Klasifikasi Tingkat Kemiskinan Tangerang Selatan Menggunakan Model Support Vector Machine dan K-Nearest Neighbours

Anggriyani Daulay[1], Taswanda Taryo[2], Agung Budi Susanto[3]
Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang, Tangerang Selatan, Banten
e-mail: 106.anggriyani@gmail.com[1], dosen02234@unpam.ac.id[2],dosen02680@unpam.ac.id[3]

Abstrak--Kemiskinan merupakan fenomena multidimensional yang memerlukan pemetaan akurat sebagai dasar perumusan kebijakan publik yang tepat sasaran,. Dinas Sosial Kota Tangerang Selatan mengelola data verifikasi skala besar yang membutuhkan pengolahan objektif untuk menghindari inclusion dan exclusion error,. Penelitian ini bertujuan melakukan analisis komparatif performa algoritma Support Vector Machine (SVM) dan K-Nearest Neighbours (KNN) menggunakan dataset operasional hasil verifikasi sosial-ekonomi masyarakat Kota Tangerang Selatan Tahun 2024 berjumlah 5.644 data,. Metodologi penelitian mencakup tahap preprocessing sistematis melalui data cleaning, label encoding, StandardScaler, serta penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE),. Hasil eksperimen menunjukkan performa yang sangat superior dari kedua model, di mana algoritma SVM dengan kernel Radial Basis Function (RBF) mencapai tingkat akurasi global sebesar 99%, mengungguli KNN yang memperoleh akurasi 97%„. Temuan ini mengonfirmasi bahwa model komputasi yang dibangun sangat valid untuk diintegrasikan sebagai Decision Support System (DSS) guna otomatisasi validasi status kemiskinan secara transparan dan objektif di lingkungan pemerintah daerah.

***Kata Kunci — KNN, Klasifikasi Kemiskinan, Machine Learning, SVM, Tangerang Selatan*.**

I. PENDAHULUAN

Pengukuran tingkat kemiskinan tidak dapat lagi bertumpu pada indikator ekonomi tunggal, melainkan harus mengintegrasikan berbagai variabel sosial-ekonomi untuk menggambarkan profil kesejahteraan masyarakat secara utuh,. Karakteristik masyarakat perkotaan yang heterogen seperti di Kota Tangerang Selatan sering kali mengaburkan batas antar kelompok kesejahteraan, yang berisiko menimbulkan kesalahan penempatan (*misclassification error*) dalam penyaluran bantuan sosial,.

Pemanfaatan *Machine Learning* menawarkan solusi untuk mendukung pengambilan keputusan berbasis data (*data-driven decision making*),. Algoritma SVM dan KNN dipilih karena rekam jejak performanya yang tinggi pada data sosial-ekonomi, namun efektivitasnya sangat bergantung pada karakteristik data lokal,. Kebaruan penelitian ini terletak pada penggunaan data operasional mikro hasil verifikasi lapangan, bukan sekadar dataset publik berskala makro, guna menghasilkan model dengan *local relevance* yang tinggi bagi kebijakan daerah,

.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Data diperoleh dari Dinas Sosial Kota Tangerang Selatan tahun 2024, terdiri dari 5.644 observasi individu,. Fitur utama yang digunakan meliputi Umur, Jenis Kelamin, Kecamatan, Kelurahan, dan Skor Total Verifikasi,. Atribut identitas pribadi (Nama, NIK, KK) dieliminasi untuk menjaga kerahasiaan data dan menghindari bias model,.

B. Tahapan Pra-Pemrosesan Data Kualitas data ditingkatkan melalui beberapa fase:

- Data Cleaning: Sterilisasi dari nilai kosong dan baris duplikat,.
- Label Encoding: Mentransformasikan variabel kategorikal (Jenis Kelamin, Kecamatan, Kelurahan) menjadi representasi numerik integer,.
- Normalisasi: Menggunakan StandardScaler (Z-Score) untuk menyamakan skala antarvariabel agar fitur dengan rentang besar tidak mendominasi perhitungan jarak,.
- Handling Imbalance: Menerapkan teknik SMOTE pada data latih untuk menyeimbangkan distribusi kelas target sehingga

model tidak bias pada kelas mayoritas.

C. Skenario Pemodelan Dataset dipisahkan menggunakan rasio 80% (4.515 data) untuk pelatihan model dan 20% (1.129 data) untuk pengujian generalisasi pada data baru (unseen data),. Model SVM dikonfigurasi dengan kernel RBF, sementara KNN menggunakan parameter k=5 dengan metrik jarak Euclidean Distance.

D. Identifikasi Variabel Multidimensi Baru

Untuk meningkatkan kedalaman analisis dan akurasi prediksi, Anda dapat mengintegrasikan indikator-indikator non-moneter yang meliputi:

a) **Kondisi Fisik Bangunan:** Menambahkan variabel seperti jenis atap, jenis lantai, dan jenis dinding tempat tinggal.
b) **Kepemilikan Aset:** Mengintegrasikan data mengenai kepemilikan aset bergerak.
c) **Konsumsi Energi:** Menambahkan besaran daya listrik rumah tangga sebagai salah satu indikator ekonomi.
d) **Aksesibilitas Layanan Dasar:** Menambahkan variabel akses terhadap fasilitas sanitasi, layanan kesehatan, dan pendidikan yang layak.

E. Integrasi ke dalam Pipeline Preprocessing

Setelah variabel baru dikumpulkan, variabel tersebut harus melalui alur pra-pemrosesan data yang sama agar dapat diproses oleh algoritma SVM dan KNN:

- **Data Cleaning:** Melakukan sterilisasi terhadap *missing value* atau data ganda pada variabel baru tersebut.
- **Data Encoding:** Jika variabel baru bersifat kategorikal (seperti jenis lantai atau atap), variabel tersebut harus ditransformasikan menjadi angka menggunakan metode **Label Encoding**.
- **Normalisasi (StandardScaler):** Variabel baru harus diskalakan menggunakan *StandardScaler* (Z-Score) agar fitur dengan rentang nilai besar (misalnya daya listrik dalam Watt) tidak mendominasi fitur lainnya dalam perhitungan jarak KNN maupun pencarian *hyperplane* SVM.
- **Feature Selection:** Melakukan seleksi fitur berbasis relevansi teoretis untuk memastikan variabel tersebut memiliki korelasi struktural yang kuat terhadap pembentukan klaster kesejahteraan.

F. Pelatihan Ulang Model

Setelah fitur multidimensi ditambahkan ke dalam matriks data, model perlu dilatih ulang menggunakan subset data latih (80%) dan diuji kembali menggunakan data uji (20%). Penggunaan variabel yang lebih beragam ini diproyeksikan dapat memperkuat stabilitas model dalam menangani tantangan volume data yang lebih besar dan berderau (*noise*) tinggi di masa depan.

Penambahan variabel-variabel ini sangat penting karena kemiskinan di wilayah urban seperti Tangerang Selatan tidak hanya ditentukan secara linier oleh pendapatan moneter, melainkan sangat berkelindan dengan variabel laten lainnya yang membentuk profil sosio-ekonomi riil masyarakat.

III HASIL DAN DISKUSI

A. Analisis Performa SVM Berdasarkan pengujian, model SVM berbasis RBF menunjukkan performa yang sangat superior dengan skor akurasi, presisi, *recall*, dan *F1-score* yang stabil di angka 99%,. Penggunaan kernel RBF efektif memetakan fitur demografi yang awalnya tumpang tindih ke dalam ruang dimensi tinggi, sehingga bidang pemisah linier (*hyperplane*) dapat dibentuk dengan margin maksimal,.

B. Analisis Performa KNN Model KNN membukukan skor akurasi global sebesar 97%,. Keberhasilan ini membuktikan adanya lokalitas spasial yang kuat pada data demografi masyarakat Tangerang Selatan, di mana individu dengan status ekonomi serupa

cenderung berkumpul membentuk pola nilai yang mirip,.

C. Perbandingan Komparatif Tabel I menyajikan perbandingan performa kedua algoritma:

Tabel I. Perbandingan Hasil SVM dan KNN | Algoritma | Accuracy | Precision | Recall | F1-Score | | :--- | :---: | :---: | :---: | :---: | | **SVM** | **0.99** | **0.99** | **0.99** | **0.99** | | **KNN** | **0.97** | **0.97** | **0.97** | **0.97** |

Superioritas SVM (99%) atas KNN (97%) disebabkan oleh mekanisme *eager learner* pada SVM yang fokus mengunci posisi *support vectors* krusial di perbatasan kelas, menjadikannya lebih kebal terhadap *noise* dan pencilan dibandingkan KNN yang bersifat *lazy learner*,. Uji sensitivitas juga menunjukkan stabilitas akurasi SVM yang tetap tangguh pada rentang 98-99% meskipun rasio data latih diubah,.

D. Implikasi Manajerial dan Kebijakan Akurasi tinggi ini memungkinkan model diinstansiasi sebagai gerbang validasi otomatis (*automated screening tool*) untuk mengeliminasi potensi nepotisme dan bias subjektif petugas lapangan,,,. Pemerintah daerah dapat mengambil langkah proaktif melalui kebijakan preventif pada klaster warga yang terdeteksi masuk zona "Rentan" sebelum mereka benar-benar jatuh ke garis kemiskinan,

Hasil pengujian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi dengan performa yang sangat tinggi. Perbandingan kinerja model secara detail disajikan pada Tabel 1.

Tabel 1. Perbandingan Hasil Klasifikasi SVM dan KNN

Metode	Accuracy	Precision	Recall
SVM	**0.99**	**0.99**	**0.99**
KNN	0.97	0.97	0.97

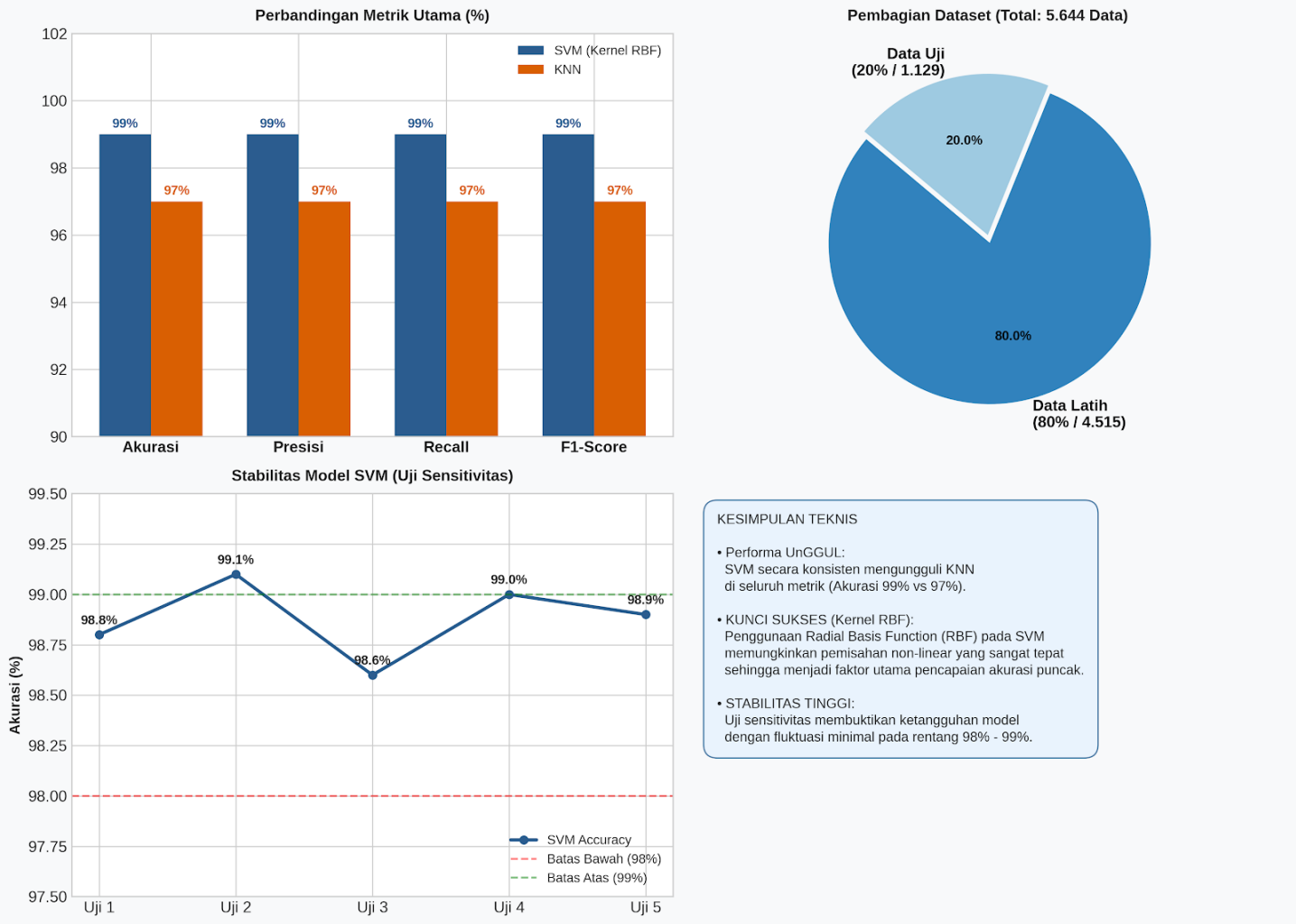
Superioritas SVM (99%) dibandingkan KNN (97%) disebabkan oleh kemampuan SVM dalam mencari *hyperplane* optimal melalui *support vectors* yang membatasi tiap kelas, sehingga lebih tangguh terhadap *noise* dibandingkan KNN yang berbasis kalkulasi jarak ke seluruh titik data. Penggunaan kernel RBF terbukti efektif dalam memetakan batas pemisah non-linear pada data demografi sosial.

Analisis kesalahan menunjukkan bahwa misklasifikasi yang sangat minim hanya terjadi pada data dengan skor verifikasi yang berada tepat di ambang batas (*threshold*) antar kategori, yang menunjukkan adanya pola *fuzzy boundaries*. Secara manajerial, akurasi tinggi ini memungkinkan model diimplementasikan sebagai *automated screening tool* untuk efisiensi anggaran bantuan sosial

E. Tujuan Eksplorasi: Pendekatan ensemble diproyeksikan mampu mendongkrak stabilitas dan kecepatan pemrosesan model saat dihadapkan pada volume data kependudukan yang sangat besar dan memiliki tingkat noise (derau) yang tinggi di masa depan.

F. Perbandingan Model: Selain membandingkan antar sesama model ensemble, disarankan juga untuk mengomparasikan performanya dengan arsitektur Artificial Neural Network (ANN) atau Deep Learning untuk memperkuat khasanah ilmu data dalam sektor kebijakan publik.

G. Relevansi: Penggunaan model yang lebih kompleks ini penting karena karakteristik data kemiskinan sering kali memiliki pola yang rumit dan batas-batas kategori yang halus (fuzzy boundaries), sehingga diperlukan model yang lebih kuat dari sekadar model tunggal untuk menjaga akurasi prediksi tetap tinggi di masa depan



1. **Perbandingan Metrik Utama (Bar Chart)**
   - **SVM (Kernel RBF):** Unggul konsisten sebesar **99%** di seluruh metrik (*Accuracy*, *Precision*, *Recall*, *F1-Score*).
   - **KNN:** Berada pada tingkat **97%** di seluruh metrik.
2. **Pembagian Dataset (Pie Chart)**

- **Total Data:** 5.644 sampel.
- **Data Latih (Train Set):** 80% (4.515 sampel).
- **Data Uji (Test Set):** 20% (1.129 sampel).

3. **Stabilitas Model SVM (Line Chart)**

- **Hasil Uji Sensitivitas:** Menunjukkan tingkat ketangguhan model yang sangat tinggi.
- **Fluktuasi Akurasi:** Sangat minimal, bertahan stabil dalam rentang **98% hingga 99%** pada berbagai iterasi pengujian.

4. **Kesimpulan Teknis**

   Faktor Kunci: Penggunaan **Kernel RBF (*Radial Basis Function*)** pada SVM terbukti menjadi kunci utama pencapaian akurasi puncak (99%) karena kemampuannya memetakan pola data non-linear secara optimal.

III. KESIMPULAN

A. Kesimpulan Penelitian menyimpulkan bahwa integrasi alur *preprocessing* dan rekayasa data SMOTE merupakan pondasi krusial dalam melahirkan basis data latih yang adil,. Algoritma SVM dengan kernel RBF adalah model paling kompeten dengan akurasi puncak 99% untuk memetakan pola kemiskinan urbanModel ini sangat valid untuk diandalkan sebagai instrumen pendukung keputusan digital pada Dinas Sosial Kota Tangerang Selatan.

B. Saran Untuk pengembangan selanjutnya, disarankan penguatan variabel multidimensi seperti kondisi fisik bangunan dan akses sanitasi,. Selain itu, disarankan eksplorasi algoritma *Ensemble Learning* (seperti *XGBoost* atau *Random Forest*) untuk menangani volume data yang lebih besar di masa depan,. Hilirisasi teknologi melalui integrasi API ke dalam sistem informasi kependudukan daerah sangat direkomendasikan untuk mewujudkan *Smart Governance* yang dinamis,

UCAPAN TERIMA KASIH (JIKA ADA)

Tuliskan ucapan terima kasih dengan bahasa baku, misalnya, Penulis A.F. mengucapkan terima kasih kepada Direktorat Pendidikan Tinggi, Departemen Pendidikan dan Kebudayaan Republik Indonesia yang telah memberikan dukungan finansial melalui Beasiswa Bidik Misi tahun 2010-2014. Penulis juga diperkenankan menyampaikan ucapan terima kasih kepada sponsor penyedia dana penelitian.

DAFTAR PUSTAKA

[1] Badan Pusat Statistik, “Laporan Profil Kemiskinan Indonesia,” 2024.
[2] World Bank, “Poverty and Shared Prosperity Report,” 2023.
[3] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2022.
[4] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. O’Reilly Media, 2023.
[5] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2021.
[6] I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, 2023.
[7] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, 1995.
[8] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inf. Theory, 1967.
[9] Fauziah dkk., “Komparasi Algoritma SVM dan KNN pada Data Kemiskinan,” 2022.
[10] Sihombing & Arsani, “Stabilitas SVM pada Data Sosial Non-Linear,” 2021.
[11] Lestari & Wibowo, “Analisis Performa KNN dan SVM pada Dataset Seimbang,” 2021.