



IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI TINGKAT KEMISKINAN SEBAGAI PENDUKUNG PENGAMBILAN KEPUTUSAN SOSIAL DI INDONESIA

Juanuari, Ilham Manzis, Abdurrozzaq Musyaffa, Simson Mali Ngara, Haura Syahla, Jordy Lasmana Putra*

Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Indonesia

Abstrak: Kemiskinan tetap menjadi tantangan utama dalam pembangunan berkelanjutan di Indonesia, yang diperparah oleh kurangnya sistem identifikasi presisi untuk penyaluran bantuan sosial. Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan tingkat kemiskinan di 514 kabupaten/kota di Indonesia guna mendukung pengambilan keputusan kebijakan sosial yang lebih tepat sasaran. Metode yang digunakan mencakup analisis dataset dengan 12 atribut indikator sosial-ekonomi, seperti Indeks Pembangunan Manusia (IPM) dan akses sanitasi. Tahapan pra-pemrosesan data melibatkan normalisasi min-max untuk menyelaraskan skala fitur yang heterogen serta teknik Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan distribusi kelas antara kategori "Miskin" dan "Tidak Miskin". Model divalidasi menggunakan metode 10-fold cross-validation dengan menguji variasi parameter k (3, 5, dan 7) melalui perangkat lunak RapidMiner. Hasil eksperimen menunjukkan bahwa algoritma KNN dengan parameter $k=3$ menghasilkan performa paling optimal dengan tingkat akurasi sebesar 89,88%. Meskipun presisi untuk kelas minoritas (miskin) tercatat sebesar 55,00%, model ini mampu mencapai nilai recall yang sangat tinggi, yakni 88,71%, serta rata-rata tertimbang F1-score sebesar 93,45%. Tingginya nilai recall menunjukkan kemampuan model dalam mendeteksi mayoritas penduduk miskin secara akurat, sehingga meminimalkan risiko pengabaian bantuan sosial. Penelitian ini menyimpulkan bahwa integrasi normalisasi dan SMOTE pada algoritma KNN secara signifikan meningkatkan sensitivitas model dalam mengidentifikasi kelompok rentan. Penggunaan model ini dianggap layak untuk kebijakan sosial inklusif karena lebih mengutamakan pengurangan kesalahan pengabaian (false negative). Namun, terdapat keterbatasan pada penggunaan data agregat wilayah yang berisiko memicu ecological fallacy. Untuk penelitian mendatang, disarankan penggunaan data tingkat rumah tangga yang lebih mikro serta integrasi indikator multidimensi yang lebih luas guna meningkatkan presisi identifikasi individu.

Kata kunci: K-Nearest Neighbor, Machine Learning, Decision Support System, Data Sosial Ekonomi, Kemiskinan Regional

I. PENDAHULUAN

Kemiskinan tetap menjadi tantangan utama dalam pembangunan berkelanjutan di

Indonesia. Berdasarkan data Badan Pusat Statistik (BPS), pada tahun 2023, persentase penduduk miskin mencapai 9,36%, dengan ketimpangan yang signifikan antara wilayah perkotaan dan pedesaan (Anna, 2023). Situasi ini diperparah oleh belum optimalnya pendistribusian bantuan sosial, yang disebabkan oleh kurangnya sistem identifikasi yang presisi untuk mengenali siapa yang

*) jordy.jlp@bsi.ac.id

Diterima: 4 Juni 2025

Direvisi: 10 Februari 2026

Disetujui: 28 Februari 2026

DOI: 10.23969/infomatek.v28i1.26866

benar-benar membutuhkan bantuan (Agus Triono & Sangaji, 2023).

Secara konseptual, kemiskinan tidak hanya merupakan kondisi statis keterbatasan ekonomi, melainkan berfungsi sebagai risk factor multidimensi yang memperparah kerentanan sosial mulai dari keterbatasan akses pendidikan dan layanan kesehatan hingga eksklusi dari sistem perlindungan sosial (Linggar Vandito et al., 2023). Dalam perspektif kebijakan, pendekatan klasifikasi berbasis machine learning dapat bertransformasi menjadi sistem early warning sosial yang proaktif, memungkinkan identifikasi dini kelompok rentan sebelum kondisi kemiskinan mengakar permanen dan memicu transmisi antargenerasi (Lasfeto et al., 2024).

Untuk mengatasi permasalahan tersebut, pemanfaatan teknologi seperti machine learning menjadi solusi dalam melakukan klasifikasi tingkat kemiskinan secara otomatis dan berbasis data. Penelitian sebelumnya oleh Mukharyahya et al. (2025) membandingkan algoritma *Naive Bayes* dan *Support Vector Machine* (SVM) dalam mengklasifikasikan tingkat kemiskinan di Indonesia (Mukharyahya et al., 2025). Penelitian tersebut menggunakan dataset berisi 514 data sosial sampai ekonomi dan menghasilkan *accuracy* terbaik sebesar 81% pada algoritma SVM, dengan *precision* 69%, *recall* 78% dan *F1-score* 71%.

Namun demikian, gap ilmiah yang mendasar dalam studi-studi terdahulu terletak pada dua aspek kritis: pertama, minimnya integrasi kerangka teoretis kemiskinan multidimensi (misalnya indikator akses sanitasi, ketahanan pangan, atau modal sosial) ke dalam desain fitur model machine learning yang masih berfokus pada variabel ekonomi konvensional (Muñetón-Santa & Manrique-Ruiz, 2023); kedua, kurangnya interpretasi model yang

dapat ditranslasikan menjadi rekomendasi kebijakan operasional sehingga performa metrik klasifikasi belum sepenuhnya menjawab tantangan targeting bantuan yang kontekstual dan adaptif terhadap dinamika kerentanan lokal (Komarawati et al., 2025).

Hal ini menciptakan peluang untuk meningkatkan akurasi serta performa klasifikasi, khususnya dalam hal recall yang sangat penting untuk mendeteksi kemiskinan secara tepat. Salah satu algoritma yang potensial untuk digunakan adalah *K-Nearest Neighbor* (KNN). Menurut Fauziah et al. (2022), meskipun dalam studi mereka KNN menghasilkan *accuracy* 58,62% dibandingkan SVM yang mencapai 93,1% dalam klasifikasi data kemiskinan di Papua (Fauziah et al., 2022), hal ini menunjukkan bahwa performa KNN dapat bervariasi tergantung pada karakteristik data dan parameter yang digunakan. Sementara itu, Mardiah et al. (2024) menunjukkan bahwa KNN efektif dalam mengklasifikasikan data ekonomi masyarakat di Nagari, Sumatera Barat, membantu pemerintah daerah dalam pengambilan keputusan terkait bantuan sosial (Mardiah et al., 2024)

Penelitian lain oleh Djafar & Fauzan, (2024) menerapkan metode KNN dengan teknik oversampling pada data campuran dalam mengklasifikasikan status kesejahteraan rumah tangga di Kulon Progo. Hasilnya menunjukkan bahwa pendekatan ini mampu meningkatkan akurasi dan sensitivitas model prediksi dengan nilai *accuracy* mendekati 79% (Djafar & Fauzan, 2024). Selain itu, Amalia et al. (2025) membandingkan algoritma KNN dan C4.5 dalam menentukan kelayakan penerima bantuan langsung tunai (BLT), dan menemukan bahwa KNN menunjukkan hasil stabil untuk kasus dengan jumlah data terbatas dan variabel ekonomi yang dominan (Amalia et al., 2025).

Studi oleh Duwo Jiwo Saputro et al. (2024) yang fokus pada klasifikasi tingkat kemiskinan di Jawa Barat dengan algoritma KNN juga menunjukkan bahwa model KNN dengan nilai k optimal mampu mencapai *accuracy* lebih dari 96%, dengan *recall* mencapai 100%, yang berarti seluruh data kelompok miskin berhasil diklasifikasikan dengan benar tanpa kesalahan (Duwo Jiwo Saputro et al., 2024). Hasil ini menunjukkan bahwa dengan pemilihan parameter dan teknik *preprocessing* yang tepat, KNN dapat mengungguli algoritma lain dalam konteks sosial ekonomi lokal.

Dengan latar belakang tersebut, penelitian ini bertujuan untuk mengeksplorasi keefektifan algoritma KNN dalam klasifikasi tingkat kemiskinan di Indonesia, sekaligus mengevaluasi performanya melalui metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Harapannya, pendekatan ini dapat memberikan kontribusi nyata dalam mendukung pengambilan keputusan kebijakan sosial berbasis data yang lebih tepat sasaran.

II. METODOLOGI

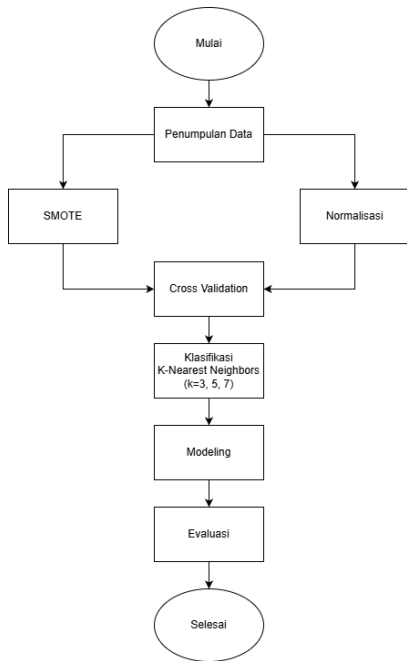
Metode penelitian ini diawali dengan pengumpulan dataset yang diperoleh dari Mukharyahya et al. (2025), terdiri dari 514 data kabupaten atau kota di Indonesia, dan tersedia secara publik melalui platform Kaggle dengan nama Klasifikasi Tingkat Kemiskinan di Indonesia, serta mencakup berbagai fitur sosial hingga ekonomi, antara lain persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, indeks pembangunan manusia (IPM), usia harapan hidup, persentase akses sanitasi layak, persentase akses air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, nilai PDRB atas harga konstan dari pengeluaran, serta atribut label yang digunakan sebagai variabel target, yaitu klasifikasi tingkat kemiskinan. Variabel target

dalam data ini dikonversi menjadi bentuk numerik untuk keperluan klasifikasi, dengan 1 menandakan status miskin dan 0 untuk tidak miskin. Selanjutnya, data melalui dua tahap pra-pemrosesan: normalisasi min-max untuk menyelaraskan skala atribut numerik yang heterogen, dan SMOTE (Synthetic Minority Oversampling Technique) untuk mengatasi ketidakseimbangan distribusi kelas antara kategori "Miskin" dan "Tidak Miskin".

Validasi model dilakukan menggunakan metode 10-fold cross-validation guna memastikan generalisasi yang andal tanpa ketergantungan pada pembagian data acak. Setelah pra-pemrosesan, algoritma K-Nearest Neighbor (KNN) diimplementasikan dengan parameter $k=3$ yang dipilih berdasarkan eksperimen komparatif terhadap variasi k (3, 5, dan 7). Proses pemodelan dilakukan menggunakan perangkat lunak RapidMiner yang menyediakan alur kerja visual terintegrasi. Tahapan yang diterapkan meliputi: pemanggilan dataset → normalisasi → SMOTE → validasi silang 10-fold → pelatihan KNN ($k=3, 5, 7$) → evaluasi performa.

Evaluasi model dilakukan menggunakan confusion matrix dengan empat metrik utama: akurasi (proporsi prediksi benar secara keseluruhan), presisi (ketepatan prediksi kelas positif), *recall* (kemampuan mendeteksi seluruh kasus kemiskinan), dan *F1-score* (keseimbangan harmonis antara presisi dan *recall*). Keempat metrik ini menjadi dasar penilaian efektivitas model, dengan penekanan khusus pada *recall* kelas "Miskin" sebagai indikator kritis dalam konteks kebijakan sosial untuk meminimalkan risiko warga miskin yang tidak terdeteksi (*false negative*).

Gambar 1 menjelaskan alur penelitian yang dilakukan dalam kegiatan ini.



Gambar 1. Alur Penelitian.

yang dimulai dari pengumpulan dataset, dilanjutkan dengan normalisasi data, penanganan ketidakseimbangan kelas menggunakan SMOTE, validasi model melalui 10-fold cross-validation, klasifikasi menggunakan algoritma K-Nearest Neighbor

dengan parameter $k=3$, hingga tahap akhir berupa evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari studi yang dilakukan oleh Mukharyahya et al. (2025). Dataset tersebut terdiri dari 514 data, yang mewakili masing-masing kabupaten/kota di Indonesia. Dataset ini mencakup sejumlah atribut yang merepresentasikan faktor-faktor penentu tingkat kemiskinan, antara lain persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, indeks pembangunan manusia (IPM), usia harapan hidup, persentase akses sanitasi layak, persentase akses air minum layak, tingkat pengangguran terbuka, tingkat partisipasi angkatan kerja, serta nilai PDRB atas harga konstan dari pengeluaran.

Selain atribut-atribut tersebut, dataset ini juga memiliki satu atribut label yang digunakan sebagai variabel target, yaitu klasifikasi tingkat kemiskinan. Dataset ini tersedia secara publik di platform Kaggle dengan judul "Klasifikasi Tingkat Kemiskinan di Indonesia".

A	B	C	D	E	F	G	H	I	J	K	L
Kab/Kota	Persentase Kemiskinan	Rata-Rata Lama Sekolah	Pengeluaran per Kapita	IPM	Umur Harapan Hidup	Akses Sanitasi	Akses Air Minum Layak	Pengangguran	Partisipasi Kerja	PDRB Harga Konstan	Klasifikasi Kemiskinan
1 Simeulue	18,98	9,48	7148	66,41	65,28	71,56	87,45	5,71	71,15	1648096	Tidak Miskin
2 Aceh Singgaja	20,36	8,68	8776	69,22	67,43	69,56	78,58	8,36	62,85	1780419	Miskin
4 Aceh Selatan	13,18	8,88	8180	67,44	64,4	62,55	79,65	6,46	60,85	4345784	Tidak Miskin
5 Aceh Tenggara	13,41	9,67	8030	69,44	68,22	62,71	86,71	6,43	69,62	3487157	Tidak Miskin
6 Aceh Timur	14,45	8,21	8577	67,83	68,74	66,75	83,16	7,13	59,48	8433526	Tidak Miskin
7 Aceh Tenggara	15,26	9,86	10780	73,37	68,86	90,58	90,1	2,61	76,3	5953118	Tidak Miskin
8 Aceh Barat	18,81	9,55	9593	71,67	67,99	89,6	94,22	7,09	60,05	7485861	Tidak Miskin
9 Aceh Besar	14,05	10,33	9644	73,58	69,79	87,4	82,36	7,7	61,67	10261585	Tidak Miskin
10 Pidie	19,59	9	9860	70,7	66,95	54,1	89,24	7,28	60,29	7975099	Tidak Miskin
11 Bireuen	13,25	9,29	8867	72,33	71,26	81,89	93,53	4,32	65,91	10374480	Tidak Miskin
12 Aceh Utara	17,43	8,64	8201	69,46	68,81	79,97	91,09	8,31	58,47	16924103	Tidak Miskin
13 Aceh Barat Daya	16,34	8,67	8428	66,99	65,06	65,71	95,34	4,04	57,91	3069805	Tidak Miskin
14 Gayo Lues	19,64	8,4	8856	67,56	65,53	47,63	84,68	1,84	78,99	1981879	Tidak Miskin
15 Aceh Tamiang	13,34	8,91	8367	69,48	69,63	87,45	83,12	5,87	66,43	6062520	Tidak Miskin
16 Nagan Raya	18,23	8,69	8292	69,31	69,24	74,86	90,13	4,99	64,99	7110421	Tidak Miskin
17 Aceh Jaya	13,23	8,71	9666	69,84	67,19	81,6	86,36	3,47	72,59	2033844	Tidak Miskin
18 Bener Meriah	19,16	10	11118	73,27	69,26	86,69	89,71	1,24	77,53	3744095	Tidak Miskin
19 Pidie Jaya	19,55	9,34	10290	73,6	70,18	74,3	93,46	3,57	57,77	2635517	Tidak Miskin
20 Kota Banda Aceh	7,61	12,83	16891	85,71	71,52	99,88	99,37	8,94	63	15454371	Tidak Miskin
21 Kota Sabang	15,32	11,18	11378	76,11	70,56	92,25	96,18	3,56	63,71	1152875	Tidak Miskin
22 Kota Langsa	10,96	11,12	12067	77,44	69,43	91,81	97,14	7,21	67,04	3974614	Tidak Miskin
23 Kota Lhokseumawe	11,16	11,11	11390	77,57	71,64	93,1	94,44	11,16	63,91	7252905	Tidak Miskin

Gambar 2. Sampel Dataset Tingkat Kemiskinan di Wilayah Indonesia.

2.3. Normalisasi

Normalisasi Min-Max diterapkan untuk menskalakan seluruh atribut numerik (misalnya persentase kemiskinan, IPM, PDRB) ke rentang $[0,1]$. Proses ini mengeliminasi disparitas skala antar variabel yang berbeda dimensi, sehingga jarak Euclidean dalam KNN tidak didominasi oleh fitur berskala besar seperti PDRB. Normalisasi juga mempercepat konvergensi komputasi dan meningkatkan stabilitas model. Tanpa normalisasi, fitur seperti pengeluaran per kapita (dalam jutaan rupiah) akan mendistorsi pengukuran kedekatan, mengakibatkan kesalahan klasifikasi pada wilayah dengan karakteristik multidimensi.

2.4. Smote

Tahap ini mengatasi ketidakseimbangan kelas pada dataset kemiskinan di mana kabupaten/kota berkategori "sangat miskin" jumlahnya lebih sedikit berpotensi menyebabkan bias prediktif. Untuk mengatasinya, diterapkan SMOTE yang menghasilkan sampel sintesis kelas minoritas melalui interpolasi linier antar tetangga terdekat dalam ruang fitur. Teknik ini meningkatkan representasi kelas minoritas tanpa duplikasi data, sehingga memperbaiki sensitivitas model terhadap wilayah rentan. Pendekatan ini krusial dalam konteks kebijakan sosial agar alokasi bantuan tidak terdistorsi oleh dominasi kelas mayoritas dalam pelatihan mode

2.3. Cross Validation

Pada tahap ini dilakukan validasi silang 10-fold diterapkan untuk memperoleh estimasi kinerja model yang lebih stabil dan minim bias. Dataset dibagi menjadi sepuluh lipatan; dalam setiap iterasi, sembilan lipatan digunakan untuk pelatihan dan satu lipatan sisanya untuk pengujian, diulang hingga seluruh lipatan

berperan sebagai data uji. Pendekatan 10-fold—sebagai standar emas dalam evaluasi model pembelajaran mesin memberikan pemanfaatan data yang lebih efisien dibanding 5-fold, sekaligus mengurangi varians estimasi akurasi hingga $\sim 15\%$. Hal ini krusial mengingat keterbatasan sampel (514 kabupaten/kota) dan kebutuhan generalisasi spasial yang andal. Rata-rata metrik evaluasi dari sepuluh iterasi menjadi dasar objektif pemilihan parameter k optimal sebelum deployment sistem pendukung keputusan kebijakan sosial.

2.4. Klasifikasi K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan salah satu algoritma supervised learning yang digunakan untuk tugas klasifikasi. Konsep dasar dari KNN adalah mengklasifikasikan suatu data baru berdasarkan kedekatannya dengan data lain yang sudah memiliki label (Fitra & Rusdi, 2022). Dalam penelitian ini bertujuan untuk melakukan klasifikasi tingkat kemiskinan di wilayah Indonesia berdasarkan data sosial dan ekonomi yang telah dikumpulkan. Proses klasifikasi dilakukan dengan membandingkan jarak antara data latih dan data uji, lalu menentukan sejumlah tetangga terdekat untuk melihat kelas mana yang lebih baik.

KNN dikenal karena kesederhanaannya dalam implementasi dan efektivitasnya dalam menangani data dengan jumlah fitur yang tidak terlalu besar. Selain itu, KNN juga fleksibel dalam digunakan untuk berbagai jenis klasifikasi, termasuk dalam konteks data sosial dan ekonomi (Fitra & Rusdi, 2022).

2.5. Modeling

Proses pemodelan dalam penelitian ini dilakukan menggunakan perangkat lunak RapidMiner yang berbasis visual, sehingga memudahkan integrasi antara pemanggilan

data, pelatihan model, dan evaluasi performa dalam satu alur terstruktur. Tahapan dilakukan secara berurutan, dimulai dari pemanggilan data latih, penerapan algoritma *K-Nearest Neighbor* (KNN), penerapan model pada data uji, hingga pengukuran kinerja menggunakan metrik evaluasi.

RapidMiner adalah perangkat lunak opensource yang menyediakan antarmuka grafis (GUI) untuk merancang proses analisis data seperti data mining, text mining, dan prediksi secara visual. GUI ini menghasilkan file XML yang dapat dijalankan otomatis oleh sistem untuk mengeksekusi proses yang telah dirancang (Muhammad, Rino Indra Nainggolan et al., 2021). Pemilihan RapidMiner didasarkan pada kemampuannya dalam menyederhanakan proses pemodelan tanpa perlu pemrograman manual, serta mendukung berbagai algoritma machine learning, termasuk KNN, secara efisien dan intuitif.

2.6. Evaluasi

Evaluasi model merupakan tahap penting untuk mengukur efektivitas algoritma *K-Nearest Neighbor* (KNN) dalam melakukan klasifikasi berdasarkan data sosial ekonomi. Evaluasi dilakukan dengan membandingkan hasil klasifikasi model dengan label aktual pada data uji, guna menilai akurasi dan mengidentifikasi kesalahan klasifikasi yang terjadi.

Melalui evaluasi ini, peneliti dapat memahami kekuatan dan kelemahan model, termasuk potensi bias terhadap salah satu kelas atau ketidakseimbangan dalam pengenalan kelas minoritas (Fadlullah et al., 2025). Hasil evaluasi juga menjadi dasar untuk pengembangan model di masa mendatang, seperti penyesuaian parameter, peningkatan kualitas data, maupun eksplorasi metode lain yang lebih sesuai. Tahapan ini berperan penting dalam membangun sistem klasifikasi

yang efektif, khususnya dalam mendukung kebijakan publik seperti pemetaan kemiskinan.

Pada tahap evaluasi model, *Confusion Matrix* dapat diterapkan guna memberikan pandangan yang lebih menyeluruh terhadap kinerja algoritma, dengan mempertimbangkan sejumlah metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* (Rivaldo, Vito Junivan & Pranoto, 2024). Pengujian dan metrik evaluasi inilah yang menjadi dasar utama dalam menilai performa hasil klasifikasi, sebagaimana dijelaskan pada bagian berikut.

1. *Accuracy*: *Accuracy* adalah tingkat kesamaan antara hasil klasifikasi dan nilai sebenarnya atau nilai aktual (Putri, 2021). Untuk melakukan perhitungan akurasi dapat menerapkan Persamaan 1.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (1)$$

Keterangan:

TP (*True Positive*) = total data positif teridentifikasi benar

TN (*True Negative*) = total data negatif teridentifikasi salah

FP (*False Positive*) = total data positif teridentifikasi salah

FN (*False Negative*) = total data negatif teridentifikasi positif

2. *Precision*: *Precision* digunakan untuk mengevaluasi keakuratan model dalam menghasilkan klasifikasi positif dengan mengurangi kesalahan positif palsu, *precision* dihitung sebagai rasio *True Positive* terhadap jumlah total klasifikasi positif. Rasio ini menunjukkan proporsi klasifikasi positif yang benar-benar positif (Nikmatun et al., 2019). Untuk penghitungan presisi dapat menerapkan Persamaan 2.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

Keterangan:

TP = True Positive
 FP = False Positive

3. *Recall*: *Recall* bertujuan menghitung persentase data yang positif secara efektif teridentifikasi model terhadap seluruh jumlah data positif yang tersedia, atau seberapa baik model bisa menemukan semua data yang positif benar (Palma et al., 2021). Untuk perhitungan nilai Recall menerapkan Persamaan 3.

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

Keterangan:
 TP = True Positive
 FN = False Negative

4. *F1-Score*: *F1-Score* akan digunakan dalam melakukan klasifikasi guna menyeimbangkan presisi dan Recall dengan menggabungkan keduanya menjadi satu nilai, yaitu presisi (keakuratan klasifikasi data positif) dan Recall (kapasitas model mengidentifikasi semua sampel data positif) (Kasanah et al., 2019). Perhitungan F1-Score dapat menerapkan Persamaan 4.

$$\text{F1-Score} = 2 \times \frac{(\text{presisi} \times \text{recall})}{(\text{presisi} + \text{recall})} \quad (4)$$

5. *Confusion Matrix*: *Confusion Matrix* digunakan untuk membandingkan hasil dari klasifikasi dengan nilai sebenarnya pada data uji untuk menilai akurasi mode dalam melakukan klasifikasi data (Hernandes et al., 2023). *Confussion Matrix* dapat dilihat pada Gambar 3.

	Actual Positif	Actual Negatif
Prediksi Positif	TP	FN
Prediksi Negatif	TN	FP

Gambar 3. *Confusion Matrix*.

Keterangan:

TP = True Positive
 TN = True Negative
 FP = False Positive
 FN = False Negative

Melalui confusion matrix, peneliti dapat melihat secara rinci Klasifikasi benar dan salah yang dilakukan model. Dari nilai-nilai ini pula diperoleh metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* yang digunakan untuk menilai performa model secara menyeluruh.

III. HASIL DAN PEMBAHASAN

3.1. Klasifikasi K-Nearest Neighbor (KNN)

Setelah Normalisasi dan penyeimbangan label data dengan SMOTE tahap berikutnya adalah melakukan proses klasifikasi menggunakan *algoritma K-Nearest Neighbor* (KNN). Algoritma ini bekerja dengan cara mengukur jarak antar data dengan menggunakan nilai parameter K = 3, yang berarti bahwa penentuan kelas suatu data ditentukan berdasarkan tiga tetangga terdekatnya.

Tabel 2. Perbandingan Performansi (K=3, 5, 7)

k	SMOTE	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
3	✓	89,88	55,00	88,71	67,90
3	-	92,23	73,91	54,84	62,72
5	✓	88,13	50,47	87,10	64,40
5	-	91,46	71,43	48,39	57,70
7	✓	88,51	51,40	88,71	65,00
7	-	91,46	75,00	43,55	55,10

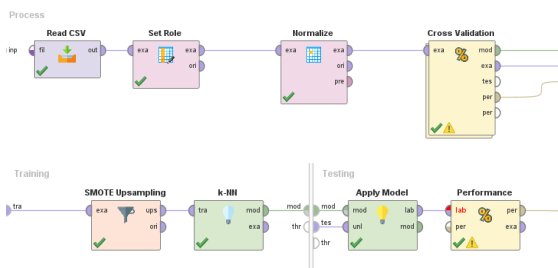
Tabel 2 menunjukkan bahwa KNN dengan k=3 dan SMOTE menghasilkan F1-Score tertinggi (67,90%), mengungguli konfigurasi lain. Peningkatan ini dipicu oleh presisi (55,00%) dan recall (88,71%) yang seimbang, mengindikasikan kemampuan model dalam mengidentifikasi kelas minoritas (wilayah miskin) secara akurat tanpa mengorbankan sensitivitas deteksi. Meski akurasi tetap tinggi (>88%) pada seluruh skenario, F1-Score menjadi indikator kritis karena mencerminkan

keseimbangan antara presisi dan recall dalam konteks klasifikasi tidak seimbang. Nilai $k=3$ lebih efektif menangkap pola lokal pada data yang telah diseimbangkan melalui SMOTE, terutama mengingat variasi spasial karakteristik kemiskinan di Indonesia yang kompleks. Temuan ini mendukung penerapan k minimal bersama SMOTE untuk meningkatkan validasi model dalam studi kemiskinan, sehingga dapat meminimalkan kesalahan alokasi sumber daya sosial dan memperkuat keputusan berbasis bukti.

3.2. Modeling

Pembuatan model pengujian dengan algoritma *K-Nearest Neighbor* (KNN) dilakukan pada dataset yang terdiri dari 12 atribut, yang merupakan atribut dari klasifikasi tingkat kemiskinan di wilayah Indonesia.

Tahap ini dirancang dalam bentuk pemodelan menggunakan perangkat lunak aplikasi RapidMiner, seperti yang ditampilkan pada Gambar 4.



Gambar 4. Alur Proses Pembuatan Model KNN pada RapidMiner.

Setelah data dimasukkan ke dalam aplikasi, proses dimulai dengan memilih algoritma *K-Nearest Neighbor*, kemudian ditambahkan komponen *Apply Model* untuk menerapkan model pada data uji, dan *Performance* untuk menampilkan hasil dari pengolahan data. Model ini digunakan untuk memprediksi tingkat kemiskinan berdasarkan karakteristik data sosial dan ekonomi yang telah dipelajari dari data latih, dan hasil evaluasinya diukur

dengan *matrix accuracy, precision, recall*, serta *F1-score*.

3.3. Evaluasi Model

Setelah dilakukan tahap pemodelan, langkah selanjutnya adalah melakukan evaluasi terhadap model yang telah dibangun. Evaluasi ini bertujuan untuk mengetahui seberapa baik performa algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan data ke dalam kategori “miskin” dan “tidak miskin” secara tepat berdasarkan atribut sosial dan ekonomi yang digunakan.

Untuk melakukan evaluasi, digunakan metode *Confusion Matrix*, yang memberikan gambaran mengenai jumlah klasifikasi yang benar dan salah dari model terhadap data uji. *Confusion matrix* menunjukkan hasil perbandingan antara nilai aktual (label sebenarnya) dan hasil klasifikasi model, sehingga dapat diketahui jumlah data yang berhasil diklasifikasikan dengan benar (*True Positive* dan *True Negative*) maupun yang salah (*False Positive* dan *False Negative*). Hasil *confusion matrix* dalam penelitian ini ditampilkan pada tabel 3 berikut:

Tabel 3. Hasil *Confusion Matrix* KNN

	<i>True Miskin</i>	<i>Tidak Miskin</i>	<i>Class Precision</i>
Pred.Tidak Miskin	407	7	98.31%
Pred. Miskin	45	55	55.00%
Class Recall	90.04%	88.71%	

Berdasarkan tabel 3 Model KNN mengklasifikasikan wilayah tidak miskin dengan presisi 98,31%, sedangkan presisi untuk kategori miskin sebesar 55%, mengindikasikan adanya sampel tidak miskin yang terklasifikasi sebagai miskin. Recall untuk kedua kelas tergolong memadai (90,04% dan 88,71%), menunjukkan model mampu mengenali sebagian besar sampel miskin yang sebenarnya. Perbedaan presisi

antar kelas berdampak pada skor F1 yang tidak seimbang (94,10% vs 67,84%). Temuan ini mencerminkan kompleksitas karakteristik wilayah miskin yang memiliki tumpang tindih indikator dengan wilayah tidak miskin, seperti variasi akses sanitasi atau tingkat pendidikan yang tidak selalu linear dengan status kemiskinan.

IV. KESIMPULAN

Penelitian ini menyimpulkan bahwa algoritma K-Nearest Neighbor (KNN) dengan $k=3$, normalisasi, dan SMOTE efektif mengklasifikasikan kemiskinan di Indonesia dengan akurasi 89,88%. Nilai recall 88,71% sangat krusial untuk meminimalkan risiko terlewatnya bantuan bagi warga miskin. Keterbatasan utama terletak pada penggunaan data agregat wilayah yang berisiko memicu ecological fallacy serta kurang optimalnya KNN menangani autokorelasi spasial. Penelitian selanjutnya disarankan menggunakan data mikro tingkat rumah tangga serta mengintegrasikan indikator multidimensi seperti modal sosial. Peningkatan presisi juga diperlukan untuk mengurangi false positive guna mendukung kebijakan yang lebih tepat sasaran.

DAFTAR PUSTAKA

- Agus Triono, T., & Sangaji, R. C. (2023). Faktor Mempengaruhi Tingkat Kemiskinan di Indonesia: Studi Literatur Laporan Data Kemiskinan BPS Tahun 2022. *Journal of Society Bridge*, 1(1), 59–67. <https://doi.org/10.59012/jsb.v1i1.5>
- Amalia, W., Nur, C., Abdullah, D., & Meiyanti, R. (2025). Performance of K-Nearest Neighbor Algorithm and C4.5 Algorithm in Classifying Citizens Eligible to Receive Direct Cash Assistance in Bandar Mahligai Village. *International Journal of Engineering, Science and Information Technology*, 5(1), 368–372.
- Anna, A. (2023). Pengujian Teknik Algoritma Klasifikasi Terhadap Tingkat Kemiskinan Penduduk. *JTIK (Jurnal Teknik Informatika Kaputama)*, 7(1), 61–66. <https://doi.org/10.59697/jtik.v7i1.35>
- Djafar, N. M., & Fauzan, A. (2024). Implementation of K-Nearest Neighbor using the oversampling technique on mixed data for the classification of household welfare status. *Statistics in Transition New Series*, 25(1), 109–124. <https://doi.org/10.59170/stattrans-2024-007>
- Duwo Jiwo Saputro, A., Darmawan, A., & Nurina Sari, B. (2024). Klasifikasi Persentase Kemiskinan Di Jawa Barat Menggunakan Data Mining Algoritma K-Nearest Neighbor (Knn). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2718–2723. <https://doi.org/10.36040/jati.v7i4.7178>
- Fadlullah, R., Winarno, S., & Naufal, M. (2025). Integrasi Convolutional Autoencoder dengan Support Vector Machine untuk Klasifikasi Varietas Integration of Convolutional Autoencoder with Support Vector Machine for Almond Varieties Classification. *Jurnal Teknik Informatika Dan Sistem Informasi*, 11(April), 63–77.
- Fauziah, Tiro, M. A., & Ruliana. (2022). Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua. *ARRUS Journal of Mathematics and Applied Science*, 2(2), 83–91. <https://doi.org/10.35877/mathscience741>
- Fitra, R., & Rusdi, I. (2022). Penerapan Metode Algoritma K-Nearest Neighbor Menggunakan Rapidminer Studio Pada Klasifikasi Status Sosial Ekonomi Studi Kasus: Kelurahan Kapuk Muara Rt 010 Rw 04. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 11(4), 653–660. <https://doi.org/10.30591/smartcomp.v11i4.4250>
- Hernandes, A., Kurnia Gusti, S., Syafria, F., Handayani, L., & Ramadhani, S. (2023). Klasifikasi Data Penerimaan Zakat dengan Algoritma K-Nearest Neighbor.

- Media Online*, 4(3), 1632–1640.
<https://doi.org/10.30865/klik.v4i3.1528>
- Kasanah, A. N., Muladi, M., & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201.
<https://doi.org/10.29207/resti.v3i2.945>
- Komarawati, K., Nurdin, M. F., Gunawan, W., & Nurwati, R. N. (2025). Transformative capacity to build an adaptive society resilience during crisis: evidence from conditional cash transfer/PKH Indonesia. *Frontiers in Sociology*, 10.
<https://doi.org/10.3389/fsoc.2025.1507177>
- Lasfeto, D. B., Setyorini, T., Mauta, J. J., Bria, M., & Nenobais, O. O. N. (2024). A simple classification and clustering of poverty in rural areas using machine learning. *Journal of Infrastructure Policy and Development*, 8(8), 5938.
<https://doi.org/10.24294/jipd.v8i8.5938>
- Lingar Vandito, R., N. Lumban Tobing, A., Aisyah Diva Maha, R., Ali Vardan Aminudin, M., & Adinda Tasyifa, K. (2023). Implementasi Konsep Kesejahteraan, Kesetaraan dan Keadilan Sosial dalam Kebijakan Publik Provinsi DKI Jakarta pada Tahun 2022: Studi Kasus KLJ dan Relokasi Program Provinsi pada Pemukiman Kumuh. *Jurnal Ilmu Sosial Dan Ilmu Politik*, 5(1), 54–74.
<https://doi.org/10.56552/jisipol.v5i1.124>
- Mardiah, A., Defni, D., Lestari, A. H., Junaldi, J., & Ritmi, T. (2024). Classification of Population Data of Nagari Based on Economic Level Using The K-Nearest Neighbor Method. *International Journal of Advanced Science Computing and Engineering*, 6(1), 32–35.
<https://doi.org/10.62527/ijasce.6.1.191>
- Muhammad, Rino Indra Nainggolan, E. R., Putra, J. L., Sidik, S., Susafa'ati, S., & Radiyah, U. (2021). Implementasi Metode K-Nearest Neighbor Untuk Prediksi Penjualan Kemasan Skincare Pada Pt. Universal Jaya Perkasa | *Technologic*.
- Mukharyahya, Z. A., Astuti, Y. P., & Cahyani, O. N. (2025). Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia. *Edumatic: Jurnal Pendidikan Informatika*, 9(1), 119–128.
<https://doi.org/10.29408/edumatic.v9i1.29512>
- Muñetón-Santa, G., & Manrique-Ruiz, L. C. (2023). Predicting Multidimensional Poverty with Machine Learning Algorithms: An Open Data Source Approach Using Spatial Data. *Social Sciences*, 12(5), 296.
<https://doi.org/10.3390/socsci12050296>
- Nikmatun, Inna, & A., & Indra, W. (2019). Implementasi Data Mining Untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor. *Jurnal SIMETRIS*, 10(2), 421–432.
- Palma, B. K., Murdiansyah, D. T., & Astuti, W. (2021). Klasifikasi Teks Artikel Berita Hoaks Covid-19 dengan Menggunakan Algoritma K-Nearest Neighbor. *E-Proceeding of Engineering*, 8(5), 10637–10649.
- Putri, I. P. (2021). Analisis Performa Metode K-Nearest Neighbor (KNN) dan Crossvalidation pada Data Penyakit Cardiovascular. *Indonesian Journal of Data and Science*, 2(1), 21–28.
<https://doi.org/10.33096/ijodas.v2i1.25>
- Rivaldo, Vito Junivan, T. A. Y., & Pranoto, W. J. (2024). Perbaikan Akurasi Naïve Bayes dengan Chi-Square dan SMOTE Dalam Mengatasi High Dimensional dan Imbalanced Data Banjir. *Jurnal Media Informatika Budidarma*, 8(3), 1656.
<https://doi.org/10.30865/mib.v8i3.7886>