

PENERAPAN ALGORITMA C4.5 DALAM KLASIFIKASI PENERIMAAN PESERTA DIDIK BARU

Harnah Syuryanny Nurindah¹, St. Syahdan¹

¹Jurusan Matematika, Fakultas Mipa, Universitas Kaltara
e-mail: harnahsyuryany991@gmail.com, stsyahdan89@gmail.com

ABSTRACT

Learners are community members who try to develop themselves through the educational process at certain paths, levels, and types. As for some of the attributes that affect the acceptance of new students including gender, school origin, math test scores, science test scores, English test scores and IQ test scores. The purpose of this study was to determine the attributes that affect the acceptance of new students using the C4.5 Algorithm method. One of the methods that can be used to analyze which attributes have the most influence on the acceptance of new students is using the C4.5 Algorithm method. The results of data analysis using the C4.5 Algorithm method indicate that the attribute that most influences the acceptance of new students is the mathematics test score which gives the next largest contribution, namely the English and Science test scores and the smallest contribution in the formation of the decision tree, namely AS (Original). School). Based on the analysis of the results of the tests carried out with an accuracy rate of 90.76%, precision 94.20% and recall 92.85%. This means that the C4.5 classification algorithm can be implemented properly in the acceptance of new students..

Keywords: C4.5 Algorithm, Classification of New Students, Decision Tree

ABSTRAK

Peserta didik adalah anggota masyarakat yang berusaha mengembangkan dirinya melalui proses pendidikan pada jalur, jenjang, dan jenis tertentu. Adapun beberapa atribut yang mempengaruhi dalam penerimaan peserta didik baru diantaranya yaitu jenis kelamin, asal sekolah, nilai tes matematika, nilai tes IPA, nilai tes bahasa inggris dan nilai tes IQ. Tujuan dari penelitian ini adalah untuk mengetahui atribut yang mempengaruhi penerimaan peserta didik baru menggunakan metode *Algoritma* C4.5. Salah satu metode yang dapat digunakan untuk menganalisis atribut mana yang paling mempengaruhi dalam penerimaan peserta didik baru dengan menggunakan metode *Algoritma* C4.5. Hasil analisis data dengan menggunakan metode *Algoritma* C4.5 menunjukkan bahwa atribut yang paling mempengaruhi dalam penerimaan peserta didik baru adalah nilai tes matematika yang memberikan kontribusi terbesar selanjutnya yaitu nilai tes bahasa inggris dan IPA serta yang memiliki kontribusi terkecil dalam pembentukan pohon keputusan yaitu AS (Asal Sekolah). Berdasarkan analisa hasil pengujian yang dilakukan dengan tingkat akurasi 90.76 %, presisi 94,20 % dan *recall* 92,85 %. Hal ini berarti algoritma klasifikasi C4.5 dapat diimplementasikan dengan baik dalam penerimaan peserta didik baru.

Kata Kunci: *Algoritma* C4.5, Klasifikasi Peserta Didik Baru, Pohon Keputusan.

1. Pendahuluan

Matematika merupakan ilmu yang memiliki peranan penting bagi kemajuan peradaban manusia. Matematika telah dikembangkan oleh para matematikawan mulai dari zaman Mesir kuno, Babylonia, hingga Yunani kuno. Pada zaman tersebut matematika dipelajari, dikembangkan, dan digunakan untuk menyelesaikan permasalahan dalam kehidupan sehari-hari, seperti masalah perdagangan, pengukuran tanah, pelukisan, konstruksi, dan astronomi. Sampai sekarang pun matematika masih digunakan, baik untuk menyelesaikan permasalahan dalam kehidupan atau membantu dalam mengembangkan disiplin ilmu lain.

Menurut ketentuan umum Undang-Undang RI tentang sistem pendidikan nasional, peserta didik adalah anggota masyarakat yang berusaha mengembangkan dirinya melalui proses pendidikan pada jalur, jenjang, dan jenis tertentu. Pada pendidikan dasar dan menengah menurut ketentuan pasal 1 Peraturan Pemerintah RI Nomor 28 dan Nomor 29 tahun 1990 disebut dengan siswa. Menurut Nurhamzah (2018) penerimaan peserta didik baru merupakan kegiatan yang sangat penting di sekolah. Dikatakan demikian karena apabila tidak ada penerimaan maka sekolah akan tidak mendapatkan peserta didik. Apabila hal tersebut terjadi, maka eksistensi sekolah pun akan terganggu. Sehingga dalam waktu yang tidak lama sekolah bisa saja ditutup karena tidak mendapatkan peserta didik sesuai dengan kuota yang telah ditentukan, penerimaan peserta didik umumnya dilaksanakan untuk kelas permulaan (kelas satu).

Algoritma C4.5 merupakan salah satu metode yang dapat digunakan dalam mengklasifikasi penerimaan peserta didik baru. Kelebihan dari metode ini adalah dapat mengolah data numerik dan diskret, menangani nilai atribut yang hilang serta menghasilkan aturan-aturan yang mudah diinterpretasikan dan performanya merupakan salah satu yang tercepat dibandingkan dengan algoritma lain. Dalam penelitian yang dilakukan oleh Latifah, dkk (2018) dengan judul penelitian Analisis dan Penerapan Algoritma C4.5 dalam Data Mining untuk Menunjang Strategi Promosi Prodi Informatika UPGRIS menyimpulkan bahwa variabel yang paling tinggi pengaruhnya terhadap hasil registrasi mahasiswa adalah asal sekolah dan jenis kelamin. Rata-rata peserta didik yang melakukan registrasi berasal dari Semarang dengan asal sekolah SMU dari jurusan IPA dan yang berasal dari luar kota rata-rata berasal dari Pati dan Batang. Dari asal sekolah SMU jurusan IPS dengan jenis kelamin perempuan berasal dari Pati sedangkan jenis kelamin Laki-laki berasal dari Batang. Akurasi dari pembentukan model ini adalah sebesar 89.33% Hal ini berarti variabel tersebut perlu digunakan sebagai pertimbangan bagi prodi Informatika dalam melakukan strategi promosi lebih banyak ke sekolah umum.

Selanjutnya dalam penelitian lainnya yang dilakukan oleh Chair, dkk (2017) dengan judul "Aplikasi Klasifikasi Algoritma C4.5 (Study Kasus Masa Study Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008)" menyimpulkan bahwa berdasarkan dari hasil penelitian yang dilakukan dalam klasifikasi masa studi mahasiswa FMIPA UNMUL angkatan 2008 (102 data) diperoleh 16 aturan yang terbentuk. Dari hasil klasifikasi yang dilakukan dalam penelitian ini diperoleh ketepatan akurasi untuk data training (75 data) adalah 100% dan untuk data testing (27 data) adalah 77,7%.

Berdasarkan latar belakang diatas, maka penulis tertarik untuk melakukan penelitian dengan judul Penerapan Algoritma C4.5 Dalam Klasifikasi Penerimaan Peserta Didik Baru.

2. Metode

Dalam penelitian ini digunakan jenis penelitian terapan dengan penjelasan bahwa penelitian terapan adalah penelitian yang dikerjakan dengan maksud untuk menerapkan,

menguji dan mengevaluasi kemampuan suatu teori yang diterapkan dalam pemecahan permasalahan praktis.

Data yang digunakan dalam penelitian ini adalah data sekunder. Menurut Firdaus, dkk (2016) data sekunder adalah data yang dikumpulkan dari tangan kedua atau sumber-sumber lain yang telah tersedia sebelum penelitian dilakukan.

Data yang digunakan dalam penelitian ini diperoleh dari jurnal penelitian yang dilakukan oleh Jannah (2020) dalam penelitiannya yang berjudul "*Klasifikasi Penerimaan Peserta Didik Baru di Madrasah Bertaraf Internasional Amanatul Ummah Pacet Menggunakan Bootstrap Aggregating Classification and Regression Tree*".

Variabel (atribut) yang digunakan dalam penelitian ini adalah variabel (atribut) yang ada dalam jurnal yang menjadi rujukan sumber data. Atribut tersebut diantaranya adalah sebagai berikut:

1. Jenis kelamin (X_1)

Dalam penelitian ini, atribut jenis kelamin sebagai atribut pertama dengan dilambangkan X_1 , atribut ini dibedakan menjadi dua kategori diantaranya :

1 = responden jenis kelamin laki-laki

2 = responden jenis kelamin perempuan

2. Asal Sekolah (X_2)

Atribut asal sekolah sebagai atribut kedua dengan dilambangkan X_2 , atribut ini dibedakan menjadi dua kategori diantaranya :

1 = responden berasal dari SMP berbasis islam

2 = responden berasal dari SMP umum

3. Nilai Tes Matematika (X_3)

Atribut nilai tes matematika sebagai atribut ketiga dengan dilambangkan X_3 , atribut ini tersebar dari 1 hingga 100.

4. Nilai Tes IPA (X_4)

Atribut nilai tes IPA sebagai atribut keempat dengan dilambangkan X_4 , atribut ini tersebar dari 1 hingga 100.

5. Nilai Tes Bahasa Inggris Inggris (X_5)

Atribut nilai tes bahasa inggris sebagai atribut kelima dengan dilambangkan X_5 , atribut ini tersebar dari 1 hingga 100.

6. Nilai Tes IQ (X_6)

Atribut nilai tes IQ sebagai atribut keenam dengan dilambangkan X_6 , atribut ini menggunakan skala kontinu.

3. Hasil dan Pembahasan

Analisis deskriptif dalam penelitian ini bertujuan untuk memberikan gambaran dari data yang digunakan dalam penelitian ini. Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari jurnal ilmiah yang ditulis oleh Jannah (2020) dengan judul *Klasifikasi Penerimaan Peserta Didik Baru Di Madrasah Bertaraf Internasional Amanatul Ummah Pacet Menggunakan Bootstrap Agregating Classification And Regression Tree*. Data diketahui bahwa jumlah seluruh peserta calon peserta didik baru sebanyak 390 peserta dengan rincian sebanyak 110 calon peserta didik baru dinyatakan telah diterima seleksi penerimaan peserta didik baru dan sebanyak 280 calon peserta didik baru dinyatakan

tidak diterima seleksi penerimaan peserta didik baru. Setelah dilakukan analisis terhadap data awal, diperoleh data sebagai berikut:

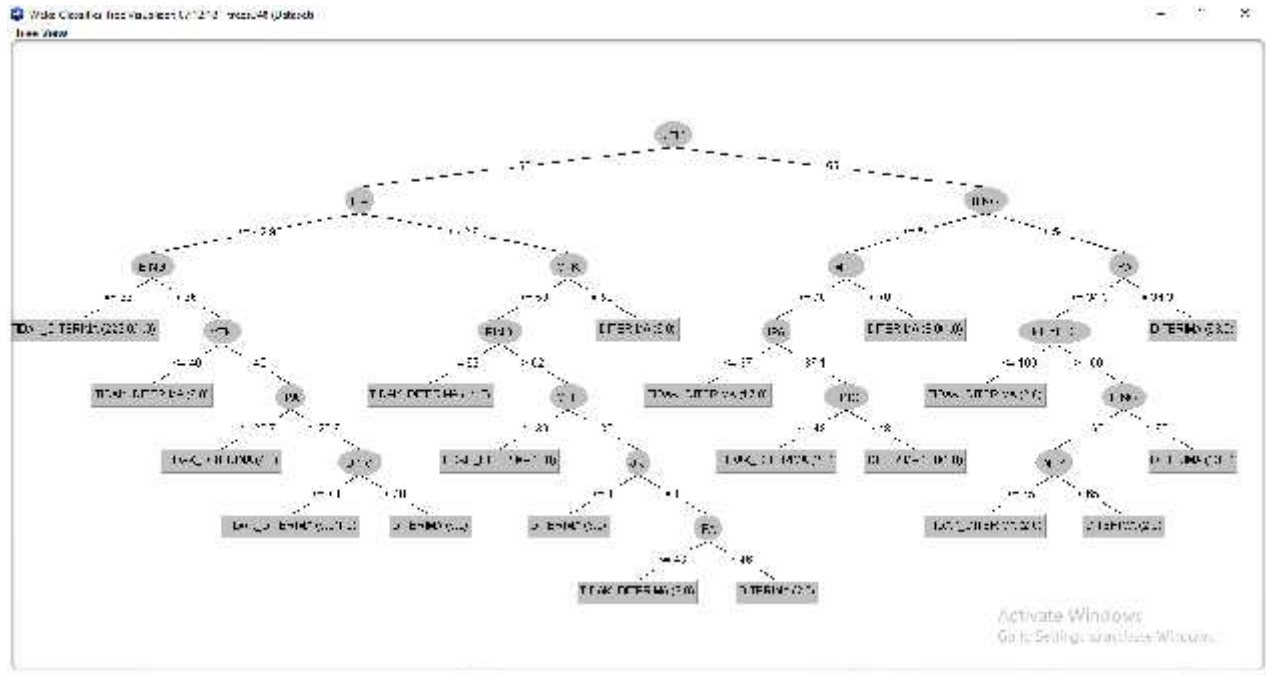
Tabel 1 Persentase Indikator Menurut Kategori Atribut

Atribut	Jumlah	Persentase
1. Jenis Kelamin (X_1)		
Laki-Laki	181	46,41%
Perempuan	209	53,59%
2. Asal Sekolah (X_2)		
SMP Berbasis Islam	319	81,80%
SMP Umum	71	18,20%
3. Nilai Tes Matematika (X_3)		
≤ 55	273	70%
> 55	117	30%
4. Nilai Tes IPA (X_4)		
$\leq 37,1$	254	65,12%
$> 37,1$	136	34,88%
5. Nilai Tes Bahasa Inggris (X_5)		
≤ 56	237	60,77%
> 56	153	39,23%
6. Nilai Tes IQ (X_6)		
≤ 113	274	70,25%
> 113	116	29,74%

Berdasarkan Tabel 1, terdapat beberapa atribut yang mempengaruhi dalam penentuan diterima atau tidaknya dalam seleksi penerimaan peserta didik baru di Madrasah Bertaraf Internasional Amanatul Ummah Pacet, atribut tersebut diantaranya yaitu, Jenis Kelamin (Laki-laki atau Perempuan), Asal Sekolah (SMP Berbasis Islam atau SMP Umum), Nilai tes Matematika, Nilai tes IPA, Nilai tes Bahasa Inggris, dan Nilai tes IQ.

Dari Tabel 1 juga terlihat bahwa untuk atribut jenis kelamin (X_1) terbagi atas dua kategori dimana jumlah persentase perempuan lebih besar dibandingkan dengan laki-laki yaitu sebesar 53,59%. Untuk atribut asal sekolah (X_2), calon peserta didik baru didominasi dari sekolah SMP berbasis islam dengan persentase sebesar 81,80%. Adapun penentuan kategori untuk atribut nilai tes matematika (X_3), nilai tes IPA (X_4), nilai tes Bahasa Inggris (X_5), dan nilai tes IQ (X_6) dilakukan berdasarkan analisis nilai *gain* terbesar dari masing-masing interval yang ada dari setiap atribut.

Setelah dilakukan perhitungan manual dari masing-masing atribut dengan menggunakan persamaan (2.1), persamaan (2.2), persamaan (2.3), dan persamaan (2.4) untuk menghasilkan pohon keputusan algoritma C4.5, adapun cara lain yang dapat dilakukan yaitu dengan menggunakan aplikasi weka pada Gambar 1 sebagai berikut :



Gambar 1 Hasil Klasifikasi Pohon Algoritma C4.5

Berdasarkan Gambar 4.1 terlihat bahwa atribut matematika yang menjadi akar pada pohon algoritma C4.5. Hal ini dikarenakan atribut matematika memiliki nilai *gain ratio* tertinggi dari atribut lainnya. Selanjutnya atribut matematika ini akan dijadikan sebagai pemisah untuk menentukan cabang pohon algoritma C4.5. Proses ini akan dilakukan secara terus menerus dan dihentikan apabila tidak ada lagi yang akan menjadi atribut serta memiliki kelas nya masing-masing.

4. Simpulan dan Saran

Berdasarkan hasil analisis dan pembahasan klasifikasi dari data penerimaan peserta didik baru dapat diperoleh kesimpulan sebagai berikut: Berdasarkan analisis algoritma C4.5 terdapat atribut yang mempengaruhi dalam pembentukan pohon keputusan tersebut diantaranya yaitu nilai tes matematika merupakan atribut yang memberikan kontribusi terbesar selanjutnya yaitu nilai tes bahasa inggris dan IPA serta yang memiliki kontribusi terkecil dalam pembentukan pohon keputusan yaitu AS (Asal Sekolah). Berdasarkan analisa hasil pengujian yang dilakukan dengan tingkat akurasi 90.76 %, presisi 94,20 % dan *recall* 92,85 %. Hal ini berarti algoritma klasifikasi C4.5 dapat diimplementasikan dengan baik dalam penerimaan peserta didik baru.

Beberapa saran dari penulis untuk pengembangan penelitian lebih lanjut yaitu :

1. Perlu adanya penelitian lebih lanjut dengan melakukan pengujian dengan metode lain seperti *Naive Bayes* dan lain sebagainya agar memperoleh perbandingan dengan tingkat akurasi yang paling tinggi dalam mengklasifikasikan penerimaan peserta didik baru.
2. Dalam penelitian selanjutnya diharapkan penelitian ini dapat dikembangkan dalam bentuk aplikasi lain seperti *Rapid Miner* dan lain sebagainya.

Daftar Pustaka

- Asroni., Respati, B.M., Riyadi, S. (November 2018). *Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta*. Jurnal Semesta Teknik, Vol.21, No.2.
- Azwanti, N. (Februari 2018). *Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT.Capella Dinamik Nusantara Cabang Muka Kuning*. Jurnal Ilmiah Ilmu Komputer, Vol.13, No.1.
- Chair, M., Nasution, Y.N., Rizki, N.A. (Februari 2017). *Aplikasi Klasifikasi Algoritma C4.5 (Studi Kasus Masa Studi Mahasiswa Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Mulawarman Angkatan 2008)*. Jurnal Informatika Mulawarman, Vol. 12, No.1.
- Febriansyah, A., Rachmanto, A. (Oktober 2016). *Jurnal Riset Akuntansi*. Vol.8, No.2.
- Fernitha, V. (2019). *Penerapan Algoritma C4.5 untuk Deteksi Penyakit Kanker Serviks*. (Skripsi). Jurusan Teknik Informatika.
- Harahap, M. (Desember 2016). *Esensi Peserta Didik dalam Perspektif Pendidikan Islam*. Jurnal Al-Thariqah, Vol.1, No.2.
- Hasibuan, E.A., Harahap, A.N. (November 2018). *Aplikasi Metode CHAID dalam Menganalisis Kecendrungan Penelitian Skripsi Mahasiswa Pada Program Studi Pendidikan Matematika*. Jurnal Riset Pendidikan Matematika, Vol.1, No.2.
- Jahari, J., Khoiruddin, H., Nurjanah, H. (Desember 2018). *Manajemen Peserta Didik*. Jurnal Islamic Education Manajemen, Vol.3, No.2.
- Jannah, D.M. (2020). *Klasifikasi Penerimaan Peserta Didik Baru Di Madrasah Bertaraf Internasional Amanatul Ummah Pacet Menggunakan Bootstrap Agregating Classification And Regression Tree*. (Skripsi). Jurusan Matematika.
- Latifah, K., Wibowo, S., Nada, N.Q. (Oktober 2018). *Analisis Penerapan Algorithma C4.5 Dalam Data Mining Untuk Menunjang Strategi Promosi Prodi Informatika UPGRIS*. Jurnal Teknik Informatika, Vol. 11, No.2.
- Mujilahwati, S. (24-25 November 2017). *Pemanfaatan Algoritma ID3 untuk Klasifikasi Penjualan Obat*. Jurnal Fakultas Teknik, 978-602-50782-0-0
- Nasrullah, A.H. (Agustus 2018). *Penerapan Metode C4.5 untuk Klasifikasi Mahasiswa Berpotensi Drop Out*. Jurnal Ilmiah, Vol.10, No.2.
- Nurhamzah, I. (2018). *Manajemen Penerimaan Peserta Didik Baru Di Madrasah Ibtidaiyah Hasyim Asy'Ari Malang*. (Skripsi). Jurusan Manajemen Pendidikan Islam.
- Rohmaniyah, V. (2020). *Peserta Didik dalam Pendidikan Islam*. Jurnal Jurusan Pendidikan Islam.