

Optimasi Penyaluran Bantuan Langsung Tunai Menggunakan Algoritma *K-Nearest Neighbor* di Desa Penedagador

Amrina Rosyida^{*1}, Bulkis Kanata², Cipta Ramadhani³

^{1,2,3}Jurusan Teknik Elektro, Fakultas Teknik, Universitas Mataram, Indonesia

Email: amrinarosyida04@gmail.com, uqikanata@unram.ac.id, ramadhani.cipta@gmail.com

Abstrak

Kemiskinan merupakan salah satu tantangan terbesar yang masih terus dihadapi semua bangsa termasuk Indonesia. Sesuai amanat konstitusi dalam pasal 34 UUD 1945, pemerintah terus berupaya melindungi dan memberdayakan masyarakat miskin melalui berbagai program strategis. Salah satu program yang ditawarkan pemerintah yaitu Bantuan Langsung Tunai. Tetapi karna ketersediaan sssdata sasaran masyarakat miskin yang kurang sesuai, menyebabkan program berjalan kurang optimal karena perlu verifikasi dan validasi ulang. Distribusi Bantuan Langsung Tunai (BLT) di Indonesia masih menghadapi kendala akibat ketidaktepatan data penerima, yang menyebabkan inefisiensi dalam penyalurannya. Kesalahan dalam data sasaran mengharuskan adanya proses verifikasi dan validasi ulang, sehingga menghambat efektivitas program. Penelitian ini bertujuan untuk mengoptimalkan penyaluran BLT dengan menerapkan metode klasifikasi *K-Nearest Neighbor* (KNN). Penelitian ini bertujuan untuk mengoptimalkan penyaluran BLT menggunakan pendekatan *K-Nearest Neighbor*, dengan 714 dataset dan empat atribut yaitu kondisi rumah, penghasilan, pekerjaan dan jumlah keluarga. Prinsip kerja KNN adalah mencari jarak terdekat antara data baru yang akan diuji dengan data latihnya. Untuk mencari jarak terdekat digunakan euclidean distance dengan 3 nilai *k* yang berbeda yaitu $k=35$, $k=45$, dan $k=55$. Hasil yang didapatkan untuk akurasi, presisi dan recall yang tertinggi yaitu menggunakan parameter $k=35$ memberikan performa terbaik dengan nilai akurasi sebesar, 95%, presisi sebesar 95% dan recall sebesar 100%. Penerapan metode ini diharapkan dapat meningkatkan efisiensi distribusi BLT dengan memastikan bahwa bantuan tepat sasaran, sehingga mendukung kebijakan pengentasan kemiskinan secara lebih efektif.

Kata kunci: *Bantuan Langsung Tunai, Data Mining, Euclidean distance, Klasifikasi, K-Nearest Neighbor*

Optimization of Direct Cash Assistance Distribution Using the K-Nearest Neighbor Algorithm in Penedagador Village

Abstract

Poverty is one of the biggest challenges that all nations including Indonesia continue to face. In accordance with the mandate of the constitution in article 34 of the 1945 Constitution, the government continues to strive to protect and empower the poor through various strategic programs. One of the programs offered by the government is Direct Cash Assistance. However, due to the availability of data targeting the poor that is not suitable, the program is not running optimally because it needs to be verified and revalidated. The distribution of Direct Cash Assistance (BLT) in Indonesia still faces obstacles due to inaccurate recipient data, which causes inefficiencies in its distribution. Errors in target data require a verification and revalidation process, thus hindering the effectiveness of the program. This study aims to optimize the distribution of BLT by applying the *K-Nearest Neighbor* (KNN) classification method. This study aims to optimize the distribution of BLT using the *K-Nearest Neighbor* approach, with 714 datasets and four attributes, namely home condition, income, work and family size. The working principle of KNN is to find the closest distance between the new data to be tested and the training data. To find the nearest distance, euclidean distance is used with 3 different *k* values, namely $k=35$, $k=45$, and $k=55$. The results obtained for the highest accuracy, precision and recall, namely using the $k=35$ parameter provide the best performance with an accuracy value of 95%, precision of 95% and recall of 100%. The application of this method is expected to increase the efficiency of BLT distribution by ensuring that assistance is on target, thereby supporting poverty alleviation policies more effectively.

Keywords: *Classification, Direct Cash Assistance, Data Mining, Euclidean Distance, K-Nearest Neighbor.*

1. PENDAHULUAN

Kebutuhan dasar masyarakat mencakup kebutuhan makanan maupun non-makanan. [1] Menurut Pomati et al. (2024) bahwa kemiskinan menyoroti kurangnya akses masyarakat ke sumber daya yang memadai baik itu material, sosial, adat serta memiliki aset dan kemampuan untuk mengamankan standar hidup dan kesejahteraan minimum. Kemiskinan juga ditandai dengan kurangnya pendidikan, kesehatan, ketidakberdayaannya ikut serta pada pembangunan dalam berbagai masalah yang berkaitan dengan pembangunan Sumber Daya Manusia. [2]

Dalam menanggulangi kemiskinan, pemerintah melakukan berbagai upaya, salah satunya yaitu melalui pengalokasian Dana Desa. Tujuan mendasar dari dana desa, sebagaimana diatur dalam Undang-Undang Republik Indonesia No. 6/2014 tentang Desa (2014), yaitu untuk meningkatkan kesejahteraan masyarakat secara umum dan mengurangi kemiskinan. [3] Dana desa adalah anggaran yang diperuntukkan bagi desa yang ditransfer melalui APBD kabupaten/kota yang digunakan untuk membiayai penyelenggaraan pemerintah, pembangunan, serta pemberdayaan masyarakat, dan kemasyarakatan. Melalui Anggaran Dana Desa maka Bupati Lombok Timur memprioritaskan anggaran untuk membantu warga Lombok Timur melalui Bantuan Langsung Tunai (BLT). [4]

Distribusi Bantuan Langsung Tunai (BLT) yang tepat sasaran merupakan aspek krusial dalam upaya pemerintah mengentaskan kemiskinan dan meningkatkan kesejahteraan masyarakat. Ketepatan dalam penyaluran BLT memastikan bahwa bantuan tersebut diterima oleh individu atau keluarga yang benar-benar membutuhkan, sehingga dapat meminimalkan ketimpangan sosial dan ekonomi. Namun, dalam praktiknya, seringkali terjadi permasalahan seperti data penerima yang tidak akurat, penyaluran yang tidak tepat sasaran, serta keterbatasan sumber daya dalam memverifikasi kelayakan penerima. Permasalahan ini dapat mengurangi efektivitas program bantuan sosial dan menimbulkan ketidakpercayaan masyarakat terhadap pemerintah. Untuk mengatasi permasalahan tersebut, berbagai metode klasifikasi data telah diterapkan dalam menentukan penerima bantuan sosial yang layak. Salah satu metode yang umum digunakan adalah algoritma Naive Bayes. Penelitian oleh Alfiah (2021) mengaplikasikan algoritma ini dalam klasifikasi penerima Program Keluarga Harapan (PKH) dan menunjukkan bahwa Naive Bayes mampu mengklasifikasikan data dengan tingkat akurasi yang memadai. [5]

Selain itu, algoritma C4.5 juga telah digunakan untuk menentukan kelayakan penerima bantuan sosial. Junaidi (2021) dalam penelitiannya mengimplementasikan algoritma C4.5 dan menemukan bahwa metode ini efektif dalam mengklasifikasikan data penerima bantuan dengan mempertimbangkan berbagai kriteria seperti pendapatan bulanan dan jumlah tanggungan keluarga. [6] Meskipun kedua metode tersebut memiliki keunggulan, mereka juga memiliki keterbatasan. Algoritma Naive Bayes, misalnya, mengasumsikan independensi antar fitur, yang dalam kenyataannya seringkali tidak terpenuhi, sehingga dapat mempengaruhi akurasi klasifikasi. Sementara itu, algoritma C4.5 dapat menghasilkan pohon keputusan yang kompleks dan sulit diinterpretasikan ketika berhadapan dengan data yang memiliki banyak atribut atau nilai yang kontinu.

Sebagai alternatif, algoritma K-Nearest Neighbor (KNN) menawarkan pendekatan yang lebih sederhana dan intuitif dalam klasifikasi data. KNN bekerja dengan mengklasifikasikan objek berdasarkan kedekatannya dengan sejumlah k tetangga terdekat dalam ruang fitur. Keunggulan utama KNN adalah kemampuannya untuk menangani data dengan distribusi yang tidak diketahui tanpa memerlukan asumsi khusus tentang distribusi data. Selain itu, KNN mudah diimplementasikan dan dapat digunakan untuk berbagai jenis masalah klasifikasi. Namun, KNN juga memiliki kelemahan, antara lain sensitivitas terhadap skala dan relevansi fitur, serta kebutuhan komputasi yang tinggi terutama pada dataset berukuran besar.

Namun, dalam pelaksanaannya juga, distribusi BLT masih menghadapi berbagai permasalahan, seperti ketidaktepatan sasaran dan kurangnya transparansi dalam pendataan penerima manfaat. Oleh karena itu, diperlukan metode klasifikasi yang efektif untuk memastikan bahwa bantuan sosial diberikan kepada individu yang benar-benar membutuhkan. Salah satu metode yang dapat digunakan dalam klasifikasi data penerima bantuan sosial adalah Data Mining. Data mining menjadi salah satu pendekatan yang digunakan dalam menganalisis dan mengidentifikasi pola dari data penerima bantuan sosial. [7] Data mining adalah suatu proses yang melibatkan penggunaan matematika, teknik statistika, machine learning, dan kecerdasan buatan untuk mengekstraksi serta mengidentifikasi informasi dan pengetahuan yang bermanfaat dari berbagai database berukuran besar. [8] KDD adalah proses terorganisir untuk mengidentifikasi pola yang valid, baru, berguna, dan dapat dimengerti dari kumpulan data besar dan kompleks. Data Mining (DM) adalah inti dari proses KDD, yang melibatkan inferring algoritma yang mengeksplorasi data, mengembangkan model dan menemukan pola yang sebelumnya tidak diketahui. Model ini digunakan untuk memahami fenomena dari data, analisis dan Prediksi. [9]

Salah satu metode dalam data mining yang dapat digunakan untuk klasifikasi penerima BLT adalah algoritma K-Nearest Neighbor (KNN). Menurut Halim dan Anraeni (2021), Algoritma K-Nearest Neighbor (KNN) merupakan algoritma klasifikasi berdasarkan kedekatan jarak suatu data dengan data yang lain. [10] Dalam konteks klasifikasi penerima bantuan sosial, KNN dapat digunakan untuk mengidentifikasi calon penerima berdasarkan karakteristik yang telah ditentukan sebelumnya, seperti tingkat pendapatan, kondisi tempat tinggal, serta jumlah tanggungan dalam keluarga. Pada algoritma KNN, data berdimensi q , jarak dari data tersebut ke data yang lain

dapat dihitung. Nilai jarak inilah yang digunakan sebagai nilai kedekatan/kemiripan antara data uji dengan data latih. Nilai K pada KNN berarti K-data terdekat dari data uji. Algoritma *K-Nearest* Algoritma KNN adalah metode regresi non-parametrik yang digunakan untuk memprediksi variabel target berdasarkan rata-rata nilai target k tetangga terdekatnya. [11]

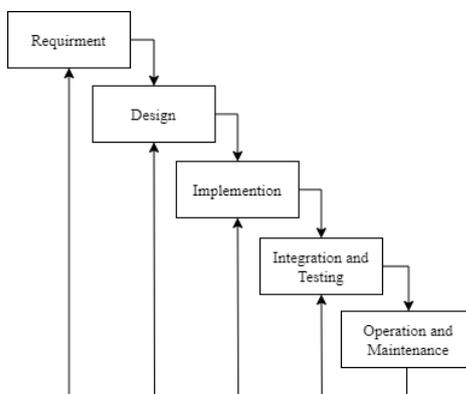
Sebelumnya, berbagai metode telah digunakan dalam klasifikasi data penerima bantuan sosial, seperti Decision Tree (DT), Support Vector Machine (SVM), dan Naïve Bayes. Decision Tree memiliki keunggulan dalam kemampuannya menginterpretasikan hasil dengan mudah serta memberikan aturan klasifikasi yang jelas, namun metode ini cenderung kurang optimal ketika menghadapi data dengan dimensi yang sangat besar. Sementara itu, Support Vector Machine menawarkan akurasi tinggi dalam klasifikasi, terutama pada data dengan margin yang jelas antara kelas yang berbeda, tetapi memerlukan komputasi yang lebih tinggi dibandingkan dengan KNN. Naïve Bayes, di sisi lain, memiliki keunggulan dalam analisis data berbasis probabilitas dan bekerja dengan baik pada data kategorikal, tetapi memiliki keterbatasan dalam menangani data yang memiliki hubungan antar variabel yang kompleks. Dibandingkan dengan metode lain, KNN memiliki kelebihan dalam kesederhanaannya serta kemampuannya dalam menangani data non-linear. [12]

Selain itu, KNN tidak memerlukan proses pelatihan yang kompleks karena klasifikasinya hanya bergantung pada data yang tersedia saat proses prediksi. Namun, kelemahan utama dari metode ini adalah kepekaannya terhadap pemilihan parameter K, serta waktu komputasi yang relatif lebih tinggi ketika bekerja dengan jumlah data yang besar. Menurut Astuti dan Nuris (2022), didapatkan hasil dari melakukan klasifikasi dengan metode *K-Nearest Neighbor* dalam menganalisis tinjauan sentimen aplikasi PeduliLindungi yang terbukti bahwa dengan metode ini mendapatkan klasifikasi nilai akurasi yang baik dimana nilai akurasi diperoleh sebesar 81,72%. [13] Sedangkan hasil penelitian Bahtiar (2023) melakukan prediksi penjualan kusen terlaris dengan metode *K-Nearest Neighbor*, dengan hasil prediksi yang didapatkan yaitu sebesar 88.89% pada data penjualan dan 80.00% pada data bahan terpakai yang bisa dikatakan sebagai hasil prediksi yang cukup baik. [14]

Dalam konteks penyaluran BLT di Desa Penedagador, penerapan algoritma KNN dapat menjadi solusi yang efektif untuk mengoptimalkan proses klasifikasi penerima bantuan. Dengan mempertimbangkan karakteristik demografis dan ekonomi masyarakat setempat, KNN dapat membantu dalam mengidentifikasi individu atau keluarga yang paling membutuhkan bantuan secara akurat. Namun, penting untuk memastikan bahwa data yang digunakan telah melalui proses normalisasi dan seleksi fitur yang tepat guna meningkatkan kinerja algoritma dan mengurangi beban komputasi. Dengan mempertimbangkan berbagai metode klasifikasi yang telah digunakan sebelumnya, pemanfaatan algoritma KNN dalam distribusi BLT dapat memberikan keuntungan dalam hal fleksibilitas dan akurasi dalam menentukan penerima bantuan. Namun, perlu dilakukan optimalisasi pada pemilihan parameter K serta peningkatan efisiensi dalam menangani data skala besar agar hasil klasifikasi tetap akurat dan efisien.

2. METODE PENELITIAN

Metode penelitian ini menggunakan pendekatan air terjun (*waterfall approach*) dalam pengembangan sistem aplikasi berbasis website ini. Pendekatan air terjun (*waterfall approach*) pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Waterfall Approach

Pada gambar diatas menggambarkan pengembangan system yang dimulai dari tahap requirement dimana pengembang harus mengetahui seluruh informasi mengenai kebutuhan software yang diinginkan oleh pengguna. Tahap design dilakukan untuk memberikan gambaran lengkap tentang apa yang harus dikerjakan dan bagaimana tampilan dari system yang ingin dibangun. Sedangkan pada tahap implementation atau coding ini akan dilakukan penulisan code, dimana pembuatan software akan dipecah menjadi modul modul kecil yang nantinya akan digabungkan dalam tahap selanjutnya. Pada tahap integration & testing akan dilakukan penggabungan modul-modul yang sudah dibuat sebelumnya dan akan dilakukan pengujian untuk mengetahui apakah software sudah sesuai dengan desain yang diinginkan. Tahap terakhir yaitu operation & maintenance, pada tahap ini software yang dibuat sudah jadi dan siapa dioperasikan oleh pengguna dan akan dilakukan pemeliharaan berupa perbaikan kesalahan, perbaikan implementasi unit sistemnya dan peningkatan jasa system sesuai kebutuhan.

2.1 Teknik Pengolahan Data

Pengumpulan data dilakukan dengan meminta data langsung diperangkat Desa Penedagandor dan melakukan wawancara atau survei langsung kelapangan. Pada tahapan ini peneliti akan mengolah data set yang didapatkan menggunakan algoritma *K-Nearest Neighbor* dengan gambaran atribut berupa: kondisi rumah, penghasilan, pekerjaan dan jumlah tanggungan.

a. Analisa Data

Setelah pengumpulan data selesai, maka akan dilakukan tahap analisa data dengan cara data selection dan transformasi. Pada tahap data selection, data atribut akan diberi inisialisasi seperti pada Tabel 1.

Tabel 1. Inisialisasi Atribut

Atribut	Keterangan
KR	Kondisi Runah
PH	Penghasilan
PK	Pekerjaan
JT	Jumlah Keluarga

Tabel 2. Hasil Proses Transformasi

Data Kondisi Rumah	Nilai Bobot
Beton	0
Tembok	1
Batu Bata	2
Batako	3
Bambu	4
Data Penghasilan	Nilai Bobot
>5.000.000	0
3.000.001-5.000.000	1
1.000.001-3.000.000	2
500.000-1.000.000	3
<500.000	4
Data Pekerjaan	Nilai Bobot
PNS	0
Wiraswasta	1
Petani	2
Buruh	3
Tidak Bekerja	4
Data Jumlah Keluarga	Nilai Bobot
1 orang	0
2-3 orang	1
4-5 orang	2
6-7 orang	3
>7 orang	4

Selanjutnya dilakukan proses transformasi yang akan mengubah skala pengukuran data asli menjadi bentuk lain sehingga data dapat digunakan. Hasil proses transformasi dapat dilihat pada Tabel 2.

Pada tabel 2 terdapat data kondisi rumah, data penghasilan, data pekerjaan dan data jumlah keluarga beserta nilai bobotnya masing masing dimana nilai bobot 0 merupakan nilai yang paling rendah dan nilai bobot 4 merupakan nilai bobot yang paling besar.

Tabel 3. Data Kelas

Kelas
Layak
Tidak Layak

Pada tabel 3 terdapat data kelas yang digunakan untuk mengklasifikasikan apakah penerima BLT antara layak dan tidak layak.

b. Inisialisasi Nilai K

Pada penginisialisasi nilai k dilakukan secara acak atau tidak tentu, karena tidak ada metode pasti untuk menemukan k terbaik untuk mencari hasil ketetanggaan yang baik juga. Maka, saat menginisialisasi nilai k (jumlah tetangga yang paling dekat), k yang digunakan adalah k dengan nilai ganjil.

c. Menghitung Euclidean Distance

Selanjutnya akan dilakukan perhitungan Euclidean Distance dari masing-masing record terhadap data testing yang diberikan. Untuk menghitung Euclidean distance menggunakan persamaan (1) sebagai berikut [15]:

$$j(t, u) = \sum_{k=1}^n \sqrt{(t - u)^2} \tag{1}$$

Menerangkan bahwa t merupakan nilai pada data training, u merupakan nilai pada data testing, n adalah banyak data, k merupakan atribut individu antara 1 dan n, dan j(t,u) adalah jarak antara data training dan data testing.

d. Pengurutan Hasil Perhitungan (*Ascending*)

Pada proses pengurutan, data akan diurutkan berdasarkan nilai jarak *Euclidean* yang terkecil ke yang terbesar (*ascending*), kemudian akan dimasukkan dalam kelompok anggota k sesuai dengan nilai yang sudah ditentukan. Jika nilai k=35 maka diambil 35 urutan paling kecil dari nilai *Euclidean* yang didapatkan.

e. Hasil Prediksi

Hasil prediksi ini didapatkan dari hasil dari data latih yang digunakan sebagai model dalam melakukan proses prediksi dari data uji.

f. Pengujian Algoritma Menggunakan *Confusion Matrix*

Confusion Matrix merupakan suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining atau sistem pendukung keputusan. Tabel 7 menunjukkan *Confusion Matrix* [16]:

Tabel 4. Confusion Matrix

		Terklasifikasi positif		Terklasifikasi negatif	
kelas		+		-	
<i>Positif</i>	+	<i>True positif</i>		<i>False negatif</i>	
<i>Negative</i>	-	<i>False positif</i>		<i>True negatif</i>	

Pada Tabel 4 menerangkan bahwa *True positif* (TP) merupakan jumlah *recod positif* yang diklasifikasikan positif, *True negative* (TN) merupakan jumlah *record positif* yang diklasifikasikan sebagai negative, *False Positif* (FP) merupakan jumlah *record negative* yang diklasifikasikan sebagai positif dan *False negative* (FN) merupakan jumlah *record negative* yang diklasifikasikan sebagai negatif.

Berdasarkan nilai TP, TN, FP dan FN maka akan diperoleh nilai akurasi, presisi dan recall dengan masing-masing menggunakan persamaan (2), (4) dan (3) [17].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{2}$$

Nilai akurasi ini menunjukkan seberapa akurat sistem bisa mengklasifikasikan data. Dengan kata lain, nilai yang akurat adalah perbandingan antara data yang diklasifikasikan benar dengan semua data.

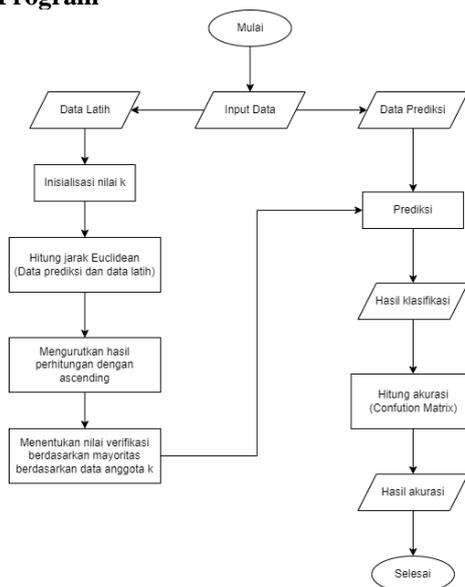
$$Presisi = \frac{TP}{FP+TP} \times 100\% \tag{3}$$

Nilai presisi ini menunjukkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasikan positif.

$$Recall = \frac{TP}{FN+TP} \times 100\% \tag{4}$$

Nilai *recall* ini menunjukkan berapa persen data kategori positif yang terklasifikasikan cocok oleh sistem.

2.2 Diagram Alur Pembuatan Program



Gambar 2. Flowchart

Pada Gambar 2 menjelaskan bahwa, sistem yang dirancang dimulai dengan pengguna akan memasukkan data dengan mengunggah *file* yang berisikan data latih ataupun data prediksi. Selanjutnya menginisialisasi nilai *k* (jumlah tetangga yang paling dekat), misalnya nilai *k* yang digunakan *k*=35. Setelah itu dilakukan perhitungan untuk menghitung dataset yang sudah dimasukkan sebagai data latih, kemudian dilakukan proses perhitungan untuk memperoleh nilai *Euclidean Distance*. Pada proses pengurutan, data akan diurutkan berdasarkan nilai *Euclidean Distance* yang terkecil, kemudian akan dimasukkan dalam kelompok anggota *k* sesuai dengan nilai yang sudah ditentukan. Jika nilai *k*=35 maka diambil 35 urutan paling kecil dari nilai *Euclidean Distance* yang didapatkan. Hasil dari data latih sebelumnya akan digunakan sebagai model dalam melakukan proses prediksi untuk mendapatkan hasil prediksi dari klasifikasi *K-Nearest Neighbor*. Proses untuk menghitung nilai akurasi, presisi dan *recall*nya dilakukan dengan menggunakan metode *confusion matrix*.

Dalam penelitian ini, diusulkan metode optimasi penyaluran Bantuan Langsung Tunai (BLT) menggunakan algoritma *K-Nearest Neighbor* (KNN). Proses klasifikasi data penerima BLT dilakukan dengan langkah-langkah sistematis untuk meningkatkan akurasi dalam pemilihan penerima yang benar-benar berhak menerima bantuan. [18] Proses dimulai dengan pengumpulan *data latih* yang berisi informasi historis penerima BLT yang telah diverifikasi sebelumnya. Data ini kemudian digunakan sebagai dasar untuk memprediksi penerima baru melalui *input data*. Setelah itu, data penerima potensial yang baru dimasukkan sebagai *data prediksi*.

Selanjutnya, sistem melakukan *inisialisasi* jumlah tetangga terdekat (nilai *k*). Untuk menentukan kemiripan antara calon penerima dengan data latih, sistem menghitung *jarak Euclidean* antara data prediksi dan data latih. Hasil perhitungan jarak ini kemudian disusun dalam urutan *ascending* untuk memudahkan pemilihan *k* tetangga terdekat. [19] Dari daftar tersebut, mayoritas kelas dari *k* tetangga yang paling dekat digunakan untuk menentukan klasifikasi calon penerima, sehingga memungkinkan verifikasi otomatis berdasarkan pola historis. Proses klasifikasi ini menghasilkan output yang kemudian diuji dengan perhitungan *Confusion Matrix* untuk mengevaluasi tingkat akurasi model KNN dalam menentukan penerima BLT secara optimal. Setelah perhitungan akurasi dilakukan, hasil akhirnya digunakan untuk menentukan penerima BLT yang benar-benar memenuhi kriteria. [20] Dengan pendekatan ini, diharapkan distribusi bantuan menjadi lebih tepat sasaran, mengurangi risiko salah sasaran, dan meningkatkan efisiensi dalam penyaluran BLT di Desa Penedagador.

Diagram Alur Proses Klasifikasi menggunakan K-Nearest Neighbor (KNN), Untuk menjelaskan proses klasifikasi dalam penelitian ini, digunakan metode *K-Nearest Neighbors* (KNN). Berikut adalah diagram alur dari proses klasifikasi menggunakan KNN:[21]

1. Preprocessing Data:

- a. **Pembersihan Data:** Data yang diperoleh dari berbagai sumber sering kali mengandung *missing values*, duplikasi, atau anomali. Oleh karena itu, proses pembersihan dilakukan dengan menghilangkan data yang tidak lengkap atau mengganti nilai yang hilang menggunakan metode interpolasi atau median.
- b. **Transformasi Data ke Bobot:** Fitur yang bersifat kategori dikonversi menjadi nilai numerik menggunakan metode *one-hot encoding* atau *label encoding* agar dapat digunakan dalam perhitungan jarak dalam KNN.
- c. **Standarisasi Data:** Karena KNN bergantung pada perhitungan jarak, data perlu dinormalisasi dengan metode Min-Max Scaling atau Z-score normalization agar setiap fitur memiliki skala yang seragam.

2. Pemilihan K dan Perhitungan Jarak:

- a) Parameter *K* menentukan jumlah tetangga terdekat yang digunakan untuk klasifikasi. Nilai *K* yang optimal dipilih berdasarkan eksperimen.
- b) Perhitungan jarak antara data uji dan data latih dilakukan menggunakan *Euclidean Distance*, *Manhattan Distance*, atau *Minkowski Distance* tergantung pada karakteristik dataset.

3. Klasifikasi dan Pengambilan Keputusan:

- a. Data uji diklasifikasikan berdasarkan mayoritas label dari *K* tetangga terdekat.

4. Evaluasi Model:

- a) Pengujian dilakukan menggunakan metode *k-fold cross-validation* untuk menghindari bias dalam penilaian akurasi.
- b) Metrik evaluasi yang digunakan mencakup *accuracy*, *precision*, *recall*, dan *F1-score*.

5. Pengujian dan Validasi:

Setelah model dievaluasi dan menunjukkan kinerja yang memuaskan, langkah selanjutnya adalah mengujinya pada data baru yang belum pernah dilihat oleh model. Hal ini untuk memastikan bahwa model memiliki generalisasi yang baik dan dapat berfungsi dengan baik pada data nyata.

Implementasi metode KNN dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan pustaka *Scikit-learn*, yang menyediakan fungsi-fungsi untuk pemrosesan data, pembelajaran mesin, serta validasi model. Alternatif lain yang dapat digunakan adalah MATLAB atau R, yang juga memiliki pustaka khusus untuk analisis data dan implementasi algoritma klasifikasi. [22] Dalam konteks penyaluran Bantuan Langsung Tunai (BLT) di Desa Penedagador, algoritma KNN dapat digunakan untuk mengklasifikasikan warga yang berhak menerima bantuan berdasarkan kriteria tertentu. Prosesnya melibatkan pengumpulan data demografis dan ekonomi warga, prapemrosesan data untuk memastikan kualitas dan konsistensi, serta implementasi model KNN untuk menentukan penerima bantuan. Evaluasi model dilakukan untuk memastikan akurasi dan efektivitas penyaluran bantuan.

3. HASIL DAN PEMBAHASAN

Implementasi algoritma K-Nearest Neighbor untuk mengklasifikasikan penerima bantuan dan desa di Desa Penedagador, akan dibahas tentang langkah-langkah atau proses yang terjadi pada sistem sehingga menghasilkan luaran Penerima bantuan dana desa yang layak dan tidak layak.

3.1 Implementasi Antarmuka

- a. Antarmuka Halaman Awal

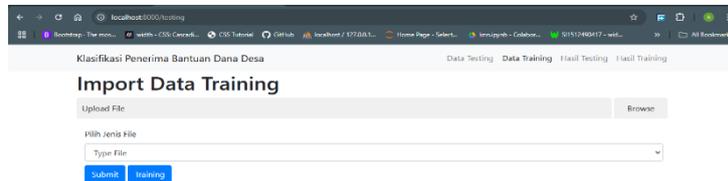
Tampilan awal atau home pada system yang dibangun dapat dilihat pada Gambar 3:



Gambar 3. Halaman Awal

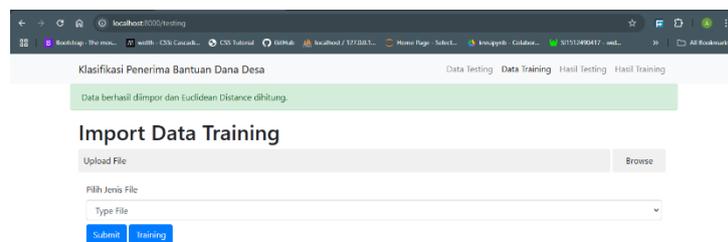
Gambar di atas merupakan tampilan awal pada system klasifikasi penerima bantuan dana desa Penedagandor. Pada halaman ini terdapat tombol start yang akan mengarahkan pengguna ke halaman selanjutnya yaitu halaman training dan halaman testing.

b. Antarmuka Halaman Training



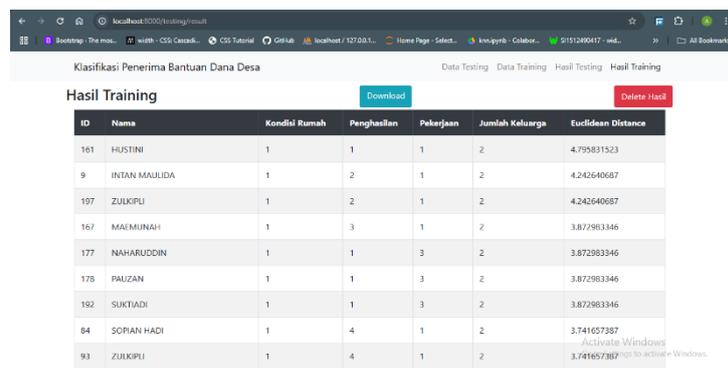
Gambar 4. Halaman Import Data

Gambar 4 merupakan tampilan halaman training yang menampilkan kolom untuk mengunggah data dari file ke database. Kolom pada halaman ini digunakan untuk memilih file serta format file yang akan diunggah. Tombol import digunakan untuk memulai proses import data dari file ke database. Tombol train digunakan untuk memulai proses pelatihan algoritma.



Gambar 5. Data Berhasil Diimport dan siap detraining

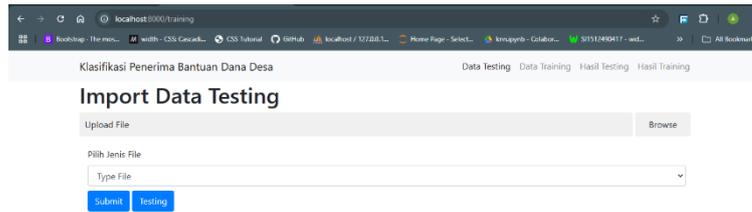
Gambar 5 menampilkan informasi bahwa proses import data dari file ke database selesai dilakukan.



Gambar 6. Halaman Hasil Training

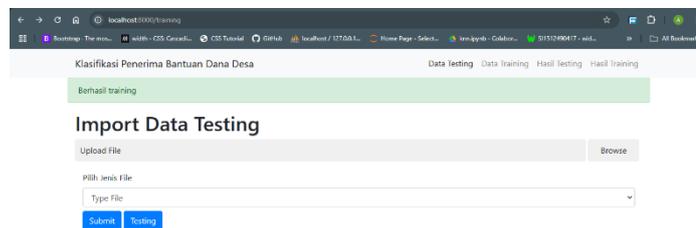
Gambar 6 merupakan halaman hasil training yang menampilkan hasil training dari data yang sudah di import sebelumnya. Halaman hasil training akan menampilkan informasi jarak Euclidean yang sudah dihitung dan diurutkan secara ascending. Pada halaman ini juga terdapat 2 tombol yaitu tombol download yang digunakan untuk mengunduh data training yang didapatkan dan tombol delete hasil untuk menghapus hasil training yang tersimpan pada database.

c. Antarmuka Halaman Testing



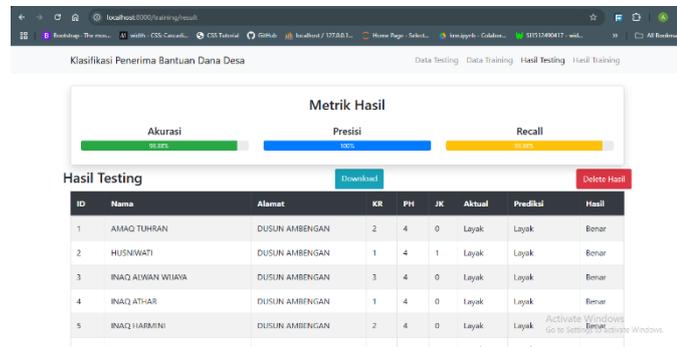
Gambar 7. Halaman Testing

Gambar 7 merupakan tampilan halaman testing yang menampilkan kolom untuk mengunggah data dari *file* ke *database*. Kolom pada halaman ini digunakan untuk memilih *file* serta *format file* yang akan diunggah. Tombol *import* digunakan untuk memulai proses *import* data dari *file* ke *database*. Tombol *testing* digunakan untuk memulai proses pengujian algoritma.



Gambar 8. Data Berhasil Diimport dan Siap ditesting

Gambar 8 menampilkan informasi bahwa proses import data dari file ke *database* selesai dilakukan.



Gambar 9. Halaman Hasil Testing

Gambar 10 merupakan halaman hasil testing yang menampilkan hasil testing dari data yang sudah diimport sebelumnya. Halaman hasil testing akan menampilkan informasi berupa hasil akurasi, presisi dan recall. Pada halaman ini juga menampilkan hasil dari pengujian data yang sudah dilakukan yaitu nilai aktual, prediksi dan hasil. Nilai aktual merupakan nilai yang sebenarnya dari data, sedangkan nilai prediksi merupakan nilai prediksi yang didapatkan dari perhitungan menggunakan algoritma K-Nearest Neighbor, dan nilai hasil merupakan hasil perbandingan dari nilai aktual dan prediksi yang didapatkan. Pada halaman testing ini juga terdapat 2 tombol yaitu tombol download yang digunakan untuk mengunduh data testing yang didapatkan dan tombol delete hasil untuk menghapus hasil testing yang tersimpan pada database.

3.2 Implementasi Algoritma K-Nearest Neighbor

a. Eksplorasi Dataset dan Preprocessing

Eksplorasi dataset merupakan tahap awal dalam penelitian ini, yang bertujuan untuk memahami struktur dan karakteristik data penerima Bantuan Langsung Tunai (BLT) di Desa Penedagador. Dataset yang digunakan mencakup variabel seperti kondisi rumah, penghasilan, pekerjaan, dan jumlah anggota keluarga. Data ini kemudian diproses melalui tahap preprocessing, yang meliputi penanganan data yang hilang, normalisasi, dan transformasi data agar sesuai dengan kebutuhan algoritma K-Nearest Neighbor (KNN). Setelah

preprocessing, data dibagi menjadi dua bagian utama, yaitu data latih (training set) dan data uji (testing set). Pembagian ini dilakukan dengan rasio 80:20, di mana 80% data digunakan untuk melatih model dan 20% digunakan untuk menguji akurasi model.

b. Inisialisasi nilai K

Pada penginisialisasi nilai k ini menggunakan k=35, k=45, dan k=55.

c. Perhitungan Jarak Euclidean

Pada perhitungan jarak Euclidean ini menggunakan masing masing k yang sudah ditentukan. Contoh perhitungan manual jarak Euclidean antara data latih penerima BLT 2020 (data 1 dan data 2) dan data uji penerima BLT 2022 (data 1).

$$\begin{aligned}
 j(1,1) &= \sqrt{(1-2)^2 + (2-4)^2 + (3-4)^2 + (2-0)^2} \\
 &= \sqrt{1 + 4 + 1 + 4} \\
 &= \sqrt{10} \\
 &= 3.162
 \end{aligned}$$

$$\begin{aligned}
 j(2,1) &= \sqrt{(1-1)^2 + (2-4)^2 + (3-4)^2 + (2-1)^2} \\
 &= \sqrt{0 + 4 + 1 + 1} \\
 &= \sqrt{6} \\
 &= 2.450
 \end{aligned}$$

Jarak Euclidean antara data uji dengan data kedua dan seterusnya dihitung menggunakan cara yang sama. Hasil perhitungan jarak Euclidean bias dilihat pada table berikut:

Tabel 5. Hasil Perhitungan Euclidean Distance

Id	Kondisi rumah	penghasilan	pekerjaan	Jumlah keluarga	Euclidean distance
1	1	2	3	2	3,162
2	1	2	3	2	2,449
3	3	2	2	2	3,162
4	2	4	4	0	2,645
5	1	3	3	2	2,236
.
.
198					(Data baru)

Tabel di atas menyajikan data mengenai beberapa karakteristik rumah tangga dan jarak Euclidean yang dihitung berdasarkan variabel tertentu. Berikut adalah penjelasan dari setiap kolom dalam tabel:[23]

- 1) **Id** – Merupakan nomor identifikasi unik untuk setiap rumah tangga dalam dataset.
- 2) **Kondisi rumah** – Menunjukkan kondisi tempat tinggal rumah tangga, yang dinyatakan dalam bentuk kategori numerik. Semakin tinggi angka, kemungkinan besar kondisi rumah semakin baik atau memiliki fasilitas yang lebih memadai.
- 3) **Penghasilan** – Mewakili tingkat pendapatan rumah tangga, yang juga dikategorikan dalam skala numerik. Nilai yang lebih tinggi dapat menunjukkan penghasilan yang lebih besar.
- 4) **Pekerjaan** – Mengacu pada jenis atau status pekerjaan kepala rumah tangga, yang dinyatakan dalam bentuk angka. Kategori ini dapat mencerminkan tingkat pekerjaan yang lebih stabil atau lebih menguntungkan.
- 5) **Jumlah keluarga** – Menunjukkan jumlah anggota keluarga dalam rumah tangga yang bersangkutan.
- 6) **Euclidean Distance** – Merupakan hasil perhitungan jarak Euclidean berdasarkan beberapa variabel dalam dataset. Nilai ini digunakan untuk mengukur kedekatan atau kemiripan suatu rumah tangga dengan rumah tangga lainnya berdasarkan atribut yang telah ditentukan.

Tabel ini dapat digunakan untuk analisis lebih lanjut dalam penelitian sosial, ekonomi, atau kebijakan publik, seperti mengelompokkan rumah tangga berdasarkan kondisi ekonomi dan sosial atau menentukan pola kesejahteraan masyarakat berdasarkan faktor-faktor yang diberikan. Dalam tabel ini juga, Skala 0 hingga 4 diterapkan pada setiap variabel, termasuk kondisi rumah, pendapatan, pekerjaan, dan jumlah keluarga . Interpretasi nilai dalam skala ini adalah sebagai berikut:

- Nilai yang lebih kecil (mendekati 0) menunjukkan bahwa kondisi rumah tangga lebih baik, dengan fasilitas yang memadai, penghasilan yang lebih tinggi, pekerjaan yang lebih stabil, dan jumlah keluarga yang lebih sedikit (yang dalam beberapa kasus dapat mencerminkan beban ekonomi yang lebih ringan).

- Nilai yang lebih besar (mendekati 4) menunjukkan kondisi rumah tangga yang kurang berkecukupan, dengan fasilitas terbatas, penghasilan lebih rendah, pekerjaan yang kurang stabil, dan jumlah keluarga yang lebih banyak (yang bisa menjadi indikator beban ekonomi lebih tinggi).

Dalam konteks Euclidean Distance , rumah tangga dengan nilai jarak yang lebih kecil memiliki karakteristik yang lebih dekat atau mirip dengan rumah tangga lainnya dalam dataset, sedangkan rumah tangga dengan nilai jarak lebih besar menunjukkan perbedaan karakteristik yang lebih signifikan. Analisis ini berguna untuk memahami pola kesejahteraan rumah tangga dan dapat membantu dalam penyusunan kebijakan sosial serta ekonomi, misalnya dalam penentuan program bantuan bagi kelompok yang lebih membutuhkan.

d. Pengurutan hasil perhitungan (*Ascending*)

Setelah diketahui jarak Euclidean antara setiap data latih dan data uji, maka akan dilakukan pengurutan jarak euclidean dari yang terkecil ke yang terbesar (*ascending*). Semakin kecil jarak euclidean maka jarak antar datanya juga semakin dekat. Hasil pengurutannya akan diberikan peringkat sesuai dengan nilai k yang diberikan. Lalu menentukan kategori atau label dari data yang telah diperingkatkan tersebut. Berikut merupakan contoh hasil pengurutan perhitungan secara *ascending* dengan menggunakan k=35:

Tabel 6. Hasil perangkingan Ascending dari k=35

Id	KR	PH	PK	JK	Euclidean Distance	Ranking
53	2	4	4	0	0	1
62	2	4	4	0	0	2
.
.
.
65	3	4	4	1	1,414	33
66	1	4	4	1	1,414	34
70	3	4	3	0	1,414	35

Tabel 6 menyajikan hasil perangkingan ascending untuk k=35, yang berarti algoritma mempertimbangkan 35 tetangga terdekat untuk menentukan klasifikasi suatu data. Setiap baris dalam tabel tersebut mencantumkan beberapa atribut, antara lain:[24]

- **Id (Identifikasi Unik):** Kode atau nomor unik yang diberikan untuk setiap data guna membedakan satu entitas dari yang lain.
- **KR (Kondisi Rumah):** Menggambarkan keadaan fisik dan kualitas hunian tempat tinggal seseorang atau keluarga, termasuk aspek struktural dan fasilitas pendukung.
- **PH (Penghasilan):** Jumlah pendapatan yang diperoleh individu atau keluarga dalam periode tertentu, yang dapat berasal dari gaji, usaha, atau sumber keuangan lainnya.
- **PK (Pekerjaan):** Jenis profesi atau bidang usaha yang dijalankan oleh individu sebagai sumber mata pencaharian utama.
- **JK (Jumlah Keluarga):** Total anggota keluarga yang tinggal dalam satu rumah tangga, mencakup kepala keluarga, pasangan, anak, serta anggota lainnya yang bergantung secara ekonomi.
- **Euclidean Distance:** Jarak Euclidean antara data yang dievaluasi dengan data lainnya dalam dataset.
- **Ranking:** Peringkat berdasarkan jarak Euclidean, diurutkan secara ascending.

Proses klasifikasi dengan KNN dimulai dengan menghitung jarak Euclidean antara data yang akan diklasifikasikan dengan seluruh data dalam dataset. Setelah itu, jarak-jarak tersebut diurutkan dari yang terkecil hingga terbesar. Kemudian, diambil sejumlah k data dengan jarak terdekat (dalam hal ini k=35). Kelas yang paling sering muncul di antara k tetangga terdekat tersebut akan menjadi prediksi kelas untuk data yang dievaluasi. [25]

Penggunaan k=35 menunjukkan bahwa algoritma mempertimbangkan 35 tetangga terdekat untuk menentukan klasifikasi. Pemilihan nilai k yang tepat sangat penting karena dapat mempengaruhi akurasi model. Nilai k yang terlalu kecil mungkin membuat model sensitif terhadap noise, sementara nilai k yang terlalu besar dapat menyebabkan model mengabaikan pola penting dalam data.

e. Hasil Prediksi

1. Prediksi Hasil Manual

Contoh hasil prediksi menggunakan k=35 pada Tabel 7:

Tabel 7. Hasil prediksi manual data uji menggunakan k=35

Id	Nama	Alamat	KR	PH	PK	JK	Klasifikasi		Hasil
							Aktual	Prediksi	
1	Amaq Tuhran	Dusun Ambengan	2	4	4	0	Layak	Layak	Benar
2	Husniwati	Dusun Ambengan	1	4	4	1	Layak	Layak	Benar
.
.
134	Siti Zulaiha	Dusun Repok	1	4	4	1	Layak	Layak	Benar
135	Sopiah	Dusun Repok	3	4	4	1	Layak	Layak	Benar
136	Sulhiah	Dusun Repok	1	4	4	0	Layak	Layak	Benar

Tabel 7 menyajikan hasil prediksi manual terhadap data uji menggunakan algoritma K-Nearest Neighbor (KNN) dengan parameter k=35. Setiap entri dalam tabel mencakup informasi seperti Id, Nama, Alamat, serta atribut KR, PH, PK, dan JK. Kolom Klasifikasi menunjukkan status aktual (Layak/Tidak Layak), sementara kolom Hasil mencantumkan prediksi algoritma dan apakah prediksi tersebut benar atau salah.

Algoritma KNN bekerja dengan mengklasifikasikan data uji berdasarkan mayoritas kelas dari k tetangga terdekat dalam data pelatihan. Dalam konteks ini, k=35 berarti algoritma mempertimbangkan 35 tetangga terdekat untuk menentukan klasifikasi. [26] Pemilihan nilai k yang tepat sangat penting, karena nilai k yang terlalu kecil dapat membuat model sensitif terhadap noise, sementara nilai k yang terlalu besar dapat mengaburkan batas antara kelas yang berbeda.

Proses prediksi manual ini melibatkan perhitungan jarak antara data uji dan seluruh data pelatihan menggunakan metrik tertentu, seperti Euclidean Distance. Setelah jarak dihitung, 35 tetangga terdekat diidentifikasi, dan klasifikasi ditentukan berdasarkan mayoritas kelas dari tetangga-tetangga tersebut. Hasil prediksi kemudian dibandingkan dengan klasifikasi aktual untuk menilai akurasi model.

2. Prediksi Dari Website K-Nearest Neighbor

Contoh hasil prediksi menggunakan k=35 pada Tabel 8 di bawah ini:

Tabel 8. Hasil prediksi website data uji menggunakan k=35

Id	Nama	Alamat	KR	PH	PK	JK	Klasifikasi		Hasil
							Aktual	Prediksi	
1	Amaq Tuhran	Dusun Ambengan	2	4	4	0	Layak	Layak	Benar
2	Husniwati	Dusun Ambengan	1	4	4	1	Layak	Layak	Benar
.
.
134	Siti Zulaiha	Dusun Repok	1	4	4	1	Layak	Layak	Benar
135	Sopiah	Dusun Repok	3	4	4	1	Layak	Layak	Benar
136	Sulhiah	Dusun Repok	1	4	4	0	Layak	Layak	Benar

Tabel 8 menyajikan hasil prediksi klasifikasi kelayakan berdasarkan algoritma K-Nearest Neighbor (K-NN) dengan nilai k=35. Algoritma K-NN merupakan metode klasifikasi yang menentukan kelas suatu data berdasarkan mayoritas kelas dari k tetangga terdekatnya. Dalam konteks ini, setiap individu dievaluasi berdasarkan empat atribut: KR, PH, PK, dan JK. [27] Pemilihan nilai k=35 berarti bahwa untuk menentukan klasifikasi setiap individu, algoritma mempertimbangkan 35 tetangga terdekat dalam ruang fitur. Pemilihan nilai k yang tepat sangat penting karena nilai k yang terlalu kecil dapat membuat model sensitif terhadap noise, sementara nilai k yang terlalu besar dapat mengaburkan batas antara kelas yang berbeda. Dalam studi ini, k=35 dipilih untuk mencapai keseimbangan antara bias dan varians, yang diharapkan dapat meningkatkan akurasi prediksi.

Evaluasi performa model dilakukan dengan membandingkan hasil prediksi dengan klasifikasi aktual. Kolom "Hasil" menunjukkan apakah prediksi tersebut benar atau tidak. Akurasi model dapat dihitung dengan

membagi jumlah prediksi benar dengan total jumlah prediksi. Selain akurasi, metrik evaluasi lain seperti precision, recall, dan F1-score juga penting untuk memberikan gambaran yang lebih komprehensif tentang kinerja model, terutama jika terdapat ketidakseimbangan kelas dalam data. [28]

Penerapan algoritma K-NN dalam berbagai domain telah menunjukkan hasil yang bervariasi. Misalnya, dalam klasifikasi penyakit jantung, algoritma K-NN dengan cross-validation telah digunakan untuk mengklasifikasikan data pasien dengan tingkat akurasi yang signifikan. Selain itu, pemilihan metode normalisasi data yang tepat, seperti Min-Max atau Z-Score, dapat mempengaruhi kinerja algoritma K-NN secara keseluruhan. [29]

Dalam konteks tabel ini, analisis lebih lanjut dapat dilakukan untuk memahami pengaruh masing-masing atribut terhadap keputusan klasifikasi. Hal ini dapat dilakukan melalui teknik seperti analisis korelasi atau pengujian hipotesis statistik untuk menentukan signifikansi kontribusi setiap atribut terhadap hasil klasifikasi. Secara keseluruhan, penggunaan algoritma K-NN dengan k=35 dalam studi ini menunjukkan pendekatan yang sistematis dalam klasifikasi kelayakan individu berdasarkan atribut yang telah ditentukan. Evaluasi kinerja model dan analisis lebih lanjut terhadap atribut dapat memberikan wawasan yang lebih mendalam dan membantu dalam pengambilan keputusan yang lebih akurat.

f. Evaluasi Hasil Klasifikasi

Evaluasi Performa Hasil Pengujian Algoritma K-Nearest Neighbor Menggunakan Confusion Matrix

Berikut adalah contoh hasil pengujian algoritma K-Nearest Neighbor menggunakan confusion matrix yang didapatkan dapat dilihat pada Tabel 9.

Tabel 9. Hasil Confussion Matrix k=35

		Kelas Prediksi	
		Layak	Tidak Layak
Kelas Aktual	Layak	127	0
	Tidak Layak	8	1
Akurasi		95%	
Presisi		95%	
Recall		100%	

Dari semua nilai k yang sudah ditentukan, nilai k terbaik yaitu k=35 dengan data latih 2021 dengan akurasi 95%, presisi 95% dan recall sebesar 100%. Selanjutnya, yaitu membandingkan beberapa tingkat akurasi, presisi, dan recall dari algoritma *K-Nearest Neighbor* dengan nilai k=35, k=45, dan k=55. Dengan data latih tahun 2021, akurasi tertinggi dicapai pada k=35, yaitu 95%. Hasil ini menunjukkan bahwa algoritma KNN mampu mengklasifikasikan penerima BLT dengan tingkat ketepatan yang tinggi, sehingga dapat dijadikan alat bantu dalam pengambilan keputusan penyaluran bantuan. Setelah itu, hasil dari perbandingan tersebut akan dicatat dan dibandingkan satu sama lain untuk menetapkan nilai k yang paling optimal dengan tingkat persentase lebih tinggi. Berikut hasil dari presentase akurasi, presisi dan recall dengan nilai k yang berbeda dapat dilihat pada tabel 10

Tabel 10. Kategori data baru dan tingkat akurasi, presisi dan recall dari berbagai nilai k menggunakan data latih tahun 2020

k	35	45	55
Kategori data baru	Layak	Layak	Layak
Akurasi	94%	93%	93%
Presisi	94%	93%	93%
recall	100%	100%	100%

Tabel 10 menunjukkan hasil prediksi kategori data baru sebagai "Layak" dengan nilai k yang berbeda (35, 45, dan 55). Akurasi dan presisi tertinggi dicapai pada k=35 dengan nilai 94%, sementara recall mencapai 100% untuk semua nilai k yang diuji.

Tabel 11 menampilkan evaluasi serupa, namun tanpa mencantumkan kategori data baru. Akurasi dan presisi tertinggi tercatat pada k=35 dengan nilai 95%, sedangkan recall mencapai 100% untuk k=35 dan 45, namun sedikit menurun menjadi 99% pada k=55. Pemilihan nilai k yang optimal sangat penting dalam algoritma KNN, karena nilai k yang terlalu kecil dapat membuat model sensitif terhadap noise, sementara

nilai k yang terlalu besar dapat mengaburkan batas antara kelas yang berbeda. Studi oleh Cholil et al. (2021) menunjukkan bahwa variasi nilai k mempengaruhi akurasi dan presisi model KNN dalam klasifikasi data penerima beasiswa. [30]

Tabel 11. Kategori data baru dan tingkat akurasi, presisi dan recall dari berbagai nilai k menggunakan data latih tahun 2020

k	35	45	55
Kategori data baru			
Akurasi	95%	93%	93%
Presisi	95%	93%	93%
recall	100%	100%	99%

Evaluasi kinerja model klasifikasi umumnya dilakukan menggunakan metrik seperti akurasi, presisi, dan recall. Akurasi mengukur persentase prediksi yang benar dari keseluruhan prediksi, presisi mengindikasikan proporsi prediksi positif yang benar, sementara recall menunjukkan kemampuan model dalam mendeteksi semua instance positif. Penggunaan confusion matrix seringkali menjadi dasar dalam perhitungan metrik-metrik ini. Dalam konteks algoritma KNN, pemilihan nilai k yang optimal dan evaluasi kinerja model menggunakan metrik yang tepat sangat krusial untuk memastikan model dapat mengklasifikasikan data dengan akurat dan andal. [31]

Penelitian ini dibandingkan dengan metode lain yang telah digunakan dalam klasifikasi penerima BLT, seperti Decision Tree dan Naive Bayes. Berdasarkan penelitian sebelumnya, Decision Tree memiliki akurasi rata-rata 92%, sedangkan Naive Bayes 89%, hal ini menunjukkan bahwa KNN dengan $k=35$ memiliki keunggulan dalam akurasi dibandingkan metode lainnya. Dalam konteks klasifikasi penerima bantuan sosial, beberapa penelitian telah membandingkan kinerja KNN dengan metode lain. Misalnya, penelitian oleh Hamonangan dan Rifai (2023) menunjukkan bahwa KNN memiliki akurasi yang lebih tinggi dibandingkan dengan Naive Bayes dalam menentukan penerima BLT. [32] Namun, penelitian lain oleh Pahrudin dan Harianto (2022) mengindikasikan bahwa algoritma C4.5 dapat menghasilkan model dengan kompleksitas lebih rendah tanpa mengorbankan akurasi. [33]

Keunggulan utama KNN adalah kemampuannya dalam menangani data non-linear dan tidak memerlukan asumsi distribusi data tertentu. Namun, KNN juga memiliki keterbatasan, terutama dalam hal waktu komputasi yang tinggi ketika dataset menjadi lebih besar, terdapat sensitif terhadap noise KNN dapat terpengaruh oleh data yang noisy atau outlier, yang dapat mengurangi akurasi model, serta memori intensif dikarenakan KNN menyimpan semua data pelatihan, kebutuhan memori dapat menjadi masalah seiring dengan bertambahnya ukuran dataset. [34] Sebaliknya, KNN pun mempunyai beragam kelemahan contohnya kebutuhan untuk menentukan nilai k yang optimal, biaya komputasi yang tinggi ketika data latih berukuran besar, serta sensitivitas terhadap fitur yang tidak relevan atau adanya *noise*. Salah satu kelemahan utama KNN adalah ketergantungannya pada pemilihan nilai k yang sesuai serta tingginya biaya komputasi saat proses klasifikasi, karena seluruh perhitungan dilakukan pada tahap klasifikasi, bukan saat pelatihan data (*lazy learning*). [35]

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *K-Nearest Neighbor* (K-NN) dalam sistem klasifikasi penerima bantuan dana desa di Desa Penedagandor menunjukkan hasil yang optimal dengan pemilihan parameter k yang tepat. Dari berbagai pengujian yang dilakukan, nilai k yang menghasilkan performa terbaik dalam hal akurasi, presisi, dan *recall* adalah $k = 35$. Model dengan parameter ini mencapai tingkat akurasi sebesar 95%, presisi tertinggi sebesar 95%, serta *recall* maksimal sebesar 100%. Dengan hasil tersebut, sistem klasifikasi yang dikembangkan dapat menjadi alat bantu yang andal dalam pengambilan keputusan terkait penyaluran bantuan dana desa. Penelitian ini memberikan kontribusi signifikan dalam optimalisasi penyaluran Bantuan Langsung Tunai (BLT) dengan memanfaatkan metode pembelajaran mesin berbasis *K-Nearest Neighbor*. Implementasi metode ini menunjukkan bahwa pemanfaatan teknologi dalam sistem pemerintahan, khususnya dalam penyaluran bantuan sosial, dapat meningkatkan efisiensi dan ketepatan sasaran. Jika diterapkan secara lebih luas dalam sistem pemerintahan, model ini berpotensi membantu pemerintah dalam menyalurkan bantuan secara lebih objektif dan transparan, mengurangi kesalahan dalam pendistribusian, serta mengoptimalkan pemanfaatan anggaran negara.

Dibandingkan dengan metode lain, algoritma K-NN memiliki keunggulan dalam kesederhanaan perhitungan dan fleksibilitas dalam menangani berbagai jenis data tanpa memerlukan asumsi distribusi

tertentu. Namun, metode ini juga memiliki keterbatasan dalam hal efisiensi komputasi ketika berhadapan dengan dataset yang sangat besar, sehingga mungkin memerlukan optimasi lebih lanjut. Sebagai saran untuk penelitian masa depan, perlu dilakukan pengujian dengan dataset yang lebih besar dan beragam agar model dapat menangani berbagai skenario yang lebih kompleks. Selain itu, perbandingan dengan metode klasifikasi lain seperti **Random Forest**, **Support Vector Machine (SVM)**, atau bahkan **Deep Learning** dapat dilakukan untuk mengevaluasi apakah ada metode yang lebih efektif dalam meningkatkan akurasi dan efisiensi klasifikasi penerima bantuan. Dengan demikian, penelitian ini dapat terus berkembang untuk mendukung sistem penyaluran bantuan yang lebih optimal dan berbasis data.

DAFTAR PUSTAKA

- [1] B. Menyhert, *Absolute Poverty Measurement with Minimum Food Needs: A New Inverse Method for Advanced Economies*, vol. 174, no. 1. Springer Netherlands, 2024. doi: 10.1007/s11205-024-03368-x.
- [2] M. Pomati, S. Nandy, S. Jose, and B. Reddy, "Multidimensional Adult and Child Poverty in India—Establishing Consensus About Socially Perceived Necessities for a New Measure of Poverty," *Indian J. Hum. Dev.*, 2024, doi: 10.1177/09737030241274084.
- [3] A. H. Ginting, I. Widianingsih, R. Mulyawan, and H. Nurasa, "Village Government's Risk Management and Village Fund Administration in Indonesia," *Sustain.*, vol. 15, no. 24, pp. 1–17, 2023, doi: 10.3390/su152416706.
- [4] Y. Jamaluddin, "Policy Actors as Determining Factors of Sustainability Village Fund Management in Indonesia," *J. Ecohumanism*, vol. 3, no. 4, pp. 689–694, 2024, doi: 10.62754/joe.v3i4.3524.
- [5] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes," *Respati*, vol. 16, no. 1, 2021, doi: 10.35842/jtir.v16i1.386.
- [6] A. Junaidi, Y. Yunita, S. Agustyani, C. I. Agustyaningrum, and Y. T. Arifin, "Klasifikasi Penerima Bantuan Sosial Menggunakan Algoritma C 4.5," *J. Tek. Komput.*, vol. 9, no. 1, 2023, doi: 10.31294/jtk.v9i1.14378.
- [7] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, "Penerapan Data Mining Untuk Analisis Stunting Pada Balita," vol. 13, no. November, pp. 1717–1728, 2024.
- [8] M. Sholikhul Fiqri and H. Dwi Bhakti, "Klasifikasi Potensi Penyakit Diabetes Mellitus Tipe Ii Pada Pasien Menggunakan Algoritme Knn (K-Nearest Neighbor)," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 4, pp. 7305–7313, 2024, doi: 10.36040/jati.v8i4.10133.
- [9] W. W. Hu, L. Sun, and L. Li, "An Exploration of the Computer Big Data Mining Service Model Under Resource Sharing," *Inf. Resour. Manag. J.*, vol. 37, no. 1, pp. 1–17, 2024, doi: 10.4018/IRMJ.340032.
- [10] A. A. D. Halim and S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, pp. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.
- [11] V. Kumar, N. Kedam, K. V. Sharma, D. J. Mehta, and T. Caloiero, "Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models," *Water (Switzerland)*, vol. 15, no. 14, 2023, doi: 10.3390/w15142572.
- [12] M. Windarti and A. Suradi, "Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa," *Telematika*, vol. 12, no. 1, 2019, doi: 10.35671/telematika.v12i1.778.
- [13] P. Astuti and N. Nuris, "Penerapan Algoritma KNN Pada Analisis Sentimen Review Aplikasi Peduli Lindungi," *Comput. Sci.*, vol. 2, no. 2, pp. 137–142, 2022, doi: 10.31294/coscience.v2i2.1258.
- [14] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *J. Inform. MULTI*, vol. 1, no. 3, pp. 203–214, 2023.
- [15] S. Diansyah, "Klasifikasi Tingkat Kepuasan Pengguna dengan Menggunakan Metode K-Nearest Neighbour (KNN)," *J. Sistim Inf. dan Teknol.*, 2022, doi: 10.37034/jsisfotek.v4i1.114.
- [16] M. Bramer, *Principles of Data Mining*.
- [17] A. A. D. Halim and S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, 2021, doi: 10.33096/ijodas.v2i1.23.
- [18] Anwar Pauji, S. Aisyah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai," *KOPERTIP J. Ilm. Manaj. Inform. dan Komput.*, vol. 4, no. 1, 2022, doi: 10.32485/kopertip.v4i1.114.
- [19] H. Franco-Lopez, A. R. Ek, and M. E. Bauer, "Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method," *Remote Sens. Environ.*, vol. 77, no. 3, 2001, doi: 10.1016/S0034-4257(01)00209-7.
- [20] A. A. , I. , Tedy Rismawan, "PROTOTYPE KUNCI OTOMATIS PADA PINTU BERDASARKAN

- SUARA PENGGUNA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR),” *Coding J. Comput. dan Apl.*, vol. 4, no. 1, 2016, doi: 10.26418/coding.v4i1.13453.
- [21] M. Manyol, S. Eke, A. J. M. Massoma, A. Biboum, and R. Mouangue, “Preprocessing Approach for Power Transformer Maintenance Data Mining Based on k -Nearest Neighbor Completion and Principal Component Analysis,” *Int. Trans. Electr. Energy Syst.*, vol. 2022, 2022, doi: 10.1155/2022/8546588.
- [22] D. A. Manalu and G. Gunadi, “IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI,” *Infotech J. Technol. Inf.*, vol. 8, no. 1, 2022, doi: 10.37365/jti.v8i1.131.
- [23] M. D. Nurwanda, A. N. Liviasari, and D. D. Pambudi, “Jurnal JPILKOM (Jurnal Penelitian Ilmu Komputer) Penerapan K-Means untuk Meningkatkan Strategi Pemasaran Melalui Segmentasi Rumah Tangga yang Efektif,” vol. 2, no. 128, pp. 128–136, 2024.
- [24] A. Darmawan, “Perancangan Pengukuran Risiko Operasional pada Perusahaan Pembiayaan dengan Metode Risk Breakdown Structure (RBS) dan Analytic Network Process (ANP),” pp. 1–145, 2011.
- [25] D. Rika Widianita, “KLASIFIKASI PENYAKIT KARDIOVASKULAR DENGAN ALGORITMA K-NEAREST NEIGHBOR,” *AT-TAWASSUTH J. Ekon. Islam*, vol. VIII, no. I, pp. 1–19, 2023.
- [26] 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.
- [27] P. T. Prasetyaningrum, P. Purwanto, and A. F. Rochim, *GameBank : Model Keterlibatan Pelanggan di Mobile Banking*, no. January. 2025.
- [28] N. Hidayah and Dodiman, “Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluasi Mahasiswa Terhadap Dosen Teknik Informatika Universitas Dayanu Ikhsanuddin,” *J. Akad. Pendidik. Mat.*, vol. 10, no. 1, pp. 8–15, 2024.
- [29] R. Fadilah *et al.*, “Pengaruh metode pengukuran jarak dan smote pada klasifikasi penilaian kredit,” vol. 7, no. 2, pp. 193–202, 2024.
- [30] 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, 2021, doi: 10.31294/ijcit.v6i2.10438.
- [31] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [32] 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.
- [33] P. Pahrudin and K. Harianto, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Warga Penerima Bantuan Sosial,” *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2276.
- [34] F. A. Kusuma, “Pemodelan Klasifikasi Anemia Aplastik Menggunakan Teknik Oversampling Dan K-Nearest Neighbors,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4326.
- [35] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN model-based approach in classification,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2888, 2003, doi: 10.1007/978-3-540-39964-3_62.