

Perbandingan Algoritma K-Nearest Neighbour dan C4.5 Decision Tree Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan

Kariyamin¹, La Ode Alyandi^{1*}, Samsul Arif¹

¹Program Studi Teknologi Informasi, Institut Teknologi dan Bisnis Muhammadiyah Wakatobi, Indonesia

Artikel Info

Kata Kunci:

C4.5
KNN
PKH
Rapid Minner

Keywords:

C4.5
KNN
PKH
Rapid Minner

Riwayat Artikel:

Submitted: 03 September 2025
Accepted: 27 Oktober 2025
Published: 29 Oktober 2025

Abstrak: Program Keluarga Harapan (PKH) merupakan bantuan sosial bersyarat yang ditujukan bagi keluarga dengan kondisi ekonomi rentan. Tantangan utama dalam implementasinya adalah memastikan ketepatan sasaran penerima bantuan. Penelitian ini bertujuan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan C4.5 dalam mengklasifikasikan kelayakan penerima PKH. Data penelitian diambil dari Dinas Sosial Kabupaten Wakatobi tahun 2023, kemudian melalui tahap praproses, pembagian data menggunakan K-Fold Cross Validation (K=7), serta pelatihan dan pengujian model dengan perangkat lunak RapidMiner. Evaluasi dilakukan berdasarkan akurasi, presisi, dan recall. Hasil analisis menunjukkan bahwa algoritma C4.5 memberikan kinerja lebih baik dengan akurasi 96,22%, presisi "YES" 95,71%, dan recall "YES" 94,89%. Sebaliknya, KNN memperoleh akurasi 81,44%, presisi "YES" 69,66%, dan recall "YES" 95,74%. Temuan ini menunjukkan bahwa C4.5 lebih efektif dan konsisten dalam mengklasifikasikan penerima bantuan dibandingkan KNN. Model pohon keputusan yang dihasilkan juga memberikan kejelasan logis terhadap faktor-faktor penentu kelayakan. Penelitian ini berpotensi dikembangkan melalui integrasi model ke dalam sistem berbasis web atau aplikasi untuk meningkatkan efisiensi proses seleksi penerima bantuan.

Abstract: The Family Hope Program (PKH) is a conditional cash assistance program aimed at households with vulnerable economic conditions. One of its main challenges lies in ensuring accurate targeting of beneficiaries. This study aims to compare the performance of the K-Nearest Neighbor (KNN) and C4.5 algorithms in classifying PKH beneficiary eligibility. Data were obtained from the Wakatobi District Social Service in 2023, then processed through data preprocessing, partitioning using K-Fold Cross Validation (K=7), and model training and testing with RapidMiner software. Evaluation was conducted based on accuracy, precision, and recall. The results indicate that the C4.5 algorithm outperformed KNN, achieving an accuracy of 96.22%, "YES" precision of 95.71%, and "YES" recall of 94.89%. In comparison, KNN achieved an accuracy of 81.44%, "YES" precision of 69.66%, and "YES" recall of 95.74%. These findings highlight that C4.5 is more effective and consistent in classifying beneficiaries than KNN. The generated decision tree model also provides a logical interpretation of the eligibility-determining factors. This research has potential for further development by integrating the model into web-based or mobile

applications to improve the efficiency of beneficiary selection processes.

Corresponding Author:

La Ode Alyandi

Email: laodealyandi@gmail.com

PENDAHULUAN

Dalam upaya menekan angka kemiskinan dan meningkatkan taraf hidup masyarakat, pemerintah Indonesia menggulirkan Program Keluarga Harapan (PKH), sebuah bantuan tunai bersyarat yang ditujukan kepada keluarga dengan kondisi sosial ekonomi paling rentan (Nuraeni et al., 2024). Program ini tidak hanya bertujuan memberikan dukungan finansial semata, tetapi juga mendorong peningkatan akses terhadap layanan pendidikan, kesehatan, dan kesejahteraan sosial (Mangunsong et al., 2024). Meski telah menunjukkan dampak positif dalam menurunkan tingkat kemiskinan, keberhasilan implementasi PKH masih menghadapi tantangan besar, terutama dalam memastikan bahwa bantuan benar-benar diterima oleh mereka yang paling membutuhkan (Wardhani et al., 2025). Di sinilah proses klasifikasi penerima bantuan menjadi sangat krusial, karena kesalahan dalam penetapan sasaran dapat menghambat efektivitas program secara keseluruhan (Hidayati et al., 2020).

Di berbagai daerah, termasuk Kabupaten Wakatobi, pelaksanaan PKH kerap menemui kendala, baik dari segi ketepatan sasaran maupun kondisi geografis yang menyulitkan proses pendataan. Tak jarang ditemukan kasus di mana keluarga tidak layak justru menerima bantuan, sementara yang seharusnya berhak justru terabaikan (Utami et al., 2023). Menghadapi tantangan ini, pemanfaatan teknologi informasi dan pendekatan berbasis data menjadi semakin penting. Salah satu solusi yang kini banyak dikembangkan adalah penggunaan algoritma pembelajaran mesin (machine learning), khususnya dalam proses klasifikasi penerima bantuan sosial (Daud et al., 2025).

Data mining, sebagai cabang ilmu komputer yang fokus pada analisis data berskala besar, menyediakan berbagai metode klasifikasi yang mampu meningkatkan akurasi penetapan penerima bantuan. Dua algoritma yang sering digunakan dalam konteks ini adalah K-Nearest Neighbor (KNN) dan C4.5 (Zaeni et al., 2020). KNN bekerja dengan menghitung jarak antar data untuk menentukan klasifikasi, sedangkan C4.5 membentuk pohon keputusan berdasarkan atribut-atribut penting dalam dataset dan menghasilkan model berbasis aturan yang mudah dipahami (Arowolo et al., 2021). Keduanya menawarkan pendekatan berbeda, dan membandingkan kinerja masing-masing menjadi penting untuk memilih metode yang paling efektif dalam konteks PKH (Ma & Zhai, 2023).

Berdasarkan penelitian sebelumnya, membandingkan kinerja tiga algoritma klasifikasi Naïve Bayes Classifier (NBC), K-Nearest Neighbor (K-NN), dan C4.5 pada data penerima Program Keluarga Harapan (PKH) menunjukkan bahwa NBC memperoleh recall tertinggi meskipun akurasi (77,51%) tidak menjadi yang paling besar. Algoritma C4.5 menghasilkan akurasi tertinggi sebesar 80,16% dan mampu mereduksi jumlah atribut dari 33 menjadi 8, sehingga mengurangi kompleksitas model. Sementara itu, K-NN dengan nilai K terbaik (K=3) mencatat akurasi 76,72%. Hasil ini menegaskan bahwa setiap algoritma memiliki keunggulan berbeda, di mana NBC lebih efektif meminimalkan kesalahan pengklasifikasian keluarga yang layak menerima bantuan, sedangkan C4.5 unggul dalam akurasi dan efisiensi atribut (Dina et al., 2023).

Penelitian terdahulu lainnya melakukan penelitian tentang komparasi antara algoritma C4.5, Naïve Bayes, dan K-Nearest Neighbor (K-NN) dalam mengklasifikasikan penerima Program Keluarga Harapan (PKH) menunjukkan hasil yang bervariasi. Algoritma C4.5 memperoleh akurasi tertinggi, diikuti oleh Naïve Bayes, sedangkan K-NN berada di posisi terakhir. Temuan ini mengindikasikan bahwa C4.5 lebih optimal untuk kasus klasifikasi data penerima bantuan sosial, sementara Naïve Bayes masih memberikan hasil yang kompetitif dengan keunggulan pada kesederhanaan pemodelan.

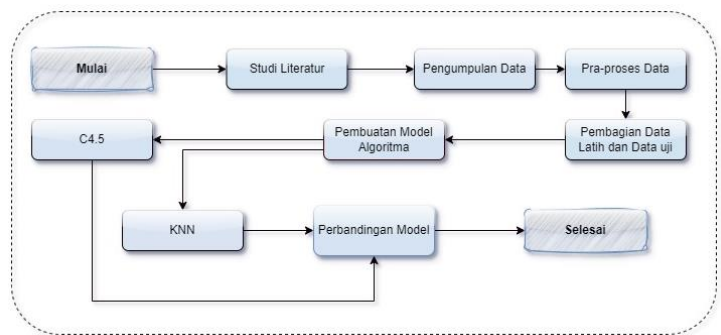
Perbedaan kinerja ini memperlihatkan bahwa pemilihan algoritma yang tepat sangat bergantung pada karakteristik data dan tujuan analisis (Fitriani, 2020).

Dalam penelitian ini, digunakan data dari Dinas Sosial Kabupaten Wakatobi untuk menguji kinerja kedua algoritma tersebut. Proses pelatihan dan pengujian model dilakukan dengan metode K-Fold Cross Validation agar hasil prediksi lebih akurat dan tidak bias. Evaluasi performa dilakukan dengan mengukur akurasi, presisi, dan recall dari masing-masing algoritma. Dengan pendekatan ini, diharapkan proses klasifikasi penerima PKH dapat dilakukan secara objektif, efisien, dan tepat sasaran, sehingga program bantuan benar-benar menyentuh keluarga yang paling membutuhkan.

Berdasarkan latar belakang dan rumusan masalah yang telah dijelaskan sebelumnya, maka tujuan dari penelitian ini adalah (1). Menganalisis kinerja algoritma K-Nearest Neighbor (KNN) dalam mengklasifikasikan penerima bantuan Program Keluarga Harapan (PKH), (2). Menganalisis kinerja algoritma C4.5 dalam mengklasifikasikan penerima bantuan Program Keluarga Harapan (PKH), (3). Membandingkan hasil kinerja kedua algoritma tersebut untuk menentukan metode yang paling efektif dalam klasifikasi penerima bantuan PKH.

METODE

Berdasarkan penelitian terdahulu, tahapan atau proses yang akan dilakukan pada penelitian untuk mencari nilai perbandingan antara algoritma KNN dan C4.5, yang akan menjadi objek perbandingan adalah Recall, precision, dan akurasi. Berikut diagram alir penelitian dilakukan.



Gambar 1. Alur Penelitian

Studi Literatur

Pencarian literatur mengarahkan penelitian menjadi kumpulan literatur yang diperiksa untuk menemukan referensi yang relevan dengan topik penelitian. Referensi tersebut diambil dari buku atau e-book, publikasi jurnal, maupun media daring yang berkaitan guna mendukung proses penelitian (Deviyanto & Wahyudi, 2018). Fokus penelitian mencakup algoritma K-Nearest Neighbor, C4.5, klasifikasi, serta studi sejenis yang telah dilakukan dalam penelitian sebelumnya (Rachma, 2022) (Halder et al., 2024).

Pengumpulan Data

Data bersumber dari Dinas Sosial Kabupaten Wakatobi, Wangi-Wangi Selatan, Yaitu data tahun 2023. Pengumpulan data juga dilakukan dengan cara wawancara dan pengumpulan dataset dari pihak pengurus PKH di kantor Dinas Sosial (Anitalia et al., 2024).

Praproses Data

Pada tahap praproses data, dilakukan proses pembersihan. Pembersihan data melibatkan penghapusan baris yang memiliki nilai Null, menggantikannya dengan nilai rata-rata yang dilakukan dengan menggunakan operator Filter Example pada RapidMiner. Dengan beberapa proses yang akan menggunakan Google colab, seperti menghitung jumlah atribut dan tipe data atribut serta hal lain sebagainya yang akan menggunakan Google Colab (Romadloni & Septiyanti, 2023).

Pembagian Data Latih Dan Data Uji

Data latih dan data uji dibagi menggunakan metode K-Fold Cross Validation dengan nilai K-fold sebesar 7. Proses ini dilakukan menggunakan operator Cross Validation di RapidMiner. Metode Cross Validation berguna untuk memperkirakan kesalahan dalam memprediksi kinerja model. Teknik ini membagi seluruh dataset menjadi data training dan data testing (Sutanty et al., 2025).

Pembuatan Model Algoritma

Dalam algoritma K-NN, dilakukan percobaan dengan mencoba berbagai nilai parameter K. Nilai K yang dicoba meliputi K = 5. Operator yang digunakan adalah operator K-NN pada kolom data training, serta operator apply model dan performance pada kolom data testing di RapidMiner. Parameter yang dipilih untuk measure types adalah mixed measure, yang digunakan untuk menghitung jarak pada atribut numerik. Untuk tipe mixed Euclidean Distance (Briouza et al., 2022).

Dalam algoritma C4.5, digunakan operator Decision Tree pada kolom data training serta operator apply model dan performance pada kolom data testing di RapidMiner. Parameter yang diterapkan mencakup apply pruning dan apply prepruning. Penggunaan parameter pruning dan prepruning ini dapat meningkatkan akurasi hasil (Elisyah, 2022).

Perbandingan Model

Setiap model klasifikasi dievaluasi menggunakan confusion matrix untuk mengukur performanya. Selain itu, dihitung juga nilai akurasi, presisi, dan Recall. Proses ini dilakukan dengan menggunakan operator performance untuk menampilkan hasil evaluasi kinerja algoritma di RapidMiner (Es-sabery et al., 2024). Setelah itu, dilakukan perbandingan nilai akurasi, presisi, dan Recall dari kedua model algoritma tersebut (K-NN dan C4.5) (Permana et al., 2021). Rumus untuk menghitung nilai akurasi dapat ditemukan pada persamaan (1), sementara rumus untuk menghitung nilai presisi dan Recall masing-masing dapat dilihat pada persamaan (2) dan persamaan (3).

$$akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

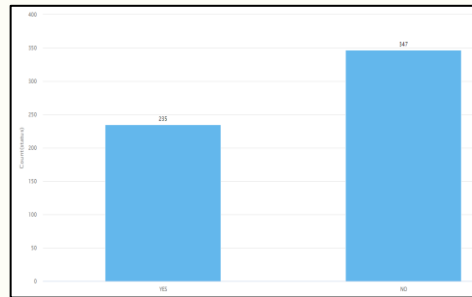
$$presisi = \frac{TP}{TP+FP} \times 100\% \quad (2)$$

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (3)$$

HASIL DAN PEMBAHASAN

Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan melalui dua langkah utama. Pertama, dilakukan sesi wawancara bersama tenaga pendamping sosial PKH dan perwakilan dari Dinas Sosial Kabupaten Wakatobi. Tujuannya adalah untuk mendapatkan gambaran menyeluruh mengenai proses penentuan penerima bantuan, termasuk kriteria dan variabel yang digunakan dalam seleksi. Setelah wawancara memberikan pemahaman yang cukup, tahap selanjutnya adalah menghimpun data penerima PKH. Dataset yang diperoleh memuat berbagai informasi sosial dan ekonomi yang dianggap relevan dalam menilai kelayakan seseorang untuk menerima bantuan. Visualisasi dari data yang berhasil dikumpulkan dapat dilihat pada Gambar 2.

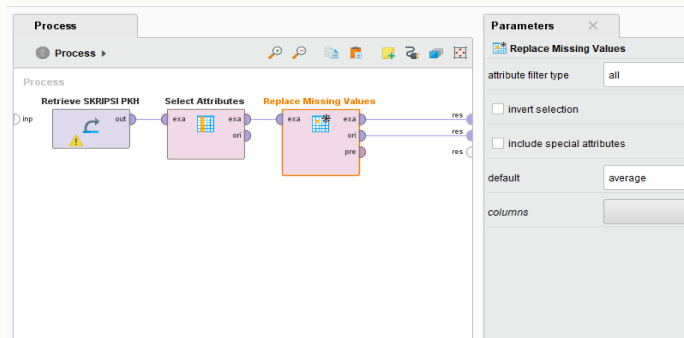


Gambar 2. Data Hasil Pengumpulan

Gambar 2 menampilkan diagram batang yang menggambarkan sebaran data sebanyak 582 entri berdasarkan dua kategori pada variabel status, yaitu "YES" dan "NO". Dari visualisasi tersebut, dapat dilihat bahwa data berstatus "NO" mendominasi dengan jumlah 347, sedangkan kategori "YES" tercatat sebanyak 235 data. Setiap data memiliki 33 atribut yang digunakan sebagai fitur, serta satu label utama, yaitu status penerimaan bantuan.

Praproses Data

Pada tahap pra-pemrosesan data, dilakukan penanganan terhadap nilai-nilai yang hilang menggunakan teknik imputasi dengan pendekatan rata-rata. Metode ini digunakan agar struktur serta karakteristik asli dari data tetap terjaga. Di samping itu, atribut yang dianggap tidak berpengaruh terhadap proses klasifikasi, seperti kolom "NO" yang hanya berfungsi sebagai penanda urutan, dihapus dari dataset karena tidak memberikan nilai informasi dalam analisis. Proses pengisian nilai kosong menggunakan rata-rata ditampilkan pada Gambar 3.



Gambar 3. Pengisian Nilai Kosong

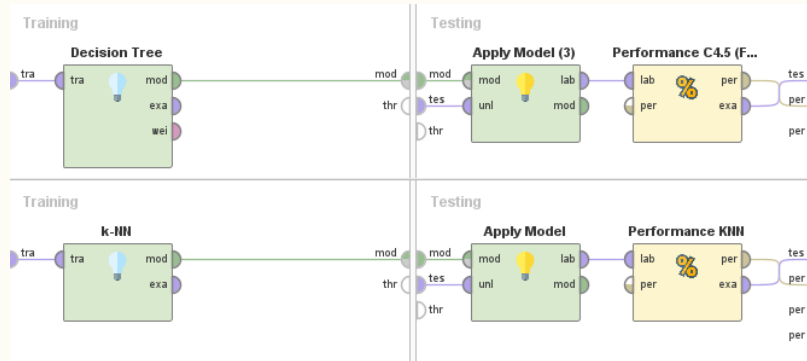
Gambar 3 memperlihatkan tahapan dalam proses mengganti nilai yang kosong dengan nilai rata-rata. Langkah pertama dilakukan dengan menggunakan fungsi select attributes untuk menyaring atribut yang memiliki data hilang. Setelah atribut yang mengandung nilai kosong berhasil diidentifikasi, proses dilanjutkan menggunakan fungsi replace missing values. Pada tahap ini, parameter yang digunakan adalah average, yang berfungsi menggantikan nilai kosong dengan rata-rata dari masing-masing atribut tersebut.

Pembagian Data Latih Dan Data Uji

Penelitian ini menerapkan metode evaluasi K-Fold Cross Validation dengan nilai K sebesar 7. Dengan metode ini, data dibagi menjadi tujuh bagian yang seimbang, di mana setiap bagian secara bergiliran digunakan sebagai data pengujian, sementara enam bagian lainnya berperan sebagai data pelatihan. Proses ini dilakukan secara berulang hingga seluruh subset telah berkesempatan menjadi data uji. Pendekatan ini memastikan bahwa seluruh data dimanfaatkan secara merata, sehingga hasil evaluasi model menjadi lebih objektif dan tidak bergantung pada satu pola pembagian tertentu. Teknik ini sangat berguna untuk menilai seberapa baik model dapat menggeneralisasi atau memprediksi data yang belum pernah dilihat sebelumnya.

Pembuatan Model

Setelah data melewati tahap pra-pemrosesan, langkah selanjutnya adalah membangun model klasifikasi dengan menggunakan tiga algoritma yang cukup dikenal, yaitu K-Nearest Neighbor (KNN) dan C4.5. Algoritma KNN bekerja dengan mengklasifikasikan data berdasarkan kedekatannya terhadap data lain yang labelnya sudah diketahui. Sementara itu, C4.5 menyusun model dalam bentuk pohon keputusan dengan mempertimbangkan seberapa besar kontribusi masing-masing atribut dalam membedakan kelas. Kedua metode ini digunakan untuk dibandingkan tingkat akurasi guna mengetahui mana yang paling efektif dalam melakukan klasifikasi pada dataset yang tersedia. Ilustrasi dari proses ini dapat dilihat pada Gambar 4.



Gambar 4. Pembuatan Model

Hasil Pembuatan Model

1. KNN

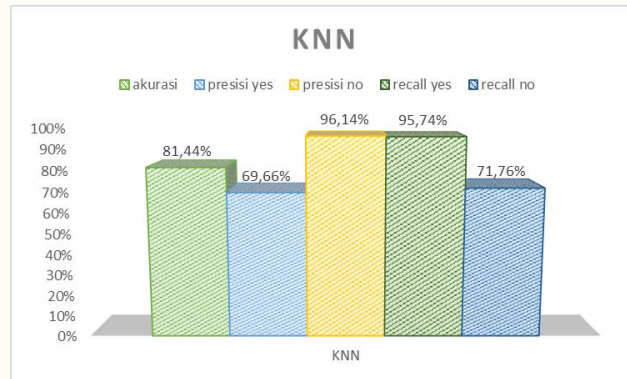
Evaluasi terhadap performa algoritma KNN dalam perangkat lunak RapidMiner difokuskan pada dua kelas, yaitu "YES" dan "NO". Penilaian dilakukan menggunakan confusion matrix serta metrik evaluasi seperti akurasi, presisi, dan recall. Evaluasi ini bertujuan untuk mengukur sejauh mana model mampu membedakan antara kedua kategori tersebut secara akurat. Hasil analisis kinerja dari algoritma KNN ditampilkan pada Gambar 5.

accuracy: 81.44% +/- 5.70% (micro average: 81.44%)			
	true YES	true NO	class precision
pred. YES	225	98	69.66%
pred. NO	10	249	96.14%
class recall	95.74%	71.76%	

Gambar 5. Hasil KNN

Algoritma KNN menunjukkan tingkat akurasi sebesar 81,44%. Presisi untuk kelas "NO" tergolong tinggi, mencapai 96,14%, sedangkan untuk kelas "YES" nilainya lebih rendah, yaitu 69,66%. Sementara itu, nilai recall pada kelas "YES" sangat baik, yakni 95,74%, yang mengindikasikan kemampuan model dalam mengenali data positif cukup kuat. Sebaliknya, recall untuk kelas "NO" masih belum optimal dan memiliki ruang untuk peningkatan.

Temuan ini mengisyaratkan bahwa model cenderung lebih responsif terhadap data penerima bantuan ("YES"), namun kurang sensitif terhadap data non-penerima ("NO"). Dalam konteks sosial PKH, hal ini dapat diartikan bahwa model KNN memiliki potensi untuk lebih cepat mengenali keluarga miskin yang layak menerima bantuan, tetapi masih berisiko mengelompokkan sebagian keluarga yang sebenarnya tidak layak sebagai penerima. Hal ini penting bagi Dinas Sosial karena dapat memengaruhi keadilan distribusi bantuan. Visualisasi dari hasil ini disajikan pada Gambar 6.



Gambar 6. Visualisasi KNN

2. C4.5

Evaluasi terhadap algoritma C4.5, yang dikenal sebagai salah satu metode decision tree paling populer, dilakukan dalam konteks klasifikasi biner antara kelas "YES" dan "NO". Proses evaluasi memanfaatkan metode K-Fold Cross Validation, disertai dengan penerapan Feature Selection (FS) untuk mengoptimalkan kinerja model. Hasil evaluasi disajikan dalam bentuk confusion matrix, serta metrik lain seperti akurasi, presisi, dan recall untuk masing-masing kelas. Visualisasi hasil evaluasi tersebut dapat dilihat pada Gambar 7.

accuracy: 96.22% +/- 2.02% (micro average: 96.22%)			
	true YES	true NO	class precision
pred. YES	223	10	95.71%
pred. NO	12	337	96.56%
class recall	94.89%	97.12%	

Gambar 7. Hasil C4.5

Model C4.5 pada Gambar 7 menunjukkan performa klasifikasi yang sangat baik, dengan tingkat akurasi mencapai 96,22% \pm 2,02%. Dari hasil evaluasi, sebanyak 223 data dengan label "YES" dan 337 data "NO" berhasil diklasifikasikan dengan tepat. Sementara itu, kesalahan klasifikasi terjadi hanya pada 12 data berlabel "YES" dan 10 data berlabel "NO". Tingkat presisi masing-masing kelas tercatat sebesar 95,71% untuk "YES" dan 96,56% untuk "NO", sedangkan recall-nya adalah 94,89% untuk "YES" dan 97,12% untuk "NO".

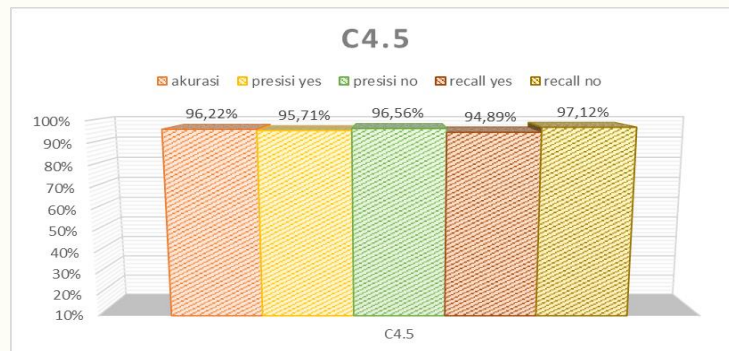
Angka-angka ini menunjukkan bahwa model memiliki kemampuan yang seimbang dalam mengenali kedua kelas dengan akurasi tinggi. Dalam konteks sosial PKH, performa ini menandakan bahwa C4.5 mampu memberikan rekomendasi yang lebih objektif dan konsisten dalam menentukan calon penerima bantuan. Hal ini sangat penting dalam praktik penyaluran bantuan sosial, di mana kesalahan dalam pengambilan keputusan dapat berdampak langsung pada keluarga miskin yang seharusnya menerima bantuan namun terlewat, atau sebaliknya, keluarga yang tidak layak namun terdata sebagai penerima.

Struktur pohon keputusan hasil dari algoritma C4.5 dapat dilihat pada Gambar 8, yang telah melalui proses feature selection untuk memastikan hanya atribut-atribut paling relevan yang digunakan dalam menentukan kelayakan penerima bantuan.



Gambar 8. Pohon Keputusan

Pohon keputusan pada Gambar 8 memperlihatkan bahwa proses klasifikasi mempertimbangkan kepemilikan aset rumah tangga sebagai salah satu indikator utama kondisi ekonomi. Pola ini mencerminkan kondisi nyata masyarakat Wakatobi, di mana jumlah dan jenis aset seperti alat elektronik, kendaraan, atau perabotan sering kali menjadi ukuran ekonomi keluarga. Semakin sedikit jumlah aset yang dimiliki, semakin besar peluang dikategorikan sebagai penerima bantuan. Dengan demikian, model ini tidak hanya menunjukkan performa komputasional yang tinggi, tetapi juga memiliki relevansi sosial dalam menggambarkan realitas kemiskinan lokal. Hasil dari algoritma C4.5 divisualisasikan secara lebih rinci pada Gambar 9.

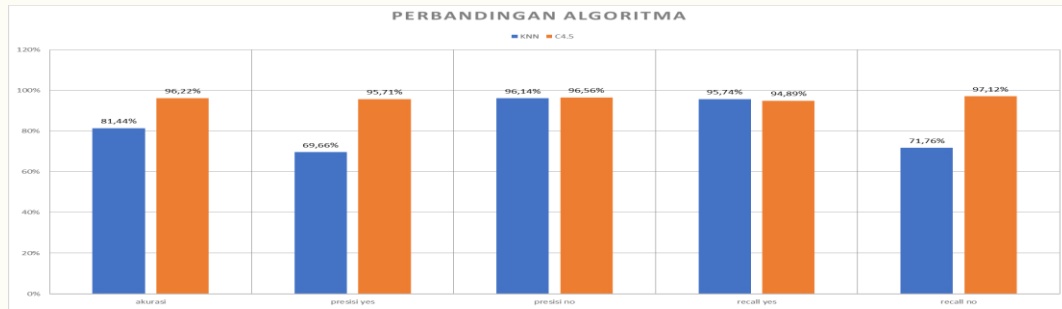


Gambar 9. Hasil C4.5

Perbandingan Algoritma

Berdasarkan hasil evaluasi, algoritma C4.5 menunjukkan performa yang lebih unggul dibandingkan KNN. C4.5 mencatatkan akurasi sebesar 96,22%, sementara KNN menghasilkan 81,44%. Perbedaan ini menunjukkan bahwa C4.5 lebih efektif dan stabil dalam mengklasifikasikan data penerima bantuan PKH.

Secara sosial, hal ini mengindikasikan bahwa penerapan algoritma C4.5 lebih mampu membantu proses seleksi calon penerima bantuan secara adil dan terukur, sehingga dapat mendukung transparansi serta efisiensi kebijakan penyaluran bantuan sosial oleh pemerintah daerah. Perbandingan visual antara kedua algoritma tersebut ditampilkan pada Gambar 10.



Gambar 10. Perbandingan Hasil Algoritma

Untuk memberikan gambaran yang lebih jelas mengenai performa masing-masing algoritma, berikut disajikan tabel perbandingan hasil evaluasi berdasarkan metrik akurasi, presisi, dan recall. Tabel ini bertujuan untuk memperkuat analisis yang dapat dijadikan acuan dalam menilai efektivitas masing-masing metode klasifikasi. Dapat dilihat pada Tabel 1.

Tabel 1 Hasil Perbandingan

Algoritma	KNN	C4.5
Akurasi	81,44%	96,22%
Presisi Yes	69,66%	95,71%
Presisi No	96,14%	96,56%
Recall Yes	95,74%	94,89%
Recall no	71,76%	97,12%

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma C4.5 memiliki kinerja yang lebih unggul dibandingkan K-Nearest Neighbor (KNN) dalam melakukan klasifikasi penerima bantuan Program Keluarga Harapan (PKH). Dengan tingkat akurasi mencapai 96,22%, algoritma C4.5 terbukti mampu menghasilkan klasifikasi yang lebih seimbang antara kategori “layak” dan “tidak layak”, serta memberikan interpretasi yang jelas melalui struktur pohon keputusan. Keunggulan ini menjadikan C4.5 sebagai metode yang tepat untuk membantu proses seleksi penerima bantuan secara lebih objektif dan transparan, karena setiap keputusan didasarkan pada atribut-atribut yang relevan seperti kepemilikan aset, kondisi ekonomi, dan faktor rumah tangga lainnya. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, dataset yang digunakan masih terbatas yaitu 582 entri saja pada wilayah Kabupaten Wakatobi, sehingga hasil model belum tentu sepenuhnya mewakili kondisi daerah lain dengan karakteristik sosial ekonomi berbeda. Kedua, faktor-faktor sosial seperti partisipasi masyarakat, dinamika keluarga, atau kebijakan lokal belum sepenuhnya dimasukkan ke dalam variabel penelitian, sehingga potensi pengaruhnya terhadap hasil klasifikasi belum dapat dievaluasi secara menyeluruh.

Untuk penelitian selanjutnya, disarankan agar dilakukan pengumpulan data dengan cakupan wilayah yang lebih luas serta penambahan variabel sosial dan ekonomi lain yang dapat mempengaruhi penentuan kelayakan penerima bantuan. Selain itu, model C4.5 dapat dikembangkan lebih lanjut melalui kombinasi dengan teknik machine learning lain atau integrasi ke dalam sistem digital berbasis aplikasi/web agar hasilnya dapat diterapkan langsung dalam proses pendataan dan penyaluran bantuan sosial secara efektif dan efisien.

DAFTAR PUSTAKA

- Anitalia, W. O. S., Yusuf, B., & Azizah, L. N. (2024). Peran Pendamping Program Keluarga Harapan (Pkh) Dalam Kegiatan Pertemuan Peningkatan Kemampuan Keluarga Penerima Manfaat Di Kelurahan Wanci Kabupaten Wakatobi. *Welvaart: Jurnal Ilmu Kesejahteraan Sosial*, 5(1), 109–116. <https://doi.org/10.52423/welvaart.v5i1.11>
- Arowolo, M. O., Adebisi, M. O., Adebisi, A. A., & Olugbara, O. (2021). Optimized Hybrid Investigative Based Dimensionality Reduction Methods For Malaria Vector Using KNN Classifier. *Journal of Big Data*, 8(29), 1-14. <https://doi.org/10.1186/s40537-021-00415-z>
- Briouza, S., Gritli, H., Khraief, N., Belghith, S., & Singh, D. (2022). EMG Signal Classification for Human Hand Rehabilitation via Two Machine Learning Techniques: kNN and SVM. *Proceedings of the 2022 5th International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2022*, 412–417. https://doi.org/10.1109/IC_ASET53395.2022.9765856
- Daud, M., Juita, R., & Suhendra, C. D. (2025). Penerapan Metode Algoritma C4.5 Untuk Klasifikasi Kelayakan Penerima Program Bantuan Pada Dinas Sosial Kabupaten Manokwari. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(1), 271–278. <https://doi.org/http://dx.doi.org/10.51454/decode.v5i1.1057>
- Deviyanto, A., & Wahyudi, M. D. R. (2018). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 3(1), 1-13. <https://doi.org/10.14421/jiska.2018.31-01>
- Dina, A., Permana, I., Muttakin, F., & Maita, I. (2023). Perbandingan Algoritma NBC, KNN, dan C4.5 Untuk Klasifikasi Penerima Bantuan Program Keluarga Harapan. *Jurnal Media Informatika Budidarma*, 7(3), 1079–1087. <https://doi.org/10.30865/mib.v7i3.6316>
- Elisyah, E. (2022). Klasifikasi Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Algoritma C4.5 (Studi Kasus: Kecamatan Kupang Timur). http://skripsi.undana.ac.id/index.php?p=show_detail&id=12441&keywords=
- Es-sabery, F., Es-sabery, I., Qadir, J., Sainz-de-Abajo, B., & Garcia-Zapirain, B. (2024). A Hybrid Hadoop-Based Sentiment Analysis Classifier For Tweets Associated With COVID-19 Utilizing Two Machine Learning Algorithms: CNN, And Fuzzy C4.5. *Journal of Big Data*, 11(176), 1-55. <https://doi.org/10.1186/s40537-024-01014-4>
- Fitriani, E. (2020). Perbandingan Algoritma C4.5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *Sistemasi*, 9(1), 103-115. <https://doi.org/10.32520/stmsi.v9i1.596>
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-Nearest Neighbor Algorithm: A Comprehensive Review And Performance Analysis Of Modifications. *Journal of Big Data*, 11(113), 1-55. <https://doi.org/10.1186/s40537-024-00973-y>
- Hidayati, D. C., Al Faraby, S., & Adiwijaya, A. (2020). Klasifikasi Topik Multi Label pada Hadis Shahih Bukhari Menggunakan K-Nearest Neighbor dan Latent Semantic Analysis. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 140-147. <https://doi.org/10.30865/jurikom.v7i1.2013>
- Ma, S., & Zhai, J. (2023). BigData Decision Tree For Continuous-Valued Attributes Based On Unbalanced Cut Points. *Journal of Big Data*, 10(135), 1-22. <https://doi.org/10.1186/s40537-023-00816-2>
- Mangunsong, U. S., Syah, A. Z., & Saputra, E. (2024). Implementasi Sistem Seleksi Calon Penerima Subsidi Program Keluarga Harapan Dengan Metode SAW Di Kelurahan Sijambi. *Decode: Jurnal Pendidikan Teknologi Informasi*, 4(3), 737–747. <https://doi.org/10.51454/decode.v4i3.559>
- Nuraeni, S., Harliana, H., & Prabowo, T. (2024). Analisis Akurasi Naïve Bayes Dan Knn Dalam Penentuan Penerima Pkh Di Lombok Utara. *Journal of Information System Management (JOISM)*,

5(2), 121–126. <https://doi.org/10.24076/joism.2024v5i2.1205>

- Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178–188. <https://doi.org/10.14421/jiska.2021.6.3.178-188>
- Rachma, C. A. (2022). Implementasi Algoritma K-Nearest Neighbor Dalam Penentuan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi Jawa Timur. <http://etheses.uin-malang.ac.id/38201/1/18610077.pdf>
- Romadloni, N. T., & Septiyanti, N. D. (2023). Optimasi Feature Selection Pada Komentar Media Sosial Terhadap Peralihan Tv Digital Menggunakan Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 151–160. <https://doi.org/10.51454/decode.v3i2.121>
- Sutanty, E., Setiyaningsih, E., & Yanti, S. N. (2025). Optimasi Hyperparameter Pada Algoritma K-Nearest Neighbor Untuk Analisis Sentimen Terhadap Pembelajaran Jarak Jauh. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(2), 348–364. <https://doi.org/10.51454/decode.v5i2.1170>
- Utami, D. K., Irawati, N., & Sumantri, S. (2023). Analysis of the k-Means Method in Clustering Acceptance of PKH Aid in Pulau Rakyat Tua Village. *Jurnal Sistemasi*, 12(3), 953–961. <https://doi.org/10.32520/stmsi.v12i3.3236>
- Wardhani, I. K., Triyanto, W. A., & Supriyono. (2025). Penerapan Metode Weighted Product dan Simple Multi Attribute Rating Technique (SMART) untuk Seleksi Penerima Program Keluarga Harapan (PKH) Application of the Weighted Product Method and Simple Multi-Attribute Rating Technique (SMART) for Selecting. *Sistemasi: Jurnal Sistem Informasi*, 14(2), 905–916. <https://doi.org/10.32520/stmsi.v14i2.5080>
- Zaeni, I. A. E., Anzani, D. R., Putra, D. S., Devi, M., Hidayati, L., Sudjono, I. (2020). Classifying the parental involvement on school from home during covid-19 using c4.5 algorithm. *Proceedings of the 2022 5th International Conference on Advanced Systems and Emergent Technologies, IC_ASET 2022*, 253–257. https://doi.org/10.1109/IC_ASET53395.2022.9765856