



Feature Selection pada Indikator Indeks Ekonomi Hijau di Indonesia dengan Machine Learning

Fonda Leviany^{1*}, Ika Nur Laily Fitriana², Nurul Nisa'a Amin³

¹Program Studi Sains Data, Fakultas Sains dan Teknologi, Universitas Terbuka

²Program Studi Statistika, Fakultas Sains dan Teknologi, Universitas Terbuka

³Program Studi Agribisnis, Fakultas Sains dan Teknologi, Universitas Terbuka

*Corresponding Author's e-mail: fonda.leviany@ecampus.ut.ac.id

Article History:

Received: September 15, 2025

Revised: September 22, 2025

Accepted: September 26, 2025

Keywords:

Feature Selection, LASSO Regression, Green Economy Index

Abstract: *Green economy policies are crucial for all countries to ensure that economic activities progress while preserving environmental sustainability. The success of such policies is measured by the Green Economy Index, which in 2020 recorded a national score of 59.17 with 15 indicators, while provincial-level indicators are still being developed. This study analyzes 18 provincial indicators to identify the main factors influencing the Green Economy Index using LASSO regression. This method was chosen for its ability to efficiently perform feature selection, address multicollinearity, and reduce overfitting risks. The dataset includes 18 indicators and index values from 34 provinces. The results show that 15 indicators significantly affect the index. The developed model demonstrates good performance with an RMSE of 1.23 for the training set and 2.29 for the testing set. The R^2 values of 95.6% (training) and 85.98% (testing) indicate strong predictive capability. Moreover, surface water quality is identified as the most influential indicator. These findings are expected to support data-driven policymaking in strengthening the green economy at the provincial level.*

Copyright © 2025, The Author (s).

This is an open access article under the CC-BY-SA license



How to cite: Leviany, F., Fitriana, I. N. L., & Amin, N. N. (2025). Feature Selection pada Indikator Indeks Ekonomi Hijau di Indonesia dengan Machine Learning. *SENTRI: Jurnal Riset Ilmiah*, 4(9), 1822–1835. <https://doi.org/10.55681/sentri.v4i9.4615>

PENDAHULUAN

Ekonomi hijau didefinisikan sebagai ekonomi yang menghasilkan peningkatan kesejahteraan manusia dan keadilan sosial, sekaligus mengurangi risiko lingkungan dan kelangkaan ekologi secara signifikan [1]. Menurut Bappenas, ekonomi hijau adalah model pembangunan ekonomi untuk menunjang pembangunan berkelanjutan dengan fokus pada investasi, kapital dan infrastruktur, lapangan kerja dan keterampilan untuk mewujudkan kesejahteraan sosial dan kelestarian lingkungan [2]. Berdasarkan definisi tersebut dan kondisi perekonomian saat ini, kebijakan ekonomi hijau sangat penting untuk diterapkan di semua negara agar aktivitas ekonomi dapat berjalan lancar dengan memperhatikan kelestarian lingkungan. Kebijakan ekonomi hijau yang telah diterapkan di masing-masing negara tersebut perlu diukur keberhasilannya untuk menentukan kebijakan selanjutnya.

Pada tahun 2011, Organisation for Economic Co-operation and Development (OECD) merumuskan *green growth indicators* yang terdiri dari 15 indikator dan dikelompokkan ke dalam 4 faktor yaitu *environmental and resource productivity*, *natural asset base*, *environmental quality of life*, dan *economic opportunity and policy response* [3]. Sementara, berdasarkan UNEP (2012), terdapat 40 indikator yang dikelompokkan ke dalam 3 kelompok meliputi *environmental*, *policy*, dan *well-being and equity*. Menurut *Global Green*

Economy Index, indikator diukur melalui 18 indikator yang dikelompokkan menjadi 4 dimensi yaitu *climate change and social equity*, *sector decarbonization*, *markets and ESG investment*, dan *environmental health* [4]. Sementara itu, di Indonesia, kinerja penerapan kebijakan ekonomi hijau diukur dengan Indeks Ekonomi Hijau Indonesia. Indeks ini diperoleh berdasarkan hasil pengukuran dari 15 indikator-indikator terpilih yang mewakili bidang lingkungan, ekonomi, dan sosial.

Sejak tahun 2011 hingga 2020, Indeks Ekonomi Hijau di Indonesia memiliki tren meningkat positif. Nilai indeks ekonomi hijau Indonesia mencapai 59,17 di tahun 2020. Terdapat 4 indikator yang memiliki indeks lebih dari 75 yaitu tutupan hutan, pengelolaan limbah, produktivitas tenaga kerja industri, dan harapan hidup. [5]. Indeks ekonomi hijau di Indonesia saat ini juga sedang dikembangkan di tingkat provinsi. Hal ini bertujuan agar penerapan ekonomi hijau di Indonesia sebagai salah satu sasaran visi Indonesia Emas 2045 dapat tercapai. Berbeda dengan indikator di tingkat nasional, indikator Indeks Ekonomi Hijau di tingkat provinsi dikembangkan dan disesuaikan lagi dengan menambahkan indikator kualitas air permukaan, kualitas udara, penggunaan indikator tutupan lahan gambut, PDRB per kapita, dan produktivitas sawit. Pengukuran dan data indeks ekonomi di tingkat provinsi diharapkan dapat mengukur perkembangan dan pencapaian ekonomi hijau di daerah. Hasil studi awal menggambarkan perkembangan pembangunan ekonomi hijau yang positif di sebagian besar provinsi. Adapun sektor prioritas yang diharapkan dapat ditingkatkan antara lain sektor energi, pertanian, pelestarian lingkungan investasi, dan penanggulangan kemiskinan yang diharapkan dapat memberikan dampak bagi pencapaian ekonomi hijau di Indonesia [2]. Namun, jumlah sektor prioritas yang diusulkan masih tergolong banyak. Sehingga, diperlukan studi lanjut dengan pendekatan berbasis data agar indeks ekonomi hijau di Indonesia dapat ditingkatkan secara efektif di tingkat provinsi.

Indikator ekonomi hijau yang telah disusun memberikan informasi mengenai ukuran kuantitatif tingkat pencapaian ekonomi hijau yang telah ditetapkan. Apabila indeks ekonomi hijau di tingkat provinsi berhasil ditingkatkan, diharapkan dapat memberikan pengaruh positif signifikan bagi kenaikan pendapatan negara [6]. Oleh karena itu, diperlukan langkah strategis untuk merumuskan kebijakan yang efektif agar Indeks Ekonomi Hijau dapat meningkat signifikan. Salah satu langkah yang dapat dilakukan oleh pemerintah adalah mengetahui indikator apa yang memberikan dampak signifikan bagi ketercapaian Indeks Ekonomi Hijau. Belum terdapatnya penelitian kuantitatif yang berfokus untuk memilih faktor utama yang mempengaruhi Indeks Ekonomi Hijau di Indonesia, membuka peluang untuk melakukan penelitian ini dengan menyeleksi variabel penting dengan pendekatan *machine learning*.

Feature selection merupakan suatu proses untuk memilih sejumlah variabel prediktor untuk menggambarkan variabel respons. Teknik ini dapat diterapkan untuk mengekstraksi informasi yang bermanfaat khususnya untuk mengetahui faktor utama yang mempengaruhi Indeks Ekonomi Hijau di tingkat provinsi dengan menghilangkan variabel yang dianggap tidak relevan. Hasil dari seleksi variabel merupakan variabel yang dianggap penting dan memiliki pengaruh signifikan terhadap indeks ekonomi hijau, sehingga dapat meningkatkan akurasi pemodelan, mempersingkat waktu pemodelan, dan menyederhanakan hasil pemodelan. Menurut Fonti & Belitser (2017) beberapa teknik *machine learning* yang dapat digunakan untuk melakukan pemilihan variabel penting adalah metode *filter* (Chi-square, ANOVA, korelasi Pearson), metode *wrapper* (*forward* dan *backward selection*), dan metode *embedded* (regresi LASSO) [7].

Pada penelitian ini, regresi LASSO dipilih karena memiliki kelebihan yaitu mampu melakukan seleksi variabel secara efisien serta menangani masalah *overfitting* yang umum terjadi pada model regresi dengan jumlah variabel prediktor yang banyak. Selain itu, LASSO dapat menangani kasus multikolinearitas antar variabel dengan baik. Regresi LASSO juga dinilai mampu menyeleksi variabel secara efektif dengan menyusutkan koefisien variabel yang kurang penting menjadi nol, sehingga variabel-variabel yang benar-benar berkontribusi terhadap indeks ekonomi hijau yang akan dipertahankan serta dapat mengurutkan variabel berdasarkan tingkat kepentingannya sehingga dapat memberikan gambaran yang jelas tentang variabel mana yang paling berpengaruh terhadap indeks ekonomi hijau.

Metode regresi LASSO pernah digunakan pada penelitian sebelumnya mengenai penentuan variabel prediktor untuk meramalkan harga minyak mentah oleh Zhang dkk (2019), memberikan hasil bahwa regresi LASSO memiliki performa yang sangat baik untuk menentukan variabel yang paling berpengaruh dalam meramalkan harga minyak mentah dunia dengan data *out-of-sample*. Penelitian dengan regresi LASSO juga pernah dilakukan oleh Siregar dkk (2024) di bidang kesehatan, dimana regresi LASSO mampu memberikan informasi mengenai variabel yang berpengaruh signifikan untuk mengetahui faktor penyebab utama kematian pasien komorbid yang terinfeksi Covid-19 [8, 9].

Berdasarkan uraian sebelumnya, diketahui bahwa regresi LASSO memiliki kemampuan untuk menyeleksi variabel yang dianggap penting dan menampilkan urutan tingkat kepentingan yang paling berpengaruh terhadap indikator ekonomi hijau di Indonesia. Pada penelitian ini, 18 indikator berperan sebagai variabel independen yang akan diolah dengan pendekatan *machine learning* untuk memperoleh informasi variabel mana yang paling berpengaruh signifikan terhadap indeks ekonomi hijau di tingkat provinsi. Selain itu, akan ditampilkan urutan indikator berdasarkan tingkatan kepentingannya. Sehingga, diharapkan hasil penelitian ini dapat memberikan masukan bagi pengambil kebijakan khususnya pemerintah untuk menyiapkan program khusus agar berfokus pada peningkatan indikator yang berpengaruh signifikan terhadap indikator ekonomi hijau di tingkat provinsi.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan metode kuantitatif dengan menggunakan data sekunder dan metode analisis Regresi Lasso. Penjelasan sumber data, variabel penelitian, dan langkah analisis diuraikan sebagai berikut.

Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari diseminasi oleh Bappenas tahun 2023 yang berjudul “Penyusunan Indeks Ekonomi Hijau di 34 Provinsi”. Data yang digunakan terdiri dari 18 indikator indeks ekonomi hijau dan 1 nilai indeks ekonomi hijau di 34 provinsi yang belum di standardisasi. Indikator-indikator tersebut mewakili 3 pilar yaitu lingkungan (6 indikator), ekonomi (8 indikator), dan sosial (4 indikator).

Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian merupakan 18 indikator ekonomi hijau di tingkat provinsi sebagai variabel independen dan 1 nilai indeks ekonomi hijau di tingkat provinsi dengan rincian sebagai berikut.

Tabel 1. Variabel Penelitian

Variabel	Satuan	Definisi Operasional
Y - Indeks ekonomi hijau	%	Indeks yang menggambarkan hasil evaluasi efektivitas transformasi ekonomi Indonesia menuju ekonomi hijau
X1 - Persentase Luas Tutupan Hutan	%	Perbandingan antara luas tutupan hutan primer, sekunder, dan hutan tanaman terhadap total luas daratan
X2 - Bauran Energi Baru Terbarukan	mg/1	Perbandingan pasokan energi primer (energi yang diberikan oleh alam dan belum mengalami proses pengolahan lebih lanjut) yang berasal dari EBT terhadap total keseluruhan pasokan energi primer
X3 - Kualitas Air Permukaan	%	Nilai konsentrasi BOD (<i>biological oxygen demand</i>) pada air permukaan sebagai salah satu parameter kualitas air
X4 - Persentase lahan gambut	%	Perbandingan tutupan lahan gambut yang bukan jenis hutan (belukar, rawa, perkebunan, permukiman dsb) terhadap total keseluruhan area lahan gambut.
X5 - Kualitas Udara	$\mu\text{g}/\text{m}^3$	Nilai konsentrasi parameter NO ₂ pada udara ambien sebagai salah satu parameter kualitas udara
X6 - Persentase penurunan emisi kumulatif	%	Total penurunan emisi GRK sejak tahun 2010 yang dihasilkan dari aksi PRK dibandingkan terhadap <i>baseline</i>
X7 - Intensitas emisi	Ton CO ₂ e/2010 juta rupiah	Perbandingan antara emisi gas rumah kaca yang terlepas ke atmosfer terhadap produk domestik bruto yang dihasilkan pada tahun tersebut.
X8 - Intensitas energi final	BOE/2010 juta rupiah	Total konsumsi energi final yang digunakan per unit produk domestik bruto yang dihasilkan.
X9 - Pendapatan Domestik Bruto per Kapita	USD/kapita	Perbandingan antara Produk Domestik Regional Bruto suatu daerah yang dibagi dengan total jumlah penduduk yang tinggal (tanpa sektor pertambangan).
X10 - Indeks Produktivitas Pertanian Padi	Ton/hektar/tahun	Jumlah produksi padi atau hasil yang diperoleh dari usahatani padi per satuan luas panen dalam periode satu tahun.
X11 - Indeks Produktivitas Pertanian Kelapa Sawit	Ton/hektar/tahun	Jumlah produksi kelapa sawit atau hasil yang diperoleh dari usahatani kelapa

		sawit per satuan luas panen dalam periode satu tahun.
X12 - Indeks Produktivitas Perikanan Budidaya	Ton/hektar/tahun	Jumlah produksi ikan yang dibudidayakan per satuan luas dalam periode satu tahun
X13 - Produktivitas tenaga kerja sektor industri	2010 IDR/orang	Perbandingan antara produk domestik bruto untuk sektor industri manufaktur terhadap total jumlah tenaga kerja di sektor tersebut.
X14 - Produktivitas tenaga kerja sektor jasa	2010 IDR/orang	Perbandingan antara produk domestik bruto untuk seluruh sektor jasa terhadap total jumlah tenaga kerja di sektor tersebut.
X15 - Rata rata lama sekolah	Tahun	Perbandingan total lama waktu pendidikan pada penduduk dengan usia 25 tahun ke atas terhadap jumlah penduduk dengan usia 25 tahun ke atas.
X16 - Angka harapan hidup	Tahun	Rata-rata tahun hidup yang akan dijalani oleh seseorang dalam situasi mortalitas yang berlaku di lingkungan masyarakat.
X17 - Tingkat kemiskinan	%	Persentase penduduk dengan tingkat pengeluaran yang berada di bawah garis kemiskinan.
X18 - Tingkat pengangguran	%	Persentase jumlah pengangguran terhadap jumlah angkatan kerja (penduduk usia 15 tahun ke atas yang bekerja atau punya pekerjaan namun sementara tidak bekerja).

Metode Analisis

Statistika deskriptif bertujuan untuk memberikan informasi mengenai gambaran umum dan rangkuman dari suatu data yang dapat disajikan dalam bentuk tabel yang memberikan ukuran pemusatan dan penyebaran data ataupun dalam bentuk grafik [10]. Berikut rumus ukuran pemusatan dan penyebaran data pada statistika deskriptif [11]:

a. Rata-rata

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

b. Standar deviasi:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad (2)$$

dengan x_i adalah data ke- i dan n adalah banyak data.

Selain itu, penelitian ini juga menggunakan metode analisis korelasi pearson untuk mengetahui hubungan linear antar dua variabel dengan nilai korelasi antara -1 hingga 1. Apabila nilai korelasi (r) mendekati 1, maka kedua variabel tersebut memiliki hubungan positif kuat. Namun, apabila nilai korelasi (r) mendekati -1, maka kedua variabel tersebut

memiliki hubungan negatif kuat. Apabila nilai korelasi (r) mendekati 0, maka hubungan antara kedua variabel tersebut lemah. Berikut rumus korelasi Pearson [11]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

Metode analisis dengan *machine learning* yaitu *Least Absolute Shrinkage and Selection Operator (LASSO) Regression* atau regresi LASSO merupakan salah satu teknik dari pemodelan regresi yang dapat digunakan untuk menyeleksi variabel penting dengan pendekatan penyusutan. Pendekatan penyusutan pada regresi LASSO ini bekerja dengan meminimalkan jumlah kuadrat residual menjadi jumlah absolut koefisien (regularisasi L1) hingga mendekati 0 [12]. Pada *cost function* dalam regresi LASSO terdapat tambahan penalti pada besaran absolut dari koefisien regresi (L1), atau dapat ditulis sebagai berikut:

$$\hat{\beta} = \min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right), \quad (4)$$

dengan n adalah banyak data, y_i adalah nilai observasi dari variabel dependen untuk data ke- i , X_i adalah vektor dari nilai observasi untuk variabel independen, β adalah vektor koefisien regresi, λ adalah parameter pealti yang mengontrol tingkat regularisasi. Regresi LASSO memiliki kelebihan di antaranya dapat mengurangi kerumitan model dan dapat menyeleksi variabel [13]. Selain itu dengan melakukan seleksi variabel dan regularisasi secara bersamaan, regresi LASSO mampu meningkatkan akurasi prediksi, mengurangi *overfitting*, dan meningkatkan interpretabilitas model [7, 12, 13].

HASIL DAN PEMBAHASAN

Statistika Deskriptif

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar. (Calisto MT, size 12, Spacing: before 0 pt; after 0 pt, Line spacing: 1)

Sebelum dilakukan analisis lebih lanjut, dilakukan eksplorasi data untuk mengetahui Gambaran umum tentang data. Tabel 2 menampilkan statistika deskriptif dari data

Tabel 2. Statistika Deskriptif

Variabel	Rata-rata	Standar Deviasi	Minimum	Median	Maksimum
Y	52,45	6,02	40,38	51,45	69,34
X1	39,41	23,26	1,37	35,65	93,17
X2	11,73	6,70	1,04	10,83	35,09
X3	4,31	3,47	1,23	2,78	14,57
X4	39,04	41,51	0,00	27,90	100,00
X5	5,46	5,67	0,21	5,50	29,60
X6	22,28	17,25	2,46	17,58	84,66
X7	79,76	143,35	0,36	37,42	727,81
X8	97,46	67,38	28,46	78,33	397,97

Variabel	Rata-rata	Standar Deviasi	Minimum	Median	Maksimum
X9	2.511,35	1.893,82	864,38	2.079,74	11.672,62
X10	4,60	0,82	3,03	4,78	5,88
X11	2,31	1,77	0,00	2,35	5,06
X12	31,59	44,34	1,03	16,10	223,83
X13	182.529.551,43	211.000.481,25	2.775.529,48	120.744.088,65	848.387.552,00
X14	52.794.751,37	44.219.223,11	23.833.451,83	42.521.951,17	290.995.334,80
X15	8,68	0,95	6,76	8,74	11,17
X16	70,20	2,51	65,29	70,11	75,08
X17	10,76	5,40	4,53	8,90	26,86
X18	5,49	1,82	3,01	5,06	9,91

Berdasarkan Tabel 2, secara umum nilai-nilai tiap variabel menunjukkan keragaman (variabilitas) yang relatif tinggi, sebagaimana ditunjukkan oleh perbedaan yang cukup lebar antara nilai minimum dan maksimum maupun standar deviasi beberapa variabel. Variabel Y (Indeks Ekonomi Hijau) memiliki rerata sebesar 52,45 persen dengan simpangan baku 6,02. Rentang nilai minimum hingga maksimum berada antara 40,38 persen dan 69,34 persen. Hal ini menandakan bahwa capaian “kinerja ekonomi hijau” di berbagai provinsi memiliki tingkat heterogenitas moderat. Selanjutnya, variabel-variabel terkait lingkungan menunjukkan karakteristik yang bervariasi. Misalnya, X1 (Persentase Luas Tutupan Hutan) memiliki rata-rata 39,41 persen dengan standar deviasi 23,26, mengindikasikan perbedaan luas tutupan hutan yang cukup tajam antar provinsi. Variabel X2 (Bauran Energi Baru Terbarukan) juga memiliki sebaran yang cukup lebar (Standar Deviasi = 6,70), meskipun rentangnya tidak sebesar X1. X3 (Kualitas Air Permukaan) dan X5 (Kualitas Udara) sama-sama menunjukkan rerata di bawah 6 (masing-masing 4,31 persen dan 5,46 $\mu\text{g}/\text{m}^3$). Keduanya memiliki sebaran yang relatif moderat (standar deviasi berturut-turut 3,47 dan 5,67), menandakan masih adanya disparitas kondisi kualitas lingkungan antarlokasi namun tidak terlalu ekstrem. Sementara itu, X4 (Persentase Lahan Gambut) memunculkan rentang ekstrem (0–100 persen) dengan standar deviasi yang tinggi (41,51), menggambarkan ada kawasan yang sama sekali tidak memiliki lahan gambut dan ada pula yang hampir seluruh areanya berupa lahan gambut.

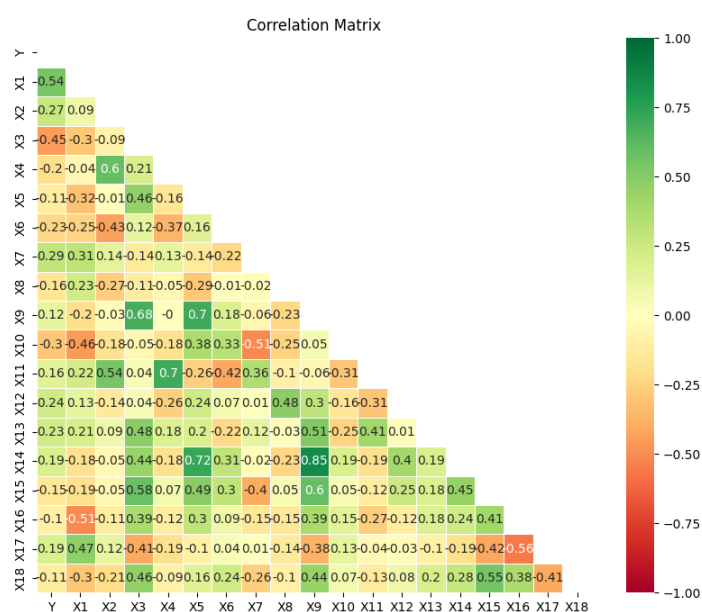
Pada X6 (Persentase Penurunan Emisi Kumulatif), nilai rata-ratanya tercatat 22,28 persen dengan standar deviasi 17,25. Pada data X6 memiliki rentang nilai 2,46–84,66 persen mengindikasikan sejumlah provinsi yang sudah cukup berhasil menurunkan emisi secara signifikan, sedangkan yang lain masih sangat minimal. X7 (Intensitas Emisi) dan X8 (Intensitas Energi Final) juga menunjukkan penyebaran yang lebar. Terutama X7, memiliki rata-rata 79,76 ton CO₂ per 2010 juta rupiah namun standar deviasi mencapai 143,35, menandakan ketimpangan intensitas emisi yang cukup ekstrem; hal ini terlihat dari data asli yang memiliki nilai minimum 0,36 hingga maksimum 727,81.

Pada aspek ekonomi, variabel X9 (PDB per kapita) berkisar antara 864,38–11.672,62 USD/kapita, hal ini menunjukkan adanya kesenjangan ekonomi antar provinsi. Variabel yang berkaitan dengan produktivitas tenaga kerja dari berbagai sektor juga

menunjukkan variasi yang cukup tinggi. Hal ini menggambarkan beragamnya kapasitas ekonomi antar provinsi di Indonesia.

Pada aspek sosial, variabel X15 (Rata-rata Lama Sekolah) terdistribusi cukup merata dengan nilai rata-rata 8,68 tahun dengan rentang pada data 6,76–11,17 tahun, menunjukkan perbedaan capaian pendidikan formal antar provinsi yang tidak terlalu mencolok. X16 (Angka Harapan Hidup) memiliki nilai yang cukup tinggi di mayoritas provinsi (rata-rata 70,20 tahun; minimum 65,29, maksimum 75,08; standar deviasi = 2,51), menandakan mayoritas berada pada tingkat harapan hidup menengah-atas. Variabel X17 (Tingkat Kemiskinan) memiliki rerata 10,76 persen (SD = 5,40), dengan provinsi tertentu masih berjuang pada level kemiskinan yang lebih tinggi (maksimum 26,86 persen). Sementara variabel X18 (Tingkat Pengangguran) cenderung lebih rendah, berkisar antara 3,01–9,91 persen.

Eksplorasi data dilanjutkan dengan melakukan analisis korelasi antar variabel untuk mengetahui hubungan antar variabel. Hasil matriks korelasi antar variabel ditampilkan pada Gambar 1.



Gambar 1. Matriks Korelasi Antar Variabel

Berdasarkan Gambar 1, dapat diketahui bahwa pada aspek lingkungan, variabel Y (Indeks Ekonomi Hijau) memiliki korelasi positif terhadap X1 (Persentase Luas Tutupan Hutan), yakni sebesar 0,54, dan X2 (Bauran Energi Baru Terbarukan), yakni sebesar 0,27. Temuan ini menunjukkan bahwa daerah dengan tutupan hutan yang relatif luas serta penggunaan energi baru terbarukan yang lebih tinggi cenderung memiliki kinerja ekonomi hijau yang lebih baik. Sebaliknya, variabel Y memiliki korelasi negatif dengan variabel X3 (Kualitas Air Permukaan), X4 (Persentase lahan gambut), X5 (Kualitas Udara), dan X6 (Persentase penurunan emisi kumulatif). Korelasi negatif dengan X3 (Kualitas Air Permukaan) dan X5 (Kualitas Udara) dapat terjadi jika wilayah yang kualitas air dan udaranya sudah baik secara alamiah belum sepenuhnya menerapkan praktik ekonomi hijau berbasis kebijakan atau inovasi hijau; sementara korelasi negatif dengan X4 (Persentase lahan gambut) dan X6 (Persentase penurunan emisi kumulatif) dapat disebabkan oleh tantangan tata kelola lahan gambut serta dinamika industri atau sektor

lain yang memangkas emisi secara signifikan, tetapi belum diiringi penguatan ekonomi hijau secara komprehensif. Dengan demikian, perbedaan arah korelasi menunjukkan bahwa kualitas lingkungan yang “baik secara alamiah” tidak selalu sejajar dengan peningkatan Indeks Ekonomi Hijau, khususnya jika belum diintegrasikan dengan kebijakan berorientasi pembangunan berkelanjutan.

Selanjutnya, pada aspek ekonomi, berdasarkan gambar 1 dapat dilihat bahwa Indeks Ekonomi Hijau memiliki korelasi positif dengan hampir seluruh variabel X7-X14, kecuali dengan variabel X8 dan X10 yang bernilai negatif. Temuan ini menunjukkan bahwa provinsi dengan intensitas emisi yang lebih tinggi, kapasitas ekonomi (PDB) yang besar, serta produktivitas di sektor-sektor kunci yang lebih optimal, cenderung memiliki Indeks Ekonomi Hijau yang lebih tinggi. Sebaliknya, X8 (Intensitas Energi Final) dan X10 (Produktivitas Pertanian Padi) justru berkorelasi negatif dengan Indeks Ekonomi Hijau (masing-masing $-0,16$ dan $-0,30$). Korelasi negatif ini menandakan bahwa provinsi dengan intensitas penggunaan energi yang belum efisien, serta didominasi sumber energi fosil, cenderung memiliki indeks ekonomi hijau yang lebih rendah.

Pada aspek sosial, nilai korelasi Indeks Ekonomi Hijau memiliki hubungan negatif dengan X15 (Rata-rata Lama Sekolah), X16 (Angka Harapan Hidup), dan X18 (Tingkat Pengangguran), serta korelasi positif dengan X17 (Tingkat Kemiskinan). Koefisien negatif terhadap X15 dan X16 mengindikasikan bahwa provinsi dengan rata-rata lama sekolah dan harapan hidup yang relatif tinggi tidak selalu diiringi oleh kinerja ekonomi hijau yang baik. Hal serupa terjadi pada X18, di mana daerah dengan pengangguran yang lebih tinggi cenderung memiliki Indeks Ekonomi Hijau yang lebih rendah. Menariknya, korelasi positif Y dengan X17 (Tingkat Kemiskinan) mengisyaratkan bahwa beberapa provinsi yang memiliki tingkat kemiskinan relatif tinggi justru memperlihatkan capaian ekonomi hijau lebih tinggi. Namun demikian, analisis korelasi tidak serta-merta mencerminkan kausalitas, oleh karena itu, dibutuhkan analisis lanjutan menggunakan analisis regresi Lasso untuk menunjukkan hubungan sebab-akibat secara lebih komprehensif.

Analisis Regresi LASSO

Pada Gambar 1, dapat pula dilihat korelasi antar variabel X, yang mana terdapat beberapa variabel X yang memiliki korelasi tinggi (multikolinieritas). Pada penelitian ini, terdapat multikolinieritas (korelasi yang tinggi antara variabel X), sehingga metode Regresi LASSO merupakan metode yang tepat untuk digunakan dalam penelitian ini agar bisa mengatasi multikolinieritas yang tinggi pada data. Dengan menerapkan metode Regresi LASSO, menjadi mungkin untuk mengidentifikasi variabel mana yang secara signifikan mempengaruhi variabel Y dan menentukan sejauh mana pengaruh variabel X terhadap Y.

Regresi LASSO diketahui efektif dalam melakukan pemilihan (seleksi) variabel dengan mengecilkan koefisien variabel yang kurang relevan menjadi nol, sehingga hanya variabel yang secara signifikan berkontribusi terhadap indeks ekonomi hijau yang dipertahankan.

Dalam penelitian ini, analisis regresi LASSO dilaksanakan dengan menerapkan penalti L1 pada total nilai kuadrat koefisien. Penentuan nilai penalti L1 yang optimal dilakukan melalui prosedur validasi silang *k-fold* ($k=5$), dan diperoleh nilai penalti L1 optimal sebesar 0,045. Selanjutnya, pemodelan dilakukan dengan regresi LASSO menggunakan penalti L1 tersebut, sehingga dihasilkan koefisien regresi LASSO yang disajikan pada Tabel 3.

Tabel 3. Koefisien Regresi LASSO

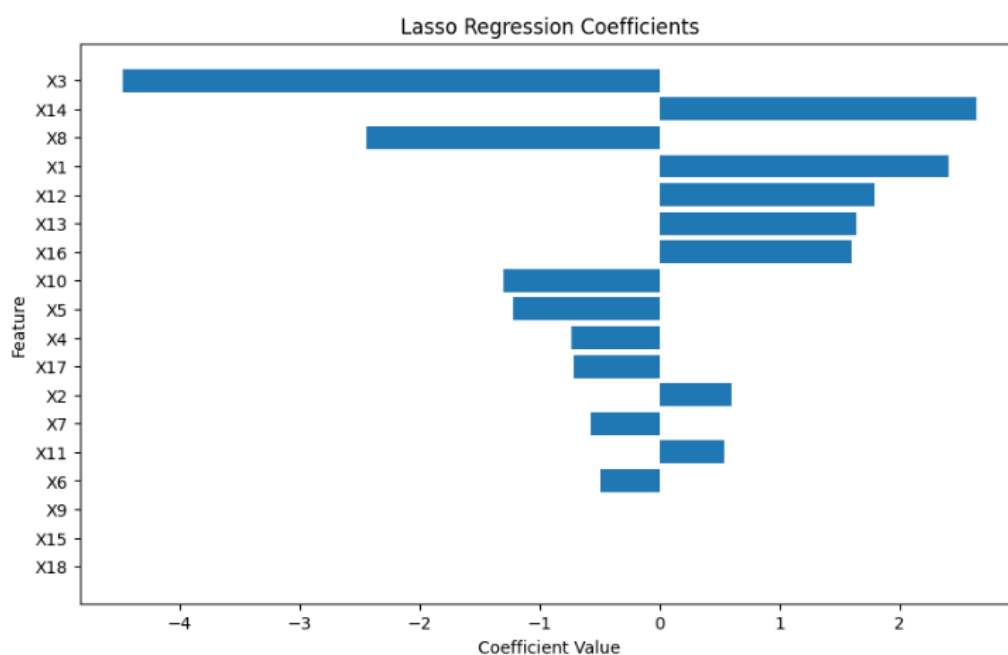
Variabel	Koefisien Regresi Lasso	Signifikansi
X1	2,40	Signifikan
X2	0,60	Signifikan
X3	-4,48	Signifikan
X4	-0,73	Signifikan
X5	-1,22	Signifikan
X6	-0,49	Signifikan
X7	-0,57	Signifikan
X8	-2,44	Signifikan
X9	0,00	Tidak Signifikan
X10	-1,31	Signifikan
X11	0,54	Signifikan
X12	1,79	Signifikan
X13	1,64	Signifikan
X14	2,64	Signifikan
X15	0,00	Tidak Signifikan
X16	1,60	Signifikan
X17	-0,72	Signifikan
X18	0,00	Tidak Signifikan

Model tersebut memiliki performansi yang baik, dengan nilai RMSE pada *training set* sebesar 1,23 dan pada *testing set* sebesar 2,29. Nilai R^2 pada *training set* sebesar 95,6% dan pada *testing set* sebesar 85,98%. Nilai R^2 menunjukkan bahwa model yang dihasilkan memiliki kemampuan prediksi yang baik, dengan proporsi variasi yang dijelaskan oleh model pada *training set* mencapai 95,6% dan pada *testing set* sebesar 85,98%. Perbedaan sebesar sekitar 9,62% antara kedua nilai tersebut menandakan adanya sedikit penurunan kinerja ketika model dihadapkan pada data baru, namun selisih ini masih tergolong kecil sehingga dapat dikatakan bahwa potensi *overfitting* dalam model relatif rendah. Dengan demikian, regresi LASSO yang dibangun dengan penalti L1 optimal sebesar 0,045 memberikan keseimbangan yang baik antara kemampuan menjelaskan data pelatihan dan ketepatan prediksi pada data uji, sekaligus memperkuat efisiensi pemodelan melalui pemangkasan koefisien yang tidak signifikan.

Berdasarkan Tabel 3, terdapat tiga variabel yang memiliki nilai koefisien regresi nol, yaitu variabel X9 (Pendapatan Domestik Bruto per Kapita), X15 (Rata rata lama sekolah), dan X18 (Tingkat pengangguran). Pada regresi LASSO. Koefisien regresi yang bernilai nol mengindikasikan bahwa variabel tersebut tidak signifikan. Hal ini terjadi karena regresi LASSO menerapkan penalti L1 pada koefisien, yang menyebabkan beberapa koefisien menyusut menjadi nol dan secara otomatis melakukan pemilihan (seleksi) fitur dengan menyingkirkan variabel yang tidak relevan. Dengan kata lain, hanya variabel yang memiliki kontribusi signifikan terhadap Indeks Ekonomi Hijau yang dapat mempertahankan koefisien bukan nol.

Selain itu, regresi LASSO juga bisa melakukan pengurutan variabel berdasarkan tingkat kepentingannya, sehingga memberikan pandangan komprehensif mengenai variabel yang paling berpengaruh terhadap indeks ekonomi hijau. Urutan tingkat kepentingan variabel didasarkan pada seberapa tinggi nilai koefisien regresinya, tanpa

memandang tanda positif atau negatif pada koefisien regresi. Visualisasi urutan variabel berdasarkan tingkat kepentingannya ditunjukkan pada Gambar 2.



Gambar 2. Urutan Koefisien Regresi LASSO

Terdapat 15 dari 18 variabel yang dinilai penting dan berkontribusi signifikan terhadap indeks ekonomi hijau. Variabel tersebut menurut tingkat kepentingan dan kontribusinya bagi indeks ekonomi hijau secara berurutan tanpa memperhatikan tanda koefisien positif dan negatif adalah: variabel X3 (kualitas air permukaan), X14 (produktivitas tenaga kerja sektor jasa), X8 (intensitas energi final), X1 (persentase luas tutupan hutan), X12 (Indeks produktivitas perikanan budidaya), X13 (produktivitas tenaga kerja sektor industri), X16 (angka harapan hidup), X10 (indeks produktivitas pertanian padi), X5 (kualitas udara), X4 (persentase lahan gambut), X17 (tingkat kemiskinan), X2 (bauran energi baru terbarukan), X7 (Intensitas emisi), X11 (Indeks produktivitas pertanian kelapa sawit), dan X6 (persentase penurunan emisi kumulatif). Di antara 15 variabel yang signifikan tersebut, 7 variabel di antaranya memiliki nilai koefisien positif, yaitu: X14 (produktivitas tenaga kerja sektor jasa), X1 (persentase luas tutupan hutan), X12 (Indeks produktivitas perikanan budidaya), X13 (produktivitas tenaga kerja sektor industri), X16 (angka harapan hidup), X2 (bauran energi baru terbarukan), dan X11 (Indeks produktivitas pertanian kelapa sawit). Artinya, apabila terdapat kenaikan nilai pada salah satu variabel, maka indeks ekonomi hijau akan meningkat. Selain itu, 8 variabel lainnya yang memiliki nilai koefisien negatif yaitu: variabel X3 (kualitas air permukaan), X8 (intensitas energi final), X10 (indeks produktivitas pertanian padi), X5 (kualitas udara), X4 (persentase lahan gambut), X17 (tingkat kemiskinan), X7 (Intensitas emisi), dan X6 (persentase penurunan emisi kumulatif). Artinya, apabila terdapat kenaikan nilai pada salah satu variabel, maka indeks ekonomi hijau akan menurun. Hampir seluruh variabel memiliki tanda koefisien yang sesuai dengan harapan dari Bappenas, dimana, ada variabel yang diharapkan meningkat nilainya dan juga menurun. Namun, terdapat 2 variabel yang memiliki nilai koefisien belum sesuai harapan, yaitu variabel X10 (indeks produktivitas pertanian padi) dan X6 (persentase penurunan emisi kumulatif). Kedua variabel tersebut

memiliki koefisien negatif, padahal Bappenas berharap bahwa nilai dari indeks produktivitas pertanian padi dan persentase penurunan emisi kumulatif meningkat sejalan dengan indeks ekonomi hijau.

Berdasarkan Gambar 2 juga diperoleh informasi bahwa di antara seluruh variabel atau indikator, kualitas air permukaan memberikan kontribusi paling signifikan terhadap indeks ekonomi hijau di Indonesia untuk tingkat provinsi. Hal ini dapat dikarenakan secara geografis Indonesia merupakan negara yang memiliki banyak sungai dan danau, serta sering dimanfaatkan oleh masyarakat yang bermatapencaharian di sektor pertanian dan perikanan. Sehingga, apabila kualitas air mengalami penurunan, tidak hanya berdampak pada lingkungan, namun juga berdampak ada kondisi ekonomi dan kesehatan masyarakat yang bergantung pada air bersih. Selain itu, kualitas air yang buruk juga berdampak pada meningkatnya biaya pengolahan air untuk memenuhi kebutuhan masyarakat. Hal ini dapat mengurangi potensi ekonomi hijau karena anggaran yang seharusnya digunakan untuk inovasi dan teknologi hijau teralihkan untuk menangani dampak negatif dari buruknya pengelolaan air. Sehingga, dapat disimpulkan bahwa kualitas air permukaan memiliki pengaruh langsung bagi kelestarian lingkungan, ekonomi, dan sosial-kesehatan masyarakat.

Hasil analisis regresi LASSO menunjukkan bahwa beberapa faktor memiliki pengaruh signifikan terhadap indeks ekonomi hijau, di antaranya adalah produktivitas tenaga kerja di sektor jasa, industri, dan perikanan budidaya, produktivitas kelapa sawit, luas hutan, bauran energi baru, serta angka harapan hidup. Sejalan dengan penelitian O'Neill *et al.* (2023) dan Erwiningsih (2023), yang menjelaskan bahwa investasi dalam energi terbarukan, pertanian berkelanjutan, dan teknologi ramah lingkungan tidak hanya berperan dalam mengurangi kerusakan lingkungan, tetapi juga mendukung pertumbuhan ekonomi dan penciptaan lapangan kerja di berbagai sektor, termasuk industri kelapa sawit di Indonesia yang memiliki kontribusi besar terhadap PDB dan memfasilitasi penerapan pertanian yang berkelanjutan [14, 15]

Indikator bauran energi terbarukan berpengaruh terhadap indeks ekonomi hijau di Indonesia dapat terlihat dari penerapan kebijakan ekonomi hijau di Indonesia yang telah berjalan dalam beberapa tahun terakhir dengan berfokus kepada pengelolaan sumber daya yang lebih efisien dan penerapan teknologi ramah lingkungan. Energi terbarukan menjadi salah satu aspek utama dalam ekonomi hijau di Indonesia karena kontribusinya dalam mengurangi emisi karbon dan mendukung pembangunan berkelanjutan. Namun, dalam penerapannya masih diperlukan upaya untuk meningkatkan kesadaran publik tentang manfaat ekonomi hijau, terutama terkait konsumsi produk ramah lingkungan seperti makanan dan energi berkelanjutan [14].

Pada sektor industri, mengutamakan keberlanjutan dalam kebijakan dan praktik industri sangat penting untuk mengurangi dampak negatif perubahan iklim, yang dapat mengancam infrastruktur dan perekonomian. Penerapan ekonomi hijau tidak hanya membantu menjaga kelestarian lingkungan, tetapi juga mendorong inovasi, diversifikasi ekonomi, dan memperkuat ketahanan ekonomi terhadap tantangan eksternal seperti fluktuasi harga minyak atau dampak perubahan iklim.

Penelitian mengenai ekonomi hijau di Indonesia menunjukkan bahwa investasi dalam energi terbarukan dan pengelolaan sumber daya alam yang lebih efisien dapat memberikan dampak positif terhadap pertumbuhan ekonomi. Hal ini termasuk dapat membantu mengatasi tantangan lingkungan global seperti perubahan iklim, serta berkontribusi pada stabilitas ekonomi melalui inovasi hijau dan peningkatan efisiensi

energi [16]. Beberapa sektor, seperti pertanian dan energi, di Indonesia mulai mengadopsi praktik yang lebih ramah lingkungan, misalnya dengan sertifikasi Indonesian Sustainable Palm Oil (ISPO), yang berfungsi untuk mengurangi deforestasi sekaligus mendorong pertumbuhan ekonomi yang berkelanjutan [17].

KESIMPULAN

Berdasarkan hasil seleksi variabel untuk menentukan faktor utama yang mempengaruhi Indeks Ekonomi Hijau tingkat provinsi yang telah dilakukan, diperoleh kesimpulan bahwa dari 18 indikator yang dianalisis, didapatkan 15 variabel indikator berpengaruh signifikan terhadap Indeks Ekonomi Hijau di tingkat provinsi. Hasil ini diperoleh setelah dilakukan seleksi variabel dengan regresi LASSO yang memberikan performansi nilai RMSE pada *training set* sebesar 1,23 dan pada *testing set* sebesar 2,29. Nilai R^2 sebesar 95,6% menunjukkan bahwa model yang dihasilkan memiliki kemampuan prediksi yang baik pada *training set* dan pada *testing set* sebesar 85,98%. Perbedaan sebesar sekitar 9,62% antara kedua nilai tersebut menandakan adanya sedikit penurunan kinerja ketika model dihadapkan pada data baru, namun selisih ini masih tergolong kecil sehingga dapat dikatakan bahwa potensi *overfitting* dalam model relatif rendah. Selain itu, hasil dari seleksi variabel dengan regresi LASSO memberikan informasi bahwa indikator yang dinilai paling berpengaruh signifikan adalah indikator kualitas air permukaan. Di Indonesia, kualitas air permukaan sangat penting mengingat Indonesia memiliki banyak sungai dan danau yang berfungsi sebagai sumber air baku, irigasi, transportasi, dan perikanan. Penurunan kualitas air permukaan di Indonesia memiliki pengaruh langsung bagi kelestarian lingkungan, ekonomi, dan sosial-kesehatan masyarakat.

Pemerintah diharapkan dapat memperkuat kebijakan ekonomi hijau dengan fokus pada faktor-faktor yang berpengaruh secara signifikan. Pemerintah perlu terus memantau dan mengevaluasi indikator ekonomi hijau di tingkat provinsi untuk memastikan kebijakan yang diambil efektif. Selain itu, peningkatan kesadaran masyarakat dan kerjasama dengan sektor swasta juga penting untuk mendukung pencapaian tujuan pembangunan berkelanjutan di Indonesia.

DAFTAR REFERENSI

- [1] UNEP, *Measuring Progress Towards a Green Economy*, Kenya: UNEP, 2012.
- [2] Bappenas, "Penyusunan Indeks Ekonomi Hijau di 34 Provinsi," Bappenas, Jakarta, 2023.
- [3] OECD, "Towards Green Growth: Monitoring Progress: OECD Indicators," OECD Publishing, 2011.
- [4] Dual Citizen LLC, "2022 Global Green Economy Index (GGEI)," Dual Citizen LLC, New York, 2022.
- [5] Bappenas, *Green Economy Index: A Step Forward to Measure the Progress of Low Carbon & Green Economy in Indonesia*, Jakarta: Bappenas, 2022.
- [6] P. C. a. P. L. L. Lumbanraja, "Analisis Variabel Ekonomi Hijau (Green Economy Variable) terhadap Pendapatan Indonesia (Tahun 2011-2020) dengan Metode SEM-PLS," *Cendekia Niaga: Journal of Trade Development and Studies*, vol. 7, no. 1, 2023.
- [7] V. Fonti and E. Belitser, "Paper in Business Analytics Feature Selection using LASSO," 2017.

- [8] A. L. Siregar, R. F. Sari and R. Widyasari, "Penerapan Analisis Lasso dalam Menentukan Penyebab Utama Faktor Kematian Pasien Komorbid yang Terinfeksi Covid-19," *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika*, vol. 7, no. 1, pp. 290-296, 2024.
- [9] Y. F. M. a. Y. W. Zhang, "Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?," *Journal of Empirical Finance*, vol. 54, pp. 97-117, 2019.
- [10] R. Cooksey, "Descriptive Statistics for Summarising Data," in *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, Singapore, Springer, 2020.
- [11] Walpole, Pengantar Statistika Edisi Ketiga, Jakarta: PT. Gramedia Pustaka Utama, 1982.
- [12] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267-288, 1996.
- [13] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, Coimbatore, India, 2016.
- [14] C. M. M. S. O. a. F. & A. O'Neill, "Food interests, preferences and behaviours: a profile of the sustainable food consumer," *Br. Food J*, vol. 13, no. 125, pp. 352-374, 2023.
- [15] W. Erwiningsih, "Analyzing drivers and mitigation of deforestation for oil palm expansion in Indonesia, 2000-2020," *Int. J. Sustain. Dev. Plan*, vol. 8, no. 11, pp. 3657-3664, 2023.
- [16] L. M. X. H. C. Y. L. a. S. & C. Zhang, "Globalization, green economy and environmental challenges: state of the art review for practical implications," *Front. Environ. Sci.*, vol. 10, pp. 1-9.
- [17] N. N. E. D. D. K. a. D. & S. Denashurya, "Overcoming barriers to ISPO certification: analyzing the drivers of sustainable agricultural adoption among farmers," *Sustainability*, vol. 15, no. 23, pp. 1-25, 2023.