

Classification Of Social Assistance Eligibility In Talang Benuang Village Using C4.5 And Naive Bayes

Felly Intania ¹; Yupianti ²; Eko Suryana ³

¹*Study Program of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu*

^{2,3} *Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu*

Email: ¹fellyintania18@gmail.com

How to Cite :

Intania, F., Yupianti, Suryana, E. (2026). Classification Of Social Assistance Eligibility In Talang Benuang Village Using C4.5 And Naive Bayes Jurnal Media Computer Science, 5(4)

ARTICLE HISTORY

Received [05 September 2026]

Revised [23 Oktober 2026]

Accepted [26 Oktober 2026]

KEYWORDS

BPNT, C4.5, Naive Bayes, Klasifikasi, Confusion Matrix.

This is an open access article under the [CC-BY-SA](#) license



ABSTRAK

Penentuan penerima Bantuan Pangan Non Tunai (BPNT) di Desa Talang Benuang masih dilakukan secara manual sehingga berisiko menimbulkan ketidaktepatan sasaran. Penelitian ini bertujuan membangun dan membandingkan model C4.5 dan Naive Bayes untuk mengklasifikasikan kelayakan penerima BPNT. Pendekatan berbasis data mengikuti CRISP-DM dengan 100 data publik berlabel sebagai data pemodelan dan 50 data warga desa sebagai data penerapan. Tahapan analisis meliputi pembersihan data, transformasi ordinal, pembentukan pipeline, penyetelan hiperparameter C4.5 melalui GridSearchCV, serta evaluasi 5-fold cross-validation menggunakan confusion matrix dan ROC-AUC. Kedua algoritma mencapai accuracy 93,00%, precision 100,00%, recall 85,42%, dan F1-score 92,13%. C4.5 dipilih sebagai model akhir karena menghasilkan aturan keputusan yang mudah ditafsirkan dan AUC 0,986, lebih tinggi daripada Naive Bayes sebesar 0,868. Penerapan pada 50 kepala keluarga menghasilkan 42 prediksi layak dan 8 tidak layak. Temuan ini menunjukkan bahwa C4.5 dapat mendukung penyaringan awal penerima BPNT secara objektif, tetapi keputusan administratif tetap memerlukan verifikasi musyawarah desa dan data acuan resmi.

ABSTRACT

The selection of Non-Cash Food Assistance (BPNT) recipients in Talang Benuang Village is still performed manually, creating a risk of mistargeted distribution. This study develops and compares C4.5 and Naive Bayes models for classifying BPNT eligibility. A data-driven approach followed CRISP-DM, using 100 labeled public records for modeling and 50 village household records for application. The analysis comprised data cleaning, ordinal transformation, pipeline construction, C4.5 hyperparameter tuning with GridSearchCV, and 5-fold cross-validation evaluated through a confusion matrix and ROC-AUC. Both algorithms achieved 93.00% accuracy, 100.00% precision, 85.42% recall, and a 92.13% F1-score. C4.5 was selected as the final model because it produced interpretable decision rules and an AUC of 0.986, exceeding the Naive Bayes AUC of 0.868. Application to 50 households yielded 42 eligible and 8 ineligible predictions. These findings indicate that C4.5 can support objective preliminary screening of BPNT recipients; however, administrative decisions still require verification through village deliberation and official ground-truth records.

PENDAHULUAN

Pemanfaatan teknologi informasi dalam pelayanan publik memungkinkan pengelolaan data dilakukan secara lebih cepat, terstruktur, dan dapat ditelusuri. Dalam penyaluran bantuan sosial, kualitas pengelolaan data menentukan apakah bantuan menjangkau rumah tangga yang benar-benar memenuhi kriteria. Transformasi digital pada layanan pemerintah tidak hanya berkaitan dengan pemindahan dokumen ke media elektronik, tetapi juga dengan penggunaan data untuk memperkuat konsistensi keputusan dan mengurangi kesalahan administratif (Aziz et al., 2021). Kebutuhan tersebut relevan bagi

Bantuan Pangan Non Tunai (BPNT), karena proses pendataan calon penerima melibatkan kondisi sosial ekonomi yang beragam dan harus diperbarui secara berkala.

Desa Talang Benuang, Kabupaten Seluma, rutin melakukan pendataan dan pengusulan penerima BPNT. Jumlah penduduk meningkat dari 1.717 jiwa pada 2024 menjadi 1.749 jiwa pada 2025, sedangkan jumlah kepala keluarga bertambah dari 520 menjadi 529. Peningkatan ini memperbesar beban pemeriksaan data. Hasil observasi dan wawancara awal menunjukkan bahwa analisis kelayakan masih dilakukan secara manual dan belum didukung klasifikasi terkomputerisasi. Kondisi tersebut berpotensi menghasilkan ketidaktepatan sasaran, baik dalam bentuk rumah tangga layak yang terlewat maupun rumah tangga kurang layak yang tetap masuk dalam daftar usulan.

Teknik klasifikasi dalam data mining dapat membentuk pola keputusan dari atribut sosial ekonomi. C4.5 menghasilkan pohon keputusan yang transparan, sedangkan Naive Bayes menawarkan proses probabilistik yang sederhana dan efisien. Susanto et al. (2026) menunjukkan bahwa C4.5 dapat membentuk aturan kelayakan bantuan sosial yang mudah dipahami. Rifat dan Akhriza (2025) melaporkan efektivitas Naive Bayes pada klasifikasi penerima Program Keluarga Harapan, sementara Attamami et al. (2023) membandingkan kedua algoritma pada bantuan jaminan kesehatan. Studi-studi tersebut menegaskan kelayakan kedua metode, tetapi konteks program, tingkat pemerintahan, komposisi atribut, dan karakteristik data berbeda dengan BPNT di tingkat desa.

Celah penelitian terletak pada belum tersedianya perbandingan langsung C4.5 dan Naive Bayes untuk BPNT di Desa Talang Benuang dengan prosedur evaluasi yang sama. Penelitian ini juga tidak berhenti pada nilai accuracy, tetapi memeriksa precision, recall, F1-score, ROC-AUC, struktur kesalahan, tingkat kepentingan fitur, dan keterbacaan aturan. Kebaruan praktisnya adalah penyatuan pipeline transformasi data, penyetelan hiperparameter, validasi silang, dan penerapan model pada data warga menjadi mekanisme penyaringan awal yang dapat diaudit.

Penelitian ini bertujuan membangun model klasifikasi kelayakan penerima BPNT menggunakan C4.5 dan Naive Bayes, membandingkan kinerja keduanya melalui 5-fold cross-validation dan confusion matrix, menentukan model yang paling sesuai, serta menerapkannya pada data warga Desa Talang Benuang. Hasilnya diharapkan menjadi dukungan analitis bagi pemerintah desa tanpa menggantikan verifikasi administratif dan musyawarah penetapan penerima.

LANDASAN TEORI

Klasifikasi dan Kelayakan Bantuan Sosial

Data mining merupakan proses memperoleh pola dan pengetahuan yang berguna dari kumpulan data untuk mendukung pengambilan keputusan (Syam et al., 2024). Salah satu tugas utamanya adalah klasifikasi, yaitu pembentukan model dari data berlabel untuk memetakan observasi baru ke kelas tertentu (Setiawan et al., 2024). Dalam penelitian ini kelas target terdiri atas “Layak” dan “Tidak Layak”, sedangkan atribut prediktor berasal dari kondisi sosial ekonomi rumah tangga. Model berfungsi sebagai alat bantu penyaringan awal; status penerima yang sah tetap ditentukan melalui prosedur pemerintah dan bukti lapangan.

Algoritma C4.5 dan Naive Bayes

C4.5 membangun pohon keputusan dengan memilih pemisah yang menurunkan ketidakpastian kelas. Implementasi menggunakan DecisionTreeClassifier ber kriteria entropy. Atribut dan ambang dengan information gain tertinggi ditempatkan pada simpul, kemudian pemisahan berlanjut secara rekursif sampai data cukup homogen atau batas kompleksitas tercapai. Kelebihan utamanya adalah hasil berbentuk aturan jika–maka yang dapat ditelusuri oleh pengguna nonteknis. Pembatasan kedalaman pohon dan jumlah minimum sampel diperlukan untuk mencegah model terlalu menyesuaikan data latih.

Naive Bayes menghitung probabilitas posterior kelas berdasarkan probabilitas prior dan likelihood setiap atribut. Penelitian ini memakai CategoricalNB karena prediktor bersifat kategorikal ordinal. Laplace smoothing dengan $\alpha = 1,0$ mencegah probabilitas nol ketika kategori tertentu tidak muncul pada salah satu kelas. Asumsi independensi antarprediktor menyederhanakan komputasi, tetapi dapat mengurangi kemampuan model apabila atribut saling berkaitan kuat. Perbandingan dengan C4.5 penting karena kedua algoritma mewakili pendekatan struktur keputusan dan probabilistik.

Evaluasi Model

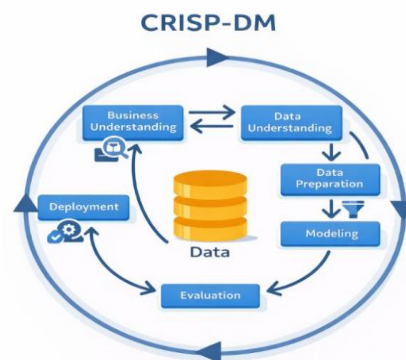
Kinerja model diuji dengan 5-fold cross-validation agar seluruh data berlabel memperoleh giliran sebagai data validasi. Confusion matrix memisahkan true positive, true negative, false positive, dan false

negative. Accuracy menunjukkan proporsi prediksi benar; precision menunjukkan ketepatan prediksi kelas Layak; recall mengukur kemampuan menemukan seluruh data yang benar-benar Layak; dan F1-score menyeimbangkan precision dengan recall. ROC-AUC digunakan untuk menilai kemampuan pemisahan kelas pada berbagai ambang keputusan. Dalam konteks bantuan sosial, false negative perlu mendapat perhatian karena merepresentasikan calon penerima layak yang berpotensi tidak lolos penyaringan.

METODE PENELITIAN

Penelitian menggunakan desain kuantitatif komparatif berbasis data dengan kerangka Cross-Industry Standard Process for Data Mining (CRISP-DM). Tahapannya meliputi pemahaman kebutuhan penentuan BPNT, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan hasil. Analisis dijalankan menggunakan Python pada Google Colab dengan pandas dan numpy untuk pengolahan data, scikit-learn untuk transformasi serta pemodelan, dan matplotlib serta seaborn untuk visualisasi.

Gambar 1. Tahapan penelitian berbasis CRISP-DM



Sumber data terdiri atas 100 rekaman publik berlabel sebagai data pemodelan dan 50 rekaman warga Desa Talang Benuang sebagai data penerapan. Data warga belum memiliki label ground truth digital pada saat penelitian. Dataset awal memuat nomor urut, pendapatan, tanggungan, status tempat tinggal, kendaraan, dan keputusan. Nomor urut dihapus karena hanya berfungsi sebagai identitas. Status tempat tinggal juga dihapus karena bernilai konstan pada seluruh data sehingga tidak memberikan daya pembeda. Tiga prediktor yang dipertahankan adalah Pendapatan, Tanggungan, dan Kendaraan.

Tabel 1. Ringkasan dataset penelitian

Sumber data	Peran	Jumlah	Layak	Tidak Layak
Dataset publik	Pemodelan dan validasi	100	48	52
Data warga Desa Talang Benuang	Penerapan model	50	Belum berlabel	Belum berlabel

Pemeriksaan kualitas data mencakup missing value, duplikasi, tipe data, distribusi kelas, dan tabulasi silang antara prediktor dengan keputusan. Tidak ditemukan missing value pada 100 data berlabel. Distribusi kelas relatif seimbang, yaitu 48 Layak dan 52 Tidak Layak. Prediktor kategorikal kemudian diubah dengan OrdinalEncoder: Pendapatan Rendah=0, Sedang=1, Tinggi=2; Tanggungan Cukup=0, Banyak=1; dan Kendaraan Cukup=0, Banyak=1. Encoder dan model disatukan dalam pipeline untuk mencegah perbedaan transformasi antara proses pelatihan dan penerapan.

Model C4.5 dibangun menggunakan DecisionTreeClassifier dengan criterion entropy. GridSearchCV menguji 36 kombinasi max_depth (3, 5, 7, atau tanpa batas), min_samples_split (2, 5,

atau 10), dan min_samples_leaf (1, 2, atau 4). Setiap kombinasi dinilai melalui 5-fold cross-validation sehingga terjadi 180 proses fit-predict. Model Naive Bayes menggunakan CategoricalNB dengan Laplace smoothing alpha=1,0. Kedua model dinilai pada pembagian validasi yang sama menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan ROC-AUC.

Tabel 2. Ruang pencarian dan parameter terbaik C4.5

Parameter	Kandidat	Terbaik
max_depth	3, 5, 7, None	3
min_samples_split	2, 5, 10	2
min_samples_leaf	1, 2, 4	1

Model akhir dipilih berdasarkan akurasi validasi silang, kemampuan pemisahan kelas, dan interpretabilitas. Jika akurasi sama, ROC-AUC dan transparansi aturan digunakan sebagai pertimbangan. Setelah dipilih, model dilatih pada seluruh 100 data berlabel dan diterapkan pada 50 data warga. Hasil prediksi disimpan dalam berkas Excel untuk diperiksa kembali oleh pemerintah desa. Pengujian fungsional dilakukan pada tahapan unggah data, validasi struktur, preprocessing, training, evaluasi, visualisasi, prediksi, dan ekspor keluaran.

HASIL DAN PEMBAHASAN

Hasil

Karakteristik Data dan Pembentukan Model

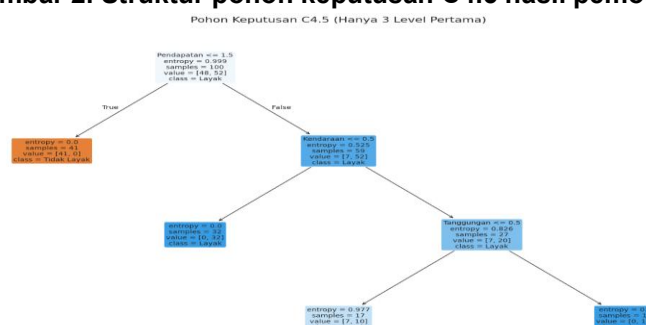
Tabulasi silang memperlihatkan pemisahan yang kuat pada atribut Pendapatan. Seluruh 21 data berpendapatan rendah dan 20 data berpendapatan sedang berlabel Layak. Pada kelompok berpendapatan tinggi, 52 dari 59 data berlabel Tidak Layak dan 7 data berlabel Layak. Pola Tanggungan dan Kendaraan lebih bercampur. Perhitungan information gain mengonfirmasi Pendapatan sebagai atribut paling informatif dengan gain 0,6888, jauh di atas Tanggungan 0,0596 dan Kendaraan 0,0538. Karena itu, Pendapatan menjadi simpul akar pohon keputusan.

Tabel 3. Distribusi keputusan menurut pendapatan

Pendapatan	Jumlah	Layak	Tidak Layak
Rendah	21	21	0
Sedang	20	20	0
Tinggi	59	7	52

GridSearchCV memilih max_depth=3, min_samples_split=2, dan min_samples_leaf=1 dengan rerata akurasi 0,93. Kedalaman tiga menghasilkan struktur ringkas dengan empat aturan terminal. Kelompok pendapatan rendah atau sedang langsung diklasifikasikan Layak. Pada kelompok pendapatan tinggi, keputusan dilanjutkan melalui Kendaraan dan Tanggungan. Struktur ini mempertahankan keterbacaan model sekaligus mencapai kinerja validasi yang tinggi.

Gambar 2. Struktur pohon keputusan C4.5 hasil pemodelan



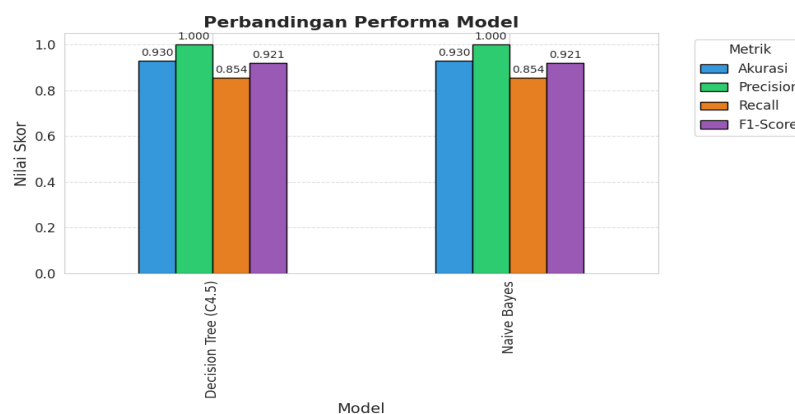
Perbandingan Kinerja Model

C4.5 dan Naive Bayes menghasilkan prediksi validasi silang yang identik. Dari 48 data Layak, 41 diklasifikasikan benar dan 7 diprediksi Tidak Layak. Seluruh 52 data Tidak Layak diklasifikasikan benar dan tidak terdapat false positive. Susunan kesalahan tersebut menghasilkan accuracy 93,00%, precision kelas Layak 100,00%, recall 85,42%, dan F1-score 92,13% pada kedua model.

Tabel 4. Perbandingan metrik evaluasi model

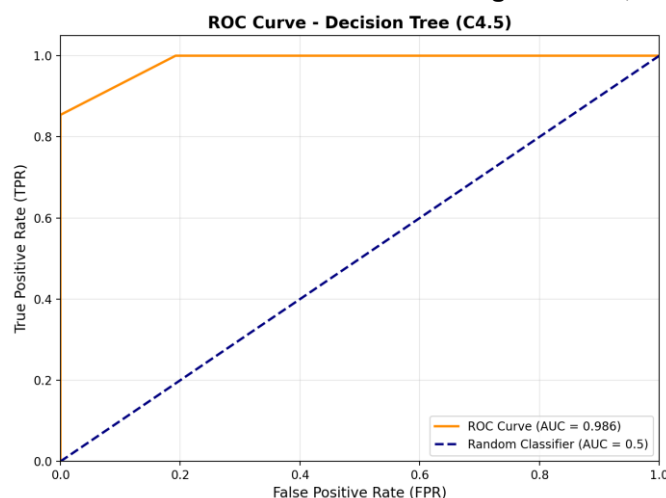
Metrik	C4.5	Naive Bayes
Accuracy	93,00%	93,00%
Precision (Layak)	100,00%	100,00%
Recall (Layak)	85,42%	85,42%
F1-score (Layak)	92,13%	92,13%
ROC-AUC	0,986	0,868

Gambar 3. Perbandingan accuracy, precision, recall, dan F1-score



Walaupun metrik berbasis label sama, kurva ROC membedakan kekuatan pemeringkatan probabilitas. C4.5 memperoleh AUC 0,986, sedangkan Naive Bayes memperoleh 0,868. Nilai C4.5 yang lebih dekat ke 1 menunjukkan kemampuan lebih baik dalam memisahkan dua kelas pada berbagai ambang. Dengan mempertimbangkan AUC, aturan yang dapat dijelaskan, dan kondisi akurasi yang sama, C4.5 dipilih sebagai model akhir.

Gambar 4. Kurva ROC model C4.5 dengan AUC 0,986



Penerapan pada Data Warga dan Pengujian Sistem

Model akhir diterapkan pada 50 kepala keluarga di Desa Talang Benuang. Sebanyak 42 KK (84%) diprediksi Layak dan 8 KK (16%) diprediksi Tidak Layak. Kedua algoritma menghasilkan distribusi prediksi yang sama. Pemeriksaan kewajaran menunjukkan seluruh prediksi Layak memiliki pendapatan rendah atau sedang, sedangkan seluruh prediksi Tidak Layak memiliki pendapatan tinggi. Konsistensi ini memperlihatkan bahwa keluaran mengikuti pola yang dipelajari dari data berlabel.

Tabel 5. Ringkasan hasil penerapan model pada data warga

Hasil prediksi	Jumlah KK	Persentase
Layak	42	84%
Tidak Layak	8	16%
Total	50	100%

Pengujian black box menunjukkan seluruh fungsi utama berjalan sesuai keluaran yang diharapkan: pustaka dapat diimpor, data 100×6 dan 50×6 terbaca, pemeriksaan missing value berhasil, transformasi ordinal berjalan, parameter terbaik ditemukan, kedua model terlatih, validasi menghasilkan metrik, visualisasi terbentuk, 50 data warga memperoleh prediksi, dan berkas hasil dapat diekspor. Waktu terlama terdapat pada GridSearchCV sekitar 1,3 detik karena mengevaluasi 36 kombinasi dalam lima fold. Training Naive Bayes berlangsung 4,21 ms, prediksi validasi silang sekitar 30–33 ms, dan prediksi 50 data warga sekitar 1,60 ms per model.

Pembahasan

Kesamaan accuracy, precision, recall, dan F1-score menunjukkan bahwa pada ambang klasifikasi baku, kedua algoritma membentuk batas keputusan yang sama untuk dataset ini. Hal tersebut terutama dipengaruhi oleh dominasi Pendapatan. Sebanyak 41 data pada kategori rendah dan sedang seluruhnya berlabel Layak, sedangkan kelompok pendapatan tinggi didominasi Tidak Layak. Ketika satu atribut menyediakan pemisahan sangat kuat, pohon keputusan dan model probabilistik dapat berakhir pada prediksi kelas yang sama meskipun mekanisme matematisnya berbeda.

Temuan bahwa C4.5 efektif mendukung klasifikasi bantuan sosial sejalan dengan Susanto et al. (2026), khususnya karena pohon keputusan menghasilkan aturan yang dapat dijelaskan. Efektivitas Naive Bayes juga selaras dengan Rifat dan Akhriza (2025) yang menggunakannya untuk penerima PKH. Dibandingkan kajian Attamami et al. (2023), penelitian ini memperlihatkan bahwa kesetaraan akurasi tidak selalu berarti kualitas pemisahan probabilitas yang sama. Selisih AUC 0,118 memperlihatkan nilai tambah evaluasi multi-metrik ketika dua model tampak setara pada confusion matrix.

Precision kelas Layak sebesar 100% berarti tidak ada data Tidak Layak yang salah diberi status Layak. Dari perspektif pengendalian ketepatan sasaran, karakteristik ini mengurangi risiko memasukkan rumah tangga yang tidak memenuhi pola kelayakan. Namun, recall 85,42% berarti tujuh dari 48 data Layak tidak terdeteksi, atau sekitar 14,58%. Kesalahan false negative tersebut memiliki konsekuensi sosial lebih serius karena rumah tangga yang seharusnya layak dapat tersisih. Oleh sebab itu, model sebaiknya digunakan sebagai penyaring dan pemberi prioritas pemeriksaan, bukan sebagai pemutus tunggal.

Interpretabilitas C4.5 menjadi keunggulan operasional. Empat aturan dengan kedalaman maksimum tiga dapat diterjemahkan ke prosedur pemeriksaan yang mudah diaudit. Pemerintah desa dapat melihat atribut yang mengarahkan keputusan, menelusuri kasus yang berada pada cabang berisiko salah, dan membandingkannya dengan dokumen serta kondisi lapangan. Naive Bayes lebih ringkas secara komputasional, tetapi probabilitas gabungan kurang intuitif bagi pengguna yang tidak terbiasa dengan model statistik.

Nilai feature importance Pendapatan sebesar 0,827 dan gain 0,6888 menegaskan dominasi indikator ekonomi. Dominasi tersebut logis karena kemampuan memenuhi kebutuhan pangan berkaitan langsung dengan pendapatan. Meskipun demikian, ketergantungan yang terlalu besar pada satu atribut dapat menyederhanakan kemiskinan yang bersifat multidimensional. Status tempat tinggal tidak berkontribusi karena nilainya konstan; Tanggungan dan Kendaraan menyediakan variasi yang lebih kecil. Penambahan pekerjaan, kualitas rumah, pengeluaran, kepemilikan aset produktif, kondisi kesehatan, dan kerentanan anggota keluarga berpotensi memperkaya representasi kelayakan.

Kebaruan penelitian terletak pada penerapan alur komparatif yang seragam untuk BPNT tingkat desa: preprocessing yang sama, penyetalan khusus C4.5, 5-fold cross-validation, pemeriksaan confusion matrix, ROC-AUC, feature importance, aturan keputusan, waktu komputasi, dan penerapan pada data warga. Integrasi tersebut menghasilkan dasar pemilihan model yang bukan hanya akurat, tetapi juga dapat dijelaskan. C4.5 dipilih bukan karena accuracy lebih tinggi, melainkan karena AUC lebih baik dan struktur keputusan lebih transparan.

Hasil prediksi 42 Layak dan 8 Tidak Layak harus ditafsirkan secara hati-hati karena 50 data warga tidak memiliki label ground truth digital. Kesesuaian pola dengan kriteria pendapatan merupakan sanity check, bukan bukti akurasi eksternal. Selain itu, data pemodelan hanya 100 rekaman publik dan mungkin tidak sepenuhnya merepresentasikan karakteristik Desa Talang Benuang. Validitas penerapan akan meningkat apabila hasil musyawarah desa didigitalkan, dipadankan dengan prediksi, dan digunakan untuk evaluasi prospektif.

Ditemukan pula kesalahan teknis pada keluaran kode awal terkait urutan class_names dalam ekstraksi aturan dan indeks probabilitas pada ROC-AUC. Setelah urutan kelas disesuaikan dengan classes_model, aturan menunjukkan kelompok pendapatan rendah atau sedang sebagai Layak; sedangkan cabang pendapatan tinggi diarahkan melalui Kendaraan dan Tanggungan. Nilai AUC C4.5 yang benar adalah 0,986. Temuan debugging ini penting karena visualisasi yang tampak meyakinkan dapat menghasilkan interpretasi keliru apabila pemetaan kelas tidak diverifikasi.

KESIMPULAN DAN SARAN

Kesimpulan

Model klasifikasi kelayakan BPNT berhasil dibangun menggunakan C4.5 dan Naive Bayes melalui pipeline preprocessing, pemodelan, dan 5-fold cross-validation. Kedua algoritma menghasilkan accuracy 93,00%, precision 100,00%, recall 85,42%, dan F1-score 92,13%. C4.5 ditetapkan sebagai model paling sesuai karena memiliki ROC-AUC 0,986, lebih tinggi daripada Naive Bayes 0,868, serta menghasilkan pohon dan aturan keputusan yang mudah ditelusuri. Pendapatan menjadi atribut dominan dengan feature importance 0,827 dan information gain 0,6888. Penerapan pada 50 KK menghasilkan 42 prediksi Layak dan 8 Tidak Layak. Dengan demikian, C4.5 dapat mendukung penyaringan awal BPNT secara lebih cepat, konsisten, dan transparan, tetapi tidak menggantikan verifikasi dokumen, pemeriksaan lapangan, dan musyawarah desa.

Saran

Keterbatasan penelitian mencakup jumlah data latih yang kecil, perbedaan sumber antara data pelatihan dan data penerapan, belum tersedianya label ground truth digital untuk warga desa, dominasi satu atribut, serta temuan bug pada pemetaan kelas dalam keluaran awal. Pemerintah desa disarankan mendigitalkan hasil penetapan BPNT dan menggunakan model sebagai daftar prioritas verifikasi. Penelitian berikutnya perlu menambah jumlah serta keragaman data, memasukkan indikator sosial ekonomi yang lebih komprehensif, melakukan validasi eksternal pada beberapa desa, mengoptimalkan ambang klasifikasi untuk menekan false negative, dan membandingkan Random Forest, Support Vector Machine, maupun metode ensemble. Sistem juga dapat dikembangkan menjadi aplikasi web dengan audit trail dan mekanisme koreksi data agar dapat digunakan tanpa ketergantungan pada notebook pemrograman.

DAFTAR PUSTAKA

- Adriansa, M., Yulianti, L., Elfianty, L., Dehasen Bengkulu, U., & Meranti Raya, J. (2022). Analisis Kepuasan Pelanggan Menggunakan Algoritma C4.5.
- Annas, A., Amalia, R. P., & Nursafitra. (2022). Pelayanan bantuan sosial. Chakti Pustaka Indonesia.
- Apandi, J., Eko Wijaya, A., Hermawan, R., Faritcan Parlaungan Siallagan, T., Udoyono, K., & Nur Ahmad, H. (2025). Perbandingan kinerja algoritma naïve bayes dan c4.5 pada sistem web klasifikasi kelayakan PKH. In STMIK Subang (Vol. 18, Number 1).
- Astono, A. D. (2021). Metodologi penelitian: Metode penelitian manajemen pendekatan kuantitatif dan kualitatif. Cahya Ghani Recovery.
- Attamami, N., Triayudi, A., & Aldisa, R. T. (2023). Analisis performa algoritma klasifikasi Naive Bayes dan C4.5 untuk prediksi penerima bantuan jaminan kesehatan. Jurnal JTIC (Jurnal Teknologi Informasi dan Komunikasi), 7(2), 262–269.

- Aziz, F. N., Mubin, F., Hidayat, R. J. P., Nurjaman, A., Romadhan, A. A., Sulistyaningsih, T., & Hijri, Y. S. (2021). Bagaimana Teknologi Informasi dan Komunikasi Bertransformasi Menjadi Inovasi Pelayanan Publik? *PERSPEKTIF*, 10(2), 616–626. <https://doi.org/10.31289/perspektif.v10i2.4905>
- Bilal Alfayyadh, M., & Assegaff, S. (2025). Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Machine Learning Untuk Klasifikasi Performa Pelajar. *Jurnal Manajemen Teknologi Dan Sistem Informasi (JMS)*, 5(2), 1095. <https://doi.org/10.33998/jms.v5i2>
- Darmawan, A. (2025). BPJS Ketenagakerjaan: Antara hak dan kewajiban. Lembaga Pembangunan Masyarakat Indonesia
- Fathudin, D., Nurhalimah, N., & Shakina, P. (2026). Analisis Sentimen Ulasan Digital untuk Pemula Menggunakan Google Colab. CV Eureka Media Aksara.
- Felicia Watratan, A., Puspita, A. B., Moeis, D., Informasi, S., & Profesional Makassar, S. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. In *journal of applied computer science and technology (jacost)* (Vol. 1, Number 1). <http://journal.isas.or.id/index.php/JACOST>
- Hanief, S., & Jepriana, I. W. (2020). Konsep algoritme dan aplikasinya dalam bahasa pemrograman C++. Yogyakarta: ANDI.
- Huda, A., Ardi, N., & Mubai, A. (2021). Pengantar coding berbasis C/C++. UNP Press.
- Indah Juwita Harahap, R., Khairani, S., & author, correspondent. (2024). Implementasi Metode K-Nearest Neighbor Untuk Klasifikasi Penyakit Tanaman Mentimun Pada Citra Daun Implementation Of The K-Nearest Neighbor Method For Classification Of Cucumber Diseases On Leaf Images (Vol. 2024, Number 2). <https://jurnal.unity-academy.sch.id/index.php/jirsi/index>
- Kohsasih, K. L., Situmorang, Z., & Artikel, I. (2022). Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular. *jurnal informatika*, 9(1), 13–17. <http://ejournal.bsi.ac.id/ejurnal/index.php/ji>
- Muis, A., Syafwan, H., Arfianto, A. Z., Simanjuntak, M. S., Rivandi, A., Trisnawan, A. B., Ramdhan, W., Triana, H., Saputra, M. H., Handoko, W., Naswin, A., & Suhadi. (2025). Data mining: Konsep, metode, dan aplikasi. PT. Faaslib Serambi Media.
- Mukhid, A. (2021). Metodologi penelitian pendekatan kuantitatif. CV. Jakad Media Publishing.
- Novella, Y., & Mardhatillah, Y. (2023). Penggunaan teknologi informasi dan komunikasi dalam pelayanan kependudukan pada dinas kependudukan dan pencatatan sipil kabupaten buleleng. *Jurnal Konstituen*, 4(2), 114–131. <https://doi.org/10.33701/jk.v4i2.2836>
- Rifat, M., & Akhriza, T. M. (2025). ntuan sosial program keluarga harapan (PKH) Kecamatan Gondanglegi menggunakan metode Naive Bayes. *Jurnal Ilmiah KOMPUTASI*, 24(2), 289-296.
- Rojabi, M. A. (2025). Cara cek status bansos PKH & BPNT 2025: Panduan lengkap dan praktis. Afdan Rojabi Publisher.
- S, U. (2023). Penerapan Data Mining Dengan Mengimplementasikan Algoritma K-Means Dalam Proses Clustering Untuk Pengelompokan Mahasiswa Calon Penerima Beasiswa KIP. *Building of Informatics, Technology and Science (BITS)*, 5(1). <https://doi.org/10.47065/bits.v5i1.3411>