INSECT: Jurnal Teknik Informatika
INFORMATICS AND SECURITY e-ISSN. 2614-431X

Klasifikasi Kelayakan Penerima Bantuan Sosial dengan Metode *K-Nearest Neighbors*

Nyimas Siti Rosyada*[1], Herri Setiawan[2], Muhammad Haviz Irvani[3]
[1,2,3] Program Studi Teknik Informatika, Universitas Indo Global Mandiri
E-mail: [1]2021110100@students.uigm.ac.id, [2]herri@uigm.ac.id, [3]*m.haviz@uigm.ac.id

Abstrak

Penyaluran bantuan sosial di Indonesia masih menghadapi berbagai kendala, seperti ketidakakuratan data penerima, birokrasi yang rumit, dan kurangnya transparansi, sehingga bantuan sering kali tidak tepat sasaran. Penelitian ini bertujuan mengoptimalkan penerapan metode K-Nearest Neighbors (K-NN) dalam klasifikasi kelayakan penerima bantuan sosial melalui pengujian beberapa rasio pembagian data latih dan uji serta variasi parameter k. Tujuan utama penelitian adalah menghasilkan model klasifikasi yang akurat dan andal sebagai rujukan dalam kebijakan penyaluran bantuan sosial di Kelurahan Kuto Batu, Palembang. Data yang digunakan meliputi atribut sosial ekonomi warga dan diolah melalui tahapan pembersihan, encoding, serta penanganan nilai hilang sebelum diterapkan ke algoritma K-NN. Pengujian dilakukan pada empat skenario pembagian data, yaitu 80/20, 70/30, 60/40, dan 50/50, untuk menentukan komposisi yang paling optimal. Hasil evaluasi menunjukkan akurasi model sebesar 97,44% pada rasio 80/20, 98,30% pada 70/30, 97,10% pada 60/40, dan 98,00% pada 50/50. Skenario pembagian 70/30 memberikan hasil terbaik dengan akurasi 98,30%, presisi 100%, recall 98%, dan F1-score 98,98%.Rasio 70/30 dipilih sebagai konfigurasi optimal karena memberikan keseimbangan antara jumlah data pelatihan yang memadai untuk membentuk pola dan data pengujian yang cukup untuk mengukur kemampuan generalisasi model. Hasil ini membuktikan bahwa metode K-NN efektif dalam membedakan penerima bantuan yang layak dan tidak layak secara objektif, serta berpotensi menjadi dasar pengembangan sistem pendukung keputusan untuk penyaluran bantuan sosial yang lebih transparan dan tepat sasaran.

Kata kunci: *K-Nearest Neighbors, Klasifikasi, Bantuan Sosial, Confusion Matrix, Akurasi.*

1. PENDAHULUAN

Pesatnya kemajuan di bidang teknologi informasi dan kecerdasan buatan (AI) pada era digital telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk cara pemerintah mengelola dan mendistribusikan bantuan sosial. Proses penyaluran bantuan sosial di Indonesia masih menghadapi sejumlah kendala mendasar, seperti subjektivitas dalam penilaian kelayakan, birokrasi yang berbelit, serta kurangnya transparansi dalam verifikasi data penerima. Tantangan ini sering kali menyebabkan bantuan tidak sampai ke sasaran yang tepat, sehingga tujuan utama program bantuan yaitu meningkatkan kesejahteraan masyarakat dan mengurangi tingkat kemiskinan tidak tercapai secara optimal [1].

Bantuan sosial (bansos) merupakan salah satu program intervensi pemerintah yang bertujuan melindungi individu, keluarga, atau kelompok masyarakat yang rentan secara sosial dan ekonomi [2] [3]. Meskipun alokasi anggaran bansos terus meningkat hingga mencapai Rp496,8 triliun pada tahun 2024, data Badan Pusat Statistik tahun 2023 menunjukkan bahwa tingkat kemiskinan nasional masih stagnan pada angka 9,36% [4]. Fakta ini mengindikasikan adanya ketimpangan antara besarnya dana yang disalurkan dan efektivitas program di lapangan. Berbagai

studi dan laporan lapangan juga menyoroti akar masalah utama, yakni ketidakakuratan data penerima bantuan [5], proses verifikasi manual yang lambat, serta birokrasi yang kompleks[6].

Banyak penelitian terdahulu telah membuktikan bahwa metode K-Nearest Neighbors (K-NN) adalah pendekatan yang menjanjikan untuk mengklasifikasikan kelayakan penerima bantuan sosial. Penelitian ini [7] telah menerapkan K-NN untuk menentukan penerima Bantuan Langsung Tunai (BLT) berdasarkan kondisi ekonomi mereka. Penelitian ini [8] juga menggunakan K-NN untuk klasifikasi bansos dengan mempertimbangkan atribut seperti pendapatan dan kepemilikan rumah. Sementara itu, penelitian ini [9] sukses mengklasifikasikan masyarakat prasejahtera dengan menggunakan data aset dan penghasilan.

Untuk menjawab tantangan tersebut, pendekatan berbasis *machine learning* mulai dikembangkan sebagai sistem pendukung keputusan yang lebih objektif dan efisien. Salah satu metode yang menonjol dalam konteks ini adalah *K-Nearest Neighbors* (K-NN). Algoritma K-NN bekerja dengan menganalisis kesamaan (jarak) antar data, di mana setiap calon penerima bantuan diklasifikasikan berdasarkan kedekatan karakteristik sosial-ekonominya terhadap data warga yang telah terverifikasi. Dengan demikian, keputusan kelayakan tidak lagi bergantung pada subjektivitas manusia, tetapi pada pola data yang terukur. Pendekatan ini diharapkan mampu meningkatkan akurasi klasifikasi, mempercepat proses verifikasi, dan meningkatkan transparansi dalam penyaluran bantuan sosial.

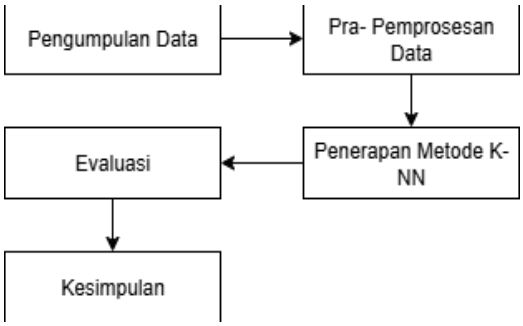
Sejumlah penelitian terdahulu telah membuktikan efektivitas K-NN dalam berbagai konteks klasifikasi penerima manfaat. Misalnya, penelitian [7] menggunakan K-NN untuk menentukan kelayakan penerima Bantuan Langsung Tunai (BLT) berdasarkan kondisi ekonomi, sementara [8] mengimplementasikannya pada klasifikasi bansos dengan mempertimbangkan atribut pendapatan dan kepemilikan rumah. Penelitian lain [9] menunjukkan bahwa K-NN mampu mengklasifikasikan masyarakat prasejahtera secara akurat berdasarkan aset dan penghasilan. Selain bidang bansos, efektivitas metode ini juga terbukti dalam seleksi keringanan Uang Kuliah Tunggal (UKT) [10], identifikasi penerima bantuan UMKM [11], klasifikasi status sosial ekonomi [12], serta seleksi penerima beasiswa [13] dan program sembako [14].

Meskipun hasil-hasil tersebut menunjukkan potensi besar, sebagian penelitian sebelumnya masih memiliki keterbatasan — terutama karena hanya menggunakan satu rasio pembagian data latih dan uji atau tidak melakukan eksplorasi terhadap variasi parameter *k* untuk menemukan konfigurasi optimal. Oleh karena itu, penelitian ini berupaya memberikan kontribusi baru dengan mengoptimalkan performa model K-NN melalui dua aspek utama: (1) pengujian beberapa rasio pembagian data latih-uji (80/20, 70/30, 60/40, dan 50/50) dan (2) evaluasi nilai parameter *k* yang berbeda.Pendahuluan menguraikan latar belakang permasalahan yang diselesaikan, isu-isu yang terkait dengan masalah yg diselesaikan, ulasan penelitan yang pernah dilakukan sebelumnya oleh peneliti lain yg relevan dengan penelitian yang dilakukan.

Fokus penelitian ini adalah mengembangkan model klasifikasi kelayakan penerima bantuan sosial yang lebih akurat dan adaptif terhadap variasi data di lapangan. Dengan pendekatan berbasis data dan algoritma K-NN yang mampu menilai kedekatan karakteristik sosial-ekonomi warga, penelitian ini diharapkan dapat menghasilkan sistem pendukung keputusan yang lebih transparan, objektif, dan efisien dalam penyaluran bantuan sosial di Kelurahan Kuto Batu.

2. METODE PENELITIAN

Penelitian ini dilakukan untuk mengkategorikan kelayakan penerima bantuan sosial menggunakan metode K-Nearest Neighbors (K-NN) di Kelurahan Kuto Batu sebagai lokasi penelitian. Pada Gambar 1 disajikan kerangka kerja pada penelitian ini.



Gambar 1. Tahapan Metode Penelitian

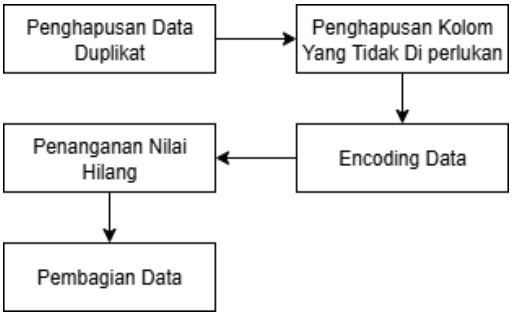
Gambar 1 sebagai *diagram alir keseluruhan penelitian*, yang meliputi lima tahap utama: pengumpulan data, pra-pemrosesan data, penerapan metode K-NN, evaluasi hasil, dan penarikan kesimpulan. Diagram ini menggambarkan urutan logis proses penelitian mulai dari input data mentah hingga diperolehnya hasil klasifikasi.

2.1 Pengumpulan Data

Data utama diperoleh dari Kelurahan Kuto Batu sebanyak 200 entri penerima bantuan sosial, data penelitian dengan atribut meliputi pekerjaan, kepemilikan rumah, kepemilikan simpanan, jenis atap, jenis dinding, jenis lantai, dan keterangan kelayakan. Data kemudian di olah menggunakan RStudio.

2. 2 Pra-Pemprosesan Data

Tahapan pra-pemrosesan data dilakukan untuk memastikan kualitas dan kelayakan data sebelum digunakan pada proses klasifikasi bisa dilihat di Gambar 2.



Gambar 2. Tahapan Pra-Pemprosesan Data

Gambar 2 memperlihatkan *subdiagram dari tahap pra-pemrosesan data*, yang merinci lima langkah penting, yaitu penghapusan data duplikat, penghapusan kolom yang tidak diperlukan, penanganan nilai hilang, *encoding* data kategori ke numerik, serta pembagian data menjadi data latih dan data uji. Diagram ini menunjukkan bagaimana data disiapkan agar sesuai untuk proses klasifikasi menggunakan algoritma K-NN.

1. Penghapusan Data Duplikat
   Pada tahap *data preprocessing*, baris data yang memiliki kesamaan nilai pada seluruh kolom dihapus karena tidak memberikan kontribusi informasi baru. Data duplikat berpotensi menimbulkan bias, memengaruhi keakuratan analisis, meningkatkan variabilitas hasil, serta mengurangi stabilitas model sehingga lebih rentan terhadap overfitting.
2. Penghapusan Kolom Yang Tidak di Perlukan
   Pada tahap ini, kolom yang tidak relevan atau memiliki kontribusi rendah terhadap tujuan analisis dihapus untuk mengurangi dimensi dataset. Langkah ini mempermudah proses analisis, meningkatkan efisiensi pengolahan data, serta mempercepat waktu komputasi,

terutama pada dataset berukuran besar, karena hanya data yang relevan yang akan diproses.

3. Encoding Data
   Encoding Data merupakan proses dalam *data preprocessing* yang mengonversi data kategorikal, seperti teks atau label, menjadi bentuk numerik agar dapat dipahami dan diolah secara optimal oleh sistem komputer.
4. Penanganan Nilai Hilang
   Pada tahap ini, data dengan nilai hilang ditangani untuk mencegah gangguan pada proses analisis. Nilai hilang dapat muncul akibat kesalahan input, survei yang tidak lengkap, atau data yang tidak tercatat. Penanganannya dilakukan melalui dua metode utama, yaitu menghapus baris jika jumlah nilai hilang relatif kecil, atau menggantinya dengan nilai modus jika jumlahnya cukup besar.
5. Pembagian Data
   Proses pembagian data menjadi *training set* dan *testing set* guna memastikan model mampu melakukan generalisasi terhadap data yang belum pernah dilihat. *Training set* digunakan untuk melatih model dalam mengenali pola atau hubungan antar variabel, sedangkan *testing set* berfungsi untuk mengukur kinerja model setelah proses pelatihan.
   Proses pembagian data menjadi *training set* dan *testing set* guna memastikan model mampu melakukan generalisasi terhadap data yang belum pernah dilihat. *Training set* digunakan untuk melatih model dalam mengenali pola atau hubungan antar variabel, sedangkan *testing set* berfungsi untuk mengukur kinerja model setelah proses pelatihan. Pembagian ini dilakukan secara acak untuk menjaga proporsionalitas distribusi kelas “Layak” dan “Tidak Layak” pada kedua subset data. Data latih digunakan untuk membantu model mempelajari pola dan karakteristik dari atribut-atribut yang mempengaruhi kelayakan penerima bantuan sosial, sedangkan data uji berfungsi untuk mengukur sejauh mana model mampu melakukan generalisasi terhadap data baru yang belum pernah dilihat sebelumnya. Pemisahan data yang seimbang sangat penting untuk menghindari bias prediksi pada salah satu kelas, sehingga hasil evaluasi dapat merefleksikan

2.3 Penerapan Metode K-NN

Metode K-Nearest Neighbors (K-NN) digunakan dalam penelitian ini untuk mengklasifikasikan kelayakan calon penerima bantuan sosial berdasarkan sejumlah atribut penting. Atribut-atribut tersebut meliputi pekerjaan, kepemilikan rumah, kepemilikan simpanan uang atau perhiasan, jenis atap, jenis dinding, jenis lantai, serta variabel target berupa keterangan kelayakan. Pemilihan K-NN didasarkan pada kemampuannya yang efektif dalam mengidentifikasi pola berdasarkan kedekatan antar data, tanpa memerlukan asumsi distribusi data tertentu.

2.4 Evaluasi

Evaluasi dilakukan menggunakan confusion matrix untuk menilai performa model K-NN setelah proses klasifikasi. Melalui confusion matrix, berbagai metrik evaluasi seperti akurasi, presisi, recall, dan F1-score dapat dihitung, sehingga memberikan gambaran komprehensif mengenai kemampuan model dalam mengklasifikasikan data [15]

True Positive (TP) merupakan jumlah kasus di mana model berhasil memprediksi kelas positif dengan benar. False Positive (FP) terjadi saat model memprediksi positif padahal kondisi sebenarnya negatif, yang disebut juga sebagai Type I Error. False Negative (FN) muncul ketika model memprediksi negatif padahal kondisi sebenarnya positif, dikenal sebagai Type II Error, yang berpotensi menyebabkan kegagalan dalam mendeteksi kondisi penting, seperti penyakit dalam bidang medis. Sedangkan True Negative (TN) adalah jumlah kasus di mana model secara tepat mengidentifikasi kelas negatif. Setelah membentuk confusion matrix, berbagai metrik penting dapat dihitung untuk menilai kinerja model secara menyeluruh:

1. Akurasi (*Accuracy*) adalah metrik yang digunakan untuk mengukur seberapa tepat model dalam memprediksi kelas data yang benar. Akurasi dihitung dengan membandingkan jumlah

prediksi yang benar (*True Positive* - TP dan *True Negative* - TN) terhadap jumlah total data yang diuji, yang mencakup TP, TN, *False Positive* (FP), dan *False Negative* (FN).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. Presisi (*Precision*) adalah metrik yang mengukur proporsi prediksi positif yang benar dari keseluruhan prediksi positif yang dihasilkan oleh model. Presisi dihitung sebagai rasio antara jumlah prediksi positif yang benar (*True Positive* - TP) dengan total prediksi positif yang dilakukan oleh model, yang mencakup TP dan *False Positive* (FP).

$$precision = \frac{TP}{TP+FP} \quad (2)$$

3. *Recall* adalah metrik yang mengukur sejauh mana model berhasil memprediksi kasus positif yang seharusnya dikenali. *Recall* dihitung sebagai rasio antara jumlah prediksi positif yang benar (*True Positive* - TP) dengan total jumlah kasus positif yang seharusnya diprediksi, yang terdiri dari TP dan *False Negative* (FN).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. *F1-Score* adalah metrik yang menggabungkan presisi dan *recall* menjadi satu nilai untuk memberikan gambaran yang lebih seimbang tentang kinerja model. *F1-Score* dihitung sebagai rata-rata harmonik antara presisi dan *recall*.

$$F1 - score = \frac{precision\ x\ recall}{precision+recall} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data untuk penelitian ini dikumpulkan langsung dari Kelurahan Kuto, fokus utama studi ini. Kami mendapatkan 200 data penerima bantuan sosial, data yang dimabil dari periode januari 2024 - januari 2025 dengan tetap menjaga kerahasiaan informasi pribadi warga. Tabel 1 ini mencakup berbagai atribut seperti Provinsi, Kabupaten/Kota, Kecamatan, Desa/Kelurahan, Kepala Keluarga, Pekerjaan, Kepemilikan Rumah, Memiliki Simpanan Uang/Perhiasan, Jenis Atap, Jenis Dinding, Jenis Lantai, dan Keterangan.

Tabel 1. Pengumpulan Data

No	Kepala Keluarga	Pekerjaan	Kepemilikan Rumah	Memiliki Simpanan Uang/Perhiasan	Jenis Atap	Jenis Dinding	Jenis Lantai	Keterangan
1	Tomi hidayat	Buruh	Kontrak/Sewa	Tidak	Asbes Seng	Kayu/ Papan	Semen	layak
2	Efen	Buruh	Kontrak/Sewa	Tidak	Asbes Seng	Kayu/ Papan	Kayu/Papan	layak
3	Sazi	Pegawai Swasta	Bebas Sewa	Ya	Genteng	Tembok	Keramik /Granit/ Ubin	tidaklayak

No	Kepala Keluarga	Pekerjaan	Kepemilikan Rumah	Memiliki Simpanan Uang/Perhiasan	Jenis Atap	Jenis Dinding	Jenis Lantai	Keterangan
4	Hibah	Wiraswsta	Milik Sendiri	Ya	Genteng	Tembok	Keramik /Granit/ Ubin	tidaklayak
:	:	:	:	:	:	:	:	:
200	Indra	Wiraswsta	Milik Sendiri	Ya	Asbes Seng	Tembok	Keramik /Granit/ Ubin	tidaklayak

3.2 Pra-Pemprosesan Data

Proses ini sangat penting untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam analisis. Pada penelitian ini menggunakan dataset yang ada di kelurahan kuto batu, diketahui bahwa dataset tersebut tidak mengandung data yang teridentifikasi sebagai duplikat dan juga tidak terdapat nilai yang hilang dalam dataset. Pada dataset warga kelurahan kuto batu, kolom seperti Nama kepala keluarga, diidentifikasi sebagai kolom yang tidak relevan untuk analisis lebih lanjut. Kolom tersebut dihapus karena informasi yang dikandungnya tidak memengaruhi hasil analisis atau pengembangan model. Adapun hasil dari penghapusan kolom yang tidak digunakan dapat dilihat pada Tabel 2.

Tabel 2. Hasil Penghapusan Kolom

No	Pekerjaan	Kepemilikan Rumah	Memiliki Simpananan Uang/Perhiasan	Jenis Atap	Jenis Dinding	Jenis Lantai
1	Buruh	Kontrak/Sewa	Tidak	AsbesSeng	Kayu/Papan	Semen
2	Buruh	Kontrak/Sewa	Tidak	AsbesSeng	Kayu/Papan	Kayu/Papan
3	Pegawai Swasta	BebasSewa	Ya	Genteng	Tembok	Keramik/Granit/Ubin
4	Wiraswasta	Milik Sendiri	Ya	Genteng	Tembok	Keramik/Granit/Ubin
5	Buruh	Menumpang	Ya	Genteng	Kayu/Papan	Kayu/Papan
6	Buruh	Kontrak/Sewa	Ya	AsbesSeng	Kayu/Papan	Kayu/Papan
7	Pedagang	Menumpang	Tidak	AsbesSeng	Kayu/Papan	Kayu/Papan
8	Tidak/belum bekerja	Kontrak/Sewa	Tidak	AsbesSeng	Kayu/Papan	Kayu/Papapn
:	:	:	:	:	:	:
200	Wiraswasta	Milik sendiri	Ya	AsbesSeng	Tembok	Keramik/Granit/Ubin

Selanjutnya encoding data merupakan proses dalam *data preprocessing* yang mengonversi data kategorikal, seperti teks atau label, menjadi bentuk numerik agar dapat dipahami dan diolah secara optimal oleh sistem komputer. Transformasi ini sangat dibutuhkan karena sebagian besar machine learning hanya dapat memproses angka untuk komputasi dan analisis statistik. Sebagai contoh, sebuah model tidak bisa langsung memahami kategori tekstual seperti "Pedagang" atau "Buruh"; khusus untuk metode K-Nearest Neighbors (KNN) yang mengandalkan perhitungan jarak, data harus berupa angka. Bisa dilihat di Tabel 3 merupakan hasil dari encoding data.

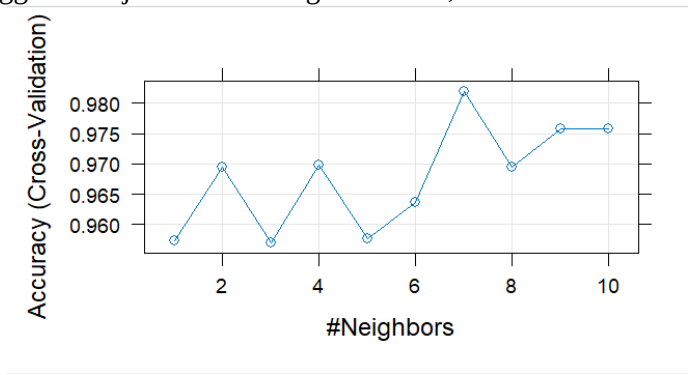
Tabel 3. Hasil Encoding Data

No	Pekerjaan	Kepemilikan Rumah	Memiliki Simpananan Uang/Perhiasan	Jenis Atap	Jenis Dinding	Jenis Lantai
1	1	1	0	1	0/	1
2	1	1	0	1	0	0
3	4	2	1	2	1	2
4	3	3	1	2	1	2
5	1	0	1	2	0	0
6	1	1	1	1	0	0
7	2	2	0	1	0	0
8	0	1	0	1	0	0
:	:	:	:	:	:0	:
200	Wiraswasta	3	1	1	1	2

Selanjutnya pembagian dataset salah satu Langkah penting dalam pengembangan model, proses ini membagi dataset menjadi dua bagian yaitu data latih dan data uji. Tujuan nya untuk memastikan bahwa model dapat melakukan generalisasi terhadap data baru dari pada hanya hanya menghafal data latih. Pada penelitian ini, dataset dibagi dengan beberapa proporsi, yaitu 80:20, 70:30, 60:40, dan 50:50, di mana persentase pertama digunakan untuk pelatihan (data latih) dan persentase kedua digunakan untuk pengujian (data uji). Setiap distribusi data dilakukan secara acak.

3.3 Penerapan Metode K-NN

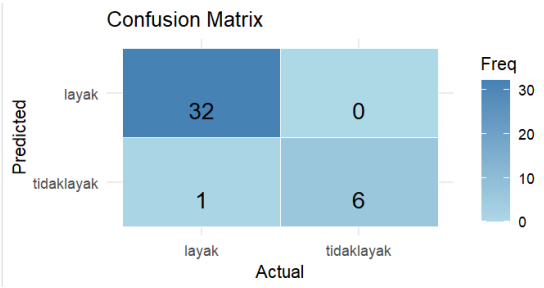
Metode K-NN digunakan dengan nilai K = 7, yang diperoleh melalui 10-fold cross-validation. Perhitungan jarak dilakukan menggunakan Euclidean Distance. Proses klasifikasi dilakukan berdasarkan mayoritas kelas dari tiga tetangga terdekat. Implementasi dilakukan menggunakan R studio menggunkan K tertinggi. Di Gambar 6 di jelaskan bahwa mendapatkan hasil akurasi tertinggi menunjukan k = 7 dengan hasil 97,80%



Gambar 3. Hasil Menentukan K tertinggi

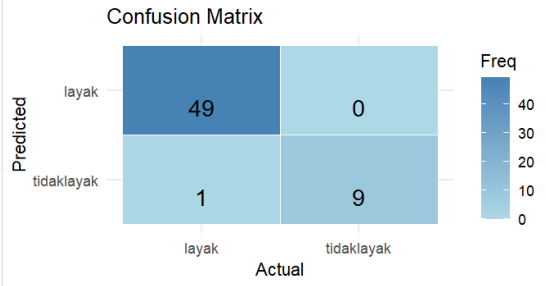
3.4 Pengujian

Setelah model KNN selesai dilatih dan nilai parameter optimalnya telah ditentukan, langkah selanjutnya adalah mengevaluasi kinerjanya. Tahap ini krusial untuk memahami seberapa baik model dapat membuat prediksi pada data yang belum pernah dilihat sebelumnya. Pada pengujian pertama dengan membagi 80/20 dapat dilihat di Gambar 4.



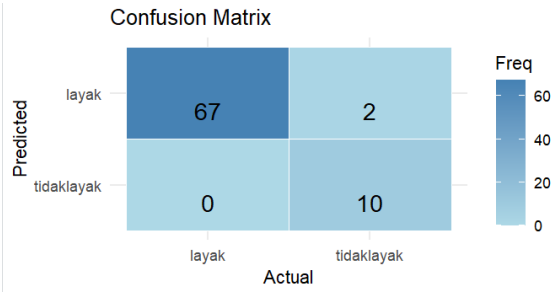
Gambar 4. Hasil Confusion Matrix 80/20

Pengujian kedua menunjukan hasil confusion matrix pada pengujian 70/30 bisa dilihat di Gambar 5.



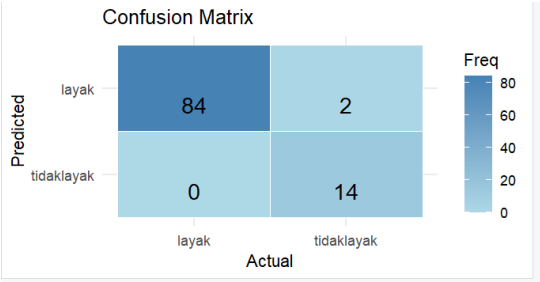
Gambar 5. Hasil Confusion Matrix 70/30

Pengujian ketiga menunjukan hasil confusion matrix dari 60/40 bisa dilihat di Gambar 6.



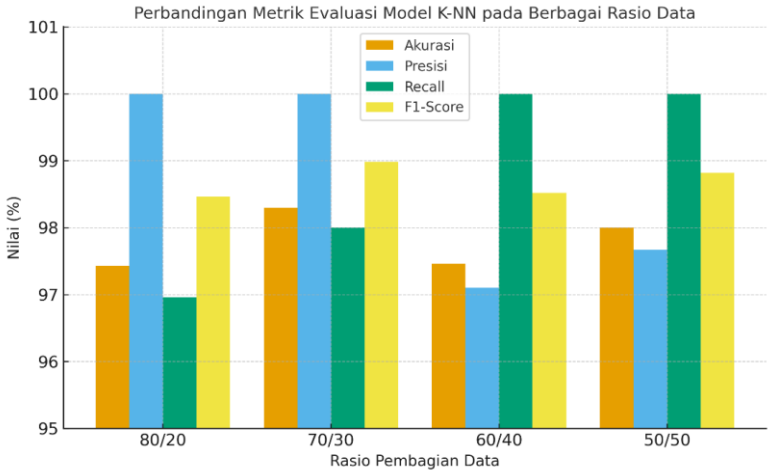
Gambar 6. Hasil Confusion Matrix 60/40

Pengujian terakhir menunjukan hasil confusion matrix dari 50/50 bisa dilihat di Gambar 7.



Gambar 7. Hasil confusion Matrix 50/50

Selanjutnya melakukan evaluasi kinerja model k-NN. Tujuan evaluasi kinerja model untuk mengevaluasi kinerja klasifikasi pada empat scenario pembagian data uji dan data latih yaitu 80/20, 70/30, 60/40, 50/50. Untuk mengetahui bagaimana perubahan proporsi data mempengaruhi kemampuan model untuk mengenali pola. Di Tabel 4 merupakan hasil evaluasi metrix yang di gunakan Akurasi, presisi, recall,dan f1-Score.



Gambar 8. Hasil perbandingan metrik evaluasi model K-NN

Grafik perbandingan metrik evaluasi (Akurasi, Presisi, Recall, dan F1-Score) pada empat rasio pembagian data (Gambar 8). Terlihat bahwa rasio 70/30 memberikan hasil paling optimal dan stabil di semua metrik performa model.

Setelah diuji dengan berbagai rasio pembagian data, performa algoritma K-NN bervariasi. Rasio 70%:30% menghasilkan kinerja model yang paling optimal, menunjukkan keseimbangan yang baik antara metrik akurasi, presisi, recall, dan F1-score.

Hal ini terjadi karena rasio tersebut memberikan data latih yang cukup bagi model untuk belajar, sekaligus menyediakan data uji yang memadai untuk mengukur kemampuan model dalam mengenali pola pada data baru. Gambar 8 menampilkan grafik akurasi hasil dari kempat skenario

Gambar 8 Hasil Grafik Akurasi Keseluruhan

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa penggunaan metode K-Nearest Neighbors (K-NN) terbukti efektif dalam mengklasifikasikan kelayakan penerima bantuan sosial di Kelurahan Kuto Batu. Pengujian dilakukan melalui beberapa skenario pembagian data untuk menemukan proporsi data pelatihan dan data pengujian yang paling optimal, yaitu pada rasio 80/20, 70/30, 60/40, dan 50/50. Hasil evaluasi menunjukkan akurasi model sebesar 97,44% pada pembagian data 80/20, 98,30% pada 70/30, 97,10% pada 60/40, dan 98,00% pada 50/50. Di antara semua skenario, pembagian data 70/30 menghasilkan akurasi tertinggi yaitu 98,30%, disertai dengan presisi sebesar 100%, recall 98%, dan nilai F1-Score sebesar 98,98%. Nilai-nilai tersebut menandakan bahwa model memiliki performa sangat baik dalam membedakan antara data penerima bantuan yang layak dan tidak layak.

5. SARAN

Berdasarkan hasil penelitian, penulis memiliki beberapa saran yang dapat digunakan sebagai bahan pertimbangan untuk penelitian selanjutnya:

1. Melakukan validasi model dengan set data yang lebih besar dan bervariasi untuk memastikan robustness model dalam berbagai skenario data.
2. Menerapkan pemantauan kinerja model secara berkala dan pembaruan data untuk menjaga akurasi seiring waktu.
3. Mengeksplorasi perbandingan kinerja K-NN dengan metode machine learning lainnya untuk mengidentifikasi potensi peningkatan efektivitas.

DAFTAR PUSTAKA

[1] K. D. Irianto, “Design of Smart Farm Irrigation Monitoring System Using IoT and LoRA,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 1, pp. 47–56, 2021.

[2] R. Fadilah, “Bantuan Sosial Sembako dan Bantuan Sosial Tunai,” *J. El-Thawalib*, vol. 2, no. 3, pp. 167–179, 2021, doi: 10.24952/el-thawalib.v2i3.3992.

[3] I. D. Utami, *Pemodelan Sistem*. Malang: Media Nusa Creative, 2021.

[4] Fithria and Annisa, “Bansos, demokrasi,dan merawat kemiskinan.” [Online]. Available: https://news.uad.ac.id/bansos-demokrasi-dan-upaya-merawat-kemiskinan%0A

[5] F. P. Lestari, M. Haekal, R. Edmi Edison, F. Ravi Fauzy, S. Nurul Khotimah, and F. Haryanto, “Epileptic Seizure Detection in EEGs by Using Random Tree Forest, Naïve Bayes and KNN Classification,” *J. Phys. Conf. Ser.*, vol. 1505, no. 1, 2020, doi: 10.1088/1742-6596/1505/1/012055.

[6] A. R. Wijaya, I. Lazulfa, M. F. Rizal, and A. Andriani, “Implementasi Metode K-Nearest Neighbor (Knn) Untuk Menentukan,” vol. 0, pp. 165–172, 2024.

[7] R. Hamonangan, R. K. Sari, S. Anwar, and T. Hartati, “Klasifikasi Algoritma KNN dalam menentukan Penerima Bantuan Langsung Tunai,” vol. 6, no. 1, pp. 198–204, 2024.

[8] P. Pahrudin and K. Harianto, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1241–1245, 2022, doi: 10.47065/bits.v4i3.2276.

[9] A. Khairi, A. F. Ghozali, and A. D. N. Hidayah, “Implementasi K-Nearest Neighbor (KNN) untuk Mengklasifikasi Masyarakat Pra-Sejahtera Desa Sapikerep Kecamatan Sukapura,” *TRILOGI J. Ilmu Teknol. Kesehatan, dan Hum.*, vol. 2, no. 3, pp. 319–323, 2021, doi: 10.33650/trilogi.v2i3.2878.

[10] P. M. Hasan, N. Ayu, and R. A. Saputra, “Klasifikasi Keringanan Ukt Mahasiswa Uho Menggunakan K-Nearest Neighbor (KNN),” vol. 8, no. 6, pp. 11939–11945, 2024.

[11] A. Zulkarnain and A. B. Pohan, “Klasifikasi Penerima Bantuan Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Algoritma K-Nearest Neighbors,” *Biomaterials*, vol. 07, no. 12, pp. 85–90, 2023.

[12] R. Fitra and I. Rusdi, “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 Pint. Komput.*, vol. 11, no. 4, pp. 653–660, 2022, doi: 10.30591/smartcomp.v11i4.4250.

[13] S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita, “Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 6, no. 2, pp. 118–127, 2021, doi: 10.31294/ijcit.v6i2.10438.

[14] H. Yozza, N. M. Azizah, L. Yulianti, and I. RAHMI, “The Classification of ‘Program Sembako’ recipients in Payobasung West Sumatra based on the K-nearest neighbors classifier,” *J. Nat.*, vol. 23, no. 2, pp. 83–91, 2023, doi: 10.24815/jn.v23i2.29738.

[15] Nazori Suhandi, Rendra Gustriansyah, and A. Destria, “Klasifikasi Penyakit TBC Menggunakan Metode UMAP dan K-NN,” *bit-Tech*, vol. 7, no. 3, pp. 843–852, 2025, doi: 10.32877/bt.v7i3.2227.