



Perbandingan Metode C4.5 dan K-Nearest Neighbor dalam Klasifikasi Status Penerima Bantuan Program Keluarga Harapan di Desa Selat

Ida Ayu Gede Basma Pujanti^{1*}, I Gusti Putu Suharta², I Made Candiasa³

Program Studi Matematika, FMIPA, Universitas Pendidikan Ganesha, Jalan Udayana Nomor 11 Singaraja, Bali, Indonesia 81116.

Email Korespondensi: ayu.basma@undiksha.ac.id

Abstrak

Kemiskinan merupakan permasalahan sosial yang kompleks dan menjadi tantangan utama bagi pemerintah Indonesia. Salah satu upaya penanggulangan kemiskinan adalah melalui Program Keluarga Harapan yang bertujuan untuk meningkatkan kesejahteraan masyarakat penerima program, yang dipilih melalui proses seleksi. Penelitian ini bertujuan untuk membangun model klasifikasi calon penerima bantuan dengan membandingkan efektivitas algoritma C4.5 dan K-Nearest Neighbor. Penelitian ini menggunakan 538 entri data dengan enam atribut, yaitu pekerjaan, kesehatan, pendidikan, kesejahteraan sosial, kepemilikan rumah, dan kepemilikan aset berupa lahan. Hasil pengujian menunjukkan bahwa algoritma C4.5 menghasilkan nilai akurasi, presisi, dan recall sebesar 100%, serta berhasil mereduksi jumlah atribut menjadi empat atribut. Sementara itu, untuk algoritma K-Nearest Neighbor dengan nilai $k = 21$ memperoleh nilai akurasi sebesar 99,07% dan presisi sebesar 98,46%. Temuan ini menunjukkan bahwa algoritma C4.5 lebih efektif daripada algoritma K-Nearest Neighbor dalam klasifikasi penerima bantuan Program Keluarga Harapan, sehingga dapat dijadikan alternatif dalam proses pengambilan keputusan.

Kata kunci: Klasifikasi; K-Nearest Neighbor; Kemiskinan; Program Keluarga Harapan.

Comparison of the C4.5 and K-Nearest Neighbor Methods in Classifying the Status of Recipients of the Family Hope Program Assistance in Selat Village

Abstract

Poverty is a complex social issue and remains a major challenge for the Indonesian government. One of the efforts to alleviate poverty is the Family Hope Program, which aims to improve the welfare of beneficiary households selected through a screening process. This study aims to develop a classification model for prospective beneficiaries by comparing the effectiveness of the C4.5 and K-Nearest Neighbor algorithms. The study utilizes 538 data entries with six attributes: employment, health, education, social welfare, home ownership, and land asset ownership. The evaluation results show that the C4.5 algorithm achieved an accuracy, precision, and recall of 100%, while also reducing the number of attributes to four. Meanwhile, the K-Nearest Neighbor algorithm with $k = 21$ achieved an accuracy of 99,07% and a precision of 98,46%. These findings indicate that the C4.5 algorithm is more effective than the K-Nearest Neighbor algorithm in classifying beneficiaries of the Family Hope Program, and thus can serve as an alternative for decision-making processes.

Keywords: Classification; K-Nearest Neighbor; Poverty; Family Hope Program.

How to Cite: Pujanti, I. A. G. B., Suharta, I. G. P., & Candiasa, I. M. (2025). Perbandingan Metode C4.5 dan K-Nearest Neighbor dalam Klasifikasi Status Penerima Bantuan Program Keluarga Harapan di Desa Selat. *Empiricism Journal*, 6(3), 1295–1304. <https://doi.org/10.36312/ej.v6i3.3256>



<https://doi.org/10.36312/ej.v6i3.3256>

Copyright© 2025, Pujanti et al.

This is an open-access article under the CC-BY-SA License.



PENDAHULUAN

Isu kemiskinan menjadi permasalahan sosial yang kompleks dan terus menjadi tantangan utama bagi pemerintah di berbagai negara termasuk Indonesia (Suryahadi et al., 2020). Banyak upaya telah dilakukan oleh pemerintah Indonesia, salah satunya adalah dengan mengadakan program bantuan sosial yang disebut Program Keluarga Harapan (PKH) yang diharapkan dapat membantu meningkatkan kesejahteraan masyarakat penerima bantuan. Program ini bertujuan untuk memberikan bantuan tunai kepada keluarga yang dikategorikan sangat miskin, sehingga diharapkan mereka dapat meningkatkan akses

terhadap pendidikan, kesehatan, dan kebutuhan dasar lainnya (Kementerian Sosial Republik Indonesia, 2021).

Proses penyaluran bantuan sosial PKH masih menghadapi kendala akibat kurangnya data yang terkini dan akurat mengenai kondisi sosial ekonomi masyarakat. Data yang tidak akurat ini menyebabkan banyak bantuan tidak tepat sasaran. Badan Pemeriksa Keuangan (BPK) menemukan bahwa pada tahun 2021 terdapat penerima bantuan yang sudah meninggal dunia namun masih tercatat sebagai penerima aktif, serta adanya data ganda dan NIK yang tidak valid dalam daftar penerima. Proses pendataan yang kurang berkesinambungan dan data yang tidak diperbarui secara berkala dapat menyebabkan beberapa keluarga yang seharusnya menerima bantuan justru terlewatkan atau sebaliknya (Anggraeni & Nugroho, 2022). Oleh karena itu, dibutuhkan metode yang efektif dan efisien dalam menganalisis data agar pemerintah dapat memperoleh hasil yang objektif dalam menentukan penerima manfaat dari program bantuan ini.

Algoritma C4.5 dan K-Nearest Neighbor (K-NN) merupakan dua pendekatan yang dapat membantu proses klasifikasi dalam penentuan penerima bantuan PKH. Di antara metode pohon keputusan, algoritma C4.5 dikenal karena kemampuannya menghasilkan model klasifikasi yang mudah dipahami dan ramah pengguna. Pohon keputusan adalah alat visual yang digunakan dalam analisis untuk membantu pengambilan keputusan dengan menggambarkan kemungkinan hasil di masa depan (Puspita et al., 2022). Algoritma C4.5 ini bekerja dengan membagi data kedalam subset berdasarkan atribut tertentu, yang nantinya membentuk struktur pohon keputusan untuk pengambilan keputusan. Keunggulan utama C4.5 adalah kemampuannya dalam mengidentifikasi pola pada data kompleks dan menghasilkan aturan klasifikasi yang jelas (Siska & Heni, 2021).

Algoritma K-NN memiliki keunggulan yaitu algoritmanya yang sederhana namun efektif, dapat mengklasifikasikan data berdasarkan kedekatan atau kemiripan dengan data lainnya (Bhatia, 2010). Dalam konteks penentuan penerima bantuan PKH, algoritma ini dapat membantu dengan memperkirakan status bantuan berdasarkan kondisi keluarga lain yang memiliki karakteristik serupa. Algoritma K-NN bergantung pada data jarak atau kemiripan, sehingga sangat bergantung pada kualitas dan keterbaruan data yang dimiliki. Algoritma K-NN menentukan kategori kelas dengan menilai kemiripan antara data baru dan data yang telah diklasifikasikan sebelumnya. Pada dasarnya, algoritma ini beroperasi dengan membandingkan data baru dengan dataset pelatihan yang sudah ada (Atma & Setyanto, 2018). Meskipun sederhana, algoritma ini dikenal memiliki akurasi tinggi dalam berbagai kasus klasifikasi, khususnya saat data memiliki pola tertentu yang tidak kompleks. Kedua algoritma dipilih karena representasi pendekatan rule-based (C4.5) dan proximity-based (K-NN) yang dapat dibandingkan untuk efisiensi klasifikasi pada data kesejahteraan sosial. Hingga saat ini belum terdapat studi klasifikasi data PKH dengan konteks data desa terpencil menggunakan kedua algoritma ini secara langsung.

Penelitian ini bertujuan untuk membandingkan dan mengembangkan model klasifikasi berbasis algoritma C4.5 dan K-NN dalam mengolah data penerima bantuan PKH. Dengan menggunakan data historis dan variabel-variabel yang relevan, penelitian ini akan mengevaluasi efektivitas kedua algoritma dalam menentukan status penerima bantuan. Temuan ini diharapkan dapat menjadi panduan bagi pemerintah dalam memilih algoritma yang paling efektif untuk pengelolaan data sosial-ekonomi, sehingga pendistribusian bantuan sosial PKH dapat lebih optimal. Selain itu, penelitian ini juga diharapkan dapat berkontribusi dalam pengembangan sistem informasi berbasis data yang transparan dan akuntabel, guna meningkatkan kepercayaan masyarakat terhadap program bantuan sosial serta memastikan bahwa bantuan diterima oleh keluarga yang benar-benar membutuhkan.

Kemiskinan merujuk pada kondisi kekurangan dan pendapatan yang tidak mencukupi, sehingga individu tidak mampu memenuhi kebutuhan dasarnya, seperti pakaian, makanan, dan tempat tinggal (Bhinadi, 2017). Penduduk yang dikategorikan miskin umumnya memiliki rata-rata pengeluaran bulanan per kapita yang berada di bawah ambang batas kemiskinan. Bappenas (dalam Melati et al., 2021) menyatakan bahwa kemiskinan merupakan permasalahan yang memiliki banyak dimensi, dipengaruhi oleh berbagai faktor yang saling berkaitan dan melibatkan berbagai sektor. Pembangunan ekonomi nasional dapat terhambat karena faktor kemiskinan yang terus meningkat dan tidak terkendali.

Klasifikasi adalah proses pelatihan suatu model atau fungsi terhadap dataset pelatihan, sehingga model tersebut mampu memprediksi kategori dari data uji yang belum dikenali (Azhari et al., 2021). Secara umum, klasifikasi merupakan proses untuk mengenali pola yang dapat menjelaskan dan membedakan kelas-kelas dalam sekumpulan dataset. Klasifikasi sangat berguna untuk memprediksi hasil dengan cara mengelompokkan data ke dalam kategori atau kelompok tertentu berdasarkan atribut atau fitur yang dimiliki. Proses klasifikasi mencakup analisis pola dari data yang ada dan memahami hubungan antara fitur-fitur tersebut dengan kelas tertentu. Setelah model selesai dilatih, model ini dapat digunakan untuk memperkirakan kelas yang paling mungkin dari data baru berdasarkan fitur-fitur yang telah dipelajari sebelumnya (Meiriza et al., 2020).

METODE

Penelitian ini menggunakan data sekunder berupa data penduduk miskin yang bersumber dari Data Terpadu Kesejahteraan Sosial (DTKS) Desa Selat. Data ini terdiri dari beberapa variabel dengan format kategorial. Sebelum melakukan tahap klasifikasi dilakukan tahap pra-pemrosesan data yaitu melakukan proses *encoding* data. Dalam proses *encoding*, data dengan format kategorial akan diubah menjadi format label angka agar proses klasifikasi dengan algoritma C4.5 dan K-NN dapat lebih efektif.

Tahap selanjutnya adalah melakukan pembagian *dataset* ke dalam data *training* dan data *testing*. Data yang akan digunakan dalam klasifikasi pada penelitian ini terdiri dari 538 entri, dengan rasio *splitting data* sebesar 80% untuk data *training* dan 20% untuk data *testing*. Pembagian data ini merupakan bagian dari proses pelatihan model klasifikasi, di mana data *training* digunakan untuk membangun model dan data *testing* digunakan untuk menguji kinerja model terhadap data baru.

Penelitian ini menggunakan dua metode untuk melakukan klasifikasi terhadap data penduduk miskin. Metode yang digunakan yaitu algoritma C4.5 dan algoritma K-Nearest Neighbor (K-NN). Proses klasifikasi dengan C4.5 diawali dengan menghitung nilai *entropy* dan *gain* dari masing-masing variabel untuk membuat aturan pohon keputusan. Aturan dari pohon keputusan yang terbentuk akan digunakan untuk membangun model klasifikasi C4.5 dan dilatih dengan data *training*. Selanjutnya, model tersebut akan digunakan dalam melakukan klasifikasi kelas dari data *testing*.

Proses klasifikasi dengan K-NN diawali dengan menghitung kedekatan antara seluruh data *testing* dan data *training*. Jarak antara data tersebut dihitung menggunakan metode penghitungan jarak *euclidean*. Kemudian, data diurutkan berdasarkan nilai jarak terdekat, lalu dipilih sebanyak *k* data teratas. Klasifikasi kelas dari data *testing* ditentukan berdasarkan mayoritas kelas dari *k* tetangga terdekatnya. Nilai *k* dapat menggunakan akar dari jumlah data yang digunakan. Selain itu, nilai *k* umumnya juga menggunakan bilangan ganjil untuk menghindari hasil imbang dalam penentuan kelas.

Setelah melalui tahapan klasifikasi, hasil klasifikasi dari kedua model akan dievaluasi menggunakan *confusion matrix*. *Confusion matrix* adalah alat yang digunakan untuk menilai kinerja model klasifikasi. Pada dasarnya, *confusion matrix* menunjukkan perbedaan antara hasil klasifikasi yang dilakukan oleh model dengan target klasifikasi yang seharusnya (Hadianto et al., 2019).

Tabel 1. Confusion Matrix

Kelas	Klasifikasi	
	Positif	Negatif
Nilai Aktual	Positif	TP
	Negatif	FP

TP merupakan data positif yang benar diklasifikasikan, TN merupakan data negatif yang benar diklasifikasikan, FP merupakan data negatif yang salah diklasifikasikan sebagai positif, dan FN merupakan data positif yang salah diklasifikasikan sebagai negatif. Tingkat efektivitas dari model klasifikasi dapat dihitung dengan metrik berikut ini.

1. Akurasi, untuk menghitung data yang berhasil diklasifikasikan dengan benar.

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

2. Presisi, untuk mengitung data yang berhasil diambil dan relevan dengan informasi yang dibutuhkan.

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

3. *Recall*, untuk menghitung data relevan yang diambil dari seluruh data yang relevan terhadap query.

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

Penelitian ini memiliki tujuan untuk membandingkan kedua model yang diperoleh dengan algoritma C4.5 dan algoritma K-NN. Perbandingan ini dilakukan untuk mendapatkan model yang lebih efektif dalam melakukan klasifikasi terhadap data penduduk miskin. Model dengan tingkat efektivitas lebih tinggi akan digunakan dalam menentukan calon penerima bantuan Program Keluarga Harapan (PKH) di Desa Selat.

Algoritma C4.5 merupakan kelompok algoritma *decision tree* yang dikembangkan dari algoritma ID3 oleh Ross Quinlan. Beberapa peningkatan pada algoritma ini mencakup kemampuan dalam menangani data yang hilang (*missing value*), memproses atribut kontinu, serta melakukan pemangkasan (*pruning*). Algoritma ini menggunakan dua jenis *input* utama yaitu data pelatihan dan data pengujian. Data pelatihan terdiri atas contoh-contoh data yang telah diverifikasi dan digunakan untuk membentuk pohon keputusan, sementara data pengujian berisi atribut-atribut yang digunakan untuk mengklasifikasikan data baru (Atma & Setyanto, 2018). Pada algoritma ini menggunakan persamaan-persamaan berikut untuk membangun pohon keputusan (Kusrini & Luthfi, 2009).

1. Menghitung nilai *entropy*.

$$Entropy(S) = - \sum_{i=1}^n p_i \times \log_2 p_i$$

S merupakan himpunan kasus, n merupakan jumlah partisi dari S , dan p_i merupakan proporsi dari S_i terhadap S .

2. Menghitung nilai *gain*.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} * Entropy(S_i)$$

A merupakan atribut, n merupakan jumlah partisi atribut A , $|S_i|$ merupakan jumlah kasus pada partisi ke- i , dan $|S|$ merupakan jumlah kasus dalam S .

Algoritma K-Nearest Neighbor

Algoritma K-Nearest Neighbor (K-NN) merupakan teknik klasifikasi yang menggunakan teknik pengelompokan data baru berdasarkan jaraknya dengan sejumlah data yang sudah ada. Proses untuk mendapatkan kelas kategori dalam algoritma ini meliputi penghitungan kemiripan antara data baru dengan seluruh data pelatihan yang telah dikategorikan. Data latih tersebut kemudian diurutkan berdasarkan tingkat kemiripan tertinggi, lalu dipilih sebanyak k data teratas. Kategori data baru ditentukan berdasarkan mayoritas kelas dari k tetangga terdekat tersebut. Nilai k pada umumnya mempergunakan bilangan ganjil agar tidak ada hasil yang imbang pada pemilihan *class* (Atma & Setyanto, 2018). Algoritma ini merupakan salah satu bagian dari algoritma *supervised learning*, hal ini karena dalam proses klasifikasinya menggunakan data pembelajaran (*train*) yang sudah memiliki label kelas sebagai acuan untuk memprediksi kelas data uji (Yandi Saputra & Primadasa, 2018). Jarak antar data baru dengan data latih dapat ditentukan dengan metode pengukuran jarak seperti jarak *Euclidean*. Jarak *Euclidean* dapat dihitung menggunakan persamaan berikut (Kusrini & Luthfi, 2009).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$d(x, y)$ merupakan jarak antara data x dengan data y , x_i merupakan nilai data x pada variabel ke- i , y_i merupakan nilai data y pada variabel ke- i , dan n merupakan jumlah atribut.

HASIL DAN PEMBAHASAN

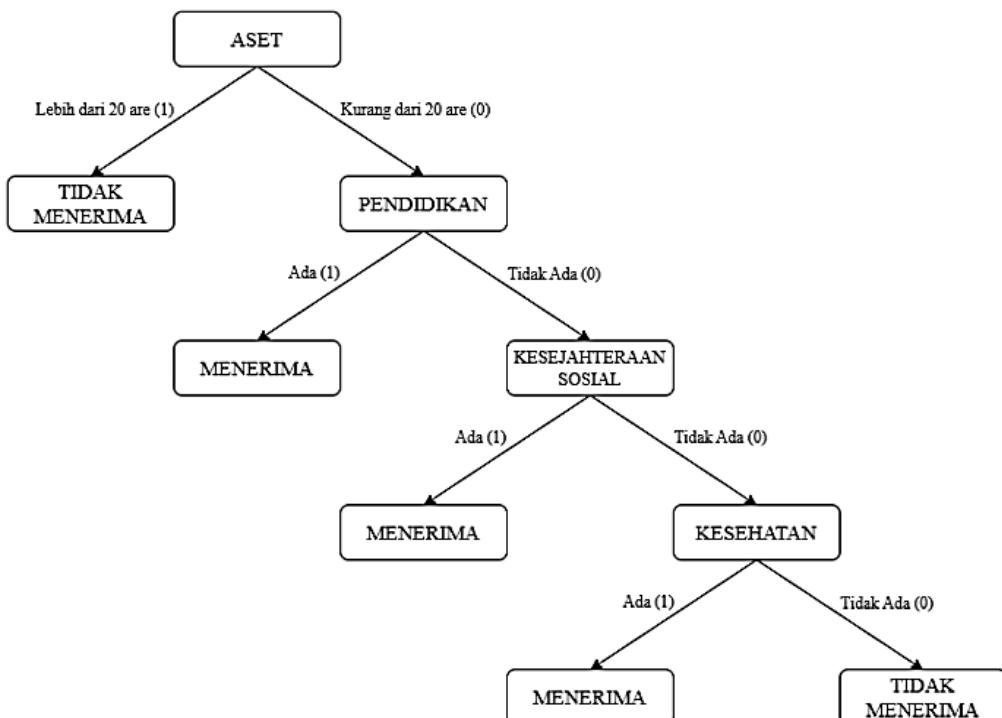
Penelitian ini menggunakan 538 entri data dengan enam variabel dan satu kelas target. Variabel-variabel yang digunakan dalam proses klasifikasi calon penerima PKH pada penelitian ini dapat dilihat pada tabel berikut.

Tabel 2. Variabel Klasifikasi

Variabel	Keterangan
X_1 Pekerjaan	1 : Buruh 2 : Pedagang 3 : Petani
X_2 Kesehatan	0 : Tidak Ada 1 : Ada
X_3 Pendidikan	0 : Tidak Ada 1 : Ada
X_4 Kesejahteraan Sosial	0 : Tidak Ada 1 : Ada
X_5 Status Rumah	0 : Bukan Hak Milik 1 : Hak Milik
X_6 Aset	0 : Kurang dari 20 are 1 : Lebih dari 20 are
Y Status PKH	0 : Tidak Menerima 1 : Menerima

Hasil Klasifikasi C4.5

Proses klasifikasi dengan metode C4.5 dilakukan dengan menghitung nilai *entropy* dan *gain* dari masing-masing variabel. Variabel yang memiliki nilai *gain* terbesar akan menjadi akar dari pohon keputusan. Perhitungan nilai *entropy* dan *gain* dilakukan beberapa kali dan akan berhenti jika: (a) semua kasus dalam node n mendapat kelas yang sama, (b) tidak ada atribut di dalam kasus yang dipartisi lagi, dan (c) tidak ada kasus di dalam cabang yang kosong.



Gambar 1. Pohon Keputusan C4.5

Gambar di atas merupakan pohon keputusan yang didapat dari proses algoritma C4.5. Aturan yang terbentuk dari pohon keputusan tersebut akan digunakan dalam melakukan klasifikasi terhadap data *testing*. Berikut merupakan tabel *confusion matrix* dari hasil klasifikasi metode C4.5.

Tabel 3. Confusion Matrix Metode C4.5

Kelas	Klasifikasi	
	Menerima	Tidak Menerima
Nilai Aktual	Menerima	64
	Tidak Menerima	0
		44

Tabel di atas menunjukkan bahwa seluruh data penduduk yang menerima PKH dengan total 64 data dapat diklasifikasikan dengan benar oleh model C4.5 yang dibangun. Selain itu, seluruh penduduk yang seharusnya tidak menerima PKH dengan total 44 data juga dapat terdeteksi dengan baik oleh model.

Pada Hasil Klasifikasi C4.5, penelitian ini menunjukkan bahwa algoritma C4.5 berhasil menghasilkan model yang sangat akurat, dengan tingkat akurasi, presisi, dan recall yang mencapai 100%. Keberhasilan ini menegaskan keunggulan C4.5 dalam mengelola data sosial yang kompleks, seperti data Program Keluarga Harapan (PKH), yang melibatkan atribut kategorikal dan numerik. C4.5, sebagai pengembangan dari algoritma ID3, bekerja dengan memilih atribut yang memberikan gain informasi terbesar di setiap node pohon keputusan, yang memungkinkan pembuatan aturan klasifikasi yang jelas dan transparan. Interpretabilitas yang ditawarkan oleh C4.5 sangat penting dalam konteks pengambilan keputusan sosial, di mana pemangku kepentingan, seperti pemerintah atau lembaga sosial, perlu memahami dengan jelas dasar dari setiap keputusan klasifikasi (Nofriani, 2020). Dengan menggunakan pohon keputusan, C4.5 memberikan gambaran visual yang memudahkan pemangku kepentingan untuk memahami kriteria yang digunakan dalam menentukan status penerima bantuan.

Selain itu, penelitian ini menunjukkan bahwa C4.5 mampu mengurangi dimensi data dari enam atribut menjadi empat atribut, tanpa mengorbankan akurasi. Teknik pengurangan dimensi yang digunakan ini mengurangi kompleksitas model, menghilangkan atribut yang tidak relevan, dan menghasilkan model yang lebih efisien. Hal ini sangat penting ketika mengelola data sosial yang sering kali memiliki banyak atribut yang saling tumpang tindih atau redundant. Dalam konteks ini, penelitian oleh Ulya et al. (2021) menunjukkan bahwa pengurangan dimensi yang tepat dapat memperbaiki kinerja algoritma klasifikasi dengan menghilangkan noise dan fitur yang tidak penting. Selain itu, kemampuan C4.5 dalam menangani data yang hilang atau tidak lengkap juga menjadi keunggulan penting, seperti yang dijelaskan oleh Nofriani (2020), yang menyatakan bahwa C4.5 dapat memproses data sosial dengan baik meskipun ada informasi yang hilang, tanpa harus menghapus data atau mengorbankan hasil klasifikasi.

Keunggulan C4.5 lainnya adalah kemampuan untuk mengklasifikasikan data dengan struktur yang lebih jelas dan terdefinisi, yang cocok dengan karakteristik dataset sosial yang biasanya memiliki batasan yang cukup tegas dalam hal pengelompokan data berdasarkan indikator sosial-ekonomi seperti pekerjaan, pendidikan, dan kesejahteraan sosial. Nofriani (2020) juga menyebutkan bahwa C4.5 seringkali memberikan performa yang lebih baik dalam kasus data yang memiliki batasan kelas yang jelas dibandingkan dengan algoritma lain yang mengandalkan kedekatan seperti K-NN. Dengan demikian, C4.5 terbukti lebih efektif dalam mengelola data sosial untuk pengambilan keputusan yang lebih transparan dan akurat.

Selain itu, kemampuan C4.5 untuk menangani atribut kontinu dan kategorikal juga memperkuat alasan mengapa algoritma ini lebih unggul dalam pengklasifikasian data yang melibatkan variabel-variabel seperti kepemilikan rumah, kesehatan, dan aset, yang bisa berupa data numerik maupun kategorikal. Pengurangan dimensi yang dilakukan oleh C4.5 juga sejalan dengan penelitian oleh Dananjaya et al. (2019), yang menunjukkan bahwa pengurangan dimensi dapat mengurangi overfitting pada data dengan banyak fitur yang tidak relevan, meningkatkan efisiensi dan kinerja model secara keseluruhan.

Hasil Klasifikasi K-Nearest Neighbor

Proses klasifikasi dengan metode K-NN diawali dengan menghitung jarak *euclidean* masing-masing data *testing* dari data *training*. Kemudian, hasil perhitungan jarak tersebut akan diurutkan dari nilai jarak terkecil hingga terbesar. Klasifikasi kelas data *testing* dilakukan dengan menghitung mayoritas kelas data *training* dari data dengan jarak *euclidean* terkecil

hingga sebanyak k tetangga terdekat. Nilai k pada penelitian ini menggunakan hasil akar dari jumlah data yang digunakan, yaitu sebanyak 21 data.

Tabel 4. Mayoritas Kelas Klasifikasi

Data ke-n	Menerima	Tidak Menerima	Aktual	Klasifikasi
1	21	0	1	1
2	21	0	1	1
3	21	0	1	1
4	0	21	0	0
5	0	21	0	0
6	20	1	1	1
7	12	9	0	1
8	20	1	1	1
9	21	0	1	1
10	0	21	0	0

Berdasarkan tabel di atas, dapat ditentukan kelas dari masing-masing data *testing*. Misalnya, pada data ke-1 diperoleh mayoritas kelas data *training* adalah menerima sebanyak 21 data, maka data *testing* ke-1 diklasifikasikan ke dalam kelas menerima (1). Berikut merupakan tabel *confusion matrix* dari hasil klasifikasi metode K-NN.

Tabel 5. Confusion Matrix Metode K-NN

Kelas	Klasifikasi	
	Menerima	Tidak Menerima
Nilai Aktual	Menerima	64
	Tidak Menerima	1
		43

Tabel di atas menunjukkan bahwa seluruh data penduduk yang menerima PKH dengan total 64 data dapat diklasifikasikan dengan benar oleh model yang dibangun. Namun, data penduduk yang seharusnya tidak menerima PKH dengan total 44 data hanya diklasifikasikan secara benar sebanyak 43 data, dan 1 data sisanya diklasifikasikan sebagai menerima. Sehingga, dapat diketahui bahwa terdapat 1 data yang tidak diklasifikasikan dengan benar oleh model K-NN yang dibangun.

Meskipun algoritma K-NN juga memberikan hasil yang cukup baik, dengan tingkat recall 100%, akurasi yang diperoleh hanya sebesar 99,07% dan presisi sebesar 98,46%. Seperti yang terlihat pada *Confusion Matrix* (Tabel 5), K-NN mampu mengklasifikasikan dengan benar mayoritas data yang menerima bantuan PKH, namun terdapat satu kesalahan dalam mengklasifikasikan data yang tidak berhak menerima bantuan. Salah satu alasan mengapa K-NN tidak mencapai hasil yang sempurna dapat dijelaskan melalui pemilihan nilai k .

Pemilihan nilai k adalah faktor krusial dalam kinerja K-NN. Berdasarkan penelitian oleh Ulya et al. (2021) dan Khoiruzzaman et al. (2021), nilai k yang terlalu kecil cenderung menyebabkan overfitting, di mana model menjadi sensitif terhadap noise dalam data, sementara nilai k yang terlalu besar dapat mengurangi akurasi karena hilangnya informasi lokal. Oleh karena itu, pemilihan nilai k yang optimal sangat penting, dan pengujian yang tepat perlu dilakukan untuk menentukan nilai k yang memberikan performa terbaik pada dataset spesifik. Pada penelitian ini, nilai k yang digunakan adalah 21, yang didasarkan pada akar jumlah data. Namun, hasil dari Wibowo et al. (2024) menunjukkan bahwa nilai k yang lebih kecil, seperti 3 atau 5, dapat memberikan hasil yang lebih baik dalam hal akurasi, karena dapat lebih sensitif terhadap pola lokal dalam data. Oleh karena itu, pemilihan nilai k yang tepat menjadi penting dalam meningkatkan kinerja model K-NN pada klasifikasi status penerima bantuan PKH.

Perbandingan Efektivitas C4.5 dan K-NN

Berdasarkan hasil klasifikasi yang telah dilakukan menggunakan kedua metode di atas, dapat dilakukan perbandingan efektivitas dengan membandingkan persentase akurasi, presisi, dan *recall* dari masing-masing metode. Tabel perbandingan efektivitas dari metode C4.5 dan metode K-Nearest Neighbor dapat dilihat sebagai berikut.

Tabel 6. Perbandingan Efektivitas C4.5 dan K-NN

Metode	Akurasi	Presisi	Recall
C4.5	100%	100%	100%
K-NN	99,07%	98,46%	100%

Berdasarkan tabel tersebut, dapat diketahui bahwa model dengan algoritma C4.5 memperoleh tingkat akurasi, presisi, dan *recall* yang sempurna yaitu sebesar 100%. Walaupun model dengan algoritma K-NN memperoleh nilai *recall* yang sama dengan model C4.5, namun tingkat akurasinya hanya sebesar 99,07% dan presisi yang diperoleh sebesar 98,46%. Hasil perbandingan ini menunjukkan bahwa model yang dibangun dengan algoritma C4.5 lebih efektif daripada model dengan algoritma K-NN dalam klasifikasi calon penerima PKH di Desa Selat.

Dalam perbandingan antara kedua algoritma, C4.5 terbukti lebih unggul dalam hal kecepatan klasifikasi dan interpretabilitas. C4.5 memberikan pemahaman yang jelas tentang aturan yang digunakan untuk klasifikasi, yang sangat penting dalam program sosial seperti PKH, di mana transparansi dan kejelasan adalah aspek yang krusial. Selain itu, pengurangan dimensi yang dilakukan oleh C4.5 memungkinkan penggunaan atribut yang lebih relevan, sehingga meningkatkan efisiensi model tanpa mengorbankan akurasi.

Sebaliknya, K-NN menawarkan kemudahan dalam implementasi, namun biaya komputasi yang lebih tinggi menjadi kendala, terutama dengan dataset yang besar dan ketika diperlukan untuk melakukan pengklasifikasian secara real-time. Seperti yang dicatat oleh Nofriani (2020), K-NN dapat menjadi kurang efisien dalam aplikasi waktunya karena memerlukan perhitungan jarak terhadap seluruh dataset pelatihan untuk setiap prediksi baru. Hal ini mengurangi efektivitasnya dalam pengambilan keputusan cepat yang dibutuhkan dalam program sosial.

Tabel 6 memperlihatkan perbandingan yang jelas antara kedua algoritma. Meskipun K-NN berhasil mencapai nilai *recall* yang sama dengan C4.5, nilai akurasi dan presisi C4.5 lebih tinggi, yang menjadikannya pilihan yang lebih tepat dalam konteks klasifikasi penerima bantuan PKH, di mana kecepatan klasifikasi dan akurasi tinggi sangat dibutuhkan untuk memastikan bantuan diterima oleh keluarga yang tepat.

KESIMPULAN

Penelitian ini membandingkan dua algoritma klasifikasi yaitu C4.5 dan K-Nearest Neighbor yang digunakan untuk mengklasifikasikan status penerima bantuan Program Keluarga Harapan di Desa Selat. Hasil penerapan algoritma C4.5 mendapatkan nilai akurasi, presisi, dan *recall* sebesar 100%. Selain itu, algoritma C4.5 juga dapat mereduksi variabel yang digunakan dalam proses klasifikasi dari 6 menjadi 4 variabel. Sedangkan, hasil penerapan algoritma K-NN dengan $k = 21$ hanya mendapatkan akurasi sebesar 99,07% dan presisi sebesar 98,46%, serta mendapatkan nilai *recall* sebesar 100%. Hasil pengujian menunjukkan bahwa model klasifikasi dengan algoritma C4.5 memiliki nilai akurasi dan presisi yang lebih tinggi, sehingga model dengan algoritma C4.5 lebih efektif daripada model dengan algoritma K-Nearest Neighbor dalam melakukan klasifikasi status calon penerima bantuan Program Keluarga Harapan di Desa Selat.

REKOMENDASI

Perlu dilakukan penelitian lanjutan dengan menggunakan variabel-variabel yang lebih kompleks serta mempertimbangkan penggabungan beberapa metode klasifikasi. Hal ini penting untuk menguji konsistensi model terhadap data yang lebih beragam dan dinamis. Dengan mengeksplorasi dan membandingkan metode-metode lain secara lebih mendalam, diharapkan dapat ditemukan pendekatan yang tidak hanya akurat tetapi juga lebih adaptif dalam memprediksi calon penerima bantuan PKH secara tepat sasaran. Hasil penelitian ini dapat menjadi referensi yang berharga bagi pihak-pihak terkait dalam mengembangkan langkah-langkah mengentaskan kemiskinan dan mengurangi penerimaan yang tidak tepat sasaran.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini. Terima kasih khusus kami sampaikan kepada Universitas Pendidikan Ganesha yang telah memberikan fasilitas dan sumber daya untuk penelitian ini. Kami juga mengucapkan terima kasih kepada rekan-rekan sejawat dan para ahli yang telah memberikan saran dan masukan berharga yang sangat membantu dalam menyempurnakan penelitian ini. Kami juga ingin mengucapkan terima kasih kepada para pemangku kepentingan yang telah membantu dalam memberikan wawasan terkait implementasi Program Keluarga Harapan (PKH) di Desa Selat.

DAFTAR PUSTAKA

- Anggraeni, A. P., & Nugroho, A. A. (2022). Evaluasi Kebijakan PKH (Program Keluarga Harapan) Di Indonesia. *Journal of Public Policy and Applied Administration*, 4(2), 39–54.
- Atma, Y. D., & Setyanto, A. (2018). Perbandingan Algoritma C4.5 dan K-NN dalam Identifikasi Mahasiswa Berpotensi Drop Out. *Metik Jurnal*, 2(2), 31–37.
- Bhatia, N. (2010). Survey of Nearest Neighbor Techniques. *International Journal of Advanced Computer Science and Applications*, 8(2).
- Bhinadi, A. (2017). *Penanggulangan Kemiskinan dan Pemberdayaan Masyarakat: Studi Kasus Daerah Istimewa Yogyakarta*. In DEEPUBLISH Publisher CV Budi Utama (hal. 89–92).
- Dananjaya, D., Werdiningsih, I., & Semiati, R. (2019). Decision support system for classification of early childhood diseases using principal component analysis and k-nearest neighbors classifier. *Journal of Information Systems Engineering and Business Intelligence*, 5(1), 13. <https://doi.org/10.20473/jisebi.5.1.13-22>
- Hadianto, N., Novitasari, H. B., & Rahmawati, A. (2019). Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode Neural Network. *Jurnal Pilar Nusa Mandiri*, 15(2), 163–170.
- Kementerian Sosial Republik Indonesia. (2021). Pedoman Pelaksanaan Program Keluarga Harapan (PKH).
- Khoiruzzaman, N., Ramadhan, R., & Junaidi, A. (2021). Hasil klasifikasi algoritma backpropagation dan k-nearest neighbor pada cardiovascular disease. *Journal of Dinda Data Science Information Technology and Data Analytics*, 1(1), 17-27. <https://doi.org/10.20895/dinda.v1i1.141>
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Buku Algoritma Data Mining (T. A. Prabawati (ed.)). CV Andi Offset (Penerbit Andi).
- Melati, A. M., Sudrajat, & Burhany, D. I. (2021). Pengaruh Belanja Pendidikan, Belanja Kesehatan dan Belanja Bantuan Sosial terhadap Kemiskinan pada Kabupaten dan Kota di Provinsi Jawa Barat. *Indonesian Accounting Research Journal*, 1(3), 422–430.
- Nofriani, N. (2020). Machine learning application for classification prediction of household's welfare status. *Jitce (Journal of Information Technology and Computer Engineering)*, 4(02), 72-82. <https://doi.org/10.25077/jitce.4.02.72-82.2020>
- Puspita, D., Aminah, S., & Arif, A. (2022). Prediction System for Credit Eligibility Using C4.5 Algorithm. *Journal of Informatics and Telecommunication Engineering*, 6(1), 148–156.
- Siska, F., & Heni, S. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 2(4), 89–95.
- Suryahadi, A., Al Izzati, R., & Suryadarma, D. (2020). The Impact of COVID-19 Outbreak on Poverty: An Estimation for Indonesia. *In SMERU Working Paper*. Vol. April.
- Ulya, S., Soeelman, M., & Budiman, F. (2021). Optimasi parameter k pada algoritma k-nn untuk klasifikasi prioritas bantuan pembangunan desa. *Techno Com*, 20(1), 83-96. <https://doi.org/10.33633/tc.v20i1.4215>
- Wang, Y., Guo, C., Xiao, C., & Yang, W. (2022). Combining imputation method and feature weighting algorithms to improve the classification accuracy of incomplete data. *Journal*

- of Physics Conference Series, 2171(1), 012038. <https://doi.org/10.1088/1742-6596/2171/1/012038>
- Wibowo, A., Kasih, P., & Farida, I. (2024). Sistem bantu penentuan konsentrasi mahasiswa menggunakan metode k-nearest neighbor classification. *Stains*, 3(1), 370-379. <https://doi.org/10.29407/stains.v3i1.4343>
- Yandi Saputra, A., & Primadasa, Y. (2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour Implementation of Classification Method to Predict Student Graduation Using K-Nearest Neighbor Algorithm. *Techno.Com*, 17(4), 9.