



PENERAPAN DATA MINING UNTUK PREDIKSI MINAT BELAJAR SISWA MENGGUNAKAN ALGORITMA C4.5

Nurul Abdilla¹, Zelvi Gustiana² & Amru Yasir³

1,2,3)Teknologi Informasi, Teknik dan Ilmu Komputer, Universitas Dharmawangsa, Indonesia

Email: nurulabdilla590@gmail.com

Abstrak

Pendidikan dalam kehidupan memegang peranan yang sangat penting untuk menjamin kelangsungan hidup seseorang dimasa depan, terlebih kepada para siswa/siswi yang sedang menempuh pendidikan. Dalam hal ini akan dilakukan penentuan seberapa besar minat siswa dalam melakukan kegiatan pembelajaran. Maka dari itu dibuatlah prediksi minat belajar menggunakan metode Algoritma C4.5. Algoritma C4.5 merupakan salah satu metode *data mining* untuk memprediksi kemampuan siswa dalam menyelesaikan bidang studi dilihat dari nilai hasil belajar siswa. Yang hasil akhirnya merupakan sebuah bentuk pohon keputusan dengan nilai akurasi sebesar 90%.

Kata Kunci : Algoritma C4.5, Minat Belajar, Pendidikan, Siswa

Abstract

Education in life plays a very important role in ensuring a person's survival in the future, especially for students who are currently studying. In this case, students will be determined how interested they are in carrying out learning activities. Therefore, a prediction of learning interest was made using the C4.5 Algorithm method. The C4.5 algorithm is a data mining method for predicting students' ability to complete a field of study based on student learning outcomes. The final result is a decision tree with an accuracy value of 90%.

Keyword : Algoritma C4.5, Minat Belajar, Pendidikan, Siswa

PENDAHULUAN

Kecenderungan yang terjadi saat ini, banyak siswa yang tidak tahu minat dan bakatnya dalam menentukan minat belajar mana yang sangat disukai selama menempuh pendidikan. Sehingga menurunnya kualitas dan minat belajar para siswa karna merasa tidak menemukan bidang atau minat yang cocok untuk dirinya. Banyak kasus yang dijumpai bahwa pemilihan minat yang tidak sesuai dengan kemampuan, dan bakat dapat mempengaruhi para siswa dalam mengikuti pembelajaran di sekolah. Dalam beberapa penelitian psikologi pendidikan, minat dan bakat siswa diketahui cukup terkait dengan prestasi akademiknya.

Sementara itu, *data mining* adalah proses yang menggunakan statistik, matematika, kecerdasan buatan, *machine learning* untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* disisi lain adalah kegiatan meliputi pengumpulan, pemakaian, data histori untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan. Untuk penentuan semacam ini, beberapa ahli telah menggunakan data mining untuk melakukannya. Mereka memanfaatkan sifat prediksi yang dimiliki pohon keputusan. Dalam data mining banyak algoritma yang dapat dipakai dalam pohon keputusan, salah satunya algoritma C4.5.

KAJIAN TEORI

Untuk mendapatkan hasil penelitian yang objektif dan memperoleh wawasan ilmu pengetahuan yang lebih luas, maka peneliti menggunakan data laporan nilai hasil belajar siswa untuk mengetahui seberapa besar minat dan bakat siswa dalam belajar. Berikut ini akan di jelaskan mengenai masing masing variabel.

1. Data Mining

Data mining adalah teknologi yang merupakan campuran teknik-teknik analisis data dengan algoritma-algoritma untuk memproses data berukuran besar.

2. Pohon keputusan (*Decision Tree*)

Decision Tree atau sering juga disebut pohon keputusan merupakan metode pengklasifikasian menggunakan struktur pohon. Pohon keputusan dimulai dari satu *node* atau simpul lalu *node* tersebut akan bercabang – cabang yang mana cabang tersebut menyatakan berbagai pilihan.

3. Metode C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan dalam data mining untuk membangun model pohon keputusan. Algoritma C4.5 menggunakan konsep pembelajaran mesin untuk menghasilkan model prediksi berdasarkan data yang diberikan.

METODE PENELITIAN

1. Metode Studi Pustaka

Studi pustaka merupakan suatu metode pengumpulan data dengan mencari informasi yang relevan dengan tujuan penelitian. Informasi ini tersedia dari berbagai referensi, termasuk buku, jurnal, situs web, dan sumber lainnya.

2. Wawancara

Salah satu metode pengumpulan data ini dengan menanyakan langsung ke responden atau narasumber untuk mendapatkan sebuah informasi. Pada penelitian ini dilakukan wawancara dengan 1 narasumber dari MTs. Islamiyah YPI Batang Kuis yaitu dengan Bapak M. Iqbal S.pd.I selaku kepala sekolah.

3. Observasi

Observasi merupakan suatu metode pengumpulan data dengan mengunjungi langsung ke lokasi penelitian guna mengumpulkan informasi yang didapat. Observasi yang dilakukan pada penelitian ini mendatangi langsung MTs. Islamiyah YPI Batang Kuis. Serta meminta keterangan data nilai siswa sebagai acuan dalam menentukan minat belajar siswa

HASIL DAN PEMBAHASAN

Langkah-langkah Algoritma C4.5:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai atribut
3. Bagi kasus di dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Hasil penelitian dapat disajikan dengan dukungan tabel, grafik atau gambar sesuai kebutuhan, untuk memperjelas penyajian hasil secara verbal. Judul tabel dan grafik atau keterangan gambar disusun dalam bentuk frase (bukan kalimat) secara ringkas.

Untuk memilih atribut akar, didasarkan pada nilai GAIN tertinggi dari atribut-atribur yang ada. Untuk mendapatkan nilai GAIN, harus ditentukan terlebih dahulu nilai ENTROPY.

Solusi :

1. Hitung Jumlah Kasus
2. Hitung Keputusan “Minat”
3. Hitung Jumlah Keputusan “Tidak Minat”
4. Hitung : - Entropy (Hasil),
 - a. Entropy (Keterampilan),
 - b. Entropy (Kerajinan),
 - c. Entropy (Kesopanan), dan
 - d. Entropy (Kehadiran)

Tabel 1 Data Testing

No	Keterampilan	Kerajinan	Kesopanan	Kehadiran	Hasil
1	Baik Sekali	Baik	Baik	Baik	Minat
2	Baik Sekali	Baik Sekali	Cukup	Baik	Minat
3	Baik Sekali	Cukup	Baik	Baik	Minat
4	Baik	Baik	Baik Sekali	Baik	Minat
5	Baik	Baik	Baik Sekali	Baik	Minat
6	Baik Sekali	Baik Sekali	Cukup	Baik	Minat
7	Baik Sekali	Cukup	Baik	Baik	Minat
8	Cukup	Baik	Cukup	Baik	Tidak Minat
9	Cukup	Cukup	Baik Sekali	Cukup	Tidak Minat
10	Baik	Baik Sekali	Baik Sekali	Baik Sekali	Minat
11	Baik Sekali	Cukup	Baik	Baik Sekali	Minat
12	Baik Sekali	Baik Sekali	Baik Sekali	Baik Sekali	Minat
13	Baik	Baik	Baik Sekali	Baik Sekali	Minat
14	Baik	Cukup	Baik	Baik Sekali	Minat
15	Cukup	Baik Sekali	Baik	Baik Sekali	Minat
16	Baik	Baik	Baik Sekali	Baik Sekali	Minat
17	Baik	Cukup	Baik	Baik Sekali	Minat
18	—	—	—	—	—
19	—	—	—	—	—
20	—	—	—	—	—
30	Baik	Baik Sekali	Baik Sekali	Cukup	Minat
31	Baik	Baik	Baik	Cukup	Minat
32	Baik	Baik	Baik	Cukup	Minat
33	Baik	Baik Sekali	Baik Sekali	Cukup	Minat
34	Baik	Cukup	Cukup	Cukup	Tidak Minat

Rumus *Entropy* :

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

Keterangan :

S = Jumlah sampel data (*Sampling*)

n = Jumlah Partisi S

p_i = Proporsi dari S_i terhadap S.

Rumus Gain :

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

Keterangan :

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

S_i = Jumlah Kasus Pada Partisi Ke-1

S = Jumlah Kasus Dalam S

Perhitungan Nilai Entropy

Entropy (Total) :

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

Tabel 2. Perhitungan Nilai Entropy

Node		Jumlah Kasus	Minat	Tidak Minat	Entropy	Gain
1	Hasil	34	30	4		

$$Entropy(Total) = \left(-\frac{4}{34} * \log_2 \left(\frac{4}{34}\right)\right) + \left(-\frac{30}{34} * \log_2 \left(\frac{30}{34}\right)\right)$$

$$Entropy(Total) = 0,522559375$$

Tabel 3. Hasil Perhitungan Entropy

Node		Jumlah Kasus	Minat	Tidak Minat	Entropy	Gain
1	Hasil	34	30	4	0,522559375	

Perhitungan Nilai Entropy :

- HASIL

$$Entropy Total = \left(-\frac{4}{34} * \log_2 \left(\frac{4}{34}\right)\right) + \left(-\frac{30}{34} * \log_2 \left(\frac{30}{34}\right)\right)$$



Entropy Total = 0,522559375

- KETERAMPILAN

$$\begin{aligned} \text{Entropy Total : Baik} &= \left(-\frac{1}{17} * \log_2 \left(\frac{1}{17}\right)\right) + \left(-\frac{16}{17} * \log_2 \left(\frac{16}{17}\right)\right) \\ &= 0,32275696 \end{aligned}$$

$$\begin{aligned} \text{Baik Sekali} &= \left(-\frac{1}{12} * \log_2 \left(\frac{1}{12}\right)\right) + \left(-\frac{11}{12} * \log_2 \left(\frac{11}{12}\right)\right) \\ &= 0,41381685 \end{aligned}$$

$$\begin{aligned} \text{Cukup} &= \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5}\right)\right) \\ &= 0,64384121 \end{aligned}$$

- KERAJINAN

$$\begin{aligned} \text{Entropy Total: Baik} &= \left(-\frac{1}{11} * \log_2 \left(\frac{1}{11}\right)\right) + \left(-\frac{10}{11} * \log_2 \left(\frac{10}{11}\right)\right) \\ &= 0,439496987 \end{aligned}$$

$$\begin{aligned} \text{Baik Sekali} &= \left(-\frac{0}{13} * \log_2 \left(\frac{0}{13}\right)\right) + \left(-\frac{13}{13} * \log_2 \left(\frac{13}{13}\right)\right) \\ &= -1 \end{aligned}$$

$$\begin{aligned} \text{Cukup} &= \left(-\frac{3}{10} * \log_2 \left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2 \left(\frac{7}{10}\right)\right) \\ &= 0,8812909 \end{aligned}$$

- KESOPANAN

$$\begin{aligned} \text{Entropy Total : Baik} &= \left(\frac{0}{17} * \log_2 \left(\frac{0}{17}\right)\right) + \left(-\frac{17}{17} * \log_2 \left(\frac{17}{17}\right)\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Baik Sekali} &= \left(-\frac{0}{12} * \log_2 \left(\frac{0}{12}\right)\right) + \left(-\frac{12}{12} * \log_2 \left(\frac{12}{12}\right)\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Cukup} &= \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) \\ &= 0,97095059 \end{aligned}$$

- KEHADIRAN

$$\begin{aligned} \text{Entropy Total: Baik} &= \left(-\frac{1}{8} * \log_2 \left(\frac{1}{8}\right)\right) + \left(-\frac{7}{8} * \log_2 \left(\frac{7}{8}\right)\right) \\ &= 0,54356444 \end{aligned}$$

$$\begin{aligned} \text{Baik Sekali} &= \left(-\frac{0}{16} * \log_2 \left(\frac{0}{16}\right)\right) + \left(-\frac{16}{16} * \log_2 \left(\frac{16}{16}\right)\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Cukup} &= \left(-\frac{3}{10} * \log_2 \left(\frac{3}{10}\right)\right) + \left(-\frac{7}{10} * \log_2 \left(\frac{7}{10}\right)\right) \\ &= 0,8812909 \end{aligned}$$

Perhitungan Nilai Gain

Menghitung Gain :

1. Hitung *Gain* (Total, Keterampilan)
2. Hitung *Gain* (Total, Kerajinan)
3. Hitung *Gain*(Total, Kesopanan)
4. Hitung *Gain* (Total, Kehadiran)

Perhitungan Nilai *Gain* :

Gain(Total, Keterampilan)

$$= \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{\text{Keterampilan}}{\text{Total}} = \text{Entropy} (\text{Keterampilan})$$

$$\begin{aligned} \text{Gain}(\text{Total, Keterampilan}) &= 0,522559375 - \left(\frac{17}{34} * 0,32275696\right) + \left(\frac{12}{34} * 0,41381685\right) + \left(\frac{5}{34} * 0,64384121\right) \end{aligned}$$

$$\text{Gain}(\text{Total, Keterampilan}) = 0,120445358235$$

$$\text{Gain}(\text{Total, Kerajinan}) = \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{\text{Kerajinan}}{\text{Total}} = \text{Entropy} (\text{Kerajinan})$$

$$\text{Gain}(\text{Total, Kerajinan}) = 0,522559375 - \left(\frac{11}{34} * 0,439496987\right) + \left(\frac{13}{34} * -1\right) + \left(\frac{10}{34} * 0,8812909\right)$$

$$\text{Gain}(\text{Total, Kerajinan}) = 0,603518909$$

$$\text{Gain}(\text{Total, Kesopanan}) = \text{Entropy}(\text{Total}) - \sum_{i=1}^n \frac{\text{Kesopanan}}{\text{Total}} = \text{Entropy} (\text{Kesopanan})$$

$$\text{Gain}(\text{Total, Kesopanan}) = 0,522559375 - \left(\frac{17}{34} * 0\right) + \left(\frac{12}{34} * 0\right) + \left(\frac{5}{34} * 0,97095059\right)$$

$$\text{Gain}(\text{Total, Kesopanan}) = 0,679772524$$

$$Gain(Total, Kehadiran) = Entropy(Total) - \sum_{i=1}^n \frac{Kehadiran}{Total} = Entropy(Kehadiran)$$

$$Gain(Total, Kehadiran) = 0,522559375 - \left(\frac{8}{34} * 0,54356444\right) + \left(\frac{16}{34} * 0\right) + \left(\frac{10}{34} * 0,80812909\right)$$

$$Gain(Total, Kehadiran) = 0,135458654$$

Tabel 4. Hasil Perhitungan Nilai Entropy dan Gain

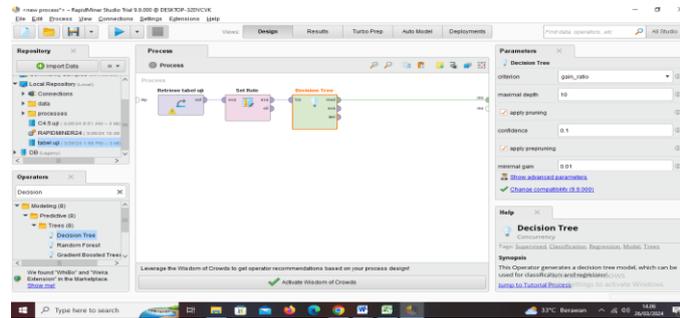
Node			Jumlah Kasus	Minat	Tidak Minat	Entropy	Gain
1	Hasil		34	30	4	0,522559375	
	Keterampilan						0,120445358
		Baik	17	16	1	0,32275696	
		Baik Sekali	12	11	1	0,41381685	
		Cukup	5	3	2	0,64384121	
	Kerajinan						0,603518909
		Baik	11	10	1	0,439496987	
		Baik Sekali	13	13	0	-1	
		Cukup	10	7	3	0,8812909	
	Kesopanan						0,679772524
		Baik	17	17	0	0	
		Baik Sekali	12	12	0	0	
		Cukup	5	2	3	0,97095059	
	Kehadiran						0,135458654
		Baik	8	7	1	0,54356444	
		Baik Sekali	16	16	0	0	
		Cukup	10	7	3	0,8812909	

Hitung Entropy dan Gain di Exel

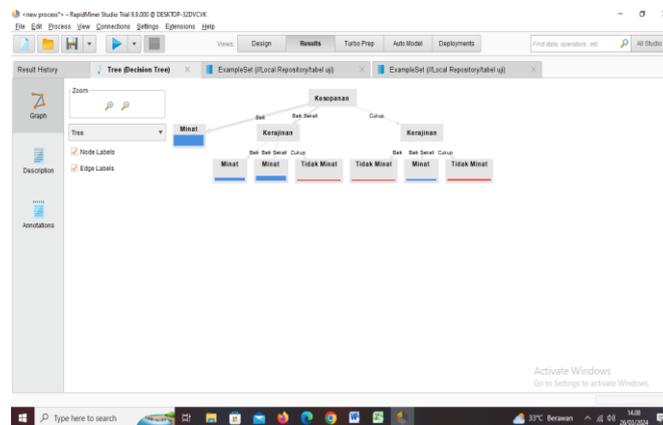
- Rumus menghitung Entropy pada excel :
- $Entropy(S) = (-kasus(Tidak(S1))/jumlah_kasus * \log((kasus(Tidak(S1))/jumlah_kasus);2)) + (-kasus(Yes(S2))/jumlah_kasus * \log((kasus(Yes(S2))/jumlah_kasus);2))$
- Rumus menghitung gain pada excel :

- $Gain(S,A) = \text{entropy total} - ((\text{jumlah_kasus_atribut1} / \text{jumlah_kasus} * \text{entropy_atribut1}) + (\text{jumlah_kasus_atribut2} / \text{jumlah_kasus} * \text{entropy_atribut1}))$

Dalam pengujian Rapid Miner ini dengan memakai decision tree peneliti telah menemukan jawaban yang sama dengan hasil perhitungan manual.



Gambar 1. Tampilan Proses Rapid Miner



Gambar 2. Tampilan hasil



Gambar 3. Tampilan Hasil Lebih Jelas

Setelah semua atribut disambungkan, sekarang pilih menu berbentuk segitiga terbalik dibagian atas sebelah kanan pada tampilan layar untuk menjalankan pengujian dan mendapatkan hasil akhir dari pohon keputusan.

Dari pengujian ini maka didapatkan *rules* sebagai berikut :

1. Jika kesopanan = Baik, Maka siswa tersebut berminat.
2. Jika kesopanan = Baik sekali. Dan kerajinan nya baik, maka siswa tersebut berminat.
3. Jika kesopanan = Baik sekali. Dan kerajinan nya baik sekali, maka siswa tersebut berminat.
4. Jika kesopanan = Baik sekali. Dan kerajina nya cukup, maka siswa tersebut tidak berminat.
5. Jika kesopanan = Cukup. Dan kerajinan nya baik, maka siswa tersebut tidak berminat.
6. Jika kesopanan = Cukup. Dan kerajinan nya baik sekali, maka siswa tersebut berminat.
7. Jika kesopanan = Cukup. Dan kerajinan nya cukup, maka siswa tersebut tidak berminat.

SIMPULAN

Berdasarkan seluruh hasil tahapan penelitian yang telah dilakukan pada Penerapan Klasifikasi *Decision Tree* dengan Algoritma C4.5 untuk menentukan minat belajar siswa dapat disimpulkan sebagai berikut :

1. Data yang diperoleh dari penelitian ini dapat diolah dengan proses data mining dengan jenis klasifikasi nya menggunakan Algoritma C4.5.
2. Hasil perhitungan Algoritma C4.5 Menetapkan kelompok yang memiliki nilai gain paling terbesar.
3. Berdasarkan proses perhitungan yang telah dilakukan melalui beberapa langkah pada proses penerapan data mining dengan Algoritma C4.5. Maka, tingkat akurasi yang dihasilkan oleh metode tersebut adalah 90.00% dan mendapatkan sebanyak 7 *rules* dari hasil pengujian ini.



4. Dengan adanya penerapan data mining algoritma C4.5 diharapkan mampu memberikan solusi dalam menentukan minat belajar siswa di MTs. Islamiyah YPI Batang Kuis.

DAFTAR PUSTAKA

- Azhari, D. Z., Damanik, I. S., & Suhendro, D. (2022). *Penerapan Algoritma C4.5 Untuk Klasifikasi Tingkat Pemahaman Siswa Pada Mata Pelajaran Matematika*. 1(1).
- Halimah, D., Lubis, M. R., & Saputra, W. (2022). Algoritma C4.5 Untuk Menentukan Klasifikasi Tingkat Pemahaman Mahasiswa Pada Matakuliah Bahasa Pemrograman. *Jurnal Teknik Mesin, Industri, Elektro Dan Informatika*, 1(3), Article 3. <https://doi.org/10.55606/jtmei.v1i3.534>
- Harahap, P. N., & Sulindawaty, S. (2019). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus Pt.Arma Anugerah Abadi Cabang Sei Rampah). *Matics: Jurnal Ilmu Komputer Dan Teknologi Informasi (Journal Of Computer Science And Information Technology)*, 11(2), Article 2. <https://doi.org/10.18860/mat.v11i2.7821>
- Hendri, N., & Arma F., (2022). Implementasi Data Mining Untuk Mengetahui Minat Siswa Terhadap Perguruan Tinggi. *Jurnal Teknik Informatika Dan Teknologi Informasi*, <https://ejournal.politeknipratama.ac.id/jtmei.v1i3.534>
- Manullang, N., Sembiring, R. W., Gunawan, I., Parlina, I., & Irawan, I. (2021). Implementasi Teknik Data Mining Untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5. *Jurnal Ilmu Komputer Dan Teknologi*, 2(2), 1–5. <https://doi.org/10.35960/ikomti.v2i2.700>
- M.Arfa C, & Ika, A.W.(2021). Sistem Informasi Berprestasi Berbasis Web Ada Smp Negeri 7 Kota Metro. *Jurnal Mahasiswa Ilmu Komputer (Jmik)*.
- Ningsih, W. R. (2020). *Penerapan Data Mining Dalam Menentukan Masa Studi Mahasiswa Menggunakan Algoritma C4.5* [Skripsi, Universitas Islam Negeri Sumatera Utara Medan]. <http://repository.uinsu.ac.id/15147/>
- Risky, E.P., (2021). Implementasi Data Mining Untuk Prediksi Efektivitas Pada Mesin



COMPTECH

Jurnal Ilmu Komputer Dan Teknologi

Vol. 2 No. 1 September 2025, pp. 24-35

<https://jurnal.compartdigital.com/index.php/comptech>

Injection Menggunakan Algoritma C4.5 [*Skripsi, Universitas Islam Negeri Syarif Hidayarullah Jakarta*]. [Https://Repository.Uinjkt.Ac.Id](https://Repository.Uinjkt.Ac.Id)

Rizky, H.P., Budi, D.S.,& Indriati (2018). Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, E-Issn: 2548-964x, [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)

Syaiful, B.,& Akhyar (2020). Metode Klasifikasi Desicion Tree Untuk Prediksi Juara English Premier League. *Jurnal Sintaksis Pendidikan Guru Sekolah Dasar, Ipa, Ips Dan Bahasa Inggris*; E-Issn: 2715-6176, [Http://Jurnal.Stkipalmaksum.Ac.Id/](http://Jurnal.Stkipalmaksum.Ac.Id/)

Zelvi Gustiana, (2020). Penerapan Algoritma C4.5 Dalam Sistem Pendukung Keputusan Evaluasi Kinerja Fasilitator Pamsismas(Studi Kasus Dikabupaten Kampar). [*Jurnal Of Information Teknologi Research*].