonesia, sedangkan angka 71,91 merupakan nilai rata-rata IPM dari ke-514 kota/kabupaten tersebut.

Tahapan setelah penetapan peubah *split* pada simpul akar adalah penentuan titik *split* optimal. Proses ini dilakukan dengan pendekatan *greedy search*, yang berarti semua nilai unik pada peubah pengeluaran per kapita akan dijadikan calon titik *split* dengan pemilihan titik berdasarkan pengurangan nilai *Sum of Squared Errors* (SSE) terbesar. Berikut disajikan ilustrasi daftar pengurangan SSE dari beberapa titik unik pada peubah pengeluaran per kapita pada Tabel 10.



Gambar 3. Simpul Akar dengan Peubah Split Pengeluaran per Kapita.

Tabel 10. Ilustrasi Hasil Penurunan SSE pada Peubah Pengeluaran per Kapita

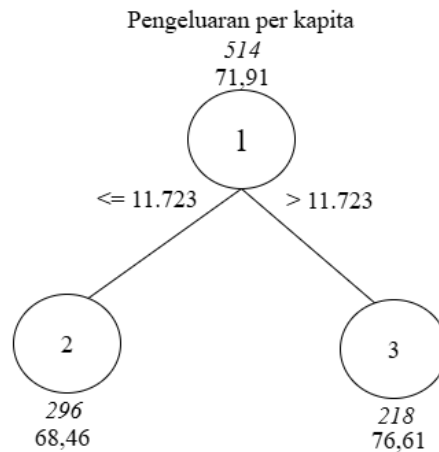
No.	Split pada peubah Pengeluaran $X_i \leq c < X_i$	SSE Total sebelum split	SSE Total Setelah split	Penurunan SSE
1	11.723	20.281,53	11.927,52	8.354,01
2	11.805	20.281,53	12.052,59	8.228,94
3	12.625	20.281,53	12.474,57	7.806,96
⋮	⋮	⋮	⋮	⋮
498	11.274	20.281,53	12.309,27	7.972,26
499	13.133	20.281,53	12.653,25	7.628,28

Berdasarkan Tabel 10, penurunan SSE paling besar diperoleh pada titik 11.723, yaitu sebesar 8.354,01. Gambar 4 menunjukkan ilustrasi pembentukan simpul akar setelah penentuan titik split. Pada Gambar 4 ini terlihat bahwa simpul akar (1) merupakan peubah split pertama dengan titik split pengeluaran per kapita 11.723 ribu rupiah, atau Rp 11,7 juta. Split ke kiri merupakan simpul dengan pengeluaran per kapita kurang dari atau sama dengan titik split. Dengan kata lain, Simpul 2 merupakan simpul dengan pengeluaran per kapita ≤ 11.723 , yaitu kurang dari atau sama dengan Rp 11,7 juta. Simpul 2 memiliki 296 amatan dengan nilai rataan IPM 68,46. Berikutnya, split ke kanan merupakan simpul dengan pengeluaran per kapita lebih dari titik split. Simpul 3 ini merupakan simpul dengan pengeluaran per kapita lebih dari 11.723, atau lebih dari Rp 11,7 juta. Simpul 3 memiliki 218 amatan dengan nilai rataan IPM 76,61.

Tahapan selanjutnya adalah pembentukan pohon secara rekursif menjadi pohon yang maksimal. pohon maksimal dihasilkan oleh proses pembentukan pohon secara rekursif, dengan total 16 simpul terminal dan Mean Square Error (MSE) sebesar 5,824.

Proses pemangkasan dilakukan secara bertahap dari pohon maksimal hingga menyisakan satu simpul berdasarkan penurunan biaya terbesar, menghasilkan 16 variasi subpohon yang kemudian dievaluasi. Pohon optimal selanjutnya dipilih menggunakan metode 10-fold cross-validation dengan membagi 514 amatan secara acak untuk menguji performa model pada

setiap iterasi secara adil. Tabel 11 merangkum hasil rata-rata MSE setiap jumlah simpul terminal yang telah melewati validasi silang 10-fold.



Gambar 4. Simpul Akar dengan Peubah *Split* dan Titik *Split*.

Tabel 11. Ringkasan Validasi Silang (10-fold Cross-Validation)

Jumlah Simpul Terminal	MSE Rataan	SE	Keterangan
16	5,824	0,995	Pohon Maksimal
15	5,884	0,995	Pohon Pemangkasan
14	5,891	0,996	Pohon Pemangkasan
13	5,948	0,998	Pohon Pemangkasan
12	6,272	1,003	Pohon Pemangkasan
11	6,893	1,013	Pohon Pemangkasan
10	6,885	1,013	Pohon Pemangkasan
9	7,351	1,019	Pohon Pemangkasan
8	7,695	1,024	Pohon Pemangkasan
7	8,099	1,031	Pohon Pemangkasan
6	8,391	1,033	Pohon Pemangkasan
5	9,371	1,068	Pohon Pemangkasan
4	10,22	1,096	Pohon Pemangkasan
3	16,18	1,299	Pohon Pemangkasan
2	24,03	2,958	Pohon Pemangkasan
1	39,48	1,118	Pohon Pemangkasan

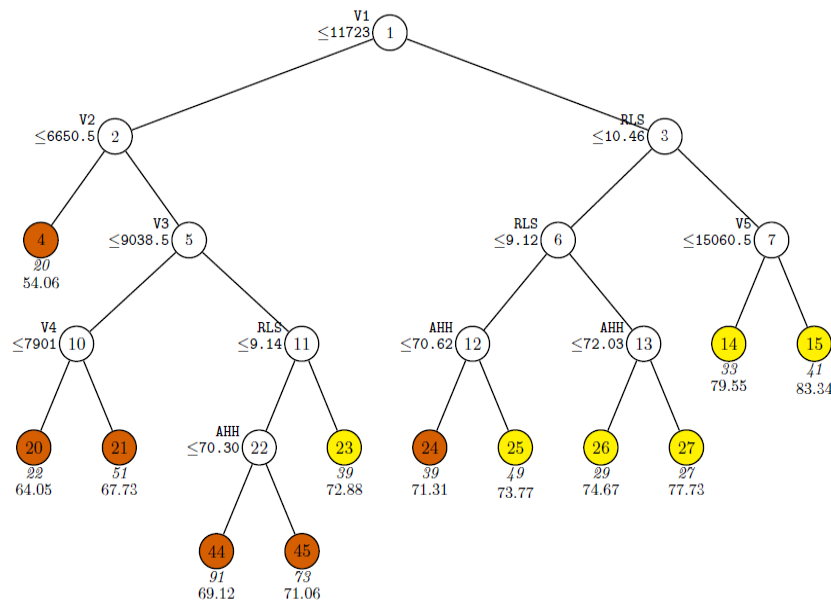
Berdasarkan Tabel 11, diketahui bahwa MSE terkecil diperoleh pada pohon maksimal dengan 16 simpul terminal, yaitu sebesar 5,824 dengan (*Standard Error*) SE sebesar 0,995. Pemilihan struktur pohon yang paling optimal juga ditentukan tidak hanya nilai MSE terkecil, namun digunakan juga pendekatan aturan 1-*Standard Error* (1-SE) yang dipertimbangkan untuk kestabilan performa model. Oleh karena itu, batas atas MSE dihitung sebagai berikut:

$$MSE_{batas\ atas} = MSE_{terkecil} + SE$$

$$MSE_{batas\ atas} = 5,824 + 0,9994 = 6,8234$$

Bedasarkan hasil batas atas MSE untuk aturan 1-SE, pohon dengan jumlah simpul lebih sedikit dengan nilai MSE yang tidak melebihi batas atas tersebut diidentifikasi. Pada Tabel 11, teridentifikasi bahwa pohon dengan 12 simpul terminal memiliki nilai MSE 6,272, yang masih berada dalam batas toleransi.

Pohon akhir yang terpilih menjadi pohon optimal memiliki 23 simpul dengan 12 simpul terminal dan MSE sebesar 6,272; pohon ini ditunjukkan pada Gambar 5.



Gambar 5. Hasil Pohon Regresi Model Konstan Algoritma GUIDE.

Selain simpul akar (1), simpul berwarna putih merupakan simpul antara yang juga merupakan peubah split. $V1, V2, V3, V4$ adalah penamaan yang diberikan oleh algoritma GUIDE pada peubah pengeluaran per kapita. Terdapat dua penduga lain yang menjadi peubah split selain pengeluaran per kapita, yaitu RLS dan AHH. Simpul terminal ditandai dengan warna berbeda. Terdapat enam simpul terminal dengan warna merah dan enam simpul dengan warna kuning. Warna merah menunjukkan simpul dengan nilai prediksi rata-ran IPM $\leq 71,91$, yang dapat dikategorikan sebagai IPM rendah; sedangkan warna kuning menunjukkan simpul dengan nilai prediksi rata-ran IPM lebih dari 71,91, yang dapat dikategorikan sebagai IPM tinggi. Angka yang berada di dalam lingkaran simpul adalah label simpul. Angka di bawah bulatan simpul menunjukkan banyaknya amatan pada simpul (angka miring) dan nilai prediksi rata-ran IPM.

Sebagai ilustrasi, Simpul terminal 4 berwarna merah merupakan kategori IPM rendah; terdiri dari 20 kota/kabupaten dengan pengeluaran per kapita $\leq 6650,5$ atau kurang dari atau sama dengan Rp 6,6 juta, dan prediksi rata-ran IPMnya adalah 54,06. Simpul terminal 15 berwarna kuning merupakan kategori IPM tinggi; terdiri dari 41 kota/kabupaten dengan rata-ran lama sekolah lebih dari 10,46 tahun dan pengeluaran per kapita $> 15.060,5$ atau lebih dari Rp 15,06 juta, dan prediksi rata-ran IPMnya adalah 83,34. Simpul terminal 44 berwarna merah merupakan kategori IPM rendah; terdiri dari 91 kota/kabupaten dengan pengeluaran per kapita lebih dari Rp 9,04 juta, rata-ran lama sekolah $\leq 9,14$ tahun, angka harapan hidup $\leq 70,3$ tahun, dan prediksi rata-ran IPMnya adalah 69,12. Secara keseluruhan, terdapat 296 kota/kabupaten dengan prediksi kategori IPM rendah, serta 218 kota/kabupaten dengan prediksi kategori IPM tinggi.

Evaluasi model dari pohon regresi model konstan dengan 12 simpul terminal ini dilakukan dengan koefisien determinasi (R^2). Perhitungan koefisien determinasi (R^2) adalah sebagai berikut:

$$R^2 = 1 - \frac{SSE}{SST} = 0,8883$$

Koefisien determinasi (R^2) yang diperoleh sebesar 0,8883. Nilai ini menunjukkan bahwa hamper 89% variasi nilai IPM dalam data dapat dijelaskan oleh model.

4. KESIMPULAN

Simpulan dari penelitian adalah algoritma GUIDE dengan model konstan berhasil mengidentifikasi pola hubungan non-linear pada IPM, di mana peningkatan pengeluaran per kapita, Rataan Lama Sekolah (RLS), dan Angka Harapan Hidup (AHH) secara signifikan mendorong kenaikan nilai IPM. Pengeluaran per kapita ditemukan sebagai faktor yang paling berpengaruh, diikuti oleh aspek pendidikan dan kesehatan, sebagaimana tercermin dalam urutan pembentukan pohon regresi. Model akhir yang dihasilkan memiliki struktur 23 simpul dengan 12 simpul terminal, serta menunjukkan performa yang andal dengan nilai R^2 sebesar 88,8% dan *Mean Squared Error* (MSE) sebesar 6,272, sehingga sangat layak digunakan sebagai instrumen prediksi maupun landasan pengambilan kebijakan pembangunan manusia.

5. DAFTAR PUSTAKA

- Alviani, L. O. (2021). Penggunaan regresi data panel pada analisis Indeks Pembangunan Manusia. *Jurnal Riset Matematika*, 1(2), 99-108.
- Isnani, Rahmi., I. H. G., and Asdi, Y. (2019). Penerapan metode pohon regresi stepwise linear dengan algoritma guide dalam menalisis pengaruh kinerja program gizi terhadap prevalensi underweight di Indonesia. *Jurnal Matematika UNAND*, 8(1), 110-119.
- Jena, M. and Dehuri, S. 2020. Decision tree for classification and regression: A state-of-the art review. *Informatica*, 44(4), 405-420.
- Liu, Y., Ma, X., Zhang, D., Geng, L., Wang, X., Zheng, W., and Chen, M. -H. 2019. Look before you leap: Systematic evaluation of tree-based statistical methods in subgroup identification. *Journal of Biopharmaceutical Statistics*, 29(6), 1082-1102.
- Loh, W. -Y., Cao, L., and Zhou, P. (2019a). Subgroup identification for precision medicine: A comparative review of 13 methods. *WIREs Data Mining Knowledge Discovery*, 9(5), e1326.
- Loh, W. -Y., Eltinge, J., Cho, M. J., and Li, Y. (2019b). Classification and regression trees and forests for incomplete data from sample surveys. *Statistica Sinica*, 29(1), 431-453.
- Loh, W. -Y., Man, M., and Wang, S. (2019c). Subgroup from regression trees with adjustment for prognostic effects and postselection inference. *Statistics in Medicine*, 38(4), 545-557.
- Loh, W. -Y., Zhang, Q., Zhang, W., and Zhou, P. (2020). Missing data, imputation and regression trees. *Statistica Sinica*, 30(4), 1697-1722.
- Mu'minah, S. and Tjenreng, M. B. Z. (2025). Desentralisasi dan ketimpangan pembangunan antar daerah. *Scientific Journal of Reflection: Economic, Accounting, Management and Business*, 8(1), 342-351.

Satria, T. A. I, Imro'ah, N., and Huda, N. M. (2023). Penerapan model log linier dalam menganalisis tabel kontingensi dua arah. *Buletin Ilmiah Matematika, Statistika dan Terapannya (BIMASTER)*, 12(6), 545-552.