

OPTIMASI ELASTIC NET REGRESSION DENGAN OPTUNA UNTUK ANALISIS FAKTOR KEMISKINAN DI INDONESIA

Yunita Nur, Kartika Maulida Hindrayani, Aviolla Terza Damaliana

Sains Data, Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur, Indonesia
kartika.maulida.ds@upnjatim.ac.id

ABSTRAK

Kemiskinan di Indonesia merupakan permasalahan kompleks yang dipengaruhi oleh berbagai faktor sosial dan ekonomi. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi tingkat kemiskinan di Indonesia tahun 2023 dengan menggunakan metode *Elastic Net Regression* yang dioptimalkan melalui algoritma *Optuna*. Pendekatan ini dipilih karena mampu mengatasi multikolinearitas dan menghasilkan model prediktif yang lebih stabil dan akurat. Data yang digunakan merupakan data sekunder dari Badan Pusat Statistik (BPS) yang mencakup delapan variabel independen: Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Produk Domestik Regional Bruto (PDRB), Indeks Pembangunan Manusia (IPM), Partisipasi Angkatan Kerja (PAK), Rata-rata Lama Sekolah (RLS), Indeks Kemahalan Konstruksi (IKK), dan Pernikahan Dini. Hasil analisis menunjukkan bahwa model terbaik, yaitu *Elastic Net Regression* yang dioptimasi dengan *Optuna*, memiliki nilai R^2 sebesar 0,7155, MAE sebesar 3,1405, dan MSE sebesar 16,1338, serta memenuhi seluruh asumsi regresi klasik. Enam variabel signifikan ditemukan, yakni IPM, AHH, AMH, dan Pernikahan Dini yang berpengaruh negatif, serta RLS dan PAK yang berpengaruh positif terhadap kemiskinan. Temuan ini menekankan pentingnya pembangunan manusia, pendidikan, literasi, akses kerja, dan pencegahan pernikahan dini dalam upaya penanggulangan kemiskinan secara berkelanjutan.

Kata kunci : *Elastic Net Regression, Kemiskinan, Optimasi Hyperparameter Optuna*

1. PENDAHULUAN

Indonesia sebagai negara berkembang masih menghadapi tantangan kemiskinan yang relatif tinggi [1]. Kemiskinan dipahami sebagai ketidakmampuan dalam memenuhi kebutuhan dasar seperti pangan, tempat tinggal, dan layanan kesehatan [2] [3]. Data dari Badan Pusat Statistik (BPS) mencatat bahwa tingkat kemiskinan mengalami penurunan dari 9,54% pada Maret 2022 menjadi 9,36% pada Maret 2023 [4] [5]. Pemerintah telah menjalankan berbagai program, seperti perluasan lapangan kerja dan peningkatan layanan publik, sebagai upaya pengentasan kemiskinan [6] [7]. Namun demikian, ketimpangan antar daerah tetap tinggi, terutama di wilayah pedesaan dan terpencil yang masih mengalami keterbatasan akses terhadap pendidikan, kesehatan, dan pekerjaan layak.

Ketimpangan tersebut mencerminkan bahwa kemiskinan di Indonesia bersifat multidimensi dan kompleks, sehingga tidak cukup hanya diatasi dengan pendekatan konvensional. Berbagai upaya yang telah dilakukan masih belum sepenuhnya mampu menjawab kompleksitas permasalahan yang melibatkan banyak faktor yang saling berkaitan. Oleh karena itu, diperlukan pendekatan analitis yang mampu menangkap hubungan antar variabel secara akurat sekaligus menghasilkan model yang andal untuk prediksi kebijakan. *Elastic Net Regression* menjadi pilihan karena kemampuannya mengatasi permasalahan multikolinearitas dan melakukan seleksi variabel secara lebih efisien [8]. Untuk meningkatkan performa model, digunakan *Optuna*, sebuah framework optimasi hyperparameter berbasis

Bayesian yang menawarkan efisiensi tinggi dibandingkan metode pencarian konvensional seperti grid search. *Optuna* memanfaatkan algoritma *Tree-structured Parzen Estimator (TPE)* untuk secara adaptif mengeksplorasi kombinasi parameter terbaik dalam ruang pencarian [9] [10]. Kombinasi keduanya diharapkan mampu menghasilkan model yang tidak hanya akurat, tetapi juga interpretatif, sehingga dapat memberikan rekomendasi kebijakan yang lebih tepat sasaran dalam mendukung program pengentasan kemiskinan di Indonesia.

Penelitian ini menganalisis delapan variabel berdasarkan temuan empiris dan pendekatan pembangunan sosial ekonomi, yaitu Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Produk Domestik Regional Bruto (PDRB), Indeks Pembangunan Manusia (IPM), Partisipasi Angkatan Kerja (PAK), Rata-Rata Lama Sekolah (RLS), Indeks Kemahalan Konstruksi (IKK), dan Pernikahan Dini. Guntur Wahyu K. dkk. dan Fitri Mudia S. dkk. dalam penelitiannya menemukan bahwa Angka Harapan Hidup dan Angka Melek Huruf berpengaruh signifikan terhadap kemiskinan, sementara Lestari dkk. menyoroti ketimpangan akses pendidikan sebagai faktor utama kemiskinan ekstrem [8] [1] [11]. Titis Hana P. dkk. menjelaskan bahwa Produk Domestik Regional Bruto dan Partisipasi Angkatan Kerja mencerminkan aspek ekonomi dan akses kerja, sedangkan Faisal Agymnastiar R. F. dkk. menekankan pentingnya Indeks Pembangunan Manusia [12] [13]. Indeks Kemahalan Konstruksi disusun oleh Badan Pusat Statistik sebagai indikator biaya pembangunan fisik daerah, dan Bappenas menyatakan pernikahan

dini sebagai bentuk kerentanan sosial yang berdampak pada kesejahteraan usia produktif [14] [15].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi kemiskinan di Indonesia pada tahun 2023 dengan menggunakan model Elastic Net Regression yang dioptimasi menggunakan Optuna. Hasil dari penelitian ini diharapkan dapat menjadi dasar bagi pemerintah dalam merumuskan kebijakan yang lebih tepat di bidang pendidikan, kesehatan, ketenagakerjaan, dan pencegahan pernikahan dini, sehingga dapat mendukung upaya pengentasan kemiskinan secara lebih efektif.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Menurut penelitian sebelumnya, Guntur Wahyu K. dkk menggunakan metode Elastic Net Regression untuk menganalisis kerentanan kemiskinan dan menemukan bahwa 10,4% keluarga tergolong miskin dan 4,8% berada dalam kategori rentan miskin. Variabel Angka Harapan Hidup terbukti signifikan, menunjukkan bahwa kesehatan masyarakat berperan penting dalam menurunkan kemiskinan. Selain itu, sektor pertanian serta pengeluaran untuk pendidikan dan kesehatan turut memberikan kontribusi. Elastic Net juga terbukti efektif dalam mengatasi multikolinearitas dan menghasilkan model yang stabil [8]. Sementara itu, Fitri Mudia S. dkk meneliti pengaruh Angka Melek Huruf terhadap kemiskinan dan menemukan bahwa peningkatan literasi sangat berkorelasi dengan penurunan tingkat kemiskinan. Studi tersebut menggunakan metode Regresi Ridge yang mampu menangani korelasi antar variabel yang tinggi dan menghasilkan model yang robust [1].

Meskipun beberapa penelitian sebelumnya telah menunjukkan faktor-faktor signifikan yang memengaruhi tingkat kemiskinan, namun studi tersebut umumnya masih terbatas pada wilayah tertentu dan belum mengintegrasikan pendekatan optimasi hyperparameter yang komprehensif. Oleh karena itu, penelitian ini berupaya melengkapi temuan terdahulu dengan menerapkan metode Elastic Net Regression yang dikombinasikan dengan Optuna, guna menganalisis faktor-faktor yang memengaruhi kemiskinan secara nasional pada tahun 2023. Pendekatan ini diharapkan mampu memberikan hasil analisis yang lebih akurat dan relevan untuk mendukung perumusan kebijakan pengentasan kemiskinan yang efektif dan tepat sasaran.

2.2. Linier Regression

Linear Regression adalah metode statistik untuk memodelkan hubungan antara satu variabel dependen (Y) dan satu atau lebih variabel independen (X). Tujuannya adalah memprediksi nilai Y berdasarkan nilai X [16]. Berikut rumus regresi linier sederhana:

$$Y = \beta_0 + \beta_i X_i + \epsilon_i \quad [1]$$

Keterangan:

Y_i = Variabel dependen yang diprediksi

X_i = Variabel independent

β_0 = Intersep (titik potong)

β_i = Koefisien regresi

ϵ_i = Komponen error atau residual

Jika regresi linier berganda (lebih dari satu variabel X), maka:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \epsilon \quad [2]$$

2.3. Ridge Regression (L2 Regularization)

Ridge Regression adalah metode regresi yang digunakan untuk mengatasi multikolinieritas, yaitu kondisi ketika variabel independen saling berkorelasi tinggi. Metode ini bekerja dengan menambahkan penalti L2 ke dalam fungsi biaya, yang berfungsi untuk mengecilkan nilai koefisien regresi tanpa menghilangkan variabel [1] [17]. Adapun penduga koefisien Ridge Regression:

$$\hat{\beta}_{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad [3]$$

Fungsi $\underset{\beta}{\operatorname{argmin}}$ pada Ridge Regression sebagai langkah optimasi untuk mencari nilai koefisien β yang meminimalkan fungsi biaya (*loss function*), yang terdiri dari kesalahan prediksi dan penalti L2 ($\sum_{j=1}^p \beta_j^2$). Di mana $\lambda \geq 0$ adalah parameter yang mengontrol kekuatan regularisasi.

2.4. Lasso Regression (L1 Regularization)

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adalah metode regresi yang digunakan untuk mengatasi multikolinieritas dan melakukan seleksi variabel secara otomatis. Teknik ini menambahkan penalti L1 pada fungsi biaya, yang mendorong sebagian koefisien menjadi nol, sehingga hanya variabel penting yang dipertahankan dalam model [1] [18]. Adapun penduga koefisien Lasso Regression:

$$\hat{\beta} (lasso) = \underset{\beta}{\operatorname{argmin}} \|y - \sum_{j=1}^p x_j \beta_j\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad [4]$$

Fungsi $\underset{\beta}{\operatorname{argmin}}$ pada Lasso Regression sebagai langkah optimasi untuk mencari nilai koefisien β yang meminimalkan fungsi biaya (*loss function*), yang terdiri dari kesalahan prediksi dan penalti.

2.5. Elastic Net Regression

Elastic Net Regression adalah metode regresi yang menggabungkan penalti L1 (Lasso) dan L2 (Ridge) untuk mengatasi permasalahan

multikolinearitas dan seleksi variabel secara bersamaan. Elastic Net dirancang untuk mengatasi keterbatasan Ridge dan Lasso Regression. Dengan menggabungkan keunggulan keduanya, Elastic Net mampu menangani multikolinearitas dan menyeleksi variabel penting pada data dengan banyak prediktor. Penalti L1 membuat beberapa koefisien menjadi nol, sedangkan penalti L2 menjaga kestabilan model pada prediktor berkorelasi tinggi. Parameter Elastic Net berada pada rentang 0 hingga 1, sehingga lebih fleksibel dibanding penalti Ridge (0) atau Lasso (1) [19]. Adapun penduga koefisien Elastic Net:

$$\hat{\beta}_{\text{Elastic Net}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \left[(1 - \alpha) \sum_{j=1}^p \beta_j^2 + \alpha \sum_{j=1}^p |\beta_j| \right] \right\} \quad [5]$$

Fungsi $\underset{\beta}{\operatorname{argmin}}$ pada *Elastic Net Regression* sebagai langkah optimasi untuk mencari nilai koefisien β yang meminimalkan fungsi biaya (*loss function*), yang terdiri dari kesalahan prediksi dan penalti kombinasi L1 (Lasso) dan L2 (Ridge). Parameter λ mengontrol kekuatan regularisasi secara keseluruhan, sementara parameter α (bernilai antara 0 dan 1) mengontrol keseimbangan antara penalti L1 dan L2.

2.6. Optimasi Hyperparameter dengan Optuna

Optuna adalah framework open-source untuk optimasi hyperparameter secara otomatis dan efisien. Optuna digunakan untuk mencari kombinasi hyperparameter terbaik pada model Elastic Net Regression dengan mengimplementasikan algoritma pencarian berbasis Bayesian, yaitu Tree-structured Parzen Estimator (TPE), yang mampu melakukan pencarian secara adaptif berdasarkan hasil trial sebelumnya [10]. Ruang pencarian hyperparameter mencakup parameter seperti α , l1_ratio , max_iter , tol , selection , dan fit_intercept . Selama proses optimasi, Optuna menjalankan serangkaian trial untuk mengevaluasi berbagai kombinasi parameter. Berdasarkan hasil dari setiap trial, algoritma akan memperbarui area pencarian dan memfokuskan eksplorasi pada konfigurasi yang paling menjanjikan. Pendekatan ini membuat pencarian lebih efisien dan akurat dibandingkan metode konvensional seperti grid search atau random search, serta dapat meningkatkan performa model secara keseluruhan [20].

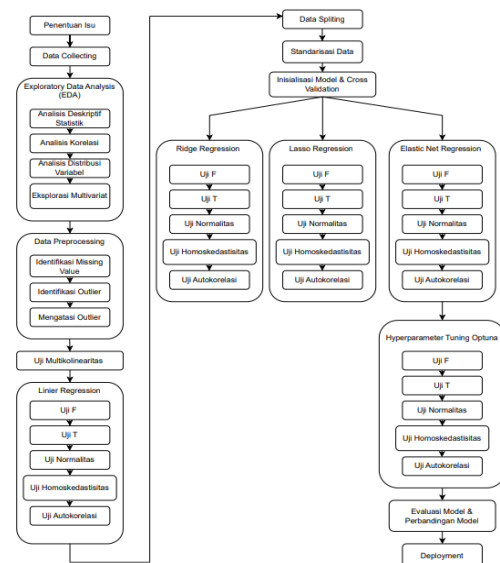
3. METODE PENELITIAN

Gambar 1 menunjukkan alur proses analisis dalam penelitian ini, yang dijelaskan secara rinci sebagai berikut

3.1. Data Collecting

Proses ini bertujuan mengumpulkan data mengenai faktor-faktor yang memengaruhi kemiskinan di kabupaten dan kota di Indonesia. Data

sekunder diperoleh dari Badan Pusat Statistik (BPS) tahun 2023, mencakup persentase penduduk miskin serta indikator sosial ekonomi yaitu Angka Harapan Hidup (AHH), Angka Melekat Huruf (AMH), Produk Domestik Regional Bruto (PDRB), Indeks Pembangunan Manusia (IPM), Partisipasi Angkatan Kerja (PAK), Rata-Rata Lama Sekolah (RLS), Indeks Kemahalan Konstruksi (IKK), dan Pernikahan Dini. Setiap variabel dipilih untuk merepresentasikan dimensi kesehatan, pendidikan, ekonomi, dan kerentanan sosial, guna menunjang analisis yang akurat dan menyeluruh.



Gambar 1 Diagram Alir

3.2. Exploratory Data Analysis (EDA)

Tahap eksplorasi data (EDA) dilakukan untuk memahami karakteristik awal dataset. Proses ini mencakup analisis deskriptif, korelasi, distribusi variabel, dan eksplorasi multivariat guna melihat pola, hubungan, dan interaksi antar variabel sebelum tahap pemodelan.

3.3. Data Preprocessing

Tahap preprocessing dilakukan untuk menyiapkan data agar lebih siap dianalisis, dengan mengidentifikasi dan menangani missing value serta outlier guna menjaga kualitas data dan mendukung hasil pemodelan yang akurat.

3.4. Uji Multikolinearitas

Uji multikolinearitas dilakukan dengan menggunakan Variance Inflation Factor (VIF) guna menilai tingkat hubungan antar variabel independen. Apabila VIF melebihi nilai 10, hal ini menunjukkan adanya multikolinearitas tinggi yang berpotensi mengganggu kestabilan model dan menghasilkan interpretasi yang kurang akurat. Pengujian ini diterapkan pada seluruh variabel independen dalam model.

3.5. Data Splitting

Pembagian data (data splitting) menggunakan rasio 80:20, di mana 80% data digunakan untuk pelatihan (training) dan 20% sisanya untuk pengujian (testing), guna mengevaluasi kinerja model secara objektif.

3.6. Standarisasi Data

Standarisasi dilakukan untuk menyamakan skala antar variabel numerik, sehingga setiap variabel memiliki kontribusi yang seimbang dalam proses pemodelan dan menghindari bias terhadap variabel dengan skala yang lebih besar.

3.7. Inisialisasi Model & Cross Validation

Tahap ini dimulai dengan membangun model Linear Regression menggunakan seluruh data sebagai baseline awal. Selanjutnya, tiga model regularisasi yaitu Ridge, Lasso, dan Elastic Net diinisialisasi menggunakan data yang telah dibagi dengan rasio 80 banding 20 untuk pelatihan dan pengujian. Proses cross validation diterapkan pada ketiga model tersebut untuk mengevaluasi kinerja model secara lebih stabil dan mengurangi risiko overfitting, sehingga hasil yang diperoleh lebih andal dan dapat digeneralisasi.

3.8. Tuning Hyperparameter Optuna

Tahap ini memanfaatkan teknik Tree-structured Parzen Estimator (TPE) dari Optuna untuk melakukan pencarian hyperparameter secara adaptif dan efisien, dengan ruang pencarian mencakup parameter seperti alpha, l1_ratio, max_iter, tol, selection, dan fit_intercept. Optimasi dilakukan dengan 3000 trial untuk memastikan eksplorasi yang menyeluruh terhadap ruang parameter. Dengan algoritma TPE, Optuna secara dinamis menyesuaikan arah pencarian berdasarkan hasil trial sebelumnya, sehingga lebih fokus pada area yang menjanjikan dan menghindari eksplorasi di wilayah yang kurang efektif. Selain itu, Optuna menggunakan teknik pruning untuk menghentikan trial yang tidak menjanjikan secara lebih awal, sehingga mempercepat proses pencarian tanpa mengorbankan kualitas hasil optimasi.

3.9. Uji Signifikansi

Uji signifikansi mengevaluasi apakah hasil model dapat digeneralisasikan ke populasi. Uji F digunakan untuk menilai pengaruh seluruh variabel independen secara bersama-sama, sedangkan uji t mengukur pengaruh masing-masing variabel secara individu. Hasil dianggap signifikan jika nilai $p < 0,05$.

3.10. Uji Asumsi

Uji asumsi bertujuan memastikan model regresi memenuhi syarat validitas. Uji normalitas residual dilakukan dengan Shapiro-Wilk, homoskedastisitas diuji menggunakan Breusch-Pagan, dan autokorelasi diuji dengan statistik Durbin-Watson. Ketiga uji ini memastikan hasil estimasi regresi dapat diinterpretasikan secara tepat.

3.11. Evaluasi Model

Tahap evaluasi model bertujuan untuk menilai keakuratan prediksi model terhadap data. Evaluasi dilakukan menggunakan tiga metrik, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), dan R-squared (R^2). MAE dan MSE digunakan untuk mengukur besar kesalahan prediksi, sedangkan R^2 menunjukkan seberapa baik model menjelaskan varians data. Hasil dari evaluasi ini digunakan untuk membandingkan kinerja seluruh model yang digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Linier Regression

Tahap Linear Regression digunakan sebagai model awal (baseline) untuk mengevaluasi pengaruh variabel independen terhadap tingkat kemiskinan sebelum diterapkannya model regularisasi. Model ini disusun menggunakan pendekatan Ordinary Least Squares (OLS), yang bertujuan meminimalkan jumlah kuadrat selisih antara nilai aktual dan nilai prediksi. Berikut persamaan linear yang diperoleh:

$$\hat{Y} = 99.9299 - 0.2317X_1 - 0.1663X_2 - 1.2620 \times 10^{-9}X_3 - 0.3615X_4 + 0.0330X_5 + 1.0587X_6 + 0.0165X_7 - 11.9678X_8$$

4.2. Ridge Regression

Tahap Ridge Regression dilakukan dengan menentukan nilai parameter penalti (α) melalui 10-fold cross-validation menggunakan fungsi RidgeCV, dengan rentang α antara 10^{-6} hingga 10^6 dalam skala logaritmik. Hasil validasi silang menghasilkan nilai α optimal sebesar 14,1747. Persamaan regresi Ridge yang diperoleh dituliskan sebagai berikut:

$$\hat{Y} = 11.1399 - 1.3877X_1 - 1.6464X_2 + 0.0113X_3 - 3.3701X_4 + 0.5748X_5 + 0.9561X_6 + 0.4302X_7 - 0.9342X_8$$

4.3. Lasso Regression

Tahap Lasso Regression dilakukan dengan menentukan parameter penalti (α) melalui 10-fold cross-validation menggunakan fungsi LassoCV. Proses ini menghasilkan nilai α optimal sebesar 0,0624. Persamaan regresi Lasso yang diperoleh dituliskan sebagai berikut:

$$\hat{Y} = 11.1399 - 1.2680X_1 - 1.6681X_2 + 0.0000X_3 - 3.6600X_4 + 0.5001X_5 + 1.0791X_6 + 0.3051X_7 - 0.9440X_8$$

Koefisien untuk variabel X_3 memiliki nilai nol, yang menegaskan bahwa variabel tersebut dieliminasi oleh penalti L1 Lasso, sehingga menunjukkan bahwa efek seleksi variabel berjalan pada model ini.

4.4. Elastic Net Regression

Tahap Elastic Net Regression dilakukan dengan menggabungkan penalti L1 (seperti pada Lasso) dan

L2 (seperti pada Ridge) secara bersamaan. Penentuan parameter α dan $l1_ratio$ dilakukan melalui 10-fold cross-validation menggunakan fungsi ElasticNetCV. Hasil validasi menghasilkan nilai α optimal sebesar 0,0440 dan $l1_ratio$ sebesar 0,1. Persamaan regresi Elastic Net yang diperoleh dituliskan sebagai berikut:

$$\hat{Y} = 11.1399 - 1.3946X_1 - 1.6256X_2 - 0.0010X_3 - 3.2680X_4 + 0.5666X_5 + 0.8674X_6 + 0.4580X_7 - 0.9250X_8$$

4.5. Elastic Net dengan Optimasi Optuna

Tahap ini mengoptimasi model Elastic Net Regression menggunakan algoritma Optuna selama 3.000 iterasi dengan validasi silang 10-fold. Proses pencarian dilakukan secara adaptif menggunakan algoritma Tree-structured Parzen Estimator (TPE) dan Hyperband Pruner untuk memangkas *trial* dengan performa kurang optimal. Ruang pencarian mencakup parameter α (α), $l1_ratio$, max_iter , tol , $selection$, dan $fit_intercept$. Parameter optimal yang diperoleh adalah:

- α (α): 0.04205
- $l1_ratio$: 0.9992
- max_iter : 100.000
- tol : 0.000327
- $selection$: random
- $fit_intercept$: True

Hasil ini menunjukkan bahwa model Elastic Net yang dioptimasi mampu menangani multikolinearitas,

menjaga akurasi, serta layak digunakan sebagai model akhir dalam analisis faktor-faktor yang memengaruhi Tingkat Kemiskinan.

4.6. Uji Signifikansi Seluruh Model

Untuk menguji signifikansi antar model dan masing-masing variabel, peneliti melakukan uji F dan uji T. Hasil uji F disajikan pada Tabel 1:

Tabel 1. Hasil uji F

Model	F-statistic	P-value
Linier Regression	102.2616	0.0
Ridge Regression	36.6544	0.0
Lasso Regression	36.6544	0.0
Elastic Net Regression	36.6544	0.0
Elastic Net Optuna	36.6544	0.0

Hasil uji F menunjukkan bahwa seluruh model memiliki p-value sebesar 0,0, yang jauh di bawah tingkat signifikansi 0,05. Hal ini menandakan bahwa semua model signifikan secara statistik dalam menjelaskan variasi tingkat kemiskinan. Model Linear Regression memiliki nilai F-statistic tertinggi (102,2616), mengindikasikan kemampuan penjelasan yang baik, meskipun performa prediktifnya lebih rendah dibandingkan model regularisasi. Sementara itu, model Ridge, Lasso, Elastic Net, dan Elastic Net Optuna menunjukkan nilai F-statistic yang identik (36,6544) dan tetap signifikan. Nilai F yang lebih rendah pada model regularisasi mencerminkan adanya penalti terhadap kompleksitas, namun seluruh model tetap valid dan relevan secara statistik.

Tabel 2. Hasil uji T linier regression

Variabel	Koefisien	Std. Error	t-Statistic	P-Value	Signifikan
Konstanta	99.9299	7.1075	14.0598	0.0000	Signifikan
AHH	-0.2317	0.0399	-5.8052	0.0000	Signifikan
AMH	-0.1663	0.0294	-5.6544	0.0000	Signifikan
PDRB	-0.0000	0.0000	-0.1855	0.8529	Non- Signifikan
IPM	-0.3615	0.0419	-8.6358	0.0000	Signifikan
PAK	0.0330	0.0181	1.8186	0.0696	Non-Signifikan
RLS	1.0587	0.2906	3.6433	0.0003	Signifikan
IKK	0.0165	0.0186	0.8874	0.3753	Non-Signifikan
Pernikahan Dini	-11.9678	2.5472	-4.6984	0.0000	Signifikan

Hasil uji t pada model Linear Regression menunjukkan bahwa AHH, AMH, IPM, RLS, dan Pernikahan Dini berpengaruh signifikan terhadap tingkat kemiskinan (p-value < 0,05). Pernikahan Dini

memiliki pengaruh negatif terbesar, diikuti IPM, AHH, dan AMH. Sementara itu, RLS berpengaruh positif signifikan, sedangkan PDRB, PAK, dan IKK tidak signifikan (p-value > 0,05).

Tabel 3. Hasil uji T ridge regression dan elastic net regression

Variabel	Koefisien	Std. Error	t-Statistic	P-Value	Signifikan
Konstanta	11.7884	0.3960	29.7662	0.0000	Signifikan
AHH	-1.8292	0.5171	-3.5372	0.0006	Signifikan
AMH	-1.7737	0.7169	-2.4742	0.0151	Signifikan
PDRB	-0.6884	0.4177	-1.6482	0.1027	Non-Signifikan
IPM	-3.9516	1.0653	-3.7093	0.0004	Signifikan
PAK	-0.3987	0.4849	-0.8223	0.4130	Non-Signifikan
RLS	1.5069	0.9784	1.5401	0.1269	Non-Signifikan
IKK	0.1707	0.5936	0.2876	0.7743	Non-Signifikan
Pernikahan Dini	-1.3767	0.5196	-2.6493	0.0095	Signifikan

Tabel 4. Hasil uji T lasso regression dan elastic net optimasi optuna

Variabel	Koefisien	Std. Error	t-Statistic	P-Value	Signifikan
Konstanta	11.1399	0.2121	52.5300	0.0000	Signifikan
AHH	-1.2873	0.2836	-4.5399	0.0000	Signifikan
AMH	-1.7833	0.3333	-5.3511	0.0000	Signifikan
PDRB	0.1389	0.2565	0.5417	0.5883	Non-Signifikan
IPM	-4.1618	0.5372	-7.7477	0.0000	Signifikan
PAK	0.6053	0.2406	2.5160	0.0123	Signifikan
RLS	1.5896	0.4728	3.3618	0.0008	Signifikan
IKK	0.1826	0.3024	0.6038	0.5463	Non-Signifikan
Pernikahan Dini	-0.9708	0.2644	-3.6719	0.0003	Signifikan

Hasil uji t pada model Ridge dan Elastic Net menunjukkan empat variabel signifikan terhadap kemiskinan, yaitu AHH, AMH, IPM, dan Pernikahan Dini. Sementara itu, model Lasso dan Elastic Net Optuna menunjukkan enam variabel signifikan, dengan IPM berpengaruh negatif paling besar, disusul AHH, AMH, dan Pernikahan Dini. PAK dan RLS berpengaruh positif. Elastic Net Optuna dinilai paling representatif karena mampu menyeimbangkan akurasi prediksi dan seleksi variabel secara efektif.

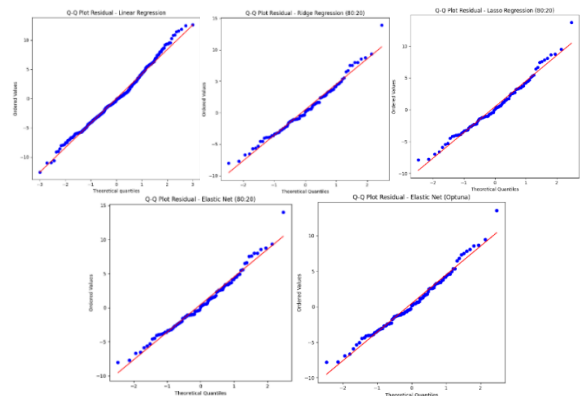
4.7. Uji Asumsi

Untuk memvalidasi seluruh model, dilakukan uji asumsi klasik yang meliputi uji normalitas, homoskedastisitas, dan autokorelasi. Hasil uji asumsi klasik ditampilkan pada Tabel 5 berikut:

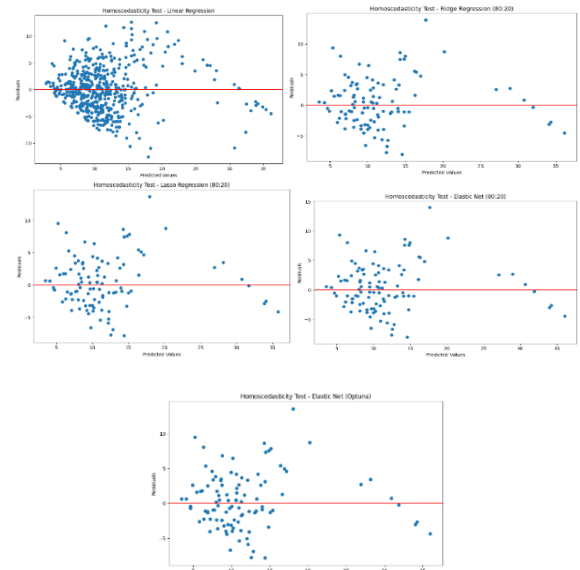
Tabel 5. Uji asumsi seluruh model

Model	Normalitas (Shapiro p-value)	Homoskedastisitas (BP p-value)	Autokorelasi (DW statistic)
Linier Regression	0.0041	0.0000	1.0119
Ridge Regression	0.1659	0.2131	2.4564
Lasso Regression	0.1728	0.2345	2.4713
Elastic Net Regression	0.1586	0.2156	2.4580
Elastic Net Optuna	0.1837	0.2267	2.4663

Berdasarkan uji asumsi, model Linear Regression tidak memenuhi tiga syarat utama, yaitu normalitas (p-value Shapiro-Wilk = 0.0041), homoskedastisitas (p-value Breusch-Pagan = 0.0000), dan bebas autokorelasi (Durbin-Watson = 1.0119), sehingga model kurang stabil. Oleh karena itu, digunakan pendekatan regularisasi seperti Lasso, Ridge, dan Elastic Net. Dari ketiganya, Elastic Net yang dioptimasi dengan Optuna menunjukkan performa terbaik, dengan seluruh uji asumsi terpenuhi (p-value > 0.05, DW mendekati 2), menandakan residual yang stabil dan model yang lebih relevan serta layak digunakan dalam analisis. Diperkuat dengan visualisasi plot dibawah ini:



Gambar 2. Visualisasi Q-Q plot normalitas residual



Gambar 3. Visualisasi uji homoskedastisitas

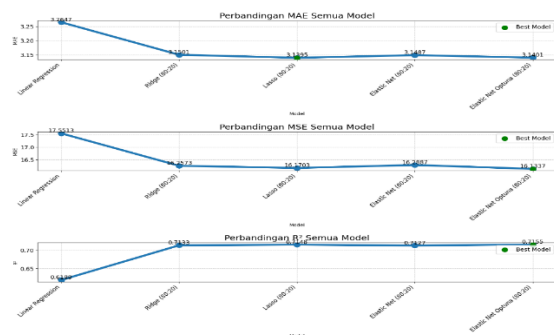
4.8. Evaluasi Model

Penelitian ini menganalisis faktor-faktor yang memengaruhi kemiskinan di Indonesia dengan membandingkan lima model regresi: Linear Regression, Ridge Regression, Lasso Regression, Elastic Net Regression, dan Elastic Net yang dioptimasi dengan Optuna. Fokus utama penelitian adalah pada model Elastic Net Optuna. Evaluasi kelima model ditampilkan pada Tabel 3.2.

Tabel 6. Evaluasi Model

Model	MAE	MSE	R2
Linier Regression	3.2647	17.5513	0.6188
Ridge Regression	3.1501	16.2573	0.7133
Lasso Regression	3.1395	16.1703	0.7148
Elastic Net Regression	3.1487	16.2887	0.7127
Elastic Net Optuna	3.1405	16.1338	0.7155

Berdasarkan evaluasi performa, Elastic Net yang dioptimasi dengan Optuna menunjukkan hasil terbaik dengan MAE terendah (3.1405), MSE terendah (16.1338), dan R^2 tertinggi (0.7155), menandakan akurasi tinggi dalam memprediksi kemiskinan. Lasso berada di posisi kedua, diikuti Ridge dan Elastic Net tanpa optimasi. Linear Regression memiliki performa terendah dengan MSE tertinggi dan R^2 terendah, menunjukkan keterbatasannya dalam menangani data kompleks. Temuan ini memperkuat bahwa Elastic Net yang dioptimasi dengan Optuna adalah model yang paling tepat digunakan dalam analisis ini karena keunggulannya dalam akurasi prediksi dan ketepatan seleksi variabel. Diperjelas dengan visualisasi linechart dibawah ini:



Gambar 4. Visualisasi evaluasi model

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa model Elastic Net yang dioptimasi dengan Optuna memiliki performa terbaik dalam memprediksi tingkat kemiskinan di Indonesia, dengan R^2 tertinggi (0,7155), MAE terendah (3,1405), dan MSE terendah (16,1338), dibandingkan Linear Regression dan model regularisasi lain. Model ini mengidentifikasi enam variabel signifikan, yaitu Indeks Pembangunan Manusia (IPM) dan Pernikahan Dini dengan pengaruh negatif kuat, serta Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-Rata Lama Sekolah (RLS), dan Partisipasi Angkatan Kerja (PAK) dengan pengaruh positif. Selain itu, model memenuhi semua prasyarat regresi klasik sehingga memiliki hasil valid untuk diinterpretasikan. Penelitian ini berkontribusi dengan mengombinasikan metode regularisasi Elastic Net dan optimasi hyperparameter Optuna untuk menghasilkan model kemiskinan yang lebih akurat dibanding regresi tradisional. Temuan ini diharapkan dapat menjadi dasar kebijakan pengentasan kemiskinan yang fokus pada peningkatan kualitas pembangunan manusia, pengurangan pernikahan dini, serta peningkatan pendidikan dan kesempatan kerja.

Namun, penelitian terbatas pada data satu tahun dan indikator sosial ekonomi tertentu sehingga belum mencakup dinamika waktu dan faktor spasial. Oleh karena itu, disarankan penelitian selanjutnya menggunakan data panel atau time series, menambah variabel seperti akses infrastruktur, distribusi bantuan sosial, dan faktor geografis. Penerapan metode machine learning lain untuk memperdalam pemahaman dan menangani kompleksitas kemiskinan.

DAFTAR PUSTAKA

- [1] F. Mudia Sari, K. Anwar Notodiputro, dan B. Sartono, "Analisis Tingkat Kemiskinan di Provinsi Sumatera Barat Melalui Pendekatan Regresi Terkendala (Ridge Regression, Lasso, dan Elastic Net)," 2021.
- [2] Trimono, Arindra Harris A, M Risqi, "Analisis Perubahan Tingkat Pengangguran di Kabupaten/Kota Kalimantan Barat," in *Prosiding Seminar Nasional Sains Data*, Surabaya, 2023.
- [3] K. Maulida Hindrayani, P. Aji Riyantoko, T. Maulana Fahrudin, S. Data UPN, dan J. Timur, "Penerapan Dashboard System Pendapatan Asli Daerah (PAD) Sektor Pariwisata Pada Kabupaten Mojokerto Menggunakan Tableau," *Journal of Computer Science and Technology JCS-TECH*, vol. 4, no. 1, hlm. 14–18, 2024.
- [4] B. P. Statistik, "Persentase Penduduk Miskin Maret 2022 Turun Menjadi 9,54 Persen," bps, 15 juli 2022. [Online]. Available: <https://www.bps.go.id/id/pressrelease/2022/07/15/1930/persentase-penduduk-miskin-maret-2022-turun-menjadi-9-54-persen.html>. [Accessed 14 mei 2025].
- [5] B. P. Statistik, "Profil Kemiskinan di Indonesia Maret 2023," bps, 17 juli 2023. [Online]. Available: <https://www.bps.go.id/id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html>. [Accessed 14 mei 2025].
- [6] Nurfika dan J.-C. Maswana, "The Role of the Secondary Sector in Poverty Alleviation in Indonesia," *The Journal of Indonesia Sustainable Development Planning*, vol. 2, no. 2, hlm. 111–128, Agu 2021, doi: 10.46456/jisdep.v2i2.113.
- [7] K. M. Hindrayani, A. Anjani, dan A. L. Nurlaili, "Penerapan Machine Learning pada Penjualan Produk UMKM: Studi Literatur," *Seminar Nasional Sains Data*, vol. 2021.
- [8] I. Y. W. Guntur Wahyu Kusuma, "Analisis Kemiskinan Dan Kerentanan Kemiskinan Dengan Regresi Ridge, Lasso, Dan Elastic-Net Di Provinsi Jawa Tengah Tahun 2017," dalam *Seminar Nasional Official Statistics*, Jakarta Timur: Politeknik Statistika STIS, 2021, hlm. 503–513.
- [9] P. K. Sahu dan T. Fatma, "Optimized Breast Cancer Classification Using PCA-LASSO Feature Selection and Ensemble Learning Strategies with Optuna Optimization," *IEEE*

- Access, vol. 13, hlm. 35645–35661, 2025, doi: 10.1109/ACCESS.2025.3539746.
- [10] S. Jeganathan, A. R. Lakshminarayanan, S. Parthasarathy, A. A. A. Khan, dan K. J. Sathick, "OptCatB: Optuna Hyperparameter Optimization Model to Forecast the Educational Proficiency of Immigrant Students based on Cat Boost Regression," *Journal of Internet Services and Information Security*, vol. 14, no. 2, hlm. 111–132, Mei 2024, doi: 10.58346/JISIS.2024.I2.008.
- [11] Putri Indah Lestari, Bernadette R, Sukanto, "Kemiskinan Ekstrem, Ketimpangan Dan Pertumbuhan Ekonomi Di Indonesia," *Jurnal Ekombis Review*, vol. 11, no. 2, hlm. 1739–1752, 2023, doi: 10.37676/ekombis.v11i12.
- [12] T. Hana Pratiwi dan R. Setya Wijaya, "Pengaruh Tingkat Pendidikan, Produk Domestik Regional Bruto, Tingkat Pengangguran Terbuka dan Upah Minimum Terhadap TPAK di Kota Semarang," *Jambura Economic Education Journal*, vol. 6, no. 2, [Daring]. Tersedia pada: <https://ejurnal.ung.ac.id/index.php/jej/index>,
- [13] F. Agymnastiar dkk., "Analisis Faktor-Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Provinsi Riau Tahun 2021-2024," *Jurnal Penelitian Ilmu-ilmu Sosial*, vol. 2, no. 10, hlm. 3025–6704, 2025, doi: 10.5281/zenodo.15449540.
- [14] B. P. Statistik, "Indeks Kemahalan Konstruksi Provinsi dan Kabupaten/Kota 2024," bps, 1 Oktober 2024. [Online]. Available: <https://www.bps.go.id/id/publication/2024/10/01/a0ce996786e332d81a43d95f/indeks-kemahalan-konstruksi-provinsi-dan-kabupaten-kota-2024.html>. [Accessed 26 Juni 2025].
- [15] B. P. P. Nasional, "Perkawinan Anak Masuk Kategori Darurat, Bappenas Susun Stranas Upaya Pencegahan Bersama," Bappenas, 18 Desember 2021. [Online]. Available: <https://www.bappenas.go.id/id/berita/perkawinan-anak-masuk-kategori-darurat-bappenas-susun-stranas-upaya-pencegahan-bersama>. [Accessed 26 Juni 2025].
- [16] K. M. Hindrayani, T. M. Fahrudin, R. Prismahardi Aji, dan E. M. Safitri, "Indonesian Stock Price Prediction including Covid19 Era Using Decision Tree Regression," dalam *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, Institute of Electrical and Electronics Engineers Inc., Des 2020, hlm. 344–347. doi: 10.1109/ISRITI51436.2020.9315484.
- [17] A. T. Damaliana, K. M. Hindrayani, dan T. M. Fahrudin, "Hybrid Holt Winter-Prophet method to forecast the number of foreign tourist arrivals through Bali's Ngurah Rai Airport", doi: 10.3390/xxxxx.
- [18] A. T. Damaliana, A. Muhaimin, dan D. A. Prasetya, "Forecasting The Occupancy Rate Of Star Hotels In Bali Using The Xgboost And Svr Methods," 2024, doi: 10.14710/JSUNIMUS.
- [19] F. Fitri, M. W. Ketrin, dan M. Almuhayar, "Modeling Total Fertility Rate In Indonesia: A Comparison Of Fourier Series Regression And Elastic Net Regression," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 3, hlm. 2017–2028, Jul 2025, doi: 10.30598/barekengvol19iss3pp2017-2028.
- [20] K. M. Hindrayani, T. Maulana, dan A. R. Prismahardi, "Determining Students Preparation for College Entrance Examinations in Indonesia From Twitter Data Using Exploratory Data Analysis," 2021.