

IMPLEMENTASI ALGORITMA C4.5 CLASIFIKASI CALON PENERIMA BEASISWA

IMPLEMENTATION ALGORITHM C4.5 CLASSIFICATION OF PROSPECTIVE SCHOLARSHIP RECIPIENTS

Yadi

¹Prodi Teknik Informatika, Institut Teknologi Pagaralam, Sumatera Selatan
Jl. Masik Siagim, No 75 Simpang Mbacang, Dempo Tengah, Kota Pagaralam
E-mail: 1*yadimkom@itpa.ac.id

ABSTRAK

Beasiswa merupakan salah satu bentuk apresiasi yang diberikan kepada mahasiswa agar dapat melanjutkan proses pembelajaran tanpa harus memikirkan biaya yang besar, *beasiswa* diberikan dapat berupa beasiswa prestasi bidang akademik maupun non akademik serta beasiswa keluarga sejahtera yang diberikan oleh pemerintah, maupun yayasan tempat mahasiswa belajar. *Beasiswa* yang diberikan berdampak pada tingkatnya keinginan masyarakat yang memiliki prestasi untuk memiliki peluang yang besar dalam melaksanakan proses menempuh pendidikan sampai pada tingkat perguruan tinggi pada kondisi ini keinginan pemerintah bahwa pendidikan wajib 12 tahun dapat dituntaskan dengan sangat baik melalui program-program yang telah tersusun secara berkelanjutan. Tujuan penelitian untuk melakukan klasifikasi calon penerima beasiswa dengan menggunakan algoritma C4.5. Metodologi penelitian mengimplementasikan algoritma C4.5 dengan hasil penelitian klasifikasi dalam prediksi calon penerima *beasiswa* dengan beberapa indikator diantaranya Pendapatan orang tua, Jumlah Tanggungan, Nilai UN, Nilai Tes. Klasifikasi serta perhitungan yang dipergunakan algoritma C4.5 dengan prediksi ketepatan akurasi mencapai 85% dan digambarkan berdasarkan pada decision tree.

Kata kunci: Algoritma C4.5, Beasiswa, Implementasi, Klasifikasi

ABSTRACT

Scholarships are a form of appreciation given to students so that they can continue the learning process without having to think about large costs, scholarships can be in the form of academic and non-academic achievement scholarships as well as prosperous family scholarships provided by the government, as well as foundations where students study. The scholarships given have an impact on the level of desire of people who have achievements to have great opportunities in carrying out the process of taking education up to the tertiary level. . The purpose of this research is to classify prospective scholarship recipients using the C4.5 algorithm. The research methodology implements the C4.5 algorithm with the results of classification research in the prediction of prospective scholarship recipients with several indicators including Parents' income, number of dependents, UN scores, and test scores. Classification and calculations used the C4.5 algorithm with prediction accuracy reaching 85% accuracy and described based on the decision tree.

Keywords: Algorithm C4.5, Clasification, Implementation, Scholarship

PENDAHULUAN

Panitia seleksi penerimaan mahasiswa baru merupakan unsur utama dalam melakukan seleksi mahasiswa serta sebagai unsur yang harus dapat melihat calon-calon Penerima beasiswa

harus mempertimbangkan beberapa faktor sebagai penentu untuk memastikan bahwa sasaran dalam penerima beasiswa memang layak. Metode data mining untuk klasifikasi dapat digunakan untuk membantu meningkatkan kecepatan dan ketepatan dalam prediksi calon penerima

beasiswa, Proses seleksi penerima beasiswa pada perguruan tinggi mengacu pada ketentuan yang telah ditetapkan oleh pemerintah sesuai dengan karakteristik dan kondisi dari perguruan tinggi pelaksana [1].

Beasiswa sebagai bentuk apresiasi yang diberikan dengan tujuan agar proses pembelajaran dapat dilanjutkan pada jenjang yang lebih tinggi [2]. Sehingga beasiswa tersebut dapat memberikan dampak positif bagi penerima, antara lain: Mereka tidak perlu lagi memikirkan tagihan biaya pendidikan, dan lebih fokus untuk belajar serta berusaha mendapatkan nilai dan prestasi yang tinggi [3].

Penerapan algoritma C4.5 dalam melakukan prediksi sudah memiliki akurasi yang cukup baik dalam melakukan penyeleksian terhadap data dengan prosedur yang terstruktur dengan baik melalui beberapa tahapan diantaranya proses klasifikasi dengan data mining [4]. Decision tree yang dihasilkan melalui perhitungan algoritma C4.5 sebagai penarik sebuah keputusan menghasilkan sebuah informasi yang dipergunakan untuk node dalam akar pohon sehingga terbentuk sebuah pohon keputusan dengan menggunakan perhitungan atribut [5].

Algoritma C4.5 terdiri dari beberapa variabel dan atribut yang dipergunakan untuk mendapatkan informasi atau pengetahuan tentang suatu aturan yang sangat mudah dipahami dengan penggambaran sebuah pohon keputusan [6][7]. Dengan klasifikasi penggunaan algoritma C4.5 sebagai pengolahan data untuk melakukan prediksi berdasarkan pada parameter-parameter yang telah ditentukan melalui tahap pruning sehingga terbentuk pohon keputusan untuk menguji akurasi algoritma [8], dengan menggunakan algoritma dalam proses pemberian rekomendasi penerima beasiswa lebih tepat sasaran serta terkonsep berdasarkan pada indikator [9] proses pencarian nilai pada pohon keputusan yang dihasilkan melalui perhitungan dengan menggunakan nilai gain dan perhitungan

entropy akurat [10]. Penerapan algoritma dengan menerapkan teknik data mining metode decision tree algoritma C4.5 dapat memberikan informasi berupa klasifikasi [11][12].

Proses klasifikasi terhadap seleksi calon penerima beasiswa yang terjadi selama ini hanya melihat dari pekerjaan orang tua saja, sehingga pada proses penerimaan beasiswa yang terjadi selama ini terdapat beberapa kendala, terlihat pada hasil pembelajaran mahasiswa yang mendapatkan beasiswa cenderung malas untuk mengikuti proses perkuliahan sehingga prestasi pada pencapaian hasil belajar belum sepenuhnya terpenuhi, maka pada kondisi ini perlu dilakukan klasifikasi calon penerima beasiswa yang baik berdasarkan beberapa indikator agar tujuan pada penyaluran beasiswa benar-benar tepat. Algoritma C4.5 dengan *decision tree* memiliki struktur kerja mulai dari root menuju paling daun, jika diuji dengan data uji [13] [14], dari nilai prediksi yang didapatkan sebagai tolok ukur untuk menarik sebuah kesimpulan berdasarkan data atribut [15]. Semakin banyak data sampel yang digunakan maka semakin kuat juga kebenarannya dalam menggunakan algoritma C4.5 [16] penerapan algoritma C4.5 juga dapat dikombinasikan dengan menggunakan beberapa bahasa pemrograman dalam mengelola data untuk melakukan prediksi [17] [18] akurasi prediksi yang dilakukan menggunakan algoritma C4.5 mendapatkan tingkat akurasi yang baik [19][20]. Klasifikasi dengan menggunakan algoritma sangat baik dalam melakukan prediksi hal ini disebabkan adanya perhitungan secara matematis dan logis berdasarkan pada kriteria-kriteria informasi dengan perhitungan sesuai pada aturan pada proses pengambilan keputusan [21], Kelebihan dari algoritma C4.5 dapat menangani atribut *continue* dan diskrit beserta pengolahan training data dengan *missing value* untuk memperbaiki informasi pada gain menggunakan gain ratio. Oleh sebab itu implementasi algoritma C4.5 dirasakan sangat baik

dalam menentukan calon penerima beasiswa.

METODE

Metode penelitian menggunakan pendekatan deskriptif kuantitatif dengan menggunakan tahapan diantaranya Pengumpulan data, Pengolahan data awal, Implementasi algoritma C4.5, Pengujian Algoritma C4.5 dan Kesimpulan. Tahapan penelitian terlihat pada Gambar 1.

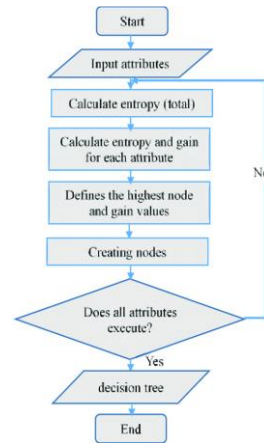


Gambar 1. Tahapan Penelitian [2]

Berdasarkan pada Gambar 1 tahapan penelitian dilakukan sebagai berikut: **1) Pengumpulan data.** Proses pengumpulan data untuk mendapatkan data awal dari permasalahan melalui pengumpulan data berupa dokumen hasil registrasi mahasiswa berupa identitas mahasiswa, nilai ujian nasional, serta ujian tes masuk. **2) Pengolahan data awal.** Pengolahan data awal untuk menentukan atribut-atribut untuk diklasifikasikan melalui perhitungan algoritma C4.5. **3) Implementasi algoritma C4.5.** Proses implementasi algoritma C4.5 dapat dilihat pada Gambar 2. Proses implementasi algoritma melakukan perhitungan entropy serta gain untuk menentukan dataset serta decision tree pada data yang telah diolah. **4) Pengujian algoritma C4.5.** Proses pengujian algoritma C4.5 untuk mengetahui seberapa besar tingkat akurasi prediksi yang dihasilkan melalui perhitungan rumus berdasarkan data-data yang telah didapatkan sehingga terlihat node pada atribut. **5) Kesimpulan.** Penarikan sebuah kesimpulan untuk mengetahui hasil akurasi calon penerima beasiswa dengan menggunakan algoritma C4.5.

Implementasi algoritma C4.5 melalui beberapa tahapan diantaranya adanya sebuah atribut yang dipergunakan sehingga perhitungan pada nilai entropy dapat di perhitungan berdasarkan pada rumus untuk melihat nilai gain pada

atribut proses implementasi algoritma C4.5 terlihat pada Gambar 2.



Gambar 2. Flowchart Algoritma C4.5 [1]

Berdasarkan pada *Flowchart Algoritma C45* proses dalam melakukan klasifikasi dimulai dengan melakukan pemilihan atribut untuk diinputkan yang selanjutnya dilakukan perhitungan pada nilai *entropy* dengan menggunakan rumus gain terhadap atribut yang telah ditetapkan, proses perhitungan entropy dan nilai gain akan membentuk sebuah nodes yang menghasilkan sebuah pohon keputusan berdasarkan pada perhitungan algoritma dengan ketentuan penginputan nilai attribute yang telah ditetapkan.

Perhitungan entropy bertujuan untuk memisahkan klasifikasi data dengan menggunakan rumus:

$$Entropy (X) = \sum_{p_j=1}^k P_j \cdot \log_2 \frac{1}{P_j} = - \sum_{p_j=1}^k P_j \cdot \log_2 P_j$$

Keterangan:

X : Kelompok Data

k : Jumlah Partisi X

P_j : Proporsi P_j terhadap K

Nilai Gain bertujuan untuk mengetahui selisih nilai dari setiap atribut dan node dengan rumus sebagai berikut:

$$Gain\ ratio (a) = Entropy (x) - \sum_{i=1}^N \frac{X_i}{X} \cdot Entropy (X_i)$$

Keterangan:

S : Kelompok Data

A : Atribut

N : Jumlah partisi atribut A

|X_i| : Jumlah kasus pada atribut ke-i

|X| : Jumlah kasus dalam X

Setelah perhitungan dengan menggunakan rumus entropy dan gain

dilakukan selanjutnya akan terbentuk sebuah node dari hasil perhitungan dengan menggunakan beberapa atribut yang akan dipergunakan untuk mendapatkan pohon keputusan.

HASIL DAN PEMBAHASAN

Metdoe dalam implementasi algoritma C45 terlihat pada penentuan atribut terlebih dahulu agar proses kalkulasi dan klasifikasi dapat dilakukan terlihat pada Tabel 1 proses klasifikasi atribut yang akan dipergunakan dalam penetapan calon penerima beasiswa.

Algoritma C4.5 salah satu algoritma yang dapat diimplementasikan untuk melakukan prediksi terhadap hasil yang akurat berdasarkan beberapa kriteria dan klasifikasi yang dipergunakan pada penelitian terdapat beberapa atribut yang dipergunakan untuk melakukan prediksi terhadap calon penerima beasiswa terlihat pada Tabel 1.

Tabel 1. Klasifikasi Atribut

No	Atribut
1	Nilai Ujian (UN)
2	Nilai Tes
3	Pendapatan Orang Tua
4	Pekerjaan Orang Tua
5	Jumlah Tanggungan

Berdasarkan pada Tabel 1 terlihat beberapa atribut yang telah ditetapkan berdasarkan pada pengumpulan data yang telah dilakukan melalui observasi dan dokumentasi, atribut ditentukan berdasarkan pada ketetapan dalam pemberian beasiswa terhadap calon penerima dengan kriteria yang harus terpenuhi pada atribut.

Pengumpulan data awal menggunakan sampel 20 calon penerima beasiswa yang dipilih berdasarkan data yang ada pada panitia penerimaan mahasiswa baru dengan indikator atribut seperti pada Tabel 1. Implementasi algoritma C4.5 pada proses awal berdasarkan data yang didapatkan dilakukan pengolahan untuk menentukan learning dataset seperti pada Tabel 2.

Tabel 2. Learning Dataset

No	Nama	Pendapatan orang tua	Pekerjaan orang tua	Jumlah tanggungan	Nilai UN	Nilai tes PMB	Ket
1	A	500.000	Tani	2	75	80	L
2	B	1.000.000	Tani	3	80	85	L

3	C	1.000.000	Tani	2	80	85	L
4	D	500.000	Buruh	1	80	75	T
5	E	1.500.000	Sopir	1	75	75	T
6	F	1.500.000	Sopir	4	75	75	L
7	G	500.000	Buruh	4	75	80	L
8	H	500.000	Buruh	3	80	80	L
9	I	1.500.000	Tani	2	80	80	L
10	J	1.000.000	Sopir	2	80	85	L
11	K	1.500.000	Ojek	3	75	85	L
12	L	500.000	Ojek	4	75	75	L
13	M	1.000.000	Tani	3	80	75	L
14	N	1.000.000	Tani	3	80	70	L
15	O	1.500.000	Sopir	2	85	70	L
16	P	1.500.000	Sopir	2	85	75	T
17	Q	1.000.000	Ojek	3	85	75	T
18	R	2.000.000	Tani	3	75	70	L
19	S	1.500.000	Tani	4	75	80	L
20	T	1.000.000	Ojek	2	75	80	T

Terlihat pada Tabel 2 dataset yang telah dipergunakan dalam melakukan klasifikasi calon penerima beasiswa terdapat 20 data kasus yang dipergunakan dalam perhitungan klasifikasi calon penerima *beasiswa* pada dataset telah terlihat tidak semua calon penerima beasiswa lulus dalam seleksi klasifikasi, terlihat ada sebanyak 5 calon penerima beasiswa yang tidak masuk dalam klasifikasi hal ini berdasarkan pada perhitungan yang telah dilakukan menggunakan algoritma C45.

Setelah dilakukan perhitungan dengan rumus entropy terlihat hasil seperti pada Tabel 3.

Tabel 3. Nilai Entropy Status

Total Kasus	Sum (L)	Sum (TL)	Total Entropy
20	15	5	0.188721876

Pada Tabel 3. Dapat diketahui bahwa nilai perhitungan entropy dengan total sebesar 0.188 dengan demikian dapat terlihat tidak semua data yang telah diklasifikasikan mendapatkan *beasiswa*.

Pemanfaatan algoritma C4.5 untuk melakukan klasifikasi dan prediksi calon penerima beasiswa, dapat dipergunakan dengan baik dengan melihat beberapa atribut indicator sebagai bahan penilaian, dari data training yang telah diolah maka terdapat beberapa pola parameter yang dipergunakan pada pemilihan indeks parameter seperti pada pendapatan orang tua merupakan parameter utama yang dapat mendukung beberapa parameter yang lainnya.

Berdasarkan pada perhitungan dengan nilai dataset terlihat bahwa nilai entropy merupakan hasil akhir dari beberapa calon penerima beasiswa yang bisa diberikan bantuan beasiswa pada learning dataset terdapat sebanyak 20 kasus dari beberapa calon penerima beasiswa hanya 5 calon yang tidak memenuhi persyaratan untuk mendapatkan beasiswa hal ini terlihat pada perhitungan total entropy.

SIMPULAN

Dari hasil penelitian maka bisa ditarik kesimpulan bahwa penerapan algoritma C4.5 untuk klasifikasi calon penerima beasiswa sudah memiliki akurasi yang baik dalam penyeleksian mahasiswa berdasarkan pada algoritma C4.5 dengan beberapa atribut yang telah ditentukan tingkat akurasi didapatkan sebesar 85% sehingga data yang telah diolah dapat dilakukan evaluasi untuk pendataan calon penerima beasiswa.

SARAN

Pada implementasi algoritma C4.5 klasifikasi calon penerimaan beasiswa bagi mahasiswa baru untuk pengembangan penelitian kedepan sebaiknya indicator pada atribut dapat dikembangkan dengan penambahan beberapa atribut agar proses seleksi dengan menggunakan algoritma C4.5 menjadi semakin akurat.

DAFTAR PUSTAKA

- [1] F. Wiza and B. Febriadi, "Classification Analysis Using C4.5 Algorithm To Predict The Level of Graduation of Nurul Falah Pekanbaru High School Students," *IJISTECH (International J. Inf. Syst. Technol.*, vol. 2, no. 2, p. 43, 2019, doi: 10.30645/ijistech.v2i2.21.
- [2] N. Y. L. Gaol, "Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5," *J. Inf. Teknol.*, vol. 2, pp. 23–29, 2020, doi: 10.37034/jidt.v2i1.22.
- [3] G. A. Putri, D. Maryono, and F. Liantoni, "Implementation of the C4.5 Algorithm to Predict Student Achievement at SMK Negeri 6 Surakarta," *IJIE (Indonesian J. Informatics Educ.*, vol. 4, no. 2, p. 51, 2020, doi: 10.20961/ijie.v4i2.47124.
- [4] E. Hasmin and S. Aisa, "Penerapan Algoritma C4.5 Untuk Penentuan Penerima Beasiswa Mahasiswa," *CogITo Smart J.*, vol. 5, no. 2, p. 308, 2019, doi: 10.31154/cogito.v5i2.219.308-320.
- [5] A. Yuliana and D. B. Pratomo, "Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik TEDC Bandung," *Semnasinotek 2017*, pp. 377–384, 2017.
- [6] S. Wahyuni, "Implementation of Data Mining to Analyze Drug Cases Using C4.5 Decision Tree," *J. Phys. Conf. Ser.*, vol. 970, no. 1, 2018, doi: 10.1088/1742-6596/970/1/012030.
- [7] D. Kurniawan, A. Anggrawan, and H. Hairani, "Graduation Prediction System On Students Using C4.5 Algorithm," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 2, pp. 358–365, 2020, doi: 10.30812/matrik.v19i2.685.
- [8] S. Haryati, A. Sudarsono, and E. Suryana, "Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu)," *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [9] A. A. F. Bahri Saiful, "Sistem Pendukung Keputusan Pemberian Beasiswa Untuk Siswa Kurang Mampu Di Smk Muhammadiyah 1 Kepanjen Menggunakan Metode Fuzzy Tsukamoto," *J. SPIRIT*, vol. 9, no. 2, pp. 48–56, 2017.
- [10] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 3, no. Juli, pp. 1–13, 2018, [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/inde>

- x.php/jurasik/article/download/60/pdf.
- [11] E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3, no. July, p. 1, 2018, doi: 10.30645/jurasik.v3i0.60.
- [12] M. A. Abdillah, A. Setyanto, and Sudarmawan, "Implementasi Decision Tree Algoritma C4 . 5 Untuk Memprediksi Kesuksesan Pendidikan Karakter," *Respati*, vol. XV, pp. 59–69, 2020, [Online]. Available: <http://jti.respati.ac.id/index.php/jurnaljti/article/view/349>.
- [13] E. Darmawan, "C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education," *J. Online Inform.*, vol. 3, no. 1, p. 22, 2018, doi: 10.15575/join.v3i1.171.
- [14] N. Anwar, A. Pranolo, and R. Kurnaiwan, "Grouping the community health center patients based on the disease characteristics using C4.5 decision tree," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 403, no. 1, pp. 1–6, 2018, doi: 10.1088/1757-899X/403/1/012084.
- [15] M. P. A. S. R. R. A. P. Pritasari, "Klasifikasi untuk Memprediksi Pembayaran Kartu Kredit Macet," *J. Teknol.*, vol. 3, no. 1, pp. 91–101, 2020.
- [16] S. N. Hermawanti, Asriyanik, and A. A. Sunarto, "Implementasi Algoritma C4.5 untuk Prediksi Kelulusan Tepat Waktu (Studi Kasus : Program Studi Teknik Informatika)," *J. Ilm. SANTIKA*, vol. 9, no. 1, pp. 853–864, 2019, [Online]. Available: <http://jurnalummi.agungprasetyo.net/index.php/santika/article/download/552/253>.
- [17] V. S. Ginting, K. Kusriani, and E. T. Luthfi, "Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 1–6, 2020, doi: 10.36294/jurti.v4i1.1101.
- [18] F. Rahman, H. Zulfia Zahro', and F. X. Ari Wibisono, "Penerapan Algoritma C4.5 Dalam Memprediksi Asal Calon Mahasiswa Berbasis Website (Studi Kasus: Fakultas Hukum Universitas Mataram)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 161–169, 2020, doi: 10.36040/jati.v4i2.2654.
- [19] V. S. Ginting, K. Kusriani, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 36–44, 2020, doi: 10.35585/inspir.v10i1.2535.
- [20] C. Anam and H. B. Santoso, "Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa," *Energy - J. Ilm. Ilmu-Ilmu Tek.*, vol. 8, no. 1, pp. 13–19, 2018, [Online]. Available: <https://ejournal.upm.ac.id/index.php/energy/article/view/111>.
- [21] M.R Santoso and P.Musa, "Rekomendasi Kesehatan Janin Dengan Penerapan Algoritma C5.0 Menggnakan Classifying Cardiotocography Dataset," *Jurnal Simantec*, vol. 9, no. 2, pp. 65–79, 2021, [Online]. Available: <https://journal.trunojoyo.ac.id/simantec/article/view/10730/5748>
- [22] Y.Kustiyahningsih and E. Rahmanita, "Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5 Untuk Penjurusan SMA," *Jurnal Simantec*, vol. 5, no. 2, pp. 101–108, 2016, [Online]. Available: <https://journal.trunojoyo.ac.id/simantec/article/view/1629/1386>