

## Klasifikasi Indeks Desa Membangun Desa di Sumatera Utara Menggunakan Metode Support Vector Machine (SVM)

### *Classification of Village Development Index in North Sumatra Using the Support Vector Machine (SVM) Method*

Yayang Arum Kemangi<sup>1</sup>, Daniel Desmanto Sihombing<sup>2</sup>, Permaisuri Siregar<sup>3</sup>, Sella Ujani<sup>4</sup>, Victor Asido Elyakim P<sup>5</sup>  
<sup>1,2,3,4,5</sup>STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara

#### Article Info

##### Genesis Artikel:

Diterima, 23 Juli 2025

Direvisi, 03 Agustus 2025

Disetujui, 15 September 2025

##### Kata Kunci:

Klasifikasi

Indeks Desa Membangun (IDM)

Support Vector Machine (SVM)

Sub-Indeks

Sumatera Utara

#### ABSTRAK

Klasifikasi status Indeks Desa Membangun (IDM) adalah komponen fundamental dalam perumusan kebijakan pembangunan desa yang terarah dan efektif. Namun, proses klasifikasi konvensional seringkali berjalan lambat dan tidak efisien, sehingga mengurangi relevansi data untuk pengambilan keputusan yang dinamis. Penelitian ini bertujuan merancang dan mengevaluasi model klasifikasi otomatis untuk status IDM di 5.417 desa di Provinsi Sumatera Utara menggunakan metode *Support Vector Machine* (SVM). Dengan memanfaatkan data sekunder tahun 2024, model ini menggunakan tiga sub-indeks utama—Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL)—sebagai variabel prediktor untuk memetakan desa ke dalam lima kategori status. Implementasi model SVM dengan *kernel Radial Basis Function* (RBF) dipilih untuk menangani hubungan non-linear yang kompleks antar variabel. Hasil evaluasi pada data uji menunjukkan kinerja yang sangat unggul dengan tingkat akurasi keseluruhan mencapai 96,77%. Kinerja model terbukti sangat kuat, terutama dalam mengidentifikasi kelas 'Berkembang' dengan *recall* sempurna (1.00) dan kelas 'Mandiri' dengan presisi sempurna (1.00). Meskipun ditemukan tantangan minor dalam membedakan kelas-kelas yang berdekatan seperti 'Tertinggal' dan 'Berkembang', tingginya nilai F1-score di semua kelas menegaskan keseimbangan yang baik antara presisi dan *recall*. Penelitian ini menyimpulkan bahwa metode SVM adalah pendekatan yang sangat andal dan valid untuk otomatisasi klasifikasi IDM, serta menawarkan implikasi signifikan sebagai alat pendukung keputusan berbasis bukti yang cepat dan akurat bagi pemerintah daerah.

#### ABSTRACT

The classification of the Village Development Index (IDM) status is a fundamental component in formulating targeted and effective village development policies. However, the conventional classification process is often slow and inefficient, thereby reducing the data's relevance for dynamic decision-making. This research aims to design and evaluate an automatic classification model for the IDM status in 5,417 villages in North Sumatra Province using the Support Vector Machine (SVM) method. By utilizing secondary data from 2024, this model uses three main sub-indices—the Social Resilience Index (IKS), the Economic Resilience Index (IKE), and the Environmental Resilience Index (IKL)—as predictor variables to map villages into five status categories. The implementation of the SVM model with a Radial Basis Function (RBF) kernel was chosen to handle the complex non-linear relationships between variables. The evaluation results on the test data show superior performance, with an overall accuracy rate reaching 96.77%. The model's performance proved to be very strong, particularly in identifying the 'Developing' class with a perfect recall (1.00) and the 'Independent' class with perfect precision (1.00). Although minor challenges were found in distinguishing between adjacent classes such as 'Disadvantaged' and 'Developing', the high F1-score across all classes confirms a good balance between precision and recall. This study concludes that the SVM method is a highly reliable and valid approach for automating IDM classification, and it offers significant implications as a fast and accurate evidence-based decision support tool for local government.

this is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



**Penulis Korespondensi:**

Daniel Desmanto Sihombing,  
Program Studi Sistem Informasi, STIKOM Tunas Bangsa,  
Jl. Jend. Sudirman, Blok A No. 1,2&3, Siantar Barat, Pematang Siantar, Indonesia.  
Email: danielsihombing346@gmail.com

**1. PENDAHULUAN**

Pembangunan desa merupakan agenda prioritas nasional yang menjadi tulang punggung pembangunan berkelanjutan di Indonesia, suatu mandat yang tertuang secara eksplisit dalam Undang-Undang Nomor 6 Tahun 2014 tentang Desa [1, 3]. Kebijakan ini menggarisbawahi urgensi untuk meningkatkan kesejahteraan masyarakat, memajukan perekonomian desa, serta memperkuat masyarakat sebagai subjek pembangunan [2][3]. Dalam rangka memonitor dan mengevaluasi efektivitas kebijakan pembangunan tersebut, pemerintah telah menetapkan Indeks Desa Membangun (IDM) sebagai instrumen pengukuran resmi dan komprehensif [4]. IDM berfungsi sebagai barometer yang tidak hanya memotret kondisi faktual desa, tetapi juga mengarahkan intervensi kebijakan agar lebih tepat sasaran [5]. Melalui kerangka IDM, setiap desa di Indonesia diklasifikasikan ke dalam lima tingkatan status Sangat Tertinggal, Tertinggal, Berkembang, Maju, dan Mandiri—berdasarkan capaian skor kompositnya [5][3]. Klasifikasi ini menjadi krusial karena setiap status merefleksikan kebutuhan dan tantangan yang berbeda, sehingga memerlukan pendekatan pembangunan yang spesifik dan terukur untuk mendorong transformasi menuju kemandirian [5].

Secara konseptual, IDM dibangun di atas tiga pilar fundamental yang merepresentasikan dimensi ketahanan desa secara holistik: Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL) [4][2]. Ketiga sub-indeks ini secara kolektif mencerminkan kemajuan desa dari berbagai aspek esensial. Indeks Ketahanan Sosial (IKS) mengukur kualitas sumber daya manusia dan modal sosial, yang mencakup dimensi kesehatan, pendidikan, dan kohesi sosial masyarakat [6][3]. Indeks Ketahanan Ekonomi (IKE) menilai kapasitas desa dalam menopang kebutuhan ekonominya, yang tercermin dari keragaman produksi, akses terhadap perdagangan dan lembaga keuangan, serta ketersediaan infrastruktur penunjang ekonomi [6][7]. Sementara itu, Indeks Ketahanan Lingkungan (IKL) mengevaluasi kualitas lingkungan hidup serta kesiapsiagaan desa dalam menghadapi potensi bencana alam [3]. Kompleksitas dari puluhan indikator yang membentuk ketiga sub-indeks ini menunjukkan bahwa status sebuah desa merupakan hasil dari interaksi dinamis berbagai faktor, sehingga analisisnya menuntut pendekatan yang cermat dan multidimensional.

Konteks geografis memainkan peran penting dalam implementasi dan analisis IDM. Provinsi Sumatera Utara, sebagai salah satu provinsi dengan jumlah desa yang signifikan dan kondisi sosial-ekonomi yang heterogen, merepresentasikan tantangan tersendiri dalam upaya pemetaan dan akselerasi pembangunan desa. Proses klasifikasi status IDM yang akurat menjadi prasyarat mutlak untuk memastikan bahwa program pembangunan yang dirancang pemerintah provinsi dapat menyentuh akar permasalahan di setiap desa. Namun, metode klasifikasi konvensional yang mengandalkan pengolahan data secara manual sering kali dihadapkan pada kendala efisiensi. Sebuah studi menunjukkan bahwa proses manual semacam ini dapat memakan waktu hingga delapan atau sembilan bulan, suatu rentang waktu yang membuat data hasil klasifikasi menjadi kurang mutakhir dan berpotensi mengurangi ketepatan dalam pengambilan kebijakan yang bersifat dinamis [3]. Keterlambatan dan potensi subjektivitas dalam proses ini menciptakan kebutuhan mendesak akan adanya sebuah sistem klasifikasi yang lebih cepat, objektif, dan dapat diandalkan.

Perkembangan pesat di bidang ilmu komputer, khususnya dalam ranah data mining dan machine learning, menawarkan solusi transformatif untuk mengatasi permasalahan tersebut. Pendekatan ini memungkinkan analisis data dalam skala besar untuk menemukan pola tersembunyi dan membangun model prediktif dengan tingkat akurasi yang tinggi [4]. Relevansi machine learning untuk analisis data IDM telah dieksplorasi secara ekstensif dalam berbagai penelitian sebelumnya. Beragam algoritma telah terbukti mampu memberikan hasil yang menjanjikan, di antaranya adalah Random Forest yang kinerjanya dioptimalisasi melalui teknik ensemble untuk mencapai akurasi tinggi [1], serta Decision Tree dengan berbagai variannya seperti C.45 [4] dan C5.0 [3]. Selain pendekatan klasifikasi, metode unsupervised learning seperti clustering dengan algoritma CLARA juga telah digunakan untuk mengelompokkan desa berdasarkan kesamaan karakteristiknya [6], sementara penelitian lain menggunakan analisis regresi untuk mengidentifikasi faktor-faktor determinan seperti dana desa yang memengaruhi skor IDM [2].

Di tengah beragamnya eksplorasi metodologis tersebut, terdapat sebuah celah penelitian (research gap) yang signifikan terkait penerapan metode Support Vector Machine (SVM). SVM merupakan salah satu algoritma supervised learning yang sangat andal, dikenal luas karena keefektifannya dalam menyelesaikan masalah klasifikasi yang kompleks, kemampuannya beroperasi di ruang berdimensi tinggi, serta fleksibilitasnya dalam memodelkan batas keputusan non-linear melalui kernel trick. Keunggulan-keunggulan ini menjadikan SVM sebagai kandidat metode yang sangat potensial untuk diterapkan pada data IDM yang bersifat multidimensional.

Dengan demikian, penelitian ini bertujuan secara spesifik untuk membangun dan mengevaluasi sebuah model klasifikasi status Indeks Desa Membangun di Provinsi Sumatera Utara menggunakan metode Support Vector Machine, dengan hanya mengandalkan tiga sub-indeks utama (IKS, IKE, dan IKL) sebagai variabel prediktor. Tujuan utamanya adalah untuk mengukur seberapa akurat SVM dapat memetakan status desa dan memvalidasi kinerjanya menggunakan metrik evaluasi standar seperti confusion matrix, akurasi, presisi, recall, dan F1-score. Kontribusi penelitian ini bersifat ganda. Secara teoretis, penelitian ini akan memperkaya literatur ilmiah mengenai aplikasi SVM dalam domain klasifikasi sosial-ekonomi. Secara praktis, model yang dihasilkan berpotensi menjadi sebuah alat bantu pendukung keputusan yang cepat, akurat, dan terukur bagi pemerintah daerah dan pemangku kepentingan lainnya dalam merumuskan strategi pembangunan desa yang lebih efektif dan berbasis bukti.

## 2. METODE PENELITIAN

### 2.1 Jenis Data

Penelitian ini mengadopsi pendekatan kuantitatif dengan memanfaatkan **data sekunder**. Data sekunder merujuk pada rangkaian data yang telah dihimpun, diolah, dan dipublikasikan oleh lembaga atau institusi resmi, bukan dikumpulkan secara primer oleh peneliti. Dalam konteks penelitian ini, data yang digunakan adalah data komprehensif Indeks Desa Membangun (IDM) untuk seluruh desa yang berada dalam cakupan wilayah administratif Provinsi Sumatera Utara untuk tahun 2024.

### 2.2 Sumber Data

Sumber utama data yang digunakan dalam penelitian ini adalah portal Satu Data Indonesia, yang merupakan inisiatif pemerintah untuk menyediakan data publik yang terpusat, terverifikasi, dan dapat diakses secara luas. Dataset spesifik yang menjadi acuan adalah "Data Indeks Desa Membangun Tahun 2024", yang diperoleh melalui *Uniform Resource Locator* (URL) berikut: <https://data.go.id/dataset/dataset/data-indeks-desa-membangun-tahun-2024>. Portal ini menyediakan data yang otoritatif dan terstruktur, yang mencakup informasi detail mengenai Status IDM serta tiga komponen sub-indeks utamanya, yaitu Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL).

### 2.3 Teknik Pengumpulan Data

Proses akuisisi data untuk penelitian ini dilaksanakan melalui beberapa tahapan yang sistematis untuk menjamin validitas dan relevansi data. Prosedur tersebut dapat diuraikan sebagai berikut:

1. **Identifikasi dan Akses Sumber:** Tahap awal adalah mengidentifikasi dan mengakses dataset IDM Tahun 2024 pada portal Satu Data Indonesia yang telah ditentukan.
2. **Verifikasi dan Seleksi Data:** Setelah data berhasil diunduh, dilakukan proses verifikasi untuk memastikan kelengkapan dan integritasnya. Pada tahap ini, dilakukan seleksi atau penyaringan data untuk hanya menyertakan variabel-variabel yang relevan dengan tujuan penelitian, yaitu tiga variabel prediktor (IKS, IKE, IKL) dan satu variabel target (Status IDM) untuk lingkup wilayah Provinsi Sumatera Utara. Proses ini krusial untuk memastikan data yang digunakan bersih dan sesuai untuk tahap pemodelan klasifikasi.

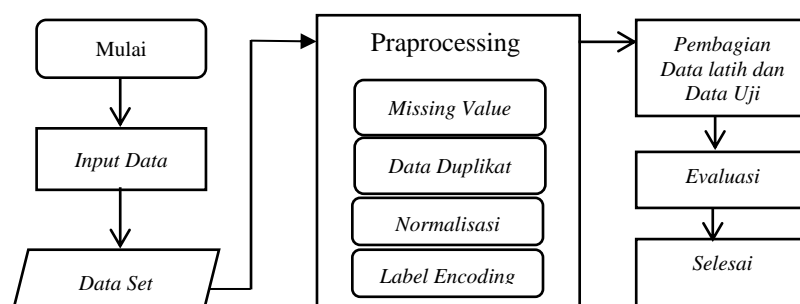
### 2.5 Variabel Penelitian

Penelitian ini didasarkan pada analisis data sekunder, yaitu dataset Indeks Desa Membangun (IDM) untuk Provinsi Sumatera Utara pada tahun 2024. Dataset komprehensif ini mencakup total 5.417 entitas desa. Seluruh tahapan penelitian, mulai dari pra-pemrosesan data, implementasi model Support Vector Machine, hingga evaluasi kinerja, dieksekusi menggunakan bahasa pemrograman Python dalam lingkungan komputasi berbasis cloud, Google Colaboratory. Rincian variabel prediktor dan variabel target yang digunakan dalam pemodelan klasifikasi ini disajikan secara lengkap pada Tabel 1.

Tabel 1. Data Penelitian

Variabel	Keterangan	Pertimbangan Pemilihan Variabel
Kelas (Y)	Mandiri	Kelima kelas yang telah ditetapkan oleh Kementerian Desa, Pembangunan Daerah Tertinggal dan Transmigrasi
	Maju	
	Berkembang	
	Tertinggal	
	Sangat Tertinggal	
IKS (X1)	Indes Ketahanan Sosial	Sub Indeks Penyusunan IDM
IKE (X2)	Indeks Ketahanan Ekonomi	Sub Indeks Penyusunan IDM
IKL (X3)	Indeks Ketahanan Ekologi	Sub Indeks Penyusunan IDM

### 2.5 Tahapan Penelitian



Penelitian ini dilaksanakan melalui serangkaian tahapan yang sistematis dan terstruktur untuk menjamin validitas dan reliabilitas hasil, sebagaimana diilustrasikan dalam diagram alur (flowchart) penelitian. Proses metodologis diawali dengan tahap akuisisi data, di mana dataset Indeks Desa Membangun (IDM) untuk seluruh desa di Provinsi Sumatera Utara diinput sebagai objek utama penelitian. Dataset mentah ini mencakup tiga variabel prediktor utama, yaitu Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL), serta satu variabel target berupa kelas Status IDM. Tahap selanjutnya merupakan fondasi krusial dalam pemodelan *machine learning*, yaitu pra-pemrosesan data, yang bertujuan untuk mempersiapkan dataset agar berkualitas dan sesuai dengan kebutuhan komputasi algoritma Support Vector Machine (SVM).

Tahap pra-pemrosesan data terdiri dari empat langkah esensial. Pertama, dilakukan penanganan terhadap *missing value* untuk memastikan tidak ada data yang hilang atau tidak lengkap yang dapat mengganggu integritas analisis. Kedua, dilakukan identifikasi dan eliminasi data duplikat guna mencegah terjadinya redundansi yang dapat menimbulkan bias pada model. Langkah ketiga adalah normalisasi data. Mengingat SVM merupakan algoritma yang sensitif terhadap skala fitur karena berbasis pada perhitungan jarak, maka proses normalisasi (atau *feature scaling*) menjadi sangat penting. Tahap ini mentransformasikan rentang nilai pada variabel IKS, IKE, dan IKL ke dalam skala yang seragam, sehingga setiap fitur memberikan kontribusi yang setara dalam proses pembentukan *hyperplane* oleh model. Terakhir, dilakukan *label encoding*, yaitu proses transformasi variabel target 'Status IDM' yang bersifat kategorikal (misalnya, teks "Maju", "Berkembang") menjadi representasi numerik yang dapat diinterpretasikan oleh model matematis SVM.

Setelah data berhasil diproses, tahapan selanjutnya adalah pembagian dataset menjadi data latih (*training data*) dan data uji (*testing data*), umumnya dengan proporsi 80:20. Data latih berfungsi sebagai bahan utama untuk "mengajari" model SVM dalam mengenali pola-pola yang membedakan setiap kelas status desa. Sementara itu, data uji, yang tidak pernah dilihat oleh model selama proses pelatihan, digunakan untuk menguji kemampuan generalisasi model. Tahap akhir adalah evaluasi model, di mana model yang telah dilatih digunakan untuk membuat prediksi pada data uji. Kinerja prediktif model kemudian diukur secara kuantitatif menggunakan berbagai metrik evaluasi standar, seperti *confusion matrix*, akurasi, presisi, *recall*, dan *F1-score* [7][1]. Hasil evaluasi ini menjadi tolok ukur utama untuk menyimpulkan efektivitas dan keandalan model SVM dalam mengklasifikasikan status IDM, yang sekaligus menandai selesainya alur penelitian ini.

## 2.6 Support Vector Machine

Metode Support Vector Machine (SVM) dipilih dalam penelitian ini sebagai algoritma klasifikasi utama karena keunggulannya yang telah terbukti dalam menangani permasalahan klasifikasi yang kompleks [7][8]. SVM merupakan salah-satu metode *supervised learning* yang bekerja dengan prinsip fundamental untuk menemukan *hyperplane* atau bidang pemisah optimal yang secara tegas membedakan data ke dalam kelas-kelas yang berbeda. Konsep utama yang membedakan SVM adalah kemampuannya untuk memaksimalkan margin, yaitu jarak antara *hyperplane* dengan titik data terdekat dari setiap kelas (*support vectors*) [9][10]. Dengan memaksimalkan margin ini, model SVM cenderung memiliki kemampuan generalisasi yang lebih superior, sehingga dapat memprediksi data baru yang belum pernah dilihat sebelumnya dengan tingkat keandalan yang tinggi [11][12].

Dalam penelitian ini, metode SVM diimplementasikan untuk melakukan klasifikasi multikelas terhadap status Indeks Desa Membangun (IDM) di Provinsi Sumatera Utara. Model ini dirancang untuk memetakan hubungan non-linear antara tiga sub-indeks utama ketahanan sosial (IKS), ekonomi (IKE), dan lingkungan (IKL) sebagai variabel prediktor dengan lima kelas status desa sebagai target luaran. Mengingat interaksi antar-variabel sosial-ekonomi dan lingkungan sering kali tidak dapat dipisahkan secara linear, maka penelitian ini secara spesifik memanfaatkan fungsi kernel Radial Basis Function (RBF). Penggunaan kernel RBF memungkinkan SVM untuk mentransformasikan data ke dalam ruang berdimensi lebih tinggi, sehingga mampu membangun batas keputusan yang sangat fleksibel dan non-linear. Pendekatan ini dinilai sangat relevan dan berpotensi menghasilkan model klasifikasi yang optimal untuk membedakan nuansa antar-status kemajuan desa yang kompleks.

## 2.7 Evaluasi Model

Evaluasi performa model Support Vector Machine (SVM) yang telah dibangun dalam penelitian ini diukur secara kuantitatif menggunakan metrik-metrik standar yang diturunkan dari *confusion matrix*. Metrik evaluasi utama yang digunakan meliputi akurasi, presisi, *recall*, dan *F1-score* [14]. Keempat metrik ini secara komprehensif digunakan untuk menilai sejauh mana kapabilitas dan keandalan model dalam melakukan klasifikasi status Indeks Desa Membangun (IDM) secara akurat.

### a. Recall

Recall mengukur seberapa besar proporsi data berlabel positif yang berhasil diklasifikasikan dengan benar sebagai positif oleh model. Rumusnya adalah:

$$\text{Recall} = \frac{TP}{TP + FN}$$

### b. Precision

Precision menunjukkan seberapa akurat prediksi positif yang dilakukan oleh model, yaitu proporsi prediksi positif yang benar-benar relevan. Rumusnya adalah:

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP + TN}$$

## c. Accuracy

Accuracy adalah rasio jumlah prediksi yang benar (baik positif maupun negatif) terhadap keseluruhan jumlah data. Rumusnya dituliskan sebagai:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

## d. F1-Score

F1-score merupakan harmonic mean dari precision dan recall. Metode ini berguna untuk menyeimbangkan keduanya, terutama saat distribusi kelas tidak seimbang. Rumusnya:

$$F_1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

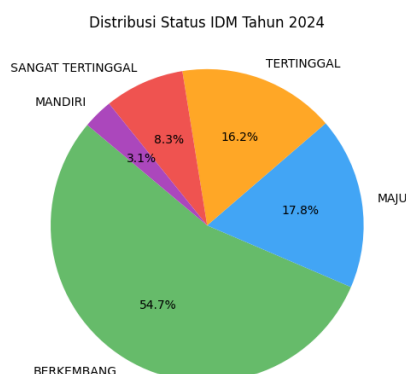
Keterangan:

- *TP*: True Positives (data positif yang diprediksi positif)
- *TN*: True Negatives (data negatif yang diprediksi negatif)
- *FP*: False Positives (data negatif yang diprediksi positif)
- *FN*: False Negatives (data positif yang diprediksi negatif)

### 3. HASIL DAN PEMBAHASAN (10 PT)

#### 3.1 Statistik Deskriptif

Analisis statistik deskriptif diterapkan pada tahap awal penelitian ini dengan tujuan untuk menyajikan ringkasan dan pemahaman mendasar mengenai karakteristik utama dari dataset yang digunakan. Pendekatan ini esensial untuk memberikan gambaran umum data sebelum melangkah ke tahap pemodelan klasifikasi. Variabel dependen (Y) yang menjadi target prediksi dalam penelitian ini adalah Status Indeks Desa Membangun (IDM), yakni sebuah variabel kategorikal yang merepresentasikan tingkatan kemajuan desa di Provinsi Sumatera Utara pada tahun 2024.



Gambar 1. Kategori IDM Provinsi Sumatera Utara 2024

Untuk memberikan gambaran yang lebih mendetail mengenai komposisi data, sebaran jumlah desa di Provinsi Sumatera Utara pada tahun 2024 untuk setiap kategori Status Indeks Desa Membangun (IDM) disajikan secara rinci pada Tabel 2.

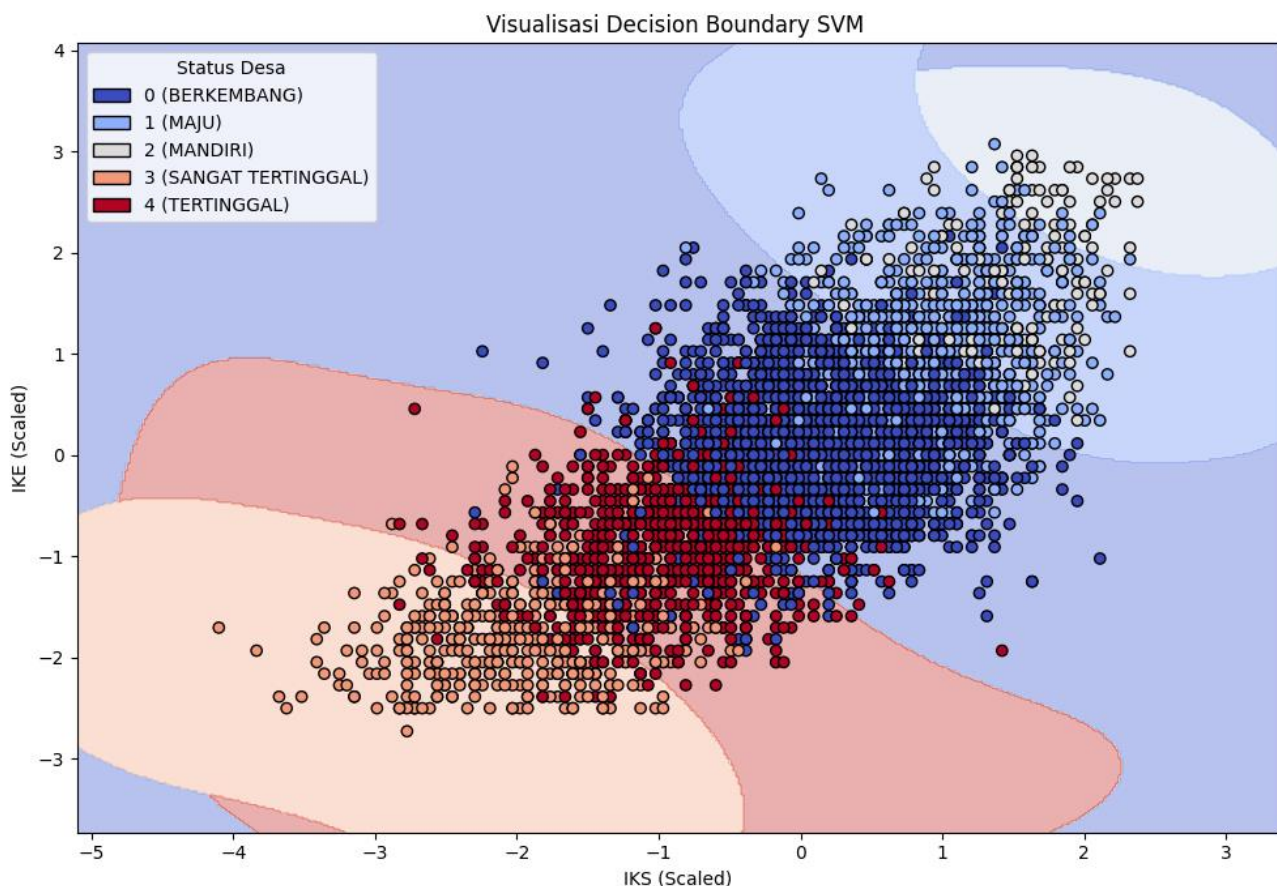
Tabel 2. Jumlah Desa Berdasarkan Kategori IDM per Kabupaten/Kota

Kabupaten/Kota	Berkembang	Maju	Mandiri	Sangat Tertinggal	Tertinggal	Jumlah
ASAHAN	125	45	4	0	3	177
BATU BARA	73	49	19	0	0	141
DAIRI	125	22	0	0	14	161
DELI SERDANG	235	118	27	0	0	380
HUMBANG HASUNDUTAN	125	9	0	0	19	153
KARO	188	46	6	0	19	259
KOTA GUNUNGSITOLI	66	16	1	0	15	98
KOTA PADANG SIDEMPUAN	35	6	0	0	1	42
LABUHANBATU	41	29	5	0	0	75
LABUHANBATU SELATAN	38	14	0	0	0	52
LABUHANBATU UTARA	55	23	0	0	4	82
LANGKAT	90	123	27	0	0	240
MANDAILING NATAL	153	53	6	33	132	377
NIAS	68	5	0	26	71	170

NIAS BARAT	11	0	0	33	61	105
NIAS SELATAN	55	1	0	264	139	459
NIAS UTARA	19	0	0	32	61	112
PADANG LAWAS	210	10	1	2	80	303
PADANG LAWAS UTARA	183	41	3	45	114	386
PAKPAK BHARAT	38	8	0	0	6	52
SAMOSIR	97	14	6	0	11	128
SERDANG BEDAGAI	74	120	43	0	0	237
SIMALUNGUN	285	84	11	0	6	386
TAPANULI SELATAN	130	72	9	0	0	211
TAPANULI TENGAH	89	4	0	9	57	159
TAPANULI UTARA	179	27	0	2	33	241
TOBA	174	24	0	1	32	231
<b>Total</b>	<b>2961</b>	<b>963</b>	<b>168</b>	<b>447</b>	<b>878</b>	<b>5417</b>

Berdasarkan Tabel tersebut, secara umum dapat disimpulkan bahwa status Indeks Desa Membangun (IDM) suatu kabupaten atau kota dipengaruhi oleh dominasi jumlah desa dalam kategori IDM tertentu. Sebagai contoh, Kabupaten Deli Serdang tergolong dalam kategori “Mandiri” karena sebagian besar desa di wilayah tersebut memiliki status IDM kategori “Mandiri” dan “Maju”. Di sisi lain, Kabupaten Nias Selatan masih tergolong dalam kategori “Tertinggal” karena mayoritas desa yang ada berada dalam kategori “Sangat Tertinggal” dan “Tertinggal”.

### 3.2 Boxplot Sub-Indeks IDM Berdasarkan Kategori Status Desa



Gambar 2. Pola Sebaran Data IDM Provinsi Sumatera Utara 2024

Analisis data eksploratif dilakukan melalui visualisasi *boxplot* Gambar 2. untuk memperoleh pemahaman mendasar mengenai karakteristik dan distribusi data dari setiap variabel predictor Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL) terhadap setiap kategori variabel target, yaitu Status IDM. Visualisasi ini krusial untuk menginvestigasi secara kualitatif daya pembeda (discriminatory power) dari masing-masing sub-indeks dan mengidentifikasi pola sebaran data yang akan menjadi dasar bagi pemodelan klasifikasi menggunakan Support Vector Machine (SVM).

Secara keseluruhan, visualisasi pada Gambar 2. secara meyakinkan menunjukkan adanya korelasi positif yang kuat dan konsisten antara nilai ketiga sub-indeks dengan tingkatan Status IDM. Terlihat jelas bahwa terdapat progresi menaik pada nilai median, yang direpresentasikan oleh garis tengah pada setiap kotak, seiring dengan meningkatnya status kemajuan desa dari ‘Sangat Tertinggal’ hingga ‘Mandiri’. Tren positif ini tidak hanya terbatas pada nilai median, tetapi juga berlaku pada keseluruhan rentang interkuartil (IQR) yang digambarkan oleh kotak. Fenomena ini mengindikasikan bahwa semakin tinggi skor ketahanan sosial, ekonomi, dan lingkungan yang dimiliki sebuah desa, semakin tinggi pula probabilitas desa tersebut untuk diklasifikasikan pada status yang lebih maju. Konsistensi pola pada ketiga sub-indeks ini memberikan justifikasi empiris yang kuat bahwa IKS, IKE, dan IKL merupakan fitur-fitur yang sangat relevan dan memiliki kapabilitas prediktif yang signifikan, sehingga penggunaannya sebagai dasar klasifikasi dalam penelitian ini sangat valid.

Meskipun tren progresi tersebut sangat jelas, analisis yang lebih mendalam pada *boxplot* juga menyingkap kompleksitas yang inheren dalam data. Teridentifikasi adanya tumpang tindih (*overlap*) sebaran data yang cukup signifikan, terutama pada kategori-kategori status IDM yang saling berdekatan. Sebagai contoh, rentang data dan bahkan sebagian dari rentang interkuartil antara kelas ‘Tertinggal’ dan ‘Berkembang’, serta antara ‘Berkembang’ dan ‘Maju’, saling bersinggungan. Tumpang tindih ini merepresentasikan sebuah “wilayah abu-abu” di mana desa-desa yang berada pada tahap transisi memiliki karakteristik sub-indeks yang sangat mirip. Secara praktis, hal ini menjelaskan mengapa tugas klasifikasi ini tidak trivial dan mengapa sebuah model dapat mengalami kesulitan dalam membedakan kelas-kelas yang bersebelahan, sebuah fenomena yang biasanya juga tercermin pada *confusion matrix* sebagai kesalahan klasifikasi.

Jika diamati secara komparatif, sebaran data pada setiap sub-indeks memberikan wawasan tambahan. Distribusi nilai untuk kategori ‘Berkembang’ cenderung memiliki rentang interkuartil yang lebih lebar dibandingkan kategori lainnya pada ketiga sub-indeks. Hal ini dapat diinterpretasikan bahwa desa-desa dalam status ‘Berkembang’ memiliki tingkat heterogenitas kondisi sosial, ekonomi, dan lingkungan yang paling tinggi. Sebaliknya, kategori ‘Mandiri’ dan ‘Sangat Tertinggal’ cenderung menunjukkan sebaran data yang lebih rapat, mengindikasikan karakteristik yang lebih homogen pada kedua ujung spektrum pembangunan tersebut.

Berdasarkan analisis visual ini, dapat disimpulkan bahwa meskipun ketiga sub-indeks merupakan prediktor yang kuat, hubungan antara fitur-fitur tersebut dengan kelas Status IDM bersifat non-linear dan kompleks. Adanya tumpang tindih data antar kelas yang berdekatan meniadakan kemungkinan adanya pemisah linear yang sederhana. Oleh karena itu, karakteristik data ini semakin memperkuat justifikasi pemilihan metode Support Vector Machine (SVM) dengan kernel Radial Basis Function (RBF). Metode ini secara teoretis dirancang untuk mampu menangani batas keputusan yang rumit dan non-linear, sehingga sangat sesuai untuk memodelkan dan memisahkan kelas-kelas dalam dataset IDM yang kompleks ini secara efektif.

### 3.3 Evaluasi Model

Evaluasi kinerja model klasifikasi Support Vector Machine (SVM) dilakukan secara kuantitatif terhadap data uji untuk mengukur kapabilitas dan keandalannya dalam memprediksi status Indeks Desa Membangun (IDM). Secara keseluruhan, model yang diusulkan menunjukkan tingkat kinerja yang sangat tinggi dan meyakinkan, dengan keberhasilan mencapai akurasi sebesar 96,77%. Capaian ini menjadi indikator utama bahwa model memiliki kemampuan generalisasi yang solid dan secara efektif mampu memetakan hubungan kompleks antara sub-indeks IKS, IKE, dan IKL dengan status kemajuan desa. Hasil evaluasi yang lebih terperinci, mencakup metrik *precision*, *recall*, dan *f1-score* untuk setiap kelas disajikan pada Tabel 3.

Tabel 3. Kinerja Klasifikasi Model SVM

Kelas	Precision	Recall	F1-Score	Support
BERKEMBANG	0.96	1.00	0.98	592
MAJU	0.98	0.93	0.95	193
MANDIRI	1.00	0.94	0.97	34
SANGAT TERTINGGAL	0.94	0.98	0.96	89
TERTINGGAL	0.99	0.91	0.95	176
<b>Accuracy</b>			<b>0.9677</b>	1084
<b>Macro Average</b>	0.97	0.95	0.96	
<b>Weighted Average</b>	0.97	0.97	0.97	

Kinerja model SVM yang berhasil mencapai akurasi 96,77% dalam penelitian ini menunjukkan tingkat efektivitas yang sangat tinggi. Apabila dibandingkan dengan temuan dari studi-studi sebelumnya yang menggunakan metode klasifikasi berbeda pada data IDM, hasil ini dapat dikatakan sangat kompetitif dan bahkan unggul. Sebagai contoh, penelitian yang menerapkan algoritma berbasis pohon keputusan seperti C4.5[4] dan C5.0 [3] pada data IDM di wilayah lain melaporkan tingkat akurasi yang lebih rendah. Demikian pula, pendekatan klasifikasi menggunakan jaringan Bayesian (*Bayesian Network*) seperti yang dilakukan oleh Utami & Wijayanto [15] juga menunjukkan bahwa setiap algoritma memiliki karakteristik kinerjanya masing-masing.

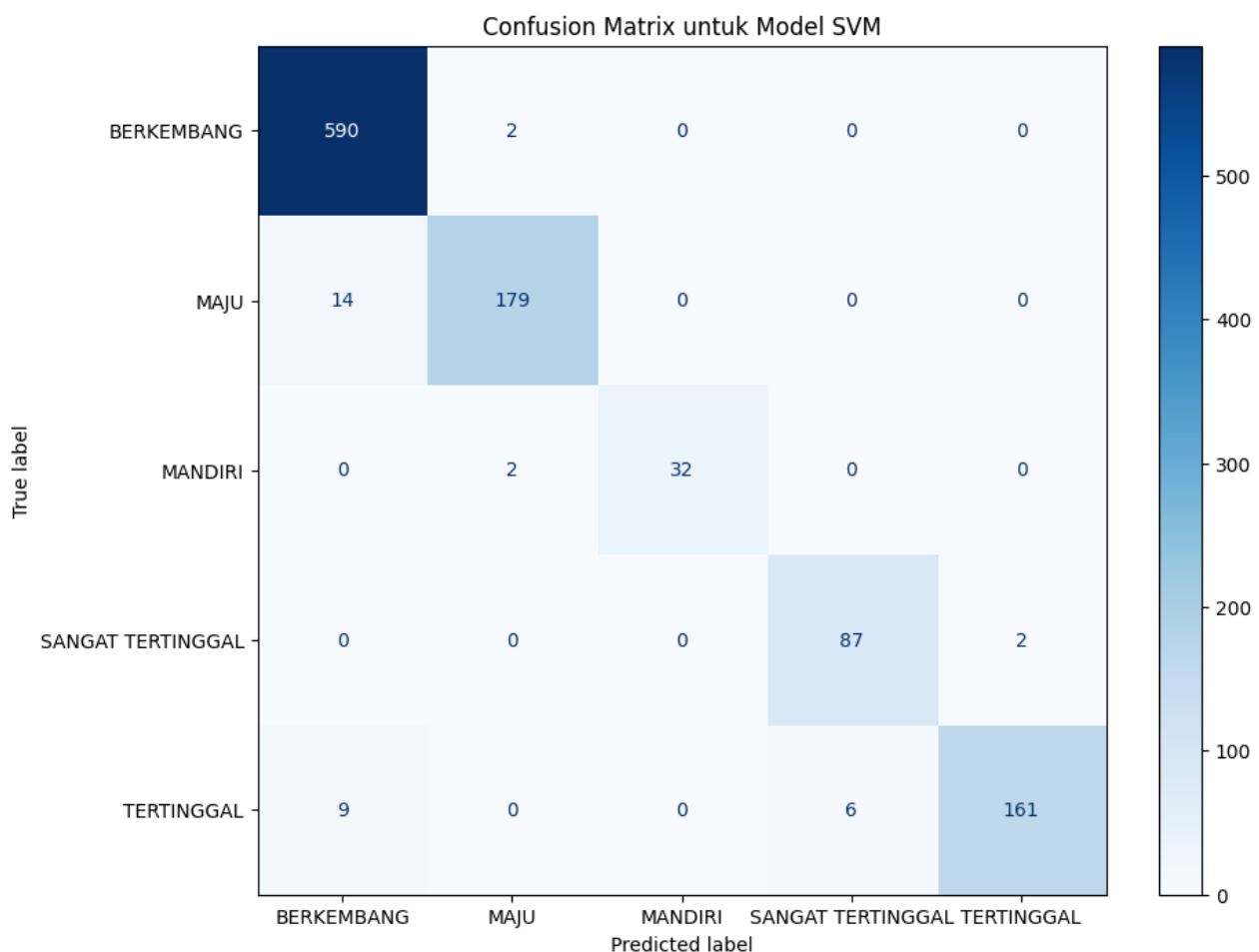
Hasil evaluasi kuantitatif terhadap kinerja model memperlihatkan kapabilitas superior sekaligus tantangan spesifik dalam proses pengenalan setiap kelas status desa. Keunggulan utama model secara signifikan termanifestasi pada kemampuannya dalam mengidentifikasi kelas ‘BERKEMBANG’, yang berhasil mencapai nilai *recall* sempurna (1.00). Capaian ini secara definitif



mengindikasikan bahwa model memiliki sensitivitas yang maksimal terhadap kelas ini, di mana seluruh desa yang secara faktual berstatus ‘Berkembang’ di dalam data uji berhasil diidentifikasi dengan benar tanpa satu pun kasus *False Negative*. Kinerja yang luar biasa ini dapat diatribusikan pada representasi data yang kuat untuk kelas ‘BERKEMBANG’, yang memiliki jumlah sampel terbesar, sehingga memungkinkan model SVM untuk mempelajari dan menggeneralisasi polanya secara ekstensif.

Di sisi lain, model juga menunjukkan tingkat keandalan prediksi yang absolut pada kelas ‘MANDIRI’, yang dibuktikan dengan pencapaian nilai presisi sempurna (1.00). Nilai ini menegaskan bahwa setiap desa yang diklasifikasikan oleh model sebagai ‘Mandiri’ dapat dipastikan 100% benar, tanpa adanya kasus *False Positive*. Tingkat kepastian yang tinggi ini memiliki nilai praktis yang sangat berharga, terutama untuk keperluan validasi kebijakan dan pemetaan desa-desa yang telah berhasil mencapai puncak spektrum pembangunan. Kombinasi dari presisi dan *recall* yang tinggi pada berbagai kelas menunjukkan bahwa model tidak hanya akurat, tetapi juga seimbang dalam kinerjanya.

Meskipun demikian, analisis yang lebih terperinci pada *confusion matrix* Gambar 3. juga menyingkap adanya tantangan klasifikasi yang inheren, terutama dalam membedakan kelas-kelas yang secara konseptual berdekatan pada spektrum pembangunan. Titik lemah utama model teridentifikasi pada nilai *recall* untuk kelas ‘TERTINGGAL’, yang tercatat sebesar 0.91. Angka ini merupakan nilai *recall* terendah di antara semua kelas dan menyiratkan bahwa model gagal mengenali sekitar 9% dari total desa yang seharusnya berstatus ‘Tertinggal’. Pengamatan pada *confusion matrix* mengonfirmasi bahwa mayoritas dari kesalahan klasifikasi ini terjadi karena model keliru mengidentifikasi desa-desa tersebut sebagai ‘BERKEMBANG’. Fenomena "tertukar" antar kelas yang bersebelahan ini dapat dijelaskan secara logis, karena desa yang berada pada ambang batas transisi sering kali memiliki profil sub-indeks (IKS, IKE, dan IKL) yang sangat mirip, sehingga menciptakan ambiguitas pada batas keputusan yang sulit dipisahkan secara sempurna oleh *hyperplane* model.



Gambar 3. Confusion Matrix

Berdasarkan *confusion matrix*, nilai *recall* yang lebih rendah ini disebabkan karena model cenderung keliru mengklasifikasikan sejumlah desa ‘Tertinggal’ sebagai ‘Berkembang’. Fenomena saling tertukar antara dua kelas ini sangat dapat dipahami, mengingat desa-desa yang berada pada ambang batas transisi dari status ‘Tertinggal’ ke ‘Berkembang’ sering kali memiliki karakteristik sub-indeks yang sangat mirip. Ambiguitas inheren pada perbatasan kelas inilah yang menjadi tantangan utama bagi algoritma untuk menarik garis pemisah yang sempurna. Hal serupa juga teridentifikasi pada kelas ‘MAJU’ (*recall* 0.93), di mana beberapa desa ‘Maju’ juga keliru diprediksi sebagai ‘Berkembang’.



Sebagai kesimpulan evaluatif, model Support Vector Machine telah menunjukkan kinerja yang sangat andal dan kuat untuk tugas klasifikasi status IDM. Tingginya nilai F1-score di semua kelas (semuanya  $\geq 0.95$ ) menegaskan bahwa model mampu menjaga keseimbangan yang sangat baik antara presisi dan *recall*. Meskipun terdapat sedikit kesalahan klasifikasi yang logis pada kelas-kelas yang berdekatan, akurasi keseluruhan yang mencapai 96,77% membuktikan bahwa model ini merupakan alat bantu prediktif yang valid dan sangat cocok untuk memetakan status kemajuan desa di Sumatera Utara berdasarkan sub-indeks IDM.

Keunggulan signifikan yang ditunjukkan oleh model SVM dalam penelitian ini dapat diatribusikan pada kemampuannya dalam menangani batas keputusan non-linear yang kompleks melalui kernel RBF. Sementara itu, penelitian lain dalam domain data kewilayahan lebih berfokus pada pendekatan *unsupervised learning* seperti *clustering* untuk mengelompokkan desa berdasarkan fasilitas kesehatan [16] atau data kemiskinan [17]. Meskipun tidak dapat dibandingkan secara langsung dari sisi akurasi, studi-studi tersebut bersama dengan penelitian analitik lainnya [18][19][20] mengonfirmasi tingginya relevansi penerapan teknik *data mining* untuk analisis pembangunan daerah. Oleh karena itu, capaian akurasi model SVM dalam penelitian ini tidak hanya memvalidasi metode yang dipilih, tetapi juga menegaskan posisinya sebagai pendekatan yang sangat andal untuk klasifikasi status IDM.

#### 4. KESIMPULAN

Penelitian ini secara komprehensif telah mengimplementasikan dan mengevaluasi metode Support Vector Machine (SVM) untuk tugas klasifikasi status Indeks Desa Membangun (IDM) di Provinsi Sumatera Utara. Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa metode SVM dengan kernel Radial Basis Function (RBF) terbukti sangat efektif dan memiliki tingkat keandalan yang tinggi untuk memetakan status kemajuan desa berdasarkan tiga sub-indeks utama: Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL). Temuan utama dari penelitian ini adalah keberhasilan model dalam mencapai akurasi keseluruhan sebesar 96,77% pada data uji, sebuah capaian yang mengonfirmasi kapabilitas generalisasi model yang sangat baik. Kinerja unggul ini didukung oleh nilai presisi, *recall*, dan F1-score yang solid di seluruh lima kelas status, yang menandakan bahwa model mampu menjaga keseimbangan antara ketepatan prediksi dan sensitivitas pengenalan kelas. Meskipun teridentifikasi adanya tantangan minor dalam membedakan kelas-kelas yang secara konseptual berdekatan, seperti 'Tertinggal' dan 'Berkembang', fenomena ini merupakan cerminan dari kompleksitas inheren data pembangunan dan bukan merupakan kelemahan fatal dari model.

Implikasi dari hasil penelitian ini sangat signifikan bagi proses pemetaan dan perencanaan pembangunan desa di Sumatera Utara. Tingkat akurasi yang tinggi menunjukkan bahwa klasifikasi status IDM dapat dilakukan secara otomatis, objektif, dan jauh lebih efisien dibandingkan dengan metode konvensional. Hal ini memberikan sebuah alat bantu yang valid bagi pemerintah daerah dan pemangku kepentingan untuk melakukan identifikasi dan pemantauan status desa secara dinamis, sehingga alokasi sumber daya dan perancangan program intervensi dapat menjadi lebih tepat sasaran. Kontribusi ilmiah utama dari penelitian ini terletak pada pembuktian empiris mengenai superioritas kinerja SVM dalam domain klasifikasi data sosial-ekonomi yang kompleks seperti IDM, sekaligus memperkaya literatur dengan menyajikan sebuah studi kasus yang solid mengenai penerapan machine learning untuk analisis pembangunan regional.

Keberhasilan penelitian ini juga membuka berbagai prospek dan arah studi lanjutan yang menjanjikan. Dari aspek metodologis, penelitian di masa depan dapat berfokus pada studi komparatif dengan algoritma machine learning lain seperti Gradient Boosting atau Jaringan Saraf Tiruan (Artificial Neural Networks) untuk menetapkan tolok ukur kinerja yang lebih definitif. Selain itu, optimisasi hiperparameter secara sistematis pada model SVM yang ada berpotensi untuk lebih mempertajam akurasi dan mengatasi ambiguitas pada batas-batas keputusan antar kelas. Dari sisi data, eksplorasi dengan penambahan variabel prediktor di luar tiga sub-indeks utama, seperti data demografi atau aksesibilitas geografis, dapat diselidiki untuk melihat dampaknya terhadap daya prediksi model. Pada akhirnya, prospek pengembangan paling relevan adalah mentransformasikan model yang telah tervalidasi ini menjadi sebuah Sistem Pendukung Keputusan (SPK) yang fungsional, yang dapat diadopsi oleh lembaga pemerintah untuk menjembatani antara riset akademik dan implementasi kebijakan berbasis data.

#### REFERENSI

- [1] A. Irma Seska Arina, V. Masinambow, and E. N. Walewangko, "PENGARUH DANA DESA DAN ALOKASI DANA DESA TERHADAP INDEKS DESA MEMBANGUN DI KABUPATEN MINAHASA TENGGARA," 2021.
- [2] M. C. Nur Fitria, N. N. Debatara, and S. W. Rizki, "Classification of Village Status in Landak Regency Using C5.0 Algorithm," *Tensor: Pure and Applied Mathematics Journal*, vol. 3, no. 1, pp. 33–42, Jun. 2022, doi: 10.30598/tensorvol3iss1pp33-42.
- [3] A. C. Adha, A. Marzuki, Y. S. Nelaz, S. H. Hendriani, and N. Purnomo, "Penerapan Data Mining Menggunakan Metode Klasifikasi Dalam Memprediksi Status Desa Berdasarkan Indeks Desa Membangun," *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1782–1788, Nov. 2024, doi: 10.33395/jmp.v13i2.14250.
- [4] A. Rifai, S. E. Permana, and R. Hamonangan, "OPTIMALISASI KLASIFIKASI INDEKS DESA MEMBANGUN MENGGUNAKAN METODE ENSEMBLE DAN ALGORITMA RANDOM FOREST," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, pp. 8226–8234, 2024.
- [5] M. R. Hendriawan and R. R. Marlina, "Pengelompokan Desa di Jawa Barat Berdasarkan Indeks Desa Membangun (IDM) Menggunakan ALgoritma Clustering Large Application (CLARA)," *JAMBURA JOURNAL OF PROBABILITY AND STATISTICS*, vol. 6, no. 1, pp. 35–41, 2025, doi: 10.34312/jjps.v4i1.27450.

- [6] A. N. Astika and N. S. Subawa, "Evaluasi Pembangunan Desa Berdasarkan Indeks Desa Membangun," *JURNAL ILMIAH MUQODDIMAH: Jurnal Ilmu Sosial, Politik Dan Humaniora*, vol. 5, no. 2, pp. 223–232, 2021, [Online]. Available: <http://jurnal.um-tapsel.ac.id/index.php/muqoddimah>
- [7] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
- [8] M. Rahardi, A. Aminuddin, F. F. Abdulloh, and R. A. Nugroho, "Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia," (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 534–539, 2022, doi: 10.14569/IJACSA.2022.0130665.
- [9] K. X. Han, W. Chien, C. C. Chiu, and Y. T. Cheng, "Application of support vector machine (SVM) in the sentiment analysis of twitter dataset," *Applied Sciences (Switzerland)*, vol. 10, no. 3, Feb. 2020, doi: 10.3390/app10031125.
- [10] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, pp. 150–55, 2021, doi: 10.30591/jpit.v6i3.2870.
- [11] Z. Alhaq, A. Mustopa, S. Mulyatun, and J. D. Santoso, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," *JOISM: JURNAL OF INFORMATION SYSTEM MANAGEMENT*, vol. 3, no. 1, pp. 16–1, 2021, doi: 10.24076/joism.2021v3i2.558.
- [12] V. Fitriyana *et al.*, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," 2023.
- [13] M. Hadi Arfian *et al.*, "Analisis Sentimen Pada Media Sosial Menggunakan Metode Support Vector Machine," *Jurnal Ilmu Teknik dan Komputer*, vol. 09, no. 01, pp. 1–6, 2025, doi: 10.22441/jitkom.v9i1.001.
- [14] F. Latuconsina, M. S. Noya van Delsen, and Yudistira, "Klasifikasi Menggunakan Metode Support Vector Machine (SVM) Multiclass pada Data Indeks Desa Membangun (IDM) di Provinsi Maluku," *Journal of Mathematics, Computations and Statistics*, vol. 7, no. 2, pp. 380–395, Oct. 2024, doi: 10.35580/jmathcos.v7i2.3624.
- [15] N. A. Utami and A. W. Wijayanto, "Classification of Village Development Index at Regency/Municipality Level Using Bayesian Network Approach with K-Means Discretization," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 12, no. 3, pp. 95–106, 2022.
- [16] A. Nikmah, C. Nisa, and M. Riefky, "Penerapan Algoritma K-Medoids untuk Pengelompokan Provinsi di Indonesia Berdasarkan Status Gizi Anak Balita," *Emerging Statistics and Data Science Journal*, vol. 3, no. 1, pp. 516–524, 2025.
- [17] K. Setiawan, Kastum, and Y. P. Pratama, "K-Means Clustering Analysis of Poverty Data in Cilacap District," *International Journal Software Engineering and Computer Science (IJSECS)*, vol. 5, no. 1, pp. 53–62, Apr. 2025, doi: 10.35870/ijsecs.v5i1.3759.
- [18] I. M. Parapat, M. T. Furqon, and Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, pp. 3163–3169, 2018.
- [19] A. S. Harahap and P. Zulvia, "Klasterisasi Desa dengan Menggunakan Algoritma K-Means pada Data Potensi Desa," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, p. 237, Dec. 2021, doi: 10.30865/jurikom.v8i6.3724.
- [20] D. T. Bachruddin and B. A. Darma, "Analisis Pembangunan Desa Berdasarkan Capaian Indeks Desa Membangun di Kabupaten Serang," *Jurnal Analis Kebijakan*, vol. 4, no. 1, pp. 1–12, 2020.