

Sistem Pendukung Keputusan Perpanjangan Kontrak Karyawan dengan Algoritma Decision tree

Muhammad Yan Handoko Putra F*, Lili Tanti

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

Email: ¹*myanhandoko@gmail.com, ²lili@potensi-utama.ac.id

Email Penulis Korespondensi: ¹*myanhandoko@gmail.com*

Submitted: 20/10/2025; Accepted: 05/11/2025; Published: 31/12/2025

Abstrak–Proses perpanjangan kontrak kerja di perusahaan sering kali menimbulkan kompleksitas karena melibatkan kesejahteraan karyawan serta kelangsungan produktivitas organisasi. PT. Duta Agung Jaya, sebagai perusahaan *outsourcing* di Kota Medan, masih mengalami hambatan dalam menentukan karyawan yang memenuhi syarat untuk perpanjangan kontrak. Hingga saat ini, proses penentuan dilakukan secara manual atau semi-komputerisasi menggunakan Microsoft Excel, yang berpotensi menimbulkan kesalahan perhitungan, kecenderungan subjektivitas, serta kurangnya pertimbangan komprehensif terhadap aspek-aspek kinerja. Variabel independen yang diterapkan mencakup absensi, pendidikan, kuantitas kerja, disiplin, masa kerja, prestasi, komunikasi, dan tanggung jawab, sedangkan variabel dependen adalah kelayakan perpanjangan kontrak karyawan (layak/tidak layak). Data penelitian terdiri dari 1000 *dataset* karyawan kontrak, yang kemudian diproses dan diuji untuk menghasilkan model keputusan yang lebih objektif. Dengan pendekatan ini, penelitian diharapkan dapat menyediakan sistem yang akurat dalam memberikan rekomendasi keputusan, mengurangi kesalahan akibat subjektivitas, serta memfasilitasi manajemen dalam mengevaluasi kinerja karyawan secara transparan dan terukur. Secara keseluruhan, hasil penelitian ini diharapkan memberikan kontribusi signifikan dalam meningkatkan kualitas pengambilan keputusan di PT. Duta Agung Jaya, serta menjadi referensi bagi penelitian serupa di bidang penerapan Sistem Pendukung Keputusan dengan algoritma ID3 pada kasus manajemen sumber daya manusia. Algoritma ID3 atau *Iterative Dichotomiser 3* merupakan salah satu algoritma yang dapat digunakan untuk membangun *decision tree* dengan mencari semua kemungkinan dalam pohon keputusan melalui struktur hierarki untuk pembelajaran terawasi.

Kata Kunci: Sistem Pendukung Keputusan; Perpanjangan Kontrak Karyawan; Algoritma ID3; Decision Tree; Evaluasi Kinerja Karyawan

Abstract– The process of extending employment contracts in companies is often a complex issue because it concerns employee welfare as well as the sustainability of company productivity. PT. Duta Agung Jaya as an outsourcing company in Medan City still faces obstacles in determining employees who are eligible for contract extension. Currently, the determination process is still carried out manually or semi-computerized using Microsoft Excel, thus creating the risk of calculation errors, subjective tendencies, and a lack of comprehensive consideration of performance factors. This algorithm was chosen because it is able to process data hierarchically and produce clear decision rules based on certain attributes. The independent variables used include absence, education, work quantity, discipline, length of service, achievement, communication, and responsibility, while the dependent variable is the eligibility of employee contract extensions (eligible/uneligible). The research data consists of 1000 contract employee datasets, which are then processed and tested to produce a more objective decision model. With this approach, the research is expected to provide an accurate system in providing decision recommendations, reducing errors due to subjectivity, and making it easier for management to evaluate employee performance in a transparent and measurable manner. Overall, the results of this research are expected to provide a real contribution in improving the quality of decision making at PT. Duta Agung Jaya, as well as become a reference for similar research in the field of DSS application with the ID3 algorithm in the case of human resource management. The ID3 algorithm or *Iterative Dichotomiser 3* is one algorithm that can be used in creating a decision tree by searching for all possibilities in the decision tree using a hierarchical structure for supervised learning.

Keywords: *Decision Support System*; Employee Contract Renewal; ID3 Algorithm; Decision Tree; Employee Performance Evaluation

1. PENDAHULUAN

Perpanjangan tenaga kerja kontrak dilakukan dalam periode tertentu, dan hanya tenaga kerja yang memiliki potensi atau prestasi yang berhak mendapatkan kesempatan untuk kontrak berikutnya. Sehingga perusahaan juga memperoleh sumber daya manusia yang optimal. Dalam istilah hukum, pekerja kontrak sering disebut “pekerja PKWT”, yang berarti pekerja dengan Perjanjian Kerja Waktu Tertentu [1]. PT. Duta Agung Jaya adalah salah satu perusahaan *outsourcing* yang berlokasi di Jl. Rawe Raya Pasar VI No.28-29, Tangkahan, Kec. Medan Labuhan, Kota Medan, Sumatera Utara 20251. PT. Duta Agung Jaya memiliki jumlah karyawan yang cukup besar, baik karyawan tetap maupun karyawan kontrak. Dalam operasionalnya, perusahaan secara berkala melakukan evaluasi terhadap karyawan kontrak untuk menentukan apakah kontrak kerja akan diperpanjang atau dihentikan. Keputusan perpanjangan kontrak karyawan ini merupakan bagian penting dari manajemen sumber daya manusia, karena secara langsung memengaruhi stabilitas tenaga kerja, efisiensi operasional, serta pencapaian target perusahaan.

Namun, proses pengambilan keputusan untuk memperpanjang kontrak karyawan di PT. Duta Agung Jaya masih dilakukan secara manual dan subjektif. Biasanya, keputusan tersebut hanya didasarkan pada penilaian pribadi atasan langsung tanpa menggunakan pendekatan sistematis atau data historis yang memadai. Hal ini

menimbulkan berbagai permasalahan, seperti potensi ketidaktepatan dalam evaluasi kinerja, ketidakkonsistenan dalam kriteria penilaian, dan kemungkinan terjadinya konflik internal karena dianggap tidak transparan atau tidak adil. Dari permasalahan tersebut, maka dilakukan penelitian untuk mendapatkan solusi dalam menentukan perpanjangan kontrak karyawan dengan menggunakan variabel absensi karyawan, prestasi karyawan, hasil kerja karyawan, loyalitas karyawan, dan kerja sama tim karyawan, yang dapat diselesaikan dengan ilmu pengambilan keputusan yaitu Sistem Pendukung Keputusan (SPK). Sistem Pendukung Keputusan (*Decision Support System*) adalah sistem berbasis komputer yang interaktif dalam membantu pengambilan keputusan dengan memanfaatkan data dan model untuk menyelesaikan masalah-masalah yang tak terstruktur [2],[3].

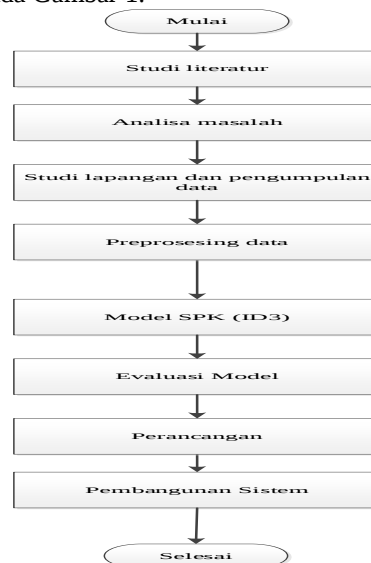
Untuk menyelesaikan masalah tersebut, penelitian ini mengusulkan penerapan sistem pendukung keputusan atau *Decision Support System* untuk membantu manajemen menghasilkan rekomendasi perpanjangan kontrak secara lebih objektif dan terukur. DSS didefinisikan sebagai sistem berbasis komputer yang interaktif untuk membantu pengambil keputusan memanfaatkan data dan model dalam menyelesaikan masalah yang tidak terstruktur.

Dalam penelitian ini, merancang sistem pendukung keputusan dalam penentuan perpanjangan kontrak karyawan dengan menggunakan algoritma *Decision tree*, yaitu *Iterative Dichotomiser 3* (ID3) [4]. Algoritma ID3 atau *Iterative Dichotomiser 3* merupakan salah satu algoritma yang dapat digunakan dalam membangun *decision tree* dengan mencari semua kemungkinan dalam pohon keputusan melalui struktur hierarki untuk pembelajaran terawasi [5],[6]. Sejumlah penelitian terdahulu menunjukkan bahwa pendekatan berbasis data dapat meningkatkan konsistensi penilaian dan ketepatan keputusan. Misalnya, penelitian I Kurniawati, Muhidin, dan Putra menerapkan teknik klasifikasi untuk kasus perpanjangan kontrak karyawan dan menunjukkan hasil evaluasi model yang tinggi (akurasi, presisi, dan recall) pada data yang diteliti [7]. pada konteks yang berbeda, penelitian yang membahas tentang perpanjangan kontrak karyawan, yaitu penelitian dengan variabel independen disiplin, absensi, tanggung jawab, kebersihan, sopan santun, umur, kerja sama tim, inisiatif kerja, dan variabel dependen perpanjangan kontrak (ya/tidak) [8]. Berdasarkan temuan dan kecenderungan pada penelitian-penelitian terdahulu, penelitian ini memprioritaskan penggunaan algoritma ID3 karena kemampuan interpretabilitasnya yang tinggi dalam menghasilkan aturan keputusan yang eksplisit. Dengan karakteristik tersebut, perusahaan dapat menelusuri serta menjelaskan dasar rekomendasi perpanjangan kontrak secara lebih transparan, konsisten, dan mudah dipertanggung jawabkan.

2. METODOLOGI PENELITIAN

2.1 Rancangan Penelitian

Dalam tahap rancangan penelitian ini, penulis memakai pemodelan diagram prosedur perancangan, model prosedur perancangan dapat dilihat pada Gambar 1.



Gambar 1. Prosedur Perancangan

Pengembangan sistem menggunakan prosedur perancangan dapat dijelaskan sebagai berikut :

- Studi Literatur
Digunakan untuk mempelajari serta menambah pemahaman peneliti tentang Sistem Pendukung Keputusan (SPK) secara umum dan metode *Decision tree* ID3 secara khusus serta membaca jurnal-jurnal ataupun referensi lainnya yang terkait dengan penelitian.
- Analisa Masalah

- Analisa masalah digunakan untuk memecahkan suatu masalah serta menganalisa data dalam melakukan suatu kajian sebelum melakukan perancangan ataupun perhitungan.
- c. Studi lapangan dan Pengumpulan Data
Dalam penelitian ini melakukan suatu observasi yang digunakan untuk lebih memahami bagaimana prosedur dalam penentuan perpanjang kontrak karyawan.
 - d. Preprocessing
Tahap preprocessing ini dilakukan dengan beberapa langkah, seperti mengisi data yang hilang/kosong, mengurangi data noise yang mengandung kesalahan atau data yang outlier, dan menangani data yang tidak konsisten
 - e. Model SPK ID3
Tahap awal dalam penelitian ini yaitu dimulai dengan menganalisa permasalahan yang terjadi dalam pemilihan. Dimulai dari menganalisa perhitungan sistem lama yang masih manual, setelah itu dilanjutkan dengan menganalisa dengan menggunakan perhitungan baru dengan metode yang digunakan yaitu *Decision tree* ID3 [9].
 - f. Evaluasi Model
Berisi langkah-langkah yang dilakukan saat pengujian peralatan secara keseluruhan, besaran-besaran yang akan diuji, dan ukuran untuk menilai apakah alat sudah bekerja dengan baik sesuai spesifikasi dengan menggunakan evaluasi model Confusion matrix. Confusion matrix adalah teknik yang dipakai dalam menilai hasil kerja dari model yang diuji terutama dalam kasus klasifikasi pada machine learning
 - g. Perancangan
Desain sistem yang digunakan dalam teori adalah pemodelan UML yang terdiri dari *use case diagram*, *class diagram*, *activity diagram* dan *sequence diagram*.
 - h. Pembangunan Sistem
Pada tahapan ini peneliti menggunakan PHP dan pengetikan coding program menggunakan java Script. Peneliti menggunakan *hardware* komputer/laptop. *Database* yang digunakan adalah MySQL.

2.2 Decision tree Learning

Decision tree merupakan salah satu teknik machine learning yang paling banyak diteliti dan digunakan secara luas dalam praktik karena struktur modelnya yang berbasis aturan sehingga relatif mudah dipahami dan dijelaskan [10]. Namun, literatur juga menegaskan bahwa decision tree dapat rentan terhadap overfitting dan berpotensi tidak stabil ketika berhadapan dengan dataset yang mengandung noise, sehingga berbagai pengembangan (misalnya pendekatan yang lebih robust atau teknik perbaikan/pengendalian kompleksitas model) kerap diperlukan [11]. Dari sisi representasi pengetahuan, struktur pohon keputusan dapat dinyatakan dalam bentuk representasi logika seperti disjunctive normal form (DNF), yaitu disjungsi (OR) dari beberapa konjungsi kondisi, yang selaras dengan karakter aturan keputusan if-then pada pohon keputusan [12]. Dalam keluarga algoritma pembentukan pohon keputusan, ID3 dikenal sebagai salah satu algoritma paling awal dan populer; ID3 membangun pohon secara rekursif dengan memilih atribut berdasarkan information gain (berbasis entropi) hingga terbentuk aturan klasifikasi [13]. *Decision tree* struktur flowchart yang mempunyai *tree* (pohon), dimana setiap simpul internal menandakan suatu atribut, setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. alur pada *decision tree* ditelusuri dari simpul ke akar ke simpul daun yang memegang prediksi kelas untuk contoh tersebut. *decision tree* mudah untuk dikonversi ke aturan klasifikasi (classification rule). Konsep data dalam *decision tree* dinyatakan dalam bentuk tabel dengan atribut dan record [14].

2.3 ID3

Iterative Dichotomiser 3 (ID3) merupakan metode dalam pembelajaran yang membangun pohon keputusan untuk menemukan solusi dari suatu masalah. Pohon keputusan yang dihasilkan melalui proses pencarian nilai terbaik (the best classifier) akan dijadikan sebagai akar (root) [15]. Dalam penelitian ini, akan dibahas model klasifikasi menggunakan *Decision tree* dengan algoritma Interactive Dichotomiser 3 (ID3) berdasarkan kasus yang relevan. Adapun langkah-langkah prosedural dalam metode Interactive Dichotomiser 3 (ID3) [16] adalah sebagai berikut:

1. Entropy(S) = $-p^+ \log_2 p^+ - p^- \log_2 p^-$ (1)
Dimana:
S : ruang (data) sample yang digunakan untuk training.
p⁺ : jumlah yang bersolusi positif (mendukung) pada data sample untuk kriteria tertentu.
p⁻ : jumlah yang bersolusi negatif (tidak mendukung) pada data sample untuk kriteria tertentu
2. Untuk suatu kumpulan data, maka dapat diukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektivitas ini disebut information gain. Secara matematis, information gain dari suatu atribut A adalah [17] :

$$\text{Information Gain} = \sum_{\text{value}(A)} \frac{|S_v|}{|S|} \log_2 \frac{|S|}{|S_v|} \quad (2)$$

$$\text{Gain}(S, A) = \text{Entropy}(S) - \text{Entropy}(S_v)$$

Dimana:

A : atribut

V : suatu nilai yang mungkin untuk atribut A

Values (A) : himp yang mungkin untuk atribut A

|Sv| : jum sampel untuk nilai v

|S| : jum seluruh sampel data

Entropy(Sv) : entropy untuk sampel-sampel yang memiliki nilai v [18]

3. HASIL DAN PEMBAHASAN

Metode Pengumpulan data yang dilakukan pada penelitian ini berupa wawancara dengan HRD dari PT. Duta Agung Jaya, kemudian melakukan pengumpulan data berdasarkan data dari PT. Duta Agung Jaya. Data yang digunakan pada penelitian ini sebanyak 1095 dengan memiliki Absensi, Pendidikan, Kuantitas kerja, disiplin kerja, masa kerja, Prestasi, komunikasi dan tanggung jawab dan variabel dependen Perpanjangan Kontrak Karyawan (Layak/Tidak Layak).

Tabel 1.Tabel Variabel

No	Nama Variabel	Keterangan
1	Nomor	Nomor merukapan urutan utk data karyawan
2	Nama Karyawan	Nama karyawan adalah nama identitas dari data karyawan kontrak
3	Alamat	Alamat karyawan adalah alamat karyawan atau tempat tinggal dari data karyawan kontrak
4	Tanggal Lahir	Tanggal lahir karyawan adalah Tanggal lahir karyawan kelahiran karyawan kontrak
5	No Telp	No telepon karyawan adalah informasi kontak dari karyawan
6	Absensi	Absensi adalah pencatatan kehadiran seseorang dalam suatu kegiatan, seperti di tempat kerja, sekolah, atau acara tertentu. Absensi digunakan untuk memantau kehadiran, keterlambatan, dan ketidakhadiran seseorang, yang dapat memengaruhi penilaian kinerja atau disiplin kerja.
7	Pendidikan	Perusahaan sering kali memeriksa latar belakang pendidikan karyawan untuk memastikan bahwa karyawan memiliki kualifikasi yang relevan dengan pengangkatan kelayakan karyawan kontrak
8	Kuantitas kerja	Kuantitas kerja adalah ukuran jumlah output atau hasil yang dihasilkan oleh seseorang atau tim dalam suatu periode waktu tertentu. Kuantitas kerja biasanya digunakan sebagai indikator produktivitas dalam dunia kerja dan dapat diukur dengan berbagai cara tergantung pada jenis pekerjaan yang dilakukan
9	Disiplin kerja	Disiplin kerja adalah sikap dan perilaku karyawan dalam menaati aturan, prosedur, serta standar yang ditetapkan dalam lingkungan kerja. Disiplin kerja mencerminkan tanggung jawab, kedisiplinan, dan komitmen seseorang terhadap pekerjaannya, yang berdampak langsung pada produktivitas dan keberhasilan organisasi.
10	Masa kerja	Masa kerja adalah periode waktu yang telah dijalani seseorang dalam suatu pekerjaan atau perusahaan sejak pertama kali bergabung hingga saat ini atau hingga ia berhenti bekerja. Masa kerja sering digunakan sebagai dasar dalam berbagai aspek ketenagakerjaan, seperti kenaikan gaji, promosi, tunjangan, pesangon, dan penghargaan.
11	Prestasi	Prestasi adalah hasil atau pencapaian yang diperoleh seseorang sebagai bentuk keberhasilan dalam bidang tertentu, baik dalam pekerjaan, akademik, olahraga, maupun aspek kehidupan lainnya. Prestasi menunjukkan usaha, kerja keras, dan dedikasi seseorang dalam mencapai suatu tujuan.
12	Komunikasi	Kemampuan berkomunikasi secara efektif baik dalam bentuk lisan maupun tulisan adalah keterampilan yang sangat dicari oleh perusahaan. Kemampuan untuk menyampaikan ide dengan jelas kepada rekan kerja, atasan, atau klien sangat penting.
13	Tanggung jawab	Tanggung jawab adalah kesadaran dan kewajiban seseorang untuk menjalankan tugas, kewajiban, atau peran yang telah dipercayakan kepadanya dengan penuh komitmen dan konsekuensi. Tanggung jawab mencerminkan kedewasaan, kepercayaan, dan integritas seseorang dalam berbagai aspek kehidupan, termasuk pekerjaan, keluarga, dan masyarakat.
14	Status	Status dalam penentuan Perpanjangan Kontrak Karyawan yaitu layak dan tidak layak

3.1. Hasil Penelitian

1. Preprocessing Data

Setelah dilakukan proses Pre-processing data yang diperoleh adalah 975 karyawan pada PT. Duta Agung Jaya. Maka proses Pre-processing dapat dilihat pada Tabel 2 berikut :

1. Missing Value

Missing Value dalam proses preprocessing merujuk pada data yang tidak tersedia atau hilang pada fitur atau variabel dalam *dataset*. Ini merupakan nilai yang tidak terisi, yang bisa terjadi karena berbagai alasan seperti kesalahan pengumpulan data, responden yang tidak menjawab pertanyaan, atau data yang tidak dapat diukur. Teknik yang sering digunakan untuk mengatasi missing value meliputi:

- Penghapusan Data (Deletion) : Menghapus baris atau kolom yang memiliki missing value jika jumlahnya sedikit.
- Imputasi: Menggantikan missing value dengan nilai pengganti, misalnya dengan rata-rata, median, modus, atau metode imputasi lanjutan seperti regresi.

Tabel 2. Statistik Missing Value

No	Fitur	Missing Value
1	Nomor	0
2	Nama Karyawan	0
3	Alamat	0
4	Tanggal Lahir	0
5	No Telp	0
7	Absensi	17
8	Pendidikan	11
9	Kuantitas kerja	32
10	disiplin kerja	23
11	Masa kerja	12
12	Prestasi	2
13	Komunikasi	12
14	tanggung jawab	3
15	Status	0

2. Duplikat data

Dalam proses *preprocessing data*, duplikasi data bisa menjadi masalah serius karena dapat menyebabkan analisis yang bias, mempengaruhi hasil model machine learning, dan meningkatkan penggunaan memori yang tidak perlu. Ada beberapa data yg duplikat yaitu pengulangan data karyawan yang sama sebanyak 8 data karyawan.

3. Seleksi Fitur

Seleksi fitur adalah proses memilih subset fitur yang paling relevan untuk meningkatkan performa model machine learning, mengurangi overfitting, dan mempercepat proses training. Ada beberapa fitur yang diseleksi diantaranya adalah no, nama karyawan, alamat, tanggal lahir dan no telp.

Dataset dibagi menjadi data training dan data testing yang merupakan data karyawan dalam penentuan perpanjangan kontrak karyawan. Proses pembagian data set dilakukan yaitu 20% untuk testing dan 80% untuk data training. Maka hasil penentuan split data sebagai berikut :

Data Testing = $20\% \times 975 = 195$

Data Training = $80\% \times 975 = 780$

1. Data Karyawan

Tabel 3. Data Karyawan (Training)

Absensi	Pendidikan	Kualitas Kerja	Disiplin	Masa Kerja	Prestasi	Komunikasi	Tanggung jawab	Status
2 hari	SMA	89	72	1.5 Tahun	80	80	72	Layak
1 hari	SMA	89	72	4.5 Tahun	80	87	80	Layak
3 hari	SMA	70	72	1.5 Tahun	87	87	82	Layak
3 hari	SMA	80	72	1.5 Tahun	72	72	84	Layak
4 hari	SMA	87	80	4 Tahun	80	87	85	Layak
3 hari	SMA	87	72	9 Bulan	87	89	72	Tidak Layak

3 hari	SMA	87	72	9 Bulan	87	89	80	Tidak Layak
3 hari	SMA	89	72	1.5 Tahun	72	89	72	Layak
4 hari	SMA	89	80	1.5 Tahun	87	70	72	Tidak Layak
4 hari	SMA	89	72	1.5 Tahun	87	80	87	Layak
1 hari	SMA	70	72	4.5 Tahun	80	87	80	Tidak Layak
3 hari	SMA	80	87	1.5 Tahun	72	80	87	Tidak Layak
4 hari	SMA	87	80	4 Tahun	80	87	89	Layak
4 hari	SMA	80	87	4.5 Tahun	70	87	87	Tidak Layak
...	...	...	...	...	...	...	...	...
2 hari	D3	80	87	4.5 Tahun	80	80	80	Tidak Layak
5 hari	SMA	80	72	4 Bulan	87	80	80	Layak
3 hari	SMA	80	80	4 Tahun	72	87	72	Layak

2. Perhitungan Entropy

Pada data penerimaan, jumlah kelas adalah 2, yaitu : ‘Layak dan ‘tidak layak’. Jumlah sampel kelas 1 (‘Layak’) adalah 414 dan jumlah sampel untuk kelas 2 (‘tidak layak’) adalah 366. jadi $p_1 = 414$ dan $p_2 = 366$. Dengan demikian entropy untuk kumpulan sampel data S adalah :

$$\begin{aligned} \text{Entropy (S)} &= - (414/780) \log_2 (414/780) - (366/780) \log_2 (366/780) \\ &= 0,4850 - (-0,5122) \\ &= 0,9972 \end{aligned}$$

Dari tabel atribut dipertahankan=Layak merupakan sample (+), dan atribut dipertahankan = Tidak Layak merupakan (-) sample pada tabel didapat

1. Absensi

Values (Absensi) = 1 hari, 2 hari, 3 hari, 4 hari, 5 hari

$$S = [414+, 366-], |S| = 780$$

$$S_{1\text{hari}} = [47+, 27-], |S| = 74$$

$$S_{2\text{hari}} = [80+, 69-], |S| = 149$$

$$S_{3\text{hari}} = [172+, 169-], |S| = 341$$

$$S_{4\text{hari}} = [75+, 75-], |S| = 150$$

$$S_{5\text{hari}} = [40+, 26-], |S| = 66$$

Hitung Entrophy $S_{1\text{hari}}, S_{2\text{hari}}, S_{3\text{hari}}, S_{4\text{hari}}, S_{5\text{hari}}$ dan information Gain untuk nilai Absensi adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\begin{aligned} \text{Entropy } S_{1\text{hari}} &= - (47/74) \log_2 (47/74) - (27/74) \log_2 (27/74) \\ &= 0,4159 - (-0,5307) \\ &= 0,9466 \end{aligned}$$

$$\begin{aligned} \text{Entropy } S_{2\text{hari}} &= - (80/149) \log_2 (80/149) - (69/149) \log_2 (69/149) \\ &= 0,4817 - (-0,5143) \\ &= 0,9960 \end{aligned}$$

$$\begin{aligned} \text{Entropy } S_{3\text{hari}} &= - (172/341) \log_2 (172/341) - (169/341) \log_2 (169/341) \\ &= 0,4980 - (-0,5019) \\ &= 0,9999 \end{aligned}$$

$$\begin{aligned} \text{Entropy } S_{4\text{hari}} &= - (75/150) \log_2 (75/150) - (75/150) \log_2 (75/150) \\ &= 0,5 - (-0,5) \\ &= 1 \end{aligned}$$

$$\begin{aligned} \text{Entropy } S_{5\text{hari}} &= - (40/66) \log_2 (40/66) - (26/66) \log_2 (26/66) \\ &= 0,4378 - (-0,5294) \\ &= 0,9672 \end{aligned}$$

Information Gain (S.Absensi) = Entrophy(S)-

$$\sum \text{Value}(S_{1\text{hari}}, S_{2\text{hari}}, S_{3\text{hari}}, S_{4\text{hari}}, S_{5\text{hari}}) \cdot \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9972 - (74/780) * 0,9466 - (149/780) * 0,9960 - (341/780) * 0,9999 - (150/780) * 1 - (66/780) * 0,9672$$

$$= 0,005872$$

2. Pendidikan

Values (pendidikan) = SMA, D3

$S = [414+, 366-]$, $|s| = 780$

$S_{SMA} = [370+, 336-]$, $|s| = 706$

$S_{D3} = [44+, 30-]$, $|s| = 74$

Hitung Entropy S_{SMA} , S_{D3} dan information Gain untuk nilai pendidikan adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\text{Entropy } S_{SMA} = - (370/706) \log_2 (370/706) - (336/706) \log_2 (336/706)$$

$$= 0,4885 - (-0,5098)$$

$$= 0,9983$$

$$\text{Entropy } S_{D3} = - (44/74) \log_2 (44/74) - (30/74) \log_2 (30/74)$$

$$= 0,4459 - (-0,5280)$$

$$= 0,9740$$

Information Gain (S.Pendidikan) = Entropy(S)-

$$\sum \text{Value}(S_{SMA}, S_{D3}) \cdot \frac{|sv|}{|s|} \text{Entropy}(Sv)$$

$$= 0,9972 - (706/780) * 0,9983 - (74/780) * 0,9740$$

$$= 0,001246$$

3. Kuantitas Kerja

Values (kuantitas kerja) = <80 , ≥ 80

$S = [414+, 366-]$, $|s| = 780$

$S_{<80} = [105+, 101-]$, $|s| = 206$

$S_{\geq 80} = [309+, 265-]$, $|s| = 574$

Hitung Entropy $S_{<80}$, $S_{\geq 80}$, dan information Gain untuk nilai kuantitas kerja adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\text{Entropy } S_{<80} = - (105/206) \log_2 (105/206) - (101/206) \log_2 (101/206)$$

$$= 0,99973$$

$$\text{Entropy } S_{\geq 80} = - (309/574) \log_2 (309/574) - (265/574) \log_2 (265/574)$$

$$= 0,9957$$

Information Gain (S.kuantitas kerja) = Entropy(S)-

$$\sum \text{Value}(S_{<80} \text{ dan } S_{\geq 80}) \cdot \frac{|sv|}{|s|} \text{Entropy}(Sv)$$

$$= 0,9972 - (206/780) * 0,99973 - (574/780) * 0,9957$$

$$= 0,000462$$

4. Disiplin Kerja

Values (Disiplin kerja) = <80 dan ≥ 80

$S = [414+, 366-]$, $|s| = 780$

$S_{<80} = [169+, 153-]$, $|s| = 322$

$S_{\geq 80} = [245+, 213-]$, $|s| = 458$

Hitung Entropy $S_{<80}$, $S_{\geq 80}$ dan information Gain untuk nilai disiplin kerja adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\text{Entropy } S_{<80} = - (169/322) \log_2 (169/322) - (153/322) \log_2 (153/322)$$

$$= 0,9982$$

$$\text{Entropy } S_{\geq 80} = - (245/458) \log_2 (245/458) - (213/458) \log_2 (213/458)$$

$$= 0,9964$$

Information Gain (S.Absensi) = Entropy(S)-

$$\sum \text{Value}(S_{<80}, S_{\geq 80}) \cdot \frac{|sv|}{|s|} \text{Entropy}(Sv)$$

$$= 0,9972 - (322/780) * 0,9982 - (458/780) * 0,9964$$

$$= 0,0000720$$

5. Masa Kerja

Values (masa kerja) = 4 Bulan, 9 Bulan, 1.5 tahun, 4 Tahun, 4.5 tahun

$S = [414+, 366-]$, $|s| = 780$

$S_{4\text{bulan}} = [37+, 27-]$, $|s| = 64$

$S_{9\text{Bulan}} = [119+, 98-]$, $|s| = 217$

$S_{1.5\text{tahun}} = [148+, 140-]$, $|s| = 288$

$S_{4\text{tahun}} = [63+, 48-]$, $|s| = 111$

$S_{4.5\text{tahun}} = [47+, 53-]$, $|s| = 100$

Hitung Entropy $S_{4\text{bulan}}$, $S_{9\text{bulan}}$, $S_{1.5\text{tahun}}$, $S_{4\text{tahun}}$, $S_{4.5\text{tahun}}$ dan information Gain untuk nilai masa kerja adalah

:

$$\text{Entropy}(S) = 0,9972$$

$$\begin{aligned}\text{Entropy } S_{4\text{bulan}} &= - (37/64) \log_2 (37/64) - (27/64) \log_2 (27/64) \\ &= 0,4570 - (-0,5252) \\ &= 0,9823\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{9\text{bulan}} &= - (119/217) \log_2 (119/217) - (98/217) \log_2 (98/217) \\ &= 0 - (0,5179) \\ &= 0,9932\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{1,5\text{tahun}} &= - (148/288) \log_2 (148/288) - (140/288) \log_2 (140/288) \\ &= 0,4935 - (-0,5058) \\ &= 0,9994\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{4\text{tahun}} &= - (63/111) \log_2 (63/111) - (48/111) \log_2 (48/111) \\ &= 0,4637 - (-0,5230) \\ &= 0,9867\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{4,5\text{tahun}} &= - (47/100) \log_2 (47/100) - (53/100) \log_2 (53/100) \\ &= 0,5119 - (-0,4854) \\ &= 0,9974\end{aligned}$$

$$\text{Information Gain (S.Masa Kerja)} = \text{Entropy}(S) -$$

$$\sum \text{Value}(S_{4\text{bulan}}, S_{9\text{bulan}}, S_{1,5\text{tahun}}, S_{4\text{tahun}}, S_{4,5\text{tahun}}) \cdot \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9972 - (67/780) * 0,9823 - (217/780) * 0,9932 - (288/780) * 0,9994 - (111/780) * 0,9867 - (100/780) * 0,9974$$

$$= 0,00302$$

6. Prestasi

$$\text{Values (prestasi)} = <80, \geq 80$$

$$S = [414+, 366-], |S| = 780$$

$$S_{<80} = [103+, 81-], |S| = 184$$

$$S_{\geq 80} = [311+, 285-], |S| = 596$$

Hitung Entropy $S_{<80}$, $S_{\geq 80}$ dan information Gain untuk nilai prestasi adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\begin{aligned}\text{Entropy } S_{<80} &= - (103/184) \log_2 (103/184) - (81/184) \log_2 (81/184) \\ &= 0,9896\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{\geq 80} &= - (311/596) \log_2 (311/596) - (285/596) \log_2 (285/596) \\ &= 0,9996\end{aligned}$$

$$\text{Information Gain (S.prestasi)} = \text{Entropy}(S) -$$

$$\sum \text{Value}(S_{<80}, S_{\geq 80}) \cdot \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9972 - (185/780) * 0,9896 - (596/780) * 0,9996$$

$$= 0,0007560$$

7. Komunikasi

$$\text{Values (komunikasi)} = <80, \geq 80$$

$$S = [414+, 366-], |S| = 780$$

$$S_{<80} = [99+, 95-], |S| = 206$$

$$S_{\geq 80} = [315+, 271-], |S| = 574$$

Hitung Entropy $S_{<80}$, $S_{\geq 80}$ dan information Gain untuk nilai prestasi adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\begin{aligned}\text{Entropy } S_{<80} &= - (99/184) \log_2 (99/184) - (95/184) \log_2 (95/184) \\ &= 0,9996\end{aligned}$$

$$\begin{aligned}\text{Entropy } S_{\geq 80} &= - (315/596) \log_2 (315/596) - (271/596) \log_2 (271/596) \\ &= 0,9959\end{aligned}$$

$$\text{Information Gain (S.komunikasi)} = \text{Entropy}(S) -$$

$$\sum \text{Value}(S_{<80}, S_{\geq 80}) \cdot \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9972 - (206/780) * 0,9996 - (574/780) * 0,9959$$

$$= 0,000405$$

8. Tanggung Jawab

$$\text{Values (tanggung jawab)} = <80, \geq 80$$

$$S = [414+, 366-], |S| = 780$$

$$S_{<80} = [178+, 161-], |S| = 339$$

$$S_{\geq 80} = [236+, 205-], |S| = 441$$

Hitung Entropy $S_{<80}$, $S_{\geq 80}$ dan information Gain untuk nilai prestasi adalah :

$$\text{Entropy}(S) = 0,9972$$

$$\text{Entropy } S_{<80} = - (178/339) \log_2 (178/339) - (161/339) \log_2 (161/339) = 0,9981$$

$$\text{Entropy } S_{\geq 80} = - (236/441) \log_2 (236/441) - (205/441) \log_2 (205/441) = 0,9964$$

$$\text{Information Gain (S.komunikasi)} = \text{Entropy}(S) -$$

$$\sum \text{Value}(S_{<80}, S_{\geq 80}) \cdot \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

$$= 0,9972 - (339/780) * 0,9981 - (441/780) * 0,9964$$

$$= 0,0000750$$

Dari ketiga nilai Information Gain diatas Gain (S, kelayakan karyawan kontrak) adalah yang terbesar sehingga atribut prestasi merupakan the best classifier dan harus diletakkan sebagai root.

Tabel 4. Hasil perhitungan entropy dan gain untuk node akar rekursi Level 0 Iterasi 1

	Node	Ket	Jumlah	Layak	Tidak Layak	Entropy	Information Gain
	Total		780	414	366	0,99727	
1	Absensi	1 Hari	74	47	27	0,94665	0,005872
		2 Hari	149	80	69	0,99606	
		3 Hari	341	172	169	0,99994	
		4 Hari	150	75	75	1	
		5 Hari	66	40	26	0,96729	
2	Pendidikan	SMA	706	370	336	0,99833	0,001246
		D3	74	44	30	0,97402	
3	Kualitas Kerja	<80	105	101	206	0,99973	0,0004
		>=80	309	265	574	0,99576	
4	Disiplin Kerja	<80	169	153	322	0,9982	0,0000720
		>=80	245	213	458	0,9964	
5	Masa Kerja	4 Bulan	64	37	27	0,98232	0,0030
		9 Bulan	217	119	98	0,51793	
		1.5 tahun	288	148	140	0,99944	
		4 tahun	111	63	48	0,98679	
		4.5 tahun	100	47	53	0,9974	
6	Prestasi	<80	103	81	184	0,9896	0,000756
		>=80	311	285	596	0,9986	
7	Komunikasi	<80	99	95	194	0,9996	0,000040
		>=80	315	271	586	0,9959	
8	Tanggung Jawab	<80	178	161	339	0,9981	0,000075
		>=80	236	205	441	0,9964	

Berdasarkan pembentukan pohon keputusan node 1 (root node), Node 2 akan dianalisis lebih lanjut. Untuk mempermudah, Tabel dibawah difilter, dengan mengambil data yang memiliki “absensi”= 3 hari. Setelah menghitung menggunakan ID3 maka diperoleh hasil Rekursi Level 1 Iterasi 2 sebagai berikut

Tabel 5. Hasil perhitungan entropy dan gain untuk node akar Rekursi Level 1 Iterasi 2

	Node	Ket	Layak	Tidak Layak	Entropy	Information Gain
	Total		414	366	0,99727	
1	Absensi	1 Hari	17	11	0,966619	0,03264
		2 Hari	12	15	0,991076	
		3 Hari	13	15	0,996317	
		4 Hari	3	6	0,918296	
		5 Hari	2	6	0,811278	
2	Pendidikan	SMA	42	46	0,998509	0,001129
		D3	5	7	0,979869	
3	Kualitas Kerja	<80	9	4	0,890492	0,034692
		>=80	9	11	0,992774	
4	Disiplin Kerja	<80	70	0	0	0,021486
		>=80	72	20	1	
5	Masa Kerja	4 Bulan	37	27	0,98232	0,005251
		9 Bulan	119	98	0,51793	
		1.5 tahun	148	140	0,99944	
		4 tahun	63	48	0,98679	
		4.5 tahun	47	53	0,9974	
6	Komunikasi	89	26	22	0,99498	0,005853
		<80	0	0	0	
		>=80	10	10	1	
7	Tanggung Jawab	<80	0	0	0	0,025209
		>=80	11	11	1	

3.2. Pembahasan

Berikut ini data testing yang akan dievaluasi :

Tabel 6. Data testing

No.	Nama karyawan	V1	V2	V3	V4	V5	V6	V7	V8	Status	Keputusan sistem
1	Mega Wati	3 Hari	SMA	>=80	<80	1.5 Tahun	>=80	>=80	<80	Layak	Layak
2	Riska Fatmawati	4 Hari	SMA	>=80	>=80	1.5 Tahun	<80	>=80	>=80	Layak	Layak
3	Edi Gunawan	4 Hari	SMA	<80	>=80	1.5 Tahun	>=80	<80	>=80	Tidak Layak	Tidak Layak
4	Luspi	1 Hari	SMA	>=80	>=80	4.5 Tahun	<80	>=80	<80	Layak	Layak
5	Lia Pradila Sari	3 Hari	SMA	>=80	>=80	4 Tahun	<80	>=80	<80	Layak	Layak
6	Maulidayani	4 Hari	SMA	>=80	>=80	4 Tahun	>=80	>=80	<80	Tidak Layak	Tidak Layak
7	Putri Dewi Septiani	4 Hari	SMA	>=80	>=80	9 Bulan	>=80	<80	>=80	admin	Layak

No. Nama karyawan	V1	V2	V3	V4	V5	V6	V7	V8	Status	Keputusan sistem
8 Afrina	2 Hari	SMA	>=80	>=80	1.5 Tahun	>=80	>=80	<80	Tidak Layak	Tidak Layak
9 Alfian	3 Hari	SMA	<80	<80	9 Bulan	>=80	<80	<80	Tidak Layak	Tidak Layak
10 Siti Nuri	2 Hari	D3	>=80	<80	9 Bulan	>=80	<80	>=80	Layak	Layak
11 Firawansyah	1 Hari	D3	>=80	<80	9 Bulan	>=80	>=80	>=80	Layak	Layak
12 Zahrina Shalhan	3 Hari	SMA	>=80	<80	4 Bulan	<80	>=80	>=80	Layak	Layak
13 Fera Muzdalifah	3 Hari	SMA	>=80	<80	4.5 Tahun	>=80	>=80	>=80	Tidak Layak	Tidak Layak
14 Jumala	2 Hari	SMA	>=80	>=80	4 Bulan	>=80	>=80	>=80	Layak	Layak
15 Safrina	3 Hari	SMA	<80	>=80	4 Tahun	>=80	>=80	<80	Layak	Layak
16 Syamsuddin	3 Hari	SMA	>=80	<80	4 Tahun	<80	<80	<80	Tidak Layak	Tidak Layak
17 Malisga Afwica	3 Hari	SMA	>=80	<80	1.5 Tahun	>=80	>=80	>=80	Layak	Layak
18 Ismayani	4 Hari	SMA	>=80	>=80	1.5 Tahun	>=80	>=80	<80	Tidak Layak	Tidak Layak
19 Miftahul Jannah	4 Hari	SMA	<80	>=80	1.5 Tahun	>=80	>=80	>=80	Tidak Layak	Tidak Layak
20 Rinta Malasari	1 Hari	SMA	>=80	>=80	4.5 Tahun	>=80	<80	>=80	Layak	Layak
21 Yunita Sari	3 Hari	SMA	>=80	>=80	1.5 Tahun	>=80	>=80	<80	Tidak Layak	Tidak Layak
22 Rohani	4 Hari	SMA	<80	>=80	1.5 Tahun	>=80	<80	<80	Tidak Layak	Tidak Layak
23 Arina	4 Hari	SMA	>=80	>=80	4 Tahun	>=80	<80	<80	Tidak Layak	Layak
24 Sari Rizki	2 Hari	SMA	>=80	>=80	1.5 Tahun	>=80	>=80	>=80	Tidak Layak	Tidak Layak
25 Khairun Nisak	3 Hari	SMA	>=80	<80	1.5 Tahun	<80	>=80	<80	Tidak Layak	Tidak Layak

Tabel evaluasi model dapat ditemukan pada tabel berikut untuk dta testing :

Tabel 7. Confusion Matrix

Confusion Matrix		Nilai aktual	
		Positif	Negatif
Nilai	Positif	12	0
Prediksi	Negatif	1	11
Jumlah		13	12
Total		25	

Confusion matrix bertujuan untuk memvisualisasikan prediksi dan keadaan aktual dari data yang dihasilkan oleh algoritma machine learning. Hal ini dilakukan dengan menghitung akurasi, presisi, recall, dan F1-score. Metrik-metrik ini sangat penting dalam mengevaluasi kinerja klasifikasi atau algoritma machine learning yang digunakan untuk melakukan prediksi. Rumus untuk keempat metrik ini ditunjukkan dalam tabel berikut

Tabel 8. Rumus confusion matrix

No	Pengukuran	Rumus	Hasil
1	Akurasi	$12 + 12 / (12 + 0 + 1 + 12)$	96%
2	Presisi	$12 / (12 + 0)$	100%
3	Recall	$12 / (12 + 1)$	92%
4	F1-Score	$2 * 0.92 * 1 / (0.92 + 1)$	96%

Bagian evaluasi model dari sistem, yang menampilkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil evaluasi ini menunjukkan bahwa sistem memiliki tingkat akurasi sebesar 90%, presisi 93% dan recall masing-masing sebesar 87%, serta F1-score sebesar 89%. Metrik ini digunakan untuk mengukur kinerja dan keandalan model ID3 dalam memprediksi keputusan perpanjangan kontrak, dan hasil yang ditampilkan menunjukkan bahwa sistem memiliki performa yang sangat baik. Berikut ini merupakan hasil pengujian unduk data baru ketikan melakukan pengujian data karyawan baru :

Tabel 9. Evaluasi Data Baru

Confusion Matrix		Nilai aktual	
		Positif	Negatif
Nilai Prediksi	Positif	13	2
	Negatif	1	14
Jumlah		14	16

Tabel 10. Perhitungan Confusion Matriks

NO	PENGUKURAN	RUMUS	HASIL
1	Akurasi	$13 + 14 / (13 + 2 + 1 + 14)$	90 %
2	Presisi	$13 / (13 + 2)$	87 %
3	Recall	$13 / (13 + 1)$	93 %
4	F1-Score	$2 * 0.93 * 0.87 / (0.93 + 0.87)$	90 %

Berdasarkan Tabel 10, perhitungan metrik evaluasi menggunakan confusion matrix pada pengujian data baru (30 data) menunjukkan bahwa sistem memiliki kinerja yang baik. Nilai akurasi sebesar 90% diperoleh dari perbandingan jumlah prediksi benar (TP + TN) terhadap seluruh data uji. Nilai presisi sebesar 87% menunjukkan bahwa sebagian besar prediksi “layak” yang dihasilkan sistem benar, sedangkan nilai recall sebesar 93% menandakan kemampuan model yang tinggi dalam menemukan karyawan yang benar-benar layak diperpanjang. Selanjutnya, nilai F1-Score 90% mengindikasikan keseimbangan yang baik antara presisi dan recall, sehingga model dinilai cukup konsisten dan efektif dalam klasifikasi kelayakan perpanjangan kontrak.

4. KESIMPULAN

Algoritma ID3 mampu menyusun keputusan berdasarkan atribut yang relevan secara hierarkis dan terstruktur. Dari hasil pengujian terhadap 25 data testing, sistem menunjukkan performa yang sangat baik, yaitu Akurasi 96%, Presisi 100%, Recall 92% dan F1-Score 96%. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat tinggi dalam mengidentifikasi karyawan yang layak atau tidak layak diperpanjang kontraknya, dengan tingkat kesalahan yang sangat rendah. Pengujian dengan Data Baru (30 Data) Saat diuji dengan data baru sebanyak 30 data, sistem juga tetap menunjukkan performa yang baik, dengan hasil Akurasi 90%, Presisi 87%, Recall 93%, F1-Score 90%. Ini membuktikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilatih sebelumnya, dan tetap konsisten dalam memberikan prediksi yang akurat. Implementasi sistem ini memberikan manfaat nyata bagi perusahaan, khususnya dalam hal mempermudah proses evaluasi dan pengambilan keputusan terkait perpanjangan kontrak karyawan secara objektif dan terukur dan mengurangi potensi bias subjektif dari manajemen.

REFERENCES

- [1] M. Ali, R. Prasetyo, and B. Sujatmiko, “Rekayasa Sistem Pendukung Keputusan dengan Algoritma ID3 dalam Penentuan Jurusan Akademik (Studi Kasus: SMA Negeri 3 Ngajuk),” *Inov. J. Ilm. Inov. Teknol. Inf.*, vol. 7, no. 1, pp. 17–23, 2022, doi: 10.33752/inovate.v7i1.3677.
- [2] M. B. P. Husaini, A. Pranata, and I. Mariami, “Sistem Pendukung Keputusan Untuk Menentukan Perpanjangan Kontrak Kerja Karyawan Menggunakan Metode Elimination and Choice Translation Reality (ELECTRE) Pada PT. Bengkel Bangun Service,” *J. Cyber Tech*, vol. 4, no. 5, pp. 1–11, 2021, doi: 10.53513/jct.v4i5.4000.
- [3] E. Turban and J. E. Aronson, *Decision Support Systems and Intelligent Systems*. Upper Saddle River, NJ, USA: Prentice Hall, 1998.
- [4] A. Abedinia and V. Seydi, “Building semi-supervised decision trees with semi-cart algorithm,” *Int. J. Mach. Learn. Cybern.*, vol. 15, pp. 4493–4510, 2024, doi: 10.1007/s13042-024-02161-z.
- [5] S. Moslehi, N. Rabiei, A. R. Soltanian, and others, “Application of machine learning models based on decision trees in classifying the factors affecting mortality of COVID-19 patients in Hamadan, Iran,” *BMC Med. Inform. Decis. Mak.*, vol. 22, p. 192, 2022, doi: 10.1186/s12911-022-01939-x.
- [6] Z. Mohammadi-Pirouz, K. Hajian-Tilaki, M. Sadeghi Haddat-Zavareh, and others, “Development of decision tree classification algorithms in predicting mortality of COVID-19 patients,” *Int. J. Emerg. Med.*, vol. 17, p. 126, 2024, doi: 10.1186/s12245-024-00681-7.
- [7] I. Kurniawati, A. Muhidin, and F. E. Putra, “Klasifikasi Perpanjangan Kontrak Karyawan pada PT Milpo Menggunakan Algoritma Naive Bayes,” *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 8, no. 1, pp. 249–259, 2024, doi: 10.33395/remik.v8i1.13358.
- [8] D. K. Widiyaningrum, R. Septiani, P. Larasati, and A. Wibowo, “Algoritma Naive Bayes untuk Menentukan Kelayakan Perpanjangan Kontrak pada PT. Gemilang,” *J. Ilmu Komput. dan Bisnis*, vol. 11, no. 2, pp. 2441–2445, 2020, doi: 10.47927/jikb.v11i2.5.
- [9] N. Ramadhani, A. W. Syahroni, and I. Darmawan, “Pencarian Klasifier Terbaik sebagai Solusi Rekrutmen Karyawan Menggunakan Algoritma Iterative Dichotomiser 3 (ID3),” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 2, pp. 324–334, 2022, doi: 10.30743/infotekjar.v6i2.4747.
- [10] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, “Decision trees: from efficient prediction to responsible AI,” *Front. Artif. Intell.*, vol. 6, p. 1124553, 2023, doi: 10.3389/fraci.2023.1124553.
- [11] R. Ofori-Boateng, M. Aceves-Martins, N. Wiratunga, and others, “Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review,” *Artif. Intell. Rev.*, vol. 57, p. 200, 2024, doi: 10.1007/s10462-024-10844-w.
- [12] V. Blanco, A. Japón, and J. Puerto, “Robust optimal classification trees under noisy labels,” *Adv. Data Anal. Classif.*, vol. 16, pp. 155–179, 2022, doi: 10.1007/s11634-021-00467-2.
- [13] M. Koren and O. Peretz, “A quantum procedure for estimating information gain in Boolean classification task,” *Quantum Mach. Intell.*, vol. 6, no. 1, p. 16, 2024, doi: 10.1007/s42484-024-00151-6.
- [14] Suhartini, M. Sadali, and Y. K. Putra, “Sistem Informasi Berbasis Web SMA Al-Mukhtariyah Mamben Lauk Berbasis PHP dan MySQL dengan Framework CodeIgniter,” *Infotek J. Inform. dan Teknol.*, vol. 3, no. 1, pp. 79–83, 2020, doi: 10.29408/jit.v3i1.1793.
- [15] S. Yang, J.-Z. Guo, and J.-W. Jin, “An improved Id3 algorithm for medical data classification,” *Comput. & Electr. Eng.*, vol. 65, pp. 474–487, Jan. 2018, doi: 10.1016/j.compeleceng.2017.08.005.



- [16] R. Buaton, H. Mawengkang, M. Zarlis, S. Effendi, A. M. H. Pardede, and others, "Decision Tree Optimization in Data Mining with Support and Confidence," in *Journal of Physics: Conference Series*, 2019, p. 12056. doi: 10.1088/1742-6596/1255/1/012056.
- [17] Y. Wang, Y. Li, Y. Song, X. Rong, and S. Zhang, "Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree," *Algorithms*, vol. 10, no. 4, p. 124, 2017, doi: 10.3390/a10040124.
- [18] scikit-learn Developers, "Decision Trees — scikit-learn User Guide (Mathematical formulation)," 2025. [Online]. Available: <https://scikit-learn.org/stable/modules/tree.html>