

# Analisa Algoritma C4.5 Terhadap Penentuan Rekomendasi Penerima Beasiswa

Wahyu Susanto<sup>1</sup>, Astriana Mulyani<sup>2\*</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Informatika, Universitas Nusa Mandiri Jakarta, DKI Jakarta, Indonesia

Email: <sup>1</sup>[12210320@nusamandiri.ac.id](mailto:12210320@nusamandiri.ac.id), <sup>2</sup>[astriana.atm@nusamandiri.ac.id](mailto:astriana.atm@nusamandiri.ac.id)

(\* : coresponding author)

**Abstrak**—Beasiswa merupakan bantuan keuangan dengan maksud untuk dipergunakan sebagai sarana melanjutkan pendidikan yang ditempuh dan biasanya diberikan oleh yayasan, perusahaan maupun lembaga pemerintah. Beasiswa berupa dana yang digunakan untuk membantu siswa kurang mampu dalam melanjutkan tugasnya menyelesaikan pendidikan. Oleh karena itu pemebrian beasiswa harus tepat sasaran kepada penerima yang benar-benar layak dan pantas mendapatkannya. Banyaknya calon penerima menjadikan proses seleksi memakan waktu yang lama. Dalam hal ini penggunaan metode data mining dapat dijadikan salah satu solusi untuk mempermudah proses seleksi. Algoritma C4.5 merupakan algoritma yang akan dipakai dalam penelitian ini. Data yang digunakan adalah nama siswa, kelas, penghasilan orang tua, tanggungan orang tua, dan nilai-rata-rata rapor. Pemrosesan data mining pada suatu data training akan menghasilkan pohon keputusan. Metode evaluasi yang dilakukan dalam pengujian ini didapatkan data nilai akurasi 90 %, hal ini bisa menjadi bukti bahwa algoritma C4.5 cukup akurat untuk memberikan rekomendasi beasiswa.

**Kata Kunci:** algoritma C4.5, beasiswa, data mining, klasifikasi, pohon keputusan

**Abstract**—Scholarships are financial assistance with the intention of being used as a means of continuing their education and are usually given by foundations, companies or government agencies. Scholarships are in the form of funds used to assist underprivileged students in continuing their duties in completing their education. Therefore, the granting of scholarships must be right on target to recipients who really deserve and deserve it. The large number of potential recipients makes the selection process take a long time. In this case, the use of data mining methods can be used as a solution to simplify the selection process. The C4.5 algorithm is the algorithm that will be used in this research. The data used are the names of students, class, parents' income, dependents of parents, and average grades of report cards. Data mining processing on a training data will produce a decision tree. The evaluation method carried out in this test obtained 90% accuracy value data, this can be evidence that the C4.5 algorithm is accurate enough to provide scholarship recommendations.

**Keywords:** classification, C4.5 algorithm, data mining, decision tree, scholarships

## 1. PENDAHULUAN

Sebagai salah satu pendukung suksesnya kegiatan di sekolah dalam menciptakan generasi yang berkualitas maka diperlukanlah biaya yang tentu saja tidak sedikit. Melihat latar belakang orang tua dari siswa di sekolah tentu saja tidak semuanya benar-benar mampu untuk membayar biaya selama pendidikan walaupun di masa sekarang biaya pendidikan sudah digratiskan oleh pemerintah namun biaya pendukung seperti biaya perlengkapan sekolah termasuk didalamnya seragam sekolah, tas, alat tulis, sepatu serta biaya transportasi tidak masuk dalam biaya yang digratiskan tersebut.

Oleh karena itu bagi para siswa dari golongan keluarga kurang mampu memiliki hak dan kesempatan untuk memperoleh biaya pendidikan dan bagi mereka yang memiliki prestasi akademik maupun non akademik juga memiliki hak untuk mendapatkan beasiswa (Heni Sulistiani, 2018). Banyaknya siswa dan siswi yang kurang mampu di SMP Al Irsyad Al Islamiyyah Purwokerto dan tidak sanggup untuk membayar biaya pendidikan sehingga proses penentuan penerimaan bantuan beasiswa dilihat dari keikutsertaan dalam program Indonesia Pintar (PIP), nilai rata-rata rapor, penghasilan dan jumlah tanggungan orang tua.

Pemerintah kabupaten Banyumas memiliki program yang bernama Kartu Banyumas Pintar (KBP) dengan tujuan untuk mengurangi halangan siswa kurang mampu untuk bersekolah dengan memberi bantuan kepada siswa tersebut untuk memperoleh pendidikan layak. Adapun Calon Penerima KBP merupakan siswa yang belum menerima bantuan beasiswa Program Indonesia Pintar dari pemerintah pusat (Radar Banyumas, 2022). Calon penerima Kartu Banyumas Pintar (KBP) diutamakan adalah siswa yang duduk di kelas VII.

Selama proses pengajuan calon peserta penerima Kartu Banyumas Pintar (KBP) di SMP Al Irsyad Al Islamiyyah Purwokerto, pihak sekolah mengalami kendala saat menentukan calon penerima KBP, hal ini disebabkan kuota yang ditetapkan oleh pemerintah daerah untuk sekolah dibatasi sedangkan total jumlah siswa tahun pelajaran 2021/2022 di kelas VII ada 298 siswa. Prosedur pengolahan data meliputi penghimpunan data, verifikasi, seleksi data dan penyusunan laporan dilakukan secara manual oleh wali kelas. Sehingga proses ini memakan waktu yang relatif lama hingga beberapa pekan.

Melihat hasil uraian dari latar belakang masalah, maka dapat diketahui rumusan masalah dalam skripsi ini, yaitu bagaimana menerapkan metode Algoritma C4.5 untuk memperkirakan calon penerima beasiswa Kartu Banyumas Pintar (KBP) di SMP Al Irsyad Al Islamiyyah Purwokerto sehingga bisa menjadi pendukung keputusan oleh pihak sekolah dalam proses pemberian beasiswa yang akan datang.

Beasiswa mendukung biaya pendidikan bagi pelajar sehingga mereka dapat melanjutkan pendidikan yang lebih tinggi dengan mempertimbangkan prestasi dan potensi akademik (Hapsari et al., 2018) walaupun di satu sisi beasiswa diberikan kepada pelajar yang kurang mampu.

Data mining memiliki makna suatu aktifitas penguraian dari sekumpulan data menjadi suatu informasi yang berpotensi secara tersirat (implisit) yang sebelumnya belum diketahui (Lailil Muflikhah, Dian Eka Ratnawati, 2018). Data mining merupakan sebuah analisis yang memiliki sifat langsung dari data besar atau kompleks dengan maksud untuk mendapatkan kecenderungan atau pola penting dan umumnya keberadaannya kurang disadari (Aprilia C et al., 2013).

Klasifikasi dapat berarti suatu proses menemukan fungsi ataupun model yang menjadi pembeda dan menggambarkan kelas ataupun konsep data (Anjar Wanto, Muhammad Noor Hasan Siregar, Agus Perdana Windarto, Dedy Hartama, Ni Luh Wiwik Sri Rahayu Ginantra, Darmawan Napitupulu, Edi Surya Negara, Muhammad Ridwan Lubis, Sarini Vita Dewi, 2020).

*Rapidminer* adalah sebuah perangkat lunak opensource dan diciptakan oleh Raif Klinkenberg dan Dr. Markus Hofmann dari *Institute of Technology Blanchardstown* bersama *rapid-i.com* dengan *Graphical User Interface* (GUI) sehingga mempermudah para penggunanya dalam memanfaatkan perangkat lunak tersebut. *Rapidminer* dibuat dengan *javascript* dan mampu dioperasikan disemua sistem operasi komputer. Aplikasi ini tidak membutuhkan keterampilan pengkodean khusus, karena semua fasilitas atau fitur sudah tersedia dan biasanya digunakan memproses data mining (Hendrian, 2018).

C4.5 adalah sebuah algoritma pembentuk pohon keputusan dan sebuah tatacara prediksi serta klasifikasi terkenal dan terkuat. Metode pohon keputusan atau yang biasa disebut *decision tree* mampu mentransformasikan fakta sangat luas menjadi suatu pohon keputusan dengan aturan-aturan yang mudah dipahami (Yunus et al., 2021).

*Decision tree* adalah metode yang dipakai untuk proses penalaran supaya diperoleh solusi dari suatu masalah. Pohon yang terbentuk tidak mesti berbentuk biner. Jika fasilitas dalam dataset memakai dua jenis nilai kategori maka pohon yang diperoleh akan berwujud pohon biner namun apabila memakai jenis numerik atau berisi lebih dari dua jenis nilai kategori maka wujud dari pohon yang didapat umumnya bukan berwujud pohon biner (Prasetyowati, 2017).

*Confusion Matrix* memiliki tujuan menganalisa kualitas dari kinerja suatu model pengelompokan dalam mengidentifikasi suatu nilai dari semua kelas. *Confusion Matrix* mengandung data terkait kelas nyata dan prediksi dari suatu metode pengelompokan (Etriyanti et al., 2020).

## 2. METODOLOGI PENELITIAN

Metode yang digunakan dalam proses pengumpulan data adalah metode pengumpulan data sekunder, untuk memperoleh sebuah data dan fakta yang diperlukan maka penulis melakukan penelitian diantaranya:

### 2.1. Observasi

Observasi dilakukan dengan tujuan untuk memperoleh data-data yang diperlukan dalam menyusun skripsi ini. Data-data yang diperlukan adalah data siswa dari kelas 7 yang nantinya akan berkaitan dengan penentuan penerimaan beasiswa.

### 2.2. Wawancara

Usaha untuk mendapatkan data yang lebih valid, maka dilakukan proses wawancara secara langsung dengan pihak kepala sekolah, guru dan staf tata usaha di SMP Al Irsyad Al Islamiyyah Purwokerto tentang bagaimana ketentuan penerimaan beasiswa di sekolah tersebut.

### 2.3. Studi Pustaka

Berdasarkan permasalahan yang telah dihimpun dari hasil observasi dan wawancara, penulis melakukan pencairan data dari berbagai sumber seperti e-journal dengan studi kasus yang hampir sama, artikel ilmiah, website, buku-buku dan referensi-referensi lainnya. Berisi penjelasan tentang tahapan penelitian yang menggambarkan urutan logis untuk mendapatkan hasil penelitian sesuai dengan harapan dan gambaran sistem. Jika ada gambar dan tabel, itu harus disajikan dengan nama tabel dan gambar yang disertai dengan nomor urut.

## 3. ANALISA DAN PEMBAHASAN

### 3.1 Analisa

Penelitian calon penerima beasiswa berasal dari siswa kelas VII SMP Al Irsyad Al Islamiyyah Purwokerto tahun pelajaran 2021/2022 yang mendapatkan rekomendasi dari wali kelas masing-masing dimana setiap kelas dibatasi maksimal 5 orang siswa sehingga dari 10 rombongan belajar (rombel) di kelas VII didapat data sebanyak 50 data siswa calon penerima beasiswa Kartu Banyumas Pintar (KBP) dengan rincian sebagai berikut:

### 3.2 Diagram Alur

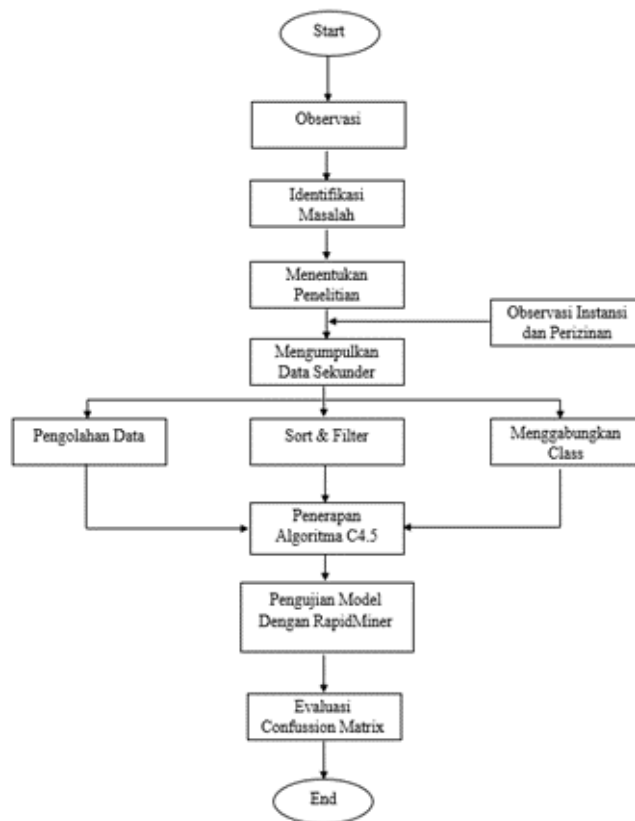
Tahapan penelitian yang dilakukan untuk mendapatkan hasil yang diharapkan dapat dijabarkan sebagai berikut:

- a. Langkah pertama peneliti menentukan lokasi riset yang nantinya dijadikan sebagai bahan penelitian.
- b. Peneliti melaksanakan observasi secara langsung ke SMP Al Irsyad Purwokerto di kabupaten Banyumas.
- c. Selanjutnya peneliti bertanya langsung kepada guru (wali kelas) mengenai ketentuan dan permasalahan pada proses penerimaan beasiswa Kartu Banyumas Pintar (KBP) di sekolah tersebut.
- d. Selanjutnya peneliti menentukan penelitian yaitu kendala pada proses penerimaan beasiswa Kartu Banyumas Pintar (KBP) di sekolah tersebut.
- e. Tahapan berikutnya peneliti meminta izin kepada kepala sekolah untuk pengambilan data yang dibutuhkan sebagai bahan penelitian.
- f. Setelah data diperoleh peneliti kemudian mengolah data menggunakan aplikasi Microsoft Excel, kemudian dilakukan sorting dan filtering agar atribut dan nilai pada data bisa dibagi secara rapi untuk mempermudah proses pengelompokan data.
- g. Data yang telah dirapikan kemudian dihitung menggunakan algoritma C4.5 dan memprediksi hasil menggunakan Confussion Matrix pada hasil akhir nilai akurasi yang diperoleh.
- h. Tahap terakhir penelitian yaitu menarik suatu kesimpulan mengenai hasil akhir penelitian yang didapat.

**Tabel I.** Data Statistik Calon Penerima Beasiswa KBP

Kategori	Simpul	Jumlah Kasus	Menerima	
			Ya	Tidak
Jumlah Tanggungan Orang Tua	1	6	0	6
	2	28	14	14
	≥3	16	15	1
Penghasilan Orang Tua	< 1.000.000	13	13	0
	1.000.000–2.000.000	29	16	13
	>2.000.000	8	0	8

Apabila tahapan di atas digambarkan menggunakan flowchart maka didapat tampilan sebagai berikut:



**Gambar 1.** Tahapan penelitian

## 4. IMPLEMENTASI

### 4.1 Perhitungan Entropy, Gain dan Nilai Akurasi

a. Entropy total

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$\begin{aligned}
 &= (- (29/50) \times \log_2(29/50)) + (- (21/50) \times \log_2(21/50)) \\
 &= ((-0.5800) \times (-0.7859)) + ((-0,4200) \times (-1.2515)) \\
 &= 0.4558 + 0.5256 \\
 &= 0.9815
 \end{aligned}$$

b. Entropy subnet

Dengan menggunakan rumus yang sama maka akan didapatkan data nilai entropy subnet sebagai berikut:

**Tabel II.** Perhitungan entropy subnet

Kategori	Simpul	Jumlah Kasus	Menerima		Entropy
			Ya	Tidak	
Jumlah Tanggungan Orang Tua	1	6	0	6	0,0000
	2	28	14	14	1,0000
	≥3	16	15	1	0,3373
Penghasilan Orang Tua	< 1.000.000	13	13	0	0,0000
	1.000.000–2.000.000	29	16	13	0,9923
	>2.000.000	8	0	8	0,0000

c. Perhitungan gain

Setelah nilai entropy diketahui maka langkah selanjutnya adalah menghitung nilai gain dengan rumus sebagai berikut:

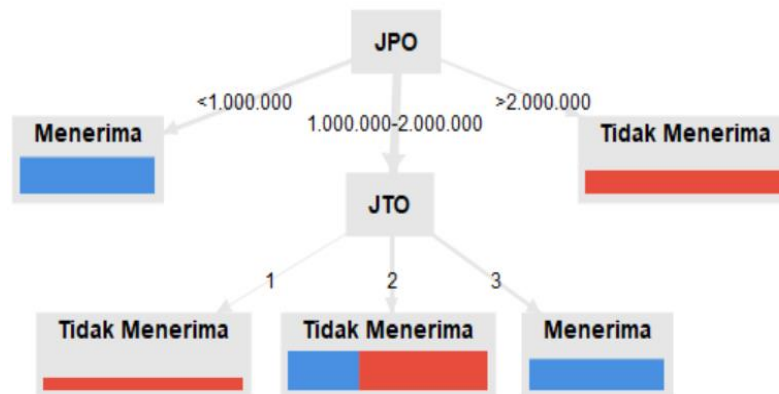
$$Gain (S, A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{s} \times Entropy (S_i)$$

Dari rumus di atas akan didapatkan data hasil perhitungan sebagai berikut:

**Tabel III.** Perhitungan gain

Kategori	Simpul	Jumlah Kasus	Menerima		Entropy	Gain
			Ya	Tidak		
Jumlah	1	6	0	6	0,0000	0.3135
Tanggungan Orang Tua	2	28	14	14	1,0000	
	≥3	16	15	1	0,3373	
Penghasilan Orang Tua	< 1.000.000	13	13	0	0,0000	0.4060
	1.000.000–2.000.000	29	16	13	0,9923	
	>2.000.000	8	0	8	0,0000	

**4.2 Proses Perhitungan Algoritma C4.5 Menggunakan Rapidminer**



**Gambar 2.** Hasil Pohon Keputusan

Apabila pohon keputusan ditampilkan dalam bentuk deskripsi atau *Tree Rule* maka akan muncul tampilan berikut:

## Tree

```
JPO = 1.000.000-2.000.000
| JTO = 1: Tidak Menerima {Menerima=0, Tidak Menerima=4}
| JTO = 2: Tidak Menerima {Menerima=5, Tidak Menerima=9}
| JTO = 3: Menerima {Menerima=11, Tidak Menerima=0}
JPO = <1.000.000: Menerima {Menerima=13, Tidak Menerima=0}
JPO = >2.000.000: Tidak Menerima {Menerima=0, Tidak Menerima=8}
```

Gambar 3. Tampilan *Rule Decision Tree*

accuracy: 90.00%

	true Menerima	true Tidak Menerima	class precision
pred. Menerima	6	1	85.71%
pred. Tidak Menerima	0	3	100.00%
class recall	100.00%	75.00%	

Gambar 4. Nilai Akurasi Menggunakan Metode *Confussion Matrix*

## PerformanceVector

```
PerformanceVector:
accuracy: 90.00%
ConfusionMatrix:
True:   Menerima      Tidak Menerima
Menerima:      6        1
Tidak Menerima: 0        3
AUC: 0.979 (positive class: Tidak Menerima)
```

Gambar 5. Nilai *Performance* Vektor

## 5. KESIMPULAN

Penerapan algoritma C4.5 dalam menentukan peserta beasiswa bagi siswa berprestasi terbukti dapat menghasilkan nilai akurasi yang tinggi dalam proses penentuan pemberian beasiswa Kartu Banyumas Pintar (KBP) dengan data awal sebanyak 50 data yang menghasilkan nilai akurasi sebesar 90,00 % dari 40 data training dan 10 data testing. Pengolahan data ini mempercepat dan mempermudah pihak sekolah dalam menentukan penerimaan beasiswa yang lebih akurat dan tepat sasaran.

Saran untuk penelitian berikutnya sebaiknya menggunakan jumlah data dan atribut yang lebih banyak agar hasil nilai akurasi dapat lebih baik dari penelitian sebelumnya. Penelitian ini bisa dikembangkan dengan menggunakan algoritma lain kemudian membandingkan hasilnya dengan hasil penelitian menggunakan algoritma C4.5 yang digunakan dalam penelitian ini.

## REFERENCES

- Anjar Wanto, Muhammad Noor Hasan Siregar, Agus Perdana Windarto, Dedy Hartama, Ni Luh Wiwik Sri Rahayu Ginantra, Darmawan Napitupulu, Edi Surya Negara, Muhammad Ridwan Lubis, Sarini Vita Dewi, C. P. (2020). *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis. [https://www.google.co.id/books/edition/Data\\_Mining\\_Algoritma\\_dan\\_Implementasi/gAnfDwAAQBAJ?hl=id&gbpv=0](https://www.google.co.id/books/edition/Data_Mining_Algoritma_dan_Implementasi/gAnfDwAAQBAJ?hl=id&gbpv=0)
- Aprilia C. D., Aji Baskoro, D., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Miner*. 139. [https://www.academia.edu/7712860/Belajar\\_Data\\_Mining\\_dengan\\_RapidMiner](https://www.academia.edu/7712860/Belajar_Data_Mining_dengan_RapidMiner)
- Etriyantri, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, 13(1), 56–67. <https://doi.org/10.35671/telematika.v13i1.881>
- Hapsari, D. T., Harini, & Nugroho, J. A. (2018). Pengaruh Beasiswa PPA dan Kebiasaan Belajar Terhadap Prestasi Belajar Mahasiswa FKIP UNS Penerima Beasiswa PPA Periode Januari – Juni 2017. *BISE: Jurnal Pendidikan Bisnis Dan Ekonomi*, 4(1), 1–15.
- Hendrian, S. (2018). Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan. *Faktor Exacta*, 11(3), 266–274. <https://doi.org/10.30998/faktorexacta.v11i3.2777>
- Heni Sulistiani, Y. T. U. (2018). Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa. *Snti*, 300–305.
- Lailil Muflikhah, Dian Eka Ratnawati, R. R. M. (2018). *Data Mining*. Universitas Brawijaya Press. [https://www.google.co.id/books/edition/Data\\_Mining/V\\_NqDwAAQBAJ?hl=id&gbpv=0](https://www.google.co.id/books/edition/Data_Mining/V_NqDwAAQBAJ?hl=id&gbpv=0)
- Prasetyowati, E. (2017). *DATA MINING Pengelompokan Data untuk Informasi dan Evaluasi*. Duta Media Publishing. [https://www.google.co.id/books/edition/DATA\\_MINING\\_Pengelompokan\\_Data\\_untuk\\_Inf/rEH2DwAAQBAJ?hl=id&gbpv=0](https://www.google.co.id/books/edition/DATA_MINING_Pengelompokan_Data_untuk_Inf/rEH2DwAAQBAJ?hl=id&gbpv=0)
- Radar Banyumas. (2022). *Tahun Ini Kuota Kartu Banyumas Pintar untuk SD Naik*. Radarbanyumas.Co.Id. <https://radarbanyumas.co.id/tahun-ini-kuota-kartu-banyumas-pintar-untuk-sd-naik/>
- Yunus, M., Ramadhan, H., Aji, D. R., & Yulianto, A. (2021). Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP). *Paradigma - Jurnal Komputer Dan Informatika*, 23(2). <https://doi.org/10.31294/p.v23i2.11395>