



## *Classification of New Employee Selection Using the C4.5 Algorithm*

### **Klasifikasi Seleksi Penerimaan Karyawan Baru Menggunakan Algoritma C4.5**

Descania Mayangsari<sup>1</sup>, Tri Wahyudi<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Indonesia

E-Mail: <sup>1</sup>descaniam@gmail.com, <sup>2</sup>triwahyudi100390@gmail.com

Received Jul 9th 2024; Revised Aug 27th 2024; Accepted Sep 29th 2024; Available Online Nov 24th 2024

Corresponding Author: Descania Mayangsari

Copyright © 2025 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

#### **Abstract**

*The process of selecting new employee candidates to determine qualified candidates for the required positions tends to be subjective and inconsistent without effective analytical tools. This research aims to make candidate selection more efficient and facilitate the identification of suitable employees. The research focuses on new employee candidates selected through CV screening, interviews, psychometric tests, and user interviews. Using the C4.5 algorithm, which analyzes large amounts of data using concepts of entropy and gain, this study provides objective and consistent analysis for the selection process of new employees. Implementation of this approach can reduce time and costs associated with the selection process, as well as improve the quality of accepted candidates. Testing results with the C4.5 algorithm showed excellent performance with an accuracy of 99.43%, precision of 99.29%, recall of 97.69%, and an AUC of 0.750.*

*Keywords: C4.5 Algorithm, Classification, Data mining, Recruitment*

#### **Abstrak**

Proses seleksi penerimaan calon karyawan baru dalam menentukan kandidat yang sesuai kualifikasi dengan posisi yang sedang dibutuhkan, tanpa alat bantu analisis yang efektif, proses ini cenderung menjadi subyektif dan tidak konsisten. Diharapkan adanya penelitian ini membuat penyeleksian calon karyawan menjadi lebih efisien dan memudahkan dalam memilih calon karyawan yang tepat. Dengan obyek penelitian calon karyawan baru yang diseleksi melalui screening CV, Wawancara, psikotest, interview user. Dengan algoritma C4.5 yang merupakan salah satu metode untuk menganalisis data dalam jumlah yang banyak menggunakan konsep entropy dan gain. Hasil dari penelitian ini membantu proses seleksi penerimaan calon karyawan baru dengan memberikan analisis yang objektif dan konsisten. Implementasinya dapat mengurangi waktu dan biaya yang dibutuhkan dalam proses seleksi, serta meningkatkan kualitas kandidat yang diterima. Hasil pengujian dengan algoritma C4.5 memiliki nilai yang bagus yaitu dengan accuracy 99.43%, precision 99.29%, recall 97.69% dan AUC 0.750.

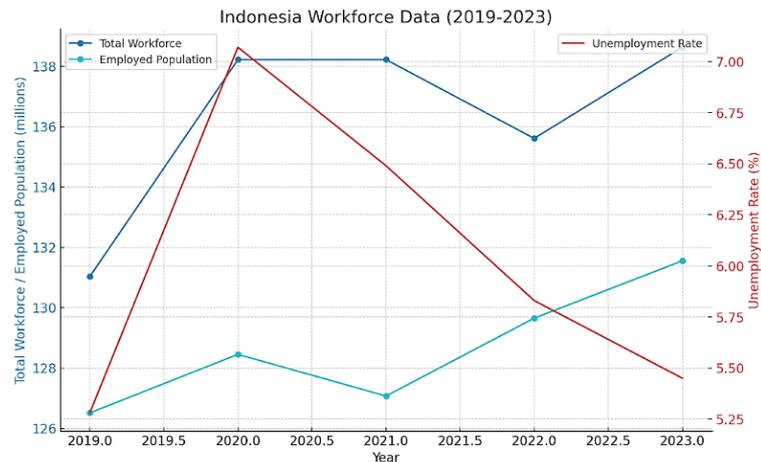
Kata Kunci: Algoritma C4.5, Data Mining, Klasifikasi, Rekrutmen

### **1. PENDAHULUAN**

Dalam era globalisasi persaingan bisnis yang semakin ketat, perusahaan dituntut untuk memiliki sumber daya manusia (SDM) yang berkualitas dan kompeten [1]. Sebagai perusahaan yang bergerak di bidang layanan outsourcing yang memiliki banyak vendor, mempunyai tantangan tersendiri dalam melakukan seleksi calon karyawan yang efektif dan efisien. Pada dasarnya, dalam suatu tujuan seleksi penerimaan karyawan adalah untuk mendapatkan sumber daya manusia yang tepat untuk suatu jabatan tertentu [2], sehingga orang tersebut mampu bekerja secara optimal [3] dan nantinya bekerja dan bertahan diperusahaan untuk jangka waktu yang lama [4].

Mengenai Sumber Daya Manusia (SDM) tenaga kerja di Indonesia dari tahun 2019 sampai 2023. Mengalami kenaikan dan penurunan yang sangat signifikan, terutama pada data pengangguran pada tahun 2019 sampai 2020 banyaknya pengangguran hingga 138 juta orang di Indonesia. Hal ini disebabkan karena adanya pandemi *covid-19* yang pada saat itu banyak perusahaan terpaksa harus mem-PHK-kan banyak pekerja bahkan tidak sedikit pula perusahaan yang gulung tikar karena perekonomian Indonesia yang sangat menurun akibat

*covid-19*. Namun masih banyak pula perusahaan yang masih bertahan ditengah pandemi *covid-19*. Gambar 1 adalah rincian data yang didapat dari Badan Pusat Statistik (BPS).



**Gambar 1.** Grafik jumlah tenaga kerja di Indonesia (Sumber: Badan Pusat Statistik)

Dari gambar 1 dapat disimpulkan untuk total SDM di Indonesia tidak terlalu banyak mengalami perubahan dari tahun 2019 sampai 2023. Begitupun dengan jumlah penduduk yang bekerja tidak banyak mengalami kenaikan atau penurunan jumlah. Namun berbeda dengan banyaknya pengangguran di Indonesia pada tahun 2019 sampai 2023. Pada tahun 2019 sampai 2020 mengalami kenaikan jumlah pengangguran dengan 127 juta orang atau 5.28% naik ke angka 138 juta orang atau 7.07% namun dari tahun 2020 sampai 2023 untuk jumlah pengangguran di Indonesia saat ini mengalami penurunan sebanyak 11 juta orang dan untuk data pengangguran di tahun 2023 sebanyak 11 juta orang atau 5.45%

Masalah dalam penelitian ini ialah adanya proses rekrutmen yang masih kurang sesuai kriteria perusahaan, sehingga perusahaan kesulitan dalam menempatkan karyawan karena ada beberapa karyawan yang tidak sesuai dengan kriteria dan kualifikasi yang dibutuhkan. Dan proses rekrutmen yang memakan waktu lumayan lama dan kurang efektif dalam mengambil suatu keputusan.

Dengan demikian dibutuhkan suatu perencanaan model otomatis dari sekumpulan data, dengan tujuan memberikan komputer kemampuan untuk belajar (*machine learning*). Maka dari itu diperlukan sebuah metode untuk menentukan perancang dalam pengambilan keputusan penerimaan karyawan baru Decision Tree Algoritma C4.5. Metode C4.5 salah satu algoritma pembelajaran mesin untuk membuat pohon keputusan. Latar belakang pemilihan metode ini untuk seleksi calon karyawan, karena kemampuannya dalam mengelola data yang kompleks dan menghasilkan aturan keputusan yang mudah dipahami. Dengan C4.5, perusahaan dapat mengidentifikasi fitur-fitur penting yang mempengaruhi keberhasilan calon karyawan dan membuat keputusan seleksi yang lebih efisien. tujuan khusus pada penelitian ini adalah membangun suatu sistem dengan mudah yang dapat mengelola data dan dapat diakses dengan mudah oleh staff HRD dalam menentukan karyawan baru pada PT. Kobus Smart Service Jakarta.

Pada penelitian sebelumnya penerapan algoritma C4.5 untuk Model Penentuan Penerima Beasiswa PIP [5]. Menentukan Jadwal Overtime menggunakan algoritma 24.5 Dengan Klasifikasi data karyawan [6]. Memprediksi minat calon mahasiswa memilih perguruan tinggi menggunakan algoritma C 4.5 [7]. Algoritma C4.5 adalah algoritma yang menghasilkan pohon keputusan, memiliki masukan berupa sampel klasifikasi [8]. Algoritma C4.5 yaitu metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan [9]. Penelitian terdahulu Algoritma C4.5 dapat membantu untuk mengambil keputusan penerimaan mahasiswa baru yang dapat lulus tepat waktu dan mahasiswa yang lulus terlambat dapat terprediksi lebih awal [10]. Algoritma C4.5 dapat digunakan sebagai metode klasifikasi dalam memprediksi profit perusahaan dengan memperhatikan nilai gain tertinggi dari tujuh atribut yaitu biaya tenaga ahli, biaya tenaga Teknik, biaya tenaga pendukung, biaya operasional kantor, biaya peralatan, biaya pelaporan dan dokumen, dan biaya mobilisasi [11].

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data Mining adalah studi tentang mengumpulkan, membersihkan, memproses, menganalisis, dan mendapatkan wawasan berguna dari data. Terdapat variasi yang luas dalam hal domain masalah, aplikasi, formulasi, dan representasi data yang dihadapi dalam aplikasi nyata. Oleh karena itu, "penambangan data" adalah istilah umum yang digunakan untuk menggambarkan berbagai aspek pemrosesan data ini [12]. Data

Mining adalah istilah yang digunakan untuk menggambarkan proses menjelajahi basis data besar untuk mencari pola-pola menarik dan yang sebelumnya tidak diketahui [13].

## 2.2 Klasifikasi

Klasifikasi adalah proses objek atau entitas yang dikelompokkan ke dalam kategori atau kelas yang memiliki karakteristik atau atribut yang sama. Tujuannya untuk menyederhanakan dan mengorganisir informasi sehingga memudahkan pemahaman, analisis, dan pengambilan keputusan. Klasifikasi sendiri memiliki penggunaan yang luas dalam berbagai bidang.

Dalam bidang rekrutmen, klasifikasi merujuk pada proses pengelompokan dan pengorganisasian kandidat berdasarkan berbagai kriteria yang relevan dengan kebutuhan perusahaan. Klasifikasi adalah proses yang paling penting untuk memastikan bahwa proses rekrutmen berjalan dengan efisien dan efektif, serta mempermudah pemilihan kandidat yang paling cocok untuk posisi tertentu. Pengklasifikasi yang dilatih dengan baik akan mampu memprediksi perilaku data yang belum pernah terlihat dari masalah yang sama adalah inti dari setiap proses klasifikasi otomatis [14].

## 2.3 Metode C4.5

C4.5 adalah algoritma pembelajaran mesin yang digunakan untuk membangun pohon Keputusan, yang merupakan model prediktif dalam analisis data dan pembelajaran mesin. Algoritma ini salah satu algoritma yang populer dan sering digunakan karena fitur – fiturnya serta kemudahan penggunaannya. Algoritma ini merupakan pengembangan dari algoritma sebelumnya yang dikenal dengan ID3 (Iterative Dichotomiser 3). Berikut adalah beberapa fitur utama dari algoritma C4.5:

### 1. Pohon Keputusan

C4.5 membangun pohon Keputusan yang digunakan untuk mengambil Keputusan berdasarkan serangkaian aturan if-then. Setiap simpul dalam pohon mewakili Keputusan berdasarkan nilai – nilai fitur tertentu. Induktor pohon keputusan adalah algoritma yang secara otomatis membangun pohon keputusan dari dataset yang diberikan. Biasanya, tujuan utamanya adalah menemukan pohon keputusan yang optimal dengan meminimalkan kesalahan generalisasi [15].

### 2. Pemilihan Fitur

C4.5 menggunakan metode pemilihan fitur yang disebut "gain ratio" untuk menentukan fitur mana yang harus digunakan untuk membagi data di setiap simpul dalam pohon keputusan. Fitur yang memiliki gain ratio tertinggi akan dipilih untuk pembagian.

### 3. Penanganan Data yang Hilang

Penanganan data yang hilang menggunakan strategi seperti mengabaikan instance yang memiliki nilai yang hilang atau menggantikan nilai yang hilang dengan nilai yang paling umum dari fitur tersebut.

### 4. Pruning

Proses pengurangan overfitting. C4.5 menggunakan proses pruning (penyisipan) pohon keputusan setelah Pembangunan. Pruning membantu menghapus cabang - cabang pohon yang tidak penting yang menyebabkan overfitting.

### 5. Klasifikasi dan Regresi

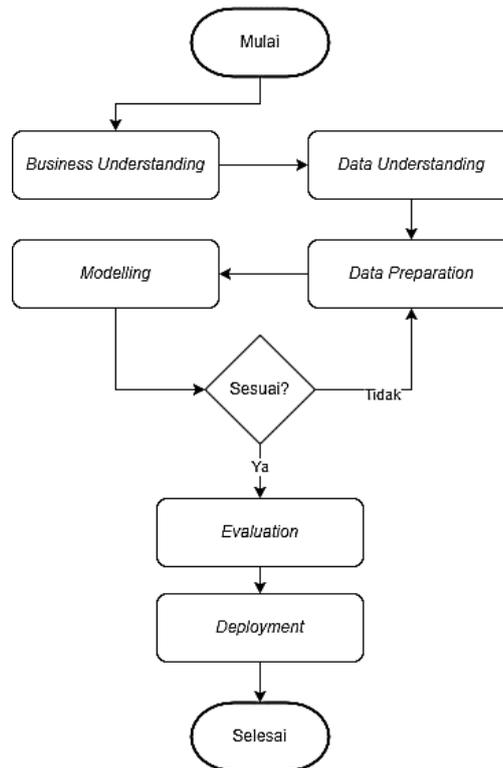
C4.5 sering digunakan untuk klasifikasi (memprediksi kelas dari instance), C4.5 juga dapat digunakan untuk regresi (memprediksi nilai numerik) dengan melakukan modifikasi pada algoritmanya.

## 2.4 RapidMiner

*RapidMiner* adalah platform analisis data yang kuat dan *user friendly* yang digunakan untuk mengekstraksi wawasan dari data. Aplikasi ini menyediakan berbagai alat dan fungsi untuk mempermudah berbagai tahap dalam proses analisis data, termasuk pembersihan data, eksplorasi data, pemodelan prediktif, evaluasi model, dan penerapan solusi analitis. *RapidMiner* adalah sistem yang mendukung perancangan dan dokumentasi keseluruhan proses penambangan data. Sistem ini tidak hanya menawarkan seperangkat operator yang hampir lengkap, tetapi juga struktur-struktur yang menggambarkan alur kontrol dari proses tersebut [16].

## 2.5 Penerapan Metode CRISP-DM

Penerapan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) memberikan kerangka kerja yang terstruktur dan memastikan bahwa semua tahapan yang relevan dalam proses data mining dipertimbangkan dengan cermat [17]. CRISP-DM sebagai *de facto* menjadi standar untuk pengembangan proyek *data mining* dan *knowledge discovery* karena paling banyak digunakan dalam pengembangan data mining [18]. Berikut alur dari penggunaan metode CRISP – DM:



**Gambar 2.** Penerapan Metodologi CRISP – DM

Berikut Penjelasan dari Metode CRISP - DM

1. *Business Understanding* (Pemahaman Bisnis)  
Bertujuan mengembangkan model untuk memprediksi penerimaan karyawan baru berdasarkan kriteria tertentu dan melakukan pemahaman yang mendalam untuk meneliti permasalahan yang sedang dihadapi
2. *Data Understanding* (Pemahaman Data)  
Pemahaman data dilakukan untuk mengetahui algoritma apa yang cocok untuk mengatasi permasalahan yang sedang diteliti.
  - a. Mengumpulkan semua data calon karyawan yang masuk
  - b. Mendeskripsikan keterangan data / label dengan menggunakan atribut join dan tidak join
3. *Data Preparation* (Persiapan Data)  
Melakukan pembersihan data untuk mengatasi missing values, outliers, dan kesalahan data lainnya
4. *Modeling* (Pemodelan):  
Dengan menerapkan algoritma C4.5
  - a. Mengimplementasikan pohon keputusan pada *rapidminer*
  - b. Membagi dataset berdasarkan atribut terpilih dan mengulang proses untuk setiap subset
5. Proses Hasil Sesuai atau Tidaknya  
Jika Tidak sesuai cek kembali pada data preparation masih ada atau tidaknya data yang *missing values*, *outliers*, dan kesalahan data lainnya. Jika sesuai lanjut Evaluasi.
6. *Evaluation* (Evaluasi)
  - a. Mengitung Gain dan Entropy untuk melakukan membenaran atas node yang menjadi akar pada pohon keputusan
  - b. Mengevaluasi model C4.5 dengan mencari nilai akurasi AUC
  - c. Mengidentifikasi potensi kekurangan atau kelebihan model
7. *Deployment* (Penerapan)  
Mengimplementasikan model kedalam sistem rekrutmen

## 2.6 Implementasi dan Pengujian

Pengujian merupakan tahapan untuk memastikan apakah sebuah sistem yang sedang di kerjakan menemui kendala atau tidak. Maka dari itu proses pengujian haruslah dilalui agar suatu sistem dapat memenuhi standar dan minim akan error atau kesalahan. Untuk Pegujian dari penelitian ini, peneliti telah melakukan beberapa proses pengujian dari data yang telah disiapkan dalam format excel, berikut adalah beberapa langkah-langkah dari proses pengujian yang telah dilakukan:

### 2.6.1 Mencari Nilai *Entropy* dan *Gain*

Pencarian nilai *Entropy* bertujuan untuk mencari akar atau mencari atribut yang akan dijadikan akar utama. Sebelum mencari nilai *entropy* dan *gain* pastikan data yang akan di uji sudah siap. Lalu untuk mencari nilai *entropy* menggunakan cara manual dengan cara:

1. Menentukan akar

Akar akan ditentukan melalui atribut yang telah dipilih dengan cara menghitung nilai *gain* dari setiap atribut, nilai *gain* yang paling tinggi nantinya akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut perlu menghitung terlebih dahulu nilai *entropy*. Untuk menghitung nilai *entropy* memiliki dapat digunakan persamaan 1 [19]:

$$Entropy(N) = \sum_{i=1}^p - p_i * \log_2 p_i \quad (1)$$

Dimana: N merupakan Himpunan Kasus; P merupakan Jumlah Partisi; dan  $p_i$  merupakan Proporsi dari  $N_i$  terhadap N. Setelah menghitung *entropy* setiap kasus maka selanjutnya adalah menghitung *gain*.

2. Menghitung nilai *Gain*

Menggunakan persamaan 2:

$$Gain(N, A) = Entropy(N) - \sum_{i=1}^p \frac{|N_i|}{|N|} * Entropy(N_i) \quad (2)$$

Dimana N merupakan Himpunan Kasus; A merupakan Atribut;  $|N_i|$  merupakan Jumlah kasus pada partisi ke i; dan  $|N|$  merupakan Jumlah kasus dalam N.

3. Ulangi langkah 1 dan 2 sampai semua atribut dihitung dan memiliki informasi dan yang akan dijadikan sebagai node (akar).

- Langkah 1: Menghitung jumlah kasus, jumlah kasus untuk keputusan *Join*, jumlah kasus untuk keputusan Tidak *join*, dan *Entropy* dari semua kasus yang dibagi berdasarkan atribut Posisi, Penempatan, Usia, Jenis Kelamin, Pendidikan, Vaksin, Sim C, *Interview HR*, *Psikotest*, *Interview User*, *Background Check*, *Sign PKWT*. Setelah itu dilakukan penghitungan *Gain* untuk masing-masing atribut.
- Langkah 2: Menghitung *Entropy* pada baris total dengan persamaan (1). Selanjutnya dilakukan penghitungan *Gain* untuk masing-masing atribut dengan persamaan (2).

Tabel 1 adalah perhitungan nilai *entropy* dan *gain*.

Menghitung *Entropy* Total :

**Tabel 1.** Perhitungan *entropy* total

Jumlah Data	Join	Tidak Join	<i>Entropy</i>
700	570	130	0,692419486

Hasil perhitungan ditunjukkan pada tabel 2.

**Tabel 2.** Hasil Perhitungan *entropy* dan *gain* dari setiap atribut

Atribut/ Node	Value	Total	Join	Tidak Join	<i>Entropy</i>	Info <i>Gain</i>
Posisi	Admin	15	14	1	0,353359335	0,14706752
	CMO	351	337	14	0,241769442	
	Field Collector	319	214	105	0,914059277	
	Sales Team Leader	15	15	0	0,000000000	
Penempatan	Aceh	43	42	1	0,159350063	0,13263705
	Sumatera Barat	84	68	16	0,702466551	
	Sumatera Utara	78	21	0	0,000000000	

Atribut/ Node	Value	Total	Join	Tidak Join	Entropy	Info Gain
	Riau	95	60	35	0,949452015	
	Jambi	34	29	5	0,602430802	
	JABODETABEK	248	204	44	0,674398887	
	Jawa Barat	48	39	9	0,69621226	
	Jawa Tengah	13	13	0	0,000000000	
	Jawa Timur	34	31	3	0,430551867	
	Kalimantan Selatan	2	2	0	0,000000000	
	Sulawesi Utara	21	21	0	0,000000000	0,07673497
Usia	Under 20	27	9	18	0,918295834	
	20 - 25	271	218	53	0,712983112	
	25 - 30	282	251	31	0,499701999	
	30 - 35	103	86	17	0,646251622	
	35 - 40	17	16	1	0,322756959	0,03759616
Jenis Kelamin	Laki - Laki	558	453	105	0,697631815	
	Perempuan	142	127	15	0,486604062	0,30990245
Pendidikan	SMA Sederajat	259	140	119	0,995252549	
	D3	63	63	0	0,000000000	
	S1	375	374	1	0,026644044	
	S2	3	3	0	0,000000000	0,06282658
Vaksin	Sudah	624	504	120	0,706274089	
	Belum	76	76	0	0,000000000	0,14680571
Sim C	Punya	487	454	33	0,357515089	
	Tidak Punya	213	126	87	0,975679787	0,30870644
Interview HR	Lanjut	625	570	55	0,429758609	
	Gugur	75	0	75	0,000000000	0,45941843
Psikotest	Disarankan	598	570	28	0,272743714	
	Tidak Disarankan	102	0	102	0,000000000	0,51240221
Interview User	Diterima	590	570	20	0,213579815	
	Ditolak	110	0	110	0,000000000	0,65377803
Background Check	Lolos	573	570	3	0,047205972	
	Tidak Lolos	127	0	127	0,000000000	

Berdasarkan tabel 2 dalam hasil perhitungan entropy dan gain dapat dibuat pohon keputusan berdasarkan nilai gain terbesar. Dimana atribut yang nilainya paling besar adalah Background Check dengan nilai 0,65377803.

### 3. Hasil dan Pembahasan

Pada akhir penelitian akan mendapatkan beberapa hasil, yaitu sebagai berikut.

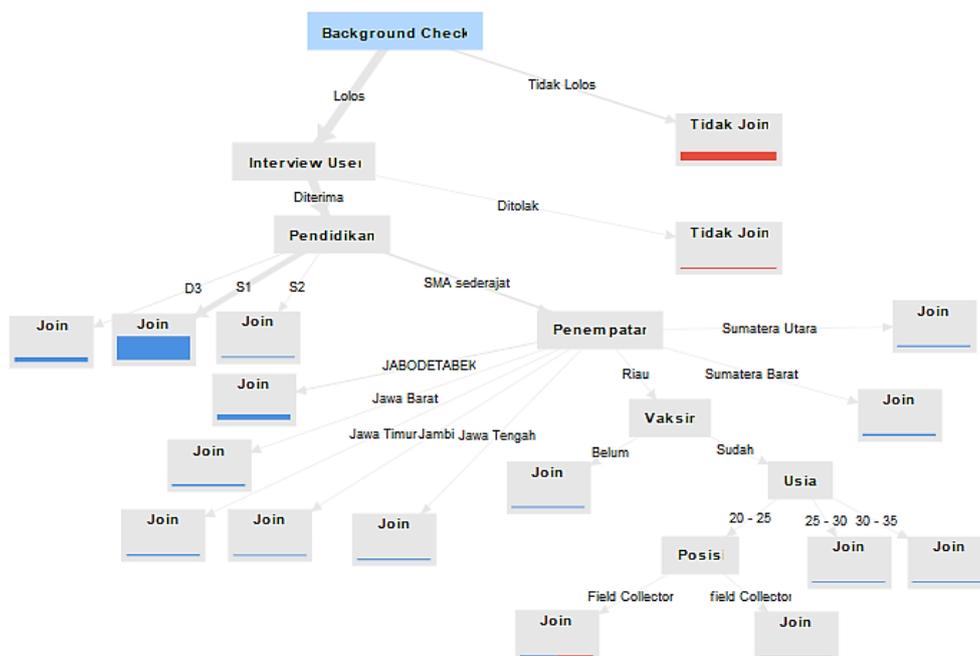
#### 3.1 Decision Tree / Pohon Keputusan

Pada gambar 3 telah berhasil mengklasifikasikan dengan menggunakan aplikasi *rapid miner* terbukti nilai perhitungan gain terbesar pada tabel 2 ditunjukkan atribut *Background Check* dengan nilai 0,65377803 yang menjadi *node* awal atau sebagai akar. Setelah dilakukannya perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir.

Jika dilihat berdasarkan hasil pohon keputusan prediksi bahwa atribut yang mempunyai pengaruh utama adalah *variable Background Check* yang menempati sebagai simpul akar. Agar lebih jelas, peneliti membuat model aturan berupa teks, seperti keterangan berikut:

1. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = D3, Maka Status *Join*
2. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = S1, Maka Status *Join*
3. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = S2, Maka Status *Join*
4. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = JABODETABEK, Maka Status *Join*
5. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Jawa Barat, Maka Status *Join*

6. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Jawa Timur, Maka Status *Join*
7. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Jambi, Maka Status *Join*
8. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Jawa Tengah, Maka Status *Join*
9. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Riau, Vaksin = Belum, Maka Status *Join*
10. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Riau, Vaksin = Sudah, Usia = 20 – 25, Posisi = Field Collector, Maka Status Tidak *Join*
11. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Riau, Vaksin = Sudah, Usia = 20 – 25, Posisi = Field Collector, Maka Status *Join*
12. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Riau, Vaksin = Sudah, Usia = 25 - 30, Maka Status *Join*
13. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Riau, Vaksin = Sudah, Usia = 30 - 35, Maka Status *Join*
14. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Sumatera Barat, Maka Status *Join*
15. Jika Nilai *Background Check* = Lolos, *Interview User* = Diterima, Pendidikan = SMA sederajat, Penempatan = Sumatera Utara, Maka Status *Join*
16. Jika Nilai *Background Check* = Lolos, *Interview User* = Ditolak, Maka Status Tidak *Join*
17. Jika Nilai *Background Check* = Tidak Lolos, Maka Status Tidak *Join*



Gambar 3. Desicion Tree

### 3.2 Evaluasi dan Akurasi

Akurasi dalam klasifikasi merupakan presentase ketepatan record data yang diklasifikasi secara benar setelah dilakukan pengujian data pada hasil klasifikasi. Dari pengujian yang telah dilakukan mendapatkan akurasi sebesar 99.43%. secara detail confusion matriks dapat dilihat pada tabel 3.

Tabel 3. Nilai Akurasi

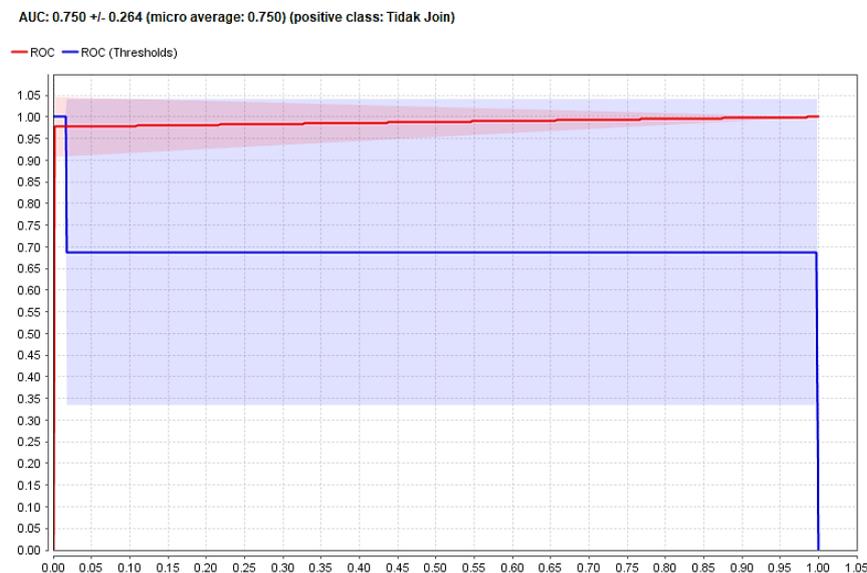
Accuracy : 99.43% +/- 1.38% (Micro Average : 99.43 %)			
	True Join	True Tidak Join	Class Precision
Pred. Join	569	3	99.48%
Pred Tidak Join	1	127	99.22%
Class Recall	99.82%	97.69%	

Berdasarkan hasil pengujian *rapidminer* diperoleh hasil dengan algoritma C4.5 dimana dari total 700 data diperoleh sebagai berikut:

1. Tingkat akurasi sebesar 99,43%. Untuk keseluruhan data Join.
2. Sebanyak 569 data diprediksi class Join ternyata sesuai, yaitu masuk kedalam class Join, sebanyak 1 data yang diprediksi class Tidak Join ternyata termasuk kedalam prediksi class Join.
3. Sebanyak 3 data yang diprediksi class Join ternyata masuk dalam class Tidak Join, kemudian 127 data di prediksi class Tidak Join sesuai yaitu termasuk kedalam prediksi class Tidak Join.
4. Nilai *Recall True Join* sebesar 99.82% sedangkan nilai *recall True Tidak join* sebesar 97.69%.
5. Nilai *Class Precision Pred. Join* yang diperoleh dari *confusion matrix* sebesar 99.48% dan nilai Nilai *Class Precision Pred. Tidak Join* yang diperoleh dari *confusion matrix* sebesar 99.22%.

### 3.3 Nilai AUC

Untuk nilai AUC dari penelitian ini adalah 0.750. Dengan nilai AUC sebesar 0,750. model C4.5 yang digunakan dalam penelitian ini memiliki kemampuan yang baik untuk membedakan antara calon karyawan yang sesuai dan tidak sesuai. Nilai 0,750 berarti bahwa model dapat membedakan dengan baik antara calon karyawan yang benar-benar memenuhi kriteria (kelas positif) dan yang tidak memenuhi kriteria (kelas negatif), dengan tingkat akurasi yang cukup baik. Evaluasi AUC merupakan salahsatu metode evaluasi yang populer digunakan dalam klasifikasi dengan kasus data tidak seimbang [20].



Gambar 4. Nilai AUC

## 4. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode C4.5 sangat efektif dalam klasifikasi dan seleksi calon karyawan baru di PT. Kobus Smart Service. Dengan tingkat *accuracy* pengujian sebesar 99.43%, *presisi* 99,29%, *recall* 97,69%, dan *AUC* 0,750. metode ini dapat diandalkan untuk membantu dalam proses pengambilan keputusan. Untuk Perhitungan *Entropy* dan *Gain* dapat disimpulkan untuk nilai gain terbesar berada pada Background Checking dengan nilai nilai 0,65377803 yang menjadi node awal atau atribut yang dijadikan sebagai akar. Tingginya akurasi menunjukkan bahwa model yang dibangun memiliki kemampuan yang sangat baik dalam mengklasifikasikan data calon karyawan berdasarkan kriteria yang telah ditentukan. Ini mengindikasikan bahwa model tersebut dapat meminimalkan kesalahan dalam proses seleksi karyawan. Penggunaan metode C4.5 dalam studi kasus ini dapat diterapkan secara praktis di perusahaan lain yang memiliki kebutuhan serupa, yakni untuk meningkatkan efisiensi dan akurasi dalam proses seleksi karyawan.

## REFERENSI

- [1] J. Setiyono dan S. Sutrimah, "Analisis Teks dan Konteks Pada Iklan Operator Seluler (XL dengan Kartu AS)," *Pedagog. J. Pendidik.*, vol. 5, no. 2, hal. 297–310, 2016, doi: 10.21070/pedagogia.v5i2.263.
- [2] N. Nuryanta, "Pengelolaan Sumber Daya Manusia (Tinjauan Aspek Rekrutmen dan Seleksi)," *el-Tarbawi*, vol. 1, no. 1, hal. 55–69, 2008, doi: 10.20885/tarbawi.vol1.iss1.art5.
- [3] Suwinardi, "Profesionalisme Dalam Bekerja," *Orbith*, vol. 13, no. 2, hal. 81–85, 2017.
- [4] J. Rahayu, E. A., & Susilowibowo, "Pengaruh Manajemen Modal Kerja Pada Industri Manufaktur yang Terdaftar di Bursa Efek Indonesia (BEI) Tahun 2011.," *J. Ilmu Manaj.*, vol. 2, no. 4, hal. 1444–1455, 2013.

- 
- [5] N. Aprilyani, I. Zulfa, dan H. Syahputra, "Penerapan Algoritma Decision Tree C4.5 Untuk Model Penentuan Penerima Beasiswa Program Indonesia Pintar (Pip) Studi Kasus Sma Negeri 3 Timang Gajah," *J. Tek. Inform. dan Elektro*, vol. 5, no. 1, hal. 96–109, 2022, doi: 10.55542/jurtie.v5i1.452.
- [6] I. Romli dan A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, hal. 694–702, 2020.
- [7] C. Nas, "Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5," *J. Manaj. Inform.*, vol. 11, no. 2, hal. 131–145, 2021, doi: 10.34010/jamika.v11i2.5506.
- [8] Y. A. Fiandra, S. Defit, dan Y. Yuhandri, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, hal. 82–89, 2017, doi: 10.29207/resti.v1i2.48.
- [9] R. Wajhillah dan I. Yulianti, "Penerapan Algoritma C4.5 Untuk Prediksi Penggunaan Jenis Kontrasepsi Berbasis Web," *Klik - Kumpul. J. Ilmu Komput.*, vol. 4, no. 2, hal. 160, 2017, doi: 10.20527/klik.v4i2.98.
- [10] Y. H. Agustin, . K., dan E. T. Luthfi, "Klasifikasi Penerimaan Mahasiswa Baru Menggunakan Algoritma C4.5 Dan Adaboost (Studi Kasus : STMIK XYZ)," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 9, no. 1, hal. 1, 2017, doi: 10.22303/csrid.9.1.2017.1-11.
- [11] E. Elisa, "Prediksi Profit Pada Perusahaan Dengan Klasifikasi Algoritma C4.5," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, hal. 179, 2018, doi: 10.20527/klik.v5i2.153.
- [12] S. Mita, Y. Yamazoe, T. Kamataki, dan R. Kato, *Metabolic activation of a tryptophan pyrolysis product, 3-amino-1-methyl-5H-pyrido[4,3-b]indole(Trp-P-2) by isolated rat liver nuclei*, vol. 14, no. 3. 1981. doi: 10.1016/0304-3835(81)90152-X.
- [13] T. Turap, T. B. Merupakan, T. B. Lebih, dan T. D. Turap, *Covariance Structure Analysis of Health-Related Indicators in Home-Dwelling Elderly Focused on Subjective Health Perceptions*.
- [14] R. Goebel, J. Siekmann, dan W. Wahlster, *Learning Classifier Systems*. 2002.
- [15] M. Rivki, A. M. Bachtiar, T. Informatika, F. Teknik, dan U. K. Indonesia, *Covariance Structure Analysis of Health-Related Indicators in Home-Dwelling Elderly Focused on Subjective Health Perceptions*, no. 112.
- [16] M. Hofmann dan R. Klinkenberg, *Data Mining and Knowledge Discovery Series*. 2014.
- [17] P. Chapman *et al.*, "Crisp-Dm," *SPSS inc*, vol. 78, hal. 1–78, 2000.
- [18] Mariscal, Gonzalo, Oscar Marban, and Covadonga Fernandez. 2010. "A Survey of Data Mining and Knowledge Discovery Process Models and Methodologies." *The Knowledge Engineering Review*, Cambridge University Press 25:2(2010): 137–66
- [19] Hermanto B, Sn A, Putra FP. 2017. Analisis Kinerja Decision Tree C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur. *Jurnal Inovasi dan Teknologi (Inovtek) Polbeng Seri Informatika*. 2 (2): 189–197.
- [20] Saifudin, A., & Wahono, R. S. (2015). Penerapan teknik ensemble untuk menangani ketidakseimbangan kelas pada prediksi cacat software. *IlmuKomputer. com Journal of Software Engineering*, 1(1), 28-37