

Regresi Logistik Biner dan *Support Vector Machine* dalam Klasifikasi Indeks Pembangunan Manusia

<http://dx.doi.org/10.28932/jutisi.v12i1.11853>

Riwayat Artikel

Received: 20 Mei 2025 | Final Revision: 06 April 2026 | Accepted: 06 April 2026

Creative Commons License 4.0 (CC BY – NC)



Rupmana Br Butar Butar ^{✉#1}, Destriana Aulia Rifaldi^{#2}, Anwar Fitrianto^{#3}, Pika Silvianti^{#4}

[#] Program studi Statistika dan Sains data, IPB University
Dramaga Kab. Bogor, Kota Bogor, 16680, Indonesia

¹rupmanabutar@apps.ipb.ac.id

²13auliarifaldi@apps.ipb.ac.id

³anwarstat@gmail.com

⁴pikasilvianti@apps.ipb.ac.id

[✉]Corresponding author: rupmanabutar@apps.ipb.ac.id

Abstrak — Regresi Logistik Biner dan Support Vector Machine (SVM) merupakan dua metode klasifikasi yang banyak digunakan dalam analisis data, terutama untuk permasalahan dengan variabel target kategorik. Dalam penelitian ini, kedua metode tersebut dibandingkan untuk mengklasifikasikan status Indeks Pembangunan Manusia (IPM) Indonesia tahun 2024. Data awal terdiri dari lima variabel prediktor, namun setelah dilakukan analisis korelasi untuk menghindari multikolinearitas, hanya tiga variabel yang digunakan dalam pemodelan. Teknik Synthetic Minority Over-Sampling Technique (SMOTE) diterapkan untuk menangani ketidakseimbangan kelas. Regresi Logistik Biner dipilih karena interpretabilitasnya yang baik, sedangkan SVM digunakan sebagai pembanding karena kemampuannya yang robust terhadap outliers. Hasil evaluasi menunjukkan bahwa Regresi Logistik Biner menghasilkan akurasi sebesar 87,85%, sedikit lebih tinggi dibandingkan SVM yang mencapai 86,92%. Oleh karena itu, Regresi Logistik Biner dinilai lebih optimal dalam mengklasifikasikan status IPM pada data yang telah diseimbangkan dan disederhanakan. Penelitian ini memberikan kontribusi dalam penerapan metode statistik dan pembelajaran mesin untuk mendukung analisis pembangunan manusia berbasis data.

Kata kunci— Biner; SMOTE; SVM.

Binary Logistic Regression and Support Vector Machine for Classifying Human Development Index

Abstract — Binary Logistic Regression and Support Vector Machine (SVM) are two widely used classification methods in data analysis, especially for problems with categorical target variables. In this study, these two methods are compared to classify the Human Development Index (HDI) status of Indonesia in 2024. The initial data consists of five predictor variables, but after conducting a correlation analysis to avoid multicollinearity, only three variables were used in the modeling. The Synthetic Minority Over-Sampling Technique (SMOTE) was applied to address class imbalance. Binary Logistic Regression was chosen due to its good interpretability, while SVM was used as a comparison due to its robustness against outliers. Evaluation results show that Binary Logistic Regression achieved an accuracy of 87.85%, slightly higher than SVM, which reached 86.92%. Therefore, Binary Logistic Regression is considered

more optimal in classifying HDI status on the data that has been balanced and simplified. This study contributes to the application of statistical methods and machine learning in supporting human development analysis based on data.

Keywords— Binary; SMOTE; SVM.

I. PENDAHULUAN

Dalam konteks pemodelan klasifikasi biner, dua metode populer yang banyak digunakan adalah Regresi Logistik Biner (Reglog Biner) dan *Support Vector Machine* (SVM). Regresi Logistik Biner merupakan salah satu pendekatan model matematis yang dapat dimanfaatkan untuk menganalisis hubungan antara satu atau lebih variabel independen dengan variabel dependen yang bersifat kategorik dikotom [1]. Keunggulan utamanya terletak pada kemampuannya untuk menginterpretasikan hubungan antara variabel prediktor dan dependen melalui *odds ratio* yang intuitif [2]. Di sisi lain, *Support Vector Machine* (SVM) merupakan metode klasifikasi diskriminatif yang bekerja dengan membangun sebuah *hyperplane* untuk memisahkan kelas secara optimal, yaitu dengan memaksimalkan margin antar kelas. SVM termasuk dalam keluarga metode jaringan saraf tiruan (*Artificial Neural Network*) dan mampu menghasilkan solusi optimal secara global. Dalam implementasinya, SVM memerlukan pemilihan parameter penting, seperti *cost parameter* (C) dan jenis kernel yang digunakan [3].

Namun, perlu diperhatikan bahwa Regresi Logistik Biner memiliki kelemahan dalam menghadapi data yang mengandung pencilan (*outlier*). Model ini mengandalkan seluruh distribusi data untuk menentukan garis keputusan, sehingga keberadaan nilai ekstrem dapat menyebabkan bias pada estimasi koefisien. Sebaliknya, SVM lebih *robust* terhadap pencilan karena hanya bergantung pada *support vectors* yang berada di sekitar margin pemisah. Penelitian oleh Huang *et al.* pada konteks prediksi hipertensi menemukan bahwa SVM dan metode *permanental classification* secara konsisten lebih baik daripada regresi logistik dalam hal performa prediksi [4]. Selain itu, Damanik *et al.* menunjukkan bahwa Metode regresi logistik biner mengindikasikan bahwa variabel Jurusan dan IPK berpengaruh terhadap lama studi mahasiswa, dengan tingkat akurasi klasifikasi sebesar 70%. Sementara itu, metode SVM memberikan akurasi klasifikasi tertinggi, yakni mencapai 90%, ketika menggunakan kernel linear, polinomial, dan RBF. Temuan ini memperkuat potensi SVM sebagai alternatif yang lebih andal dalam konteks data yang mengandung *outlier* [5]. Melihat perbedaan karakteristik kedua metode tersebut, membandingkan performa Regresi Logistik Biner dan SVM dalam mengklasifikasi menjadi relevan dan penting karena kedua metode ini memiliki pendekatan yang berbeda dalam memodelkan data klasifikasi. Regresi Logistik Biner menawarkan interpretabilitas yang baik, sementara SVM cenderung menghasilkan akurasi klasifikasi yang lebih tinggi terutama pada data yang kompleks [6]. Oleh karena itu, perlu dilakukan evaluasi kuantitatif performa kedua metode menggunakan metrik seperti akurasi, *precision*, *recall*, dan untuk menentukan metode yang paling sesuai dalam melakukan klasifikasi.

Penelitian ini bertujuan untuk membandingkan performa Regresi Logistik Biner dan SVM dalam mengklasifikasikan kabupaten/kota di Indonesia ke dalam kategori IPM tinggi dan rendah berdasarkan variabel sosial ekonomi, seperti angka harapan hidup, harapan lama sekolah, rata-rata lama sekolah, dan pengeluaran per kapita. Data dari Badan Pusat Statistik (BPS) Indonesia tahun 2024 digunakan untuk penelitian ini, diharapkan dapat memberikan gambaran yang komprehensif tentang kelebihan dan kekurangan masing-masing metode, serta memberikan rekomendasi penggunaan metode klasifikasi yang paling efektif dalam analisis IPM di Indonesia. Regresi logistik biner merupakan metode statistik tradisional yang menghasilkan model *interpretable* dan sangat cocok untuk data dengan hubungan linier antara prediktor dan logit dari probabilitas. Sementara itu, SVM merupakan metode pembelajaran mesin yang unggul dalam menangani data berdimensi tinggi dan pola *non-linier* dengan margin klasifikasi yang optimal. Dengan membandingkan keduanya, kelebihan dan kelemahan masing-masing dalam hal akurasi, interpretabilitas, serta ketahanan terhadap multikolinearitas dan data yang tidak seimbang dapat dievaluasi. Hasil perbandingan ini akan membantu dalam memilih model yang paling efektif dan sesuai dengan karakteristik data IPM yang dianalisis.

Indeks Pembangunan Manusia (IPM) merupakan salah satu isu penting dalam konteks pembangunan di Indonesia. IPM digunakan sebagai tolok ukur dalam menilai kualitas pembangunan suatu negara, serta dalam menentukan apakah negara tersebut tergolong maju, berkembang, atau tertinggal. IPM dibentuk berdasarkan tiga dimensi utama yang hingga kini masih konsisten digunakan untuk merepresentasikan kualitas hidup manusia. Ketiga dimensi tersebut meliputi umur panjang dan hidup sehat, tingkat pengetahuan, serta taraf hidup yang layak [7]. IPM mencerminkan sejauh mana penduduk di suatu daerah mampu mengakses dan merasakan manfaat dari pembangunan, terutama dalam aspek pendapatan, layanan kesehatan, pendidikan, dan aspek kesejahteraan lainnya [8]. Selama periode 2015 hingga 2022, Indeks Pembangunan Manusia (IPM) di Indonesia menunjukkan tren peningkatan setiap tahunnya. Pada tahun 2015, nilai IPM Indonesia tercatat sebesar 69,55 dan terus naik hingga mencapai 72,91 pada tahun 2022. Meskipun mengalami peningkatan secara nasional, masih terdapat tantangan berupa ketimpangan IPM antar kabupaten/kota. Ketidakmerataan pembangunan ini dipengaruhi oleh berbagai faktor, baik yang berasal dari indikator penyusun IPM seperti umur harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pendapatan per kapita maupun faktor-faktor lain di luar indikator tersebut [9]. Setiap indikator penyusun IPM

dapat dimanfaatkan untuk menilai sejauh mana peningkatan kualitas hidup manusia telah tercapai (Badan Pusat Statistik 2024).

Badan Pusat Statistik (BPS) mengklasifikasikan tingkat Indeks Pembangunan Manusia (IPM) ke dalam empat kategori yaitu sangat tinggi, tinggi, sedang, dan rendah sesuai dengan standar klasifikasi yang umum digunakan dalam literatur pengukuran pembangunan manusia. Pengukuran IPM sendiri mencakup tiga dimensi utama pembangunan manusia, yaitu umur panjang dan hidup sehat, pengetahuan, serta standar hidup layak. Dimensi umur panjang dan hidup sehat diukur melalui angka harapan hidup saat lahir, dimensi pengetahuan diukur melalui gabungan rata-rata lama sekolah dan harapan lama, sedangkan dimensi standar hidup layak diukur menggunakan pendapatan kotor nasional per kapita yang disesuaikan dengan paritas daya beli. Masing-masing indikator tersebut kemudian dinormalisasi ke dalam rentang 0–1 menggunakan formula indeks, dan nilai akhir IPM dihitung menggunakan rata-rata geometrik dari ketiga dimensi [10]. Pendekatan ini mengacu pada metodologi yang dikembangkan oleh *United Nations Development Programme* (UNDP) dan telah diadopsi dalam penghitungan IPM di Indonesia oleh BPS. Secara umum, IPM Indonesia menunjukkan tren pertumbuhan yang positif. Pada tahun 2023, pertumbuhan IPM diperkirakan mencapai 0,84 persen, meningkat signifikan dibandingkan kenaikan sebesar 0,48 persen pada tahun 2021. Sejalan dengan itu, perkembangan IPM di berbagai daerah juga menunjukkan perbaikan secara keseluruhan, termasuk di wilayah-wilayah yang sebelumnya memiliki kinerja pembangunan yang rendah. Meskipun demikian, kesenjangan IPM antar 38 provinsi serta kabupaten/kota di Indonesia masih menjadi tantangan yang perlu diatasi [11].

Dalam mengkaji perkembangan IPM di berbagai wilayah, penting untuk mempertimbangkan kualitas data yang digunakan. Beberapa studi sebelumnya menunjukkan bahwa data IPM yang diperoleh dari BPS berpotensi mengandung pencilan (*outlier*). Misalnya, penelitian oleh Puspitasari et al. menemukan keberadaan *outlier* pada data IPM tahun 2019 di Provinsi Jawa Timur, baik pada variabel dependen maupun pada variabel-variabel prediktornya, seperti angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pendapatan per kapita [12]. Permasalahan kualitas data seperti keberadaan *outlier* tersebut menjadi hambatan dalam menghasilkan analisis IPM yang akurat, padahal akurasi ini sangat dibutuhkan untuk mendukung pemahaman yang komprehensif terhadap proses pembangunan yang bersifat *multidimensional* dan berkelanjutan.

Oleh karena itu, IPM perlu diklasifikasikan untuk mempermudah identifikasi dan pemetaan tingkat kesejahteraan serta kualitas hidup penduduk di berbagai wilayah. Dengan mengelompokkan IPM ke dalam kategori-kategori tertentu, misalnya rendah dan tinggi sehingga pengambil kebijakan dapat lebih mudah menentukan prioritas intervensi pembangunan, alokasi sumber daya, dan evaluasi program yang tepat sasaran. Klasifikasi ini juga membantu dalam mengungkap ketimpangan pembangunan antar daerah, sehingga strategi pembangunan dapat diarahkan secara lebih terfokus dan adil. Selain itu, dalam konteks analisis data, klasifikasi IPM memungkinkan penggunaan metode prediktif seperti regresi logistik atau SVM untuk mengidentifikasi faktor-faktor utama yang memengaruhi kualitas pembangunan manusia di suatu wilayah.

II. METODE PENELITIAN

A. Data

Penelitian ini menggunakan data sekunder dengan tipe *cross-section* yang diperoleh dari Badan Pusat Statistik (BPS) yang bertujuan untuk mengevaluasi dan membandingkan kinerja Regresi Logistik Biner dan *Support Vector Machine* (SVM) dalam klasifikasi Indeks Pembangunan Manusia (IPM) provinsi-provinsi di Indonesia pada tahun 2024. Unit analisis dalam penelitian ini adalah kabupaten/kota yang ada di Indonesia yaitu sebanyak 541 data. Variabel yang digunakan dalam penelitian ini terdiri atas variabel prediktor (X) dan variabel respon (Y). Pemilihan *variable predictor* (X) didasarkan pada tiga dimensi utama penyusun Indeks Pembangunan Manusia (IPM) yang ditetapkan oleh *United Nations Development Programme* (UNDP) yaitu kesehatan, pendidikan, dan standar hidup layak, variabel prediktor yang digunakan tertera dalam Tabel 1. Variabel respon biner yang digunakan adalah status IPM, yaitu hasil kategorisasi dari nilai IPM tahun 2024 ke dalam dua kelas yaitu kelas 1 = IPM Tinggi ($IPM \geq 70$) dan Kelas 0 = IPM Rendah ($IPM < 70$), sesuai dengan klasifikasi dari BPS dan pendekatan yang digunakan oleh Safitri et al. [13].

TABEL 1
VARIABEL PREDIKTOR PENELITIAN

No	Variabel	Keterangan
1	Rata-rata Lama Sekolah (X_1)	Tahun rata-rata pendidikan penduduk
2	Harapan Lama Sekolah (X_2)	Ekspektasi jumlah tahun sekolah yang akan dijalani anak-anak
3	Pengeluaran Per Kapita (X_3)	Disesuaikan dalam ribuan rupiah
4	Angka Harapan Hidup (X_4)	Umur harapan hidup saat lahir (dalam tahun)

No	Variabel	Keterangan
5	Jumlah Penduduk Miskin (X_5)	Jumlah penduduk di bawah garis kemiskinan

B. Tahapan Penelitian

Adapun tahapan analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Menggabungkan data dari berbagai sumber BPS berdasarkan provinsi sebagai unit analisis. Data yang digunakan mencakup nilai IPM tahun 2024 dan lima prediktor utama. Data kemudian diperiksa untuk mendeteksi nilai yang hilang (*missing value*). Beberapa pendekatan seperti *multiple imputation* telah terbukti efektif dalam mengatasi *missing values*, namun karena proporsi data hilang relatif kecil, penelitian ini menggunakan metode *listwise deletion* sebagai pendekatan yang sederhana dan tepat dalam konteks ini. Selanjutnya, variabel target dibentuk dengan mengklasifikasikan IPM menjadi dua kelas berdasarkan ambang batas yaitu tinggi dan rendah [13].
2. Melakukan eksplorasi data melalui statistik deskriptif (mean, min, max, standar deviasi) serta visualisasi berupa pie chart.
3. Melakukan pemeriksaan terhadap distribusi kelas target untuk memastikan keseimbangan antara kelas IPM tinggi dan rendah. Jika distribusi kelas target tidak seimbang, diterapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk menyeimbangkan jumlah data pada kelas minoritas sehingga model klasifikasi yang dihasilkan tidak bias ke kelas mayoritas. Proses oversampling ini hanya dilakukan pada data latih untuk menjaga validitas evaluasi model. SMOTE efektif diterapkan dalam konteks klasifikasi data tidak seimbang.
4. Melakukan uji signifikansi parameter secara simultan dan parsial menggunakan uji G (ratio likelihood test) dengan persamaan sebagai berikut:

$$G = -2 \ln \frac{L_o}{L_p} \quad (1)$$

Dimana L_p merupakan model dengan distribusi *chi-square* yang memiliki derajat bebas sebanyak parameter dalam model, sedangkan L_o merupakan fungsi *likelihood* semua prediktor. Kemudian uji parsial menggunakan uji wald dengan statistik uji nya sebagai berikut:

$$W = \frac{(\hat{\beta}_i)^2}{[SE(\hat{\beta}_i)]^2} \quad (2)$$

Dimana $SE(\hat{\beta}_i)$ merupakan galat baku dari koefisien $\hat{\beta}_i$.

5. Melakukan analisis regresi logistik biner yang dilanjutkan dengan uji kecocokan model menggunakan MacFadden R2. Model untuk regresi logistik biner melibatkan lebih dari dua variabel variabel prediktor dituliskan sebagai berikut:

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)} \quad (3)$$

Dimana k adalah jumlah variabel prediktor.

6. Menentukan kesimpulan dengan odds rasio [7].

Model Regresi Logistik dan SVM yang dibangun dilatih menggunakan 80% data latih dan diuji menggunakan 20% data uji. Regresi logistik digunakan untuk mengestimasi probabilitas status IPM tinggi berdasarkan variabel prediktor. Parameter model diestimasi menggunakan *Maximum Likelihood Estimation* (MLE) dan diuji signifikansinya dengan uji Wald dan *Likelihood Ratio Test* (G). Model SVM digunakan untuk mengklasifikasikan data ke dalam dua kelas menggunakan *hyperplane optimal*. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, serta *confusion matrix*. Validasi tambahan dilakukan menggunakan teknik *5-fold cross-validation*. Adapun fungsi linear dari SVM adalah sebagai berikut:

$$g(x) = \text{sign}(f(x)) \quad (4)$$

Dimana $f(x) = W^t x + b$. Beberapa jenis kernel yang biasanya digunakan dalam SVM tertera dalam Tabel 2.

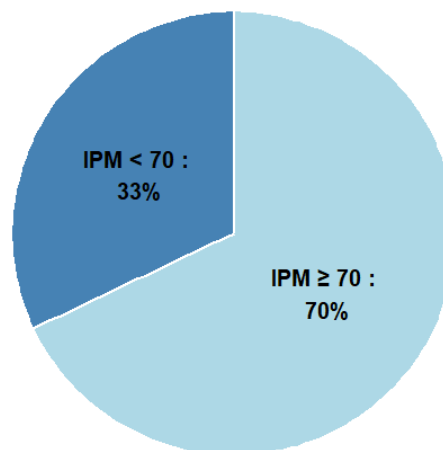
TABEL 2
JENIS KERNEL DALAM SVM

No	Nama Kernel	Fungsi
1	Linear	$K(x, y) = x, y$
2	Polinomial	$K(x, y) = (x \cdot y + c)^2$
3	Gaussian RBF	$K(x, y) = \exp\left(\frac{-\ x - y\ ^2}{2 \cdot \sigma^2}\right)$
4	Sigmoid	$K(x, y) = \tanh(\sigma(x, y) + c)$
5	Invers Multikuadrik	$K(x, y) = \frac{1}{\sqrt{\ x - y\ ^2 + c^2}}$

Dimana x adalah data latih dan y adalah data uji dengan c adalah koefisien [14].

III. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif

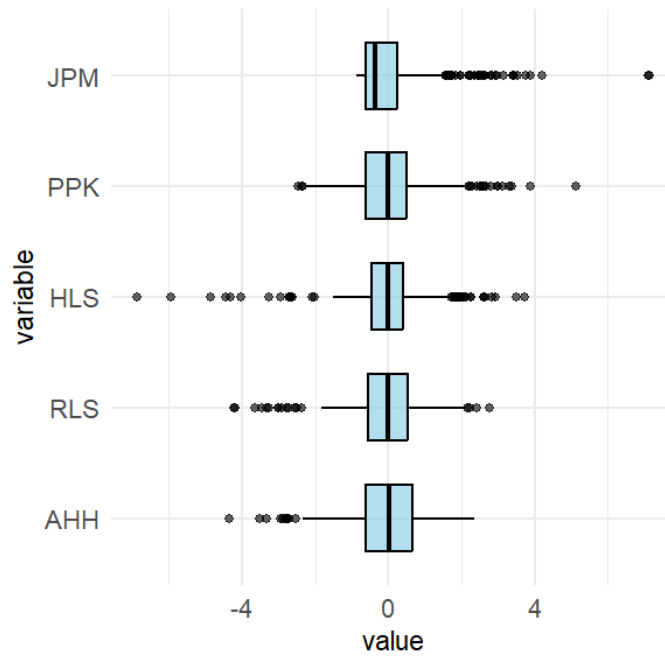


Gambar 1. Persentase IPM

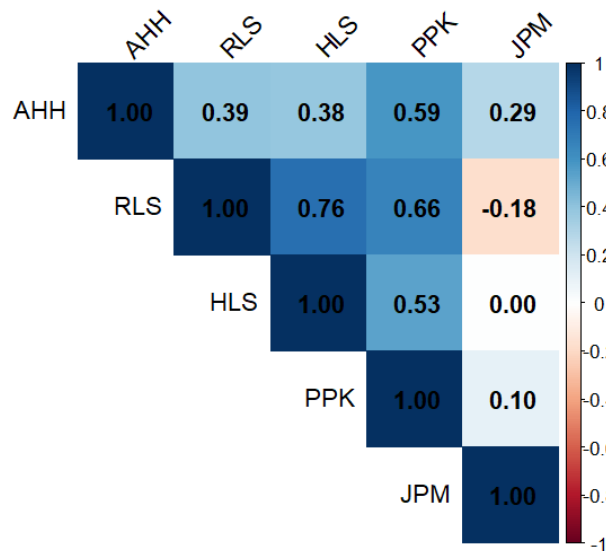
Berdasarkan Gambar 1 yang menunjukkan diagram lingkaran distribusi persentase daerah berdasarkan kategori Indeks Pembangunan Manusia (IPM). Sebesar **67%** daerah memiliki $IPM \geq 70$, yang menunjukkan tingkat pembangunan manusia yang tergolong tinggi. Sementara itu, **33%** daerah lainnya memiliki $IPM < 70$, yang menandakan bahwa sepertiga dari total daerah masih berada dalam kategori pembangunan manusia menengah ke bawah. Visualisasi ini memberikan gambaran umum bahwa mayoritas wilayah telah mencapai standar pembangunan manusia yang cukup baik, namun masih terdapat proporsi signifikan yang memerlukan perhatian lebih dalam peningkatan kualitas hidup, pendidikan, dan kesejahteraan penduduk.

Berdasarkan Gambar 2 menunjukkan bahwa setiap variabel prediktor atau variabel prediktor memiliki pencilan (*outlier*). Hal ini merupakan salah satu alasan mengapa penelitian ini membandingkan dua metode yaitu logistik biner dan SVM. Ini dikarenakan SVM lebih robust terhadap pencilan (*outlier*).

Berdasarkan hasil visualisasi matriks korelasi pada Gambar 3, terlihat bahwa sebagian besar variabel prediktor memiliki hubungan positif satu sama lain dengan kekuatan bervariasi, di mana korelasi tertinggi terjadi antara Rata-rata Lama Sekolah (RLS) dan Harapan Lama Sekolah (HLS) sebesar 0,76 yang menunjukkan hubungan yang kuat. Korelasi sedang juga terlihat antara RLS dan Pengeluaran Per Kapita (PPK) sebesar 0,66 serta antara HLS dan PPK sebesar 0,53. Sementara itu, Angka Harapan Hidup (AHH) memiliki korelasi sedang dengan PPK (0,59) dan korelasi lemah dengan variabel pendidikan (RLS dan HLS).



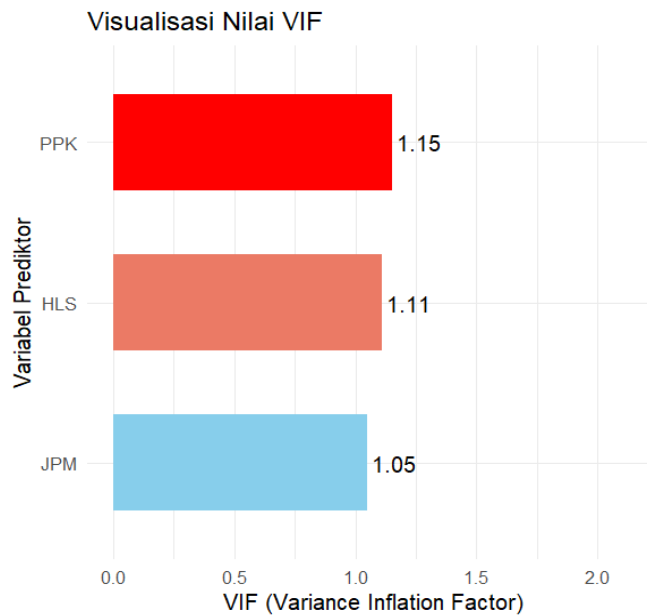
Gambar 2. Cek Penculan Variabel Prediktor



Gambar 3. Korelasi Antar Variabel

Jumlah Penduduk Miskin (JPM) tampak tidak berkorelasi kuat dengan variabel lain, bahkan memiliki korelasi negatif lemah dengan RLS (-0,18), menunjukkan bahwa JPM dapat memberikan informasi yang unik dalam model prediksi. Dengan memperhatikan korelasi yang cukup tinggi antara RLS dan HLS, pemilihan hanya HLS dalam model merupakan langkah yang tepat untuk menghindari potensi multikolinearitas, sehingga model menjadi lebih stabil dan interpretatif. Oleh karena itu dalam pemodelan regresi logistic biner dipilih tiga variabel yang relevan dengan nilai korelasi yang tidak terlalu tinggi diantaranya Harapan Lama Sekolah (HLS), Pengeluaran Per Kapita (PPK), dan Jumlah Penduduk Miskin (JPM).

Gambar 4 di atas menampilkan visualisasi nilai Variance Inflation Factor (VIF) untuk tiga variabel prediktor dalam model regresi logistic biner, yaitu PPK, HLS, dan JPM. Nilai VIF tertinggi adalah PPK sebesar 1.15, diikuti HLS sebesar 1.11, dan JPM sebesar 1.05. Seluruh nilai ini jauh di bawah ambang batas umum sebesar 5, yang menunjukkan tidak adanya indikasi multikolinearitas antar variabel.



Gambar 4. Nilai VIF Variabel Prediktor yang Dipilih

B. Uji Signifikansi Parameter Secara Simultan

Uji Simultan (Uji G) dilakukan untuk menguji apakah ketiga variabel prediktor secara bersama-sama berpengaruh signifikan terhadap variabel dependen. Hasil uji signifikansi parameter secara simultan dengan R dapat dilihat dalam Tabel 3 di bawah ini. Uji dilakukan dengan membandingkan model null dan model penuh.

TABEL 3
UJI SIGNIFIKANSI PARAMETER SIMULTAN

Model	Residual DF	Residual Deviance	Df	Deviance	P-Value
IPM status ~ 1	579	804,05			
IPM status ~ HLS + PPK + JPM	576	345,50	3	458,55	$< 2,2e - 16^{***}$

Hasil uji G yang tercantum dalam Tabel 3 menunjukkan nilai deviance sebesar 458.55 dengan derajat kebebasan 3 dan nilai p-value kurang dari 0,05. Maka dari itu dapat disimpulkan bahwa model dengan variabel HLS (Harapan Lama Sekolah), PPK (Pengeluaran per Kapita), dan JPM (Jumlah Penduduk Miskin) secara simultan berkontribusi signifikan dalam memodelkan status IPM.

C. Uji Signifikansi Parameter Secara Parsial

Uji parsial (Uji wald) dilakukan untuk mengevaluasi signifikansi masing-masing variabel prediktor secara individu berpengaruh terhadap variable dependen. Hasil uji signifikansi parameter secara parsial dengan R dapat dilihat dalam Tabel 4.

TABEL 4
UJI SIGNIFIKANSI PARAMETER PARSIAL

Variabel	Estimate	Std.error	Z Value	P-Value
Intercept	-37,55	3,75	-10,02	$< 2e - 16^{***}$
HLS	1,75	0,23	7,586	$< 3,3e - 14^{***}$
PPK	0,0014	0,000127	11,027	$< 2e - 16^{***}$
JPM	-0,001973	0,002316	-0,852	0,394

Berdasarkan Tabel 4, Nilai z dan p-value menunjukkan bahwa dua variabel, yaitu ($z = 7,586, p < 0,05$) dan PPK ($z = 11,027, p < 0,05$), berpengaruh signifikan terhadap status IPM. Sebaliknya, variabel JPM memiliki nilai p-value sebesar 0,394 yang lebih besar dari 0,05, sehingga secara parsial tidak signifikan dalam memengaruhi status IPM. Meskipun demikian, karena uji simultan menunjukkan bahwa ketiga variabel secara bersama signifikan, maka keberadaan JPM tetap relevan dalam model secara keseluruhan.

D. Uji Kecocokan Model

Uji kecocokan model (McFadden's R^2) mengukur seberapa baik model menjelaskan variasi dalam data, dibandingkan dengan model yang hanya memprediksi rata-rata atau intercept (model null). Nilai McFadden's R^2 berada di antara 0 dan 1, dengan semakin tinggi nilainya menunjukkan model yang lebih baik dalam menjelaskan data. Dalam model regresi logistic biner pada kasus ini dihasilkan nilai McFadden's R^2 sebesar 0.5703 menunjukkan bahwa model ini mampu menjelaskan sekitar 57% variasi data. Ini menunjukkan model yang cukup baik, karena nilai di atas 0.4 umumnya dianggap menunjukkan model yang efektif dalam regresi logistik. Secara keseluruhan, model ini memiliki kinerja yang baik dalam menjelaskan data.

E. Model Regresi Logistik Biner

Berdasarkan analisis regresi logistic biner yang telah dilakukan, maka estimasi untuk koefisien variabel prediktor dapat dilihat dalam Tabel 5.

TABEL 5
KOEFSISIEN VARIABEL PREDIKTOR

Variabel	Estimate (β)	$exp(\beta)$
Intercept	-37,55	$4,91e - 17$
HLS	1,75	5,75
PPK	0,0014	1,0014
JPM	-0,001973	0,998

Berdasarkan nilai – nilai koefisien variabel prediktor pada Tabel 5, maka model logit dari regresi logistik biner adalah: $log\ log\left(\frac{p}{1-p}\right) = -37,55 + 1,75HLS + 0,0014PPK - 0,001973JPM$. Dengan p adalah probabilitas bahwa status IPM berada dalam kategori tinggi (IPM_status = 1). Model ini menunjukkan bahwa setiap peningkatan satu satuan pada variabel HLS (Harapan Lama Sekolah) akan meningkatkan log odds peluang suatu daerah masuk kategori IPM tinggi sebesar 1,75, sedangkan setiap kenaikan satu satuan pada PPK (Pendapatan Per Kapita) juga akan meningkatkan log odds tersebut sebesar 0,0014. Sebaliknya, setiap peningkatan satu satuan pada JPM (Jumlah Penduduk Miskin) justru menurunkan log odds peluang suatu daerah tergolong IPM tinggi sebesar 0,001973. Dengan demikian, arah dan besar koefisien pada model ini memberikan gambaran kontribusi masing-masing variabel prediktor terhadap probabilitas suatu wilayah memiliki status IPM tinggi.

F. Interpretasi Odds ratio

Berdasarkan nilai odds ($exp\ exp(\beta)$) yang dihasilkan pada Tabel 5, menandakan pengaruh masing – masing variabel prediktor. Dimana HLS (Harapan Lama Sekolah) memiliki pengaruh paling kuat dan signifikan terhadap status IPM. Setiap peningkatan 1 satuan HLS akan meningkatkan peluang (odds) suatu wilayah memiliki status IPM tinggi sebesar 5.75 kali lipat, dengan asumsi variabel lain tetap. PPK (Pengeluaran Per Kapita) juga berpengaruh signifikan secara statistik, namun karena satuannya besar, efeknya per unit sangat kecil (odds naik sekitar 0.0001%). JPM (Jumlah Penduduk Miskin) memiliki arah pengaruh negatif, tetapi tidak signifikan secara statistik dan hanya menurunkan odds IPM tinggi dalam jumlah yang sangat kecil. Secara keseluruhan, variabel HLS adalah faktor utama dalam memengaruhi peluang peningkatan status IPM, sedangkan PPK dan JPM memberikan kontribusi yang relatif kecil terhadap model.

Berdasarkan hasil *confusion matrix* pada data pengujian, model mampu mencapai tingkat akurasi sebesar 87,85%, yang secara signifikan melampaui *No Information Rate* (67,29%). Nilai *p-value* yang sangat kecil ($8,212e-07$) mengindikasikan bahwa perbedaan ini tidak terjadi secara kebetulan, sehingga dapat disimpulkan bahwa performa model tergolong sangat baik dan signifikan secara statistik.

G. Hasil Analisis Menggunakan SVM

Hasil klasifikasi Indeks Pembangunan Manusia (IPM) menggunakan SVM dengan parameter SVM-Type adalah c-classification, kernel menggunakan jenis radial. Analisis ini menggunakan data latih sebesar 80% dan data uji 20%. Confusion matrix yang dihasilkan tertera dalam Tabel 6 menggunakan data uji.

TABEL 6
CONFUSION MATRIX SVM

	Aktual 0	Aktual 1
Prediksi 0	29 (TN)	8 (FN)
Prediksi 1	6 (FP)	64 (TP)

Berdasarkan Tabel 6, diperoleh rincian hasil klasifikasi sebagai berikut: True Negative (TN) = 29, yang berarti terdapat 29 data berstatus IPM rendah yang berhasil diklasifikasikan dengan benar sebagai IPM rendah. False Positive (FP) = 6, menunjukkan bahwa terdapat 6 data berstatus IPM rendah yang keliru diklasifikasikan sebagai IPM tinggi. True Positive (TP) = 64, menandakan ada 64 data berstatus IPM tinggi yang secara tepat diklasifikasikan sebagai IPM tinggi. Sementara itu, False Negative (FN) = 8, mengindikasikan bahwa 8 data berstatus IPM tinggi salah diklasifikasikan menjadi IPM rendah. Ringkasan hasil evaluasi model melalui *confusion matrix* pada data pengujian disajikan pada Tabel 7 berikut ini.

TABEL 7
METRIK EVALUASI SVM

Metrik	Interpretasi
Accuracy	(86,92%) Model secara menyeluruh mampu mengklasifikasikan data dengan baik.
Sensitivity/Recall kelas 0	(82,86%) model berhasil mendeteksi 82,86% dari semua kasus yang benar-benar termasuk dalam kategori IPM rendah
Specificity/Recall kelas 1	(88,89%) Model cukup andal dalam mengidentifikasi daerah yang tergolong dalam kategori IPM tinggi.
Precision kelas 0	(78,38%) Dari semua prediksi yang diklasifikasikan sebagai IPM rendah, 78,38% memang benar-benar rendah.
Precision kelas 1	(91,43%) Sebagian besar prediksi untuk IPM tinggi benar-benar sesuai.
Balanced Accuracy	(85,87%) Model tidak bias terhadap salah satu kelas.

Model SVM berhasil mengklasifikasikan status IPM dengan akurasi sebesar 86,92%. Ini berarti sekitar 87 dari 100 observasi diklasifikasikan dengan benar. Akurasi ini secara signifikan lebih baik dibandingkan dengan No Information Rate (NIR) sebesar 67,29%, yang menunjukkan bahwa model memiliki kinerja prediksi yang jauh lebih baik daripada sekadar menebak kelas mayoritas secara acak. Hal ini diperkuat dengan nilai p-value sebesar $2,77e - 06$, yang menandakan bahwa peningkatan akurasi ini sangat signifikan secara statistik.

IV. SIMPULAN

Penelitian ini membandingkan performa model Regresi Logistik Biner dan Support Vector Machine (SVM) dalam mengklasifikasikan status Indeks Pembangunan Manusia (IPM) Indonesia tahun 2024. Dari lima variabel predictor awal, hanya tiga variabel yang digunakan setelah dilakukan analisis korelasi untuk menghindari multikolinearitas. Teknik SMOTE diterapkan untuk menangani ketidakseimbangan kelas dalam data. Regresi Logistik dipilih karena sifatnya yang interpretatif, sementara SVM digunakan sebagai pembanding karena kemampuannya yang robust terhadap outliers dan efektif dalam menangani data berdimensi tinggi. Hasil evaluasi menunjukkan bahwa model Regresi Logistik Biner menghasilkan akurasi sebesar 87,85%, sedikit lebih tinggi dibandingkan dengan SVM yang mencapai 86,92%. Ini menunjukkan bahwa Regresi Logistik tetap mampu memberikan hasil klasifikasi yang lebih optimal setelah melalui prapemrosesan data. Penelitian selanjutnya disarankan untuk mengeksplorasi metode klasifikasi lain seperti Random Forest, XGBoost, atau model berbasis neural network serta menggunakan data IPM dari beberapa tahun atau analisis longitudinal untuk memberikan pemahaman yang lebih menyeluruh. Penggunaan teknik interpretabilitas model juga penting untuk menjelaskan kontribusi masing-masing variabel prediktor terhadap hasil klasifikasi, guna mendukung pengambilan kebijakan berbasis data.

DAFTAR PUSTAKA

- [1] W. Alwi, E. Ermawati, and S. Husain, "Analisis Regresi Logistik Biner Untuk Memprediksi Kepuasan Pengunjung Pada Rumah Sakit Umum Daerah Majene," *Jurnal Matematika dan Statistika serta Aplikasinya*, vol. 6, 2018.
- [2] A. Agresti, *Statistical Methods for the Social Sciences*, 5th ed. Hoboken, NJ, USA: John Wiley & Sons, 2018.
- [3] R. I. Arumnisa and A. W. Wijayanto, "Perbandingan metode ensemble learning: Random forest, support vector machine, AdaBoost pada klasifikasi indeks pembangunan manusia (IPM)," *SISTEMASI: Jurnal Sistem Informasi*, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>.
- [4] H. H. Huang, T. Xu, and J. Yang, "Comparing logistic regression, support vector machines, and permanent classification methods in predicting hypertension," in *BMC Proceedings*, BioMed Central Ltd., 2014.
- [5] S. M. S. Damanik, D. Ispriyanti, and S. Sugito, "Klasifikasi Lama Studi Mahasiswa Fsm Universitas Diponegoro Menggunakan Regresi Logistik Biner Dan Support Vector Machine (SVM)," *Jurnal Gaussian*, vol. 4, 2015.
- [6] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021.
- [7] D. M. Ferrari, M. L. Mananohas, and Y. A. R. Langi, "Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Pada Kabupaten/Kota Se-Sulawesi Tahun 2022 Menggunakan Regresi Logistik Biner," *d'Cartesian*, vol. 13, 2024.
- [8] C. A. Pamungkas and W. W. Widiyanto, "Klasifikasi indeks pembangunan manusia di Indonesia tahun 2022 dengan support vector machine," *JUISIK*, vol. 2, no. 3, 2022.
- [9] R. Anggara, Tessa Octavia Mukhti, Yenni Kurniawati, and Dina Fitria, "Perbandingan Metode Naïve Bayes Dan K-Nearest Neighbors Dalam Mengklasifikasikan Indeks Pembangunan Manusia Menurut Kabupaten/ Kota di Indonesia Tahun 2022," *UNP Journal of Statistics and Data Science*, vol. 2, no. 4, pp. 483-488, 2024.

- [10] Kaul Inge, L. Goldstone, B. Hausner, S. Menon, and J. Wei, "Human development report," Oxford University Press : United Nations Development Programme (UNDP), New York, 1990.
- [11] T. M. Sapaat, A. L. Sh. P. Lopian, and S. Y. L. Tumangkeng, "Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Provinsi Sulawesi Utara Tahun (2005-2019)," *Jurnal Berkala Ilmiah Efisiensi*, vol. 20, 2020.
- [12] M. T. S. Puspitasari, Y. Susanti, and S. S. Handajani, "Model regresi robust untuk indeks pembangunan manusia di Jawa Timur dengan estimasi M," *PRISMA: Prosiding Seminar Nasional Matematika*, vol. 4, pp. 659–665, 2021.
- [13] I. Safitri, A. Satria, and R. M. Badri, "Implementasi Algoritma K-Nearest Neighbor Dalam Klasifikasi Indeks Pembangunan Manusia Provinsi Sumatera Selatan Tahun 2023," *Digital Transformation Technology*, vol. 4, no. 2, pp. 768–775, 2024.
- [14] D. A. Rifaldi and A. Ahdika, "K-Medoids and Support Vector Machine in Predicting the Level og Building Damage in Earthquake Insurance Modeling," *Statistika: Statistics and Economy Journal*, vol. 104, no. 3, pp. 351–363, 2024.