KAMPUS AKADEMIK PUBLISING
Jurnal Sains Student Research (JSSR)
Vol.4, No.6 Desember 2026
e-ISSN: 3025-9851, p-ISSN : 3025-986X, Hal 42-55
DOI : https://doi.org/10.61722/jssr.v4i6.12770

PENERAPAN METODE K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI TINGKAT KESEJAHTERAAN PENDUDUK DI DESA NATAM KEC. BADAR KUTACANE ACEH TENGGARA

Suci Handayani [1*], Hevlie Winda Nazry S[2]
[1,2] Universitas Muhammadiyah Sumatera Utara
*Penulis Korespondensi: sucihan07@gmail.com, hevliewindanazry@umsu.ac.id

Abstract. *Determining the welfare level of residents is an important aspect in assisting village governments with decision-making, particularly in ensuring that social assistance is distributed to the appropriate beneficiaries. However, the process of classifying the welfare level of residents in Natam Village, Badar District, Southeast Aceh Regency is still carried out manually, making it time-consuming, prone to subjectivity, and unable to optimally utilize the available socio-economic data. This study aims to implement the K-Nearest Neighbor (KNN) method to classify the welfare level of residents based on six socio-economic variables, namely income, number of dependents, education level, occupation, residential status, and asset ownership. The study employed a quantitative approach with an experimental method, while the system was developed using the Waterfall model. The dataset consisted of 300 household records, which were divided into 249 training data (83%) and 51 testing data (17%). Before the classification process, the data underwent preprocessing, including categorical data transformation into numerical values and normalization using the Min-Max Scaling method. The classification process was performed using Euclidean Distance with K values of 3, 5, and 7, and the model was evaluated using a Confusion Matrix based on accuracy, precision, recall, and F1-score. The results indicate that the K-Nearest Neighbor (KNN) method is capable of classifying residents into low, medium, and high welfare categories with excellent performance. The best performance was achieved at K = 3, resulting in an accuracy of 98.04%, precision of 97.44%, recall of 99.12%, and an F1-score of 98.22%, while increasing the K value to 5 and 7 led to a decrease in classification accuracy due to class imbalance in the dataset. Based on these findings, the developed system can assist the Natam Village government in classifying residents' welfare levels more quickly, objectively, and accurately, thereby helping the decision-making process and improving the determination of eligible beneficiaries for social assistance programs..*

Keywords: *K-Nearest Neighbor (KNN), Data Mining, Classification, Residents' Welfare Level, Euclidean Distance, Confusion Matrix*

Abstrak. Penentuan tingkat kesejahteraan penduduk merupakan salah satu aspek penting dalam mendukung pengambilan keputusan pemerintah desa, khususnya dalam penyaluran bantuan sosial agar tepat sasaran. Namun, proses klasifikasi tingkat kesejahteraan penduduk di Desa Natam, Kecamatan Badar, Kabupaten Aceh Tenggara masih dilakukan secara manual sehingga membutuhkan waktu yang relatif lama, berpotensi menimbulkan subjektivitas, serta belum memanfaatkan data sosial ekonomi penduduk secara optimal. Penelitian ini bertujuan menerapkan metode K-Nearest Neighbor (KNN) untuk mengklasifikasikan tingkat kesejahteraan penduduk berdasarkan enam variabel sosial ekonomi, yaitu penghasilan, jumlah tanggungan, tingkat pendidikan, jenis pekerjaan, status tempat tinggal, dan kepemilikan aset. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen serta pengembangan sistem menggunakan model Waterfall. Data yang digunakan berjumlah 300 data kepala keluarga, yang dibagi menjadi 249 data training (83%) dan 51 data testing (17%). Sebelum proses klasifikasi dilakukan, data melalui tahapan preprocessing yang meliputi transformasi data kategorik ke bentuk numerik dan normalisasi menggunakan metode Min-Max Scaling. Proses klasifikasi dilakukan menggunakan Euclidean Distance dengan pengujian nilai K = 3, K = 5, dan K = 7, kemudian dievaluasi menggunakan Confusion Matrix berdasarkan nilai accuracy, precision, recall, dan F1-Score. Hasil penelitian menunjukkan bahwa metode K-Nearest Neighbor (KNN) mampu mengklasifikasikan tingkat kesejahteraan penduduk ke dalam kategori kesejahteraan rendah, kesejahteraan sedang, dan kesejahteraan tinggi dengan tingkat kinerja yang sangat baik. Nilai K = 3 menghasilkan performa terbaik dengan accuracy sebesar 98,04%, precision sebesar 97,44%, recall sebesar 99,12%, dan F1-Score sebesar 98,22%,

Naskah Masuk:11 September 2026; Revisi: 12 September 2026; Diterima: 13 September 2026; Terbit: 22 September 2026

sedangkan peningkatan nilai K menjadi 5 dan 7 menghasilkan penurunan tingkat akurasi akibat pengaruh ketidakseimbangan jumlah data antar kelas. Berdasarkan hasil tersebut, sistem yang dikembangkan mampu membantu pemerintah Desa Natam dalam mengklasifikasikan tingkat kesejahteraan penduduk secara lebih cepat, objektif, dan akurat sehingga dapat membantu proses pengambilan keputusan serta penentuan sasaran bantuan sosial yang lebih tepat..

Kata kunci: K-Nearest Neighbor (KNN), Data Mining, Klasifikasi, Tingkat Kesejahteraan Penduduk, Euclidean Distance, Confusion Matrix.

1. LATAR BELAKANG

Kesejahteraan masyarakat merupakan komponen penting dalam pola pembangunan ekonomi, dan keberhasilan pembangunan ekonomi diukur dengan meningkatnya tingkat kesejahteraan masyarakat. Kesejahteraan masyarakat merupakan salah satu indikator penting dalam menilai kualitas hidup masyarakat (Nathasya, 2024). Indikator kesejahteraan masyarakat merupakan alat yang memberikan arah atau data untuk mengukur keberhasilan pembangunan ekonomi. (Sastrawan et al., 2024). Tingkat kesejahteraan dapat dilihat dari berbagai aspek seperti penghasilan, pendidikan, pekerjaan, kondisi tempat tinggal, serta kepemilikan asset. Penghasilan dan kepemilikan asset merupakan faktor internal yang mempengaruhi tingkat kesejahteraan. Keluarga yang memiliki aset lebih sejahtera dibandingkan keluarga yang tidak memiliki aset. Sebab itu, keluarga yang memiliki aset yang lebih banyak, cenderung lebih sejahtera jika dibandingkan dengan keluarga yang memiliki aset terbatas (Lukluli, 2022). Oleh karena itu, pemerintah secara rutin melaksanakan program bantuan sosial untuk meningkatkan kesejahteraan masyarakat, khususnya bagi masyarakat yang kurang mampu.

Namun, dalam praktiknya masih terdapat permasalahan dalam proses penentuan tingkat kesejahteraan penduduk, khususnya di tingkat desa. Berdasarkan hasil observasi di Desa Natam Kecamatan Badar, proses pengelompokan tingkat kesejahteraan penduduk masih dilakukan secara manual oleh perangkat desa. Proses ini menimbulkan beberapa permasalahan, antara lain yaitu proses pengolahan data membutuhkan waktu yang lama karena dilakukan secara manual, penentuan kategori kesejahteraan masih bersifat subjektif sehingga berpotensi tidak konsisten, data penduduk yang tersedia belum dimanfaatkan secara optimal untuk menghasilkan informasi klasifikasi. Faktor-faktor seperti kurangnya data yang akurat, ketidakmerataan distribusi, dan kurangnya mekanisme evaluasi yang efektif seringkali menjadi penyebab utama ketidaktepatan sasaran ini. Sebagai akibatnya, sebagian besar bantuan sosial tidak mencapai individu atau keluarga yang paling membutuhkan, sementara sebagian lainnya mungkin menerima bantuan meskipun tidak sepenuhnya memenuhi kriteria yang ditetapkan (Purbaningrum & Adinugraha, 2023). Permasalahan tersebut menyebabkan hasil klasifikasi kesejahteraan menjadi kurang akurat dan tidak efisien. Padahal, dengan jumlah data penduduk yang terus meningkat, dibutuhkan suatu sistem yang mampu mengolah data secara cepat, objektif, dan sistematis. Oleh karena itu, permasalahan utama dalam penelitian ini adalah belum adanya sistem yang mampu mengklasifikasikan tingkat kesejahteraan penduduk secara cepat, objektif, dan berbasis data di Desa Natam.

Seiring perkembangan teknologi, pengolahan data dapat dilakukan menggunakan teknik data mining. Salah satu metode yang dapat digunakan adalah metode K-Nearest Neighbor (KNN). Algoritma ini merupakan salah satu teknik yang dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing(Argina, 2020). Algoritma ini diperlukan suatu sistem klasifikasi untuk mejadikan sebuah sistem yang mampu mencari informasi. Metode K-

Nearest Neighbor (KNN) dipilih dalam penelitian ini karena merupakan salah satu metode klasifikasi yang mampu mengelompokkan data berdasarkan tingkat kemiripan antar data menggunakan perhitungan jarak. Metode ini memiliki proses yang sederhana, mudah diimplementasikan, serta mampu menghasilkan klasifikasi yang baik pada data yang telah memiliki kelas sebelumnya. Dalam penelitian ini, data yang digunakan berupa data sosial ekonomi penduduk yang terdiri atas beberapa atribut, seperti tingkat pendapatan, jumlah tanggungan keluarga, tingkat pendidikan, jenis pekerjaan, kondisi tempat tinggal, dan kepemilikan aset. Karakteristik data tersebut sesuai dengan prinsip kerja KNN yang mengklasifikasikan suatu data berdasarkan kedekatannya dengan data lain yang memiliki karakteristik serupa (Raden Ayu Salma, 2025), sehingga metode ini dinilai tepat untuk digunakan dalam mengklasifikasikan tingkat kesejahteraan penduduk di Desa Natam.

Dengan menerapkan metode KNN, proses klasifikasi tingkat kesejahteraan penduduk dapat dilakukan secara otomatis berdasarkan data yang tersedia, sehingga diharapkan dapat meningkatkan akurasi, mengurangi subjektivitas, dan mempercepat proses pengambilan keputusan.

Berdasarkan permasalahan tersebut, maka penelitian ini dilakukan dengan judul **"Penerapan Metode K-Nearest Neighbor (KNN) untuk Klasifikasi Tingkat Kesejahteraan Penduduk di Desa Natam Kecamatan Badar Kutacane Aceh Tenggara."**

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dalam bidang data mining. Pendekatan kuantitatif dipilih karena penelitian ini memanfaatkan data numerik yang berasal dari indikator sosial ekonomi penduduk untuk menghasilkan klasifikasi tingkat kesejahteraan masyarakat. Metode eksperimen diterapkan untuk menguji penerapan algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan tingkat kesejahteraan penduduk. Selain itu, penelitian ini juga menggunakan pendekatan studi kasus, dengan fokus pada Desa Natam, Kecamatan Badar, Kabupaten Aceh Tenggara. Pemilihan lokasi ini didasarkan pada permasalahan dalam penentuan tingkat kesejahteraan penduduk yang masih dilakukan secara manual, yang berpotensi menyebabkan kesalahan dalam pengelompokan data masyarakat.

Penelitian dilaksanakan selama beberapa bulan pada tahun 2026, dimulai dari pengumpulan data hingga penyusunan laporan penelitian. Proses penelitian dilakukan secara bertahap untuk memastikan setiap tahap menghasilkan data yang akurat. Variabel yang digunakan dalam penelitian ini mencakup penghasilan, jumlah tanggungan, tingkat pendidikan, jenis pekerjaan, kondisi tempat tinggal, dan kepemilikan aset. Variabel-variabel ini berfungsi sebagai dasar dalam proses analisis untuk menentukan tingkat kesejahteraan penduduk. Setiap variabel direpresentasikan dalam bentuk atribut pada dataset yang akan digunakan dalam proses klasifikasi menggunakan metode KNN.

Teknik pengumpulan data yang digunakan adalah dokumentasi, yang meliputi pengumpulan data dari dokumen administrasi penduduk di Kantor Desa Natam. Data tersebut disusun dalam bentuk dataset yang terstruktur untuk memudahkan analisis. Metode evaluasi sistem menggunakan Confusion Matrix untuk mengukur kinerja model klasifikasi yang dibangun dalam penelitian ini. Evaluasi dilakukan untuk mengetahui tingkat akurasi, precision, dan recall dari sistem klasifikasi, sehingga dapat memberikan informasi yang akurat tentang kondisi sosial ekonomi masyarakat di Desa Natam. Hasil

evaluasi ini diharapkan dapat membantu pemerintah desa dalam pengambilan keputusan terkait program peningkatan kesejahteraan masyarakat.

3. HASIL DAN PEMBAHASAN

Pembagian Data Training dan Testing

Data yang digunakan dalam penelitian ini berjumlah 300 data penduduk Desa Natam yang telah melalui proses pengumpulan data sesuai dengan variabel penelitian yang telah ditentukan, yaitu penghasilan kepala keluarga, jumlah tanggungan, tingkat pendidikan, jenis pekerjaan, status tempat tinggal, dan kepemilikan aset. Data tersebut kemudian dibagi menjadi data training dan data testing menggunakan metode holdout dengan perbandingan 83% data training dan 17% data testing, sebagaimana ditunjukkan pada tabel berikut.

Tabel Pembagian Data Training dan Testing

Keterangan	Jumlah Data	Persentase
Total Data Penduduk	**300**	**100%**
Data Training (Data Latih)	249	83%
Data Testing (Data Uji)	51	17%

Sebanyak 249 data digunakan sebagai data training yang berfungsi sebagai acuan atau referensi dalam proses perhitungan jarak Euclidean pada algoritma KNN, sedangkan 51 data digunakan sebagai data testing untuk menguji kemampuan sistem dalam mengklasifikasikan tingkat kesejahteraan penduduk secara objektif. Pengujian dilakukan dengan tiga variasi nilai K, yaitu K = 3, K = 5, dan K = 7, untuk mengetahui pengaruh nilai K terhadap tingkat akurasi klasifikasi.

Hasil Pengujian dengan Nilai K = 3

Pengujian klasifikasi pada nilai K = 3 dilakukan terhadap seluruh 51 data testing menggunakan 249 data training sebagai acuan. Hasil evaluasi sistem ditampilkan pada Tabel berikut.

Tabel Hasil Evaluasi Klasifikasi K = 3

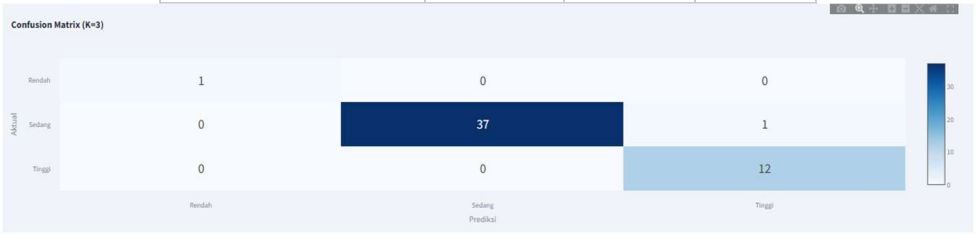
Kelas	TP	FP	FN	Precision	Recall	F1-Score
Rendah	1	0	0	100.00%	100.00%	100.00%
Sedang	37	0	1	100.00%	97.37%	98.67%
Tinggi	12	1	0	92.31%	100.00%	96.00%
Rata-rata (Macro Avg)	—	—	—	**97.44%**	**99.12%**	**98.22%**

Nilai akurasi keseluruhan yang diperoleh pada pengujian dengan K = 3 adalah sebesar 98.04%, dengan precision rata-rata 97.44%, recall rata-rata 99.12%, dan F1-Score rata-rata 98.22%. Sebanyak 50 dari 51 data uji berhasil diklasifikasikan dengan benar oleh sistem.

Tabel Confusion Matrix K = 3

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah	**1**	0	0

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Sedang	0	**37**	1
Tinggi	0	0	**12**



Gambar Visualisasi Confusion Matrix Heatmap K = 3

Berdasarkan Tabel di atas, model KNN dengan nilai K = 3 menghasilkan tingkat akurasi sebesar 98.04%, yang merupakan hasil terbaik di antara seluruh nilai K yang diujikan. Dari 51 data uji, sebanyak 50 data berhasil diklasifikasikan dengan benar dan hanya 1 data yang mengalami kesalahan klasifikasi, yaitu 1 data dengan label aktual "Sedang" yang diprediksi sebagai "Tinggi". Nilai precision rata-rata sebesar 97.44%, recall sebesar 99.12%, dan F1-Score sebesar 98.22% menunjukkan bahwa model mampu mengenali pola kesejahteraan penduduk dengan sangat baik pada nilai K yang kecil, karena tetangga terdekat yang digunakan lebih sedikit sehingga lebih sensitif terhadap kemiripan data secara lokal.

Hasil Pengujian dengan Nilai K = 5

Pengujian klasifikasi pada nilai K = 5 dilakukan terhadap seluruh 51 data testing menggunakan 249 data training sebagai acuan. Hasil evaluasi sistem ditampilkan pada Tabel berikut.

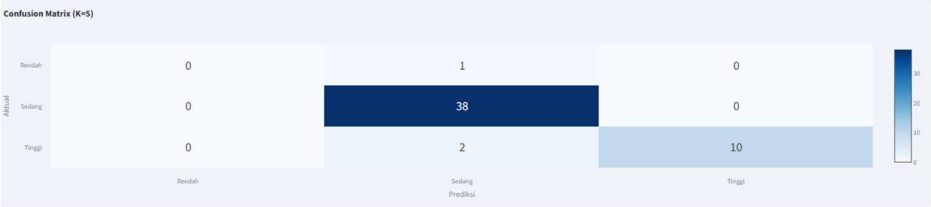
Tabel Hasil Evaluasi Klasifikasi K = 5

Kelas	TP	FP	FN	Precision	Recall	F1-Score
Rendah	0	0	1	0.00%	0.00%	0.00%
Sedang	38	3	0	92.68%	100.00%	96.20%
Tinggi	10	0	2	100.00%	83.33%	90.91%
Rata-rata (Macro Avg)	—	—	—	**64.23%**	**61.11%**	**62.37%**

Nilai akurasi keseluruhan yang diperoleh pada pengujian dengan K = 5 adalah sebesar 94.12%, dengan precision rata-rata 64.23%, recall rata-rata 61.11%, dan F1-Score rata-rata 62.37%. Sebanyak 48 dari 51 data uji berhasil diklasifikasikan dengan benar oleh sistem.

Tabel Confusion Matrix K = 5

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah	**0**	1	0
Sedang	0	**38**	0
Tinggi	0	2	**10**



Gambar Visualisasi Confusion Matrix Heatmap K = 5

Berdasarkan Tabel di atas, model KNN dengan nilai K = 5 menghasilkan tingkat akurasi sebesar 94.12%, mengalami penurunan dibandingkan dengan K = 3. Dari 51 data uji, sebanyak 48 data berhasil diklasifikasikan dengan benar dan 3 data mengalami kesalahan klasifikasi. Kesalahan paling signifikan terjadi pada kelas "Rendah", di mana satu-satunya data dengan label aktual "Rendah" salah diklasifikasikan sebagai "Sedang", sehingga nilai precision dan recall untuk kelas tersebut menjadi 0,00%. Hal ini menyebabkan nilai precision rata-rata turun signifikan menjadi 64.23% dan recall menjadi 61.11%. Penurunan performa ini disebabkan oleh jumlah data pada kelas "Rendah" yang sangat sedikit (data minoritas), sehingga ketika nilai K diperbesar menjadi 5, suara mayoritas tetangga lebih didominasi oleh kelas "Sedang" yang memiliki jumlah data jauh lebih banyak.

Hasil Pengujian dengan Nilai K = 7

Pengujian klasifikasi pada nilai K = 7 dilakukan terhadap seluruh 51 data testing menggunakan 249 data training sebagai acuan. Hasil evaluasi sistem ditampilkan pada Tabel berikut.

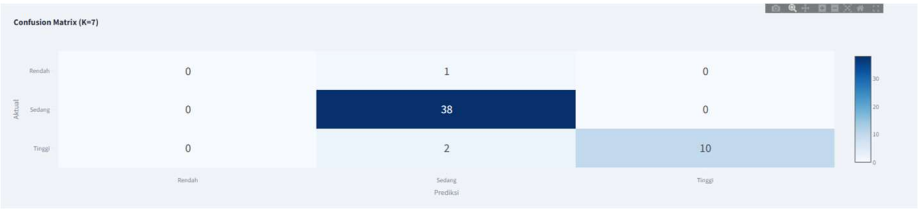
Tabel Hasil Evaluasi Klasifikasi K = 7

Kelas	TP	FP	FN	Precision	Recall	F1-Score
Rendah	0	0	1	0.00%	0.00%	0.00%
Sedang	38	3	0	92.68%	100.00%	96.20%
Tinggi	10	0	2	100.00%	83.33%	90.91%
Rata-rata (Macro Avg)	—	—	—	**64.23%**	**61.11%**	**62.37%**

Nilai akurasi keseluruhan yang diperoleh pada pengujian dengan K = 7 adalah sebesar 94.12%, dengan precision rata-rata 64.23%, recall rata-rata 61.11%, dan F1-Score rata-rata 62.37%. Sebanyak 48 dari 51 data uji berhasil diklasifikasikan dengan benar oleh sistem.

Tabel Confusion Matrix K = 7

Aktual \ Prediksi	Rendah	Sedang	Tinggi
Rendah	**0**	1	0
Sedang	0	**38**	0
Tinggi	0	2	**10**



Gambar Visualisasi Confusion Matrix Heatmap K = 7

Berdasarkan Tabel di atas, model KNN dengan nilai K = 7 menghasilkan hasil yang identik dengan K = 5, yaitu akurasi sebesar 94.12%, precision 64.23%, recall 61.11%, dan F1-Score 62.37%. Kesamaan hasil ini terjadi karena pola tetangga terdekat pada data uji yang sama tidak berubah secara signifikan ketika nilai K diperbesar dari 5 menjadi 7, mengingat keterbatasan jumlah data pada kelas "Rendah". Sama seperti pada K = 5, kesalahan klasifikasi tetap terjadi pada data kelas "Rendah" yang diprediksi sebagai "Sedang", serta 2 data kelas "Tinggi" yang diprediksi sebagai "Sedang".

Perbandingan Hasil Pengujian Nilai K

Untuk mengetahui nilai K yang menghasilkan performa klasifikasi terbaik, dilakukan perbandingan hasil evaluasi dari ketiga nilai K yang diujikan. Ringkasan perbandingan tersebut disajikan pada tabel berikut.

Tabel Perbandingan Hasil Pengujian Nilai K = 3, K = 5, dan K = 7

Nilai K	Akurasi	Precision	Recall	F1-Score	Data Benar
K = 3	**98.04%**	97.44%	99.12%	98.22%	50/51
K = 5	94.12%	64.23%	61.11%	62.37%	48/51
K = 7	94.12%	64.23%	61.11%	62.37%	48/51

Berdasarkan Tabel 4.8, dapat disimpulkan bahwa nilai K = 3 menghasilkan performa klasifikasi terbaik di antara ketiga nilai K yang diujikan, dengan tingkat akurasi sebesar 98,04%, precision 97,44%, recall 99,12%, dan F1-Score 98,22%. Sementara itu, nilai K = 5 dan K = 7 menghasilkan performa yang identik, yaitu akurasi sebesar 94,12% dengan precision dan recall yang jauh lebih rendah dibandingkan K = 3. Penurunan performa pada K = 5 dan K = 7 disebabkan oleh ketidakseimbangan jumlah data antar kelas (class imbalance), khususnya pada kelas "Rendah" yang hanya memiliki sangat sedikit data dibandingkan kelas "Sedang" dan "Tinggi". Semakin besar nilai K yang digunakan, semakin besar pula kemungkinan suara mayoritas tetangga didominasi oleh kelas dengan jumlah data lebih banyak, sehingga data dari kelas minoritas cenderung salah diklasifikasikan ke kelas mayoritas terdekat.

Dengan demikian, nilai K = 3 dipilih sebagai nilai K terbaik untuk digunakan dalam sistem klasifikasi tingkat kesejahteraan penduduk Desa Natam, karena mampu memberikan keseimbangan terbaik antara sensitivitas terhadap data minoritas dan akurasi klasifikasi secara keseluruhan.

Perhitungan Manual Klasifikasi Tingkat Kesejahteraan Menggunakan Metode KNN

Bagian ini menyajikan contoh perhitungan manual proses klasifikasi tingkat kesejahteraan penduduk menggunakan metode K-Nearest Neighbor (KNN) dengan nilai K = 3 dan perhitungan jarak menggunakan rumus Euclidean Distance. Contoh perhitungan dilakukan terhadap satu data penduduk dengan nama Saiful sebagai data uji, dengan mengacu pada tiga data training sebagai tetangga terdekat.

Langkah 1: Transformasi Data Kategorik ke Numerik (Encoding)

Sebelum proses klasifikasi dilakukan, seluruh data kategorik terlebih dahulu ditransformasi menjadi nilai numerik menggunakan tabel encoding yang telah ditetapkan. Proses ini diperlukan karena algoritma KNN bekerja berdasarkan perhitungan jarak matematis antar data. Tabel berikut menunjukkan proses encoding untuk data uji Saiful.

Tabel Encoding Data Uji Saiful ke Nilai Numerik

Variabel	**Nilai Kategori**	**Nilai Numerik**	**Keterangan**
Penghasilan	< Rp 1.000.000	**1**	Terkecil → skor terendah
Jumlah Tanggungan	3 – 5 Orang	**2**	Tingkat menengah
Pendidikan	SD	**2**	Jenjang ke-2
Pekerjaan	Buruh/Pegawai Swasta	**2**	Kelas ke-2
Tempat Tinggal	Kontrak	**2**	Status ke-2
Kepemilikan Aset	Tidak Memiliki Aset	**1**	Terkecil → skor terendah

Setelah proses encoding, data uji Saiful direpresentasikan dalam vektor fitur berikut: X_Saiful = [1, 2, 2, 2, 2, 1]. Demikian pula untuk data training, proses yang sama diterapkan sehingga diperoleh representasi numerik seperti yang ditampilkan pada Tabel berikut.

Tabel Representasi Numerik Data Training dan Data Uji

ID	**Data**	**X_1 Pgh**	**X_2 Tang**	**X_3 Pdd**	**X_4 Pkj**	**X_5 Tgl**	**X_6 Ast**	**Label**
T_1	Ahmad Fauzi (Training)	1	3	2	2	2	1	**Rendah**
T_2	Darmawi (Training)	2	2	3	3	4	3	**Sedang**
T_3	Hendri (Training)	3	1	6	4	4	4	**Tinggi**
D	**Saiful (Data Uji)**	1	2	2	2	2	1	**?**

Langkah 2: Normalisasi Data dengan Metode Min-Max Scaling

Setelah encoding, dilakukan normalisasi data menggunakan metode Min-Max Scaling agar setiap atribut memiliki skala nilai yang seragam dalam rentang [0, 1]. Normalisasi diperlukan agar atribut dengan rentang nilai besar tidak mendominasi perhitungan jarak dibandingkan atribut dengan rentang nilai kecil. Rumus Min-Max Scaling yang digunakan adalah sebagai berikut:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Keterangan: x adalah nilai asli atribut, $X_{\min}$ adalah nilai minimum atribut pada seluruh data training, $X_{\max}$ adalah nilai maksimum atribut pada seluruh data training, dan x' adalah nilai hasil normalisasi.

Berikut adalah nilai $X_{\min}$ dan $X_{\max}$ untuk setiap atribut yang diperoleh dari data training:

$X_{\min}$= [1, 1, 2, 2, 2, 1] (Penghasilan, Tanggungan, Pendidikan, Pekerjaan, Tempat Tinggal, Aset)

$X_{\max}$ = [3, 3, 6, 4, 4, 4]

Contoh perhitungan normalisasi untuk atribut Penghasilan data uji Saiful (x = 1):

x' = (1 - 1) / (3 - 1) = 0 / 2 = 0,0000

Proses yang sama diterapkan pada seluruh atribut untuk semua data. Hasil normalisasi lengkap ditampilkan pada tabel berikut.

Tabel Hasil Normalisasi Data Training dan Data Uji

ID	Data	X_1	X_2	X_3	X_4	X_5	X_6	Label
T_1	Ahmad Fauzi	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	**Rendah**
T_2	Darmawi	0.5000	0.5000	0.2500	0.5000	1.0000	0.6667	**Sedang**
T_3	Hendri	1.0000	0.0000	1.0000	1.0000	1.0000	1.0000	**Tinggi**
D	**Saiful (Uji)**	0.0000	0.5000	0.0000	0.0000	0.0000	0.0000	**?**

Langkah 3: Perhitungan Jarak Euclidean Distance

Setelah seluruh data dinormalisasi, langkah selanjutnya adalah menghitung jarak Euclidean antara data uji Saiful dengan setiap data training. Rumus Euclidean Distance yang digunakan adalah sebagai berikut:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$

Keterangan: d(x, y) adalah jarak Euclidean antara data uji x dan data training y, x_i adalah nilai atribut ke-i pada data uji setelah normalisasi, dan y_i adalah nilai atribut ke-i pada data training setelah normalisasi.

Perhitungan jarak data uji Saiful terhadap T_1 (Ahmad Fauzi, Label: Rendah):

$d(D, T_1) = \sqrt{[(0.0000-0.0000)^2 + (0.5000-1.0000)^2 + (0.0000-0.0000)^2 + (0.0000-0.0000)^2 + (0.0000-0.0000)^2 + (0.0000-0.0000)^2]}$

$d(D, T_1) = \sqrt{[0.000000 + 0.250000 + 0.000000 + 0.000000 + 0.000000 + 0.000000]}$

$d(D, T_1) = \sqrt{0.250000} = 0.500000$

Perhitungan jarak data uji Saiful terhadap T_2 (Darmawi, Label: Sedang):

$d(D, T_2) = \sqrt{[(0.0000-0.5000)^2 + (0.5000-0.5000)^2 + (0.0000-0.2500)^2 + (0.0000- 0.5000)^2 +(0.0000-1.0000)^2 + (0.0000-0.6667)^2]}$

$d(D, T_2) = \sqrt{[0.250000 + 0.000000 + 0.062500 + 0.250000 + 1.000000 + 0.444444]}$

$d(D, T_2) = \sqrt{2.006944} = 1.416667$

Perhitungan jarak data uji Saiful terhadap T_3 (Hendri, Label: Tinggi):

$d(D, T_3) = \sqrt{[(0.0000-1.0000)^2 + (0.5000-0.0000)^2 + (0.0000-1.0000)^2 + (0.0000-1.0000)^2 + (0.0000-1.0000)^2 + (0.0000-1.0000)^2]}$

$d(D, T_3) = \sqrt{[1.000000 + 0.250000 + 1.000000 + 1.000000 + 1.000000 + 1.000000]}$

$d(D, T_3) = \sqrt{5.250000} = 2.291288$

Langkah 4: Menentukan K Tetangga Terdekat (K = 3)

Setelah seluruh jarak dihitung, data training diurutkan berdasarkan nilai jarak dari yang terkecil (terdekat) hingga terbesar (terjauh). Dengan nilai K = 3, diambil 3 data training dengan jarak terkecil sebagai tetangga terdekat dari data uji Saiful.

Tabel Hasil Pengurutan Jarak dan Pemilihan K Tetangga Terdekat

Rank	Tetangga	Jarak Euclidean (d)	Label
1	Ahmad Fauzi (T_1)	0.500000	**Rendah**
2	Darmawi (T_2)	1.416667	**Sedang**
3	Hendri (T_3)	2.291288	**Tinggi**

Langkah 5: Majority Voting dan Penentuan Hasil Klasifikasi

Setelah K = 3 tetangga terdekat diperoleh, dilakukan proses voting mayoritas untuk menentukan kategori kesejahteraan data uji Saiful. Setiap tetangga memberikan satu suara sesuai label kelasnya. Kelas yang mendapat suara terbanyak ditetapkan sebagai hasil prediksi klasifikasi.

Pada dataset real dengan 249 data training dan nilai K = 3, tiga tetangga terdekat dari data uji Saiful menghasilkan komposisi voting sebagai berikut:

Tabel Hasil Majority Voting untuk Data Uji Saiful

Kelas	Jumlah Suara	Persentase	Keterangan
Rendah	**2**	67%	Mayoritas → Hasil Prediksi
Sedang	**1**	33%	Suara minoritas
Tinggi	**0**	0%	Tidak terpilih

Berdasarkan Tabel 4.13, kelas 'Rendah' memperoleh suara terbanyak (2 dari 3 suara), sehingga sistem mengklasifikasikan data uji Saiful ke dalam kategori Kesejahteraan Rendah. Hasil prediksi ini sesuai dengan label aktual data Saiful, sehingga data ini dinyatakan sebagai prediksi yang Benar (True Positive).

Hasil Prediksi Saiful = RENDAH (Sesuai label aktual)

Proses yang sama diterapkan secara berulang oleh sistem terhadap seluruh 51 data testing dengan mengacu pada 249 data training. Seluruh hasil perhitungan jarak, pemilihan tetangga terdekat, dan voting mayoritas dilakukan secara otomatis oleh sistem menggunakan algoritma KNN yang diimplementasikan dalam modul utils/knn_engine.py, sehingga menghasilkan nilai akurasi sebesar 98,04% pada pengujian dengan nilai K = 3.

PEMBAHASAN

Hasil pengujian menunjukkan bahwa metode K-Nearest Neighbor (KNN) dengan perhitungan jarak Euclidean Distance dapat diterapkan secara efektif untuk mengklasifikasikan tingkat kesejahteraan penduduk Desa Natam berdasarkan variabel sosial ekonomi yang telah ditentukan, yaitu penghasilan, jumlah tanggungan, tingkat pendidikan, jenis pekerjaan, status tempat tinggal, dan kepemilikan aset. Tingkat akurasi tertinggi sebesar 98,04% diperoleh pada nilai K = 3, yang menunjukkan bahwa algoritma

KNN bekerja optimal ketika jumlah tetangga terdekat yang dipertimbangkan relatif kecil pada kasus dataset penelitian ini.

Analisis terhadap confusion matrix pada ketiga nilai K menunjukan bahwa kesalahan klasifikasi secara konsisten terjadi pada batas antara kelas "Sedang" dan "Tinggi", serta antara kelas "Rendah" dan "Sedang". Hal ini wajar terjadi mengingat batas antar kategori kesejahteraan bersifat kontinu dan tidak selalu memiliki pemisahan yang tegas pada kondisi sosial ekonomi riil di lapangan. Selain itu, jumlah data pada kelas "Rendah" yang sangat sedikit dibandingkan kelas "Sedang" dan "Tinggi" turut memengaruhi kemampuan model dalam mengenali pola data minoritas tersebut, khususnya pada nilai K yang lebih besar.

Secara keseluruhan, hasil penelitian ini menjawab rumusan masalah yang telah ditetapkan, yaitu metode KNN dapat diterapkan dengan baik dalam mengklasifikasikan tingkat kesejahteraan penduduk Desa Natam, dengan tingkat akurasi terbaik dicapai pada nilai K = 3 sebesar 98,04%. Hasil klasifikasi berdasarkan variabel sosial ekonomi yang digunakan juga menunjukkan bahwa sistem mampu mengelompokkan penduduk ke dalam tiga kategori kesejahteraan, yaitu rendah, sedang, dan tinggi, secara akurat dan dapat dipertanggungjawabkan secara metodologis.

4. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian dan pengujian sistem yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Metode K-Nearest Neighbor (KNN) berhasil diterapkan dalam mengklasifikasikan tingkat kesejahteraan penduduk di Desa Natam, Kecamatan Badar, Kutacane, Aceh Tenggara, berdasarkan enam variabel sosial ekonomi, yaitu penghasilan, jumlah tanggungan, tingkat pendidikan, jenis pekerjaan, status tempat tinggal, dan kepemilikan aset, dengan menggunakan perhitungan jarak Euclidean Distance dan proses voting mayoritas untuk menentukan kategori kesejahteraan.
2. Tingkat akurasi metode KNN dalam mengklasifikasikan kesejahteraan penduduk bervariasi tergantung pada nilai K yang digunakan. Nilai K = 3 menghasilkan akurasi terbaik sebesar 98,04% dengan precision 97,44%, recall 99,12%, dan F1-Score 98,22%, sedangkan nilai K = 5 dan K = 7 masing-masing menghasilkan akurasi sebesar 94,12% dengan precision dan recall yang lebih rendah. Dengan demikian, nilai K = 3 merupakan nilai K optimal untuk diterapkan pada kasus klasifikasi kesejahteraan penduduk Desa Natam.
3. Hasil klasifikasi berdasarkan variabel sosial ekonomi yang digunakan menunjukkan bahwa dari 51 data uji, sistem berhasil mengklasifikasikan penduduk ke dalam tiga kategori kesejahteraan, yaitu rendah, sedang, dan tinggi, dengan tingkat ketepatan yang tinggi pada nilai K = 3, di mana hanya 1 dari 51 data yang mengalami kesalahan klasifikasi.
4. Kesalahan klasifikasi yang terjadi pada nilai K = 5 dan K = 7 sebagian besar disebabkan oleh ketidakseimbangan jumlah data antar kelas (class imbalance), khususnya pada kelas "Rendah" yang memiliki jumlah data yang sangat terbatas dibandingkan kelas "Sedang" dan "Tinggi", sehingga ketika nilai K diperbesar, data minoritas cenderung salah diklasifikasikan ke kelas mayoritas terdekat.

5. Sistem klasifikasi kesejahteraan penduduk berbasis web yang dibangun menggunakan Streamlit dan MySQL telah berhasil mengimplementasikan seluruh alur proses penelitian, mulai dari pengelolaan data penduduk, proses klasifikasi dengan algoritma KNN, evaluasi menggunakan Confusion Matrix, hingga penyajian hasil dalam bentuk laporan yang dapat dicetak, sehingga dapat digunakan sebagai alat bantu pengambilan keputusan oleh pihak pemerintah Desa Natam.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan penelitian maupun sistem selanjutnya, antara lain:

1. Penelitian selanjutnya disarankan untuk menambah jumlah data pada kelas minoritas, khususnya kelas "Rendah", agar distribusi data antar kelas menjadi lebih seimbang sehingga dapat meningkatkan akurasi klasifikasi pada nilai K yang lebih besar.
2. Penelitian selanjutnya dapat menerapkan teknik penanganan data tidak seimbang (imbalanced data), seperti metode oversampling (contohnya SMOTE) atau pembobotan kelas, untuk mengatasi permasalahan class imbalance yang ditemukan pada penelitian ini.
3. Sistem yang dikembangkan dapat ditingkatkan dengan menambahkan fitur prediksi otomatis untuk data penduduk baru tanpa perlu menjalankan ulang seluruh proses klasifikasi secara manual.
4. Penelitian selanjutnya dapat membandingkan metode KNN dengan metode klasifikasi lain, seperti Naive Bayes, Decision Tree, atau Support Vector Machine (SVM), untuk mengetahui metode yang paling optimal dalam mengklasifikasikan tingkat kesejahteraan penduduk.
5. Pihak pemerintah Desa Natam disarankan untuk memanfaatkan sistem ini secara berkala dalam melakukan pemutakhiran data kesejahteraan penduduk, sehingga data yang dihasilkan dapat digunakan sebagai dasar pengambilan kebijakan yang lebih tepat sasaran.
6. Pengembangan sistem selanjutnya dapat mempertimbangkan migrasi antarmuka pengguna ke platform yang lebih scalable serta penambahan fitur keamanan data, mengingat data yang dikelola bersifat sensitif dan berkaitan langsung dengan kondisi sosial ekonomi masyarakat.

DAFTAR REFERENSI

Alfiana Erlangga, D., Kurnia Azzahra, D., & Adi Nugraha, D. (2025). Implemntasi Algoexplorer Aplikasi Data Mining Berbasis Web. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(5), 8073–8080. https://doi.org/10.36040/jati.v9i5.14975

Alfiatul Karima, R., & Fatah, Z. (2024). Jurnal Ilmiah Multidisiplin Ilmu Implementasi Metode K-Nearest Neighbor Untuk Klasifikasi Penyakit Paru-Paru Pada Anak. *Jurnal Ilmiah Multidisiplin Ilmu*, *1*(6), 13–14.

Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neigbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, *1*(2), 29–33. https://doi.org/10.33096/ijodas.v1i2.11

Arif, A. F. N., & Nurwati, N. (2022). Pengaruh Konsentrasi Penduduk Indonesia Di Pulau Jawa Terhadap Kesejahteraan Masyarakat. *Jurnal Ilmu Kesejahteraan Sosial HUMANITAS*, *4*(I), 54–70. https://doi.org/10.23969/humanitas.v4ii.3920

Bustamam, N., Yulyanti, S., & Dewi, K. S. (2021). Analisis Faktor-Faktor…{Bustamam, dkk }|. *Jurnal Ekonomi KIAT*, *32*(1), 85. https://journal.uir.ac.id/index.php/kiat

Enyakit, K. U. K. L. P., Elitus, D. I. M., Asus, S. T. K., & Esa, W. A. D. (2022). *P a k-n n (k-nn) k p d m s k : w d j*. *16*(2), 133–142.

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Lubis, J. B., Situmorang, A., Larosa, F. G. N., Rumapea, H., Sinaga, I., Komputer, F. I., & Artikel, H. (2023). Penerapan Metode K-Nearest Neighbor Dalam Mengklasifikasi Status Ekonomi Suatu Penduduk. *Jurnal Ilmiah Sistem Informasi*, *3*(1), 59–64. http://ojs.fikom-methodist.net/index.php/methosisfo

Lukluli, Q. S. A. (2022). *Independent : Journal Of Economics E-ISSN : 2798-5008*. *2*, 169–177.

Mulia, R. A., & Putri, R. P. (2022). The Influence of Poverty Levels and Gross Regional Domestic Product on Community Welfare. (in Indonesian). *Jiee: Jurnal Ilmiah Ekotrans & Erudisi*, *2*(1), 22–33.

Nathasya, H. (2024). No TitleΕΛΕΝΗ. *Edu Research Indonesian Institute For Corporate Learning And Studies (IICLS)*, *5*(1), 70–80.

Nurdin, N. (2024). Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor. *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. *Building of Informatics, Technology and Science (BITS)*, *4*(3), 1241–1245. https://doi.org/10.47065/bits.v4i3.2276

Purbaningrum, D., & Adinugraha, H. H. (2023). *KETIDAKTEPATAN SASARAN BANTUAN SOSIAL DI DESA*. (September 2022).

Putri, & Nur. (2023). Penggunaan Bahasa Python Untuk Analisis Dan Visualisasi Data Penduduk Di Desa Sumberjo, Nganjuk. *Jurnal Pengabdian Kepada Masyarakat*, *3*(3), 206–217. https://jurnalfkip.samawa-university.ac.id/karya_jpm/index

Raden Ayu Salma, A. W. (2025). *9995-Article Text-29377-1-10-20250310*. *12*(1), 294–305.

Rahmadini, R., Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, & Tuti Meutia. (2023). Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Mahasiswa Akuntansi Samudra*, *4*(4), 223–235. https://doi.org/10.33059/jmas.v4i4.7074

Ramdani, A., Sofyan, C. D., Ramdani, F., Tama, M. F. A., & Rachmatsyah, M. A. (2022). Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat Dalam. *Jurnal Ilmiah Sistem Informasi (JUSI)*, *1*(2), 39–47.

Reynaldi Valerian, F., Syarief, M., & Abdul Fatah, D. (2025). Klasifikasi Tingkat Obesitas Menggunakan Metode Gbm Dan Confusion Matrix. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(2), 2242–2249. https://doi.org/10.36040/jati.v9i2.13062

Rohman, C. S., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *6*(2), 118–127. http://ejournal.bsi.ac.id/ejurnal/index.php/ijcit

Saputri, M. K., & Rahmawati, F. (2022). *TINGKAT KESEJAHTERAAN MASYARAKAT MASA PANDEMI COVID- Banyak penelitian terdahulu yang menjelaskan tentang Indeks Pembangunan manusia dalam hal ini Badan Pusat Statistik Provinsi Jawa Timur selalu aktif dalam merilis laporan tentang perkembangan Indeks Pem*. *5*, 55–73.

Sastrawan, B., Samsi, A., & Seran, G. G. (2024). Pelayanan Pemerintah Bidang Kesejahteraan Masyarakat. *Karimah Tauhid*, *3*(1), 473–479. https://doi.org/10.30997/karimahtauhid.v3i1.11642

Siregar, Y. S., Oktaviana Sembiring, B., Rahayu, E., Hasdiana, H., & Franchitika, R. (2024). Pemanfaatan Aplikasi MySQL untuk Membantu Siswa SMK Swasta Nur Azizi dalam Pengolahan Data. *Jurnal Pengabdian Masyarakat*, *3*(2), 229–240. https://doi.org/10.70340/japamas.v3i2.185

Srirahayu, A., & Pribadie, L. S. (2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, *14*(1). https://doi.org/10.36982/jiig.v14i1.2981

Syahril, M., Erwansyah, K., & Yetri, M. (2020). J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan Data Mining Untuk Menentukan Pola. ■, *118*(1), 118–136.

Taqwanur, & Mega Bilqis Suryawantiningtyas. (2022). G-Tech : Jurnal Teknologi Terapan. *G-Tech : Jurnal Teknologi Terapan*, *6*(2), 295–305.

Ujianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, *4*(1), 33–43. https://doi.org/10.24246/itexplore.v4i1.2025.pp33-43

Yuni Utami, D., Nurlelah, E., & Nur Hasan, F. (2021). Comparison of Neural Network Algorithms, Naive Bayes and Logistic Regression To Find The Highest Accuracy In Diabetes. *JITE (Journal of Informatics and Telecommunication Engineering)* , *5*(1), 53–64. http://ojs.uma.ac.id/index.php/jite