Klasifikasi Tingkat Kedalaman Kemiskinan di Indonesia Menggunakan Support Vector Machine dan Regresi Logistik

1st Astikhatul Mufaidah
Program Studi S1 Sains Data
Universitas Telkom
Surabaya, Indonesia
astichatul@student.telkomuniversity.ac.id

2nd Rifdatun Ni'mah
Program Studi S1 Sains Data
Universitas Telkom
Surabaya, Indonesia
rifdatun@telkomuniversity.ac.id

3rd Amalia Nur Alifah
Program Studi S1 Sains Data
Universitas Telkom
Surabaya, Indonesia
amalialifah@telkomuniversity.ac.id

***Abstrak* — Kemiskinan merupakan masalah kompleks yang masih menjadi tantangan utama di Indonesia, dengan dampak yang luas terhadap kesejahteraan masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kedalaman kemiskinan menggunakan dua model machine learning yakni Support Vector Machine (SVM) dan Regresi Logistik, serta mengidentifikasi faktor-faktor yang secara signifikan memengaruhinya. Dataset yang digunakan mencakup variabel sosial-ekonomi dari berbagai wilayah, seperti Bantuan Sosial, Rata-Rata Lama Sekolah, dan Jumlah Penduduk. Hasil analisis menunjukkan bahwa model SVM dan Regresi Logistik sama-sama menghasilkan performa klasifikasi yang tinggi, dengan akurasi 99%. Regresi Logistik pada penelitian ini digunakan untuk mengetahui faktor-faktor yang berpengaruh secara signifikan terhadap tingkat kedalaman kemiskinan melalui pendekatan uji signifikansi statistik. Regresi Logistik menunjukkan bahwa tiga variabel yang paling signifikan adalah Bantuan Sosial, Pendapatan Asli Daerah, dan Rata-rata Lama Sekolah, Temuan ini diharapkan dapat menjadi dasar bagi pengembangan kebijakan yang lebih tepat sasaran dalam upaya pengentasan kemiskinan di Indonesia**

***Kata kunci*— Analisis Data, Indeks Kedalaman Kemiskinan, Kemiskinan, Regresi Logistik, Support Vector Machine.**

I. PENDAHULUAN

Kemiskinan merupakan permasalahan multidimensi yang masih menjadi tantangan utama di Indonesia [1]. Penyebabnya meliputi ketimpangan pembangunan, kelembagaan yang tidak adil, serta kebijakan yang tidak mendukung secara merata [2]. Faktor internal masyarakat, seperti keterbatasan akses dan budaya, juga turut memperburuk kondisi [3]. Indeks Kedalaman Kemiskinan digunakan untuk mengukur seberapa jauh rata-rata pengeluaran penduduk miskin dari garis kemiskinan. Indeks ini menjadi indikator penting dalam menyusun kebijakan yang lebih tepat sasaran [4].

Pada Maret 2023, angka kemiskinan Indonesia tercatat 9,36%, melebihi target APBN sebesar 7,5–8,5% [5]. Pemerintah telah menggelontorkan anggaran besar melalui subsidi dan bantuan sosial, namun efektivitasnya masih dipertanyakan . Contohnya adalah subsidi LPG 3 kg yang hanya dinikmati oleh 33,1% rumah tangga miskin [6]. Selain itu, bantuan sosial seperti bantuan pangan dan iuran JKN juga kerap tidak tepat sasaran. Hal ini menunjukkan perlunya pendekatan berbasis data untuk memastikan penyaluran kebijakan yang lebih efisien.

Pemanfaatan teknologi seperti Big Data dan machine learning kini menjadi solusi strategis dalam mengatasi ketidaktepatan kebijakan sosial. Melalui klasifikasi berbasis data, penerima bantuan dapat diidentifikasi secara lebih tepat. Model seperti Regresi Logistik dan Support Vector Machine (SVM) telah terbukti mampu mengevaluasi dampak kebijakan terhadap kemiskinan. Penelitian sebelumnya [7] menunjukkan bahwa SVM memiliki performa tinggi dalam mengklasifikasikan status kemiskinan. Temuan ini memperkuat potensi pendekatan machine learning sebagai alat bantu pengambilan keputusan dalam perumusan kebijakan publik.

Support Vector Machine (SVM) dan Regresi Logistik menawarkan keunggulan dalam menangani data besar dan kompleks. Data kemiskinan yang mencakup variabel sosial-ekonomi seperti PAD, PDRB, dan bantuan sosial, sangat cocok dianalisis dengan metode ini. Dengan dukungan data berkualitas, kedua model dapat membantu merancang skenario kebijakan yang lebih spesifik dan berdampak. Analisis berbasis machine learning juga mampu menangkap pola tersembunyi yang sulit dijangkau oleh metode statistik konvensional. Ini menjadikan pendekatan ini relevan untuk evaluasi program pemerintah secara komprehensif [8].

Penelitian ini mengembangkan model klasifikasi untuk mengelompokkan tingkat kedalaman kemiskinan di Indonesia, sekaligus mengevaluasi variabel-variabel yang berpengaruh secara signifikan. Support Vector Machine (SVM) diterapkan untuk membangun model prediksi, sedangkan Regresi Logistik digunakan dalam analisis signifikansi faktor penentu. Analisis difokuskan pada data tahun 2023, sehingga hasilnya relevan untuk menggambarkan kondisi kemiskinan terkini. Hasil yang

diperoleh diharapkan menjadi dasar dalam merancang kebijakan pengentasan kemiskinan yang lebih terarah dan berbasis data. Dengan pendekatan ini, penelitian turut berkontribusi terhadap integrasi metode teknologi dalam perumusan kebijakan sosial yang efektif dan berkelanjutan.

II. KAJIAN TEORI

2.1 Kemiskinan

Kemiskinan adalah keadaan yang tidak memiliki kehendak individu dan serba terbatas. Orang-orang ini dianggap miskin karena tingkat pendidikan yang rendah, produktivitas kerja, pendapatan, dan kesejahteraan hidup yang buruk, semua bagian dari lingkaran ketidakberdayaan yang menunjukkan bahwa mereka termasuk dalam kategori miskin [9]. Kemiskinan memiliki banyak aspek dandefi nisi. Dimensi ekonomi adalah yang paling sering disebut sebagai dimensi penilaian. Dalam menilai kriteria kemiskinan, baik BPS maupun World Bank menggunakan standar yang menekankan keadaan hidup orang yang dibawah rata-rata [10].

2.2 Indeks Kedalaman Kemiskinan

Penilaian terhadap perbedaan antara pengeluaran rata-rata penduduk dan garis kemiskinan dilakukan melalui pengembangan indeks kedalaman kemiskinan oleh Badan Pusat Statistik (BPS), yang juga dikenal sebagai indeks gap kemiskinan. Nilai indeks yang tinggi mengindikasikan adanya perbedaan signifikan antara garis kemiskinan dan pengeluaran rata-rata penduduk. Garis kemiskinan sendiri merujuk pada jumlah minimum uang yang diperlukan untuk memenuhi kebutuhan dasar, baik pangan maupun non-pangan [9]. Pada rumus 2.2a dan 2.2b merupakan rumus Indeks Kedalaman Kemiskinan (Poverty Gap Index) dalam studi ekonomi dan sosial untuk mengukur sejauh mana rata-rata pendapatan masyarakat miskin berada di bawah garis kemiskinan.

$$P_1 = \frac{1}{n}\sum_{i=1}^{q}\left[\frac{z - y_1}{z}\right] \quad (2.2a)$$

Keterangan:
n: Jumlah total penduduk
q: Jumlah penduduk miskin (di bawah garis kemiskinan)
z: Garis kemiskinan (nilai batas)
yi: Pengeluaran rata-rata per kapita setiap individu yang berada di bawah garis kemiskinan selama sebulan, (i = 1,2,3,..., q), yi < z.

2.3 Transformasi Logaritma

Transformasi logaritma atau log-transform merupakan salah satu teknik pra pemrosesan data yang digunakan untuk mengatasi permasalahan skewness (kemiringan distribusi), mengurangi pengaruh outlier, serta memperbaiki asumsi norma litas pada data numerik. Transformasi ini umumnya diterapkan pada data dengan distribusi miring ke kanan (right-skewed), di mana sebagian besar nilai berada di kisaran kecil, namun terdapat beberapa nilai ekstrem yang sangat besar [11].

$$X' = log_b\,(X + c) \quad (2.3)$$

Keterangan:
• X : nilai asli dari variabel (harus $X + c \geq 0$)
• X' : nilai hasil transformasi logaritma
• b : basis logaritma, biasanya b = 10 (logaritma basis sepuluh) atau $b = e$ (logaritma natural)
• c : konstanta positif kecil untuk menghindari log(0), biasanya c = 1

Jika menggunakan logaritma natural (basis e), maka bentuk transformasinya menjadi:

$$X' = log_b\,(X + c) \quad (2.2b)$$

2.4 Pelabelan Data

Pelabelan data adalah tahap awal dalam pembelajaran mesin yang bertujuan mengelompokkan data berdasarkan kriteria tertentu [12]. Dalam penelitian ini, pelabelan dilakukan pada variabel *Indeks Kedalaman Kemiskinan (P1)* untuk membentuk dua kelas, yaitu:

TABEL I
PELABELAN DATA

Indeks Kedalaman Kemiskinan (P1)	**Label**
P_1 < Rata-rata	0
P_1 ≥ Rata-rata	1

Karena belum terdapat standar resmi pemerintah untuk klasifikasi P1, maka pendekatan rata-rata digunakan sebagaimana penelitian terdahulu [13] [14]. Metode ini dinilai sederhana, dapat membagi data secara seimbang, serta cocok untuk diterapkan dalam model machine learning.

2.5 Support Vector Machine

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dengan prinsip *Structural Risk Minimization* (SRM), yaitu meminimalkan kesalahan klasifikasi dengan memaksimalkan margin antar kelas. Model SVM membangun sebuah *hyperplane* sebagai batas keputusan untuk memisahkan dua kelas data dalam ruang berdimensi tinggi [15].

Jika data tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel untuk memetakan data ke ruang fitur berdimensi lebih tinggi agar dapat dipisahkan secara linear. Beberapa jenis kernel yang digunakan antara lain [16]:

Kernel Linier:

$$K(x, z) = x^T\, z \quad (2.4a)$$

Kernel Polinomial:

$$K(x, z) = (x^T\, z + c)^d \quad (2.4b)$$

Kernel Radial Basis Function (RBF):

$$K(x, z) = exp\left(\frac{\|x - z\|^2}{2\sigma^2}\right) \quad (2.4c)$$

2.6 Regresi Logistik

Regresi logistik adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen kategorik dan satu atau lebih variabel independen. Model ini digunakan untuk memperkirakan probabilitas suatu kejadian berdasarkan nilai-nilai prediktor, menggunakan fungsi logit sebagai bentuk transformasinya. Model regresi logistik dirumuskan sebagai [17]:

$$log\left(\frac{1-P(Y\leq j)}{P(Y\leq j)}\right)=a_j+\sum_{i=1}^{n}\beta_i X_i \qquad (2.5)$$

Dalam model regresi logistik, $P(Y \leq j)$ merepresentasikan probabilitas kumulatif bahwa variabel dependen Y berada pada kategori j atau lebih rendah. Parameter α_j merupakan intersep atau konstanta untuk kategori j, sedangkan β adalah koefisien regresi yang menunjukkan pengaruh dari variabel independen X_i. Adapun n menunjukkan jumlah total variabel independen yang digunakan dalam model.

2.7 Estimasi Parameter

Maximum Likelihood Estimation (MLE) adalah metode yang digunakan un tuk mencari estimasi parameter β dengan cara memaksimalkan fungsi likelihood. Fungsi likelihood untuk model regresi logistik biner didefinisikan sebagai berikut :

$$L(\beta)=\prod_{i=1}^{n}[\pi(X_i)^{y_i}(1-\pi(X_i))^{1-y_i})]$$

Keterangan:
• yi adalah nilai pengamatan pada variabel ke-i,
• $\pi(xi)$ adalah probabilitas untuk variabel prediktor ke-i.

2.8 Uji Signifikansi

Uji signifikansi dilakukan untuk mengetahui apakah variabel independen berpengaruh secara statistik terhadap variabel dependen. Terdapat dua jenis uji :

- **Uji Parsial**: Menggunakan **Uji Wald** atau **Likelihood Ratio** untuk menilai signifikansi masing-masing variabel secara individu. Nilai p < 0,05 menunjukkan variabel signifikan.
- **Uji Serentak**: Menggunakan **Likelihood Ratio Test** untuk menguji apakah seluruh variabel independen secara bersama-sama berpengaruh signifikan terhadap model.

Variabel dengan nilai p > 0,05 umumnya dianggap tidak signifikan dan dapat dipertimbangkan untuk dikeluarkan dari model.

2.9 Odd Ratio (OR)

Odd Ratio (OR) untuk mengukur seberapa besar pengaruh suatu variabel independen terhadap kemungkinan terjadinya suatu kejadian. Nilai OR dihitung dari eksponensial koefisien regresi [18]:

$$OR = e^{\beta}$$

Interpretasinya:

- OR > 1: variabel meningkatkan peluang kejadian,
- OR < 1: variabel menurunkan peluang kejadian,
- OR = 1: tidak ada pengaruh.

2.10 Pengukuran Kinerja Klasifikasi

Evaluasi model klasifikasi dilakukan untuk mengukur akurasi prediksi, salah satunya menggunakan confusion matrix [19]**.** Matriks ini terdiri dari empat komponen: True Positive (TP)**,** True Negative (TN)**,** False Positive (FP)**,** dan False Negative (FN)**,** yang menggambarkan jumlah prediksi benar dan salah pada masing-masing kelas [20]. Adapun pada Tabel 2.8 merupakan bentuk dari confusion matrix [21]:

TABEL II
CONFUSION MATRIX

Confusion Matrix	**Positif (+)**	**Negatif (-)**
Positif	TP	FN
Negatif	FP	TN

2.11 Stratified K-Fold

Pada proses pengembangan model machine learning, validasi silang (cross validation) merupakan teknik penting untuk mengevaluasi performa model secara objektif. Salah satu bentuk validasi silang yang umum digunakan pada kasus klasifikasi adalah StratifiedKFold. Metode ini membagi dataset ke dalam beberapa lipatan (fold) secara berimbang, dengan memastikan bahwa setiap fold mempertahankan proporsi kelas target yang serupa dengan distribusi aslinya. Hal ini berbeda dari K-Fold biasa yang membagi data secara acak tanpa mempertimbangkan distribusi kelas [22]

III. METODE

3.1 Sumber Data

Data sekunder yang digunakan dalam penelitian ini diperoleh dari sumber resmi, yaitu Kementerian Keuangan dan Badan Pusat Statistik (BPS) untuk periode tahun 2023. Adapun daftar lengkap variabel yang digunakan, sumber data, satuan pengukuran, dan unit observasinya, disajikan pada Tabel 3.1 berikut:

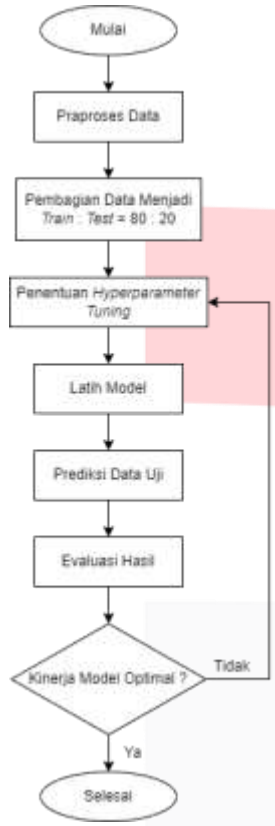
TABEL III
DAFTAR VARIABEL, SUMBER DATA, DAN UNIT OBSERVASI

Variabel	Sumber Data	Satuan Data	Unit Observasi
Belanja Subsidi (X1)	Kemenkeu	Miliar	Kabupaten/Kota
Bantuan Sosial (X2)	Kemenkeu	Miliar	Kabupaten/Kota
PendapatanAsliDaerah (PAD) (X3)	Kemenkeu	Miliar	Kabupaten/Kota
Rata-rata LamaSekolah (RLS) (X4)	BPS	Tahun	Kabupaten/Kota
Tingkat Pengangguran Terbuka (TPT) (X5)	BPS	Persen (%)	Kabupaten/Kota
Laju PDRB (X6)	BPS	Persen (%)	Kabupaten/Kota
Jumlah Penduduk (X7)	BPS	Jiwa	Kabupaten/Kota

Luas Wilayah (X8)	BPS	Km2	Kabupaten/Kota
Indeks Kedalaman Ke miskinan (Y)	BPS	Indeks (0-15)	Kabupaten/Kota

3.2 Teknik Analisis Data

Pada Gambar I ditampilkan flowchart pemodelan SVM :

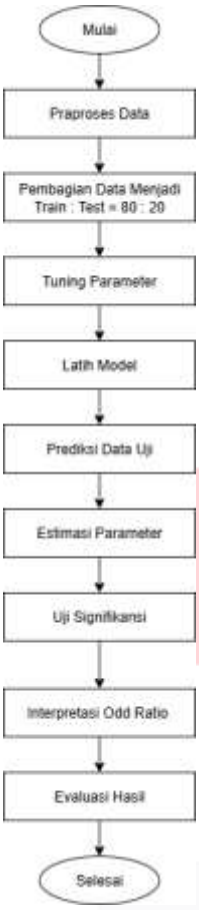


GAMBAR I
FLOWCHART PEMODELAN SVM

Berdasarkan flowchart pada Gambar I berikut langkah-langkah pemodelan SVM dalam penelitian ini :

1. Data yang tersedia diproses untuk memastikan kualitasnya. Proses ini meliputi pembersihan data (mengatasi nilai yang hilang, menghapus data duplikat,dan transformasi data) untuk memastikan bahwa variabel independen dan dependen siap digunakan dalam model.
2. Dataset yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Data latih dimanfaatkan untuk membangun dan menyesuaikan parameter model, sementara data uji digunakan untuk mengevaluasi performa model dalam mengklasifikasikan kategori Indeks Kedalaman Kemiskinan, apakah termasuk dalam kategori rendah atau tinggi.
3. Tuning Parameter ini bertujuan untuk menentukan kombinasi parameter optimal yang menghasilkan performa terbaik pada model klasifikasi. Tuning parameter dilakukan untuk meminimalkan kesalahan prediksi dan meningkatkan akurasi model. Proses tuning dilakukan menggunakan pendekatan GridSearchCV.
4. Model dilatih menggunakan data latih yang telah disiapkan. Dalam proses ini, parameter model diestimasi berdasarkan hubungan antara variabel independen dan dependen untuk membangun model klasifikasi yang mampu memprediksi tingkat kedalaman kemiskinan.
5. Model yang telah melalui proses pelatihan kemudian diterapkan pada data uji untukmemprediksikategoriIndeksKedalamanKemiskinan. Prediksitersebut selanjutnya dibandingkan dengan data aktual pada data uji guna mengevaluasi tingkat akurasi model.
6. Hasil prediksi dievaluasi untuk menilai kinerja model. Evaluasi ini dapat menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan confusion matrix.
7. Kemudian dilakukan evaluasi terhadap kinerja model yang telah dilatih. Jika hasil evaluasi menunjukkan bahwa kinerja model belum optimal, maka lang kah selanjutnya adalah melakukan tuning model lebih lanjut atau penyesuaian terhadap hyperparameter. Namun, jika kinerja model telah mencapai hasil yang optimal, maka proses selesai.

Pada Gambar II ditampilkan flowchart pemodelan Regresi Logistik :



GAMBAR II
FLOWCHART PEMODELAN REGRESI LOGISTIK

Berdasarkan flowchart pada Gambar II berikut langkah-langkah pemodelan Regresi Logistik dalam penelitian ini :

1. Data yang tersedia diproses untuk memastikan kualitasnya. Proses ini meliputi pembersihan data (mengatasi nilai yang hilang, menghapus data duplikat,dan transformasi data) untuk memastikan bahwa variabel independen dan dependen siap digunakan dalam model.
2. Dataset yang telah melalui tahap pra-pemrosesan kemudian dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%. Data latih dimanfaatkan untuk membangun dan menyesuaikan parameter model, sementara data uji digunakan untuk mengevaluasi performa model dalam mengklasifikasikan kategori Indeks Kedalaman Kemiskinan, apakah termasuk dalam kategori rendah atau tinggi.
3. Tuning Parameter ini bertujuan untuk menentukan kombinasi parameter optimal yang menghasilkan performa terbaik pada model klasifikasi. Tuning parameter dilakukan untuk meminimalkan kesalahan prediksi dan meningkatkan akurasi model. Proses tuning dilakukan menggunakan pendekatan GridSearchCV.
4. Membangun model regresi logistik menggunakan data latih. Pada tahap ini, parametermodel(koefisienregresi)diestimasimelalui pendekatanmaksimum likelihood.
5. Model yang telah dilatih kemudian digunakan untuk memprediksi data uji. Hasil prediksi ini dibandingkan dengan data aktual untuk mengevaluasi kinerja model.
6. Parameter dari model regresi logistik, seperti koefisien regresi untuk setiap variabel independen, diestimasi. Parameter ini digunakan untuk mengukur pengaruh variabel independen terhadap variabel dependen.
7. Dilakukan pengujian terhadap koefisien regresi untuk menentukan apakah variabel independen memiliki pengaruh yang signifikan terhadap variabel dependen. Uji ini dilakukan secara parsial (untuk masing-masing variabel independen) dan secara serentak (untuk semua variabel independen).
8. Menghitung Odds Ratio menggunakan hasil estimasi parameter,yang memberikan interpretasi peluang terjadinya kejadian tertentu berdasarkan perubahan variabel independen.
9. Mengevaluasi kinerja model regresi logistik menggunakan metrik seperti aku rasi, presisi, recall, F1-score, dan confusion matrix untuk menilai seberapa baik model memprediksi data.

IV. HASIL DAN PEMBAHASAN

3.3 Proses Pengolahan dan Analisis Data

3.3.1 Prepocessing Data

Tahap pemodelan klasifikasi dilakukan untuk memprediksi kelas indeks kedalaman kemiskinan berdasarkan indikator sosial dan ekonomi di Indonesia, menggunakan platform Google Colab. Pada tahap prapemrosesan, dilakukan pembersihan dan konversi data numerik ke format float akibat penggunaan tanda koma dan spasi, khususnya pada variabel Belanja Subsidi, Bantuan Sosial, dan Pendapatan Asli Daerah. Selanjutnya, dilakukan pengecekan nilai hilang dan ditemukan bahwa seluruh data lengkap (non-null), sehingga 514 baris data dapat digunakan untuk pelatihan model.

Langkah kedua yakni pemahaman karakteristik data. Berikut merupakan ringkasan dari statistik deskriptif fitur numerik:

1. Belanja Subsidi (Miliar): Min = 0, Median = 0,08, Max = 923,28. Sebagian besar wilayah menganggarkan belanja subsidi yang sangat rendah (bahkan nol). Hanya daerah tertentu seperti DKI Jakarta yang mengalokasikan subsidi besar untuk sektor transportasi umum atau energi.
2. Bantuan Sosial (Miliar): Min = 0, Median = 4,06, Max = 727,27. Daerah dengan jumlah penduduk tinggi seperti Kabupaten Bogor atau Bekasi cende rung memiliki anggaran

bansos besar karena banyaknya Keluarga Penerima Manfaat (KPM).

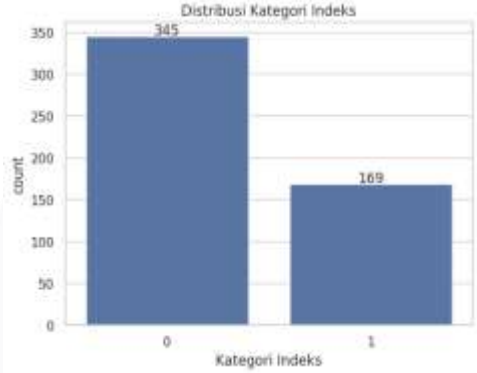
3. Pendapatan Asli Daerah (PAD) (Miliar): Min = 4,35, Median = 331,96, Max = 8.189,96. DKI Jakarta mencatat PAD sangat besar dari pajak daerah, sedangkan daerah seperti Kabupaten Pegunungan Bintang (Papua) memiliki PADdi bawah Rp 10 miliar karena minimnya potensi pendapatan lokal.
4. Rata-rata Lama Sekolah (RLS): Min = 1,71 tahun, Median = 8,53 tahun, Max = 13,04 tahun. Median RLS menunjukkan sebagian besar penduduk hanya menyelesaikan pendidikan sampai jenjang SMP. Wilayah seperti Yogyakarta memiliki RLS tinggi, sedangkan daerah tertinggal seperti Intan Jaya di Papua jauh lebih rendah akibat keterbatasan akses pendidikan.
5. Laju Pertumbuhan PDRB (%): Min =-10,37%, Median = 4,02%, Max = 42,41%. Beberapa daerah seperti Bali mengalami kontraksi ekonomi besar saat pandemi, sementara wilayah pembangunan seperti IKN di Kalimantan Timur menunjukkan pertumbuhan pesat.
6. Tingkat Pengangguran Terbuka (TPT) (%): Min = 0%, Median = 4,02%, Max=11,65%. Kota industri seperti Batam atau Cilegon memiliki TPT tinggi akibat fluktuasi pasar kerja. Di sisi lain, daerah rural mungkin mencatat 0% karena tidak tercatat secara formal, bukan karena lapangan kerja tinggi.
7. LuasWilayah (Km2): Min=10,43, Median=1.943,20, Max=1.475.147,10. Jakarta Pusat termasuk wilayah paling kecil, sedangkan kabupaten seperti Kutai Kartanegara termasuk yang terluas, berdampak pada kesenjangan akses pelayanan.
8. Jumlah Penduduk (Jiwa): Min = 25.377, Median = 662.456, Max = 5.495.372. Kota besar seperti Surabaya, Bandung, atau Kabupaten Bogor memiliki jumlah penduduk sangat tinggi, sementara kabupaten terpencil se perti Supiori (Papua) sangat sedikit.
9. Indeks Kedalaman Kemiskinan (IKM) (%): Min = 0,11, Median = 1,44, Max = 14,01. Sebagian besar wilayah berada di bawah rata-rata nasional (1,95), namun terdapat beberapa daerah dengan tingkat kedalaman kemiskinan ekstrem seperti Asmat dan Lembata.

Langkah ketiga adalah pelabelan data yang dimulai dengan menghitung rata rata indeks kedalaman kemiskinan (P_1), selanjutnya digunakan sebagai dasar untuk membedakan kategori. Hasil klasifikasi berdasarkan nilai tersebut disajikan pada Tabel IV:

TABEL IV
KLASIFIKASI KATEGORI INDEKS KEDALAMAN KEMISKINAN

Kelas	Deskripsi
0	Menunjukkan wilayah dengan tingkat kemiskinan yang lebih rendah daripada rata-rata ($P_1 < 1.95$).
1	Menunjukkan wilayah dengan tingkat kemiskinan yang lebih tinggi daripada rata-rata ($P_1 < 1.95$).

Dalam memberikan gambaran yang lebih jelas mengenai proporsi distribusi wilayah berdasarkan kategori tersebut, visualisasi data disajikan dalam bentuk diagram batang pada Gambar III Visualisasi ini bertujuan untuk menunjukkan sebaran jumlah wilayah dalam masing-masing kategori indeks kedalaman kemiskinan.



GAMBAR III
DISTRIBUSI KATEGORI INDEKS

Variabel Kategori Indeks merupakan variabel hasil kategorisasi dari nilai Indeks Kedalaman Kemiskinan terhadap rata-ratanya. Kategori 0 mengindikasikan bahwa wilayah tersebut memiliki nilai indeks yang lebih rendah dari rata-rata, sedangkan kategori 1 mengindikasikan nilai indeks di atas rata-rata. Berdasarkan visualisasi pada Gambar III, diketahui bahwa mayoritas wilayah (sebanyak 345 entri atau sekitar 67,1%) termasuk ke dalam kategori 0, sementara sisanya (169 entri atau 32,9%) tergolong dalam kategori 1. Dalam kasus ini, kedua kelas tersebut tidak seimbang jumlah data pada kelas 0 lebih banyak dibandingkan kelas 1 maka teknik yang digunakan untuk mengatasi imbalance ini dengan penggunaan parameter stratify=y saat proses pembagian data bertujuan untuk menjaga proporsi kelas yang seimbang pada data pelatihan dan data pengujian. Dengan demikian, komposisi kelas dalam kedua subset tersebut akan mencerminkan distribusi asli dari dataset secara keseluruhan, sehingga model tidak bias terhadap kelas mayoritas

Sebagai bagian dari analisis distribusi kategori indeks kedalaman kemiskinan, pemetaan wilayah dilakukan guna menunjukkan persebaran spasial dari masing masing kategori. Visualisasi seperti pada Gambar 4.2 memberikan

gambaran geo grafis mengenai wilayah-wilayah yang tergolong dalam kategori indeks kemiskinan rendah maupun tinggi, sehingga mempermudah pemahaman terhadap pola sebaran wilayah berdasarkan tingkat kedalaman kemiskinan.

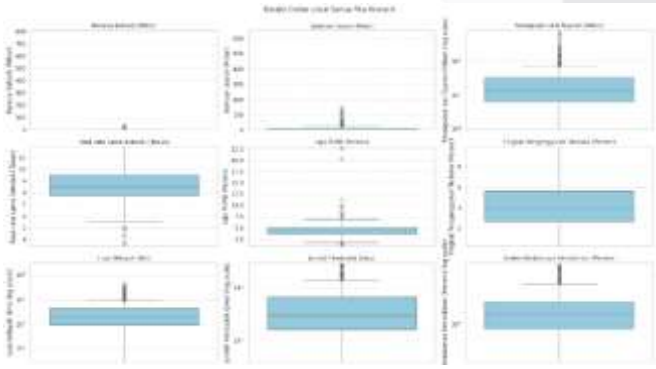
GAMBAR IV
PEMETAAN WILAYAH BERDASARKAN KATEGORI INDEKS KEDALAMAN KEMISKINAN

Visualisasi diatas menunjukkan distribusi spasial wilayah berdasarkan kategori indeks kedalaman kemiskinan. Warna biru mewakili kategori 0, yaitu wilayah dengan tingkat kemiskinan lebih rendah daripada rata-rata, sedangkan warna hijau menunjukkan kategori 1, yaitu wilayah dengan tingkat kemiskinan lebih tinggi daripada rata-rata.

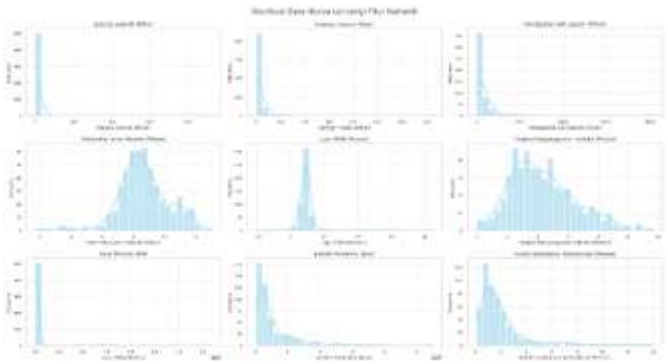
Dari peta tersebut dapat diamati bahwa:

1. Wilayah kategori 0 (rendah) umumnya tersebar di bagian barat Indonesia, seperti sebagian besar wilayah di Sumatera, Jawa, dan Kalimantan bagian barat.
2. Wilayah kategori 1 (tinggi) lebih banyak terdapat di kawasan Indonesia timur, khususnya di Papua, Maluku, dan sebagian Sulawesi, yang mengindikasikan tingkat kedalaman kemiskinan yang lebih parah di daerah-daerah tersebut.

Langkah keempat dilakukan eksplorasi terhadap distribusi dan potensi kebera daan nilai yang menyimpang (outlier) pada setiap variabel numerik menggunakan visualisasi boxplot dan histogram sebagaimana ditunjukkan pada Gambar V dan VI:



GAMBAR V
BOXPLOT UNTUK MENDETEKSI OUTLIER SETIAP FITUR



GAMBAR VI
HISTOGRAM KURVA UNTUK MELIHAT POLA DISTRIBUSI SETIAP FITUR

Visualisasi boxplot dan histogram dalam penelitian ini digunakan untuk menganalisis persebaran serta mendeteksi pencilan (outlier) pada fitur numerik. Evaluasi distribusi mengacu pada lima angka statistik deskriptif (minimum, Q1, median, Q3, maksimum) dan Interquartile Range (IQR). Beberapa temuan penting:

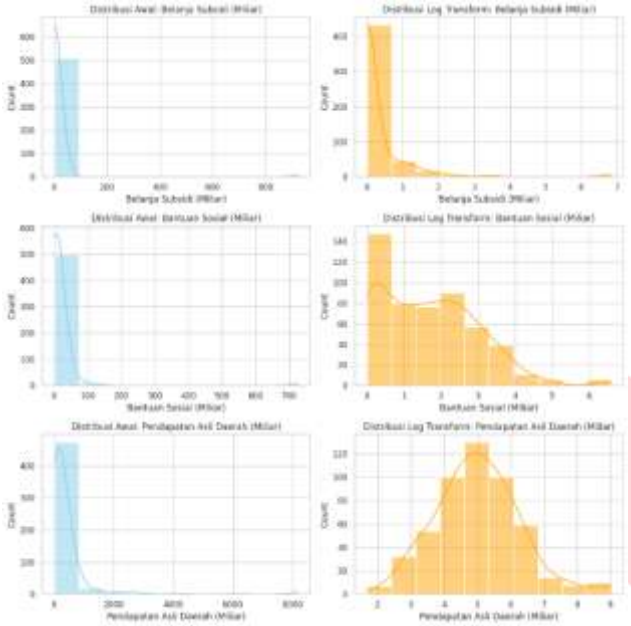
1. **Belanja Subsidi**: Mayoritas wilayah tidak menerima subsidi (nilai Q1, median, dan minimum = 0). Namun, terdapat beberapa wilayah dengan anggaran subsidi sangat besar (outlier), misalnya DKI Jakarta. Data sangat skewed ke kanan, sehingga perlu dilakukan transformasi logaritmik.
2. **Bantuan Sosial**: Terdapat rentang nilai yang sangat lebar, menunjukkan ketimpangan alokasi bansos antarwilayah. Banyak outlier ditemukan di atas nilai ambang IQR.
3. **PAD (Pendapatan Asli Daerah)**: Distribusi tidak merata, dengan outlier di wilayah ber-PAD tinggi seperti DKI Jakarta.
4. **Rata-rata Lama Sekolah, Laju PDRB, TPT, Luas Wilayah, Jumlah Penduduk, dan IKM** juga menunjukkan distribusi tidak normal dan keberadaan outlier, mencerminkan ketimpangan nyata antarwilayah di Indonesia.

Beberapa variabel seperti Belanja Subsidi, Bantuan Sosial, dan PAD memiliki distribusi yang sangat miring ke kanan, sehingga perlu dilakukan transformasi logaritmik. Transformasi ini berfungsi untuk mengurangi skewness, menstabilkan varians, dan mengurangi dampak outlier. Dengan log transformasi, nilai besar dikecilkan secara proporsional lebih besar dibanding nilai kecil, sehingga distribusi menjadi lebih simetris dan outlier berkurang. Hal ini membantu menciptakan pola data yang lebih representatif

Langkah terakhir setelah transformasi, dilakukan standardisasi data menggunakan StandardScaler, yang mengubah skala fitur menjadi memiliki rata-rata nol dan standar deviasi satu. Ini penting agar algoritma seperti SVM dan Regresi Logistik dapat bekerja optimal karena keduanya sensitif terhadap skala fitur.

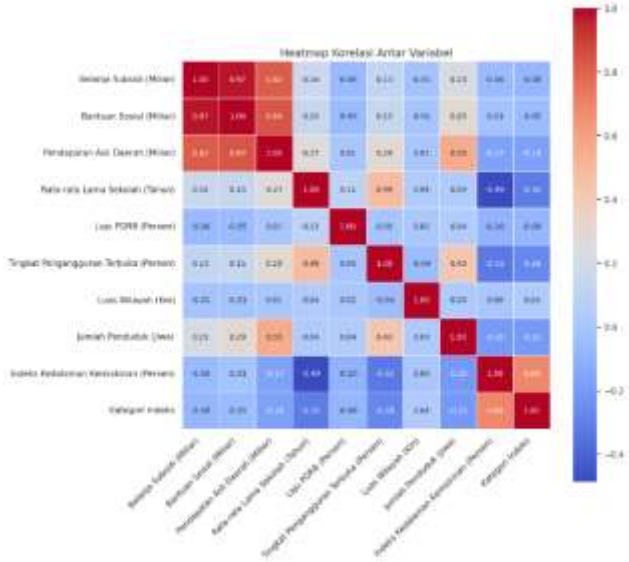
3.3.2 Exploratory Data Analysis (EDA)

Langkah pertama dalam EDA ini adalah menyajikan visualisasi hasil transformasi logaritma untuk menunjukkan perubahan bentuk distribusi pada beberapa variabel numerik seperti pada Gambar VII.



GAMBAR VIi
VISUALISASI HASIL LOG-TRANSFORM

Kemudian untuk memahami hubungan antar variabel numerik yang digunakan, analisis korelasi dilakukan melalui visualisasi heatmap. Tujuannya adalah untuk mengevaluasi kekuatan dan arah hubungan antar variabel, serta mendeteksi potensi multikolinearitas. Gambar VIII berikut menyajikan matriks korelasi antar variabel:



GAMBAR VIII
HEATMAP UNTUK MELIHAT KORELASI ANTAR VARIABEL

Gambar VIII menyajikan peta korelasi antar variabel numerik menggunakan koefisien Pearson. Fokus utamanya adalah mengidentifikasi seberapa kuat hubungan setiap fitur numerik terhadap Indeks Kedalaman Kemiskinan (IKM). Analisis ini mengacu pada hasil eksplorasi data sebelumnya seperti distribusi (histogram) dan boxplot. Hasil utama yang ditemukan:

1. Belanja Subsidi (r = -0,06): Korelasi negatif sangat lemah. Sebagian besar daerah tidak menerima subsidi (nilai Q1 dan median = 0), sehingga distribusinya tidak merata dan kurang efektif menurunkan IKM.

2. Bantuan Sosial (r = -0,01): Hubungan hampir netral. Ketidakefektifan bansos ini dapat disebabkan oleh distribusi tidak merata atau sifatnya yang jangka pendek.

3. Pendapatan Asli Daerah (PAD) (r = -0,17): Korelasi negatif lemah. PAD tinggi belum sepenuhnya digunakan untuk program penanggulangan kemiskinan secara tepat sasaran.

4. Rata-rata Lama Sekolah (RLS) (r = -0,49): Ini adalah korelasi negatif terkuat. Pendidikan yang lebih tinggi cenderung menurunkan kedalaman kemiskinan, sejalan dengan temuan distribusi sebelumnya.

5. Laju PDRB (r = -0,10): Hubungan lemah menunjukkan pertumbuhan ekonomi belum inklusif, terutama di wilayah industri dan tambang.

6. Tingkat Pengangguran Terbuka (TPT) (r = -0,32): Korelasi sedang, meskipun terbalik dari teori umum. Di Indonesia, TPT tinggi tidak selalu berarti IKM tinggi, karena terjadi di wilayah urban dengan fasilitas lebih baik.

7. Luas Wilayah (r = 0,00): Tidak ada hubungan linier, namun secara distribusi ada tantangan logistik di wilayah yang sangat luas.

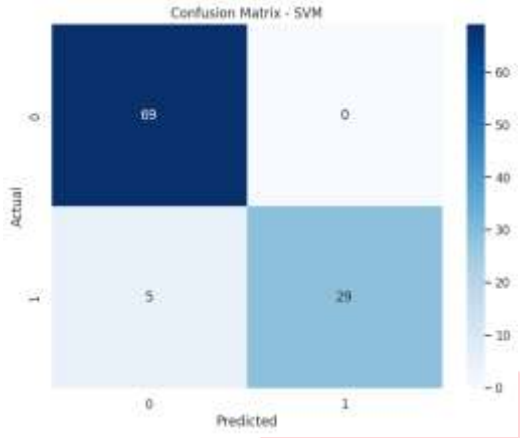
8. Jumlah Penduduk (r = -0,20): Wilayah berpenduduk besar cenderung memiliki IKM lebih rendah, mungkin karena efek aglomerasi dan efisiensi distribusi sumber daya.

3.3.3 Tuning Parameter

Pada proses ini dilakukan pencarian konfigurasi terbaik untuk model SVM dan Regresi Logistik menggunakan metode GridSearchCV dengan validasi silang. Untuk model SVM, kombinasi parameter terbaik adalah kernel linear dengan nilai C = 10 dan gamma = scale, yang menghasilkan akurasi validasi silang tertinggi sebesar 99,03%. Sementara itu, model Regresi Logistik juga menunjukkan performa optimal pada parameter C = 10, penalty = l2, dan solver = lbfgs, dengan akurasi yang sama. Hasil ini menunjukkan bahwa kedua model mampu mengklasifikasikan data dengan sangat baik setelah dilakukan penyetelan parameter secara optimal.

4.2 Pemodelan

3.3.4 Model Klasifikasi Support Vector Machine (SVM)



GAMBAR IX
CONFUSION MATRIX SVM

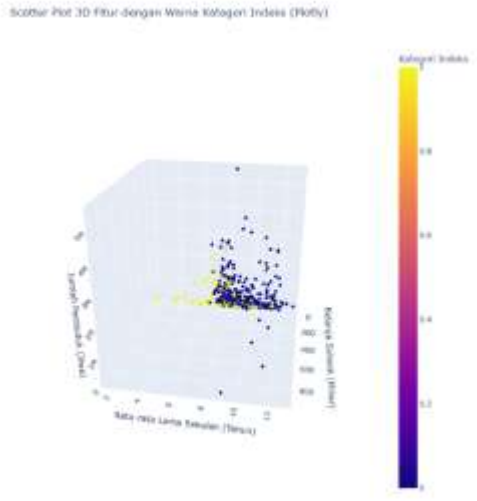
Evaluasi model SVM pada data uji menunjukkan performa sangat baik dengan akurasi 95%. Model mengklasifikasikan semua data kelas 0 dengan benar, namun terdapat 5 kesalahan klasifikasi pada kelas 1, menyebabkan recall kelas 1 hanya 85%. Meski begitu, precision kelas 1 mencapai 100%, dengan F1-score masing-masing 0,97 (kelas 0) dan 0,92 (kelas 1). Rata-rata makro dan weighted F1 masing-masing sebesar 0,94 dan 0,95, menunjukkan kinerja stabil meski data tidak seimbang. Secara keseluruhan, model efektif mengenali kedua kelas, meski masih kurang akurat pada kategori kemiskinan tinggi.

Setelah pelatihan model SVM, dilakukan evaluasi kontribusi fitur menggunakan metode Permutation Importance. Nilai importance dan deviasi standarnya disajikan dalam Tabel V:

TABEL V
KLASIFIKASI KATEGORI INDEKS KEDALAMAN KEMISKINAN

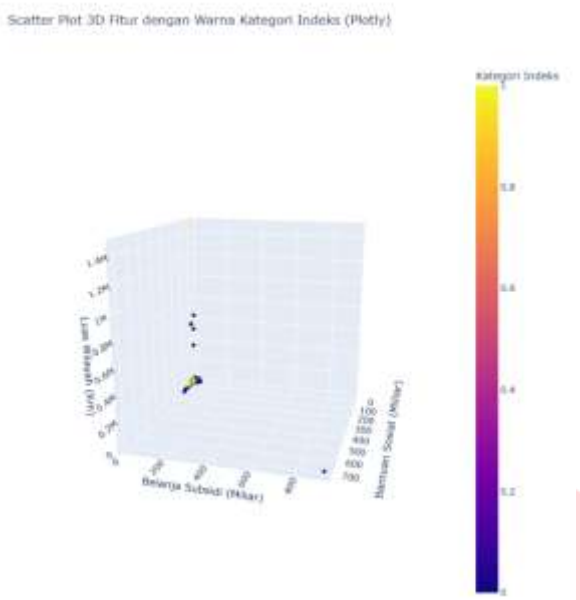
Fitur	Importance	Std Dev
Belanja Subsidi (Miliar)	0.000971	0.005228
Bantuan Sosial (Miliar)	-0.001942	0.008464
Pendapatan Asli Daerah (Miliar)	0.000000	0.000000
Rata-rata Lama Sekolah (Tahun)	0.000000	0.000000
Laju PDRB (Persen)	0.000000	0.000000
Tingkat Pengangguran Terbuka (Persen)	0.000000	0.000000
Luas Wilayah (Km2)	-0.012621	0.008738
Jumlah Penduduk (Jiwa)	0.000000	0.004342

Permutation Feature Importance menunjukkan bahwa sebagian besar fitur tidak berkontribusi signifikan terhadap performa model SVM, ditandai dengan nilai importance nol pada beberapa variabel seperti PAD, RLS, PDRB, TPT, dan jumlah penduduk. Fitur Luas Wilayah dan Bantuan Sosial bahkan berdampak negatif terhadap performa model. Hanya Belanja Subsidi yang memberikan pengaruh positif meski kecil (importance = 0,00097), menunjukkan kontribusi terbatas dalam membedakan kategori kemiskinan. Evaluasi model dan analisis fitur menunjukkan beberapa variabel dengan kontrbusi tinggi dalam klasifikasi. Oleh karena itu,dilakukan visualisasi untuk mengamati hubungan spasial antara fitur-fitur tersebut dan distribusi kelas target.



GAMBAR X
VISUALISASI RUANG FITUR 3D DENGAN FITUR BELANJA SUBSIDI, RATA-RATA LAMA SEKOLAH, DAN JUMLAH PENDUDUK

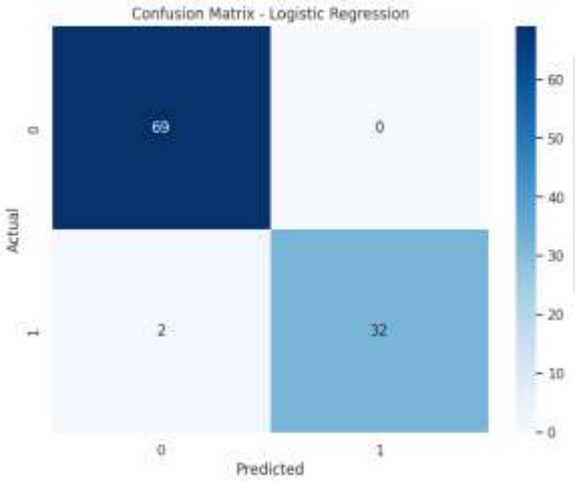
Gambar 4.10 Scatter Plot 3D di atas menunjukkan hubungan antara rata-rata lama sekolah, belanja subsidi, dan jumlah penduduk, dengan pewarnaan berdasarkan kategori indeks kedalaman kemiskinan. Terlihat bahwa sebagian besar titik data berada di kisaran rata-rata lama sekolah 6–9 tahun dan belanja subsidi di bawah 200 miliar. Titik berwarna kuning, yang menunjukkan kategori kemiskinan tinggi cenderung berada pada daerah dengan lama sekolah rendah dan subsidi terbatas. Sebaliknya, warna biru tua yang menandakan kemiskinan rendah umumnya muncul di wilayah dengan lama sekolah lebih tinggi. Pemilihan fitur rata-rata lama sekolah dan jumlah penduduk dilakukan karena keduanya berpengaruh besar terhadap tingkat kemiskinan. Pendidikan yang lebih tinggi meningkatkan kualitas SDM dan peluang ekonomi, sementara jumlah penduduk memengaruhi distribusi sumber daya. Oleh karena itu, analisis terhadap kedua fitur ini penting untuk memahami pola kemiskinan secara menyeluruh.



GAMBAR XI
VISUALISASI RUANG FITUR 3D DENGAN FITUR SIGNIFIKAN

Eksplorasi scatter plot 3D menunjukkan bahwa kombinasi fitur Rata-rata Lama Sekolah, Jumlah Penduduk, dan Bantuan Sosial memberikan pemisahan kelas kemiskinan yang lebih jelas secara visual dibandingkan fitur lain. Kombinasi dengan Belanja Subsidi dan Luas Wilayah tidak efektif karena distribusi data yang tidak merata dan outlier. Pemilihan fitur utama didasarkan pada justifikasi teoritis dan keterbacaan visual yang tinggi, di mana Rata-rata Lama Sekolah mencerminkan kualitas SDM dan Jumlah Penduduk berkaitan dengan distribusi sumber daya serta beban demografis.

3.3.5 Model Klasifikasi Regresi Logistik



GAMBAR XII
CONFUSION MATRIX REGRESI LOGISTIK

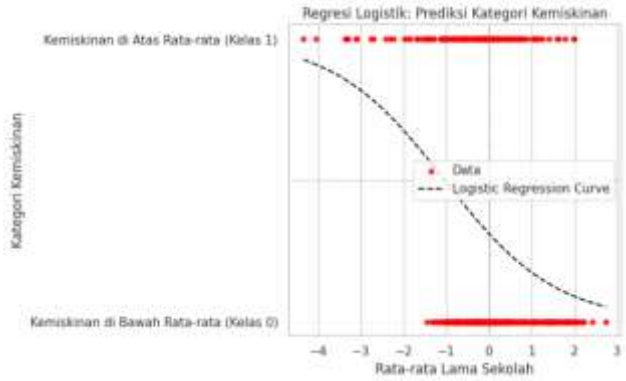
Evaluasi model Regresi Logistik pada data uji menunjukkan performa yang sangat baik, dengan akurasi mencapai 98%. Berdasarkan confusion matrix, model berhasil mengklasifikasikan seluruh 69 data kelas negatif (kelas 0) dengan benar, tanpa kesalahan (true negative sempurna). Untuk kelas positif (kelas 1), sebanyak 32 dari 34 data berhasil diprediksi dengan benar, sementara 2 data diklasifikasikan secara keliru sebagai kelas negatif (false negative). Nilai precision untuk kelas 1 mencapai 1.00, menunjukkan bahwa semua prediksi kelas positif adalah benar. Recall-nya berada pada angka 0.94, yang berarti model mampumendeteksisebagian besar data kelas positif dengan baik, meskipun masih ada beberapa yang terlewat.

Nilai f1-score untuk kedua kelas juga tinggi, yaitu 0.99 untuk kelas negatif dan 0.97 untuk kelas positif, mengindikasikan keseimbangan yang baik antara presisi dan sensitivitas. Secara keseluruhan, model Regresi Logistik menunjukkan kinerja yang sangat andal dalam klasifikasi dua kelas, dengan performa yang sedikit lebih baik dibanding model SVM, khususnya dalam mengenali kelas minoritas (positif). Selain evaluasi performa model, analisis lebih lanjut dilakukan terhadap koefi sien regresi dari setiap variabel prediktor, sebagaimana ditunjukkan pada tabel VI berikut:

TABEL VI
HASIL REGRESI LOGISTIK

Variabel	Coef	Std. Err	z	$P > \|z\|$
const	3.5068	0.828	4.235	0.000
Belanja Subsidi (Miliar)	-0.0104	0.023	-0.443	0.658
Bantuan Sosial (Miliar)	0.0349	0.009	3.814	0.000
Pendapatan Asli Daerah (Miliar)	-0.0027	0.001	-2.541	0.011
Rata-rata Lama Sekolah (Tahun)	-0.4086	0.102	-3.995	0.000
Laju PDRB (Persen)	-0.0274	0.034	-0.801	0.423
Tingkat Pengangguran Terbuka (Persen)	-0.0530	0.069	-0.764	0.445
Luas Wilayah (Km2)	2.539e-06	2.07e-06	1.225	0.220
Jumlah Penduduk (Jiwa)	-4.104e-07	4.47e-07	-0.917	0.359

Berdasarkan hasil regresi logistik, terdapat tiga variabel yang berpengaruh signi f ikan, yaitu Rata-rata Lama Sekolah, Pendapatan Asli Daerah, dan Bantuan Sosial. Rata-rata lama sekolah dan PAD berpengaruh negatif, menunjukkan bahwa pening katan pendidikan dan kapasitas fiskal daerah berperan dalam menurunkan tingkat kemiskinan. Sebaliknya, bantuan sosial berpengaruh positif, mengindikasikan bah wa daerah yang menerima bantuan sosial dalam jumlah besar umumnya adalah daerah dengan tingkat kemiskinan yang lebih tinggi. Variabel lain tidak menunjukk an pengaruh yang signifikan secara statistik.

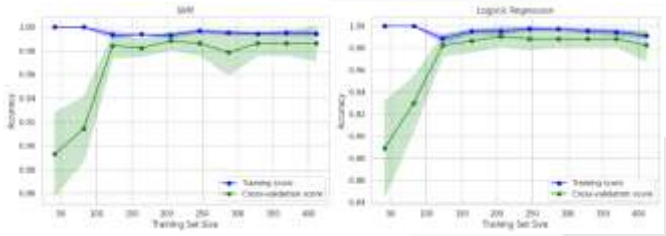


GAMBAR XIII
KURVA REGLOG PREDIKSI KATEGORI KEMISKINAN

Gambar di atas merupakan visualisasi kurva regresi logistik sebagai contoh prediksi kategori kemiskinan berdasarkan salah satu variabel, yaitu **Rata-rata Lama Sekolah**. Titik merah merepresentasikan data aktual, sedangkan garis putus-putus hitam menunjukkan kurva prediksi model regresi logistik. Terlihat bahwa semakin tinggi nilai rata-rata lama sekolah, semakin besar kemungkinan wilayah tersebut masuk ke dalam kategori kemiskinan rendah (kelas 0)**.** Sebaliknya, wilayah dengan rata-rata lama sekolah yang rendah cenderung diklasifikasikan ke dalam kategori kemiskinan tinggi (kelas 1)**.** Ini menunjukkan bahwa pendidikan berperan penting dalam menurunkan kedalaman kemiskinan.

3.3.6 Evaluasi Hasil

Dalam mengevaluasi stabilitas dari dua model, yaitu Support Vector Machine (SVM) dan Regresi Logistik apakah overfitting atau underfitting digunakan visualisasi learning curve seperti Gambar XIV di bawah ini:



GAMBAR XIV
KURVA REGLOG PREDIKSI KATEGORI KEMISKINAN

Gambar XIV menunjukkan kurva pembelajaran SVM dan Regresi Logistik. Kedua model memiliki akurasi pelatihan tinggi (0,98–1,00), namun SVM menunjukkan akurasi validasi yang stabil dan meningkat setelah 150 data pelatihan, menandakan generalisasi yang baik. Sebaliknya, Regresi Logistik menunjukkan fluktuasi pada validasi dan sedikit overfitting, terutama di data menengah. Evaluasi dilakukan dengan StratifiedKFold, memastikan proporsi kelas tetap seimbang dan hasil akurasi lebih andal. Secara keseluruhan, SVM lebih unggul dalam hal stabilitas dan kemampuan generalisasi dibandingkan Regresi Logistik.

4.3 Hasil Analisis Faktor Signifikan

Model regresi logistic ini bertujuan untuk memodelkan hubungan antara variabel kategori biner (Kategori Indeks) dengan beberapa variabel prediktor numerik. Koefisien regresi logistik (β) dapat ditransformasikan menjadi odds ratio (OR) agar lebih mudah diinterpretasikan. Rumus odds ratio adalah sebagai berikut:

$$\text{Odds Ratio} = e^{\beta}, \text{ dengan } e \approx 2{,}718$$

Odds ratio mengukur perubahan peluang kejadian (dalam hal ini: status miskin) untuk setiap peningkatan satu satuan pada variabel independen, dengan asumsi variabel lain konstan. Nilai OR > 1 menunjukkan peningkatan peluang, sedangkan OR menunjukkan penurunan peluang.

Interpretasi Odds Ratio:

1. Belanja Subsidi: Memiliki OR sebesar 0,9896, yang berarti setiap pening katan 1 miliar rupiah pada belanja subsidi menurunkan peluang suatu dae rah masuk ke dalam kategori miskin sebesar sekitar 1,04%. Namun, karena p =0,658 (lebih besar dari 0,05), pengaruh ini tidak signifikan secara statistik.
2. Bantuan Sosial: OR sebesar 1,0355 menunjukkan bahwa setiap kenaikan 1 miliar rupiah dalam bantuan sosial meningkatkan peluang suatu daerah tergolong miskin sebesar 3,55%. Efek ini signifikan secara statistik (p = 0,000), mengindikasikan bahwa bantuan sosial erat kaitannya dengan status kemiskinan daerah.
3. Pendapatan Asli Daerah (PAD): Dengan OR sebesar 0,9973, setiap tam bahan 1 miliar rupiah dalam PAD menurunkan peluang suatu daerah masuk kategori miskin sebesar 0,27%. Meskipun efek ini relatif kecil, hasilnya signifikan secara statistik (p = 0,011 < 0,05).
4. Rata-rata Lama Sekolah: OR sebesar 0,6647 menunjukkan bahwa setiap tambahan 1 tahun rata-rata lama sekolah menurunkan peluang suatu daerah tergolong miskin sebesar 33,53%. Efek ini sangat signifikan secara statis tik (p = 0,000) dan menegaskan pentingnya pendidikan dalam mengurangi kemiskinan.
5. Laju PDRB: Memiliki OR sebesar 0,9729, yang berarti bahwa peningkat an 1% dalam laju PDRB akan menurunkan peluang miskin sebesar 2,71%. Namun, pengaruh ini tidak signifikan secara statistik (p = 0,423).
6. Tingkat Pengangguran Terbuka (TPT): Dengan OR sebesar 0,9483, pe ningkatan 1% pada TPT menurunkan peluang suatu daerah tergolong mis kin sebesar 5,17%, namun efek ini juga tidak signifikan secara statistik (p = 0,445).
7. Luas Wilayah: OR sebesar 1,0000025 mengindikasikan bahwa setiap tam bahan 1 Km22 hanya meningkatkan peluang miskin sebesar

0,00025%, dan pengaruh ini tidak signifikan (p = 0,220).

8. Jumlah Penduduk: Memiliki OR sebesar 0,9999996, artinya setiap tam bahan satu jiwa akan menurunkan peluang miskin sebesar sekitar 0,00004%. Efek ini tidak signifikan secara statistik (p = 0,359) dan dampaknya sangat kecil secara praktis.

V. KESIMPULAN

1. Berdasarkan hasil klasifikasi menggunakan metode Support Vector Machine (SVM) dan Regresi Logistik, diperoleh bahwa kedua model memberikan performa yang sangat baik. Baik model SVM maupun Regresi Logistik menghasilkan akurasi yang sama tinggi, yaitu sebesar 99%. Meskipun akurasinya setara, SVM menunjukkan performa yang lebih stabil berdasarkan evaluasi metrik klasifikasi (precision, recall, dan f1-score) serta kurva pembelajaran yang mencerminkan kemampuan generalisasi yang kuat. Sementara itu, Re gresi Logistik meskipun akurat, menunjukkan fluktuasi yang lebih besar pada validasi silang. Dengan demikian, kedua metode layak digunakan untuk kla sifikasi tingkat kemiskinan, namun SVM lebih unggul dari sisi konsistensi performa pada data validasi.

2. Berdasarkan analisis regresi logistik, terdapat tiga variabel yang signifikan memengaruhi klasifikasi tingkat kedalaman kemiskinan, yaitu Bantuan So sial, Pendapatan Asli Daerah, dan Rata-rata Lama Sekolah. Bantuan Sosial berpengaruh positif, menunjukkan bahwa daerah dengan bantuan sosial le bih besar cenderung masuk kategori kemiskinan tinggi, kemungkinan karena 68 bansos lebih banyak dialokasikan ke daerah miskin. Sementara itu, PAD dan Rata-rata Lama Sekolah berpengaruh negatif, artinya daerah dengan kapasitas f iskal dan tingkat pendidikan yang lebih tinggi cenderung memiliki tingkat kemiskinan yang lebih rendah.

REFERENSI

[1] T. S. Lestari and D. A. N. Sirodj, "Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest," *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 160–167, Feb. 2022, doi: 10.29313/jrs.v1i2.525.

[2] R. Djaenal, J. E. Kaawoan, and I. Rachman, "Implementasi Kebijakan Program Bantuan Pangan Non Tunai (Bpnt) Dinas Sosial Dalam Menanggulangi Kemiskinan Di Kelurahan Tosa Kecamatan Tidore Timur Kota Tidore," *Jurnal Governance*, vol. 1, no. 2, 2021.

[3] N. Ovaliani, G. Ayu Putu Candra Dewi, R. Irmawan, S. Nisa Adiyani, and F. Perdina Waani, "MENAVIGASI KEMISKINAN DI KEPULAUAN: STUDI KASUS DI KEPULAUAN ROMANG, MALUKU BARAT DAYA," vol. 3, no. 1, Jun. 2023, [Online]. Available: https://journal.unhas.ac.id/index.php/DPMR/

[4] I. K. Wardani, Y. Susanti, and S. Subanti, "PEMODELAN INDEKS KEDALAMAN KEMISKINAN DI INDONESIA MENGGUNAKAN ANALISIS REGRESI ROBUST," 2021.

[5] Kemenkeu, "KERANGKA EKONOMI MAKRO DAN POKOK - POKOK KEBIJAKAN FISKAL," 2021. Accessed: Nov. 28, 2024. [Online]. Available: https://fiskal.kemenkeu.go.id/files/kemppkf/file/kem_ppkf_2021.pdf

[6] DPR RI, "Kajian efektivitas beberapa program pengentasan kemiskinan dan pemberdayaan masyarakat dalam APBN (Analisis Ringkas Cepat).," p. 6, 2024, Accessed: Nov. 27, 2024. [Online]. Available: https://berkas.dpr.go.id/pa3kn/analisis-ringkas-cepat/public-file/analisis-ringkas-cepat-public-48.pdf

[7] L. Nuzula, A. Prahutama, A. R. Hakim, D. Statistika, F. Sains, and D. Matematika, "KLASIFIKASI STATUS KEMISKINAN RUMAH TANGGA DENGAN METODE SUPPORT VECTOR MACHINES (SVM) DAN CLASSIFICATION AND REGRESSION TREES (CART) MENGGUNAKAN GUI R (Studi Kasus di Kabupaten Wonosobo Tahun 2018)," *JURNAL GAUSSIAN*, vol. 9, no. 4, pp. 525–534, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/

[8] M. Anjas Aprihartha, "Implementasi Metode Support Vector Machine (SVM) pada Klasifikasi Status Penerima Bantuan Pangan Non Tunai," *JSI : Jurnal Sistem Informasi (E-Journal*, vol. 16, no. 2, 2024.

[9] D. Nurmin, L. Dwi Nurul Khasanah, S. Anggraeni, and D. Andi Nohe, "PENENTUAN KETEPATAN KLASIFIKASI INDEKS KEDALAMAN KEMISKINAN DI INDONESIA DENGAN MODEL LOGIT," 2022.

[10] F. A. Setyowati and I. S. Melati, "Identifikasi Faktor Penyebab Kemiskinan di Ka bupaten Wonosobo Berdasarkan Klasifikasi Perkotaan dan Perdesaan," *Economic Education Analysis Journal*, vol. 9, no. 3, pp. 875–891, 2020, doi: 10.15294/eeaj.v9i3.42413.

[11] S. Osama, H. Shaban, and A. A. Ali, "Gene reduction and machine learning algorithms for cancer classification based on microarray gene expression data: A comprehensive review," Mar. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.eswa.2022.118946.

[12] P. A. Telnoni, Suryatiningsih, and E. Rosely, "Pelabelan Data Dengan Latent Dirichlet Allocation dan K-Means Clustering pada Data Twitter Menggunakan Bahasa Indonesia," *Jurnal Elektro dan Telekomunikasi Terapan*, vol. 7, no. 2, pp. 885–892, Mar. 2021, doi: 10.25124/jett.v7i2.3442.

[13] N. P. N. Hendayanti and M. Nurhidayati, "Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia," 2020.

[14] RAKHMASARI N.M, "IMPLEMENTASI METODE SUPPORT VECTOR MACHINE (SVM)

PADA KLASIFIKASI DAN KARAKTERISASI TINGKAT KEDALAMAN KEMISKINAN PROVINSI JAWA TIMUR," 2022.

[15] A. W. M. Gaffar, A. M. Halis, P. Purnawansyah, and S. R. Jabir, "Penerapan Algoritma Support Vector Machine untuk Klasifikasi Stunting pada Balita di Kabupaten Enrekang," *Jurnal Minfo Polgan*, vol. 13, no. 1, pp. 286–292, Apr. 2024, doi: 10.33395/jmp.v13i1.13620.

[16] P. Fremmuzar and A. Baita, "Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter," *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 57–66, Sep. 2023, doi: 10.34010/komputika.v12i2.9460.

[17] K. Karina, R. Efendi, L. Chairani, and I. M. Sari, "Implementasi Regresi Logistik Ordinal Pada Sistem Pembelajaran Daring Di Era COVID-19 Terhadap Kesehatan Mental Guru SD di Kota Pekanbaru," *Jurnal Sains Matematika dan Statistika*, vol. 7, no. 1, p. 65, Mar. 2021, doi: 10.24014/jsms.v7i1.11786.

[18] D. W. . Hosmer, Stanley. Lemeshow, and R. X. . Sturdivant, *Applied logistic regression*. Wiley, 2013.

[19] Lestari, T. S. & Sirodj, D. A. N. (2022), 'Klasifikasi penipuan transaksi kartu kredit menggunakan metode random forest', Jurnal Riset Statistika 1(2), 160–167.

[20] E. Argarini Pratama and C. M. Hellyana, "PERBANDINGAN 3 ALGORITMA KLASIFIKASI DATA MINING DALAM PRO-KONTRA BAHAYA ROKOK ELEKTRIK," 2022.

[21] H. Apriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," 2020. [Online]. Available: https://journal-computing.org/index.php/journal-ita/index

[22] S. Prusty, S. Patnaik, and S. K. Dash, "SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer," *Frontiers in Nanotechnology*, vol. 4, Aug. 2022, doi: 10.3389/fnano.2022.972421.