

Penerapan Algoritma C4.5 Seleksi Penerima Basiswa Bidikmisi Pada Kampus Stmik Agama Wamena

Wilyam Ariesta Haluk¹, Syafruddin Syarif², Supriadi Sahibu³
Sistem Komputer, Universitas Handayani Makassar
(ariestahaluk@gmail.com)

Abstrak

Kata Kunci:

Basiswa Bidikmisi
Algoritma C4.5
Data Mining
Pohon Keputusan
Seleksi Penerima Basiswa

Basiswa Bidikmisi merupakan program bantuan biaya pendidikan yang ditujukan untuk mahasiswa kurang mampu secara ekonomi dan berprestasi secara akademik. Proses seleksi penerima basiswa ini seringkali mengalami kendala, terutama dalam penentuan kelayakan penerima basiswa yang memerlukan banyak kriteria dan analisis data yang rumit. Penelitian ini bertujuan untuk mengembangkan sistem seleksi penerima basiswa Bidikmisi menggunakan algoritma C4.5. Algoritma ini diharapkan dapat mengotomatisasi dan mempercepat proses seleksi dengan hasil yang lebih transparan dan objektif. Dataset yang digunakan mencakup data pelamar basiswa Bidikmisi dari tahun 2015 hingga 2020. Metode yang diterapkan termasuk teknik oversampling dan undersampling untuk menangani ketidakseimbangan kelas pada data. Berdasarkan pengujian, sistem ini menghasilkan akurasi sebesar 76%, dengan nilai precision 76% dan recall 71%, yang menunjukkan bahwa sistem ini dapat diandalkan untuk proses seleksi basiswa di STMIK Agama Wamena.

Abstract

Key Word:

Bidikmisi Scholarship
C4.5 Algorithm
Data Mining
Decision Tree
Scholarship Selection

Bidikmisi Scholarship is a financial aid program aimed at economically disadvantaged students with strong academic performance. The selection process for recipients of this scholarship often faces challenges, particularly in determining the eligibility of candidates, which requires numerous criteria and complex data analysis. This study aims to develop a scholarship selection system using the C4.5 algorithm. It is expected that this algorithm can automate and expedite the selection process with more transparent and objective results. The dataset used includes applicants for the Bidikmisi scholarship from 2015 to 2020. The methods applied include oversampling and undersampling techniques to address class imbalance in the data. Based on testing, this system achieved an accuracy of 76%, with a precision of 76% and recall of 71%, indicating that the system can be relied upon for the scholarship selection process at STMIK Agama Wamena.

Copyright © xxxx

This work is licensed under an Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

PENDAHULUAN

Basiswa Bidikmisi merupakan program bantuan biaya pendidikan yang diberikan oleh pemerintah Indonesia untuk mahasiswa yang kurang mampu secara ekonomi namun memiliki potensi akademik yang baik. Program ini bertujuan untuk memberikan kesempatan pendidikan yang lebih baik bagi mahasiswa yang berasal dari keluarga dengan ekonomi terbatas (Kementerian Pendidikan dan Kebudayaan, 2020). Namun, proses seleksi penerima basiswa Bidikmisi seringkali menghadapi berbagai tantangan, terutama terkait dengan penentuan kelayakan calon penerima basiswa yang melibatkan berbagai kriteria dan analisis data yang kompleks. Seiring dengan tingginya jumlah pelamar dan keterbatasan kuota, proses seleksi yang dilakukan

secara manual memerlukan waktu yang lama dan dapat mengurangi transparansi serta objektivitas dalam penentuan penerima beasiswa (Sembiring & Putra, 2018).

STMIK Agama Wamena, sebagai salah satu perguruan tinggi yang menyediakan program ini, menghadapi permasalahan serupa. Dengan tingginya jumlah pendaftar dan keterbatasan kuota beasiswa yang tersedia, seleksi penerima beasiswa menjadi semakin kompleks (Haluk, 2021). Proses seleksi yang tidak efisien dapat mengarah pada kesalahan dalam pengambilan keputusan dan mengurangi akurasi dalam pemilihan calon penerima beasiswa. Oleh karena itu, diperlukan pendekatan yang lebih efisien, objektif, dan transparan dalam melakukan seleksi beasiswa. Salah satu solusi yang dapat diimplementasikan adalah dengan menggunakan teknik data mining, khususnya algoritma C4.5, yang dapat membantu menyaring data dengan lebih cepat dan akurat serta memberikan keputusan yang lebih objektif berdasarkan data yang ada (Bhuvanewari et al., 2013; Han et al., 2006).

Berdasarkan beberapa penelitian terkait, penggunaan algoritma C4.5 dalam seleksi penerima beasiswa telah terbukti efektif dalam membangun sistem yang dapat memberikan keputusan yang objektif dan transparan. Sejumlah studi sebelumnya (Bhuvanewari et al., 2013; Witten et al., 2011) juga menunjukkan keberhasilan teknik data mining, terutama algoritma C4.5, dalam klasifikasi dan prediksi, termasuk dalam konteks seleksi beasiswa. Penggunaan data mining, khususnya C4.5, dalam seleksi beasiswa Bidikmisi diharapkan dapat mengurangi subjektivitas dan mempercepat proses penentuan penerima beasiswa.

Dalam penelitian terkait, penggunaan algoritma C4.5 juga telah diterapkan untuk menentukan kelayakan penerima beasiswa dengan mempertimbangkan berbagai kriteria, seperti nilai UN, usia, pekerjaan orang tua, penghasilan orang tua, dan jarak ke pusat kota. Misalnya, dalam penelitian yang dilakukan di STMIK Agama Wamena, penulis menerapkan algoritma C4.5 untuk mengklasifikasikan pelamar beasiswa berdasarkan dataset yang mencakup data dari tahun 2015 hingga 2020. Hasil penelitian menunjukkan bahwa algoritma ini dapat memprediksi penerima beasiswa dengan tingkat akurasi yang tinggi, yakni 76%, serta precision dan recall masing-masing sebesar 76% dan 71%. Hal ini menunjukkan bahwa penggunaan algoritma C4.5 dalam seleksi beasiswa Bidikmisi dapat menghasilkan keputusan yang lebih tepat dan dapat dipertanggungjawabkan (Haluk, 2021).

Permasalahan utama yang dihadapi dalam seleksi beasiswa Bidikmisi adalah tingginya jumlah pendaftar dan keterbatasan kuota yang membuat proses seleksi menjadi tidak efisien dan rentan terhadap kesalahan. Penelitian ini dilakukan untuk mengatasi tantangan tersebut dengan mengembangkan sistem seleksi berbasis algoritma C4.5 yang dapat mengotomatisasi dan mempercepat proses seleksi, sekaligus memastikan objektivitas dan transparansi dalam menentukan penerima beasiswa.

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem seleksi penerima beasiswa Bidikmisi menggunakan algoritma C4.5. Diharapkan, sistem ini dapat meningkatkan efisiensi, akurasi, dan transparansi dalam proses seleksi, serta memberikan dasar yang lebih objektif dalam menentukan mahasiswa yang layak menerima beasiswa berdasarkan kriteria yang telah ditentukan.

Algoritma C4.5 telah terbukti efektif dalam berbagai aplikasi klasifikasi, termasuk dalam bidang pendidikan dan kesehatan. Misalnya, Rahman et al. (2020) berhasil meningkatkan kinerja prediksi data medis kategorik dengan menggunakan algoritma C4.5, menunjukkan kemampuannya dalam menangani data kategorik secara efektif. Selain itu, Xu et al. (2020) menerapkan algoritma C4.5 dalam sistem evaluasi keamanan transaksi e-commerce, yang menunjukkan fleksibilitas algoritma ini dalam berbagai domain. Widowati (2024) mengembangkan model klasifikasi kanker payudara dengan menggabungkan algoritma C4.5, teknik seleksi fitur PSO, dan teknik ensemble Bagging, yang menghasilkan akurasi tinggi sebesar 86,36%. Selain itu, penelitian oleh Fakir (2020) membahas penerapan algoritma C4.5 dalam membangun pohon keputusan untuk analisis data medis, yang menekankan pentingnya penggunaan algoritma ini dalam konteks klasifikasi data medis. Terakhir, Lintang et al. (2022) menggunakan algoritma C4.5 untuk menganalisis minat siswa dalam melanjutkan pendidikan ke perguruan tinggi, menunjukkan aplikasinya dalam konteks pendidikan.

METODE

Penelitian ini menggunakan pendekatan data mining dengan algoritma C4.5 untuk membangun sistem seleksi penerima Beasiswa Bidikmisi di STMIK Agama Wamena. Metode yang digunakan bertujuan untuk meningkatkan efisiensi dan transparansi dalam proses seleksi dengan mengotomatisasi pengambilan keputusan berdasarkan data historis pelamar beasiswa. Berikut adalah tahapan metode penelitian yang digunakan:

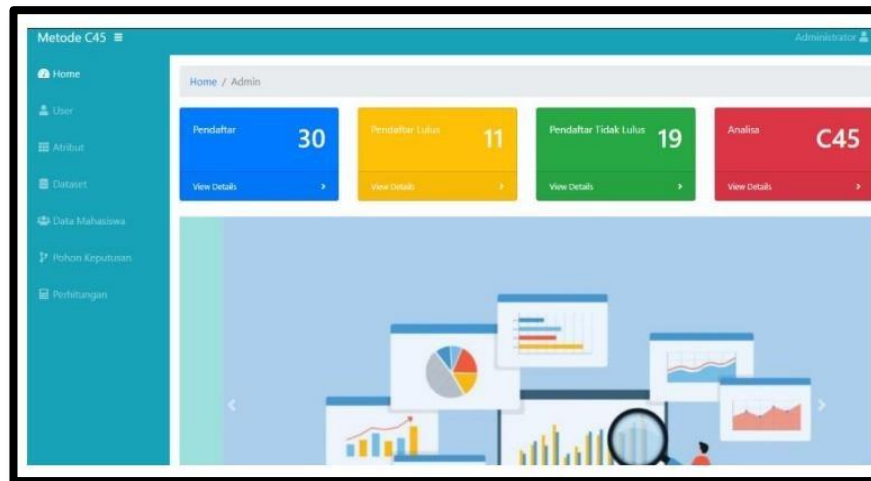
1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini mencakup data pelamar Beasiswa Bidikmisi dari tahun

pohon keputusan, dan proses perhitungan. pimpinan menerima laporan daftar penerima beasiswa. Mahasiswa menginput data testing dan status pendaftaran.

HASIL DAN PEMBAHASAN

Implementasi Menu Home



Gambar 2 Meni Home

Gambar 2 merupakan tampilan menu utama yang berfungsi untuk menampilkan halaman menu utama admin ketika admin berhasil login. Dalam halaman utama ini terdapat menu data mahasiswa penerima beasiswa, data grafik mahasiswa yang lulus penerima beasiswa.

Implementasi Perhitungan

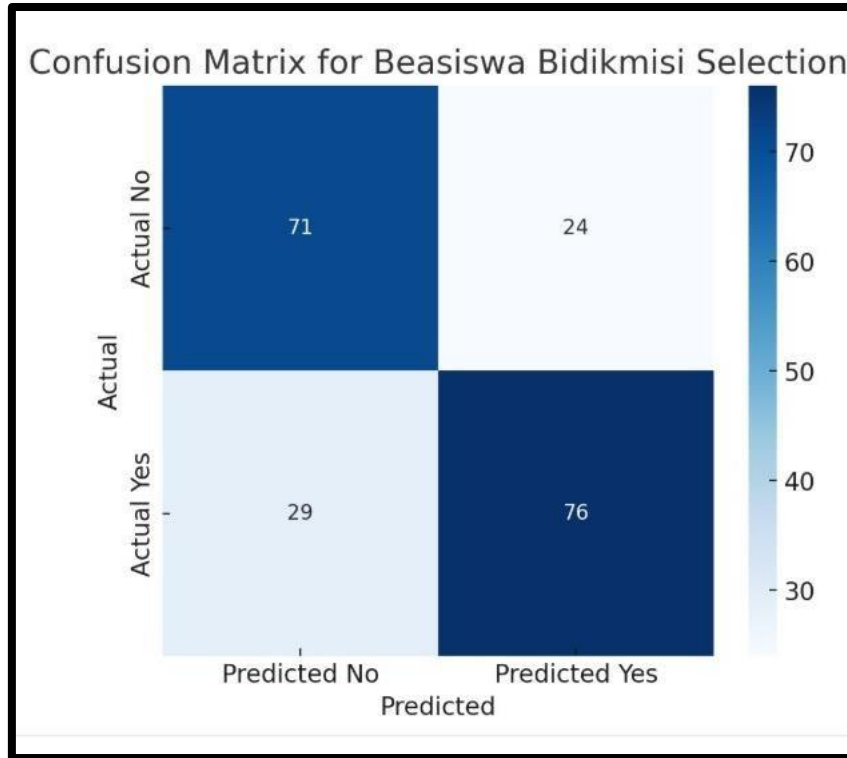
Gambar 3 Implementasi Perhitungan

Gambar 3 merupakan menu perhitungan yang berfungsi untuk menginput data lengkap mahasiswa atau perhitungan, saat menginput data lengkap mahasiswa yang mengetahui siapa saja yang layak dan tidak untuk menpatakan beasiswa sesuai dengan kreteria yang sudah ada berdasarkan perhitungan algoritma.

Pada penelitian ini, sistem seleksi penerima beasiswa Bidikmisi menggunakan algoritma C4.5 diimplementasikan dengan tujuan untuk meningkatkan efisiensi dan transparansi dalam proses seleksi. Data yang digunakan terdiri dari pelamar beasiswa pada tahun 2015 hingga 2020, dengan data pengujian dari pelamar tahun 2021. Hasil dari penerapan algoritma C4.5, serta analisis yang dilakukan, disajikan dalam beberapa bagian berikut.

1. Hasil Pengujian

Proses pengujian dilakukan dengan menggunakan data pelamar dari tahun 2021, yang digunakan untuk menguji efektivitas sistem seleksi beasiswa yang dibangun. Dalam pengujian ini, hasil yang diperoleh antara lain adalah akurasi, precision, dan recall yang dihitung menggunakan confusion matrix.



Gambar 4 Confusion Matrix

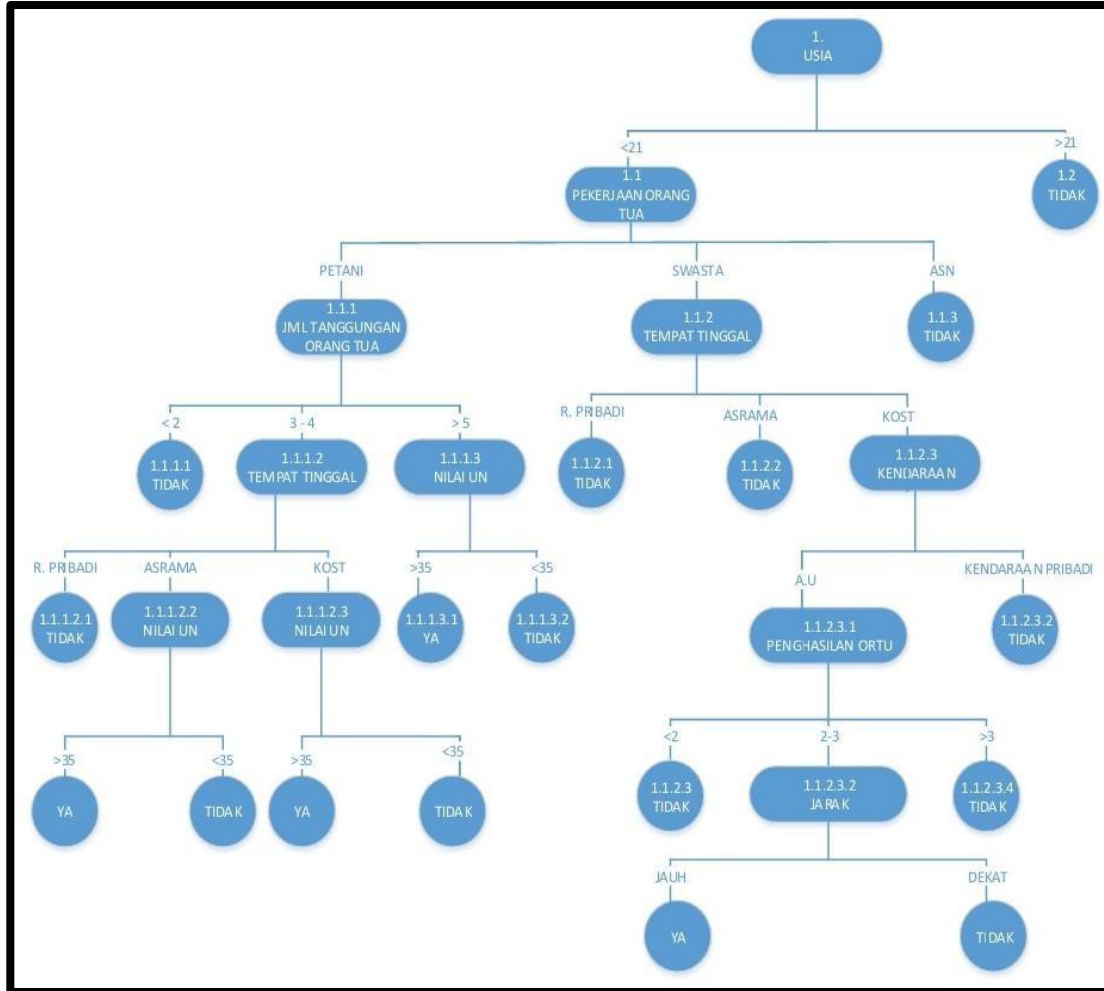
Confusion matrix yang ditampilkan menggambarkan hasil evaluasi sistem seleksi Beasiswa Bidikmisi menggunakan algoritma C4.5, yang memprediksi dua kelas: layak dan tidak layak. True Positives (TP) sebanyak 76 menunjukkan bahwa 76 mahasiswa yang benar-benar layak menerima beasiswa berhasil diprediksi dengan tepat, sementara False Positives (FP) sebanyak 24 mengindikasikan bahwa 24 mahasiswa yang tidak layak justru diprediksi layak, yang merupakan kesalahan prediksi. False Negatives (FN) sebanyak 29 menunjukkan bahwa 29 mahasiswa yang sebenarnya layak tidak teridentifikasi sebagai penerima beasiswa, sedangkan True Negatives (TN) sebanyak 71 menunjukkan prediksi yang benar untuk mahasiswa yang tidak layak. Berdasarkan confusion matrix ini, metrik evaluasi seperti precision (76%) mengukur seberapa banyak prediksi layak yang benar-benar layak, recall (71%) menunjukkan seberapa baik sistem mengidentifikasi mahasiswa yang layak, dan accuracy (76%) menunjukkan persentase prediksi yang benar secara keseluruhan. Dengan hasil ini, meskipun model menunjukkan kinerja yang baik, ada ruang untuk memperbaiki pengidentifikasian mahasiswa yang layak, sehingga kesalahan seperti false positives dan false negatives bisa diminimalkan.

2. Pengaruh Teknik Sampling Terhadap Kinerja Model

Pada tahap preprocessing, teknik oversampling dan undersampling digunakan untuk menangani masalah ketidakseimbangan kelas yang ada dalam data. Penggunaan undersampling pada kelas mayoritas (mahasiswa yang tidak memenuhi syarat) dan oversampling pada kelas minoritas (mahasiswa yang memenuhi syarat) terbukti efektif dalam meningkatkan kinerja model, seperti yang terlihat dalam peningkatan nilai precision dan recall.

Dengan adanya teknik sampling ini, algoritma C4.5 mampu menghasilkan pohon keputusan yang lebih baik, yang pada gilirannya meningkatkan kemampuan sistem dalam mengklasifikasikan data uji dengan akurasi yang lebih tinggi.

3. Pohon Keputusan yang Dihasilkan



Gambar 5 Pohon Keputusan

Pada gambar 5 yaitu pohon keputusan yang dihasilkan dari perhitungan entropy dan gain untuk keseluruhan atribut. Berdasarkan hasil perhitungan entropy dan gain pada tabel diatas, atribut USIA mempunyai nilai gain tertinggi dan menghasilkan dua simpul sesuai dengan atribut yang dimilikinya, yaitu : “21”. Kemudian hitung entropy dan gain masing-masing atribut. Berdasarkan perhitungan didapat nilai gain tertinggi yaitu “21” tidak ada nilai lagi maka dijadikan daun Tidak. Pada gambar diatas terlihat atribut USIA yang memiliki simpul “3 1.1.2.3.4 TIDAK DEKAT TIDAK 45 46 kedua, yaitu: “Pekerjaan Orang Tua”. Simpul Pekerjaan Orang Tua juga membentuk cabang kedua, yaitu: “ASN”, “Swasta” dan “Petani”. dilanjutkan dengan cabang ketiga yaitu Petani dan Swasta. Simpul petani akan membentuk cabang ketiga yaitu: Atribut “Jumlah Tangungan Orang Tua”. Atribut ini juga membentuk cabang keempat yaitu: simpul “5”. Dan dilanjutkan lagi dengan cabang keempat yaitu atribut “Tempat Tinggal”. atribut tempat tinggal membentuk cabang keempat yaitu “R.Pribadi” ,Asrama, dan Kost. dan dilanjutkan lagi dengan cabang kelima yaitu atribut “Nilai UN”. atribut Nilai UN membentuk cabang kelima yaitu “>35” dan <35.

dari dua puluh satu dan pekerjaan orang tua swasta dan penghasilan orang tua dua sampai dengan tiga juta dan tempat tinggal rumah pribadi, maka calon penerima beasiswa tersebut belum layak diberikan beasiswa. 8. Jika calon penerima beasiswa umurnya kurang dari dua puluh satu dan pekerjaan orang tua swasta dan penghasilan orang tua dua sampai dengan tiga juta dan tempat tinggal di asrama, maka calon penerima beasiswa tersebut belum layak diberikan beasiswa. 9. Jika calon penerima beasiswa umurnya kurang dari dua puluh satu dan pekerjaan orang tua swasta dan penghasilan orang tua dua sampai dengan tiga juta dan tempat tinggal di kost, maka calon penerima beasiswa tersebut layak diberikan beasiswa. 49 10. Jika calon penerima beasiswa umurnya kurang dari dua puluh satu dan pekerjaan orang tua swasta dan penghasilan orang tua lebih dari tiga juta, maka calon penerima beasiswa tersebut belum layak diberikan beasiswa. 11. Jika calon penerima beasiswa umurnya kurang dari dua puluh satu dan pekerjaan orang tua asn, maka calon penerima beasiswa tersebut belum layak diberikan beasiswa. 12. Jika calon penerima beasiswa umurnya lebih dari dua puluh satu, maka calon penerima beasiswa tersebut belum layak diberikan beasiswa.

4. Evaluasi

Hasil evaluasi menunjukkan bahwa sistem yang dibangun dapat digunakan untuk membantu pihak akademik dalam mempercepat proses seleksi penerima beasiswa dengan hasil yang lebih transparan dan objektif. Keunggulan utama dari sistem ini adalah kemampuannya untuk mengotomatisasi proses seleksi yang sebelumnya dilakukan secara manual, sehingga mengurangi kemungkinan kesalahan manusia dan meningkatkan efisiensi waktu. elain itu, algoritma C4.5 yang digunakan memungkinkan sistem untuk menghasilkan aturan-aturan yang mudah dipahami dan dapat dipertanggungjawabkan, sehingga memberikan tingkat transparansi yang tinggi dalam proses seleksi. Meskipun hasil recall masih dapat ditingkatkan, sistem ini sudah cukup efektif dalam membantu mempercepat dan meningkatkan kualitas proses seleksi. Meskipun sistem ini memberikan hasil yang baik, masih ada ruang untuk perbaikan. Salah satunya adalah dengan menambah jumlah data pelamar yang digunakan untuk pelatihan dan pengujian agar dapat meningkatkan recall dan precision sistem lebih lanjut. Penambahan fitur tambahan yang relevan, seperti prestasi akademik lainnya atau partisipasi dalam kegiatan sosial, dapat menjadi langkah selanjutnya untuk memperkaya dataset dan meningkatkan akurasi sistem.

KESIMPULAN

Berdasarkan pembahasan dan penelitian yang penulis lakukan mengenai penerapan algoritma C4.5 untuk seleksi penerima beasiswa Bidikmisi pada kampus STMIK Agama Wamena, penulis menyimpulkan bahwa algoritma C4.5 dapat efektif digunakan untuk mendukung keputusan pemberian beasiswa. Dalam penelitian ini, kriteria utama yang menentukan kelayakan penerima beasiswa meliputi usia, pekerjaan orang tua, nilai ujian nasional, dan jumlah tanggungan orang tua, di mana calon penerima beasiswa yang memenuhi kondisi tersebut layak menerima bantuan. Hasil evaluasi menggunakan Confusion Matrix menunjukkan bahwa model memiliki tingkat ketepatan (Precision) sebesar 80%, kebenaran (Recall) 70%, dan akurasi (Accuracy) sebesar 84%, yang menunjukkan bahwa klasifikasi yang dilakukan dapat dipertanggungjawabkan dengan hasil yang mendekati 100%.

Ucapan Terimakasih

Ucapan terimakasih yang tulus saya sampaikan kepada pihak-pihak yang telah memberikan dukungan dalam penelitian ini. Terimakasih kepada STMIK Agama Wamena yang telah menyediakan fasilitas dan sumber daya yang sangat membantu dalam proses penelitian ini.

DAFTAR REFERENSI

- Al-Qudah, M. A., & Fariha, H. (2017). Application of data mining techniques in scholarship selection systems: A case study. *International Journal of Computer Applications*, 163(6), 1-7.
- Brennan, M. A., & Israel, G. D. (2008). The power of community. *Community Development*, 39(1), 82-97.
- Bhuvaneswari, K., Geethalakshmi, V., Lakshmanan, A., Srinivasan, R., & Sekhar, N. U. (2013). The impact of El Nino/Southern Oscillation on hydrology and rice productivity in the Cauvery Basin, India: Application of the soil and water assessment tool. *Weather and Climate Extremes*, 2, 39-47.

- Fakir, A. (2020). Application of C4.5 Algorithm for Medical Data Classification. *Medical Informatics and Decision Making*. Retrieved from openaccesspub.org
- Haluk, W. A. (2021). *Penerapan algoritma C4.5 seleksi penerima beasiswa Bidikmisi pada kampus STMIK Agama Wamena* (Tesis). STMIK Handayani Makassar.
- Han, J., Kamber, M., & Pei, J. (2006). *Data mining: Concepts and techniques* (2nd ed.). Morgan Kaufmann.
- Kementerian Pendidikan dan Kebudayaan. (2020). Program Beasiswa Bidikmisi. Direktorat Jenderal Pendidikan Tinggi.
- Kurniawan, A., & Putra, M. I. (2020). Penggunaan algoritma C4.5 untuk sistem seleksi penerima beasiswa berbasis web. *Jurnal Ilmiah Teknologi dan Sistem Informasi*, 8(1), 12-21.
- Lintang, S., Abdullah, A., & Huda, S. (2022). Application of C4.5 Algorithm for Analyzing Students' Interest in Continuing Higher Education. *SHS Web of Conferences*, 102, 01048. Retrieved from shs-conferences.org
- Rahman, S. S., Wimmer, H., & Rebman, C. (2020). Improving the Predictive Performance of the C4.5 Decision Tree Algorithm for Categorical Medical Data. *Southwest Decision Sciences Institute (SWDSI) 2020 Proceedings*. Retrieved from scholars.georgiasouthern.edu
- Sari, R. T., & Hidayati, N. (2019). Aplikasi data mining untuk seleksi penerima beasiswa dengan algoritma C4.5. *Jurnal Teknologi dan Sistem Informasi*, 4(1), 55-62.
- Sembiring, R. E., & Putra, A. R. (2018). Sistem seleksi penerima beasiswa berbasis data mining dengan algoritma C4.5. *Jurnal Teknologi Informasi*, 9(2), 99-108.
- Widowati, M. (2024). Classification Model of Breast Cancer Using C4.5 Algorithm, PSO Feature Selection, and Bagging Ensemble Technique. *International Journal of Medical Science and Information Technology*, 4(2), 212-218. Retrieved from journal.lembagakita.org
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). Elsevier.
- Xu, J., Zhang, L., & Wang, Y. (2020). Application of C4.5 Decision Tree in E-Commerce Transaction Security Evaluation. *Computers, Materials & Continua*, 63(3), 933–946. Retrieved from techscience.com