

KLASIFIKASI PENERIMA BANTUAN LANGSUNG TUNAI DANA DESA MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

Annisa Rahma Putri, Asep Jamaludin, Garno

Informatika, Universitas Singaperbangsa Karawang

Jl. H.S. Ronggowaluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

2010631170055@student.unsika.ac.id

ABSTRAK

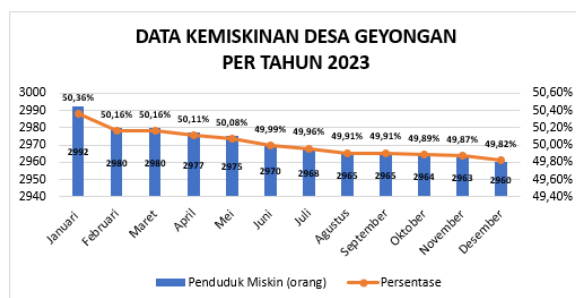
Desa Geyongan memiliki tingkat kemiskinan yang tinggi, mencapai 49,82% per 2023. Untuk mengatasi masalah ini, pemerintah mengeluarkan program Bantuan Langsung Tunai Dana Desa (BLT-DD). Namun, ada kesalahan dalam pembagian yang dianggap tidak adil karena data kelayakan penerima tidak sesuai. Oleh karena itu, penelitian ini dilakukan untuk mengklasifikasi penerima BLT menggunakan metode *K-Nearest Neighbor* dengan tujuan memastikan bantuan tepat sasaran. Pendekatan KDD (*Knowledge Discovery in Database*) digunakan untuk memodelkan algoritma *K-Nearest Neighbor* dengan bantuan aplikasi RapidMiner. Penelitian ini menggunakan empat skenario pembagian data latih dan uji (90:10, 80:20, 70:30, dan 60:40) dengan variasi nilai K (1, 3, 5, 7, 9, 11, 13, 15, 17, dan 19) untuk mencapai akurasi terbaik. Hasilnya menunjukkan bahwa skenario pembagian data 80:20 dengan nilai K=9 memberikan *accuracy* tertinggi, yaitu 87,31%, dengan *precision* 89,06%, *recall* 85,07%, dan *F-Measure* 87,02%.

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

1. PENDAHULUAN

Kemiskinan merupakan permasalahan mendasar yang dihadapi oleh Indonesia. Istilah kemiskinan muncul ketika individu atau sekelompok orang tidak mampu mencapai tingkat kesejahteraan ekonomi yang merupakan kebutuhan minimum untuk hidup layak, seperti pangan, tempat tinggal, pendidikan, pakaian, dan pekerjaan [1].

Desa Geyongan merupakan salah satu desa yang terletak di Kecamatan Arjawinangun, Kabupaten Cirebon, Jawa Barat. Berdasarkan data yang diperoleh dari pemerintahan Desa Geyongan pada tahun 2023, jumlah penduduk Desa Geyongan sebanyak 5.941 jiwa. Data persentase rata-rata penduduk yang masuk dalam kategori miskin per Desember 2023 mencapai 49,82%, sebagaimana yang terlihat dalam Gambar 1. Angka ini cukup tinggi dibandingkan dengan persentase nasional yang sebesar 9,36%. Hal ini menunjukkan bahwa Desa Geyongan merupakan desa yang masih menghadapi permasalahan kemiskinan.



Gambar 1. Data Kemiskinan Desa Geyongan Per Tahun 2023

Sebagai solusi untuk tingkat kemiskinan yang tinggi, pemerintah meluncurkan program Bantuan

Langsung Tunai (BLT). Program ini mencakup berbagai bentuk bantuan, termasuk pemberian uang tunai dengan atau tanpa syarat tertentu. Program Bantuan Langsung Tunai (BLT) merujuk pada ketentuan yang tercantum dalam Peraturan Pemerintah Republik Indonesia Nomor 60 Tahun 2014 tentang Dana Desa yang bersumber dari Anggaran Pendapatan dan Belanja Negara. Sesuai dengan ketentuan dalam Peraturan Menteri Keuangan Nomor 94/PMK.07/2021 mengenai Percepatan Penyaluran Dana Desa dan Bantuan Langsung Tunai (BLT) Desa, program ini berfokus pada pemberian uang tunai Rp300.000 per bulan kepada masyarakat miskin [2].

Dikarenakan berasal dari dana desa maka pembagian BLT telah diatur dalam Peraturan Bupati Kabupaten Cirebon Nomor 25 Tahun 2020 Tentang Perubahan Atas Peraturan Bupati Cirebon Nomor 3 Tahun 2020 Tentang Tata Cara Pembagian dan Penetapan Rincian Dana Desa Setiap Desa di Kabupaten Cirebon. Dengan adanya aturan tersebut pembagian BLT seharusnya tepat sasaran, dimana masyarakat tidak mampu menjadi prioritas utama. Namun, dalam pelaksanaannya ada beberapa kesalahan pembagian yang dirasa tidak adil.

Dalam wawancara yang dilakukan dengan seorang penduduk Desa Geyongan, ditemukan ketidaksesuaian antara kriteria yang telah ditetapkan dan status penerima BLT. Seorang pejabat desa menyatakan bahwa penyebab masalah ini adalah penggunaan metode manual dalam proses pemilihan penerima BLT. Data yang dikumpulkan oleh relawan desa dari berbagai tingkat harus divalidasi secara manual, yang memakan waktu lama dan memiliki risiko ketidaktepatan sasaran.

Untuk menangani masalah tersebut, akan dilakukan proses *data mining* menggunakan algoritma KNN untuk mengklasifikasi calon penerima BLT. Berbagai kriteria akan diolah untuk menghasilkan nilai, yang akan dibandingkan dengan data pelatihan guna menentukan kelayakan calon penerima BLT. Penelitian ini menggunakan algoritma KNN karena secara efisien dapat menangani data pelatihan yang besar, tahan terhadap data pelatihan yang *noise*, dan dapat memberikan hasil data yang akurat [3].

Terdapat beberapa algoritma yang dapat digunakan dalam proses klasifikasi seperti *Decision Trees*, *Neural Networks*, *Support Vector Machines*, *KNN*, *Naïve Bayes*, dan masih banyak yang lainnya. Berbagai penelitian telah dilakukan untuk mengetahui perbandingan dari masing-masing algoritma. Seperti pada penelitian yang dilakukan oleh Purwanto dan Nugroho pada tahun 2023 mengenai Analisa Perbandingan Kinerja Algoritma C4.5 dan Algoritma *K-Nearest Neighbors* Untuk Klasifikasi Penerima Beasiswa. Penelitian tersebut mengambil sampel sebanyak 940 calon penerima beasiswa di Universitas Muhammadiyah Pringsewu sebagai data yang digunakan untuk klasifikasi. Klasifikasi ini dilakukan menggunakan algoritma *K-Nearest Neighbor* (KNN) dan C4.5. Penelitian ini mengembangkan metode untuk mendapatkan skema penerimaan beasiswa yang optimal dengan pemerataan tertinggi bagi penyelenggara universitas. Proses perhitungan dilakukan menggunakan bantuan alat RapidMiner. Berdasarkan pengujian yang dilakukan didapat hasil bahwa algoritma KNN memiliki performa yang lebih unggul dengan presisi sebesar 98,08%, akurasi 98,30%, dan recall 98,00% serta hasil AUC sebesar 1,000 yang dimana dapat dikategorikan sebagai *Excellent Classification*. Sedangkan dengan penggunaan algoritma C4,5 memperoleh akurasi sebesar 97,23% dengan nilai precision 94,43%, nilai recall 100,00% dan hasil AUC 0,956 [4].

Selanjutnya, penelitian yang dilakukan oleh Nurmahadudin dan Gunawan pada tahun 2019 mengenai klasifikasi air PDAM menggunakan algoritma *K-Nearest Neighbor* dan *K-Means*. Algoritma *K-Nearest Neighbor* (KNN) merupakan salah satu metode yang terdapat dalam klasifikasi yang bersifat *supervised learning* sedangkan *K-Means* merupakan algoritma *clustering* yang bersifat *unsupervised*. Pada penelitian tersebut kedua algoritma digunakan untuk klasifikasi kualitas air pada PDAM Banjarmasin. Parameter kualitas air yang digunakan meliputi mikrobiologi, kimia anorganik, parameter fisik, serta parameter kimia. Kualitas air yang diuji diklasifikasikan menjadi layak dan tidak layak. Berdasarkan hasil pengujian diperoleh akurasi dengan KNN lebih baik yaitu 88% dibandingkan dengan *K-Means* yaitu 76% [1].

Berdasarkan temuan hasil penelitian di atas, penerapan Algoritma *K-Nearest Neighbor* terbukti

memiliki performa terbaik dibandingkan dengan dua algoritma lainnya. Oleh karena itu, penelitian ini akan difokuskan pada penerapan Algoritma KNN untuk mengklasifikasi penerima BLT di Desa Geyongan, sehingga penentuan penerima BLT dapat lebih akurat dan tepat sasaran. Proses klasifikasi akan menggunakan kriteria yang telah ditetapkan oleh Desa Geyongan. Metodologi yang akan digunakan adalah *Knowledge Discovery in Databases* (KDD), yang membantu dalam penambangan data terstruktur dan terarah dari awal hingga akhir. Proses klasifikasi akan dilakukan menggunakan perangkat RapidMiner.

2. TINJAUAN PUSTAKA

2.1. Bantuan Langsung Tunai

Bantuan Langsung Tunai (BLT) adalah program pemerintah berupa bantuan uang tunai atau bantuan lain kepada masyarakat yang berada dalam kondisi ekonomi kurang mampu, dengan tujuan membantu mereka mengatasi kesulitan akibat Pandemi Covid-19. Pada tingkat pemerintahan desa, terselenggara program BLT-DD memanfaatkan Dana Desa untuk memberikan bantuan kepada warga tidak mampu di desa. Program ini dilaksanakan untuk membantu mereka dalam memenuhi kebutuhan dasar. Pengaturan terkait program ini dijabarkan dalam Peraturan Menteri Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Nomor 14 Tahun 2020, yang merupakan revisi ketiga dari Peraturan Menteri Desa, Pembangunan Daerah Tertinggal, dan Transmigrasi Nomor 11 Tahun 2019 mengenai Prioritas Penggunaan Dana Desa Tahun 2020. Peraturan ini mengatur tentang penyelenggaraan BLT yang bersumber dari dana desa [5].

2.2. Data Mining

Data mining merupakan alat yang memfasilitasi pengguna mengakses data dengan cepat, terutama data dalam jumlah besar. Lebih khusus lagi, *data mining* adalah suatu proses mengelola informasi yang besar dan sebelumnya tidak diketahui diekstraksi atau digali dari *database* yang besar. Proses ini melibatkan penggunaan analisis statistik pada data guna mengidentifikasi pola-pola yang sebelumnya tidak diketahui. Tujuannya adalah menghasilkan pengetahuan yang bermanfaat dan dapat dipahami, yang selanjutnya dapat digunakan untuk mengambil keputusan bisnis yang krusial [6].

2.3. Klasifikasi

Klasifikasi adalah metode pengelompokan objek berdasarkan karakteristik yang dimilikinya. Proses klasifikasi dapat dijalankan dengan berbagai metode, baik melalui tindakan manual manusia maupun dengan dukungan teknologi. Metode manual melibatkan partisipasi manusia tanpa bantuan dari algoritma komputer. Di sisi lain, klasifikasi yang menggunakan teknologi melibatkan penggunaan beberapa algoritma untuk membantu dalam proses pengelompokan tersebut [7].

2.4. Algoritma K-Nearest Neighbor

KNN merupakan suatu algoritma yang berfungsi untuk mengklasifikasikan objek berdasarkan data latih yang memiliki jarak terdekat dengan objek tersebut. Cara kerja KNN melibatkan pencarian jarak terdekat antara data yang sedang diuji dengan K tetangga terdekat dalam data latih. Klasifikasi baru diprediksi berdasarkan mayoritas klasifikasi dari tetangga-tetangga tersebut. Penetapan nilai optimal untuk parameter K dalam algoritma ini tergantung pada sifat-sifat data [8].

Sebelum menghitung jarak antara data dan objek, langkah awalnya adalah menentukan nilai K, yang menunjukkan jumlah tetangga yang akan digunakan dalam proses klasifikasi. Kemudian, untuk menghitung jarak antara dua titik, yaitu titik pada data latih dan titik pada data uji, digunakan rumus *Euclidean* dengan persamaan sebagai berikut:

$$d(X1, X2) = \sum_{i=1}^n \sqrt{(X1i - X2i)^2} \quad (1)$$

Dimana:

d (X1,X2)= Jarak *Euclidean*

X1 = data 1

X2 = data 2

i = fitur ke-

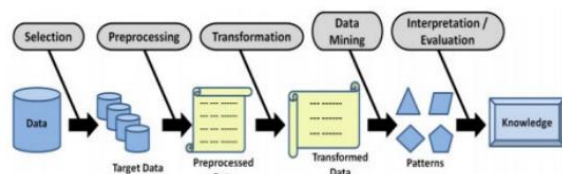
n = jumlah fitur

2.5. Confusion Matrix

Confusion matrix adalah teknik yang dipakai untuk mengevaluasi tingkat keakuratan dalam data mining. Tabel matriks dalam *confusion matrix* terdiri dari dua kelas, yaitu kelas positif dan negatif. Dalam menangani permasalahan klasifikasi, evaluasi menggunakan *confusion matrix* dianggap sebagai solusi optimal selama proses klasifikasi. Setelah melalui proses pengujian data, outputnya kemudian dimasukkan ke dalam tabel *confusion matrix* untuk menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F-Measure* [9].

2.6. Knowledge Discovery in Database

Konsep KDD (*Knowledge Discovery in Database*) dapat diartikan sebagai proses komprehensif dalam mengekstraksi atau mengidentifikasi informasi potensial kumpulan data yang besar. *Output* yang dihasilkan biasanya dianalisis untuk membentuk dasar pengetahuan, yang selanjutnya dapat digunakan sebagai landasan untuk pengambilan keputusan [10]. tahapan dalam KDD melibatkan serangkaian langkah seperti berikut :



Gambar 2. Tahapan KDD

2.7. RapidMiner

RapidMiner adalah sebuah aplikasi yang dikembangkan oleh Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* bersama Ralf Klinkenberg. Perangkat lunak ini dirancang sebagai sumber terbuka dan dikodekan menggunakan bahasa pemrograman Java di bawah lisensi *GNU Public Licence*, memungkinkannya untuk berjalan pada berbagai sistem operasi. Fokus utama RapidMiner adalah pada pengolahan data, khususnya dalam konteks *data mining* [11].

3. METODE PENELITIAN

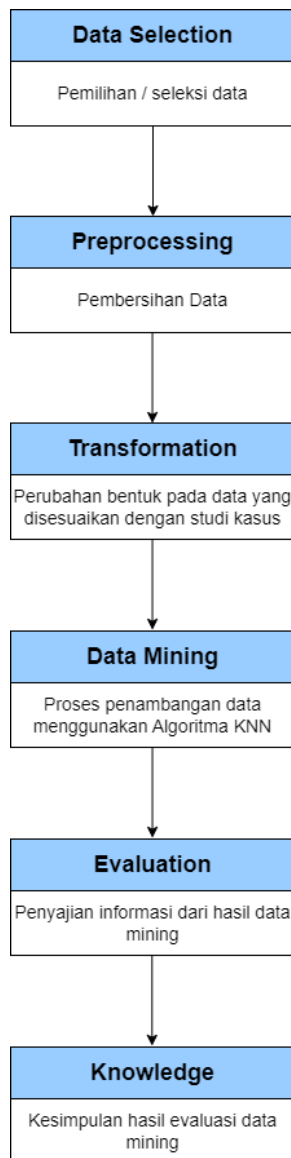
3.1. Metodologi Penelitian

Penelitian ini menggunakan metode *Knowledge Discovery in Database*. Metode ini terdiri dari lima tahapan yang harus dilalui yaitu *data selection*, *preprocessing*, *transformation*, *data mining*, *evaluation*, dan *knowledge*. KDD digunakan pada penelitian ini karena termasuk ke dalam metode yang tahapannya terstruktur, dapat digunakan untuk skala data yang besar, dan banyak penelitian yang menggunakan metode ini dengan memperoleh hasil yang baik.

3.2. Rancangan Penelitian

Penelitian ini akan dilakukan menggunakan tahapan yang sesuai dengan model yang sebelumnya telah dibahas yaitu *Knowledge Discovery in Database*. Berikut merupakan rincian dari rancangan penelitian yang akan dilaksanakan.

- *Data selection*, proses pemilihan data yang akan digunakan. Penelitian ini menggunakan data calon penerima Bantuan Langsung Tunai Dana Desa Geyongan Tahun 2023. Atribut yang digunakan disesuaikan dengan kebutuhan dalam proses seleksi penerima BLT yang ada di Desa Geyongan.
- *Preprocessing*, proses pembersihan data pada tahap ini, data yang tidak lengkap, duplikat, atau memiliki masalah lainnya diolah agar proses penambangan data menghasilkan *output* yang optimal
- *Transformation*, proses pengubahan bentuk atau format data. Proses pengubahan ini dilakukan dengan tujuan Menyusun data untuk menyesuaikan proses *data mining*.
- *Data mining*, proses penambangan data. Penelitian ini menggunakan Algoritma *K-Nearest Neighbor* dengan tujuan klasifikasi penerima Bantuan Langsung Tunai di Desa Geyongan.
- *Evaluation*, tahap ini dapat dikatakan sebagai tahap penyajian informasi dari hasil penambangan data. Informasi direpresentasikan dalam format yang dapat dipahami oleh pihak yang terlibat
- *Knowledge*, tahap ini berisi kesimpulan dari evaluasi data mining yang dihasilkan.



Gambar 3. Rancangan Penelitian

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

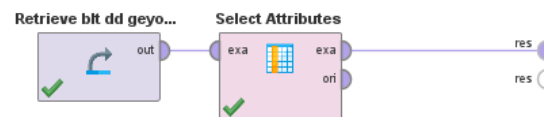
Data yang digunakan dalam penelitian ini merupakan data Calon Penerima Bantuan Langsung Tunai Dana Desa (BLT DD) Geyongan tahun 2023, disajikan dalam format *Excel* dengan total 671 records. Data ini mencakup 22 atribut, yang kemudian hanya dipilih 9 atribut relevan. Acuan yang digunakan dalam pemilihan atribut ini adalah dari Peraturan Kuwu Nomor 3 Tahun 2023 Pasal 3, yang menjadi landasan untuk menentukan calon penerima Bantuan Langsung Tunai di Desa Geyongan. Berikut merupakan atribut terpilih yang akan digunakan pada proses selanjutnya.

Tabel 1. Pemilihan Atribut

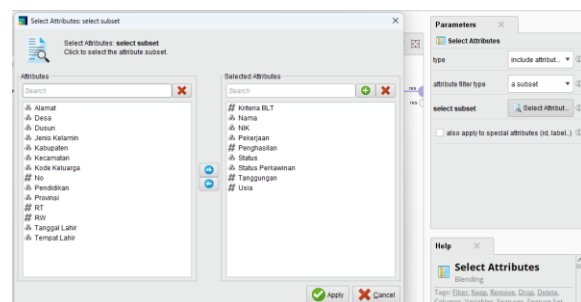
No	Atribut	Jenis	Keterangan	Detail
1.	Nama	Kategorikal	-	-
2.	NIK	Numerik	-	-
3.	Penghasilan	Numerik	Digunakan Atribut	Atribut Utama

No	Atribut	Jenis	Keterangan	Detail
4.	Usia	Numerik	Digunakan Atribut	Atribut Utama
5.	Status Perkawinan	Kategorikal	Digunakan Atribut	Atribut Utama
6.	Tanggungan	Numerik	Digunakan Atribut	Atribut Utama
7.	Pekerjaan	Kategorikal	Digunakan Atribut	Atribut Utama
8.	Kriteria BLT	Numerik	Digunakan Atribut	Atribut Utama
9.	Status	Kategorikal	Digunakan Atribut	Label/Class

Pemilihan atribut dilakukan menggunakan RapidMiner dengan bantuan operator *Select Attributes*.



Gambar 4. Pemilihan Atribut Dengan RapidMiner



Gambar 5. Pemilihan Atribut Relevan

Atribut yang terpilih akan diolah menggunakan metode *K-Nearest Neighbor* dalam proses *data mining*. Setelah pemilihan atribut, target data sudah tersedia untuk tahap berikutnya.

Gambar 6. Potongan Hasil Data Selection

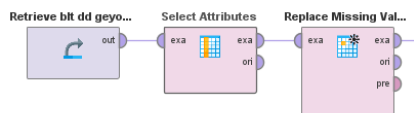
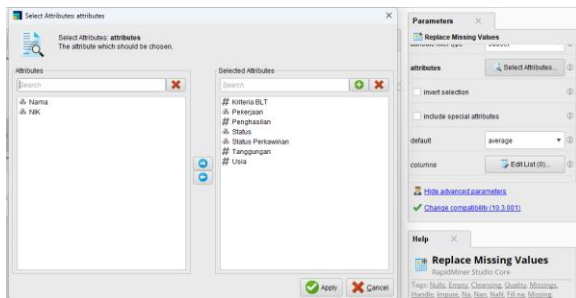
4.2. Preprocessing

Dalam kasus ini, dengan data yang terdiri dari 671 dan hanya 23 nilai yang hilang, penting untuk memilih metode yang paling efisien dan efektif untuk membersihkan data. Meskipun opsi mengabaikan *tuple* atau mengisi atribut kosong secara manual dapat menjadi solusi, namun keduanya memerlukan waktu dan upaya yang signifikan, terutama dalam skala data yang besar seperti ini. Oleh karena itu, menggunakan nilai rata-rata sebagai pengganti untuk nilai yang hilang menjadi solusi yang lebih tepat.

Name	Type	Missing	Statistics
Tanggungan	Integer	1	Min: 0, Max: 7, Average: 2.169
Pendidikan	Nominal	1	Label: Tamal 5-3 (terdapat) (1), Value: Tamal 5-3 (terdapat) (316), Tamal 5-3 (tidak terdapat) (151), [20 more]
Pekerjaan	Nominal	16	Label: Petajar (1), Value: Wiraswasta (231), Buruh Tani (151), [20 more]
Penghasilan	Integer	20	Min: 0, Max: 20000000, Average: 2622887.865
Kriteria BLT	Integer	20	Min: 1, Max: 5, Average: 2.092
Status	Nominal	3	Label: Layak (165), Value: Tidak layak (332), Layak (171), [1 more]

Gambar 7. Missing Value

Digunakan operator *Replace Missing Value* pada RapidMiner untuk mengisi *missing value* pada data dengan nilai rata rata dari masing masing atribut.

Gambar 8. Hapus Missing Value Menggunakan Operator *Replace Missing Value*

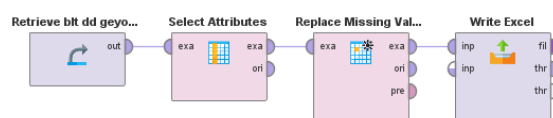
Gambar 9. Memilih Data

Setelah dilakukan pengisian nilai kosong menggunakan nilai rata rata dari masing masing atribut maka sudah tidak ada lagi *missing value* pada data.

Name	Type	Missing	Statistics
Status Perkawinan	Nominal	0	Label: Belum Kawin (17), Value: Kawin (303), Janda/Duda (101), [1 more]
Tanggungan	Integer	0	Min: 0, Max: 7, Average: 2.169
Pekerjaan	Nominal	0	Label: Petajar (1), Value: Wiraswasta (247), Buruh Tani (151), [20 more]
Penghasilan	Integer	0	Min: 0, Max: 20000000, Average: 2622887.865
Kriteria BLT	Integer	0	Min: 1, Max: 5, Average: 2.092
Status	Nominal	0	Label: Layak (165), Value: Tidak layak (332), Layak (171), [1 more]

Gambar 10. Hasil *Replace Missing Value*

Dilakukan *export* data ke dalam bentuk excel menggunakan operator *Write Excel* untuk digunakan pada tahap transformasi data.

Gambar 11. *Export Data*

4.3. Transformation

Metode *data mining* memerlukan data yang sesuai dengan kebutuhan algoritma saat dieksekusi. Algoritma K-NN, misalnya, memerlukan data

numerik untuk mencari pengetahuan atau pola. Oleh karena itu, pada tahap ini, dilakukan transformasi format data tanpa mengubah makna aslinya. Data yang ditransformasi yaitu nilai nilai pada attribute usia, status perkawinan, tanggungan, pekerjaan, penghasilan, dan kriteria BLT. Proses penentuan bobot sangat fleksibel, disesuaikan dengan studi kasus dan pemahaman pakar.

Pembobotan ini juga membantu dalam menentukan jarak antara data latih dan uji, dengan bobot yang diberikan berdasarkan relevansi tetangga terdekat. Pemilihan nilai K pada saat proses data mining yang optimal dalam KNN sangat penting untuk memastikan akurasi klasifikasi yang tinggi, di mana nilai K yang terlalu kecil dapat membuat model terlalu sensitif terhadap noise, sementara nilai K yang terlalu besar dapat mengurangi sensitivitas terhadap data relevan. Kombinasi antara pembobotan yang tepat dan nilai K yang optimal akan memastikan bahwa bantuan disalurkan kepada penerima yang benar-benar membutuhkan, sesuai dengan kriteria yang telah ditetapkan secara objektif.

a. Atribut Usia

Tabel 2. Transformasi *Record* Atribut Usia

Kategori	Nilai Bobot
18 – 30 tahun	1
31 – 50 tahun	0.6
51 – 70 tahun	0.3
>71 tahun	0

b. Atribut Status Perkawinan

Tabel 3. Transformasi *Record* Atribut Status Perkawinan

Kategori	Nilai Bobot
Belum kawin	1
Kawin	0.5
Janda/duda	0

c. Atribut Tanggungan

Tabel 4. Transformasi *Record* Atribut Tanggungan

Kategori	Nilai Bobot
<1 – 2 orang	1
3 – 4 orang	0.6
5 – 6 orang	0.3
>6 orang	0

d. Atribut Pekerjaan

Tabel 5. Transformasi *Record* Atribut Pekerjaan

Kategori	Nilai Bobot
Pekerjaan formal: - Guru swasta - Dosen swasta - Karyawan Perusahaan Swasta - Karyawan Honorer - Karyawan Perusahaan Pemerintah - Pegawai Negeri Sipil - Perangkat Desa - TNI - POLRI	1

Pekerjaan informal: • Buruh harian lepas • Buruh jasa perdagangan hasil bumi • Buruh tani • Buruh usaha jasa transportasi dan perhubungan • Pedagang barang kelontong • Pedagang keliling • Petani • Sopir • Wiraswasta	0.5
Tidak bekerja: • Ibu Rumah Tangga • Pelajar • Purnawirawan/Pensiunan • Belum bekerja	0

e. Atribut Penghasilan

Tabel 6. Transformasi *Record* Atribut Penghasilan

Kategori	Nilai Bobot
>5.000.000	1
3.000.001 – 5.000.000	0.75
1.000.001 – 3.000.000	0.5
500.000 – 1.000.000	0.25
<500.000	0

f. Atribut Kriteria BLT

Tabel 7. Transformasi *Record* Atribut Kriteria BLT

Kategori	Nilai Bobot
Keluarga Miskin/tidak mampu yang berdomisili di Desa tidak punya NIK/KK	1
Punya penyakit kronis/menahun	0.75
Kehilangan mata pencaharian	0.5
Belum terdata DTKS	0.25
Masuk DTKS belum dapat JPS (Tidak termasuk penerima PKH, Kartu Sembako, Kartu Pra Kerja, Bansos Tunai, dan Program Bansos Pemerintah Lainnya)	0

Berikut adalah potongan data hasil dari seluruh proses *data transformation* dapat dilihat pada Gambar 12.

Usia	Status Per.	Tanggung	Pekerjaan	Penghasil	Kriteria BL	Status
0,6	0,5	1,0	0,5	0,5	0,3	Tidak layak
0,3	0,5	0,0	0,5	0,5	0,3	Layak
0,6	0,5	1,0	0,5	0,8	0,3	Tidak layak
0,6	0,5	0,6	0,5	0,5	0,3	Tidak layak
0,6	0,5	1,0	0,5	0,8	0,3	Tidak layak
0,6	0,5	0,3	0,5	0,5	0,3	Tidak layak
0,3	0,0	0,6	0,5	0,5	0,0	Layak
0,6	0,5	0,6	0,5	0,5	0,3	Tidak layak
0,6	0,5	0,6	1,0	1,0	0,3	Tidak layak
0,6	0,5	0,6	0,5	1,0	0,3	Tidak layak
0,6	0,5	0,6	0,5	0,8	0,3	Tidak layak
0,6	0,5	1,0	0,5	0,5	0,3	Tidak layak
0,3	0,5	0,6	0,5	0,3	0,3	Layak
0,6	0,5	0,6	0,5	1,0	0,3	Tidak layak
0,3	0,5	1,0	0,5	0,3	0,0	Layak
0,0	0,5	1,0	0,5	0,3	0,8	Layak
0,6	0,5	0,6	0,5	0,5	0,3	Tidak layak
0,6	0,5	1,0	0,5	0,3	0,3	Layak

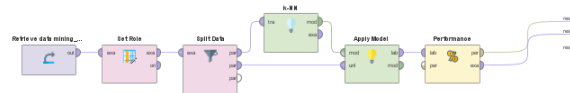
Gambar 12. Potongan Hasil Proses Data Transformation

4.4. Data Mining

Tahap utama dalam metodologi KDD adalah proses *data mining*, dimana tujuannya adalah

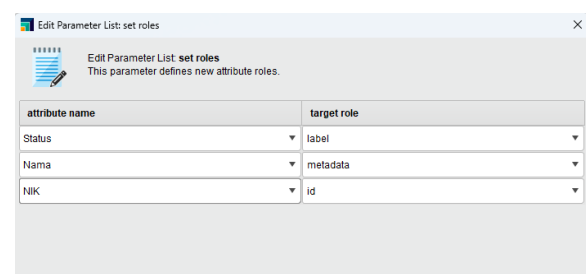
menemukan pola, pengetahuan, dan informasi menarik dalam data yang telah melewati langkah-langkah sebelumnya. Pada tahap ini, fokus pada pemilihan tugas data mining menggunakan teknik klasifikasi dengan memanfaatkan algoritma K-NN dan dijalankan menggunakan alat bantu RapidMiner.

RapidMiner sering digunakan sebagai alat untuk mengekstraksi pola data sesuai dengan tujuan pemrosesan data. Meskipun demikian, tidak semua algoritma dan metode yang tersedia dapat sepenuhnya sesuai untuk mengolah dataset yang sedang diproses. Oleh karena itu, perlu dilakukan penyesuaian pola data sebelumnya untuk memastikan kecocokan yang optimal.



Gambar 13. Pemodelan K-NN pada Rapidminer

- Retrieve* pada penelitian ini digunakan untuk mengakses *dataset* yang telah disimpan di repositori RapidMiner. Sebelumnya, *dataset* dimasukkan secara manual ke dalam repositori RapidMiner
- Set Role* digunakan untuk menetapkan peran (role) dari atribut dalam *dataset*. Atribut Status ditetapkan sebagai label untuk proses klasifikasi, NIK diberi role id sebagai penanda atau pengenal unik yang digunakan untuk mengidentifikasi entitas data, dan Nama diberi role metadata yang memberikan deskripsi atau konteks dari NIK.



Gambar 14. Memberikan Role Untuk Beberapa Attributes

- Split Data* dilakukan untuk merancang skenario data mining. Pada tahap ini, *dataset* yang telah diakses melalui operator *retrieve* akan dipartisi menjadi dua bagian yaitu data latih dan data uji. Penelitian ini menggunakan empat skenario pembagian data yaitu 90% : 10%, 80% : 20%, 70% : 30%, dan 60% : 40% untuk menemukan pemodelan terbaik.
- K-NN pada Gambar 13 adalah operator inti, di mana operator ini bertanggung jawab untuk menghasilkan model K-NN yang akan digunakan dalam proses klasifikasi. Pada operator K-NN ini, nilai K yang akan diuji meliputi 1, 3, 5, 7, 9, 11, 13, 15, 17, dan 19. Algoritma K-NN akan melakukan klasifikasi pada data uji dengan mempertimbangkan kedekatan data uji terhadap

data latih sebagai tetangga, dengan nilai-nilai K yang telah ditentukan sebelumnya.

- e. *Apply Model* digunakan untuk menerapkan model yang telah dikonfigurasi sebelumnya. Dalam konteks model yang telah dibuat, hal ini mencakup penggunaan K-NN serta skenario pembagian data yang telah diatur sebelumnya.
- f. *Performance* berfungsi sebagai alat evaluasi kinerja dalam tugas klasifikasi. Selain digunakan untuk evaluasi dalam konteks klasifikasi, operator ini juga dapat diterapkan untuk mengevaluasi teknik selain klasifikasi. Di sisi lain, operator ini menentukan teknik yang akan digunakan dan melakukan perhitungan kriteria umum yang terkait dengan teknik tersebut.

4.5. Evaluation

Evaluasi merupakan langkah krusial dalam mengidentifikasi pola dan pengetahuan yang diperoleh dari proses *data mining*. Pada tahap akhir ini, model prediksi dinilai untuk menilai sejauh mana kesesuaian hipotesis dengan hasil aktual. Model penelitian dieksekusi melalui pengujian pada data latih dan data uji. Pemodelan dilakukan dengan empat skenario pembagian data latih dan data uji yang menghasilkan nilai akurasi sebagai berikut:

- a. Skenario pembagian data berupa 90% data latih dan 10% data uji.

Tabel 8. Skenario 90% : 10%

Nilai K	Accuracy	Precision	Recall	F-Measure
1	77,61%	82,76%	70,59%	76,19%
3	85,07%	90,00%	79,41%	84,37%
5	85,07%	90,00%	79,41%	84,37%
7	83,58%	79,49%	91,18%	84,93%
9	85,07%	80,00%	94,12%	86,49%
11	83,58%	79,49%	91,18%	84,93%
13	85,07%	85,29%	85,29%	85,29%
15	83,58%	79,49%	91,18%	84,93%
17	86,57%	85,71%	88,24%	86,96%
19	86,57%	85,71%	88,24%	86,96%

- b. Skenario pembagian data berupa 80% data latih dan 20% data uji.

Tabel 9. Skenario 80% : 20%

Nilai K	Accuracy	Precision	Recall	F-Measure
1	78,36%	81,67%	73,13%	77,16%
3	84,33%	88,33%	79,10%	83,46%
5	85,07%	88,52%	80,60%	84,37%
7	86,57%	87,69%	85,07%	86,36%
9	87,31%	89,06%	85,07%	87,02%
11	85,82%	87,50%	83,58%	85,50%
13	86,57%	87,69%	86,07%	86,87%
15	86,57%	87,69%	86,07%	86,87%
17	82,84%	83,35%	83,58%	83,46%
19	84,33%	84,85%	73,58%	78,81%

- c. Skenario pembagian data berupa 70% data latih dan 30% data uji.

Tabel 10. Skenario 70% : 30%

Nilai K	Akurasi	Precision	Recall	F-Measure
1	73,27%	72,83%	75,25%	74,02%
3	82,18%	78,26%	89,11%	83,33%
5	82,67%	83,00%	82,18%	82,59%
7	82,67%	83,67%	81,19%	82,41%
9	82,18%	82,18%	82,18%	82,18%
11	80,69%	79,81%	82,18%	80,98%
13	81,19%	81,19%	81,19%	81,19%
15	76,73%	74,11%	82,18%	77,94%
17	78,71%	77,36%	81,19%	79,23%
19	80,20%	78,50%	83,17%	80,77%

- d. Skenario pembagian data berupa 60% data latih dan 40% data uji.

Tabel 11. Skenario 60% : 40%

Nilai K	Akurasi	Precision	Recall	F-Measure
1	73,88%	72,22%	77,61%	74,82%
3	80,22%	76,47%	87,31%	81,53%
5	81,72%	82,44%	80,60%	81,51%
7	81,34%	82,81%	79,10%	80,91%
9	80,22%	80,45%	79,85%	80,15%
11	79,85%	80,77%	78,36%	79,55%
13	79,85%	79,41%	80,60%	80,00%
15	80,22%	80,00%	80,60%	80,30%
17	80,60%	79,71%	82,09%	80,88%
19	80,97%	80,29%	82,09%	81,18%

Dari 4 model yang telah dibuat menggunakan pembagian data uji dan data latih serta penggunaan beberapa nilai K diketahui bahwa nilai *accuracy* tertinggi yang didapat adalah pada penggunaan dataset dengan data latih 80%, data uji 20%, dan nilai K= 9 yakni sebesar 87,31%. Selanjutnya untuk menghitung evaluasi, penelitian ini menggunakan teknik *confusion matrix*.

Tabel 12. *Confusion Matrix*

	<i>True (Tidak Layak)</i>	<i>True (Layak)</i>
Prediksi Tidak Layak	60 (TN)	10 (FN)
Prediksi Layak	7 (FP)	57 (TP)

Tabel *confusion matrix* di atas diperoleh dari proses menggunakan model K-NN dengan data latih 80%, uji 20%, dan nilai K = 9, yang memiliki akurasi terbesar. Akurasi yang berhasil dicapai adalah 87,31%. Uraian penjelasan prediksi tabel *confusion matrix* diatas dari yang diperoleh dari 134 records data uji adalah:

- Sebanyak 57 calon penerima diprediksi layak dan aktualnya juga layak.
- Sebanyak 7 calon penerima diprediksi layak tetapi aktualnya tidak layak.
- Sebanyak 60 calon penerima diprediksi tidak layak dan aktualnya juga tidak layak.
- Sebanyak 10 calon penerima diprediksi tidak layak tetapi aktualnya layak.

Berikut merupakan perhitungan jika dilakukan secara manual :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} = \frac{57+60}{57+7+10+60} = \frac{117}{134}$$

$$= 0,8731 \times 100\% = 87,31\%$$

$$Recall = \frac{TP}{TP+FN} = \frac{57}{57+10} = \frac{57}{67}$$

$$= 0,8507 \times 100\% = 85,07\%$$

$$Precision = \frac{TP}{TP+FP} = \frac{57}{57+7} = \frac{57}{64}$$

$$= 0,8906 \times 100\% = 89,06\%$$

$$F-Measure = \frac{2 \times p \times r}{p+r} = \frac{2 \times 89,06\% \times 85,07\%}{89,06\% + 85,07\%}$$

$$= 87,02\%$$

4.6. Knowledge

Dari 4 model K-NN yang telah dibuat menggunakan empat skenario pembagian data yakni 90%:10%, 80%:20%, 70%:30%, 60%:40% diketahui bahwa akurasi tertinggi diperoleh ketika menggunakan data latih 80%, uji 20%, dan nilai K = 9 dengan akurasi sebesar 87,31%. Meskipun hasil klasifikasi model ini tidak mencapai 100% keakuratan.

5. KESIMPULAN DAN SARAN

Berdasarkan analisis temuan dari penelitian yang dilakukan dengan studi kasus di Desa Geyongan, dapat disimpulkan bahwa ketidaksesuaian antara validitas data kriteria penerima yang telah ditetapkan dengan penentuan status kelayakan penerima Bantuan Langsung Tunai di Desa Geyongan, dapat ditangani melalui penggunaan metode data mining dengan teknik klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN). Pemodelan data menggunakan algoritma *K-Nearest Neighbor* menggunakan *Software* RapidMiner telah berhasil mendapatkan akurasi tertinggi sebesar 87,31% dengan skenario pembagian data 80:20 dan nilai K=9. Hal ini menunjukkan bahwa KNN merupakan metode yang efektif dan akurat dalam mengklasifikasi penerima BLT, yang diharapkan dapat memperbaiki ketidaktepatan sasaran penerima bantuan di Desa Geyongan.

K-Nearest Neighbors (KNN) dapat meningkatkan akurasi dalam penentuan penerima Bantuan Langsung Tunai (BLT) dengan mengurangi

subjektivitas yang ada dalam metode manual. KNN menggunakan data historis dan kriteria objektif untuk mengklasifikasikan calon penerima, memastikan bahwa keputusan didasarkan pada data konkret dan mengurangi bias individu. Dengan memvalidasi dan memverifikasi data secara otomatis, KNN meningkatkan konsistensi dan akurasi penentuan penerima bantuan. Algoritma ini juga tahan terhadap data noise dan dapat menggunakan pembobotan serta nilai K yang optimal untuk hasil yang lebih sensitif dan akurat. Hal ini memastikan bahwa bantuan disalurkan secara tepat sasaran kepada mereka yang benar-benar membutuhkan, sesuai dengan kriteria yang telah ditetapkan.

Disarankan agar pemerintah desa mengadopsi teknologi ini secara berkelanjutan dan mengintegrasikannya dengan sistem informasi yang ada untuk memastikan distribusi bantuan yang lebih adil dan tepat sasaran, serta melakukan pelatihan kepada relawan dan pejabat desa terkait penggunaan teknologi ini untuk meningkatkan efisiensi dan efektivitas dalam penyaluran BLT.

DAFTAR PUSTAKA

- [1] G. Rudi Cahyono, P. Negeri Banjarmasin, and U. Lambung Mangkurat, "Klasifikasi Kualitas Air PDAM Menggunakan Algoritma KNN Dan K-Means," 2019.
- [2] I. Sofi, "Efektivitas Bantuan Langsung Tunai Dana Desa Dalam Pemulihan Ekonomi Di Desa," Jakarta, 2021.
- [3] C. A. Rachma, "Implementasi Algoritma K-Nearest Neighbor Dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur," Malang, 2022.
- [4] A. Purwanto *et al.*, "Analisa Perbandingan Kinerja Algoritma C4.5 Dan Algoritma K-Nearest Neighbors Untuk Klasifikasi Penerima Beasiswa," 2023. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- [5] M. U. L. Pulana, "Implementasi Program Bantuan Langsung Tunai (BLT) Dana Desa Dalam Rangka Penanganan Dampak Bencana Covid 19 Di Desa Kanaungan Kecamatan Labakkang," 2020.
- [6] C. Zai and T. Komputer, "Implementasi Data Mining Sebagai Pengolahan Data," 2022.
- [7] A. P. Wibawa, M. Guntur, A. Purnama, M. Fathony Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, 2018.
- [8] A. Pauji, S. Aisyah, A. Surip, R. Saputra, and I. Ali, "Implementasi Algoritma K-Nearest Neighbor Dalam Menentukan Penerima Bantuan Langsung Tunai," *KOPERTIP : Jurnal Ilmiah Manajemen Informatika dan Komputer*, vol. 4, no. 1, pp. 21–27, Jun. 2022, doi: 10.32485/kopertip.v4i1.114.

- [9] B. Santosa and A. Umam, *Data Mining dan Big Data Analytics (2nd ed.)*. Penebar Media Pustaka, 2018.
- [10] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Deepublish Publisher, 2020.
- [11] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan, “Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner,” *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 1, Apr. 2021, doi: 10.30813/jbase.v4i1.2729.