

OPTIMALISASI DECISION TREE UNTUK PENENTUAN KELAYAKAN PENERIMA BANTUAN KESEJAHTERAAN

Olanda¹, Bambang Irawan², Willy Prihartono³, Nisa Dienwati Nuris⁴.

Program Studi Teknik Informatika¹²⁴
Program Studi Komputerisasi Akuntansi³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
olanda027028062003@gmail.com

(*) Corresponding Author : olanda027028062003@gmail.com
Published : 20 Juli 2026

Abstract—Determining the eligibility of social welfare assistance (KESRA) recipients is a critical process to ensure that aid distribution is fair and accurately targeted; however, manual decision-making often suffers from subjectivity, limited data analysis, and potential classification errors. This research aims to apply data mining techniques using the Decision Tree algorithm to classify the eligibility of KESRA assistance recipients based on socio-economic characteristics. The study adopts a supervised learning approach comprising several stages, including data preprocessing to clean and prepare the dataset, data transformation to suit the modeling requirements, baseline Decision Tree model development as an initial benchmark, and model optimization using GridSearchCV and pruning techniques to improve performance and reduce overfitting. The dataset is divided into training and testing subsets using an 80:20 ratio to ensure objective evaluation. Model performance is assessed using accuracy, precision, recall, and F1-score metrics and further analyzed through a confusion matrix to identify misclassification patterns. The results demonstrate that the baseline Decision Tree model achieves an accuracy of 90%, while the optimized model shows improved performance with a more compact and interpretable tree structure. Furthermore, feature importance analysis indicates that income is the most influential variable in determining eligibility for KESRA assistance. Based on these findings, it can be concluded that an optimized Decision Tree model is effective and reliable as a decision support system for objectively and accurately determining social welfare assistance eligibility using a data-driven approach.

Keywords : *Decision Tree, Social Welfare Assistance (KESRA), Classification, Data Mining, GridSearchCV*

Abstrak—Penentuan kelayakan penerima bantuan kesejahteraan (KESRA) merupakan proses penting untuk memastikan bahwa bantuan diberikan secara tepat sasaran dan adil, namun dalam praktiknya penilaian yang dilakukan secara manual sering menghadapi kendala berupa subjektivitas, keterbatasan analisis data, serta potensi kesalahan dalam pengambilan keputusan. Penelitian ini bertujuan untuk menerapkan metode data mining menggunakan algoritma Decision Tree dalam mengklasifikasikan kelayakan penerima bantuan KESRA berdasarkan karakteristik sosial dan ekonomi masyarakat. Metode penelitian yang digunakan adalah supervised learning melalui beberapa tahapan, yaitu data preprocessing untuk membersihkan dan menyiapkan data, transformasi data agar sesuai dengan kebutuhan pemodelan, pembentukan model Decision Tree baseline sebagai pembandingan awal, serta optimasi model menggunakan GridSearchCV dan teknik pruning guna meningkatkan performa dan menghindari overfitting. Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk memastikan evaluasi yang objektif. Kinerja model dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, serta dianalisis menggunakan confusion matrix untuk mengidentifikasi kesalahan klasifikasi. Hasil penelitian menunjukkan bahwa model Decision Tree baseline mampu mencapai tingkat akurasi sebesar 90%, sedangkan model hasil optimasi menunjukkan peningkatan performa dengan struktur pohon keputusan yang lebih sederhana dan mudah diinterpretasikan. Selain itu, analisis feature importance menunjukkan bahwa variabel pendapatan merupakan faktor paling dominan dalam menentukan kelayakan penerima bantuan KESRA. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma Decision Tree yang telah dioptimasi dapat digunakan secara efektif sebagai sistem pendukung keputusan dalam penentuan kelayakan penerima bantuan kesejahteraan secara objektif, akurat, dan berbasis data.

Kata Kunci : Decision Tree, Bantuan Kesejahteraan (KESRA), Klasifikasi, Data Mining, GridSearchCV

INTRODUCTION

Perkembangan teknologi informasi yang berlangsung sangat cepat telah membawa perubahan besar dalam berbagai aspek kehidupan masyarakat masa kini. Disiplin informatika tidak lagi terbatas pada pembuatan perangkat lunak ataupun sistem informasi, tetapi telah berkembang dan dimanfaatkan di berbagai bidang strategis seperti pemerintahan, dunia usaha, pendidikan, kesehatan, hingga layanan sosial. Salah satu peran penting dari teknologi informatika adalah kemampuannya mengolah serta menganalisis data dalam jumlah besar sehingga dapat menghasilkan informasi yang relevan untuk mendukung pengambilan keputusan. Dengan memanfaatkan teknik *data mining* dan *machine learning*, berbagai organisasi mampu menemukan pola tersembunyi dalam data yang kemudian digunakan untuk meningkatkan efisiensi, efektivitas operasi, serta ketepatan dalam pemberian layanan publik [1]

Dalam ranah pelayanan *publik* dan program bantuan sosial, kemajuan teknologi informatika memberikan peluang besar untuk meningkatkan *transparansi*, *akuntabilitas*, serta ketepatan sasaran berbagai program pemerintah Ladianto, (2023). Kesejahteraan sosial merupakan pilar penting dalam pembangunan nasional dan menjadi tanggung jawab negara untuk memastikan bahwa seluruh warga dapat hidup layak. Program bantuan kesejahteraan dirancang guna membantu masyarakat kurang mampu dari segi ekonomi, pendidikan, kesehatan, maupun sosial. Namun demikian, pelaksanaan program bantuan sosial di Indonesia masih sering menghadapi kendala berupa ketidaktepatan sasaran, di mana warga yang sebenarnya mampu justru menerima bantuan, sementara mereka yang membutuhkan tidak terakomodasi [3].

Ketidaktepatan tersebut umumnya disebabkan oleh tidak tersedianya sistem pendukung keputusan yang objektif dan berbasis data dalam proses seleksi penerima bantuan. Penilaian yang masih dilakukan secara manual dan mengandalkan subjektivitas petugas menyebabkan hasil yang tidak konsisten dan rentan terhadap *bias*. Padahal, dengan tersedianya data sosial ekonomi yang semakin lengkap di era digital, proses seleksi semestinya dapat dilakukan dengan metode yang lebih modern, terstruktur, dan berbasis

data sehingga menghasilkan keputusan yang lebih adil dan akurat [4]

Sejumlah penelitian sebelumnya telah mengkaji penerapan data mining untuk klasifikasi penerima bantuan sosial. Di Indonesia, studi menunjukkan bahwa algoritma *Decision Tree* mampu memberikan tingkat akurasi yang tinggi dalam menentukan kelayakan penerima bantuan berdasarkan indikator sosial ekonomi Fitriza et al., (2024) Penelitian lain menguatkan bahwa *Decision Tree* dapat menghasilkan visualisasi pemetaan keputusan yang mudah dipahami oleh pemangku kebijakan, sehingga sangat *relevan* untuk digunakan dalam pelayanan publik Satria et al., (2024). Meski demikian, peningkatan performa model masih dapat dilakukan, di antaranya melalui teknik optimasi seperti *pruning* guna mengurangi *overfitting*. Penelitian ini berfokus pada pengembangan sistem klasifikasi untuk menentukan kelayakan penerima bantuan kesejahteraan yang lebih akurat, *efisien*, dan *objektif* dengan memanfaatkan algoritma *Decision Tree* yang telah dioptimalkan menggunakan teknik *cost-complexity pruning*.

Teknik ini dipilih karena mampu menyederhanakan struktur pohon tanpa menurunkan akurasi, bahkan dapat meningkatkan kemampuan model dalam melakukan *generalisasi* terhadap data baru Huyghe et al., (2024) Melalui penelitian ini, diharapkan tersedianya solusi yang dapat membantu pemerintah daerah melakukan seleksi penerima bantuan secara lebih transparan dan berbasis data. Pendekatan yang digunakan adalah metode eksperimental dengan menerapkan algoritma *Decision Tree* berbasis CART, yang menggunakan *Gini Index* sebagai ukuran *impurity*. Optimasi dilakukan dengan *cost-complexity pruning* untuk menyeimbangkan tingkat kerumitan pohon keputusan dengan tingkat kesalahan klasifikasi.

Data penelitian terdiri atas *atribut* sosial ekonomi seperti pendapatan, jumlah tanggungan, kondisi rumah, dan kepemilikan aset, kemudian dievaluasi menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Apabila penelitian ini menghasilkan model yang optimal, maka model tersebut berpotensi digunakan sebagai sistem pendukung keputusan untuk membantu pemerintah menyalurkan bantuan dengan lebih tepat sasaran. Pendekatan berbasis data ini dapat meminimalkan kesalahan distribusi bantuan sekaligus meningkatkan kepercayaan masyarakat terhadap program-program kesejahteraan sosial Rajul Hakim, Muhammad Adnan, (2024) Selain itu, penelitian ini memberikan kontribusi bagi

pengembangan ilmu *data mining* dan *machine learning*, khususnya dalam penerapan algoritma *Decision Tree* pada konteks kesejahteraan sosial.

MATERIALS AND METHODS

Tahapan penelitian merupakan bagian penting dalam proses pengembangan sistem berbasis data mining dan machine learning karena setiap langkah memiliki peran yang saling berkaitan dalam menghasilkan model yang akurat dan optimal. Proses penelitian yang terstruktur membantu peneliti dalam memahami alur pengolahan data mulai dari identifikasi masalah hingga implementasi sistem pendukung keputusan. Selain itu, penggunaan metode yang tepat pada setiap tahapan mampu meningkatkan kualitas hasil analisis serta mempermudah proses evaluasi model yang digunakan. Dalam penelitian ini digunakan algoritma *Decision Tree Classification (CART)* yang dipadukan dengan teknik optimasi *cost-complexity pruning* untuk meningkatkan performa model klasifikasi. Adapun tahapan penelitian yang dilakukan dapat dilihat pada gambar berikut.



Gambar 1 Alur Penelitian

Gambar tersebut menunjukkan alur penelitian yang digunakan dalam proses pengembangan sistem klasifikasi berbasis algoritma *Decision Tree Classification (CART)*. Tahapan pertama dimulai dari identifikasi masalah dan tujuan penelitian. Pada tahap ini peneliti menentukan permasalahan utama yang akan diselesaikan serta menetapkan tujuan penelitian agar proses analisis dapat berjalan secara terarah dan sistematis. Setelah permasalahan berhasil diidentifikasi, tahapan berikutnya adalah pengumpulan data sosial. Data dikumpulkan dari berbagai sumber yang relevan sesuai kebutuhan penelitian untuk memperoleh informasi yang dapat digunakan dalam proses analisis dan pemodelan data.

Tahap selanjutnya adalah pra-pemrosesan data atau preprocessing data. Pada proses ini dilakukan pembersihan data, penghapusan data duplikat, penanganan missing value, transformasi data, dan normalisasi agar data menjadi lebih terstruktur dan siap digunakan pada proses klasifikasi. Setelah preprocessing selesai, data diproses menggunakan algoritma *Decision Tree Classification (CART)*. Algoritma *CART* digunakan untuk membangun model klasifikasi berbentuk pohon keputusan yang mampu menghasilkan aturan keputusan berdasarkan atribut data yang digunakan.

Setelah model *CART* terbentuk, dilakukan proses optimasi model menggunakan teknik *cost-complexity pruning*. Teknik ini bertujuan untuk mengurangi kompleksitas pohon keputusan agar model tidak mengalami overfitting serta mampu meningkatkan kemampuan generalisasi terhadap data baru. Tahapan berikutnya adalah evaluasi performansi model untuk mengukur tingkat akurasi, presisi, recall, maupun performa klasifikasi lainnya. Hasil evaluasi kemudian dianalisis dan diinterpretasikan untuk mengetahui kualitas model yang dihasilkan serta melihat pola-pola penting yang diperoleh dari proses klasifikasi.

Tahap berikutnya adalah implementasi model ke dalam sistem pendukung keputusan. Pada tahap ini model yang telah dioptimasi diterapkan ke dalam sistem sehingga dapat digunakan untuk membantu proses pengambilan keputusan secara otomatis dan lebih efektif. Tahap terakhir adalah penarikan kesimpulan dan rekomendasi berdasarkan hasil penelitian yang telah dilakukan. Kesimpulan berisi hasil akhir penelitian, sedangkan rekomendasi diberikan sebagai saran untuk pengembangan sistem maupun penelitian selanjutnya agar dapat menghasilkan performa yang lebih baik.

RESULTS AND DISCUSSION

tahap berikutnya difokuskan pada proses optimasi untuk menyederhanakan struktur pohon tanpa mengorbankan akurasi, serta meningkatkan kemampuan model dalam menentukan kelayakan penerima bantuan KESRA.



Gambar 2 Model Optimasi GirdSearchCV

Pada Gambar 2 Model Optimasi *GridSearchCV* Optimasi model dilakukan menggunakan *GridSearchCV* untuk mencari kombinasi *hyperparameter* terbaik pada *Decision Tree*, meliputi pemilihan kriteria pemisahan, pembatasan kedalaman pohon, serta pengaturan jumlah minimum sampel pada pemecahan *node* dan daun. Proses ini menggunakan *cross-validation* untuk memastikan performa yang lebih stabil dan tidak bias. Setelah pengujian seluruh kombinasi *hyperparameter*, *GridSearchCV* memberikan konfigurasi terbaik yang menghasilkan model lebih akurat, sederhana, dan mampu melakukan generalisasi dengan baik dalam menentukan kelayakan penerima bantuan KESRA secara objektif.

Berikut Output Setelah Menjalankan Optimasi *GridSearchCV*:



Gambar 3 Output Optimasi *GridSearchCV*

Seperti yang dijelaskan pada Gambar 3 Output Optimasi *GridSearchCV* Melalui proses *GridSearchCV*, diperoleh kombinasi *hyperparameter* terbaik, yaitu *criterion='gini'*, *max_depth=3*, *min_samples_leaf=2*, dan *min_samples_split=4*. Penggunaan Gini Index memungkinkan proses komputasi berjalan lebih cepat serta mempermudah pemisahan *node*, sementara pembatasan kedalaman pohon hingga level 3 membantu mencegah model menjadi terlalu kompleks dan rentan mengalami *overfitting*. Selain itu, penetapan jumlah minimum sampel pada *node* daun maupun saat pemecahan *node* memastikan bahwa setiap keputusan diambil berdasarkan data yang memadai sehingga struktur pohon tetap efisien. Dengan konfigurasi tersebut, model yang dihasilkan menjadi lebih stabil, mudah dipahami, dan mampu melakukan generalisasi secara efektif pada data baru. Konsistensi hasil prediksi pun meningkat, menjadikan model ini dapat diandalkan untuk mendukung penentuan kelayakan penerima bantuan KESRA secara objektif dan berbasis data. Pendekatan ini memberikan keuntungan signifikan dalam penerapan sistem pendukung keputusan di tingkat desa.

Tahap *Evaluation* (*Evaluation Model*)

Ujicoba



		Model		Baseline	
		precision	recall	f1-score	support
0		0.00	0.94	0.00	22
1		0.96	0.91	0.94	88
accuracy		0.96	0.91	0.91	85
macro avg		0.98	0.91	0.96	85
weighted avg		0.97	0.91	0.91	85

e

Gambar 4 Output Ujicoba Model Baseline

Seperti yang sudah ditampilkan pada Gambar 4 Output Ujicoba Model Baseline Evaluasi model *baseline* menggunakan *accuracy_score* dan *classification_report* menunjukkan akurasi 90%, dengan *precision* dan *recall* yang cukup baik pada kedua kelas, sehingga model sudah mampu mengidentifikasi kelayakan penerima bantuan secara cukup akurat. Namun, *recall* pada kelas "Layak" masih perlu ditingkatkan untuk mengurangi kasus warga yang seharusnya menerima bantuan tetapi tidak terpilih (*false negative*). Oleh sebab itu, model *baseline* perlu dioptimalkan menggunakan *GridSearchCV* agar performanya meningkat tanpa menimbulkan risiko *overfitting*.

Ujicoba Model Optimal



		precision	recall	f1-score	support
0		0.94	0.91	0.92	22
1		0.99	0.96	0.97	88
accuracy		0.96	0.91	0.94	85
macro avg		0.97	0.94	0.94	85
weighted avg		0.97	0.91	0.91	85

Gambar 4 Evaluasi Model Optimal

Seperti yang sudah ditampilkan pada Gambar 4 Output Ujicoba Model Optimal Model *Decision Tree* yang telah dioptimalkan menggunakan *GridSearchCV* menghasilkan akurasi sebesar 94,3%, dengan *precision* dan *recall* yang tinggi pada kedua kelas. Pada kategori Tidak Layak, model mencatat *precision* 0,94 dan *recall* 0,91, sedangkan pada kategori Layak *precision* mencapai 0,95 dan *recall* 0,96, sehingga performa model meningkat dibanding *baseline*. Nilai *macro* dan *weighted average* yang berada di kisaran 0,94 menunjukkan konsistensi performa meskipun distribusi kelas tidak seimbang. Dengan hasil tersebut, model optimal dinilai mampu memberikan prediksi yang lebih kuat dan reliabel dalam mendukung penentuan kelayakan penerima bantuan KESRA secara objektif.

Jika Dibandingkan dengan model *baseline*, terjadi hasil yang cukup signifikan :

Tabel 1 Perbandingan Ujicoba model *baseline* & Model Optimal

Metrik	Baseline	Model Optimal
Accuracy	0.90	0.943
Precision	0.90	0.94
Recall	0.91	0.94
F1-Score	0.90	0.94

Seperti Yang sudah dijelaskan pada Tabel 4.2 Perbandingan Ujicoba model *Baseline* & Model Optimal Penerapan optimasi pada model *Decision Tree* melalui *GridSearchCV* dan *cost-complexity pruning* terbukti memberikan peningkatan performa yang signifikan. Akurasi model meningkat dari 90% menjadi 94,3%, disertai kenaikan nilai *precision*, *recall*, dan *F1-score* yang semuanya berada di atas 0,94 untuk kedua kelas. Hal ini menunjukkan bahwa kemampuan generalisasi model semakin baik sekaligus mengurangi potensi *overfitting*. *Confusion matrix* turut memperlihatkan penurunan jumlah kesalahan prediksi, sehingga model semakin tepat dalam membedakan warga yang layak dan tidak layak menerima bantuan. Selain memberikan hasil prediksi yang lebih akurat, model teroptimasi ini juga menghasilkan struktur pohon yang lebih ringkas dan mudah dipahami, sehingga sangat mendukung penerapannya sebagai sistem pendukung keputusan di tingkat desa. Dengan performa yang konsisten serta tingkat *interpretabilitas* yang tinggi, model ini dianggap layak digunakan dalam proses penentuan kelayakan penerima bantuan KESRA secara objektif dan lebih tepat sasaran.

Evaluasi Cross-Validation

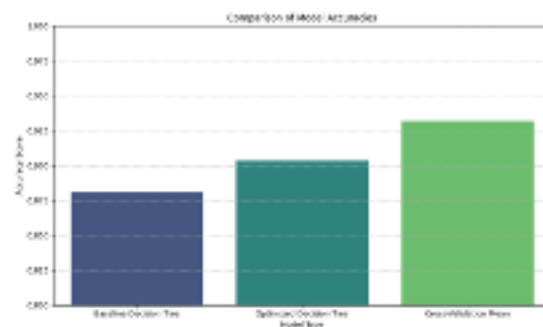


Gambar 5 Evaluasi Cross-Validation

Berdasarkan Gambar 5 Evaluasi *Croos-Validation* Untuk memastikan bahwa model *Decision Tree* yang telah dioptimasi tidak mengalami *overfitting*, dilakukan evaluasi tambahan menggunakan metode *Repeated Stratified K-Fold Cross Validation* dengan 10 fold dan 5 pengulangan. Metode ini digunakan untuk menguji model secara berulang dengan pembagian data yang berbeda-beda, sehingga hasil evaluasi menjadi lebih stabil dan tidak bergantung pada satu pembagian data saja. Hasil pengujian menunjukkan bahwa model memiliki rata-rata akurasi sebesar 93.21% dengan standar deviasi sebesar 3.67% Nilai rata-rata yang tinggi menunjukkan bahwa model memiliki

performa yang baik, sedangkan nilai standar deviasi yang rendah menunjukkan bahwa model memiliki konsistensi yang tinggi pada berbagai skenario pembagian data. Dengan demikian, dapat disimpulkan bahwa model *Decision Tree* yang telah dioptimasi memiliki kemampuan generalisasi yang baik serta tidak mengalami *overfitting*, sehingga layak digunakan dalam proses penentuan kelayakan penerima bantuan KESRA.

Output Cross Validation



Gambar 6 Output Cross Validation

Berdasarkan Gambar 6 *Output Cross Validation* grafik *Comparison of Model Accuracies*, terlihat bahwa model *Baseline Decision Tree* memperoleh akurasi sekitar 88%, sedangkan setelah dilakukan proses optimasi menggunakan *GridSearchCV* dan *cost-complexity pruning*, akurasi meningkat menjadi sekitar 90%. Selanjutnya, hasil evaluasi menggunakan metode *Repeated Stratified K-Fold Cross Validation* menunjukkan rata-rata akurasi sebesar 93,21%. Peningkatan tersebut menunjukkan bahwa model yang telah dioptimasi memiliki kemampuan *generalisasi* yang lebih baik dibandingkan model awal. Nilai rata-rata akurasi *cross validation* yang lebih tinggi dibandingkan model *baseline* menunjukkan bahwa model mampu mempertahankan performanya secara konsisten pada berbagai pembagian data. Hal ini menandakan bahwa model tidak hanya bekerja baik pada satu data uji tertentu, tetapi juga stabil ketika diuji pada kombinasi data pelatihan dan pengujian yang berbeda. Selain itu, standar deviasi sebesar 3,67% menunjukkan bahwa variasi performa model relatif rendah, sehingga model dapat dikategorikan cukup stabil dan tidak mengalami *overfitting* secara signifikan.

Hasil tersebut sejalan dengan penelitian Huyghe et al., (2024) yang menjelaskan bahwa penerapan optimasi dan teknik *pruning* pada *Decision Tree* mampu meningkatkan stabilitas model serta menghasilkan kemampuan generalisasi yang lebih baik pada data baru. Penelitian Balcan & Sharma, (2024) juga menyebutkan bahwa *cost-complexity pruning* efektif dalam mengurangi kompleksitas

pohon keputusan tanpa menurunkan performa klasifikasi secara signifikan. Jika dibandingkan dengan penelitian terbaru Perhan et al., (2025), penelitian ini menunjukkan pengembangan pada aspek evaluasi *stabilitas* model. Penelitian tersebut menekankan pentingnya *reliabilitas* model klasifikasi bantuan sosial, namun evaluasi performa masih berfokus pada hasil akurasi pengujian utama. Sementara itu, penelitian ini menerapkan *Repeated Stratified K-Fold Cross Validation* untuk mengukur konsistensi performa model pada berbagai pembagian data, sehingga hasil evaluasi menjadi lebih representatif terhadap kemampuan *generalisasi* model.

Selain itu, Siddique et al., (2024) menjelaskan bahwa evaluasi model klasifikasi sosial sebaiknya tidak hanya menggunakan satu kali pembagian data karena hasil evaluasi dapat dipengaruhi oleh distribusi *dataset* tertentu. Pendekatan *cross validation* yang digunakan dalam penelitian ini membantu mengurangi potensi *bias* evaluasi dan memastikan bahwa model memiliki performa yang konsisten pada berbagai skenario pengujian. Dengan demikian, hasil *cross validation* menunjukkan bahwa model *Decision Tree* yang telah dioptimasi tidak hanya memiliki tingkat akurasi yang baik, tetapi juga stabil, konsisten, dan memiliki kemampuan generalisasi yang memadai untuk diterapkan dalam proses klasifikasi kelayakan penerima bantuan KESRA secara objektif dan berbasis data.

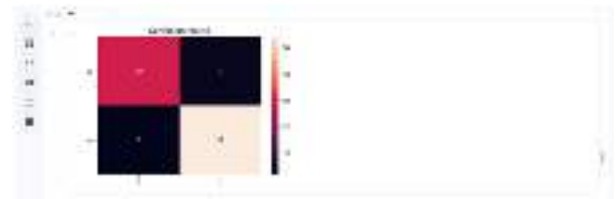
Confusion Matrix



Gambar 4.18 Confusion Matrix

Seperti yang sudah ditampilkan pada Gambar 4.18 Confusion Matrix Visualisasi *confusion matrix* dibuat menggunakan *sns.heatmap()* untuk menampilkan jumlah prediksi benar dan salah pada masing-masing kelas dalam bentuk peta panas yang mudah dianalisis. Pengaturan ukuran grafik serta penambahan judul membantu tampilan menjadi lebih jelas dan informatif. Hasil visualisasi ini penting untuk mengevaluasi kemampuan model dalam membedakan kategori layak dan tidak layak menerima bantuan. Selain akurasi secara keseluruhan, *confusion matrix* memberi pemahaman lebih detail mengenai pola kesalahan prediksi dan kontribusi masing-masing kelas terhadap performa model secara menyeluruh.

Berikut Merupakan Output setelah menjalankan Confusion Matrix :



Gambar 7 Output Confusion Matrix

Gambar 7 Output Confusion Matrix Menampilkan Bahwa *Confusion matrix* menunjukkan bahwa model optimal berhasil mengklasifikasikan sebagian besar data dengan benar. Pada kelas Tidak Layak, terdapat 29 prediksi benar dan hanya 3 kesalahan prediksi, sedangkan pada kelas Layak terdapat 54 prediksi benar dengan hanya 2 data yang salah diklasifikasikan. Hasil tersebut menegaskan bahwa model memiliki tingkat kesalahan rendah dan sangat baik dalam mengenali penerima bantuan yang layak, sehingga semakin memperkuat bahwa model *Decision Tree* yang telah dioptimasi memberikan prediksi yang stabil dan *reliabel* untuk mendukung penentuan kelayakan penerima bantuan KESRA..



Gambar 8 Visualisasi Pohon Keputusan

Seperti yang sudah ditampilkan Gambar 4.20 Visualisasi Pohon Keputusan pohon keputusan dibuat menggunakan *plot_tree()* untuk menampilkan struktur model *Decision Tree* hasil optimasi. Pengaturan ukuran tampilan dan penambahan nama fitur serta *label* kelas membantu pohon keputusan terlihat jelas, mudah dibaca, dan lebih informatif dalam menunjukkan hasil klasifikasi. Visualisasi ini penting karena menunjukkan alur pengambilan keputusan model, termasuk *fitur* yang paling berpengaruh dalam penentuan kelayakan bantuan. Dengan tampilan yang mudah dipahami, hasil model dapat diinterpretasikan langsung oleh peneliti maupun petugas terkait dalam proses pendukung keputusan KESRA.

Berikut Merupakan Output Setelah Menjalankan Visualisasi Pohon Keputusan :

Visualisasi yang ditampilkan Gambar 7 Output Visualisasi Pohon Keputusan memperlihatkan bahwa variabel *Aset* berperan sebagai pemecah pertama pada *root node*, sehingga menjadi faktor paling dominan dalam membedakan warga yang layak dan tidak layak menerima bantuan.

Pemecahan selanjutnya melibatkan variabel seperti *STATUS*, *Pendapatan*, serta atribut lainnya yang bersama-sama membentuk struktur klasifikasi bertingkat hingga menghasilkan *node* dengan tingkat *impurity* yang semakin rendah pada kedalaman pohon yang lebih dalam.

Pohon keputusan dengan kedalaman lima level ini menghasilkan beberapa *leaf node* dengan klasifikasi sempurna (*gini* = 0) dan menunjukkan kecenderungan model mengelompokkan lebih banyak data pada kategori "Tidak Layak", sesuai dengan distribusi awal *dataset*. Visualisasi ini memberikan gambaran yang jelas mengenai mekanisme pengambilan keputusan model berdasarkan fitur yang tersedia, serta menunjukkan alur klasifikasi yang digunakan dalam menentukan kelayakan penerima bantuan KESRA.

CONCLUSION

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai optimalisasi *Decision Tree* untuk penentuan kelayakan penerima bantuan kesejahteraan (KESRA) di Desa Wotgali, Kecamatan Plered, Kabupaten Cirebon, dapat ditarik beberapa kesimpulan yang menjawab rumusan masalah dan tujuan penelitian sebagai berikut:

Metode *Decision Tree* dengan pendekatan *CART (Classification and Regression Tree)* terbukti efektif dalam melakukan klasifikasi data untuk menentukan kelayakan penerima bantuan kesejahteraan (KESRA). Penerapan metode ini dilakukan secara sistematis melalui tahapan *Knowledge Discovery in Database (KDD)*, yang meliputi proses seleksi data, *preprocessing* untuk menangani *missing value* dan transformasi data kategorikal, hingga pembagian data menggunakan teknik *stratified sampling*. Hasil pengujian model awal (*baseline*) menunjukkan tingkat akurasi sebesar 90% dengan nilai *precision* dan *recall* yang cukup tinggi pada kedua kelas. Hal ini menunjukkan bahwa algoritma *Decision Tree* mampu mengidentifikasi pola hubungan antar variabel sosial ekonomi seperti pendapatan, jumlah anak, dan kepemilikan aset dalam menentukan kelayakan penerima bantuan. Dengan demikian, metode ini sudah cukup baik digunakan sebagai dasar dalam proses klasifikasi, meskipun masih memiliki potensi *overfitting* akibat kompleksitas struktur pohon yang terbentuk.

Proses optimasi model menggunakan *GridSearchCV* dengan pendekatan *cross-validation* serta penerapan teknik *cost-complexity pruning* terbukti mampu

meningkatkan kualitas model secara signifikan. Hasil optimasi menghasilkan parameter terbaik, yaitu *criterion* = *gini*, *max_depth* = 3, *min_samples_split* = 4, dan *min_samples_leaf* = 2, yang membentuk struktur pohon keputusan yang lebih sederhana, tidak terlalu dalam, dan lebih mudah diinterpretasikan. Selain itu, hasil analisis *feature importance* menunjukkan bahwa variabel pendapatan dan jumlah anak menjadi faktor paling dominan dalam menentukan kelayakan penerima bantuan dengan kontribusi yang sangat tinggi ($\geq 96\%$). Sementara itu, variabel lain yang bersifat administratif memiliki pengaruh yang relatif kecil terhadap hasil klasifikasi. Temuan ini menunjukkan bahwa model tidak hanya akurat secara statistik, tetapi juga relevan secara konseptual karena selaras dengan kondisi nyata di lapangan.

Model *Decision Tree* yang telah dioptimalkan menunjukkan peningkatan performa yang signifikan dibandingkan dengan model *baseline*. Hal ini ditunjukkan dengan meningkatnya akurasi dari 90% menjadi 94,3%, Selain mencapai akurasi sebesar 94,3%, model juga telah diuji menggunakan metode *cross-validation* dan menunjukkan hasil yang stabil dengan *standar deviasi* yang rendah. Hal ini membuktikan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik terhadap data baru. serta nilai *precision*, *recall*, dan *F1-score* yang berada pada rentang tinggi dan seimbang (0,91–0,96). Nilai *recall* yang tinggi pada kelas "Layak" menunjukkan bahwa model mampu meminimalkan kesalahan dalam mengklasifikasikan masyarakat yang seharusnya menerima bantuan (*false negative*), yang merupakan aspek krusial dalam konteks distribusi bantuan sosial. Selain itu, model hasil optimasi memiliki struktur yang lebih sederhana sehingga tidak mengalami *overfitting*, lebih stabil dalam melakukan prediksi, serta lebih mudah dipahami oleh pengguna *non-teknis* seperti petugas Puskesmas. Secara keseluruhan, model ini layak digunakan sebagai sistem pendukung keputusan yang objektif, akurat, dan transparan dalam menentukan penerima bantuan, sehingga dapat meningkatkan ketepatan sasaran dan kualitas penyaluran bantuan sosial di tingkat desa.

REFERENCE

- [1] W. D. S. Rajul Hakim, Muhammad Adnan, "Analysis Of Machine Learning Utilization In Identifying Social Assistance Recipients In Aceh Province," *JTEP Journal of tourism economics and policy*, vol. 4, no. 4, pp. 554–563, 2024.
- [2] P. Ladianto, "Implementation of Data Mining for Social Assistance Funds Using the C4 . 5 Algorithm in the Empat Lawang Regency

- Government," *Jurnal Ilmu Komputer dan Sistem Informasi [JIKSI]* Vol., vol. 4, no. 3, pp. 77–85, 2023.
- [3] A. I. Perhan, I. Yustiana, and I. Sanjaya, "Implementation of Machine Learning Using Decision Tree Method for Social Assistance Recipient Classification," *bit-Tech*, vol. 8, no. 1, pp. 901–909, 2025, doi: 10.32877/bt.v8i1.2755.
- [4] R. D. Mandasari and H. Hartana, "Implementation Of Decision Tree Algorithm For Classification Of Eligibility In Social Assistance Fund Distribution," *TIERS Information Technology Journal*, vol. 5, no. 1, pp. 34–40, 2024, doi: 10.38043/tiers.v5i1.5378.
- [5] D. Fitriza, Atus Amadi Putra, Dodi Vionanda, and Zilrahmi, "Classification of Poor Households in West Sumatra Province using Decision Tree Algorithm C4.5," *UNP Journal of Statistics and Data Science*, vol. 2, no. 2, pp. 145–150, 2024, doi: 10.24036/ujsds/vol2-iss2/157.
- [6] C. Satria, P. W. Sugijanto, A. Anggrawan, I. N. Y. Sumadewa, A. D. Dayani, and R. Anggriani, "Multi-Algorithm Approach to Enhancing Social Assistance Efficiency Through Accurate Poverty Classification," *Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 24, no. 1, pp. 167–178, 2024, doi: 10.30812/matrik.v24i1.4275.
- [7] J. Huyghe, J. Trufin, and M. Denuit, "Boosting cost-complexity pruned trees on Tweedie responses: the ABT machine for insurance ratemaking," *Scandinavian Actuarial Journal*, vol. 2024, no. 5, pp. 417–439, 2024, doi: 10.1080/03461238.2023.2258135.
- [8] M. F. Balcan and D. Sharma, "Learning Accurate and Interpretable Decision Trees," *Proceedings of Machine Learning Research*, vol. 244, pp. 288–307, 2024, doi: 10.24963/ijcai.2025/1205.
- [9] S. Siddique, M. A. Haque, R. George, K. D. Gupta, D. Gupta, and M. J. H. Faruk, "Survey on Machine Learning Biases and Mitigation Techniques," *Digital*, vol. 4, no. 1, pp. 1–68, 2024, doi: 10.3390/digital4010001.