



Klasifikasi Kelayakan UMR 98 Kota di Indonesia Berdasarkan Harga Pangan Pokok Menggunakan K-Nearest Neighbor

Eliza Tata Amalia^{1*}, Farid Akbar Siregar²

^{1,2}Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Email: ^{1*}elizatataamaliaa@gmail.com, ²faridakbar@umsu.ac.id

Abstrak

Upah Minimum Regional (UMR) merupakan instrumen kebijakan pengupahan yang idealnya mencerminkan kemampuan daya beli pekerja terhadap kebutuhan pangan pokok, namun disparitas harga pangan antarwilayah di Indonesia menyebabkan nominal UMR yang sama dapat memiliki tingkat kelayakan yang berbeda. Penelitian ini bertujuan membangun model klasifikasi untuk menilai tingkat kelayakan UMR pada 98 kota di Indonesia berdasarkan harga tujuh komoditas pangan pokok (beras, minyak goreng, telur, daging ayam, gula pasir, cabai, dan UMR) menggunakan algoritma K-Nearest Neighbor (KNN). Data dikumpulkan dari Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional/Badan Pangan Nasional, Badan Pusat Statistik, dan Kementerian Ketenagakerjaan Republik Indonesia periode 2025-2026, kemudian diproses melalui tahapan Knowledge Discovery in Database meliputi seleksi, pembersihan, transformasi Min-Max Normalization, penerapan algoritma KNN, dan evaluasi menggunakan Confusion Matrix dengan pembagian data latih dan data uji sebesar 80:20. Pengujian terhadap beberapa nilai K menunjukkan bahwa K=3 memberikan kinerja terbaik dengan akurasi 80,00%, presisi 100,00%, recall 42,86%, dan F1-Score 60,00%. Nilai presisi sempurna mengindikasikan model tidak pernah salah mengklasifikasikan kota berkategori Tidak Layak sebagai Layak, sementara nilai recall yang relatif rendah disebabkan oleh ketidakseimbangan kelas pada data latih yang didominasi kategori Tidak Layak. Model klasifikasi ini diimplementasikan dalam aplikasi berbasis Python Tkinter yang dapat digunakan sebagai alat bantu evaluasi kelayakan UMR oleh pemangku kebijakan daerah.

Kata Kunci: K-Nearest Neighbor, Upah Minimum Regional, Harga Pangan Pokok, Klasifikasi, Data Mining.

PENDAHULUAN

Upah Minimum Regional (UMR) merupakan instrumen kebijakan pengupahan yang ditetapkan pemerintah daerah untuk melindungi daya beli pekerja, dengan asumsi dasar bahwa nominal UMR mampu mencukupi kebutuhan hidup layak, termasuk kebutuhan pangan pokok. Namun demikian, disparitas harga pangan pokok antarwilayah di Indonesia menyebabkan nominal UMR yang relatif tinggi pada suatu kota belum tentu mencerminkan kelayakan yang sebanding apabila dibandingkan dengan beban biaya pangan di kota tersebut. Kertiasih (2017) menunjukkan bahwa tingkat inflasi dan pertumbuhan ekonomi daerah berpengaruh signifikan terhadap penetapan besaran UMR, sementara Febrianti dkk. (2026) membuktikan bahwa pertumbuhan ekonomi, inflasi, angkatan kerja, dan kemiskinan secara simultan memengaruhi besaran UMR di tingkat provinsi. Temuan tersebut menegaskan bahwa UMR merupakan besaran yang dipengaruhi banyak faktor ekonomi makro, sehingga penilaian kelayakannya terhadap kebutuhan riil masyarakat memerlukan pendekatan analisis berbasis data yang komprehensif.

Penelitian mengenai faktor-faktor penentu UMR juga telah dilakukan dengan pendekatan statistik. Saldi dkk. (2025) menganalisis faktor-faktor penentu UMR menggunakan regresi linear berganda dan menemukan bahwa variabel ekonomi makro memiliki kontribusi yang bervariasi antarwilayah, sedangkan Fahira dkk. (2025) melakukan evaluasi UMR menggunakan pendekatan statistik deskriptif dan inferensial untuk melihat kesenjangan penetapan upah antarkota. Kedua penelitian tersebut masih berfokus pada analisis hubungan antarvariabel ekonomi, dan belum menghasilkan model klasifikasi yang secara langsung dapat mengategorikan status kelayakan UMR suatu kota berdasarkan harga pangan pokok secara otomatis dan terukur.

Pendekatan klasifikasi berbasis algoritma K-Nearest Neighbor (KNN) telah banyak diterapkan pada berbagai persoalan sosial dan ekonomi karena kesederhanaan logika dan akurasinya yang kompetitif. Pahrudin dan Harianto (2022) menerapkan KNN untuk mengklasifikasikan warga penerima bantuan sosial,

sementara Triayudi (2023) dan Nurdin (2024) masing-masing menggunakan algoritma yang sama untuk mengklasifikasikan penerima dana bantuan sosial dan memprediksi masyarakat kurang mampu. Pada ranah pendidikan, Muhaimin dkk. (2024) menerapkan KNN untuk mengklasifikasikan prestasi akademik siswa berdasarkan nilai rapor dan kedisiplinan, sedangkan Affandi (2024) memanfaatkan KNN untuk klasifikasi sentimen pada media sosial. Penelitian-penelitian tersebut membuktikan fleksibilitas algoritma KNN dalam menangani berbagai kasus klasifikasi berbasis data numerik maupun tekstual, namun penerapannya pada kasus klasifikasi kelayakan UMR berdasarkan harga pangan pokok lintas kota di Indonesia belum banyak dieksplorasi.

Penerapan algoritma klasifikasi berbasis machine learning untuk persoalan penilaian kelayakan dan risiko juga ditemukan pada lingkup Indonesian Journal of Multidisciplinary Scientific Studies (IJOMSS), misalnya Al Hakim dan Al-Khowarizmi (2026) yang menerapkan algoritma Random Forest untuk mengklasifikasikan tingkat risiko kardiovaskular berdasarkan pola aktivitas harian dan indikator kesehatan, dengan tahapan praproses data dan evaluasi model berbasis Confusion Matrix yang serupa dengan kerangka kerja klasifikasi pada umumnya. Penelitian tersebut memperkuat relevansi pendekatan klasifikasi berbasis fitur numerik sebagai alat bantu pengambilan keputusan pada berbagai domain, termasuk domain kebijakan pengupahan yang menjadi fokus penelitian ini.

Berdasarkan uraian tersebut, penelitian-penelitian terdahulu umumnya berfokus pada analisis hubungan statistik antarvariabel penentu UMR atau pada penerapan algoritma KNN untuk klasifikasi pada domain sosial dan pendidikan, sedangkan penelitian yang secara khusus mengintegrasikan data harga pangan pokok dan besaran UMR ke dalam model klasifikasi berbasis KNN untuk menilai kelayakan UMR antarkota di Indonesia masih terbatas. Penelitian ini bertujuan untuk membangun dan mengevaluasi model klasifikasi kelayakan UMR pada 98 kota di Indonesia berdasarkan harga tujuh komoditas pangan pokok menggunakan algoritma K-Nearest Neighbor, menerapkan tahapan Min-Max Normalization untuk menyetarakan skala fitur, mengevaluasi kinerja model menggunakan Confusion Matrix pada berbagai nilai hyperparameter K, serta mengimplementasikan model ke dalam aplikasi berbasis Python Tkinter yang dapat digunakan sebagai alat bantu evaluasi kebijakan pengupahan oleh pemangku kepentingan daerah.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental yang menerapkan tahapan Knowledge Discovery in Database (KDD) untuk membangun model klasifikasi kelayakan Upah Minimum Regional (UMR) berbasis algoritma K-Nearest Neighbor (KNN). Tahapan KDD yang diterapkan meliputi selection, preprocessing (cleaning dan transformation), data mining, dan interpretation/evaluation.

Lokasi dan Waktu Penelitian

Penelitian dilaksanakan secara daring melalui pengumpulan data sekunder dan pengembangan sistem pada lingkungan komputer peneliti, dengan rangkaian kegiatan penelitian dilaksanakan pada tahun 2026.

Sumber dan Teknik Pengumpulan Data

Data penelitian merupakan data sekunder yang dikumpulkan melalui teknik studi dokumentasi dari tiga sumber resmi, yaitu Pusat Informasi Harga Pangan Strategis (PIHPS) Nasional/Badan Pangan Nasional (Bapanas) untuk data harga komoditas pangan pokok, Badan Pusat Statistik (BPS) untuk data pendukung wilayah administratif, serta Kementerian Ketenagakerjaan Republik Indonesia (Kemnaker RI) untuk data besaran UMR masing-masing kota, mencakup periode 2025-2026.

Atribut dan Variabel Data

Dataset penelitian terdiri atas 98 kota administratif di Indonesia dengan tujuh atribut numerik, yaitu harga beras, minyak goreng, telur ayam, daging ayam, gula pasir, dan cabai, serta besaran UMR kota yang bersangkutan. Ketujuh atribut tersebut digunakan sebagai fitur dalam perhitungan jarak Euclidean, sedangkan label kelas target berupa kategori kelayakan UMR yang terdiri atas dua kelas, yaitu "Layak" dan "Tidak Layak", yang ditentukan berdasarkan perbandingan antara besaran UMR dengan total kebutuhan biaya pangan pokok masing-masing kota.

Tahapan Praproses Data

Tahap preprocessing diawali dengan pembersihan data untuk memastikan tidak terdapat nilai yang hilang (missing value) maupun duplikasi data pada 98 baris dataset. Selanjutnya dilakukan transformasi Min-Max Normalization terhadap seluruh atribut numerik untuk menyetarakan rentang skala data, mengingat atribut UMR memiliki rentang nominal jutaan rupiah sedangkan atribut harga pangan memiliki rentang nominal ribuan rupiah, sehingga tanpa normalisasi atribut UMR akan mendominasi perhitungan jarak Euclidean secara tidak proporsional. Transformasi Min-Max Normalization dirumuskan sebagai berikut:

$$x' = (x - x_{min}) / (x_{max} - x_{min})$$

dengan x' adalah nilai hasil normalisasi, x adalah nilai atribut asli, x_{min} adalah nilai minimum, dan x_{max} adalah nilai maksimum pada atribut tersebut. Hasil transformasi memetakan seluruh nilai fitur mentah ke dalam domain tertutup interval 0 hingga 1, sehingga ketujuh dimensi fitur memiliki bobot kontribusi yang seimbang terhadap perhitungan jarak.

Implementasi Algoritma K-Nearest Neighbor

Algoritma KNN diimplementasikan menggunakan pustaka Pandas dan NumPy untuk manipulasi data, serta Matplotlib dan Seaborn untuk visualisasi hasil. Proses klasifikasi dilakukan dengan menghitung jarak Euclidean antara setiap data uji terhadap seluruh data latih menggunakan persamaan:

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

dengan $d(x, y)$ adalah jarak Euclidean antara titik data uji x dan data latih y , serta x_i dan y_i adalah nilai atribut ke- i dari masing-masing titik data pada ketujuh dimensi fitur yang telah dinormalisasi. Setelah jarak dihitung terhadap seluruh data latih, hasil jarak diurutkan dari yang terkecil hingga terbesar, kemudian sejumlah K tetangga terdekat dipilih untuk menentukan kelas prediksi melalui mekanisme voting mayoritas (majority voting).

Populasi dan Sampel Penelitian

Populasi penelitian mencakup seluruh 98 kota administratif di Indonesia. Dataset dibagi menjadi data latih dan data uji dengan proporsi 80:20, menghasilkan 78 kota sebagai data latih dan 20 kota sebagai data uji. Pembagian data dilakukan secara acak dengan mempertahankan representasi sebaran geografis kota dari berbagai pulau di Indonesia.

Teknik Evaluasi Model

Evaluasi kinerja model klasifikasi dilakukan menggunakan Confusion Matrix yang membandingkan kelas aktual dengan kelas hasil prediksi pada data uji, menghasilkan empat komponen yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan keempat komponen tersebut, dihitung empat metrik evaluasi, yaitu Akurasi, Presisi, Recall, dan F1-Score, dengan persamaan sebagai berikut:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Presisi} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-Score} = 2 \times (\text{Presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall})$$

Selain evaluasi pada $K=3$, dilakukan pula pengujian sensitivitas model terhadap variasi nilai hyperparameter K ($K=3, 5, 7, 9, 15$, dan 17) untuk mengetahui pengaruh jumlah tetangga terhadap kinerja klasifikasi.

Alat dan Bahan Penelitian

Penelitian menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Matplotlib, dan Seaborn untuk pengolahan data dan implementasi algoritma, serta pustaka Tkinter untuk pengembangan antarmuka pengguna grafis (Graphical User Interface) berarsitektur multi-frame yang terdiri atas modul autentikasi login, dashboard ringkasan data, pengelolaan data kota dan harga pangan, modul analisis KNN dengan parameter K yang dapat disesuaikan pengguna, serta modul pelaporan hasil klasifikasi dan evaluasi model.

HASIL DAN PEMBAHASAN

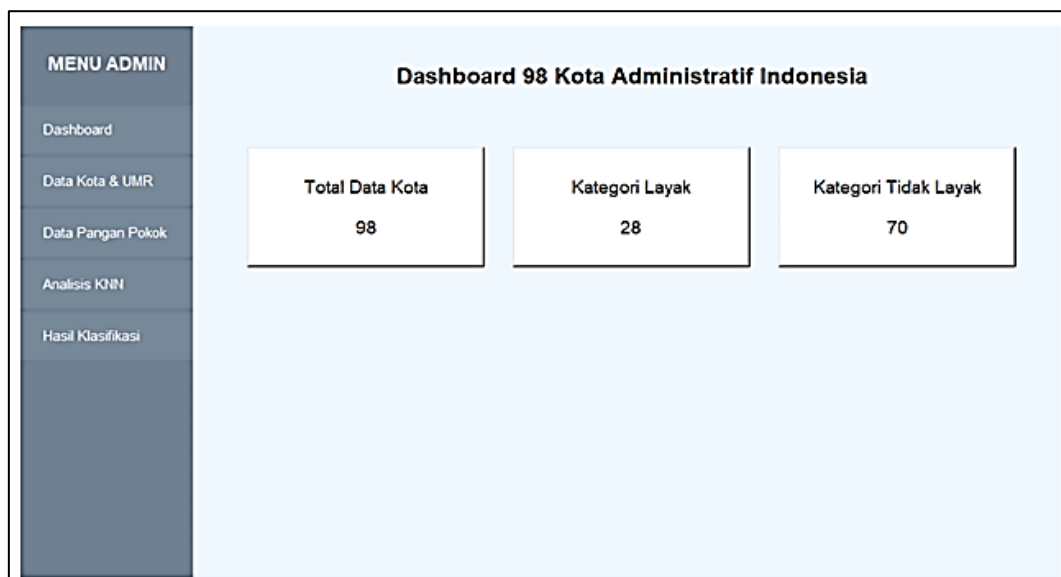
Deskripsi Dataset dan Pembagian Data

Dataset penelitian terdiri atas 98 kota administratif di Indonesia dengan tujuh atribut numerik (harga beras, minyak goreng, telur, daging ayam, gula pasir, cabai, dan UMR) serta satu atribut kelas target kelayakan. Berdasarkan hasil inspeksi data, distribusi kelas pada dataset menunjukkan ketidakseimbangan (class imbalance), yaitu 28 kota berkategori "Layak" dan 70 kota berkategori "Tidak Layak". Dataset selanjutnya dibagi menjadi 78 kota sebagai data latih dan 20 kota sebagai data uji dengan proporsi 80:20. Seluruh atribut numerik ditransformasikan menggunakan Min-Max Normalization sehingga bernilai pada rentang 0 hingga 1 sebelum digunakan dalam perhitungan jarak Euclidean pada algoritma K-Nearest Neighbor (KNN).

Implementasi Antarmuka Sistem

Model klasifikasi yang dihasilkan diimplementasikan ke dalam aplikasi desktop berbasis Python Tkinter dengan arsitektur multi-frame, terdiri atas halaman login, dashboard ringkasan data, pengelolaan data kota dan UMR, pengelolaan data harga pangan pokok, modul analisis KNN dengan parameter K yang dapat disesuaikan pengguna, serta halaman pelaporan hasil klasifikasi dan evaluasi model. Tangkapan layar masing-masing modul disajikan pada Gambar 1 hingga Gambar 6.

Gambar 1. Halaman Login Sistem



Gambar 2. Antarmuka Dashboard Ringkasan 98 Kota

Kota	UMR	Label Aktual
Banda Aceh	Rp 3,460,672	Tidak Layak
Sabang	Rp 3,460,672	Tidak Layak
Lhokseumawe	Rp 3,460,672	Tidak Layak
Langsa	Rp 3,460,672	Tidak Layak
Subulussalam	Rp 3,460,672	Tidak Layak
Medan	Rp 3,769,082	Tidak Layak
Pematangsiantar	Rp 2,890,000	Tidak Layak
Sibolga	Rp 3,195,000	Tidak Layak
Tanjungbalai	Rp 3,059,000	Tidak Layak
Binjai	Rp 2,890,000	Tidak Layak
Tebing Tinggi	Rp 2,890,000	Tidak Layak
Padangsidempuan	Rp 3,189,000	Tidak Layak
Gunungsitoli	Rp 2,890,000	Tidak Layak
Padang	Rp 3,124,000	Tidak Layak
Bukittinggi	Rp 3,124,000	Tidak Layak
Padang Panjang	Rp 3,124,000	Tidak Layak
Sawahlunto	Rp 3,124,000	Tidak Layak
Solok	Rp 3,124,000	Tidak Layak
Pariaman	Rp 3,124,000	Tidak Layak
Payakumbuh	Rp 3,124,000	Tidak Layak
Pekanbaru	Rp 3,451,524	Tidak Layak
Dumai	Rp 3,067,295	Layak
Jambi	Rp 3,353,591	Tidak Layak
Sungaijenuh	Rp 3,353,591	Tidak Layak
Palembang	Rp 3,677,591	Layak
Prabumulih	Rp 3,677,591	Tidak Layak
Pagar Alam	Rp 3,456,874	Tidak Layak
Lubuklinggau	Rp 3,456,874	Tidak Layak
Bengkulu	Rp 2,807,020	Tidak Layak
Bandar Lampung	Rp 3,103,631	Tidak Layak
Metre	Rp 2,716,000	Tidak Layak

Gambar 3. Antarmuka Pengelolaan Data Kota dan UMR

MENU ADMIN	Kelola Data Harga Pangan Pokok					
	Kota	Beras	Minyak	Telur	Ayam	Gula
Dashboard	Banda Aceh	15000	22000	32000	45000	18000
Data Kota & UMR	Selang	13300	22300	32300	46000	19500
	Lhotseumawe	14800	21500	31500	44000	18500
Data Pangan Pokok	Lampung	14700	21000	31000	43500	18000
	Subulussalam	15200	22000	32000	45500	19000
Analisis KNN	Medan	15100	21450	32050	45150	18100
	Pematangsiantar	14900	21000	31500	39000	17800
Hasil Klasifikasi	Sibolga	15300	21800	32500	41000	18500
	Tanjungbalai	15000	21200	31800	40000	18000
Logout	Binjai	14900	21000	31500	39500	17800
	Tebing Tinggi	14950	21100	31600	39500	17900
	Pedangsidempuan	15200	21500	32000	40500	18200
	Gumutukali	15800	22500	33500	43000	19000
	Padang	16000	20500	30000	38000	18500
	Bukittinggi	16200	20000	29500	37000	18000
	Padang Panjang	16300	20200	29800	37500	18200
	Sawah Lunto	16100	20000	29500	37000	18000
	Solok	16000	19800	29000	36500	17800
	Pariaman	16400	20500	29900	37500	18300
	Payakumbuh	16200	20000	29500	37000	18000
	Pelamban	16000	20000	31000	37000	18000
	Dumai	16350	19800	31750	32700	17750
	Jambi	15500	19500	29000	36000	17500
	Sungapenuh	15800	20000	29500	37000	18000
	Palembang	15200	19800	29500	35000	17500
	Prabumulih	15500	20800	30400	36500	17950
	Pagar Alam	15400	20500	30000	37500	17500
	Lubuklinggau	15800	20900	30500	38000	17800
	Bengkulu	15800	20500	31000	38000	18000
	Bandar Lampung	14900	19000	29000	35000	17000
	Metro	14800	19900	28500	34500	16800

Gambar 4. Antarmuka Pengelolaan Data Harga Pangan Pokok

MENU ADMIN	Proses Klasifikasi K-Nearest Neighbor	
	Masukkan Nilai K:	
Dashboard	3	
Data Kota & UMR		
Data Pangan Pokok		
Analisis KNN		
Hasil Klasifikasi		
Logout	<button>JALANKAN ANALISIS</button>	

Gambar 5. Antarmuka Modul Analisis KNN

MENU ADMIN	Laporan Hasil Klasifikasi & Evaluasi Model		
	Kota	Kelas Aktual	Hasil Prediksi KNN
Dashboard	Samarinda	Tidak Layak	Tidak Layak
Data Kota & UMR	Bontang	Layak	Tidak Layak
	Tarakan	Layak	Layak
Data Pangan Pokok	Manado	Layak	Layak
	Tomohon	Layak	Tidak Layak
Analisis KNN	Sitang	Layak	Tidak Layak
	Kotamobagu	Layak	Layak
Hasil Klasifikasi	Pala	Tidak Layak	Tidak Layak
	Makassar	Layak	Tidak Layak
Logout	Pangare	Tidak Layak	Tidak Layak
	Palopo	Tidak Layak	Tidak Layak
	Kendari	Tidak Layak	Tidak Layak
	Baubau	Tidak Layak	Tidak Layak
	Gorontalo	Tidak Layak	Tidak Layak
	Ambon	Tidak Layak	Tidak Layak
	Tual	Tidak Layak	Tidak Layak
	Ternte	Tidak Layak	Tidak Layak
	Tidore Kepulauan	Tidak Layak	Tidak Layak
	Jayapura	Tidak Layak	Tidak Layak
	Sorong	Tidak Layak	Tidak Layak
		Akurasi: 80.00%	Presisi: 100.00%
		Recall: 42.86%	F1-Score: 60.00%

Gambar 6. Antarmuka Laporan Hasil Klasifikasi dan Evaluasi Model

Hasil Klasifikasi dan Evaluasi Model

Proses klasifikasi dengan nilai K=3 menghasilkan matriks konfusi sebagaimana disajikan pada Tabel 1 di bawah ini.

Tabel 1. Matriks Konfusi Hasil Klasifikasi (K=3)

	Prediksi: Layak	Prediksi: Tidak Layak
Aktual: Layak	TP = 3	FN = 4
Aktual: Tidak Layak	FP = 0	TN = 13

Sumber: Hasil pengolahan data (2026)

Berdasarkan matriks konfusi tersebut, dihitung empat metrik evaluasi sebagaimana disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Klasifikasi (K=3)

Metrik	Nilai
Akurasi	80,00%
Presisi	100,00%
Recall	42,86%
F1-Score	60,00%

Sumber: Hasil pengolahan data (2026)

Dari 20 kota data uji, terdapat 7 kota yang secara aktual berkategori "Layak", terdiri atas 3 kota yang berhasil diklasifikasikan dengan tepat (True Positive), yaitu Tarakan, Manado, dan Kotamobagu, serta 4 kota yang mengalami kesalahan klasifikasi (False Negative), yaitu Bontang, Tomohon, Bitung, dan Makassar, yang secara aktual berkategori "Layak" namun diprediksi sebagai "Tidak Layak". Seluruh 13 kota yang secara aktual berkategori "Tidak Layak" berhasil diklasifikasikan dengan tepat tanpa satu pun kesalahan (False Positive = 0), sebagaimana dirangkum pada Tabel 3.

Tabel 3. Kota Uji Berkategori Aktual "Layak" dan Status Prediksinya

Kota	Kelas Aktual	Hasil Prediksi KNN	Status
Tarakan	Layak	Layak	Sesuai (TP)
Manado	Layak	Layak	Sesuai (TP)
Kotamobagu	Layak	Layak	Sesuai (TP)
Bontang	Layak	Tidak Layak	Mismatch (FN)
Tomohon	Layak	Tidak Layak	Mismatch (FN)
Bitung	Layak	Tidak Layak	Mismatch (FN)
Makassar	Layak	Tidak Layak	Mismatch (FN)

Sumber: Hasil pengolahan data (2026)

Simulasi Perhitungan Tiga Tetangga Terdekat

Untuk memverifikasi konsistensi logika algoritma, dilakukan simulasi perhitungan manual jarak Euclidean pada dua kota uji dengan K=3, yaitu Kota Tarakan (True Positive) dan Kota Bontang (False Negative), sebagaimana disajikan pada Tabel 4.

Tabel 4. Simulasi Perhitungan Tiga Tetangga Terdekat (K=3)

Kota Uji	Peringkat	Kota Latih Terdekat	Kelas	Jarak Euclidean
Tarakan	1	Balikpapan	Layak	0,3553
	2	Palangkaraya	Tidak Layak	0,3970
	3	Pangkal Pinang	Layak	0,4686
Bontang	1	Balikpapan	Layak	0,1163
	2	Palangkaraya	Tidak Layak	0,2383
	3	Pontianak	Tidak Layak	0,2586

Sumber: Hasil pengolahan data (2026)

Pada Kota Tarakan, dua dari tiga tetangga terdekat berlabel "Layak" (Balikpapan dan Pangkal Pinang), sehingga mekanisme voting mayoritas menetapkan prediksi "Layak" yang sesuai dengan label aktual. Pada Kota Bontang, hanya satu dari tiga tetangga terdekat yang berlabel "Layak" (Balikpapan), sedangkan dua tetangga lainnya, yaitu Palangkaraya dan Pontianak, berlabel "Tidak Layak". Karena suara "Tidak Layak" lebih banyak, Kota Bontang diprediksi "Tidak Layak" meskipun label aktualnya "Layak", sehingga tergolong False Negative. Simulasi ini menunjukkan bahwa kesalahan klasifikasi pada model bukan disebabkan oleh kesalahan implementasi algoritma, melainkan oleh kerapatan lokal data latih yang berlabel "Tidak Layak" di sekitar titik koordinat kota-kota yang mengalami mismatch.

Pengujian Sensitivitas Nilai Hyperparameter K

Pengujian dilakukan terhadap enam nilai K (3, 5, 7, 9, 15, dan 17) untuk mengetahui pengaruh jumlah tetangga terhadap kinerja klasifikasi, sebagaimana disajikan pada Tabel 5.

Tabel 5. Hasil Pengujian Nilai K

Nilai K	Akurasi	Presisi	Recall	F1-Score
3	80,00%	100,00%	42,86%	60,00%
5	75,00%	100,00%	28,57%	44,44%
7	70,00%	100,00%	14,29%	25,00%
9	70,00%	100,00%	14,29%	25,00%
15	65,00%	0,00%	0,00%	0,00%
17	65,00%	0,00%	0,00%	0,00%

Sumber: Hasil pengolahan data (2026)

Hasil pengujian menunjukkan bahwa nilai akurasi tertinggi diperoleh pada K=3 sebesar 80,00%, dan terus mengalami penurunan seiring bertambahnya nilai K, hingga mencapai nilai terendah sebesar 65,00% pada K=15 dan K=17. Nilai presisi tercatat sempurna (100,00%) pada rentang K=3 hingga K=9, namun turun drastis menjadi 0,00% pada K=15 dan K=17 karena model pada nilai K yang besar tidak lagi menghasilkan satu pun prediksi "Layak" yang benar akibat voting mayoritas yang semakin didominasi oleh kelas "Tidak Layak". Nilai F1-Score tertinggi diperoleh pada K=3 sebesar 60,00%, menunjukkan bahwa K=3 merupakan parameter paling optimal untuk digunakan pada klasifikasi kelayakan UMR dengan algoritma KNN pada dataset ini.

Pembahasan

Nilai presisi sempurna (100,00%) dengan False Positive bernilai nol merupakan keunggulan utama model yang dibangun. Tidak ada satu pun kota berkategori "Tidak Layak" yang diprediksi sebagai "Layak", yang mengindikasikan bahwa model bersifat konservatif dan memiliki kepastian tinggi: setiap kali model memberikan luaran prediksi "Layak", hasil tersebut dapat dipertanggungjawabkan kesesuaiannya dengan kondisi aktual. Dari perspektif kebijakan publik, karakteristik ini penting untuk mencegah penetapan kelayakan yang keliru, yaitu mencegah pemerintah daerah terjebak pada asumsi palsu bahwa UMR suatu kota telah mencukupi kebutuhan pangan pokok pekerjaanya.

Kendala utama pemodelan terletak pada nilai recall yang relatif rendah (42,86%) akibat ditemukannya empat kasus False Negative pada data uji. Faktor utama yang memicu timbulnya False Negative adalah ketidakseimbangan kelas (class imbalance) pada dataset, di mana populasi kota "Tidak Layak" (70 kota) jauh lebih dominan dibandingkan kota "Layak" (28 kota). Ketidakseimbangan proporsi pada data latih menggeser batas keputusan (decision boundary) mendekati area kelas positif, sehingga sampel kelas minoritas lebih mudah terdesak oleh kelas mayoritas pada mekanisme voting K-Nearest Neighbor. Hal ini sejalan dengan temuan Perangin-angin dkk. (2020) yang juga melaporkan sensitivitas algoritma K-Nearest Neighbor terhadap ketidakseimbangan distribusi kelas pada data latih dan merekomendasikan pendekatan level data untuk menanganinya. Untuk mengatasi kendala tersebut, penelitian lanjutan dapat menerapkan teknik penyeimbangan data seperti Synthetic Minority Over-sampling Technique (SMOTE) atau pendekatan algoritmik seperti Distance-Weighted KNN, yang memberikan bobot lebih besar pada tetangga dengan jarak lebih dekat sehingga pengaruh kelas mayoritas pada area perbatasan dapat dikurangi.

Pengujian sensitivitas terhadap nilai K memperkuat temuan tersebut, karena recall dan F1-Score konsisten menurun seiring bertambahnya nilai K, yang menunjukkan bahwa semakin besar jumlah tetangga yang dilibatkan dalam voting, semakin besar pula pengaruh kelas mayoritas "Tidak Layak" terhadap keputusan klasifikasi. Meskipun memiliki keterbatasan pada recall, pencapaian akurasi keseluruhan sebesar 80,00% pada K=3 tergolong baik untuk kasus klasifikasi dengan dataset berskala menengah dan kelas yang tidak seimbang, sebagaimana juga dilaporkan pada berbagai penerapan algoritma KNN untuk klasifikasi kelayakan penerima bantuan sosial yang turut menghadapi kendala serupa (Triayudi, 2023; Pahrudin & Harianto, 2022; Rosyada dkk., 2025).

Hasil penelitian ini menegaskan bahwa kombinasi tujuh fitur kuantitatif harga pangan pokok dan UMR, setelah melalui tahapan Min-Max Normalization, memiliki daya pembeda yang cukup kuat dalam mengklasifikasikan kelayakan UMR antarkota. Model yang diimplementasikan ke dalam aplikasi berbasis Python Tkinter terbukti dapat dioperasikan secara interaktif melalui parameter K yang dapat disesuaikan pengguna, sehingga berpotensi dimanfaatkan sebagai alat bantu evaluasi awal bagi pemangku kebijakan daerah dalam menilai kecukupan UMR terhadap kebutuhan pangan pokok masyarakat, dengan catatan bahwa keterbatasan recall akibat ketidakseimbangan kelas perlu diperhitungkan dalam interpretasi hasil, khususnya pada kota-kota yang berada di area perbatasan antarkelas.

KESIMPULAN

Penelitian ini berhasil membangun model klasifikasi kelayakan Upah Minimum Regional (UMR) pada 98 kota di Indonesia berdasarkan harga tujuh komoditas pangan pokok menggunakan algoritma K-Nearest Neighbor. Tahapan Min-Max Normalization terbukti esensial dalam menyetarakan rentang skala atribut

harga pangan dan UMR yang berbeda jauh, sehingga mencegah dominasi atribut bernilai nominal besar pada perhitungan jarak Euclidean. Hasil pengujian dengan $K=3$ menghasilkan kinerja terbaik, yaitu akurasi 80,00%, presisi 100,00%, recall 42,86%, dan F1-Score 60,00%. Nilai presisi yang sempurna ($FP=0$) menunjukkan bahwa model tidak pernah salah mengklasifikasikan kota berkategori Tidak Layak sebagai Layak, sehingga bersifat konservatif dan andal digunakan sebagai alat bantu evaluasi awal kelayakan UMR. Sebaliknya, nilai recall yang relatif rendah disebabkan oleh ketidakseimbangan kelas pada data latih yang didominasi kategori Tidak Layak, sehingga menggeser batas keputusan dan menyebabkan sebagian kota berkategori Layak yaitu Bontang, Tomohon, Bitung, dan Makassar—terklasifikasi keliru sebagai Tidak Layak. Pengujian sensitivitas terhadap nilai K menunjukkan bahwa kinerja model menurun secara konsisten seiring bertambahnya nilai K , dengan $K=3$ tetap menjadi parameter paling optimal berdasarkan nilai F1-Score tertinggi. Model klasifikasi yang dihasilkan telah diimplementasikan ke dalam aplikasi berbasis Python Tkinter dengan arsitektur multi-frame yang memungkinkan pengguna mengoperasikan analisis klasifikasi secara interaktif. Penelitian selanjutnya disarankan menerapkan teknik penyeimbangan data seperti Synthetic Minority Over-sampling Technique (SMOTE) atau pendekatan Distance-Weighted K-Nearest Neighbor untuk meningkatkan nilai recall, serta menambahkan variabel makroekonomi lain seperti inflasi daerah, Indeks Harga Konsumen, dan komponen Kebutuhan Hidup Layak non-pangan agar model klasifikasi lebih representatif terhadap kondisi riil kelayakan hidup di masing-masing kota.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, atas dukungan fasilitas dan bimbingan akademik selama penelitian ini dilaksanakan.

DAFTAR PUSTAKA

- Affandi, R. (2024). *Analisis Sentimen Komentar Akun Instagram Public Figure Politik Terkait Pembatalan Piala Dunia U-20 dengan Algoritma K-Nearest Neighbors (KNN)* [Skripsi, Universitas Medan Area]. <http://repositori.uma.ac.id/handle/123456789/24401>
- Al Hakim, A. H., & Al-Khowarizmi. (2026). Klasifikasi Tingkat Risiko Kardiovaskular Berbasis Pola Aktivitas Harian Menggunakan Random Forest. *Indonesian Journal of Multidisciplinary Scientific Studies (IJOMSS)*, 4(6), 449–454. <https://doi.org/10.33151/ijomss.v4i6.1132>
- Annisa, K., Ginting, B. S., & Syari, M. A. (2022). Penerapan Data Mining Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode Clustering Pada PDAM Langkat. *Algoritma: Jurnal Ilmu Komputer dan Informatika*, 6(1), 1–8. <https://doi.org/10.30829/algoritma.v6i1.11624>
- Aryani, Y., Aqil, I., & Paramita, B. (2024). Penerapan Unified Modeling Language (UML) pada Digitalisasi Sistem Informasi Perpustakaan. *DIGITECH: Jurnal Teknologi Informasi*, 4(2), 1032–1040.
- Fahira, D., Simanjuntak, L. F., Zahira, T. A., & Pangaribuan, H. S. (2025). Evaluasi Upah Minimum Regional (UMR) dengan Statistik. *ARU Journal (Accounting Research Unit)*, 6(2), 126–135. <https://doi.org/10.30598/arujournalvol6iss2pp126-135>
- Febrianti, V. A., Purnomo, S. D., & Fatmawati, A. (2026). Pengaruh Pertumbuhan Ekonomi, Inflasi, Angkatan Kerja dan Kemiskinan terhadap Upah Minimum Regional di Provinsi Jawa Barat. *Jurnal Ilmiah Universitas Batanghari Jambi*, 26(2), 707–714. <https://doi.org/10.33087/jiubj.v26i2.6569>
- Kertiasih, L. (2017). Pengaruh Tingkat Inflasi dan Tingkat Pertumbuhan Ekonomi terhadap Upah Minimum Regional (UMR) di Provinsi Bali pada Tahun 2008-2015. *Jurnal Pendidikan Ekonomi Undiksha*, 9(1), 148–157.
- Muhaimin, A., Hariyadi, M. A., & Imamudin, M. I. (2024). Klasifikasi Prestasi Akademik Siswa Berdasarkan Nilai Rapor dan Kedisiplinan dengan Metode K-Nearest Neighbor. *JIKOMSI: Jurnal Ilmu Komputer dan Sistem Informasi*, 7(1), 193–202. <https://doi.org/10.55338/jikomsi.v7i1.2865>
- Nuridin, N. (2024). Analisa Data Mining dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor. *JITET: Jurnal Informatika dan Teknik Elektro Terapan*, 12(2), 1090–1098. <https://doi.org/10.23960/jitet.v12i2.4131>
- Pahrudin, P., & Harianto, K. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial. *BITS: Building of Informatics, Technology and Science*, 4(3), 1241–1245. <https://doi.org/10.47065/bits.v4i3.2276>

- Perangin-angin, R., Harianja, E. J. G., & Jaya, I. K. (2020). Pendekatan Level Data untuk Menangani Ketidakseimbangan Data Menggunakan Algoritma K-Nearest Neighbor. *Jurnal TIMES*, 9(1), 22–32. <https://doi.org/10.51351/jtm.9.1.2020615>
- Romzi, M., & Kurniawan, B. (2021). Implementasi Pemrograman Python Menggunakan Visual Studio Code. *Jurnal Informatika dan Komputer*, 11(2), 1–9.
- Saldi, A. A., Manik, I. T., Junita, I., Sitorus, N. E., Sagala, R., & Pinem, Y. N. (2025). Analisis Faktor-Faktor Penentu Upah Minimum Regional Menggunakan Regresi Linear Berganda. *Griya Journal of Mathematics Education and Application*, 5(4), 1930–1938. <https://doi.org/10.29303/griya.v5i4.879>
- Triayudi, A. (2023). Penerapan Data Mining Untuk Klasifikasi Penerima Dana Bantuan Sosial Dengan Menggunakan Algoritma K-Nearest Neighbor. *BITS: Building of Informatics, Technology and Science*, 5(2), 532–542. <https://doi.org/10.47065/bits.v5i2.3972>
- Ustadatin, F., Muqtadir, A., & Arifia, A. (2023). Implementasi Metode Weighted Moving Average (WMA) Pada Prediksi Harga Bahan Pokok. *Komputika: Jurnal Sistem Komputer*, 12(2), 203–210. <https://doi.org/10.34010/komputika.v12i2.10304>