

Implementasi Algoritma C4.5 Untuk Klasifikasi Tingkat Kenaikan Jabatan Pada PT Siprama Cakrawala Depok

Rahayu Ningsih^{1*}, Muhammad Devan Basunanda², Wati Erawati³

^{1,2,3}Universitas Bina Sarana Informatika
e-mail: ¹rahayu.ryh@bsi.ac.id, ²basunandadevan@gmail.com, ³wati.wti@bsi.ac.id

Diterima 03-11-2025	Direvisi 27-05-2025	Disetujui 14-06-2025
------------------------	------------------------	-------------------------

Abstrak - Dalam era modern, kenaikan jabatan menjadi tantangan bagi perusahaan karena proses promosi yang sering kali dianggap subjektif. Penelitian ini bertujuan untuk membangun model klasifikasi kenaikan jabatan menggunakan Algoritma C4.5 guna meningkatkan objektivitas keputusan promosi karyawan. Data dikumpulkan dari database karyawan PT Siprama Cakrawala serta melalui kuesioner kepada manajer, supervisor, dan karyawan. Variabel yang digunakan mencakup kinerja, pengalaman kerja, tingkat pendidikan, sikap kerja, dan keterlibatan dalam proyek. Data dianalisis menggunakan RapidMiner dengan teknik Decision Tree, kemudian divalidasi menggunakan metode Cross Validation untuk mengukur akurasi model. Hasil penelitian menunjukkan bahwa model yang dihasilkan memiliki tingkat akurasi sebesar 88.91%, dengan classification error sebesar 11.09%. Variabel pengalaman kerja memiliki pengaruh paling dominan dengan nilai information gain tertinggi dan menjadi root node dalam pohon keputusan. Precision model menunjukkan bahwa algoritma C4.5 cukup andal dalam mengklasifikasikan karyawan yang layak dan belum layak mendapatkan promosi jabatan, dengan nilai precision kelayakan sebesar 91.11% dan precision belum layak sebesar 87.50%. Dengan hasil ini, model yang dikembangkan dapat digunakan sebagai alat bantu dalam proses pengambilan keputusan promosi karyawan.

Kata Kunci: *Algoritma C4.5, Data Mining, Klasifikasi Karyawan*

Abstract - In the modern era, promotions have become a challenge for companies due to the often subjective nature of the promotion process. This research aims to build a promotion classification model using the C4.5 Algorithm to enhance the objectivity of employee promotion decisions. Data were collected from the employee database of PT Siprama Cakrawala as well as through questionnaires to managers, supervisors, and employees. The variables used include performance, work experience, education level, work attitude, and project involvement. Data were analyzed using RapidMiner with the Decision Tree technique, then validated using the Cross Validation method to measure the model's accuracy. The research results show that the generated model has an accuracy rate of 88.91%, with a classification error of 11.09%. The work experience variable has the most dominant influence with the highest information gain value and becomes the root node in the decision tree. The model's precision indicates that the C4.5 algorithm is quite reliable in classifying employees who are eligible and not eligible for promotion, with an eligibility precision value of 91.11% and a non-eligibility precision of 87.50%. With these results, the developed model can be used as a decision-making aid in the employee promotion process.

Keywords: C4.5 Algorithm, Data Mining, Employee Classification

PENDAHULUAN

Pada era modern saat ini, kenaikan jabatan dalam perusahaan sering menjadi salah satu isu yang dihadapi organisasi. Berdasarkan survey yang dilakukan oleh *Society For Human Resource Management (SHRM)* pada tahun 2023, sebanyak 62% karyawan merasa bahwa proses promosi di tempat kerja masih bersifat subjektif, terutama karena kurangnya kriteria yang jelas dan transparansi dalam penilaian (SHRM, 2023). PT Siprama Cakrawala adalah sebuah perusahaan yang bergerak di bidang *Outsourcing* dan beralokasi pada Kota Depok, menghadapi tantangan serupa dalam menentukan

kriteria kenaikan jabatan bagi karyawan mereka. Penelitian ini akan melibatkan berbagai pihak pada perusahaan, termasuk departemen Sumber Daya Manusia, Manager Divisi, dan para karyawan.

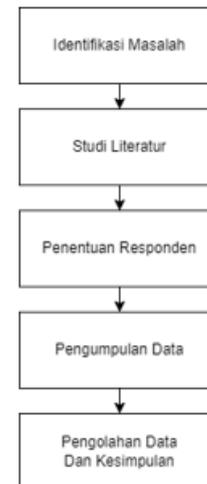
Penelitian ini juga di dukung oleh penelitian yang sudah ada sebelumnya, yaitu : Bisnis simpan pinjam (Nugroho & Amrullah, 2023), sering kali kurang mempertimbangkan kemampuan debitur, sehingga penelitian ini mengimplementasikan algoritma K-NN dengan K-Fold Cross Validation untuk memprediksi kelayakan pinjaman, menghasilkan akurasi 93% setelah pembobotan atribut menggunakan metode Rank Order Centroid (ROC). Penguasaan bahasa Inggris di

masyarakat masih rendah akibat metode pembelajaran konvensional dan kurangnya praktik langsung. (Fanny Irnanda & Perdana Windarto, n.d.) Penelitian ini mengatasi masalah tersebut dengan menerapkan algoritma C4.5 berbasis data mining untuk menganalisis faktor dominan seperti praktik (C2), menggunakan RapidMiner sebagai tools analisis, dan mencapai akurasi 95% dalam mengidentifikasi atribut paling berpengaruh untuk meningkatkan kecakapan berbahasa Inggris. Proses rekrutmen karyawan sering kali bersifat subjektif dan tidak konsisten (Devia, 2023), sehingga perusahaan kesulitan mendapatkan kandidat yang sesuai kriteria. (Mayangsari & Wahyudi, 2024) Penelitian ini mengatasi masalah tersebut dengan menerapkan algoritma C4.5 berbasis data mining untuk menganalisis atribut seperti background check dan pendidikan, menggunakan RapidMiner sebagai alat analisis, dan mencapai akurasi 99,43% dalam memprediksi kelayakan kandidat. Proses penempatan jabatan fungsional ASN sering kali bersifat subjektif dan tidak terstandarisasi, sehingga berpotensi menimbulkan ketidaksesuaian kompetensi. (Isra, 2022) Penelitian ini mengatasi masalah tersebut dengan menerapkan algoritma C4.5 berbasis pohon keputusan untuk menganalisis faktor kunci seperti pendidikan dan pengalaman, menggunakan RapidMiner sebagai alat analisis, dan mencapai akurasi 92,54% dalam memprediksi kelayakan penempatan jabatan fungsional. Karyawan PT WS tidak puas dengan penilaian kinerja manual dan subjektif, sehingga penelitian ini mengusulkan sistem berbasis web dengan Algoritma C4.5 untuk klasifikasi dan Metode 360° melibatkan multi-penilai, yang berhasil diuji dengan tingkat penerimaan pengguna sebesar 83,4% (Rochmat Asyhari et al., 2024). PT. Alfa Scorpii menghadapi kesulitan dalam memprediksi preferensi konsumen terhadap sepeda motor Yamaha karena kurangnya sistem analisis data yang efektif. Untuk mengatasi hal ini, penelitian menerapkan Algoritma C4.5 guna membangun pohon keputusan berdasarkan perhitungan *entropy* dan *gain*, serta mengembangkan sistem berbasis web yang mampu mengklasifikasikan preferensi konsumen secara akurat dan cepat (Widodo et al., 2022). omosi jabatan yang kompleks dan subjektif membutuhkan pendekatan berbasis data, sehingga penelitian ini membandingkan algoritma Decision Tree C4.5 dan C5.0 menggunakan dataset 54.808 data melalui RapidMiner dan Python, dengan hasil menunjukkan C4.5 lebih akurat (92,54%) dalam mengidentifikasi karyawan berpotensi berdasarkan kriteria seperti pencapaian KPI dan penghargaan (Abidin, Zaenal; Nurhana, Eka; Permata, Permata; Ulum, 2023). Dinas Perhubungan Sulawesi Tengah kesulitan menyeleksi karyawan sesuai kepribadian, sehingga penelitian ini mengimplementasikan algoritma C4.5 untuk membangun sistem prediksi kepribadian berbasis pohon keputusan dengan akurasi 79.167%,

menggunakan atribut seperti umur dan jabatan (Oktafiqurahman et al., 2023). BKAD Kabupaten Banyumas mengalami kesulitan dalam mengelola KIB peralatan dan mesin kantor yang menghambat proses pengadaan, sehingga penelitian ini menerapkan algoritma C4.5 melalui RapidMiner untuk membangun sistem prediksi berbasis pohon keputusan dengan akurasi sempurna 100% berdasarkan kriteria tahun pembelian, kondisi, dan bahan barang (Iftitah & Setyadi, 2023). Penelitian ini mengatasi keterlambatan rekomendasi kenaikan pangkat PNS akibat sistem manual dengan mengembangkan sistem berbasis web menggunakan algoritma Naïve Bayes dan PHP yang menghasilkan rekomendasi otomatis dengan akurasi 84,24%, precision 81,34%, dan recall 73,22%, yang telah diuji melalui confusion matrix dan kuisioner kepuasan pengguna (Angreni & Susanti, 2024).

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan data mining, khususnya Algoritma C4.5 untuk membangun model klasifikasi kenaikan jabatan karyawan. Berikut adalah langkah langkah yang dilakukan dalam penelitian ini dapat dilihat di Gambar 1:



Gambar 1. Tahapan Penelitian

Data yang digunakan dalam penelitian ini berasal dari database karyawan PT Siprama Cakrawala serta pengamatan langsung pada lingkungan kerja dan interaksi antar karyawan demi mendapatkan pemahaman yang lebih mendalam mengenai faktor-faktor yang memengaruhi keputusan kenaikan jabatan. Pengumpulan data juga akan dilakukan melalui kuesioner kepada manager, supervisor, dan para karyawan untuk melengkapi informasi yang mungkin tidak tersedia dalam database. Data yang akan dikumpulkan meliputi:

1. Kinerja karyawan (penilaian hasil kerja).
2. Lama pengalaman kerja (dalam tahun).
3. Tingkat pendidikan.

4. Penilaian sikap dan attitude dalam bekerja (berdasarkan supervisor).
5. Keterlibatan dalam proyek perusahaan.

HASIL DAN PEMBAHASAN

A. Hasil

Data yang diperoleh dalam penelitian ini merupakan data kuesioner karyawan yang diperoleh dari PT Siprama Cakrawala, data tersebut berisi kode responden, nama responden, jabatan, serta 19 pertanyaan yang terdiri dari Kinerja(x1.1 – x1.4), Pengalaman(x2.1 – x2.3), Pendidikan(x3.1 – x3.3), Sikap dan Attitude(x4. 1 – x4.5), dan Keterlibatan Proyek(x5.1 – x5.4) menggunakan *skala likert* sebesar 1-5 poin. Penelitian akan dibantu oleh 4 manager dan 1 supervisor untuk menentukan kriteria Rekomendasi(Y) dan validasi data responden seperti pengisian kuesioner dengan nilai yang asal-asalan dan banyak yang tidak di isi. Data yang dihasilkan dari para karyawan berjumlah 129 data valid dan 21 data yang tidak valid, dan penelitian menggunakan sebanyak 20 *Random Data Sample* dari 129 data tersebut untuk dijadikan bahan perhitungan

B. MENENTUKAN NILAI ATRIBUT

Dalam data sampel yang telah didapat hal berikutnya yang dalam penelitiannya di lakukan adalah menentukan *Root Node* atau atribut utama demi mendapatkan atribut yang paling banyak atau memiliki kemampuan terbaik untuk membagi data menjadi kelompok yang lebih homogen, penelitian ini menggunakan nilai *gain* yang paling besar. Kemudian akan dilakukan perhitungan nilai *entropy* dan *gain* masing masing atribut. Dengan cara ini pohon keputusan yang dihasilkan akan lebih efisien dan akurat dalam mengklasifikasikan data. Dalam hal ini kita perlu menghitung *entropy* total dari data yang telah didapat, berikut adalah cara perhitungannya:

Jumlah Kasus : 20

Layak (S1) : 11

Belum (S2) : 9

Rumus Entropy:

$$E(S) = - \sum_{i=1}^k p_i \log_2(p_i)$$

$S = 20, S1 = 11, S2 = 9$

$p(\text{Layak}) = S1/S = 11/20 = 0.55$

$p(\text{Belum}) = S2/S = 9/20 = 0.45$

$$E(S) = -(0.55 \cdot \log_2(0.55) + 0.45 \cdot \log_2(0.45))$$

$$E(S) = -(0.55 \cdot (-0.8625) + 0.45 \cdot (-1.1520)) = 0.992774454$$

Setelah itu hitung entropy untuk tiap atribut dengan cara:

X1 = Kinerja

Sub kelompok yang terdapat dalam dataset:

Baik ($S = 3, S1 = 1, S2 = 2$)

$$E(\text{Baik}) = -(1/3 \cdot \log_2(1/3) + 2/3 \cdot \log_2(2/3))$$

$$E(\text{Baik}) = -(0.3333 \cdot \log_2(0.3333) + 0.6667 \cdot \log_2(0.6667))$$

$$\log_2(0.3333) = -1.585 \quad |$$

$$\log_2(0.6667) = -0.585$$

$$E(\text{Baik}) = -(0.3333 \cdot (-1.585) + 0.6667 \cdot (-0.585))$$

$$= \mathbf{0.918295834} \quad | \text{ rounding } E(\text{Baik}) = 0.918$$

$$\text{Sangat Baik } (S = 17, S1 = 10, S2 = 7)$$

$$E(\text{Sangat Baik}) = -(10/17 \cdot \log_2(10/17) + 7/17 \cdot \log_2(7/17))$$

$$E(\text{Sangat Baik}) = -(0.5882 \cdot \log_2(0.5882) + 0.4118 \cdot \log_2(0.4118))$$

$$\log_2(0.5882) = -0.767 \quad |$$

$$\log_2(0.4118) = -1.274$$

$$E(\text{Baik}) = -(0.3333 \cdot (-0.767) + 0.6667 \cdot (-1.274))$$

$$= \mathbf{0.977417818} \quad | \text{ rounding } E(\text{Sangat Baik}) = 0.9774$$

Setelah itu hitung *Weighted Entropy* dan *Gain* nya, berikut formula yang penelitian gunakan:

$$E(A) = \sum \left(\frac{|S_i|}{|S|} |S_i| \right) * E(S_i)$$

S = Dataset Awal | A = Atribut | S_i = Subset Data dari nilai A

$|S_i|$ = Jumlah Kasus dalam subset S_i

$|S|$ = Total Jumlah Kasus

$E(S_i)$ = Entropy subset S_i

$$\text{Untuk kategori Baik: } \frac{3}{20} \cdot 0,918295834 = \mathbf{0.137744375}$$

$$\text{Untuk kategori Sangat Baik: } \frac{17}{20} \cdot 0.977417818 = \mathbf{0,831805565}$$

$$\text{Weighted Entropy: } E(A) = 0.137744375 + 0,831805565 = \mathbf{0,96954994}$$

$$\text{Gain} = E(S) - E(A)$$

Dari perhitungan diatas penelitian mendapatkan hasil *entropy* total dengan nilai 0.992774454, *weighted entropy* Kinerja(X1) dengan nilai 0,96954994 dan *Gain* Kinerja(X1) dengan nilai 0,023224514. Proses serupa juga diterapkan untuk atribut Pengalaman(X2), Pendidikan(X3), Sikap(X4), serta Keterlibatan Proyek(X5). Penelitian ini juga membuat tabel perhitungan *entropy* dari tiap atribut dan sub kelompok dari dataset yang telah didapat. Berikut adalah format serta hasil perhitungan *entropy* dan *gain* dari tiap atribut:

Tabel 2 format serta hasil perhitungan *entropy* dan *gain* dari tiap atribut

Klasifikasi Penilaian	Jumlah Kasus (S)	Layak (S1)	Belum (S2)	ENTROPY	GAIN
Total	20	11	9	0,992774454	

Kinerja					0,0232 24514
Baik	3	1	2	0,9182 95834	
Sangat Baik	17	10	7	0,9774 17818	
Pengalaman					0,7500 36805
Kurang	6	0	6	0	
Baik	5	2	3	0,9709 50594	
Sangat Baik	9	9	0	0	
Pendidikan					0,0599 30706
Cukup	1	0	1	0	
Klasifikasi Penilaian	Jumlah Kasus(S)	Laya k(S1)	Belu m(S2)	ENTR OPY	GAIN
Sangat Baik	19	11	8	0,9819 40787	
Sikap					0,4224 16175
Kurang	1	0	1	0	
Cukup	1	1	0	0	
Baik	9	2	7	0,7642 04507	
Sangat Baik	9	8	1	0,5032 58335	
Keterlibatan Proyek					0,3345 74283
Kurang	14	5	9	0,9402 85959	
Cukup	1	1	0	0	
Baik	5	5	0	0	

Dari tabel 2 dapat dilihat nilai *gain* yang tertinggi ada pada atribut Pengalaman dengan nilai **0,750036805** maka Pengalaman akan menjadi akar pertama atau *root node*, kemudian akan dilanjutkan perhitungan *node* ke 2 dengan perhitungan pada tabel dibawah ini.

Klasifikasi Penilaian	Jumlah Kasus(S)	Laya k(S1)	Belu m(S2)	ENTR OPY	GAIN
Total	5	2	3	0,9709 50594	
Kinerja					0
Sangat Baik	5	2	3	0,9709 50594	
Pendidikan					0
Sangat Baik	5	2	3	0,9709 50594	

Klasifikasi Penilaian	Jumlah Kasus(S)	Laya k(S1)	Belu m(S2)	ENTR OPY	GAIN
Sikap					0,9709 50594
Baik	3	0	3	0	
Sangat Baik	2	2	0	0	
Keterlibatan Proyek					0,9709 50594
Kurang	3	0	3	0	
Baik	2	2	0	0	

Tabel 3. Hasil perhitungan entropy dan gain dari tiap atribut node 2

Pada tabel 3 merupakan perhitungan terakhir dari Algoritma C4.5 karena dapat dilihat bahwa node 2, meskipun terdapat beberapa atribut yang sudah dilakukan pemisahan, hasil pemisahan tidak memperbaiki keadaan atau menurunkan entropy lebih lanjut, yang menandakan tidak ada pemisahan yang signifikan lagi. Dengan ini klasifikasi akar terakhir dari pohon keputusan yaitu atribut Pendidikan dan Sikap yang dapat dilihat dari tabel 2.

B. Pemodelan dengan Algoritma C4.5

Setelah data siap, Algoritma C4.5 diterapkan untuk membangun model klasifikasi. Penelitian menggunakan software RapidMiner untuk membangun pohon keputusan berdasarkan atribut-atribut yang memengaruhi tingkat promosi jabatan. Langkah-langkah yang dilakukan oleh penelitian adalah sebagai berikut :

1. Importing Data

Pada tahap ini akan dilakukan pemrosesan data decision tree dengan menggunakan RapidMine Studio 10.5.0 dengan metode decision tree. Data yang diolah menggunakan MS. Excel 2019, sehingga di aplikasi RapidMiner Studio 10.5.0 menggunakan operator read excel yang berfungsi untuk membaca file yang akan diolah. Kemudian masukan operator Decision Tree untuk membuat data yang diolah menghasilkan decision tree. Dapat dilihat di Gambar 1.



Gambar 1. User Interface Work Area dari RapidMiner Studio 10.5.0

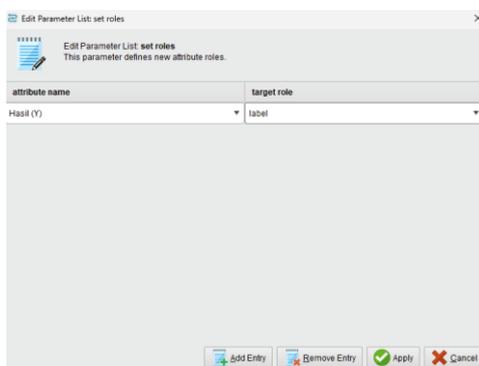
Selanjutnya kita perlu ke parameter Read Excel untuk menjalankan fungsi Import Configuration Wizard. Pilih lokasi file dan setelah data terupload di aplikasi RapidMiner Studio 10.5.0, berikut adalah tampilan *example set* yang merupakan *dataset* yang sudah untuk di proses dengan aplikasi RapidMiner 10.5.0. Dapat dilihat di Gambar 2.

Kinerja (X1)	Pengalaman (X2)	Pendidikan (X3)	Sikap (X4)	Kematangan Pro.	Hasil (Y)
1 Baik	Kurang	Sangat Baik	Sangat Baik	Kurang	Belum Layak
2 Sangat Baik	Baik	Sangat Baik	Baik	Kurang	Belum Layak
3 Sangat Baik	Kurang	Sangat Baik	Baik	Kurang	Belum Layak
4 Sangat Baik	Sangat Baik	Sangat Baik	Baik	Layak	
5 Baik	Kurang	Sangat Baik	Kurang	Kurang	Belum Layak
6 Sangat Baik	Kurang	Sangat Baik	Baik	Kurang	Belum Layak
7 Sangat Baik	Kurang	Cukup	Baik	Kurang	Belum Layak
8 Sangat Baik	Sangat Baik	Sangat Baik	Baik	Kurang	Belum Layak
9 Sangat Baik	Baik	Sangat Baik	Sangat Baik	Baik	Layak
10 Sangat Baik	Baik	Sangat Baik	Baik	Kurang	Belum Layak
11 Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Layak	
12 Sangat Baik	Baik	Sangat Baik	Baik	Kurang	Belum Layak
13 Sangat Baik	Sangat Baik	Sangat Baik	Cukup	Layak	
14 Sangat Baik	Kurang	Sangat Baik	Baik	Kurang	Belum Layak
15 Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik	Layak
16 Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik	Layak
17 Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Kurang	Layak
18 Baik	Sangat Baik	Sangat Baik	Sangat Baik	Layak	
19 Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Kurang	Layak

Gambar 2. User Interface Example Set dari RapidMiner Studio 10.5.0

2. Set Role

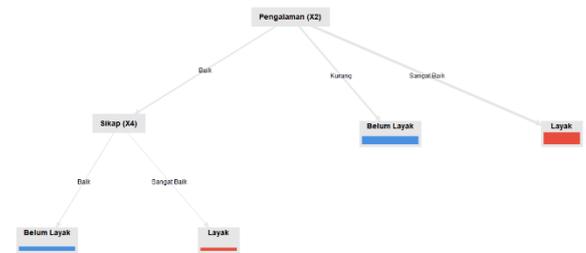
Langkah berikutnya yang akan dilakukan yaitu menambahkan fungsi *set roles* untuk mendefinisikan peran atau *role* setiap atribut dalam dataset. Dengan menentukan role, RapidMiner dapat memahami bagaimana menggunakan atribut tersebut dalam proses pemodelan dan analisis. Dalam kasus ini penelitian menggunakan *role* Label pada atribut Hasil(Y) untuk menjadi target atau variabel yang akan diprediksi seperti layak atau belum layaknya seseorang untuk mendapat promosi jabatan. Berikut adalah tampilan dari *set roles* dari aplikasi RapidMiner Studio 10.5.0. Dapat dilihat di Gambar 3.



Gambar 3. User Interface set roles dari RapidMiner Studio 10.5.0

3. Decision Tree

Langkah berikutnya yang akan dilakukan yaitu melihat hasil pengujian, dimana nantinya proses akan dimulai dari koneksi *dataset*, operator RapidMiner dan *decision tree* yang dihasilkan. Untuk itu perlu penambahan fungsi *decision tree* kedalam *Work Area RapidMiner*. Berikut adalah

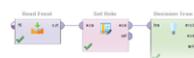


tampilan *Work Area* yang telah ditambahkan fungsi *decision tree*. Dapat dilihat di Gambar 4.

Gambar 4. Work Area yang telah ditambahkan fungsi *decision tree*

4. Graph View

Proses ini dilanjutkan dengan menjalankan atau *run* dan hasil pohon keputusan yang dibentuk akan langsung terlihat dalam bentuk *graph view decision tree* maupun *depenelitian* atau *description view*, hasil dari pohon keputusan ini akan sama dengan hasil perhitungan manual yang telah dilakukan diatas serta bisa dibuat lebih simpel dan mudah dipahami dengan fungsi *pre-pruning*. Namun dalam kasus ini hal tersebut tidak diperlukan karena hasil akan sangat jelas. Dapat dilihat pada Gambar 5.



Gambar 5. Graph View yang telah ditambahkan fungsi *decision tree*

Berdasarkan graph view dari decision tree diatas, dapat disimpulkan bahwa atribut yang menentukan kelayakan seseorang adalah Pengalaman(X3). Jika seseorang memiliki Pengalaman yang “Sangat Baik”, maka kemungkinan besar hasil akhirnya adalah “Layak” untuk promosi jabatan. Jika Pengalaman hanya “Baik”, faktor Sikap akan menjadi penentu proses kenaikan jabatan dan otomatis belum layak jika faktor Pengalaman “Kurang” persis seperti perhitungan manual yang sebelumnya penelitian lakukan.

C. Persiapan dan Evaluasi Model
A. Persiapan Data Uji

Pada tahap ini, data yang akan digunakan untuk pengujian model adalah 109 data para responden yang sudah dipisahkan dari data latih untuk menghindari bias dalam evaluasi performa model. Data uji dipersiapkan dengan memastikan formatnya sesuai dengan kebutuhan algoritma, terdiri dari fitur-fitur prediktor dan label target. Dapat dilihat pada Gambar 6.

Row No.	Hasil (Y)	Kinerja (X1)	Pengalaman...	Pendidikan...	Sikap (X4)	Keterlibatan...
1	Layak	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik
2	Layak	Baik	Sangat Baik	Sangat Baik	Sangat Baik	Baik
3	Belum Layak	Sangat Baik	Baik	Kurang	Sangat Baik	Kurang
4	Layak	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Kurang
5	Belum Layak	Sangat Baik	Baik	Sangat Baik	Kurang	Kurang
6	Layak	Sangat Baik	Sangat Baik	Sangat Baik	Baik	Kurang
7	Belum Layak	Sangat Baik	Kurang	Sangat Baik	Baik	Baik
8	Layak	Sangat Baik	Cukup	Sangat Baik	Sangat Baik	Baik
9	Layak	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Kurang
10	Layak	Sangat Baik	Kurang	Sangat Baik	Baik	Sangat Baik
11	Layak	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik	Kurang
12	Layak	Baik	Cukup	Sangat Baik	Sangat Baik	Sangat Baik
13	Layak	Sangat Baik	Baik	Sangat Baik	Sangat Baik	Baik
14	Belum Layak	Sangat Baik	Kurang	Sangat Baik	Baik	Kurang
15	Layak	Sangat Baik	Cukup	Sangat Baik	Sangat Baik	Cukup

Gambar 6. Tampilan didalam operator *Cross Validation*

Pada gambar diatas kita bisa melihat tampilan dari operator *Cross Validation*, yang bertujuan untuk membagi data menjadi 10 bagian atau *10 folds* agar dapat dievaluasi secara lebih akurat menggunakan *10-Fold Cross Validation*. Dalam operator ini, proses terbagi menjadi dua bagian yaitu *Training* dan *Testing*. Pada bagian *Training*, penelitian menggunakan operator *decision tree* untuk membangun model klasifikasi, sedangkan bagian *Testing*, operator *Apply Model* dan *Performance* ditambahkan untuk mengevaluasi performa *decision tree*. Untuk visualisasi *work area* yang telah ditambahkan operator *Cross Validation* bisa dilihat pada gambar 7



Gambar 7. Tampilan *Work Area* dengan *Cross Validation*

B. Uji dan Evaluasi Model

Langkah selanjutnya adalah menerapkan model yang telah dilatih pada data uji menggunakan operator *Apply Model* di RapidMiner. Dapat dilihat pada Gambar 8.



Gambar 8 Tampilan *Example Set Apply Model*

Setelah model diterapkan pada data uji, langkah berikutnya adalah mengevaluasi kinerja model menggunakan operator *Performance (Classification)*.

Operator ini menghasilkan metrik evaluasi seperti akurasi (*Accuracy*), presisi (*Precision*), dan *recall*. Selain itu, confusion matrix akan digunakan untuk memeriksa kesalahan klasifikasi yang terjadi seperti jumlah prediksi yang benar dan salah (*True Positive, True Negative, False Positive, False Negative*). Metrik-metrik ini memberikan gambaran yang lebih lengkap mengenai kinerja model dalam mengklasifikasikan data uji. Berikut adalah tampilan tabel *Performance Vector* dari model yang telah dibuat.

	true Layak	true Belum Layak	class precision
pred. Layak	41	4	91.11%
pred. Belum Layak	8	56	87.50%
class recall	83.67%	93.33%	

Tabel 4 tampilan tabel *Performance Vector* Dari data tabel 4, penelitian ini akan mencari nilai akurasi dan error klasifikasi (*Classification Error*) dengan formula ini.

$$\text{Akurasi} = \frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah total sampel}} = \frac{TP+TN}{TP+TN+FP+FN}$$

True Positive (TP) = 41 (prediksi “Layak” yang benar)

True Negative (TN) = 56 (prediksi “Belum Layak” yang benar)

False Positive (FP) = 4 (prediksi “Layak” yang salah)

False Negative (FN) = 8 (prediksi “Belum Layak” yang salah)

Maka:

$$\text{Akurasi} = \frac{(41+56)}{(41+4+8+56)} = \frac{97}{109} \times 100\% = 88.91\%$$

$$\text{Error} = \frac{12}{109} \times 100\% = 11.09\%$$

Maka penelitian mendapat hasil akurasi keseluruhan model dengan nilai 88,91% dan hasil error klasifikasi dengan nilai 11.09%. Dan berikut cara untuk menghitung presisi (*Precision*) dan recal dari data tabel diatas.

$$\text{Precision(Layak)} = \frac{TP(\text{Layak})}{TP(\text{Layak})+FP(\text{Layak})} = \frac{41}{41+4} = \frac{41}{45} = 0.9111(91.11\%)$$

$$\text{Precision(Belum)} = \frac{TP(\text{Belum})}{TP(\text{Belum})+FP(\text{Belum})} = \frac{56}{56+8} = \frac{56}{64} = 0.8750(87.50\%)$$

$$\text{Recall(Layak)} = \frac{TP(\text{Layak})}{TP(\text{Layak})+FP(\text{Layak})} = \frac{41}{41+8} = \frac{41}{49} = 0.8367(83.67\%)$$

$$\text{Recall(Belum)} = \frac{TP(\text{Belum})}{TP(\text{Belum})+FP(\text{Belum})} = \frac{56}{56+4} = \frac{56}{60} = 0.9333(93.33\%)$$

Dari data diatas maka dapat di interpretasikan dengan Akurasi 88.91% menunjukkan bahwa model mampu mengklasifikasikan 88.91% dari data dengan benar,

Classification Error 11.09% menunjukkan bahwa hanya 11.09% prediksi yang salah dan menunjukkan keakuratan model. Dengan *precision* kelayakan sebesar 91.11%, mengindikasikan bahwa model memiliki tingkat akurasi yang tinggi terhadap indentifikasi karyawan yang layak mendapat promosi jabatan. Serta *precision* belum layak sebesar 87.50% menunjukkan bahwa dari semua prediksi, model cukup dapat diandalkan untuk meminimalisir kesalahan klasifikasi karyawan yang belum layak untuk promosi jabatan. Model ini dapat digunakan sebagai alat bantu keputusan dalam penilaian promosi karyawan dengan performa yang baik, perusahaan dapat meningkatkan efisiensi dalam mengidentifikasi karyawan yang layak dipromosikan serta mengurangi risiko kesalahan dalam pengambilan keputusan, yang sangat berguna untuk PT Siprama Cakrawala.

KESIMPULAN DAN SARAN

A. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dalam penelitian ini maka dapat ditarik beberapa kesimpulan sebagai berikut: Berdasarkan pengujian yang telah dilakukan dalam penelitian, dapat ditarik kesimpulan bahwa variabel-variabel yang relevan dalam penelitian ini meliputi Kinerja, Pengalaman, Pendidikan, Sikap, serta Keterlibatan proyek. Penentuan variabel-variabel sudah dipertimbangkan oleh bantuan Supervisor dan Manager demi mendapat variabel-variabel yang relevan dan sesuai dengan standar perusahaan. Pada data *entropy* dan *gain* yang telah didapat, variabel Pengalaman adalah variabel yang paling menentukan dari model *Decision Tree* ini. Melalui tahap pembagian kuesioner, transformasi hasil kuesioner ke format excel, normalisasi hasil kuesioner, reduksi data, serta merubah hasil normalisasi menjadi format *dataset* untuk menjadi data *training* pada *RapidMiner*, penelitian dapat mengolah dan mempersiapkan data historis karyawan agar dapat digunakan dalam *Algoritma C4.5*. Metrik evaluasi lainnya seperti *Precision* dan *Recall*, menunjukkan bahwa model lebih unggul dalam mengenali karyawan yang tidak layak (Belum Layak) dibandingkan yang layak untuk promosi. *Recall* untuk kategori Belum Layak mencapai 93.33%, yang berarti model sangat baik dalam memastikan bahwa karyawan benar-benar tidak layak untuk promosi, teridentifikasi dengan benar, dan *Recall* untuk kategori Layak sebesar 83.67% dinyatakan cukup baik oleh Supervisor dan para Manager untuk membantu klasifikasi karyawan yang layak mendapatkan hak kenaikan jabatan serta mengurangi subjektifitas pada proses kenaikan jabatan pada perusahaan. Model *Decision Tree* ini menunjukkan performa yang baik dalam memprediksi kelayakan promosi jabatan. Dengan tingkat akurasi 88.91%, model mampu memberikan prediksi yang andal dalam mengklasifikasikan data. Dengan margin *Error*

11.09%, margin tersebut dapat diterima dalam konteks *Real-World* yang sering kali mengandung *noise* atau ketidakakuratan dalam data yang diperoleh. Dan berdasarkan hasil akurasi yang menyentuh nilai 88.91%, *Algoritma C4.5* terbukti mampu membangun model klasifikasi yang akurat. Oleh karena itu, hipotesis H0 dan H1 dapat dinyatakan bernilai positif terhadap model algoritma yang telah ditetapkan, serta *information gain* dari atribut Pengalaman(X2) yang tinggi dalam model *Decision Tree* menunjukkan bahwa atribut ini dinyatakan sebagai salah satu pembagi utama (*Root Node*), maka H2 dianggap valid. Secara Keseluruhan, model *Decision Tree* ini layak digunakan untuk membantu pengambilan keputusan promosi karyawan dan telah di setujui oleh Supervisor beserta para Manager untuk dilakukan implementasi lebih lanjut pada perusahaan. Jika akurasi lebih tinggi diperlukan, beberapa penyesuaian seperti menambahkan lebih banyak data serta meningkatkan *pre-processing* dapat diperhitungkan

B. SARAN

Setelah penelitian menyelesaikan pembahasan pada penelitian ini, maka pada bab penutup penelitian mengemukakan saran-saran sesuai dengan hasil pengamatan dalam pembahasan penelitian ini. Adapun saran-saran yang penelitian berikan setelah meneliti permasalahan ini: Untuk meningkatkan performa model *Decision Tree* ini, beberapa langkah dapat dilakukan, yang pertama penambahan data dari segi jumlah maupun variabel yang relevan, seperti evaluasi kinerja tahunan dan pelatihan yang diikuti dapat memberikan konteks yang lebih kaya pada model. Untuk meningkatkan performa model *Decision Tree* ini, beberapa langkah dapat dilakukan, yang pertama penambahan data dari segi jumlah maupun variabel yang relevan, seperti evaluasi kinerja tahunan dan pelatihan yang diikuti dapat memberikan konteks yang lebih kaya pada model. Selanjutnya, proses data *pre-processing* seperti menangani ketidakseimbangan data menggunakan teknik *oversampling* atau *undersampling*, serta normalisasi data lebih mendalam untuk memahami pola secara lebih konsisten. Selain itu *Tuning Parameter* model, seperti mengatur kedalaman maksimal pohon atau kriteria pemisahan dapat mengurangi resiko *overfitting* atau *underfitting*.

REFERENSI

Abidin, Zaenal; Nurhana, Eka; Permata, Permata; Ulum, F. (2023). ANALISIS PERBANDINGAN ALGORITMA DECISION TREE C4.5 DAN C5.0 PADA DATA KARYAWAN BERPOTENSI PROMOSI JABATAN. In *Jurnal Teknoinfo* (Vol. 17, Issue 2). www.kaggle.com

- Angreni, D. S., & Susanti, M. (2024). Implementasi Data Mining Untuk Rekomendasi Kenaikan Pangkat Pegawai Negeri Sipil Menggunakan Algoritma Naïve Bayes Pada Biro Administrasi Pimpinan Sekretariat Daerah Provinsi Sulawesi Tengah. *Journal Of Social Science Research*, 4(1), 9661–9674.
- Devia, E. (2023). Penerapan Decision Tree Dengan Algoritma C4.5 Untuk Menentukan Rekomendasi Kenaikan Jabatan Karyawan. *Jurnal Information System*, 3(1), 028–037.
- Fanny Irnanda, K., & Perdana Windarto, A. (n.d.). *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat*.
- Iftitah, A., & Setyadi, R. (2023). Penerapan Algoritma C4.5 Untuk Analisis Pengadaan Peralatan dan Mesin Kantor. *Journal of Information System Research (JOSH)*, 4(2), 434–442.
<https://doi.org/10.47065/josh.v4i2.2673>
- Isra, M. (2022). Behavior Analysis and Prediction of Civil Services Staff in Occupational Functional Positions Using C4.5 Algorithm. *Jurnal Informasi Dan Teknologi*, 4(1), 58–63.
<https://doi.org/10.37034/jidt.v4i1.186>
- Mayangsari, D., & Wahyudi, T. (2024). Klasifikasi Seleksi Penerimaan Karyawan Baru Menggunakan Algoritma C4.5. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(1), 26–34.
<https://doi.org/10.57152/malcom.v5i1.1525>
- Nugroho, A., & Amrullah, A. (2023). EVALUASI KINERJA ALGORITMA K-NN MENGGUNAKAN K-FOLD CROSS VALIDATION PADA DATA DEBITUR KSP GALIH MANUNGGAL. In *JINTEKS* (Vol. 5, Issue 2).
- Oktafiqurahman, A., Kusri, K., & Nasiri, A. (2023). Implementasi Algoritma C4.5 Untuk Tes Kepribadian Penerimaan Karyawan Di Dinas Perhubungan Provinsi Sulawesi Tengah. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomsin)*, 11(1), 11.
<https://doi.org/10.30646/tikomsin.v11i1.719>
- Rochmat Asyhari, M., Antonius Alijoyo, F., & Vernanda, D. (2024). Penerapan Algoritma C4.5 dan Metode 360 Pada Sistem Informasi Penilaian Karyawan. *Jurnal Tekno Kompak*, 18(1), 13–27.
- SHRM. (2023). *SHRM State of the Workplace Report 2023-24*.
- Widodo, T., Setiawan, D., Syahputri, A., Informasi, S., & Triguna Dharma, S. (2022). Data Mining Menentukan Minat Konsumen Memilih Sepeda Motor Idaman Dengan Algoritma C4.5. *Jurnal Sistem Informasi TGD*, 1(6), 826–836.
<https://ojs.trigunadharma.ac.id/index.php/jsi>