



Available online at : <http://bit.ly/InfoTekJar>  
**InfoTekJar : Jurnal Nasional Informatika dan  
 Teknologi Jaringan**

ISSN (Print) 2540-7597 | ISSN (Online) 2540-7600



# Pencarian Klasifier Terbaik sebagai Solusi Rekrutmen Karyawan Menggunakan Algoritma Iterative Dichotomiser 3 (ID3)

Nilam Ramadhani, Abd. Wahab Syahroni, Irwan Darmawan

Program Studi Informatika, Fakultas Teknik, Universitas Madura Jl. Raya Panglegur km 3.5 Pamekasan, Indonesia

### KEYWORDS

Data Mining, Decision Tree, Iterative Dichotomiser 3 (ID3), Employee recruitment.

### CORRESPONDENCE

Phone: 081249925468

E-mail: nilam\_ramadhani@unira.ac.id

### ABSTRACT

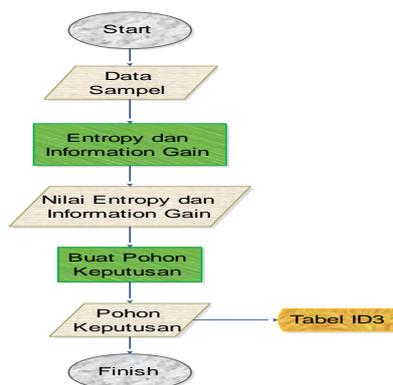
CV Aliyah Mandiri Pamekasan is a growing distribution fabric and apparel company. The worker is drawn through the employee recruitment process by passing a series of tests. But sometimes there is a subjectivity factor that becomes problem in recruitment process even criterias have been established. To avoid that, needed a system tahat can help HRD manager to decide which applicants are eligible for admission. One of them is Decision Tree. This method can find discrete approximation function and resistant to noise data. Also able to study disjunctive expressions. The Iterative Dychotomizer 3 (ID3) algorithm can build a top-down decision tree using information gain as a measure of the effectiveness of an attribute in classifying incomplete data sample sets. The results of the algorithm implementation can provide the best classifier knowledge on employee recruitment, so that it provides decision support considerations.

## PENDAHULUAN

Pada perusahaan yang sedang berkembang seperti CV Aliyah Mandiri terkadang ada permasalahan yang terjadi, misal seperti proses rekrutmen karyawan. Hal ini dikarenakan banyaknya peminat pendaftar kerja dan kebutuhan kriteria karyawan yang cukup ketat. Sehingga diperlukan sebuah cara atau metode seleksi rekrutmen karyawan yang memudahkan dan akurat. Salah satu metode yang bisa diterapkan untuk permasalahan yang berkenaan dengan data yang banyak adalah data mining [1],[2]. Data mining dengan klasifikasi menggunakan algoritma dipilih karena bisa memberikan hasil pengetahuan terbaik sebagai pertimbangan pada saat seleksi dan klasifikasi[3]. Penelitian sebelumnya yang menggunakan algoritma ID3 terkait seleksi cukup banyak. Dari penelitian tersebut didapat hasil klasifikasi yang memberikan pengetahuan sehingga memudahkan dalam pengambilan keputusan[4]. Pada penelitian ini diberikan perangkaan pada hasil akhir klasifikasi. Hal ini akan memudahkan pengambil keputusan untuk melihat hasil akhir seleksi berdasarkan skor setelah melalui proses klasifikasi algoritma ID3[5].

## METHOD

Pada penelitian ini, tahapan yang dilakukan menggunakan algoritma ID3 seperti pada gambar 1.



Gambar 1. Tahapan penelitian yang akan dilakukan

Berdasarkan pada gambar 1, tahapan metode yang dilakukan dapat diurai sebagai berikut[6] :

1. Pengambilan sampel data  
 Pada tahapan ini, diambil data set yang akan dilakukan pengujian sebagai sampel dalam penerapan algoritma data mining.
2. Menghitung nilai *Entropy* dan *Information Gain* untuk

menentukan *the best classifier*.

Tahapan ini diperlukan agar didapatkan informasi dalam penentuan root dan node sesuai dengan nilai gain yang didapat. Tahapan ini mengikuti alur algoritma ID3[7].

3. Melakukan Konstruksi Pohon Keputusan.  
Tahap ini akan memberikan gambaran hasil dari pohon keputusan yang sudah dapat menentukan jalur klasifikasi.
4. Melakukan operasi konjungsi dan Penentuan Peringkat.  
Tahap ini melakukan operasi konjungsi berupa kaidah IF THEN dengan konjungsi operator AND terhadap atribut node yang ditemukan[8]. Pada tahapan ini juga dilakukan penentuan peringkat dari keseluruhan dataset sampel, sehingga ditemukan perangnya.

## HASIL DAN PEMBAHASAN

Pada bab ini akan dijabarkan hasil dan pembahasan dari penelitian yang dilakukan sesuai tahapan dari metode penelitian.

### A. Penentuan Sampel Data dan Atribut Kriteria Penilaian

Data set sampel yang digunakan adalah data dari beberapa pelamar yang mengikuti rekrutmen karyawan baru bagian Pemasaran di CV Aliyah Mandiri Pamekasan. Data set sampel yang dipakai sebanyak 54 record. Atribut data set terdiri dari 4 kriteria penilaian. Kriteria penilaian tersebut merupakan ketentuan yang sudah ditetapkan oleh perusahaan CV Aliyah Mandiri. Masing-masing kriteria dibagi lagi menjadi beberapa kelompok atribut untuk memudahkan dalam penentuan klasifikasinya[9]. Empat kriteria tersebut adalah sebagai berikut :

1. Administrasi di kelompokkan dalam 3 katagori (Bagus, Cukup, Rendah).
2. Kesehatan di kelompokkan dalam 3 katagori (Baik, Sedang, Buruk).
3. Akademik di kelompokkan dalam 3 katagori (Tinggi, Menengah, Rendah).
4. Psikotes di kelompokkan dalam 2 katagori (Memuaskan, Tdk Memuaskan).

Pada tabel 1 disajikan data set sampel yang digunakan pada penelitian ini. Dataset yang digunakan sejumlah 54 record. Atribut data yang dipakai adalah Administrasi, Kesehatan, Akademik, dan Psikotes.

Tabel 1. Sampel data set yang dipakai

Sampel	Administrasi	Kesehatan	Akademik	Psikotes	Diterima
1	Bagus	Baik	Tinggi	Memuas	?
2	Bagus	Baik	Tinggi	Tdk	?
3	Bagus	Baik	Meneng	Memuas	?
4	Bagus	Baik	Meneng	Tdk	?
5	Bagus	Baik	Rendah	Memuas	?
6	Bagus	Baik	Rendah	Tdk	?
7	Bagus	Sedang	Tinggi	Memuas	?
8	Bagus	Sedang	Tinggi	Tdk	?
9	Bagus	Sedang	Meneng	Memuas	?
10	Bagus	Sedang	Meneng	Tdk	?
11	Bagus	Sedang	Rendah	Memuas	?
12	Bagus	Sedang	Rendah	Tdk	?
13	Bagus	Buruk	Tinggi	Memuas	?
14	Bagus	Buruk	Tinggi	Tdk	?
15	Bagus	Buruk	Meneng	Memuas	?
16	Bagus	Buruk	Meneng	Tdk	?
17	Bagus	Buruk	Rendah	Memuas	?
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
44	Kurang	Sedang	Tinggi	Tdk	?
45	Kurang	Sedang	Meneng	Memuas	?
46	Kurang	Sedang	Meneng	Tdk	?
47	Kurang	Sedang	Rendah	Memuas	?
48	Kurang	Sedang	Rendah	Tdk	?
49	Kurang	Buruk	Tinggi	Memuas	?
50	Kurang	Buruk	Tinggi	Tdk	?
51	Kurang	Buruk	Meneng	Memuas	?
52	Kurang	Buruk	Meneng	Tdk	?
53	Kurang	Buruk	Rendah	Memuas	?
54	Kurang	Buruk	Rendah	Tdk	?

Dari keempat kriteria penilaian dikelompokkan untuk menentukan tiap-tiap atributnya serta untuk menentukan batasan (*threshold*) nilai dari atribut tersebut[10].

- a. Kriteria Administrasi di kelompokkan dalam beberapa atribut yaitu pendidikan akhir minimal SMA sederajat, IPK minimal 2.50 atau nilai ijasah rata-rata 6.50, pengalaman kerja minimal 1 tahun dan batas umur 20 tahun s/d 28 tahun. Hasil matrikulasi pada kriteria Administrasi disajikan pada tabel 2.

Tabel 2. Atribut Kriteria Administrasi

No	Administrasi	Poin / Nilai			
		Bagus/3	Cukup/2	Kurang/1	
1	Pendidikan Akhir	S1	D3	SMA	
2	IPK / Nilai Ijasah	S1 & D3	>3.00	>2.75 s/d 3.00	2.50 s/d 2.75
		SMA	>8.50	7.00 s/d <8.50	6.50 s/d <7.00
3	Pengalaman kerja	> 3 th	> 2 th s/d 3 th	1 th s/d 2 th	
4	Umur	20 th s/d 23 h	>23 th s/d 26 th	>26 th s/d 28 th	

- b. Kriteria Kesehatan di kelompokkan dalam beberapa atribut yaitu tidak mengalami cacat permanen, tinggi badan minimal 160 cm dan berpenampilan menarik. Matrikulasi kriteria Kesehatan bisa dilihat pada tabel 3.

Tabel 3. Atribut Kriteria Kesehatan

No	Kesehatan	Poin / Nilai		
		Baik/ 3	Sedang/ 2	Buruk/ 1
1.	Tidak mengalami cacat permanen	tidak cacat permanen	berkaca mata	Alat bantu pendengaran
2.	Tinggi badan untuk nilai minimal 160 cm	> 174 cm	168 cm s/d 174 cm	160 s/d 167 cm
3.	Berpenampilan menarik	sangat menarik	menarik	cukup menarik

- c. Kriteria Akademik, di kelompokkan dalam beberapa atribut yaitu Bahasa Inggris, Komputer dan Pengetahuan Umum. Hasil matrikulasi kriteria Akademik disajikan pada tabel 4.

Tabel 4. Atribut Kriteria Akademik

No	Akademik	Poin / Nilai		
		Tinggi/ 3	Menengah/ 2	Rendah/ 1
1.	Bahasa Inggris	85 s/d 100	65 s/d 84	< 65
2.	Komputer	85 s/d 100	65 s/d 84	< 65
3.	Pengetahuan	85 s/d 100	65 s/d 84	< 65

- d. Kriteria Psikotes, di kelompokkan dalam beberapa atribut yaitu tes kepribadian dan tes tulis. Hasil matrikulasi kriteria Psikotes disajikan pada tabel 5.

Tabel 5. Atribut Kriteria Psikotes

No	Psikotes	Poin / Nilai	
		Memuaskan/ 3	Tidak Memuaskan/1
1.	Tes Kepribadian	60 s/d 100	<60
2.	Tes Tulis	60 s/d 100	<60

**B. Menghitung nilai Entropy dan Information Gain**

Untuk menghitung nilai Entropy dan Information Gain dalam menentukan the best classifier, ada beberapa parameter yang digunakan [11], yaitu :

1. S = ruang (data) sampel yang di gunakan untuk tranning.
2. p+ = jumlah yang bersolusi positif pada data sampel untuk kriteria tertentu.
3. p- = jumlah yang bersolusi negatif pada data sampel untuk kriteria tertentu.

Dari 54 data set sampel, tidak semua dipakai pada saat proses mining. Hanya beberapa sampel data saja yang kira-kira berguna dan sebarannya tidak terlalu acak. Karena

data yang terlalu acak akan membuat proses *mining* memakan waktu lama dan tingkat hubungannya pun rendah. Oleh karena itu perlu sebuah batasan (*threshold*) yang sudah dimanipulasi untuk proses *mining* ini.

1) Proses Perhitungan Pada Node 1

Pada perhitungan awal di node 1, diambil 10 dataset seperti yang disajikan pada tabel 6.

Sampel	Administrasi	Kesehatan	Akademik	Psikotes	Di terima
S3	Bagus	Baik	Menengah	Memuaskan	Ya
S9	Bagus	Sedang	Menengah	Memuaskan	Ya
S15	Bagus	Buruk	Menengah	Memuaskan	Tidak
S17	Bagus	Buruk	Rendah	Memuaskan	Tidak
S22	Cukup	Baik	Menengah	Tdk Memuaskan	Tidak
S25	Cukup	Sedang	Tinggi	Memuaskan	Ya
S26	Cukup	Sedang	Tinggi	Tdk Memuaskan	Tidak
S27	Cukup	Sedang	Menengah	Memuaskan	Tidak
S38	Kurang	Baik	Tinggi	Tdk Memuaskan	Tidak
S41	Kurang	Baik	Rendah	Memuaskan	Tidak

Pada data set tersebut, jumlah sampel kelas 1 (“Ya”) adalah 3 dan jumlah sampel untuk kelas 2 (“Tidak”) adalah 7. Jadi p1 = 3 dan p2 = 7. Dengan demikian *entropy* untuk kumpulan sampel data S adalah 10.

$$Entropy(S) = -p+ \log_2 p+ - p- \log_2 p- \\ = - (3/10) \log_2 (3/10) - (7/10) \log_2 (7/10) \\ = 0,881290899$$

Values (Administrasi) = Bagus, Cukup, Kurang.

Tabel 7. Value administrasi node 1

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	7	10
SBagus	2	2	4
SCukup	1	3	4
SKurang	0	2	2

Hitung *Entropy* (S) = SBagus, SCukup, SKurang dan IG untuk kriteria Administrasi :

$$Entropy(S) = - (3/10) \log_2 (3/10) - (7/10) \log_2 (7/10) \\ = 0,881290899$$

$$Entropy(SBagus) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) \\ = 1$$

$$Entropy(SCukup) = - (1/4) \log_2 (1/4) - (3/4) \log_2 (3/4) \\ = 0,811278124$$

$$Entropy(S_{Kurang}) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

$$IG(S, Administrasi) = Entropy(S) - \sum \in Values (Bagus, Cukup, Kurang) |Sv|/|S| Entropy(Sv) = 0,881290899 - (4/10) 1 - (4/10) 0,811278124 - (2/10) 0 = 0,156779649$$

Values (Kesehatan) = Baik, Sedang, Buruk

Tabel 8. Value kesehatan node 1

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	7	10
SBaik	1	3	4
SSedang	2	2	4
SBuruk	0	2	2

Hitung Entropy(S) = SBaik, SSedang, SBuruk dan IG untuk kriteria Kesehatan adalah:

$$Entropy(S) = - (3/10) \log_2 (3/10) - (7/10) \log_2 (7/10) = 0,881290899$$

$$Entropy(SBaik) = - (1/4) \log_2 (1/4) - (3/4) \log_2 (3/4) = 0,811278124$$

$$Entropy(SSedang) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 1$$

$$Entropy(SBuruk) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

$$IG(Kesehatan) = Entropy(S) - \sum \in Values (Baik, Sedang, Buruk) |Sv|/|S| Entropy(Sv) = 0,881290899 - (4/10) 0,811278124 - (4/10) 1 - (2/10) 0 = 0,156779649$$

Values (Akademik) = Tinggi, Menengah, Rendah

Tabel 9. Value akademik node 1

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	7	10
STinggi	1	2	3
SMenengah	2	3	5
SRendah	0	2	2

Hitung Entropy(S) = STinggi, SMenengah, SRendah dan IG untuk kriteria Akademik adalah:

$$Entropy(S) = - (3/10) \log_2 (3/10) - (7/10) \log_2 (7/10) = 0,881290899$$

$$Entropy(STinggi) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,918295834$$

$$Entropy(SMenengah) = - (2/5) \log_2 (2/5) - (3/5) \log_2 (3/5) = 0,970950594$$

$$Entropy(SRendah) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

$$IG(Akademik) = Entropy(S) - \sum \in Values (Tinggi, Menengah, Rendah) |Sv|/|S| Entropy(Sv) = 0,881290899 - (3/10) 0,918295834 - (5/10) 0,970950594 - (2/10) 0 = 0,128326852$$

Values (Psikotes) = Memuaskan, Tdk Memuaskan

Tabel 10. Value psikotes node 1

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	7	10
SMemuaskan	3	4	7
STdk Memuaskan	0	3	3

Hitung Entropy(S) = SMemuaskan, STdk Memuaskan dan IG untuk kriteria Psikotes adalah:

$$Entropy(S) = - (3/10) \log_2 (3/10) - (7/10) \log_2 (7/10) = 0,881290899$$

$$Entropy(SMemuaskan) = - (3/7) \log_2 (3/7) - (4/7) \log_2 (4/7) = 0,985228136$$

$$Entropy(STdk Memuaskan) = - (0/3) \log_2 (0/3) - (3/3) \log_2 (3/3) = 0$$

$$IG(Psikotes) = Entropy(S) - \sum \in Values (Memuaskan, Tdk Memuaskan) |Sv|/|S| Entropy(Sv) = 0,881290899 - (7/10) 0,985228136 - (3/10) 0 = 0,191631204$$

Setelah perhitungan nilai IG dari keempat atribut (Administrasi, Kesehatan, Akademik dan Psikotes) maka menghasilkan nilai sebagai berikut:

$$IG(S, Adiministrasi) = 0,156779649$$

$$IG(S, Kesehatan) = 0,156779649$$

$$IG(S, Akademik) = 0,128326852$$

$$IG(S, Psikotes) = 0,191631204$$

Jadi dari keempat nilai IG diatas Gain (S, Psikotes) adalah yang terbesar sehingga atribut Psikotes merupakan *the best classifier* dan harus diletakkan sebagai *root*. Setelah mendapat *the best classifier* langkah selanjutnya adalah setiap nilai pada atribut Psikotes akan di cek apakah perlu dibuat *subtree* di langkah/perhitungan berikutnya atau tidak. Atribut Psiko terdapat 2 sampel (Memuaskan, tdk memuaskan). Untuk nilai "S Tidak Memuaskan = [0+, 3-]" salah satu kelasnya ada yang kosong, berarti fungsi ini akan berhenti dan membentuk satu tunggal simpul *root*. Sedangkan untuk nilai "S Memuaskan = [3+, 4-]" salah satu kelasnya tidak ada yang kosong, sehingga perlu memanggil fungsi ID3 dan dibuat *subtree* dilangkah berikutnya[12].

## 2) Proses Perhitungan Pada Node 2

Pada tabel 11 disajikan data sampel Psikotes yang masih bisa dibentuk menjadi subtree.

Tabel 11. Data sampel node 2

Sampel	Administrasi	Kesehatan	Akademik	Psikotes	Diterima
S3	Bagus	Baik	Menengah	Memuaskan	Ya
S9	Bagus	Sedang	Menengah	Memuaskan	Ya
S15	Bagus	Buruk	Menengah	Memuaskan	Tidak

S17	Bagus	Buruk	Rendah	Memuaskan	Tidak
S25	Cukup	Sedang	Tinggi	Memuaskan	Ya
S27	Cukup	Sedang	Menengah	Memuaskan	Tidak
S41	Kurang	Baik	Rendah	Memuaskan	Tidak

Jumlah sampel kelas 1 (“ya”) adalah 3 dan jumlah sampel untuk kelas 2 (“tidak”) adalah 4. jadi  $p_1 = 3$  dan  $p_2 = 4$ . Dengan demikian *entropy* untuk kumpulan sampel data S adalah 7.

Values (Administrasi) = Bagus, Cukup, Kurang.

Tabel 12. Values administrasi node 2

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	4	7
SBagus	2	2	4
SCukup	1	1	2
SKurang	0	1	1

Hitung *Entropy* (S) = SBagus, SCukup, SKurang dan IG untuk nilai kriteria Administrasi adalah :

$$Entropy(S) = - (3/7) \log_2 (3/7) - (4/7) \log_2 (4/7) = 0,985228136$$

$$Entropy(SBagus) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 1$$

$$Entropy(SCukup) = - (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2) = 1$$

$$Entropy(SKurang) = - (0/0) \log_2 (0/0) - (0/1) \log_2 (0/1) = 0$$

$$IG(S, Administrasi) = Entropy(S) - \sum \in Values (Bagus, Cukup, Kurang) |Sv|/|S| Entropy(Sv) = 0,985228136 - (4/7) 1 - (2/7) 1 - (1/7) 0 = 0,128085279$$

Values (Kesehatan) = Baik, Sedang, Buruk

Tabel 13. Values kesehatan node 2

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	4	7
SBaik	1	1	2
SSedang	2	1	3
SBuruk	0	2	2

Hitung *Entropy*(S) = SBaik, SSedang, SBuruk dan IG untuk kriteria Kesehatan adalah:

$$Entropy(S) = - (3/7) \log_2 (3/7) - (4/7) \log_2 (4/7) = 0,985228136$$

$$Entropy(SBaik) = - (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2) = 1$$

$$Entropy(SSedang) = - (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) = 0,918295834$$

$$Entropy(SBuruk) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

$$IG(Kesehatan) = Entropy(S) - \sum \in Values (Baik, Sedang, Buruk) |Sv|/|S| Entropy(Sv) = 0,985228136 - (2/7) 1 - (3/7) 0,918295834 - (2/7) 0 = 0,305958493$$

Values (Akademik) = Tinggi, Menengah, Rendah

Tabel 14. Values akademik node 2

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	3	4	7
STinggi	1	0	1
SMenengah	2	2	4
SRendah	0	2	2

Hitung *Entropy*(S) = STinggi, SMenengah, SRendah dan IG untuk nilai kriteria Akademik adalah:

$$Entropy(S) = - (3/7) \log_2 (3/7) - (4/7) \log_2 (4/7) = 0,985228136$$

$$Entropy(STinggi) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0$$

$$Entropy(SMenengah) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 1$$

$$Entropy(SRendah) = - (0/2) \log_2 (0/2) - (2/2) \log_2 (2/2) = 0$$

$$IG(Akademik) = Entropy(S) - \sum \in Values (Tinggi, Menengah, Rendah) |Sv|/|S| Entropy(Sv) = 0,985228136 - (1/7) 0 - (4/7) 1 - (2/7) 0 = 0,413799565$$

Setelah perhitungan nilai IG dari ketiga atribut (Administrasi, Kesehatan. Fisik dan Psikotes) maka menghasilkan nilai sebagai berikut:

$$IG(S, Administrasi) = 0,128085279$$

$$IG(S, Kesehatan) = 0,305958493$$

$$IG(S, Akademik) = 0,413799565$$

Dari ketiga nilai IG diatas *Gain* (S. Akademik) adalah yang terbesar sehingga atribut Akademik merupakan *the best classifier*. Setelah mendapat *the best classifier* langkah selanjutnya adalah setiap nilai pada atribut Akademik akan di cek apakah perlu dibuat *subtree* di langkah/perhitungan berikutnya atau tidak. Atribut Akademik terdapat 3 sampel (tinggi, menengah dan rendah). Untuk nilai “S Tinggi = [1+, 0-]” dan Untuk nilai “SRendah = [0+, 2-]” salah satu kelasnya ada yang kosong, berarti fungsi ini akan berhenti dan membentuk satu tunggal simpul *root*. Sedangkan untuk nilai “S Menengah = [2+, 2]” salah satu kelasnya tidak ada yang kosong, berarti perlu memanggil fungsi ID3 dan dibuat *subtree* di level/perhitungan berikutnya dan harus diletakkan sebagai simpul dibawah simpul “psiko” pada cabang nilai “Memuaskan”.

### 3) Proses Perhitungan Pada Node 3

Pada tabel 15 disajikan data sampel Psikotes yang masih bisa dibentuk menjadi subtree.

Tabel 15. Data sampel node 3

Sampel	Administrasi	Kesehatan	Akademik	Psikotes	Diterima
S3	Bagus	Baik	Menengah	Memuaskan	Ya
S9	Bagus	Sedang	Menengah	Memuaskan	Ya
S15	Bagus	Buruk	Menengah	Memuaskan	Tidak
S27	Cukup	Sedang	Menengah	Memuaskan	Tidak

Jumlah sampel kelas 1 (“ya”) adalah 2 dan jumlah sampel untuk kelas 2 (“tidak”) adalah 2. jadi  $p_1 = 2$  dan  $p_2 = 2$ . Dengan demikian *entropy* untuk kumpulan sampel data S adalah 4, *Values* (Administrasi) = Bagus, Cukup, Kurang

Tabel 16. Values administrasi node 3

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	2	2	4
Sbagus	2	1	3
Scukup	0	1	1
Skurang	0	0	0

Hitung *Entropy* (S) = SBagus, SCukup, SKurang dan IG untuk kriteria Administrasi adalah :

$$Entropy(S) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 1$$

$$Entropy(S_{Bagus}) = - (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) = 0,918295834$$

$$Entropy(S_{Cukup}) = - (0/1) \log_2 (0/1) - (1/1) \log_2 (1/1) = 0$$

$$Entropy(S_{Kurang}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0$$

$$IG(S, Administrasi) = Entropy(S) - \sum \in Values (Bagus, Cukup, Kurang) |S_v|/|S| Entropy(S_v)$$

$$= 1 - (3/4) 0,918295834 - (1/4) 0 - (0/4) 0 = 0,311278124$$

*Values* (Kesehatan) = Baik, Sedang, Buruk.

Tabel 17. Values kesehatan node 3

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	2	2	4
Sbaik	1	0	1
SSedang	1	1	2
Sburuk	0	1	1

Hitung *Entropy*(S) = SBaik, SSedang, Sburuk dan IG untuk kriteria Kesehatan adalah:

$$Entropy(S) = - (2/4) \log_2 (2/4) - (2/4) \log_2 (2/4) = 1$$

$$Entropy(S_{Baik}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0$$

$$Entropy(S_{Sedang}) = - (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2) = 1$$

$$Entropy(S_{Buruk}) = - (0/1) \log_2 (0/1) - (1/1) \log_2 (1/1) = 0$$

$$IG(Kesehatan) = Entropy(S) - \sum \in Values (Baik, Sedang, Buruk) |S_v|/|S| Entropy(S_v)$$

$$= 1 - (1/4) 0 - (2/4) 1 - (1/4) 0 = 0,5$$

Setelah perhitungan nilai IG dari kedua atribut (Administrasi dan Kesehatan Fisik) maka menghasilkan nilai sebagai berikut:

$$IG(S, Adiministrasi) = 0,311278124$$

$$IG(S, Kesehatan) = 0,5$$

Dari kedua nilai IG diatas *Gain* (S, Kesehatan) adalah yang terbesar sehingga atribut Kesehatan Fisik merupakan *the best classifier*. Setelah mendapat *the best classifier* langkah selanjutnya adalah setiap nilai pada atribut Kesehatan Fisik akan di cek apakah perlu dibuat *subtree* di langkah/perhitungan berikutnya atau tidak. Atribut Kesehatan Fisik terdapat 3 sampel (baik, sedang dan buruk). Untuk nilai “SBaik = [1+, 0-] dan SBuruk = [0+, 1-]” salah satu kelasnya ada yang kosong, berarti fungsi ini akan berhenti dan membentuk satu tunggal simpul *root*. Sedangkan untuk nilai “SSedang = [1+, 1-]” salah satu kelasnya tidak ada yang kosong, berarti perlu memanggil fungsi ID3 dan dibuat *subtree* di langkah/perhitungan berikutnya dan harus diletakkan sebagai simpul dibawah simpul “Akademik” pada cabang nilai “Menengah”.

4) Proses Perhitungan Pada Node 4

Pada tabel 18 disajikan data set yang masih bisa dibentuk menjadi subtree node 4.

Tabel 18. Data set subtree node 4

Sampel	Administrasi	Kesehatan	Akademik	Psikotes	Diterima
S9	Bagus	Sedang	Menengah	Memuaskan	Ya
S27	Cukup	Sedang	Menengah	Memuaskan	Tidak

Jumlah sampel kelas 1 (“ya”) adalah 1 dan jumlah sampel untuk kelas 2 (“tidak”) adalah 1. jadi  $p_1 = 1$  dan  $p_2 = 1$ . Dengan demikian *entropy* untuk kumpulan sampel data S adalah 2. *Values* (Administrasi) = Bagus, Cukup, Kurang.

Tabel 19. Values administrasi node 4

Entropy(S)	Ya(P+)	Tidak(P-)	Total Sampel
S	1	1	2
Sbagus	1	0	1
Scukup	0	1	1
Skurang	0	0	0

Hitung *Entropy* (S) = SBagus, SCukup, SKurang dan IG untuk kriteria Administrasi adalah :

$$Entropy(S) = - (1/2) \log_2 (1/2) - (1/2) \log_2 (1/2) = 1$$

$$Entropy(S_{Bagus}) = - (1/1) \log_2 (1/1) - (0/1) \log_2 (0/1) = 0$$

$$Entropy(S_{Cukup}) = - (0/1) \log_2 (0/1) - (1/1) \log_2 (1/1) = 0$$

$$Entropy(S_{Kurang}) = - (0/0) \log_2 (0/0) - (0/0) \log_2 (0/0) = 0$$

$$IG(S, Administrasi) = Entropy(S) - \sum \in Values (Bagus, Cukup, Kurang) |S_v|/|S| Entropy(S_v)$$

$$= 1 - (1/2) 0 - (1/2) 0 - (0/2) 0 = 1$$

Dari perhitungan diatas didapat nilai IG dari kesatu atribut (Administrasi) IG (S, Adiministrasi) = 1.

Jadi dari kesatu nilai IG diatas *Gain* (S, Administrasi) adalah yang terbesar sehingga atribut Administrasi merupakan *the best classifier* dan harus diletakkan sebagai simpul dibawah simpul "Kesehatan Fisik" pada cabang nilai "Sedang".Setelah mendapat *the best classifier* langkah selanjutnya adalah setiap nilai pada atribut Administrasi akan di cek apakah perlu dibuat *subtree* di langkah/perhitungan berikutnya atau tidak.Atribut Administrasi terdapat 3 sampel (bagus, cukup dan kurang).Untuk nilai "SBagus = [1+, 0-], SCukup= [0+, 1-] dan SKurang = [0+ , 0-]"salah satu kelasnya ada yang kosong, berarti fungsi ini akan berhenti dan membentuk satu tunggal simpul *root*[9].

**C. Melakukan Konstruksi Pohon Keputusan**

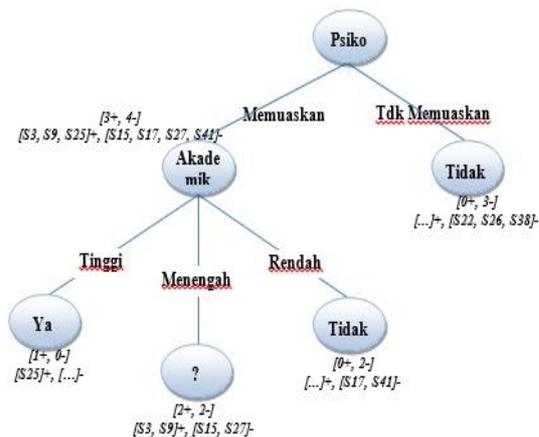
Setelah menemukan nilai Entropy dan Information Gain, maka langkah selanjutnya adalah membentuk pohon keputusan[13].

1. Pohon Keputusan pada node 1  
 Pada perhitungan di node 1,nilai Information Gain terbesar ada pada atribut Psikotes sehingga Psikotes diletakkan sebagai root. Hasil pohon keputusan pada node 1 atau root bisa dilihat pada gambar 2.



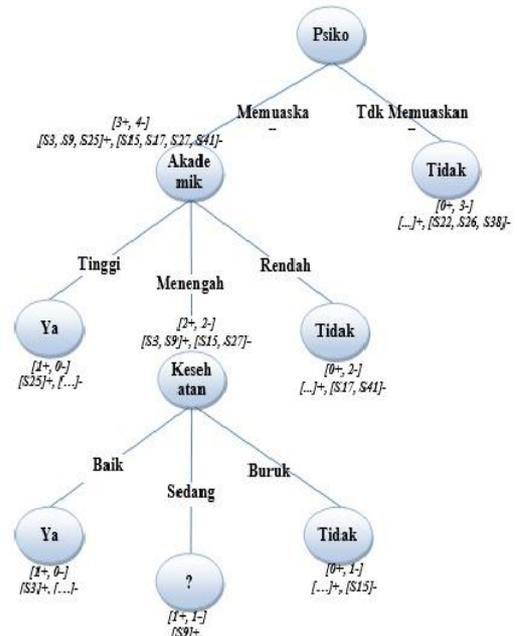
Gambar 2. Pohon Keputusan root

2. Pohon Keputusan pada node 2  
 Pada node 2,hasil perhitungan Information Gain terbesar pada atribut Akademik. Sehingga Pohon yang bisa dibentuk seperti pada gambar 3.



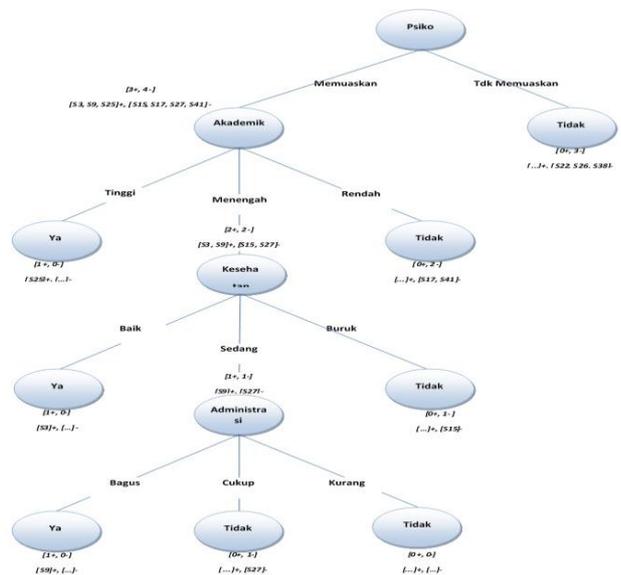
Gambar 3. Pohon Keputusan node 2

3. Pohon Keputusan pada node 3  
 Pada node 3, hasil perhitungan Information Gain terbesar pada atribut Kesehatan. Sehingga Pohon Keputusan yang dapat dibentuk seperti pada gambar 4.



Gambar 4. Pohon Keputusan Node 3

4. Pohon Keputusan pada node 4  
 Pada node 4,hasil perhitungan Information Gain terdapat pada atribut Administrasi, sehingga hasil pohon keputusan bisa dilihat pada gambar 5.



Gambar 5. Pohon Keputusan Node 4

**D. Melakukan Operasi Konjungsi dan Penentuan Peringkat**

Setelah ditemukan jalur klasifikasi terhadap seluruh atribut yang digambarkan pada pohon keputusan, maka tahapan selanjutnya adalah melakukan operasi konjungsi pada setiap simpul sampai ditemukan *leaf node*[14]. Hasil

konjungsi pada 54 sampel data set, sebagiannya adalah sebagai berikut :

- S1 IF Psikotes = “Memuaskan” ^ Akademik = “Tinggi” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Ya
  - S2 IF Psikotes = “Tdk Memuaskan” ^ Akademik = “Tinggi” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Tidak
  - S3 IF Psikotes = “Memuaskan” ^ Akademik = “Menengah” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Ya
  - S4 IF Psikotes = “Tdk Memuaskan” ^ Akademik = “Menengah” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Tidak
  - S5 IF Psikotes = “Memuaskan” ^ Akademik = “Rendah” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Tidak
  - S6 IF Psikotes = “Tdk Memuaskan” ^ Akademik = “Rendah” ^ Kes. Fisik = “Baik” ^ Administrasi = “Bagus” ^ THEN Diterima = Tidak
- 
- S54 IF Psikotes = “Tdk Memuaskan” ^ Akademik = “Rendah” ^ Kes. Fisik = “Buruk” ^ Administrasi = “Kurang” ^ THEN Diterima = Tidak

Setelah melakukan *Operasi Conjunction* ( ^ ) pada setiap simpul sampai ditemukan *leaf node*, maka langkah selanjutnya melakukan perhitungan peringkat. Tujuannya agar setiap keputusan “ya” atau “tidak” ditemukan poin dari yang tertinggi sampai ke poin yang terendah[15]. Perhitungan peringkat yang digunakan seperti yang disajikan pada persamaan 1.

$$P(\text{Hasil Perhitungan}) = ((P_{\text{Administrasi}} \cdot L)/J) + ((P_{\text{KesFisik}} \cdot L)/J) + ((P_{\text{Akademik}} \cdot L)/J) + ((P_{\text{Psiko}} \cdot L)/J) \quad (1)$$

Dimana :

$P_{\text{Administrasi}}$  = Jumlah poin pada atribut administrasi.

$P_{\text{Kes. Fisik}}$  = Jumlah poin pada atribut kesehatan fisik

$P_{\text{Akademik}}$  = Jumlah poin pada atribut akademik

$P_{\text{Psiko}}$  = Jumlah poin pada atribut psiko

$L$  = Jumlah tingkatan level pada masing-masing atribut

$J$  = Jumlah katagori pada masing-masing atribut

$P$  = Hasil perhitungan

Dari keseluruhan proses, maka hasil pemeringkatan bisa dilihat pada table 20.

Table 20. Hasil akhir pemeringkatan

Sampel	Kriteria dan Jumlah Katagori/kriteria				Poin Perkatagori				Di Terima	Peringkat ketulusan
	Admi.	Kesh.	Akad.	Psik.						
	3	3	3	2						
S1	Bagus	Baik	Tinggi	Memuaskan	3	3	3	3	Ya	53,5
S19	Cukup	Baik	Tinggi	Memuaskan	2	3	3	3	Ya	53,2
S37	Kurang	Baik	Tinggi	Memuaskan	1	3	3	3	Ya	52,8
S7	Bagus	Sedang	Tinggi	Memuaskan	3	2	3	3	Ya	52,5
S25	Cukup	Sedang	Tinggi	Memuaskan	2	2	3	3	Ya	52,2
S43	Kurang	Sedang	Tinggi	Memuaskan	1	2	3	3	Ya	51,8
S13	Bagus	Buruk	Tinggi	Memuaskan	3	1	3	3	Ya	51,5
S31	Cukup	Buruk	Tinggi	Memuaskan	2	1	3	3	Ya	51,2
S49	Kurang	Buruk	Tinggi	Memuaskan	1	1	3	3	Ya	50,8
S3	Bagus	Baik	Menengah	Memuaskan	3	3	2	3	Ya	50,5
S21	Cukup	Baik	Menengah	Memuaskan	2	3	2	3	Ya	50,2
S39	Kurang	Baik	Menengah	Memuaskan	1	3	2	3	Ya	49,8
S9	Bagus	Sedang	Menengah	Memuaskan	3	2	2	3	Ya	49,5
S27	Cukup	Sedang	Menengah	Memuaskan	2	2	2	3	Tidak	49,2
S45	Kurang	Sedang	Menengah	Memuaskan	1	2	2	3	Tidak	48,8
S15	Bagus	Buruk	Menengah	Memuaskan	3	1	2	3	Tidak	48,5
S33	Cukup	Buruk	Menengah	Memuaskan	2	1	2	3	Tidak	48,2
S51	Kurang	Buruk	Menengah	Memuaskan	1	1	2	3	Tidak	47,8
S5	Bagus	Baik	Rendah	Memuaskan	3	3	1	3	Tidak	47,5
S23	Cukup	Baik	Rendah	Memuaskan	2	3	1	3	Tidak	47,2
S41	Kurang	Baik	Rendah	Memuaskan	1	3	1	3	Tidak	46,8
S11	Bagus	Sedang	Rendah	Memuaskan	3	2	1	3	Tidak	46,5
S29	Cukup	Sedang	Rendah	Memuaskan	2	2	1	3	Tidak	46,2
S47	Kurang	Sedang	Rendah	Memuaskan	1	2	1	3	Tidak	45,8
S17	Bagus	Buruk	Rendah	Memuaskan	3	1	1	3	Tidak	45,5
S35	Cukup	Buruk	Rendah	Memuaskan	2	1	1	3	Tidak	45,2
S53	Kurang	Buruk	Rendah	Memuaskan	1	1	1	3	Tidak	44,8
S2	Bagus	Baik	Tinggi	Tdk Memuaskn	3	3	3	1	Tidak	26,5
S20	Cukup	Baik	Tinggi	Tdk Memuaskn	2	3	3	1	Tidak	26,2
S38	Kurang	Baik	Tinggi	Tdk Memuaskn	1	3	3	1	Tidak	25,8
S8	Bagus	Sedang	Tinggi	Tdk Memuaskn	3	2	3	1	Tidak	25,5
S26	Cukup	Sedang	Tinggi	Tdk Memuaskn	2	2	3	1	Tidak	25,2
S44	Kurang	Sedang	Tinggi	Tdk Memuaskn	1	2	3	1	Tidak	24,8
S14	Bagus	Buruk	Tinggi	Tdk Memuaskn	3	1	3	1	Tidak	24,5
S32	Cukup	Buruk	Tinggi	Tdk Memuaskn	2	1	3	1	Tidak	24,2
S50	Kurang	Buruk	Tinggi	Tdk Memuaskn	1	1	3	1	Tidak	23,8

S4	Bagus	Baik	Menengah	Tdk Memuaskan	3	3	2	1	Tidak	23,5
S22	Cukup	Baik	Menengah	Tdk Memuaskan	2	3	2	1	Tidak	23,2
S40	Kurang	Baik	Menengah	Tdk Memuaskan	1	3	2	1	Tidak	22,8
S10	Bagus	Sedang	Menengah	Tdk Memuaskan	3	2	2	1	Tidak	22,5
S28	Cukup	Sedang	Menengah	Tdk Memuaskan	2	2	2	1	Tidak	22,2
S46	Kurang	Sedang	Menengah	Tdk Memuaskan	1	2	2	1	Tidak	21,8
S16	Bagus	Buruk	Menengah	Tdk Memuaskan	3	1	2	1	Tidak	21,5
S34	Cukup	Buruk	Menengah	Tdk Memuaskan	2	1	2	1	Tidak	21,2
S52	Kurang	Buruk	Menengah	Tdk Memuaskan	1	1	2	1	Tidak	20,8
S6	Bagus	Baik	Rendah	Tdk Memuaskan	3	3	1	1	Tidak	20,5
S24	Cukup	Baik	Rendah	Tdk Memuaskan	2	3	1	1	Tidak	20,2
S42	Kurang	Baik	Rendah	Tdk Memuaskan	1	3	1	1	Tidak	19,8
S12	Bagus	Sedang	Rendah	Tdk Memuaskan	3	2	1	1	Tidak	19,5
S30	Cukup	Sedang	Rendah	Tdk Memuaskan	2	2	1	1	Tidak	19,2
S48	Kurang	Sedang	Rendah	Tdk Memuaskan	1	2	1	1	Tidak	18,8
S18	Bagus	Buruk	Rendah	Tdk Memuaskan	3	1	1	1	Tidak	18,5
S36	Cukup	Buruk	Rendah	Tdk Memuaskan	2	1	1	1	Tidak	18,2
S54	Kurang	Buruk	Rendah	Tdk Memuaskan	1	1	1	1	Tidak	17,8

Dari hasil pemeringkatan dapat diketahui perankingan dari seluruh dataset uji. Jadi dari tabel tersebut sudah bisa memberikan gambaran, calon karyawan yang dipertimbangkan lolos seleksi dan tidak lolos seleksi.

## KESIMPULAN

Dari hasil implementasi algoritma ID3 didapat pohon keputusan yang memberi pengetahuan pengklasifikasi terbaik seleksi karyawan berdasarkan kriterianya. Hasil klasifikasi yang diperoleh dapat dijadikan pertimbangan dalam pengambilan keputusan karena sesuai dengan kebutuhan perusahaan.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada pihak CV Aliyah Mandiri Pamekasan karena sudi memberikan data dan keperluan lainnya hingga selesainya penelitian ini. Penulis juga ucapkan terimakasih kepada Universitas Madura serta pihak yang tidak bias kami sebutkan satu persatu atas bantuannya sehingga penelitian ini selesai.

## DAFTAR PUSTAKA

- [1] N. Ramadhani and N. Fajarianto, "Sistem Informasi Evaluasi Perkuliahan dengan Sentimen Analisis Menggunakan Naïve Bayes dan Smoothing Laplace," vol. 02, pp. 228–234, 2020, doi: <https://doi.org/10.21456/vol10iss2pp228-234>.
- [2] Jianpei Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufman Publisher, 2012.
- [3] H. A. Efendi, Muhamad Subhan Wibawa, "Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik ( Diabetes Prediction using ID3 Algorithm with Best Attribute Selection )," *Juita*, vol. VI, pp. 29–35, 2018, doi: 10.30595/juita.v6i1.2412.
- [4] S. T. Ahmed, R. Al-Hamdani, and M. S. Croock, "Developed third iterative dichotomizer based on feature decisive values for educational data mining," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, pp. 209–217, 2019, doi: 10.11591/ijeecs.v18.i1.pp209-217.
- [5] R. Sovia, A. Muhammad, S. Arlis, and S. Defit, "Analysis of sales levels of pharmaceutical products by using data mining algorithm C45," vol. 22, no. 1, pp. 476–484, 2021, doi: 10.11591/ijeecs.v22.i1.pp476-484.
- [6] T. E. Dagogo-george, H. A. Mojeed, and A. O. Balogun, "Tree-based homogeneous ensemble model with feature selection for diabetic retinopathy prediction," vol. 8, no. October, pp. 297–303, 2020, doi: 10.14710/jtsiskom.2020.13669.
- [7] T. Arifin and A. Herliana, "Optimizing decision tree using particle swarm optimization to identify eye diseases based on texture analysis," *J. Teknol.*

- dan Sist. Komput.*, vol. 8, no. 1, pp. 59–63, 2020, doi: 10.14710/jtsiskom.8.1.2020.59-63.
- [8] I. P. P. Iswara, F. Farhan, W. Kumara, and A. A. Supianto, “Rekomendasi Pengambilan Mata Kuliah Pilihan Untuk Mahasiswa Sistem Informasi Menggunakan Algoritme Decision Tree,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 3, pp. 341–348, 2019, doi: <http://dx.doi.org/10.25126/jtiik.201963892>.
- [9] D. R. Nurajijah, “Algoritma Naïve Bayes , Decision Tree , dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah Naïve Bayes,” vol. 7, no. January, pp. 77–82, 2019, doi: 10.14710/jtsiskom.7.2.2019.77-82.
- [10] I. Choina, R. Aulia, and A. Zakir, “Penerapan Algoritma ID3 Untuk Menyeleksi Pegawai Kontrak Di Kantor Pengadilan Kota Langsa,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 1, p. 47, 2020, doi: <https://doi.org/10.24114/cess.v5i1.15146>.
- [11] J. Adiana, E. Sedyono, and O. Dwi, “Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 Untuk Sistem Informasi Pengukuran Kinerja PNS,” vol. 02, pp. 75–86, 2014, doi: <https://doi.org/10.21456/vol4iss2pp75-86>.
- [12] Tukino, “Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia,” vol. 01, pp. 39–46, 2019, doi: <https://doi.org/10.21456/vol9iss1pp39-46>.
- [13] T. D. Utama, S. W. Sihwi, and A. Doewes, “Implementasi Algoritma Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha UNS,” *J. Itsmart*, vol. 3, no. 2, pp. 74–82, 2014, doi: <https://doi.org/10.20961/itsmart.v3i2.700>.
- [14] A. Nurkholis, I. S. Sitanggang, and I. S. Sitanggang, “Optimalisasi model prediksi kesesuaian lahan kelapa sawit menggunakan algoritme pohon keputusan spasial Optimization for prediction model of palm oil land suitability using spatial,” vol. 8, no. July, pp. 192–200, 2020, doi: 10.14710/jtsiskom.2020.13657.
- [15] T. E. Dagogo-George, H. A. Mojeed, A. O. Balogun, M. A. Mabayoje, and S. A. Salihu, “Tree-based homogeneous ensemble model with feature selection for diabetic retinopathy prediction,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 4, pp. 297–303, 2020, doi: 10.14710/jtsiskom.2020.13669.