

The Application of Decision Tree C4.5 to Determine College Majors for Students After Madrasah Aliyah (Case Study: Madrasah Aliyah Misbahul Ulum)

Tasya Mulyani^{1*}, Bustami²

^{1,2} Universitas Malikussaleh, Indonesia

*Corresponding Author Email: tasya.200170003@mhs.unimal.ac.id

ABSTRAK

Received: 19 November 2024
Revised: 30 December 2024
Accepted: 31 December 2024
Available online: 1 January 2025

Kata Kunci:

algoritma C4.5, decision tree, php, sistem

Penelitian ini bertujuan (1) Untuk mengaplikasikan Decision Tree C4.5 dalam menentukan jurusan kuliah siswa pasca – Madrasah Aliyah (2) Untuk mengetahui hasil penerapan Metode Decision Tree C4.5 dalam menentukan jurusan kuliah siswa pasca – Madrasah Aliyah (3) Untuk merancang sistem pendukung keputusan agar mempermudah dan mempercepat dalam proses menganalisa data siswa untuk menentukan jurusan kuliah siswa pasca-Madrasah Aliyah. Sistem ini dikembangkan menggunakan Bahasa pemrograman php dengan menerapkan metode *Decision Tree* algoritma C4.5. Decision Tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon dengan konsep menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan dengan menentukan nilai entropy, split info dan gain ratio penelitian ini menghasilkan hasil yang objektif. Melibatkan siswa dari kelas 12 semester ganjil, penelitian ini menghasilkan berupa hasil identifikasi jurusan tektik untuk setiap siswa.

ABSTRACT

Keywords:

C4.5 algorithm, decision tree, php, system

This research aims (1) To apply Decision Tree C4.5 in determining post-Madrasah Aliyah students' college majors (2) To find out the results of applying the C4.5 Decision Tree Method in determining post-Madrasah Aliyah students' college majors (3) To design a system decision support to simplify and speed up the process of analyzing student data to determine post-Madrasah Aliyah students' college majors. This system was developed using the PHP programming language by applying the C4.5 Decision Tree algorithm method. Decision Tree is a machine learning algorithm that uses a set of rules to make decisions with a tree-like structure with the concept of presenting the algorithm with conditional statements, which includes branches to represent decision-making steps that can lead to profitable results by determining the value of entropy, split info and The gain ratio of this research produces objective results. Involving students from class 12 odd semesters, this research resulted in the identification of engineering majors for each student.

1. INTRODUCTION

Perkembangan sistem komputerisasi tidak terlepas dari perkembangan teknologi yang hampir mempengaruhi seluruh aspek kehidupan manusia. Dalam perkembangan sistem komputerisasi, pemanfaatan teknologi informasi dalam bidang pengetahuan sangat diperlukan. Salah satu kelebihan dari sistem komputerisasi adalah dapat menghasilkan informasi segala macam secara relavan, tepat waktu dan akurat. Pendidikan sekarang memerlukan aplikasi yang dapat mengolah data dengan cepat, supaya mempermudah siswa untuk mendapatkan informasi dan membantu serta memudahkan guru dalam hal mengolah data siswa, sehingga guru mengetahui dan dapat membimbing keputusan kuliah yang diinginkan para siswa di sekolah.

Oleh karena itu, dibutuhkan sistem pendukung keputusan untuk menentukan pilihan terbaik dari semua kriteria.

Membutuhkan waktu yang cukup lama untuk mengevaluasi keputusan kuliah siswa, karena proses mengevaluasinya masih dilakukan secara manual, jadi hal tersebut membuat hasil prosesnya menjadi lambat, maka dari itu diperlukannya sistem pendukung keputusan yang terkomputerisasi agar membantu proses pengambilan keputusan yang lebih cepat dan akurat.

Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan. Dimana setiap cabang mewakili hasil untuk atribut, sedangkan jalur dari daun ke akar mewakili aturan untuk klasifikasi.

Oleh karena itu, Metode Decision Tree diperlukan untuk mengidentifikasi rencana kuliah siswa. Penerapan Metode Decision Tree juga merupakan bentuk pengembangan teknologi dalam pendidikan. Metode Decision tree biasanya meniru kemampuan berpikir manusia saat membuat keputusan, sehingga mudah dimengerti.

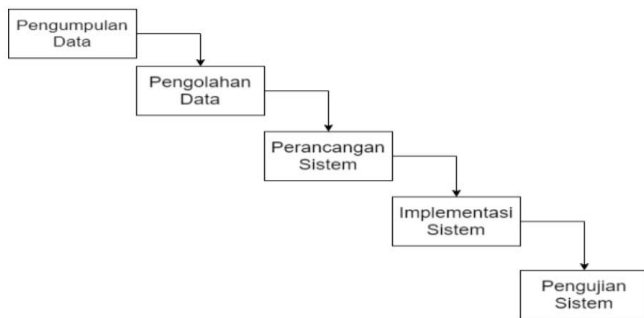
Dengan permasalahan diatas penulis merasa tertarik melakukan penelitian dengan metode decision tree.

2. RESEARCH METHODS

2.1 Tempat Dan Waktu Penelitian

Tempat penelitian ini dilakukan di Madrasah Aliyah Swasta Misbahul Ulum yang beralamat di Jalan Tgk.Chieq Dipaloh Kec.Muara Satu Kota Lhokseumawe, dengan waktu penelitian dimulai dari bulan Januari 2024.

2.2 Langkah Penelitian



Gambar 1. Langkah Penelitian

Dalam penelitian ini sudah disusun berbagai Langkah-langkah penelitian yang nantinya akan dilakukan secara sistematis. Langkah Penelitian yang dilakukan adalah:

(1) Pengumpulan Data

Pada tahapan ini akan dijelaskan bagaimana proses pengambilan data yang digunakan untuk keperluan penelitian, seperti data apa saja yang digunakan dalam penelitian, kemudian bagaimana data tersebut digunakan.

(2) Metode Pengolahan Data

Pengolahan data menggunakan berbagai metode statistic dasar seperti mencari Mean, Modus, Median dan Persentase, hasil Prediksi menggunakan berbagai software seperi Excel dan Google Spreadsheet.

(3) Perancangan Sistem

Perancangan sistem menggunakan Diagram Konteks, DFD dan ERD. Yang dimana perancangan ini akan mmembantu proses pengembangan aplikasi yang menggunakan bahasa pemrograman.

(4) Implementasi Sistem

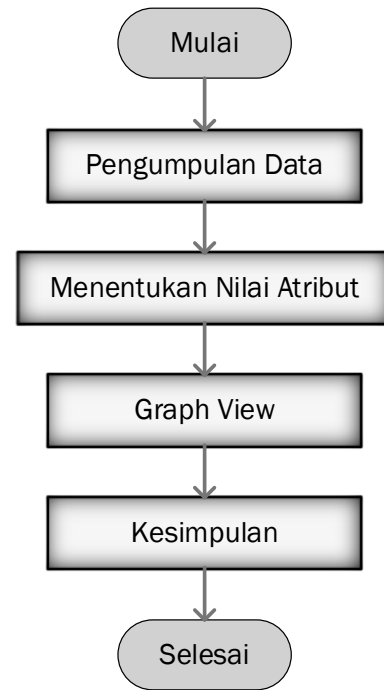
Implementasi sistem adalah proses mengembangkan aplikasi menggunakan bahasa pemrograman. Dalam hal ini akan digunakan bahasa pemrograman seperti PHP.

(5) Pengujian Sistem

Pengujian system dilakukan dengan melakukan tahapan tahapan testing and debugging program untuk memastikan apakah dapat berjalan dengan baik sesuai dengan rancangan yang sudah dibuat sebelumnya.

2.3 Metode Penelitian

Metode penelitian ini menggunakan alur penelitian yang dimulai dari pengumpulan data dan diakhiri dengan kesimpulan. Dapat dilihat pada gambar berikut:



Gambar 2. Metodologi Penelitian

Keterangan:

(1) Pengumpulan Data

Pengumpulan data penting untuk dilakukan sehingga diperoleh informasi yang berkaitan dengan objek penelitian guna menunjang proses penelitian.

(2) Menentukan Nilai Atribut

Dalam data sampel ditentukan terlebih dahulu node terpilih, yaitu dengan menghitung nilai informasi gain masing-masing atribut untuk menentukan node terpilih, gunakan nilai informasi gain yang paling besar. Kemudian akan dilakukan perhitungan nilai entropy dan gain masing-masing atribut.

(3) Graph View

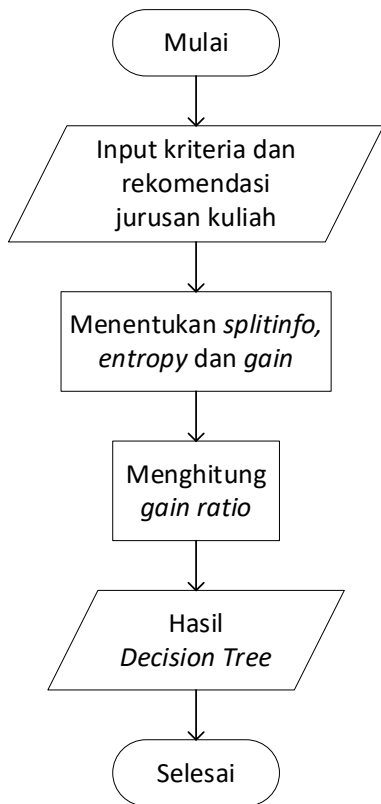
Graph view menunjukkan hasil dari klasifikasi dengan percabangan yang dapat dihasilkan kesimpulan.

(4) Kesimpulan

Aturan-aturan (rules) yang didapat dari hasil pohon keputusan dan deskripsi tree diatas untuk menganalisa faktor rekomendasi karir berdasarkan atribut.

2.4 Skema Sistem

Berikut gambaran skema sistem dari penerapan decision tree c4.5 untuk menentukan jurusan kuliah siswa pasca – Madrasah Aliyah.



Gambar 3. Skema Sistem

Keterangan:

- (1) Langkah pertama yaitu input nilai kriteria dan rekomendasi jurusan kuliah
- (2) Selanjutnya, aplikasi akan menentukan SplitInfo, Entropy dan Gain
- (3) Lalu, aplikasi akan menghitung gain ratio
- (4) Selesai, aplikasi akan menampilkan hasil nilai Decision Tree

3. RESULT AND DISCUSSION

3.1 Analisa Perancangan

Dalam penelitian ini, kriteria untuk menentukan jurusan kuliah siswa pasca-Madrasah Aliyah di Madrasah Aliyah Misbahul Ulum diidentifikasi melalui metode *Decision Tree*, dengan mempertimbangkan 5 mata pelajaran yaitu Matematika, Bahasa Inggris, Biologi, Fisika, dan Kimia. Siswa yang menjadi subjek penelitian berasal dari kelas 12 semester 1(ganjil) dengan jumlah total siswa 24 orang.

4.1 Tahapan Metode *Decision Tree* C4.5

4.1.1 Data siswa/i

Untuk nilai mata pelajaran siswa kelas 12 semester 1(ganjil) dataset yang digunakan sebagai berikut:

Tabel 1. Dataset nilai siswa

Nama Siswa	MTK	Bahasa Inggris	Biologi	Fisika	Kimia	Jurusan Kuliah Teknik
Agustina	96	96	96	96	96	Teknik Kimia
Alfia Rahmi	96	96	96	85	85	Teknik Informatika
Aulia Farissa	96	96	87	85	82	Teknik Informatika
Ayu Nabila	94	83	89	95	82	Teknik Arsitektur
Cut Fara Dina	97	97	97	97	97	Teknik Industri
Cut Rahil	96	86	97	97	97	Teknik Kimia
Cut Rahimah	97	97	86	97	97	Teknik Industri
Dinda Septia Annisa	97	97	97	97	97	Teknik Industri
Farhatul Fadillah	96	90	87	96	86	Teknik Sipil
Misratul Husna	97	97	97	95	95	Teknik Kimia
Muamar	96	96	85	96	85	Teknik Sipil
Muhammad Alwi Zahrial	95	84	84	95	83	Teknik Arsitektur
Muhammad Ausid Addari	97	87	87	97	97	Teknik Mesin
Muhammad Husnul Qamar	93	93	82	95	82	Teknik Arsitektur
Muhammad Nazar	96	96	96	96	96	Teknik Industri
Muhammad Rafi Akbar	97	86	86	87	97	Teknik Material
Muhammad Ramadhan	93	93	88	83	82	Teknik Informatika
Muhammad Rusdi	95	84	82	95	82	Teknik Sipil
Muhammad Saprizal	96	96	86	87	96	Teknik Material
Muliana	96	96	86	87	96	Teknik Material
Mutia	93	93	88	82	81	Teknik Informatika
Novi	97	97	97	96	96	Teknik Material
Nur Fira	93	93	88	82	81	Teknik Informatika
Nurbaiti	93	93	82	93	81	Teknik Arsitektur

Berikut adalah langkah-langkah untuk perhitungan Algoritma C4.5:

Atribut: Nilai Mapel Matematika, B.Inggris, Biologi, Fisika dan Kimia.

Label Kelas: Jurusan Kuliah (Teknik: Sipil, Industri, Kimia, Informatika, Arsitektur, Mesin dan Material).

4.1.2 Menghitung *entropy* awal

Sebelum membagi dataset, kita hitung *entropy* awal (*entropy* dari seluruh dataset). Terdapat 7 jurusan berbeda, jadi kita hitung distribusi *probabilitas* dari setiap jurusan.

Distribusi jurusan:

- **Teknik Sipil:** 3 siswa
- **Teknik Industri:** 4 siswa
- **Teknik Kimia:** 3 siswa
- **Teknik Informatika:** 5 siswa
- **Teknik Mesin:** 2 siswa
- **Teknik Arsitektur:** 4 siswa
- **Teknik Material:** 3 siswa

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

$$Entropy(S) = -\left(\frac{3}{24} \log_2 \frac{3}{24} + \frac{4}{24} \log_2 \frac{4}{24} + \frac{3}{24} \log_2 \frac{3}{24} + \frac{5}{24} \log_2 \frac{5}{24} + \frac{2}{24} \log_2 \frac{2}{24} + \frac{4}{24} \log_2 \frac{4}{24} + \frac{3}{24} \log_2 \frac{3}{24}\right)$$

$$= -\left(\frac{3}{24} (-3) + \frac{4}{24} (-2,585) + \frac{3}{24} (-3) + \frac{5}{24} (-2,263) + \frac{2}{24} (-3,585) + \frac{4}{24} (-2,585) + \frac{3}{24} (-3)\right)$$

$$= 2,758$$

Setelah *Entropy*(S) dihitung, selanjutnya menghitung *Gain* dan *Gain Ratio* untuk setiap atribut.

4.1.3 Mata pelajaran matematika

Hitung *information gain* untuk atribut Matematika. Kita bagi nilai menjadi dua kategori:

Nilai tinggi (≥ 90)

Nilai rendah (< 90)

Menghitung *Entropy* Tiap Cabang:

(1) *Entropy* cabang tinggi (≥ 90):

Terdapat 24 siswa dengan nilai Matematika tinggi:

$$E(\text{tinggi}) = -\left(\frac{3}{24} \log_2 \frac{3}{24} + \frac{4}{24} \log_2 \frac{4}{24} + \frac{3}{24} \log_2 \frac{3}{24} + \frac{4}{24} \log_2 \frac{4}{24} + \frac{2}{24} \log_2 \frac{2}{24} + \frac{4}{24} \log_2 \frac{4}{24} + \frac{3}{24} \log_2 \frac{3}{24}\right)$$

$$= -\left(\frac{3}{24} (-3) + \frac{4}{24} (-2,585) + \frac{3}{24} (-3) + \frac{4}{24} (-2,585) + \frac{2}{24} (-3,585) + \frac{4}{24} (-2,585) + \frac{3}{24} (-3)\right)$$

$$= 2,758$$

(2) *Entropy* cabang rendah (< 90):

Terdapat 0 siswa dengan nilai Matematika rendah, sehingga $E(\text{rendah}) = 0$

setelah menghitung *entropy*, selanjutnya menghitung *gain* atribut Matematika:

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

$$Gain(S, Matematika) = Entropy(S) - \left(\frac{24}{24} * E(\text{tinggi}) + \frac{0}{24} * E(\text{rendah})\right)$$

$$= 2,758 - \left(\frac{24}{24} * 2,748 + \frac{0}{24} * 0\right)$$

$$= 2,758 - (2,758+0)$$

$$= 0$$

Setelah *gain* didapatkan, selanjutnya adalah menghitung *split info* dan *gain ratio*

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

$$SplitInfo(\text{Matematika}) = -\left(\frac{24}{24} \log_2 \left(\frac{24}{24}\right) + \frac{0}{24} \log_2 \left(\frac{0}{24}\right)\right)$$

$$= -\left(\frac{24}{24} (0) + \frac{0}{24} (0)\right)$$

$$= 0$$

Dan untuk *Gain Rationya* adalah sebagai berikut

$$Gain Ratio = \frac{Gain(A)}{SplitInfo(S, A)} = \frac{0}{0} = 0$$

4.1.4 Mata pelajaran bahasa inggris

Hitung *information gain* untuk atribut Bahasa Inggris. Kita bagi nilai menjadi dua kategori:

Nilai tinggi (≥ 90)

Nilai rendah (< 90)

Menghitung *Entropy* Tiap Cabang:

1. *Entropy* cabang tinggi (≥ 90):

Terdapat 18 siswa dengan nilai Bahasa Inggris tinggi:

$$E(\text{tinggi}) = -\left(\frac{2}{18} \log_2 \frac{2}{18} + \frac{4}{18} \log_2 \frac{4}{18} + \frac{2}{18} \log_2 \frac{2}{18} + \frac{5}{18} \log_2 \frac{5}{18} + \frac{0}{18} \log_2 \frac{0}{18} + \frac{2}{18} \log_2 \frac{2}{18} + \frac{3}{18} \log_2 \frac{3}{18}\right)$$

$$= -\left(\frac{2}{18} (-3,170) + \frac{4}{18} (-2,170) + \frac{2}{18} (-3,170) + \frac{5}{18} (-1,848) + \frac{0}{18} (0) + \frac{2}{18} (-3,170) + \frac{3}{18} (-2,585)\right)$$

$$= 2,483$$

2. *Entropy* cabang rendah (< 90):

Terdapat 6 siswa dengan nilai Bahasa Inggris rendah:

$$E(\text{rendah}) = -\left(\frac{1}{6} \log_2 \frac{1}{6} + \frac{0}{6} \log_2 \frac{0}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{0}{6} \log_2 \frac{0}{6} + \frac{2}{6} \log_2 \frac{2}{6} + \frac{2}{6} \log_2 \frac{2}{6} + \frac{0}{6} \log_2 \frac{0}{6}\right)$$

$$= -\left(\frac{1}{6} (-2,621) + \frac{0}{6} (0) + \frac{1}{6} (-2,621) + \frac{0}{6} (0) + \frac{2}{6} (-1,621) + \frac{2}{6} (-1,621) + \frac{0}{6} (0)\right)$$

$$= 1,954$$

Setelah menghitung *entropy*, selanjutnya menghitung *gain* atribut Bahasa Inggris:

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

$$Gain(S, \text{Bahasa Inggris}) = Entropy(S) - \left(\frac{18}{24} * E(\text{tinggi}) + \frac{6}{24} * E(\text{rendah})\right)$$

$$= 2,758 - \left(\frac{18}{24} * 2,483 + \frac{6}{24} * 1,954\right)$$

$$= 2,758 - (1,862+0,4885)$$

$$= 0,4075$$

Setelah *gain* didapatkan, selanjutnya adalah menghitung *split info* dan *gain ratio*

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

$$SplitInfo(\text{Bahasa Inggris}) = -\left(\frac{18}{24} \log_2 \left(\frac{18}{24}\right) + \frac{6}{24} \log_2 \left(\frac{6}{24}\right)\right)$$

$$= -\left(\frac{18}{24} (-0,415) + \frac{6}{24} (-2)\right)$$

$$= 0,811$$

Dan untuk *Gain Rationya* adalah sebagai berikut

$$Gain Ratio = \frac{Gain(A)}{SplitInfo(S, A)} = \frac{0,4075}{0,811} = 0,502$$

4.1.5 Mata pelajaran biologi

Hitung *information gain* untuk atribut Biologi. Kita bagi nilai menjadi dua kategori:

Nilai tinggi (≥ 90)

Nilai rendah (< 90)

Menghitung *Entropy* Tiap Cabang:

1. *Entropy* cabang tinggi (≥ 90):

Terdapat 8 siswa dengan nilai Biologi tinggi:

$$E(\text{tinggi}) = -\left(\frac{0}{8} \log_2 \frac{0}{8} + \frac{3}{8} \log_2 \frac{3}{8} + \frac{3}{8} \log_2 \frac{3}{8} + \frac{1}{8} \log_2 \frac{1}{8} + \frac{0}{8} \log_2 \frac{0}{8} + \frac{0}{8} \log_2 \frac{0}{8} + \frac{1}{8} \log_2 \frac{1}{8}\right)$$

$$= -\left(\frac{0}{8}(0) + \frac{3}{8}(-1,415) + \frac{3}{8}(-1,415) + \frac{1}{8}(-3) + \frac{0}{8}(0) + \frac{0}{8}(0) + \frac{1}{8}(-3)\right)$$

$$= 1,81125$$

2. *Entropy* cabang rendah (< 90):

Terdapat 6 siswa dengan nilai Biologi rendah:

$$E(\text{rendah}) = -\left(\frac{3}{16} \log_2 \frac{3}{16} + \frac{1}{16} \log_2 \frac{1}{16} + \frac{0}{16} \log_2 \frac{0}{16} + \frac{4}{16} \log_2 \frac{4}{16} + \frac{2}{16} \log_2 \frac{2}{16} + \frac{4}{16} \log_2 \frac{4}{16} + \frac{2}{16} \log_2 \frac{2}{16}\right)$$

$$= -\left(\frac{3}{16}(-2,415) + \frac{1}{16}(-4) + \frac{0}{16}(0) + \frac{4}{16}(-2) + \frac{2}{16}(-3) + \frac{4}{16}(-2) + \frac{2}{16}(-3)\right)$$

$$= 2,45$$

Setelah menghitung *entropy*, selanjutnya menghitung *gain* atribut Biologi:

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

$$Gain(S, \text{Biologi}) = Entropy(S) - \left(\frac{8}{24} * E(\text{tinggi}) + \frac{16}{24} * E(\text{rendah})\right)$$

$$= 2,758 - \left(\frac{8}{24} * 1,81125 + \frac{16}{24} * 2,45\right)$$

$$= 2,758 - (0,60375 + 1,6333)$$

$$= 0,52095$$

Setelah *gain* didapatkan, selanjutnya adalah menghitung *split info* dan *gain ratio*

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

$$SplitInfo(\text{Biologi}) = -\left(\frac{8}{24} \log_2 \left(\frac{8}{24}\right) + \frac{16}{24} \log_2 \left(\frac{16}{24}\right)\right)$$

$$= -\left(\frac{8}{24}(-1,585) + \frac{16}{24}(-0,585)\right)$$

$$= 0,9183$$

Dan untuk *Gain Rationya* adalah sebagai berikut

$$Gain Ratio = \frac{Gain(A)}{SplitInfo(S, A)} = \frac{0,52095}{0,9183} = 0,567$$

4.1.6 Mata Pelajaran Fisika

Hitung *information gain* untuk atribut Fisika. Kita bagi nilai menjadi dua kategori:

Nilai tinggi (≥ 90)

Nilai rendah (< 90)

Menghitung *Entropy* Tiap Cabang:

1. *Entropy* cabang tinggi (≥ 90):

Terdapat 17 siswa dengan nilai Fisika tinggi:

$$E(\text{tinggi}) = -\left(\frac{3}{17} \log_2 \frac{3}{17} + \frac{4}{17} \log_2 \frac{4}{17} + \frac{3}{17} \log_2 \frac{3}{17} + \frac{0}{17} \log_2 \frac{0}{17} + \frac{2}{17} \log_2 \frac{2}{17} + \frac{4}{17} \log_2 \frac{4}{17} + \frac{1}{17} \log_2 \frac{1}{17}\right)$$

$$= -\left(\frac{3}{17}(-2,502) + \frac{4}{17}(-2,0875) + \frac{3}{17}(-2,502) + \frac{0}{17}(0) + \frac{2}{17}(-3,087) + \frac{4}{17}(-2,0875) + \frac{1}{17}(-4,0875)\right)$$

$$= 2,46834$$

2. *Entropy* cabang rendah (< 90):

Terdapat 7 siswa dengan nilai Fisika rendah:

$$E(\text{rendah}) = -\left(\frac{0}{7} \log_2 \frac{0}{7} + \frac{0}{7} \log_2 \frac{0}{7} + \frac{0}{7} \log_2 \frac{0}{7} + \frac{5}{7} \log_2 \frac{5}{7} + \frac{0}{7} \log_2 \frac{0}{7} + \frac{0}{7} \log_2 \frac{0}{7} + \frac{2}{7} \log_2 \frac{2}{7}\right)$$

$$= -\left(\frac{0}{7}(0) + \frac{0}{7}(0) + \frac{0}{7}(0) + \frac{5}{7}(-0,4855) + \frac{0}{7}(0) + \frac{0}{7}(0) + \frac{2}{7}(-1,8074)\right)$$

$$= 0,86247$$

Setelah menghitung *entropy*, selanjutnya menghitung *gain* atribut Fisika:

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

$$Gain(S, \text{Fisika}) = Entropy(S) - \left(\frac{17}{24} * E(\text{tinggi}) + \frac{7}{24} * E(\text{rendah})\right)$$

$$= 2,758 - \left(\frac{17}{24} * 2,46834 + \frac{7}{24} * 0,86247\right)$$

$$= 2,758 - (1,74757 + 0,25139)$$

$$= 0,75904$$

Setelah *gain* didapatkan, selanjutnya adalah menghitung *split info* dan *gain ratio*

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

$$SplitInfo(\text{Fisika}) = -\left(\frac{17}{24} \log_2 \left(\frac{17}{24}\right) + \frac{7}{24} \log_2 \left(\frac{7}{24}\right)\right)$$

$$= -\left(\frac{17}{24}(-0,498) + \frac{7}{24}(-1,778)\right)$$

$$= 0,871$$

Dan untuk *Gain Rationya* adalah sebagai berikut

$$Gain Ratio = \frac{Gain(A)}{SplitInfo(S, A)} = \frac{0,75904}{0,871} = 0,871$$

4.1.7 Mata Pelajaran Kimia

Hitung *information gain* untuk atribut Kimia. Kita bagi nilai menjadi dua kategori:

Nilai tinggi (≥ 90)

Nilai rendah (< 90)

Menghitung *Entropy* Tiap Cabang:

1. *Entropy* cabang tinggi (≥ 90):

Terdapat 12 siswa dengan nilai Kimia tinggi:

$$E(\text{tinggi}) = -\left(\frac{0}{12} \log_2 \frac{0}{12} + \frac{4}{12} \log_2 \frac{4}{12} + \frac{3}{12} \log_2 \frac{3}{12} + \frac{0}{12} \log_2 \frac{0}{12} + \frac{2}{12} \log_2 \frac{2}{12} + \frac{0}{12} \log_2 \frac{0}{12} + \frac{3}{12} \log_2 \frac{3}{12}\right)$$

$$= -\left(\frac{0}{12}(0) + \frac{4}{12}(-1,585) + \frac{3}{12}(-2) + \frac{0}{12}(0) + \frac{2}{12}(-2,585) + \frac{0}{12}(0) + \frac{3}{12}(-2)\right)$$

$$= 1,95916$$

2. *Entropy* cabang rendah (< 90):

Terdapat 12 siswa dengan nilai Kimia rendah:

$$E(\text{rendah}) = -\left(\frac{3}{12} \log_2 \frac{3}{12} + \frac{0}{12} \log_2 \frac{0}{12} + \frac{0}{12} \log_2 \frac{0}{12} + \frac{5}{12} \log_2 \frac{5}{12} + \frac{0}{12} \log_2 \frac{0}{12} + \frac{4}{12} \log_2 \frac{4}{12} + \frac{0}{12} \log_2 \frac{0}{12}\right)$$

$$= -\left(\frac{3}{12}(-2) + \frac{0}{12}(0) + \frac{0}{12}(0) + \frac{5}{12}(-1,2631) + \frac{0}{12}(0) + \frac{4}{12}(-1,585) + \frac{0}{12}(0)\right)$$

$$= 1,5555$$

Setelah menghitung *entropy*, selanjutnya menghitung *gain* atribut Kimia:

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

$$Gain(S, \text{Kimia}) = Entropy(S) - \left(\frac{12}{24} * E(\text{tinggi}) + \frac{12}{24} * E(\text{rendah})\right)$$

$$= 2,758 - \left(\frac{12}{24} * 2,483 + \frac{12}{24} * 1,954\right)$$

$$= 2,758 - (1,95916 + 1,5555)$$

$$= -0,75666$$

4.1.8 Menghitung *Split Info* dan *Gain Ratio*

Setelah *gain* didapatkan, selanjutnya adalah menghitung *split info* dan *gain ratio*

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

$$SplitInfo(Kimia) = -(\frac{12}{24} \log_2(\frac{12}{24}) + \frac{12}{24} \log_2(\frac{12}{24})) = -(\frac{12}{24} (-1) + \frac{12}{24} (-1)) = 1$$

Dan untuk *Gain Rationya* adalah sebagai berikut

$$Gain Ratio = \frac{Gain (A)}{SplitInfo(S, A)} = \frac{-0,75666}{1} = -0,756$$

Berikut tabel hasil perhitungan untuk semua atribut:

Tabel 2. Hasil perhitungan

Atribut	<i>Split Point</i>	<i>Gain</i>	<i>Entropy</i>	<i>Split Info</i>	<i>Gain Ratio</i>
MTK		0		0	0
	≥ 90		2,758		
	< 90		0		
Bahasa Inggris		0,4075		0,811	0,502
	≥ 90		2,483		
	< 90		1,954		
Biologi		0,52095		0,9183	0,567
	≥ 90		1,81125		
	< 90		2,45		
Fisika		0,75904		0,871	0,871
	≥ 90		2,46834		
	< 90		0,86247		
Kimia		-0,75666		1	-0,756
	≥ 90		1,95916		
	< 90		1,5555		

Dari tabel diatas dapat dilihat nilai *gain ratio* tertinggi yang akan dijadikan root pertama adalah mata pelajaran Fisika dengan nilai 0,871.

4. CONCLUSION

Berdasarkan hasil penelitian dan pembahasan yang telah dibahas pada bab sebelumnya maka kesimpulan yang dapat diambil dari sistem penerapan decision tree untuk menentukan jurusan kuliah siswa pasca-Madrasah Aliyah Misbahul Ulum sebagai berikut:

(1) Penelitian ini berhasil merancang sebuah sistem untuk mengklasifikasikan jurusan kuliah siswa pasca-Madrasah

Aliyah Misbahul Ulum. Sistem ini dikembangkan dengan menggunakan bahasa pemrograman PHP.

(2) Algoritma C4.5 berhasil diimplementasikan pada sistem untuk menentukan jurusan kuliah siswa. Algoritma ini mampu mengolah data numerik dan kategorikal sehingga cocok untuk menganalisa nilai siswa.

(3) Sistem ini memberikan kemudahan bagi pengguna dalam menentukan jurusan kuliah siswa, sehingga dapat mendukung pengambilan keputusan dalam perencanaan lanjut kuliah.

REFERENCES

- [1] Badrul, M. (2021). Penerapan Metode Profile Matching Untuk Menunjang Keputusan Seleksi Pegawai Baru. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 8(1), 75–82. <https://doi.org/10.30656/prosisko.v8i1.2815>
- [2] Darnila, E., Al-Kautsar, H., & Iksan, Y. (2020). Analysis of Patient Disease Trends Based on Medical Record Data Using the C4. 5 Algorithm. *Login: Jurnal Teknologi ...*, 14(1), 7–12. <http://www.login.seaninstitute.org/index.php/Login/article/download/11/18>
- [3] Devia, E. (2023). Penerapan Decision Tree Dengan Algoritma C4.5 Untuk Menentukan Rekomendasi Kenaikan Jabatan Karyawan. *Jurnal Information System*, 3(1), 28–37. <https://doi.org/10.61488/jis.v3i1.257>
- [4] Effendy, E., Siregar, E. A., Fitri, P. C., & Damanik, I. A. S. (2023). Mengenal Sistem Informasi Manajemen Dakwah (Pengertian Sistem, Karakteristik Sistem). *Jurnal Pendidikan Dan Konseling*, 5(2), 4343–4349.
- [5] Fikry, M., Rizal, Fadlisyah, Nurdin, Bustami, Hamdhana, D., & Qamal, M. (2020). Data Mining for Processing of Research and Community Service by Lecturer Using Decision Tree Method. *International Journal of Psychosocial Rehabilitation, Conference Special Issue*, 24(1475–7192), 367–371.
- [6] Kadir, A. M., Yahya, M., & Mappalotteng, A. M. (2021). Pengembangan Sistem Rekomendasi Pemilihan Program Studi di Universitas Negeri Makassar Menggunakan Metode Decision Tree. *Seminar Nasional LP2M UNM*, 2750–2581.
- [7] Tahir, M. A. (2019). Perancangan Aplikasi Data Mining Menggunakan Metode Decision Tree Untuk Analisa Pemberian Kredit Pada Bri Unit LalabataRilau. *Jurnal Ilmiah Sistem Informasi Dan Teknik Informatika*, 2(1), 1–10.
- [8] Yahya, V. Y., Yudhistira, F., & Manurung, L. (2023). Perancangan Sistem Informasi Pemesanan Sparepart Pengolahan Kelapasawit Pada Cv Ltsu Berbasis Java. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 7(1), 711–717. <https://doi.org/10.30998/semnasristek.v7i1.6404>