

Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online

Kiki Aidi Saputra^{*}, Jaya Tata Hardinata, Muhammad Ridwan Lubis, Sundari Retno Andani, Iham Syahputra Saragih

Program Studi Sistem Informasi STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

Email: *kiki.23aidi@gmail.com

Abstrak—Penelitian ini bertujuan untuk mengetahui pengaruh tingkat kepuasan siswa terhadap media pembelajaran online di SMA YPI Swasta Dharma Budi. Media internet memiliki peranan penting untuk belajar bagi kalangan siswa. Populasi penelitian ini menerapkan Algoritma C4.5 dalam penerapan tingkat kepuasan siswa terhadap media pembelajaran online. Jenis penelitian ini yaitu Klasifikasi dengan konsep data mining dengan melibatkan sebanyak 100 data siswa di SMA YPI Swasta Dharma Budi yang dikategorikan dengan : Puas dan Tidak Puas. Dalam pengambilan data menggunakan kuesioner dan variabel yang digunakan dalam penelitian ini ada empat yaitu minat belajar, cara belajar, sajian pembelajaran, media pembelajaran. Metode yang digunakan dalam penelitian ini yakni metode klasifikasi.

Kata Kunci: Data Mining; Internet; Algoritma C4.5; Klasifikasi

Abstract—This study aims to determine the effect of the level of student satisfaction on online learning media This study aims to determine the effect of the level of student satisfaction on online learning media in SMA YPI Swata Dharma Budi. Internet media has an important role to learn for students. This research population applies the C4.5 Algorithm in applying the level of student satisfaction to online learning media. This type of research is the classification with the concept of data mining involving as many as 100 student data in SMA YPI Swasta Dharma Budi which is categorized by: Satisfied and Dissatisfied. In collecting data using a questionnaire and the variables used in this study there were four namely interest in learning, ways of learning, learning, learning media. The method used in this study is the classification method.

Keywords: Data Mining; Internet; C4.5 Algorithm; Classification

1. PENDAHULUAN

Perkembangan teknologi informasi telah membawa pengaruh yang besar terhadap dunia pendidikan. Berbagai inovasi selalu diusahakan oleh para pelaku pendidikan, misalnya terkait materi pendidikan, dan kebutuhan guru saat mengajar. Media pembelajaran berbasis internet secara konseptual adalah pembelajaran tatap muka dengan dukungan media pembelajaran online menggunakan internet. Seiring dengan perkembangan tersebut metode pembelajaran menggunakan media internet juga banyak mengalami perkembangan. Sehubungan dengan masalah rendahnya minat belajar siswa, upaya pemecahan masalah untuk merangsang minat belajar siswa melalui penyediaan fasilitas belajar salah satunya adalah media internet. Internet di bidang pendidikan sangat berguna dalam proses belajar mengajar di sekolah, di mana para siswa dapat melengkapi ilmu pengetahuannya, sedangkan guru dapat mencari bahan ajar yang *up to date* melalui internet.

Media Pembelajaran adalah bahan, alat, maupun metode/teknik yang digunakan dalam kegiatan belajar mengajar dengan maksud agar proses interaksi komunikasi edukatif antara guru dan siswa dapat berlangsung secara efektif dan efisien. dengan pemanfaatan internet sebagai sumber belajar merupakan salah satu upaya untuk membantu kegiatan belajar. Selain itu internet juga dapat diakses oleh para siswa tanpa batas ruang dan waktu. Berdasarkan dengan hasil penelitian Prima Lestari menjelaskan bahwa penggunaan media internet memberi pengaruh terhadap minat belajar siswa yang menjadikan minat belajar tinggi.

Manfaat yang diperoleh dari hasil penelitian ini adalah untuk memberikan sumbangan pemikiran dalam pengembangan pembelajaran menggunakan media pembelajaran online dalam bidang pendidikan. Hasil penelitian ini dapat dapat digunakan sebagai sebagai bahan evaluasi guru dalam melaksanakan proses pembelajaran. Berdasarkan penjelasan diatas, peneliti menggunakan *data mining* dengan metode Algoritma C4.5 untuk menyelesaikan masalah tersebut. *Data Mining* merupakan proses untuk menemukan pengetahuan yang ditambang atau informasi dengan mencari pola dari sejumlah data yang volumenya sangat besar. Banyak penelitian terdahulu yang dilakukan dalam melakukan prediksi tingkat kepuasan dengan menggunakan *data mining* algoritma C4.

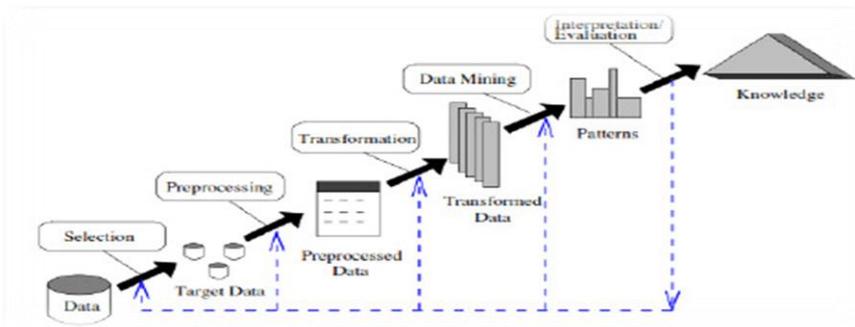
2. METODOLOGI PENELITIAN

2.1 Data Mining

Data Mining berasal dari berbagai ilmu pengetahuan, yang meliputi *machine learning* atau *pattern recognition*, statistik atau kecerdasan buatan, dan sistem basis data (Muflikhah *et al.*, 2018). “*Data Mining* adalah proses penggalian dan pertambangan pengetahuan dari sejumlah data yang besar, database atau repository database lainnya. Tujuan utama dari penambangan data ini untuk menemukan pengetahuan baru yang tersembunyi dari database tersebut” [2].

Defenisi lain dari *Data mining* merupakan salah satu proses inti yang terdapat dalam *Knowledge Discovery in Database* (KDD). Banyak orang memperlakukan data mining sebagai sinonim dari *Knowledge Discovery in Database*

(KDD), karena sebagian besar pekerjaan dalam *Knowledge Discovery in Database* (KDD) difokuskan pada data mining. Namun, langkah-langkah lain merupakan proses-proses penting yang menjamin kesuksesan dari aplikasi KDD [3].



Gambar 1. Skema dari proses KDD [3]

Data mining disebut dengan *Knowledge Discovery in Database* (KDD), yaitu kegiatan yang meliputi pengumpulan, pemakaian data untuk menemukan keteraturan dan serangkaian proses untuk menggali nilai tambah dari sekumpulan data dengan pola dan hubungan dalam set data yang berukuran besar.

Salah satu tugas yang dapat dilakukan dengan *data mining* adalah pengklasifikasian. Klasifikasi merupakan bagian dari suatu *data mining* dengan sistem penyusunan dalam kelompok, golongan dalam standar yang ditetapkan. Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Klasifikasi adalah proses menemukan sekumpulan model yang menggambarkan serta membedakan kelas-kelas data. Tujuan dari klasifikasi adalah agar model yang dihasilkan dapat digunakan untuk memprediksi kelas dari suatu data yang tidak mempunyai label kelas [4]

2.2 Pohon Keputusan (*Decision Tree*)

“*Decision tree* adalah struktur flowchart yang menyerupai tree (pohon), di mana setiap simpul internal menandakan suatu tes atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas” [5]. Pohon keputusan adalah salah satu metode klasifikasi yang kuat dan terkenal. Metode *Decision Tree* mengubah fakta besar menjadi pohon keputusan yang mewakili aturan, sehingga aturan tersebut dapat dengan mudah dipahami oleh manusia. *Decision Tree* juga berfungsi untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input dan variabel tujuan [6].

Pohon keputusan dikembangkan untuk membantu pengambil keputusan membuat serangkaian keputusan yang melibatkan peristiwa ketidakpastian. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan-kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan.

2.3 Algoritma C4.5

Algoritma *C4.5* merupakan metode yang dapat digunakan untuk melakukan pembentukan pohon keputusan. *Decision tree* tersebut mampu menghasilkan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Secara umum metode *C4.5* untuk membangun *decision tree* adalah sebagai berikut [7]:

- a. Pilih atribut sebagai akar. Langkah-langkah untuk memperoleh atribut sebagai akar adalah dengan menghitung jumlah kasus dan jumlah target atribut. Setelah itu menghitung nilai *entropy* yang digunakan untuk menentukan seberapa informatif sebuah input atribut untuk menghasilkan output atribut. Rumus dasar dari *entropy* tersebut

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

dimana:

S : himpunan kasus

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

Persamaan di atas nantinya digunakan untuk menghitung *entropy* total pada kasus ini. Adapun jumlah kasus (S) yang nantinya akan dihitung dengan total kasus awal adalah 274, sedangkan untuk n (jumlah partisi S) adalah banyaknya respon tidak puas dan respon puas dalam kasus ini. Setelah menghitung *entropy* setiap kasus, maka digunakan informasi *gain* untuk pemisahan obyek. Dengan menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

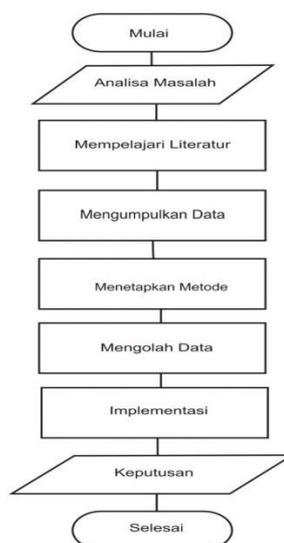
$|S|$: jumlah kasus dalam S

Untuk S (himpunan kasus) adalah sama yaitu 274, sedangkan atribut (A) yang digunakan dalam kasus ini meliputi atribut tangible, reliability, responsiveness, assurance, dan empathy. Sedangkan n adalah jumlah partisi dari masing-masing atribut. Setelah semua atribut dihitung menggunakan persamaan diatas, maka atribut yang memiliki nilai informasi tertinggi dibanding atribut yang lain dijadikan sebagai node (akar).

- b. Buat cabang. Setelah diperoleh atribut yang mempunyai nilai gain tertinggi, maka atribut tersebut digunakan sebagai node. Node ini memiliki instance sehingga instance dijadikan sebagai cabang dari node.
- c. Bagi kasus dalam cabang. Setiap nilai pada instance memiliki nilai yang berbeda. Nilai instance ini diklasifikasikan berdasarkan makna dari nilai instance tersebut agar menjadi lebih sederhana. Tetapi, jika nilai instance tidak dapat disederhanakan lagi maka perlu melakukan perhitungan lebih lanjut.

2.4 Rancangan Penelitian

Pada rancangan penelitian ini penulis menguraikan metodologi dan kerangka penelitian kerja yang digunakan dalam menyelesaikan masalah peneliti. Rancangan penelitian dapat dilihat dalam rancangan *flowchart* pada gambar 1.



Gambar 2. Flowchart Rancangan Penelitian

Pada Gambar 2. menjelaskan rancangan penelitian yang dilakukan untuk mengklasifikasikan tingkat tingkat kepuasan siswa terhadap media pembelajaran online dengan menggunakan algoritma C4.5. Pengumpulan data yang digunakan untuk penelitian terdapat beberapa metode yang terdiri dari :

1. Penelitian Kepustakaan (*Library Research*) yaitu memanfaatkan perpustakaan, buku, prosiding atau jurnal sebagai media untuk referensi dalam menentukan faktor, parameter dan label yang digunakan dalam penelitian.
2. Penelitian Lapangan (*Field Work Research*) yaitu penelitian yang dilakukan secara langsung ke lapangan dengan menggunakan beberapa teknik sebagai berikut :
 - a. Kuisioner
Penulis membuat kuisioner yang dibagikan pada 100 orang siswa, kemudian data yang diperoleh pada kuisioner tersebut sebagai bahan awal untuk menentukan tingkat kepuasan siswa terhadap media pembelajaran online di SMA YPI Swasta Dharma Budi.
 - b. Studi Literatur
Penulis memperoleh informasi dengan mengumpulkan data, mempelajari data, validasi data dan mencari referensi terkait dengan kasus penelitian. Hasil dari studi literatur ini adalah tersusun dan korelasi referensi yang baik dan benar dengan penelitian.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data yang digunakan akan diolah dari hasil kuisioner, dengan jumlah kuisioner kepada siswa dengan jumlah 100 sampel data. Setiap faktor atau variabel terdiri dari pertanyaan yang diajukan ke siswa yang

diberikan. Kuisisioner yang telah dilakukan selanjutnya mencari rata-rata dari faktor atau variabel yang digunakan. Data yang digunakan merupakan jenis *statistik deskriptif* dengan siswa SMA YPI Swata Dharma Budi. Kuisisioner yang diberikan menggunakan *linker 4* yang terdiri dari SB (Sangat Baik), B (Baik), C (Cukup) dan K (Kurang).

Data hasil dari kuisisioner selanjutnya diolah menggunakan *RapidMiner* dengan operator *Performance* yang berfungsi sebagai validasi dan reabilitas data untuk mencari keakuratan data. Data yang akurat dilakukan pengolahan data untuk mencari hasil dari masalah penelitian dengan menggunakan *RapidMiner versi 5.0* dan mengambil keputusan dari hasil klasifikasi dengan algoritma C4.5. berikut ini sampel data hasil rekapitulasi kuisisioner yang diolah di *Microsoft Excel* pada Tabel 1:

Tabel 1. Data Hasil Rekapitulasi Kuisisioner Penelitian

| Responden | C1 | C2 | C3 | C4 | Kepuasan |
|-----------|-------------|-------|-------|-------------|------------|
| M1 | Sangat Baik | Baik | Baik | Sangat Baik | Puas |
| M2 | Baik | Baik | Baik | Baik | Puas |
| M3 | Baik | Cukup | Cukup | Baik | Tidak Puas |
| M4 | Baik | Baik | Baik | Baik | Puas |
| M5 | Sangat Baik | Baik | Baik | Sangat Baik | Puas |
| M6 | Baik | Cukup | Cukup | Baik | Tidak Puas |
| M7 | Sangat Baik | Cukup | Baik | Baik | Puas |
| M8 | Baik | Cukup | Cukup | Baik | Tidak Puas |
| M9 | Sangat Baik | Cukup | Baik | Sangat Baik | Puas |
| M10 | Sangat Baik | Cukup | Cukup | Baik | Puas |
| s/d | | | | | |
| M100 | Baik | Baik | Baik | Baik | Puas |

3.1 Penerapan Algoritma C4.5

Perhitungan algoritma C4.5 dimulai dengan memilih atribut akar terlebih dahulu dengan mencari jumlah kasus keseluruhan, jumlah kasus keputusan puas dan jumlah kasus keputusan tidak puas. Menghitung *entropy* dari semua kasus yang dibagi berdasarkan minat belajar (C1), cara belajar (C2), sajian pembelajaran (C3), dan media pembelajaran (C4). Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan ditunjukkan oleh tabel berikut ini.

Tabel 2. Perhitungan *Node 1*

| | Jumlah (s) | Puas (Si) | Tidak Puas (Si) | Entropy | Gain |
|--------------|------------|-----------|-----------------|------------------|--------------------|
| Total | 100 | 76 | 24 | 0,7950403 | |
| C1 | | | | | 0,351682348 |
| SB | 44 | 41 | 3 | 0,3591016 | |
| B | 40 | 34 | 6 | 0,6098403 | |
| C | 7 | 1 | 6 | 0,5916728 | |
| K | 9 | 0 | 9 | 0,0000000 | |
| C2 | | | | | 0,083701169 |
| SB | 5 | 5 | 0 | 0,0000000 | |
| B | 48 | 38 | 10 | 0,7382849 | |
| C | 44 | 33 | 11 | 0,8112781 | |
| K | 3 | 0 | 3 | 0,0000000 | |
| C3 | | | | | 0,353960518 |
| SB | 11 | 11 | 0 | 0,0000000 | |
| B | 60 | 56 | 4 | 0,3533593 | |
| C | 24 | 9 | 15 | 0,9544340 | |
| K | 5 | 0 | 5 | 0,0000000 | |
| C4 | | | | | 0,210233979 |
| SB | 30 | 24 | 6 | 0,7219281 | |
| B | 61 | 52 | 9 | 0,6036522 | |
| C | 7 | 0 | 7 | 0,0000000 | |
| K | 2 | 0 | 2 | 0,0000000 | |

Kolom *entropy* pada baris total dihitung dengan persamaan (1) sebagai berikut :

$$Entropy (total) = - (76/100)*LOG_2(76/100) - (24/100)*LOG_2(24/100)$$

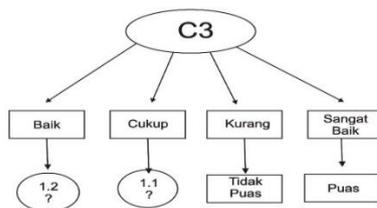
$$Entropy (total) = 0,7950403$$

Kemudian pada nilai *gain* pada baris C3 dihitung persamaan (2) sebagai berikut :

$$Gain (total,C3) = (0,7950403) - ((11/100)*0 + (60/100)*0,3533593 + (24/100)*0,9544340+(5/100)*0)$$

$$Gain (total,C3) = 0,353960518$$

Dari hasil perhitungan pada tabel 3.1 diperoleh atribut yang menjadi *node* (akar) adalah C3 memiliki gain tertinggi yaitu 0,353960518, dimana terdiri dari 4 sub atribut yaitu SB, B, C, dan K. Berdasarkan nilai *entropy* dari keempat sub atribut diatas, sub atribu SB dan K yang memperoleh keputusan. Maka perlu dilakukan perhitungan lebih lanjut untuk menentukan *node* akar selanjutnya, dapat digambarkan pohon keputusan dari tabel diatas sebagai berikut :

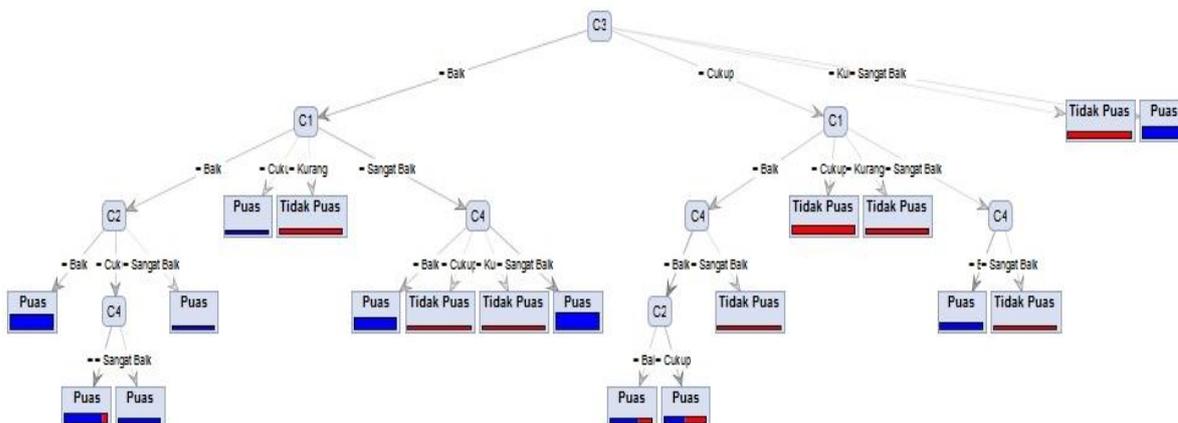


Gambar 3. Pohon Keputusan 1

3.2 Pohon Keputusan Menggunakan Aplikasi

Setelah analisis kepuasan siswa terhadap media pembelajaran online di SMA YPI Swasta Dharma Budi menggunakan algoritma C4.5 dihitung dan pembuatan pohon keputusan menggunakan perhitungan manual, maka tahap berikutnya adalah pembuktian dari analisis dan perhitungan tersebut menggunakan aplikasi RapidMiner.

Adapun hasil dari pembahasan pohon keputusan menggunakan aplikasi RapidMiner dapat dilihat pada gambar 4.



Gambar 4. Pohon Keputusan Hasil Analisis

Setelah dilakukan perhitungan dan pengujian data pada masing-masing tribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir. Dari gambar dapat terlihat bahwa pohon keputusan hasil analisis menggunakan aplikasi RapidMiner sama dengan hasil analisis dengan perhitungan manual.

1. Tampilan deskripsi *decision tree*

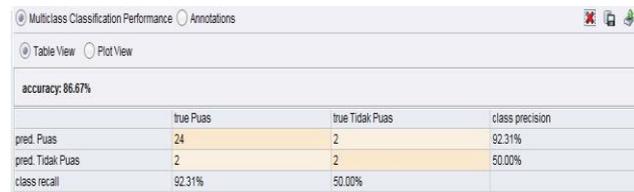
```

Result Overview | PerformanceVector (Performance) | ExampleSet (Read Exce
Graph View | Text View | Annotations
Tree
C3 = Baik
| C1 = Baik
| | C2 = Baik: Puas {Puas=15, Tidak Puas=0}
| | C2 = Cukup
| | | C4 = Baik: Puas {Puas=8, Tidak Puas=1}
| | | C4 = Sangat Baik: Puas {Puas=3, Tidak Puas=0}
| | C2 = Sangat Baik: Puas {Puas=1, Tidak Puas=0}
| C1 = Cukup: Puas {Puas=1, Tidak Puas=0}
| C1 = Kurang: Tidak Puas {Puas=0, Tidak Puas=2}
| C1 = Sangat Baik
| | C4 = Baik: Puas {Puas=11, Tidak Puas=0}
| | C4 = Cukup: Tidak Puas {Puas=0, Tidak Puas=1}
| | C4 = Kurang: Tidak Puas {Puas=0, Tidak Puas=1}
| | C4 = Sangat Baik: Puas {Puas=17, Tidak Puas=0}
C3 = Cukup
| C1 = Baik
| | C4 = Baik
| | | C2 = Baik: Puas {Puas=2, Tidak Puas=1}
| | | C2 = Cukup: Puas {Puas=3, Tidak Puas=3}
| | C4 = Sangat Baik: Tidak Puas {Puas=0, Tidak Puas=1}
| C1 = Cukup: Tidak Puas {Puas=0, Tidak Puas=6}
| C1 = Kurang: Tidak Puas {Puas=0, Tidak Puas=3}
| C1 = Sangat Baik
| | C4 = Baik: Puas {Puas=4, Tidak Puas=0}
| | C4 = Sangat Baik: Tidak Puas {Puas=0, Tidak Puas=1}
C3 = Kurang: Tidak Puas {Puas=0, Tidak Puas=4}
C3 = Sangat Baik: Puas {Puas=11, Tidak Puas=0}
    
```

Gambar 5. Deskripsi *decision tree*

Gambar di atas menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*decision tree*) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa penggunaan data mining algoritma C4.5 baik digunakan dalam proses menggali data (*data mining process*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*decision tree*).

2. Tingkat Akurasi



| | true Puas | true Tidak Puas | class precision |
|------------------|-----------|-----------------|-----------------|
| pred. Puas | 24 | 2 | 92.31% |
| pred. Tidak Puas | 2 | 2 | 50.00% |
| class recall | 92.31% | 50.00% | |

Gambar 6. Nilai Akurasi Algoritma C4.5



```
PerformanceVector:
accuracy: 86.67%
ConfusionMatrix:
True: Puas Tidak Puas
Puas: 24 2
Tidak Puas: 2 2
```

Gambar 7. Nilai Performance Vector Algoritma C4.5

Berdasarkan pengolahan data menggunakan *software RapidMiner* didapat nilai akurasi sistem sebesar 86.67%, artinya bahwa *rule* yang dihasilkan tingkat kebenaran mendekati 100%. Dimana model yang telah dibentuk diuji tingkat akurasinya dengan memasukan atau uji yang berasal dari data *training* dengan menggunakan *split validation* pada aplikasi *rapidminer* 5.3 untuk menguji tingkat akurasi.

4. KESIMPULAN

Berdasarkan seluruh hasil penelitian yang telah dilakukan pada Penerapan Algoritma C4.5 pada Klasifikasi tingkat kepuasan siswa terhadap media pembelajaran online. Permasalahan menentukan faktor kepuasan siswa terhadap media pembelajaran online dapat diselesaikan menggunakan teknik *data mining*. Yaitu dengan Algoritma C4.5. Menghasilkan 20 (dua puluh) *rules* dan Tingkat akurasi yang dihasilkan oleh metode tersebut adalah 86.67% dan faktor paling dominan adalah sajian pembelajaran (C3) dengan nilai gain sebesar 0,353960518.

REFERENCES

- [1] L. Muflikhah, D. E. Ratnawati, and R. R. M. Putri, "Data Mining."
- [2] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT . Arupadhatu Adisesanti," vol. 2, no. 1, pp. 36–41, 2017.
- [3] S. Lorena, W. Zarman, and I. Hamidah, "Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik," *Pros. Semin. Nas. Apl. Sains dan Teknol.*, no. November, pp. 263–272, 2014, doi: 10.5829/idosi.weasj.2015.6.2.22162.
- [4] A. Fristi, "Penerapan algoritma c4.5 untuk mengukur tingkat kepuasan mahasiswa," vol. 3, no. 2, pp. 1–7, 2019.
- [5] Rismayanti, "Implementasi algoritma c4.5 untuk menentukan penerima beasiswa di stt harapan medan," vol. 12, no. 2, pp. 116–120, 2016.
- [6] Baiq Andriskha Candra, "Prediksi Kepuasan Mahasiswa Terhadap Tingkat Pelayanan Menggunakan Algoritma C4.5 (Decision Tree) (Studi Kasus : Fakultas Teknik Universitas Hamzanwadi) Baiq," vol. 1, no. 1, pp. 32–39, 2018.
- [7] S. Kasus and S. Pringsewu, "ANALISIS KEPUASAN MAHASISWA TERHADAP PELAYANAN AKADEMIK MENGGUNAKAN METODE ALGORITMA C4 . 5," vol. 02, no. 01, pp. 1–11, 2016.