

Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5

Ikhsan Romli¹, Ahmad Turmudi Zy²

^{1,2}Teknik Informatika, Universitas Pelita Bangsa
Jl. Inspeksi kalimalang arah Deltamas, Tlp 021 2851 818
Ikhsan.romli@pelitabangsa.ac.id

Abstract

Technological development and scientific advancement are very important and influential parts in all fields. By the development and advancement of technology at this time, being experts in a field is a must because it is required to know more and learn about the technology that is currently developing. Information technology and informatics are needed to support performance. Large and small companies also need fast and accurate information to make easier decisions making. Therefore, data mining classification techniques are needed to solve these problems. The classification used in data mining is Decision tree because it is technique that is widely used and produces output with existing rules, so that it can present employee data to determine the overtime schedule. This study uses the C4.5 algorithm to determine the overtime schedule. The test results of the overtime schedule with the C4.5 algorithm with the Confusion matrix have good accuracy, precision and recall values, namely 91% accuracy, 86.05% precision and 92.5% recall.

Keywords: overtime schedule, Data mining, Decision tree, C4.5 Algorithm

Abstrak

Perkembangan teknologi dan kemajuan ilmu pengetahuan merupakan bagian yang sangat penting dan berpengaruh disegala bidang. Dengan adanya perkembangan dan kemajuan teknologi di masa ini, maka ahli di bidangnya merupakan suatu keharusan karena dituntut untuk semakin tau dan mempelajari teknologi yang berkembang sekarang ini. Teknologi informasi dan ilmu informatika sangat di butuhkan untuk menunjang kinerja. Perusahaan besar maupun kecil juga membutuhkan informasi yang cepat dan akurat untuk memudahkan dalam mengambil suatu keputusan. Oleh karena itu dibutuhkan teknik klasifikasi data mining dalam memecahkan masalah tersebut. Klasifikasi yang digunakan dalam data mining adalah Decision tree dikarenakan merupakan teknik yang banyak digunakan dan menghasilkan output dengan aturan yang ada, dengan hal tersebut dapat menyajikan data karyawan untuk menentukan jadwal overtime. Dalam penelitian ini menggunakan algoritma C4.5 untuk menghasilkan menentukan jadwal overtime. Hasil uji coba jadwal overtime dengan algoritma C4.5 dengan Confusion matrix memiliki nilai accuracy, precision, dan recall yang bagus yaitu dengan accuracy 91%, precision 86.05% dan recall 92.5%.

Kata kunci: jadwal overtime, Data mining, Decision tree, Algoritma C4.5

1. PENDAHULUAN

Setiap perusahaan menginginkan karyawannya dapat bekerja secara maksimal sesuai dengan keahliannya masing-masing, banyak sekali faktor-faktor yang mempengaruhi kinerja seorang karyawan, seperti umur, jenis kelamin, tingkat kedisiplinan dll. Berdasarkan faktor-faktor tersebut, maka harus disusun sebuah jadwal kerja yang sesuai dengan kriteria masing-masing. Namun akan muncul beberapa kendala apabila pembuatan jadwal kerja itu dilakukan manual, salah satunya jumlah karyawan yang besar. Dengan jumlah karyawan yang mencapai ratusan orang akan sangat sulit

membuat jadwal kerja dengan mempertimbangkan kriteria dari masing-masing karyawan.

Untuk mengatasi kendala-kendala tersebut maka diperlukan data mining yang merupakan teknik baru dimana terdapat penggabungan antara metode analisis tradisional dengan algoritma untuk memproses data berjumlah besar. Salah satu manfaat data mining dalam perusahaan yaitu dapat digunakan sebagai bahan untuk mendukung pengambilan keputusan dan menunjang program kegiatan perusahaan. Salah satunya untuk menentukan jadwal overtime pada Perusahaan yang memiliki staff dengan bagian dan sistem penggajian yang berbeda-beda. Dalam pekerjaan tersebut maka pihak Perusahaan juga harus dapat menentukan staff yang akan mengerjakannya. Di dalam menentukan jadwal staff pihak perusahaan harus menentukan juga target dari masing-masing staff dalam menyelesaikan pekerjaan tersebut. Untuk menghitung lama kerja staff setiap harinya termasuk lembur.

Dalam data mining, terdapat berbagai fungsi salah satunya merupakan fungsi klasifikasi. Teknik yang dapat digunakan untuk menentukan klasifikasi adalah algoritma Decision Tree C4.5 yang merupakan teknik klasifikasi data dan prediksi dengan metode pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar dan paling bawah dinamakan daun. Beberapa penelitian terkait pengklasifikasian data karyawan dengan menggunakan algoritma C4.5 telah dilakukan oleh Budanis dan Achmad pada tahun 2011 dengan judul klasifikasi data karyawan untuk menentukan jadwal kerja menggunakan metode decision tree, dalam penelitian tersebut telah dibuktikan bahwa pohon keputusan menghasilkan rule-rule untuk membentuk suatu jadwal kerja baru yang dilakukan pengujian dengan dataset yang diperoleh nilai akurasi sebesar 87 % dari rule-rule yang telah terbentuk [1]. Kemudian pada tahun 2016 Luvia dan Yuni melakukan penelitian dengan menerapkan algoritma C4.5 untuk klasifikasi predikat keberhasilan mahasiswa di AMIK Tunas Bangsa [2]. Lalu di tahun 2017 beberapa penelitian dengan menerapkan algoritma C4.5 pernah dilakukan oleh Kurniawan dan Tosy [3], Ferdian dan Hansun [4], Saefudin dan Lestari [5]. Pada tahun 2019, Algoritma C4.5 juga pernah digunakan dalam memprediksi kepuasan pelanggan yaitu penelitian dalam memprediksi kepuasan penumpang Bus Rapi Transit (BRT) Trans Semarang dengan menggunakan algoritma yang sama [6], dilanjutkan oleh Yulia dengan melakukan prediksi kepuasan mahasiswa terhadap kinerja Dosen di Kota Batam [7]. kemudian pada tahun 2020, algoritma C4.5 pernah dipakai dalam menentukan kepuasan pelanggan selaku wajib pajak terhadap layanan pajak di Kantor Pelayanan Pajak [8].

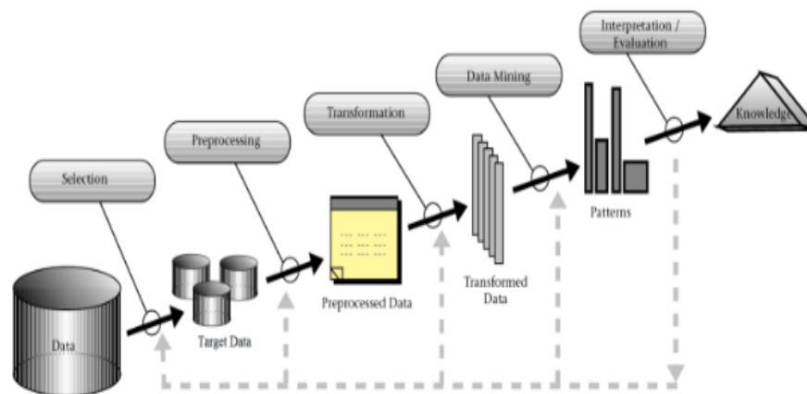
Berdasarkan latar belakang dari permasalahan tersebut maka penelitian ini menggunakan hasil ketepatan akurasi Algoritma Decision Tree C4.5 dalam mengklasifikasikan karyawan dalam menentukan jadwal overtime pada Perusahaan dan menentukan pengaruh atribut dari posisi, jenis kelamin, agama, lama kerja, dan kesehatan terhadap klasifikasi karyawan untuk menentukan jadwal overtime perusahaan.

2. METODOLOGI PENELITIAN

2.1. Tahap Tahap Data Mining

Data Mining merupakan bagian yang terintegrasi dari Knowledge Discovery in Databases (KDD). Bertujuan untuk proses transformasi data mentah menjadi informasi berguna. Jika digambarkan secara detail tahapan KDD menjadi 5 tahap. Berikut Tahapannya:

- Seleksi
- Preprocessing
- Transformasi
- Data Mining
- Interpretasi & Evaluasi



Gambar 1. Tahapan KDD

2.2. Pengertian Klasifikasi

Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan [9]. Klasifikasi merupakan suatu teknik menemukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan.

Ketika membahas klasifikasi, kita berbicara tentang taksonomi. Taksonomi (Tassein = mengklasifikasi + nomos = ilmu pengetahuan, hukum) muncul pertama sebagai ilmu mengelompokkan organism hidup, tetapi kemudian dikembangkan sebagai ilmu klasifikasi pada umumnya, termasuk disini prinsip – prinsip klasifikasi. Dengan demikian, klasifikasi (taksonomi) adalah proses menempatkan objek tertentu dalam satu set kategori, berdasarkan masing – masing menempatkan objek [10]. Ada beberapa algoritma yang dapat digunakan untuk dapat menangani masalah klasifikasi data, a.l. Decision Tree, KNN (K-Nearest Neighbor), ANN (Artificial Neural Network), Naive Bayes, SVM (Support Vector Machine) [11]. Dalam penelitian ini, algoritma yang digunakan untuk mengklasifikasikan yaitu Decision Tree dengan seri C4.5.

2.3. Pengertian Algoritma C4.5

Di tahun 1970 hingga 1980 J. Ross Quinlan membuat sebuah algoritma Decision Tree (Pohon Keputusan) yang dikenal dengan ID3 (Iterative Dichotomiser) yang menjadi cikal bakal awal mula perkembangan algoritma Decision Tree. Kemudian, Quinlan mengembangkan algoritma ID3 menjadi C4.5 [12]. Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan. Kelebihannya adalah dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, dapat menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma lainnya [13]. Algoritma ini dapat divisualisasikan dalam bentuk gambar pohon keputusan.

Menurut Kusriani dan Luthfi [14], secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- a) Pilih atribut sebagai akar.
- b) Buat cabang untuk tiap-tiap nilai.
- c) Bagi kasus dalam cabang.
- d) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Dari beberapa definisi tentang algoritma C4.5 diatas maka penulis menyimpulkan bahwa algoritma C4.5 adalah algoritma yang memiliki banyak manfaat dan kelebihan. Kelebihannya adalah dapat mengolah data numerik dan diskret, dapat menangani nilai atribut yang hilang, dapat menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain.

2.4. Entropy

Menurut Suyanto [13], secara istilah *Entropy* adalah keberbedaan atau keragaman. Dalam data mining, *Entropy* didefinisikan sebagai suatu parameter untuk mengukur heterogenitas (keberagaman) dalam suatu himpunan data. Semakin heterogen himpunan suatu data, semakin besar pula nilai *Entropy*-nya. Secara matematis *Entropy* dirumuskan sebagai berikut:

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

Dimana n adalah jumlah nilai yang terdapat pada atribut target (jumlah kelas). Sedangkan P_i menyatakan porsi atau rasio antara jumlah sampel di kelas i dengan jumlah sampel pada himpunan data.

2.5. Gain Ratio

Algoritma C4.5 merupakan perbaikan dari ID3 menggunakan Gain Ratio untuk diperbaharui information gain. Maka dengan rumus:

$$GainRatio(S,A) = (Gain(S,A)) / (Splitinfo(S,A)) \quad (2)$$

Dimana :

S = Ruang/Data Sample yang digunakan untuk data training

A = Atribut

Gain(S.A) = information gain pada atribut A

Splitinfo(S.A) = split information pada atribut A

Dengan atribut yang nilai gain ratio yang paling tertinggi dipilih sebagai atribut test untuk simpul. Pendekatan ini menerapkan normalisasi pada information gain dengan menggunakan apa yang disebut dengan split information, dengan rumus:

$$\text{Splitinfo}(S.A) = - \sum_{i=1}^i \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

Dimana :

S = Ruang (data) sample yang digunakan untuk training.

A = Atribut.

Si = Jumlah sample untuk atribut i.

3. HASIL DAN PEMBAHASAN

3.1. Proses Pemilihan Data

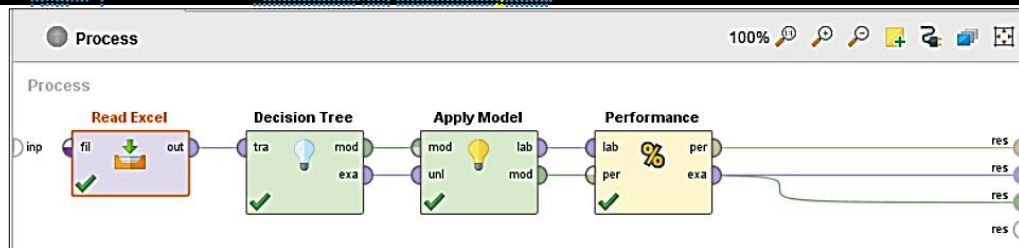
Dataset yang sudah ada selanjutnya dilakukan pemilihan terhadap parameter yang akan dianalisis. Parameter yang diambil adalah atribut dari data jadwal overtime yang telah didapat sebelumnya dan akan digunakan untuk menjadi masukan atau variabel input. Parameter tersebut ada sebanyak 9 (sembilan),

Tabel 1. Tabel Pemilihan Dataset

No.	Atribut	Indikator	Detail Penggunaan
1	No	√	Id
2	NIP	X	-
3	Nama	√	Nilai Model
4	Bagian	√	Nilai Model
5	Agama	√	Nilai Model
6	Lama Kerja	√	Nilai Model
7	Umur	X	-
8	Kesehatan	√	Nilai Model
9	Status Lembur	√	Label Target

3.2. Proses Pengujian

Pada pengujian ini dilakukan eksperimen dengan menggunakan algoritma C4.5, dengan tujuan untuk mengetahui akurasi dari algoritma C4.5 terhadap dataset. Berikut ini ditunjukkan pada gambar 2 terkait model pengujian dengan menggunakan algoritma decision tree serta proses validasi dengan menghubungkan building block performance dengan menggunakan tools Rapidminer



Gambar 2. Model Pengujian dan Validasi Menggunakan Algoritma C4.5

3.3. Perhitungan Root Node

Tahapan-tahapan dalam menentukan simpul akar pada decision tree C4.5 yaitu menghitung nilai *Entropy*, gain, splitinfo dan gain ratio masing-masing atribut bagian, jenis kelamin, agama, lama kerja, kesehatan.

$$Entropy \text{ [Bagian-DIREKTUR]} = -(0/2) * (\log_2(0/2)) + -(2/2) * (\log_2(2/2)) = 0,0000$$

$$Entropy \text{ [Bagian-Checker]} = -(15/17) * (\log_2(15/17)) + -(2/17) * (\log_2(2/17)) = 0,5226$$

$$Entropy \text{ [Bagian-Kuangan]} = -(12/12) * (\log_2(12/12)) + -(0/12) * (\log_2(0/12)) = 0,0000$$

$$Entropy \text{ [Bagian-Outsourcing]} = -(4/30) * (\log_2(4/30)) + -(26/30) * (\log_2(26/30)) = 0,5665$$

$$Entropy \text{ [Bagian-Staff]} = -(29/39) * (\log_2(29/39)) + -(10/39) * (\log_2(10/39)) = 0,8213$$

$$Gain \text{ [Bagian]} = 0,9710 - ((2/100) * 0,0000) + ((17/100) * 0,5226) + ((12/100) * 0,0000) + ((30/100) * 0,5665) + ((39/100) * 0,8213) = 0,3919$$

$$Splitinfo[\text{Bagian}] = -(2/100) * (\log_2(2/100)) + -(17/100) * (\log_2(17/100)) + -(12/100) * (\log_2(12/100)) + -(30/100) * (\log_2(30/100)) + -(39/100) * (\log_2(39/100)) = 1,9654$$

$$Gain \text{ Ratio}[\text{Bagian}] = Gain / Splitinfo = 0,3919/1,964 = 0,1994$$

$$Entropy \text{ [Jenis Kelamin - L]} = -(41/64) * (\log_2(41/64)) + -(23/64) * (\log_2(23/64)) = 0,9422$$

$$Entropy \text{ [Jenis Kelamin - P]} = -(19/36) * (\log_2(19/36)) + -(17/36) * (\log_2(17/36)) = 0,9978$$

$$Gain \text{ [Jenis Kelamin]} = 0,9710 - ((64/100) * 0,9422) + ((36/100) * 0,9978) = 0,0088$$

$$Splitinfo \text{ [Jenis Kelamin]} = -(64/100) * (\log_2(64/100)) + -(36/100) * (\log_2(36/100)) = 0,9427$$

$$Gain \text{ Ratio [Jenis Kelamin]} = Gain / Splitinfo = 0,0088/0,9427 = 0,0093$$

$$Entropy \text{ [Agama- Islam]} = -(51/85) * (\log_2(51/85)) + -(34/85) * (\log_2(34/85)) = 0,9710$$

$$Entropy \text{ [Agama- Kristen]} = -(9/15) * (\log_2(9/15)) + -(6/15) * (\log_2(6/15)) = 0,9710$$

$$\text{Gain [Agama]} = 0,9710 - ((85/100) * 0,9710) + ((15/100) * 0,9710) = 0,0000$$

$$\text{Splitinfo [Agama]} = (-(85/100) * (\log_2(85/100))) + (-(15/100) * (\log_2(15/100))) = 0,6098$$

$$\text{Gain Ratio [Agama]} = \text{Gain} / \text{Splitinfo} = 0,0000/0,6098 = 0,0000$$

$$\text{Entropy [Lama Kerja- Islam]} = (-(31/58) * (\log_2(31/58))) + (-(27/58) * (\log_2(27/58))) = 0,9966$$

$$\text{Entropy [Lama Kerja - Kristen]} = (-(29/42) * (\log_2(29/42))) + (-(13/42) * (\log_2(13/42))) = 0,8926$$

$$\text{Gain [Lama Kerja]} = 0,9710 - ((58/100) * 0,9966) + ((42/100) * 0,8926) = 0,0180$$

$$\text{Splitinfo [Lama Kerja]} = (-(58/100) * (\log_2(58/100))) + (-(42/100) * (\log_2(42/100))) = 0,9815$$

$$\text{Gain Ratio [Lama Kerja]} = \text{Gain} / \text{Splitinfo} = 0,0180/0,9815 = 0,0184$$

$$\text{Entropy [Kesehatan -Baik]} = (-(48/78) * (\log_2(48/78))) + (-(30/78) * (\log_2(30/78))) = 0,9612$$

$$\text{Entropy [Kesehatan - Tidak Baik]} = (-(12/22) * (\log_2(12/22))) + (-(10/22) * (\log_2(10/22))) = 0,9940$$

$$\text{Gain [Kesehatan]} = 0,9710 - ((78/100) * 0,9612) + ((22/100) * 0,9940) = 0,0025$$

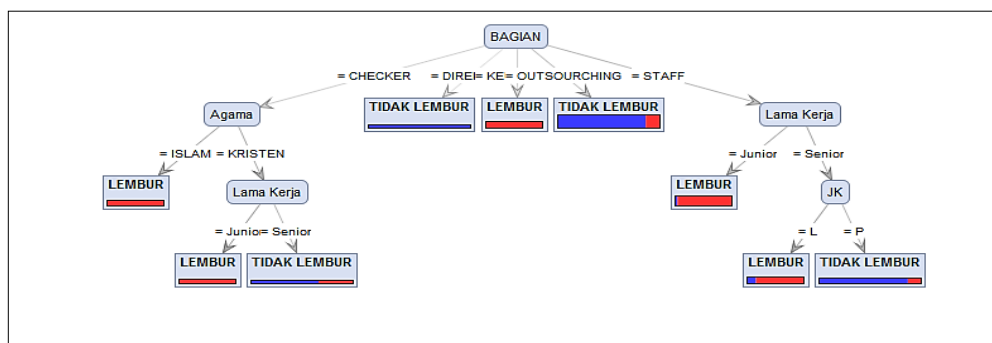
$$\text{Splitinfo [Kesehatan]} = (-(78/100) * (\log_2(78/100))) + (-(22/100) * (\log_2(22/100))) = 0,7602$$

$$\text{Gain Ratio [Kesehatan]} = \text{Gain} / \text{Splitinfo} = 0,0025/0,7602 = 0,0033.$$

Setelah semua simpul akar ditemukan maka hal tersebut akan digunakan dalam menentukan rule model pohon keputusan.

3.4. Hasil Rule Model Decision Tree

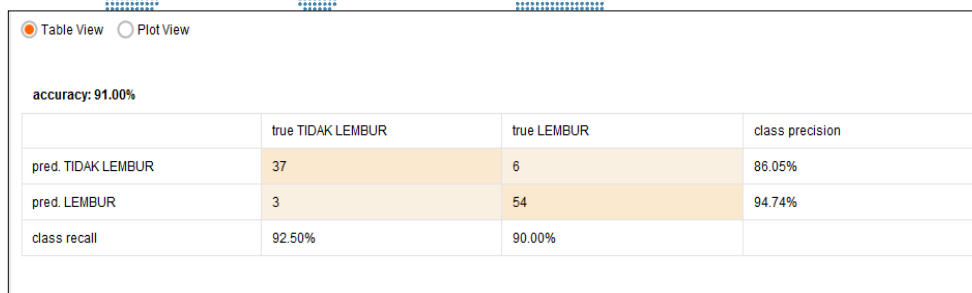
Berikut ini ditampilkan pada gambar 3 tentang hasil rule model pohon keputusan yang dapat membantu seorang pengambil keputusan dalam menentukan jadwal overtime.



Gambar 3. Pohon Keputusan Penentuan Jadwal Overtime

3.5. Perhitungan Confusion Matrix

Berikut ini ditunjukkan pada gambar 4 tentang hasil perhitungan akurasi, recall dan precision.



accuracy: 91.00%

	true TIDAK LEMBUR	true LEMBUR	class precision
pred. TIDAK LEMBUR	37	6	86.05%
pred. LEMBUR	3	54	94.74%
class recall	92.50%	90.00%	

Gambar 4. Hasil Perhitungan Confusion Matrix

Proses perhitungan confusion matrix ditunjukkan secara detil di bawah ini.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\% = (37 + 54) / (37 + 54 + 6 + 3) \times 100\% = 91 / 100 \times 100\% = 0.91 \times 100\% = 91\%$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \times 100\% = 37 / (37 + 6) \times 100\% = 37 / 43 \times 100\% = 0.86 \times 100\% = 86\%$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \times 100\% = 37 / (37 + 3) \times 100\% = 37 / 40 \times 100\% = 0.925 \times 100\% = 92.5\%$$

Hasil perhitungan manual terhadap accuracy sebesar 91%, kemudian precision yang diperoleh nilainya mencapai 86% dan perhitungan Recall 92,5% , hal ini mendekati dengan hasil perhitungan dari Rapidminer.

4. SIMPULAN

Berdasarkan hasil pengujian dataset terhadap algoritma C4.5 dalam menentukan jadwal overtime, maka dapat diambil kesimpulan bahwa hasil pohon keputusan tersebut dapat berkontribusi dalam membantu pengambilan keputusan di Perusahaan tersebut untuk menentukan jadwal lembur secara lebih cepat, akurat dan objektif berdasarkan hasil rule model yang telah terbentuk. Hasil pengujian dengan algoritma C4.5 memiliki nilai accuracy, precision, dan recall yang bagus yaitu dengan accuracy 91%, precision 86.05% dan recall 92.5%.

DAFTAR PUSTAKA

- [1] F. Dwi Meliani Achmad, Budanis, Slamet, "Klasifikasi Data Karyawan Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree," *J. IPTEK*, vol. 16, no. 1, pp. 18–23, 2012.
- [2] Y. S. Luvia, D. Hartama, A. P. Windarto, and Solikhun, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di AMIK Tunas Bangsa," *JURASIK*, vol. 1, no. 1, pp. 75–79, 2016, doi: 10.1134/S1068162020050106.
- [3] T. C. Kurniawan, "Penerapan Algoritma C4.5 Dalam Penerimaan Calon Karyawan PT. Telkom Akses Area Lampung Berbasis Website," *Semin*.



- Nas. IIB Darmajaya*, pp. 491–501, 2017.
- [4] F. F. Harryanto and S. Hansun, “Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE,” *J. Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, pp. 95–103, 2017.
 - [5] S. L. Saefudin, “Sistem Pendukung Keputusan Untuk Penilaian Kinerja Karyawan,” *J. Pengemb. Ris. dan Obs. Tek. Inform.*, vol. 2, no. September, pp. 40–43, 2015.
 - [6] A. S. Febriarini and E. Z. Astuti, “Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang,” *Eksplora Inform.*, vol. 8, no. 2, pp. 95–103, 2019, doi: 10.30864/eksplora.v8i2.156.
 - [7] Yulia and A. D. Putri, “Data Mining Menggunakan Algoritma C4.5 Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Di Kota Batam,” in *SNISTEK*, 2019, no. 2, pp. 235–240, doi: <https://doi.org/10.33884/cbis.v7i2.1373>.
 - [8] I. Romli, F. Kharida, and C. Naya, “Penentuan Kepuasan Pelanggan Terhadap Pelayanan Kantor Pelayanan Pajak Menggunakan C4.5 dan PSO,” *RESTI*, vol. 4, no. 2, pp. 296–302, 2020.
 - [9] M. F. Arifin and D. Fitriana, “Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada,” *InComTech*, vol. 8, no. 2, pp. 87–102, 2018, doi: 10.22441/incomtech.v8i1.2198.
 - [10] F. Gorunescu, *Data Mining: Concepts, Models, and Techniques*, 12th ed. Warsaw, Poland: Springer, 2011.
 - [11] F. Santoso, A. Syukur, and A. Z. Famani, “Algoritma C4 . 5 Dengan Particle Swarm Optimization Untuk Klasifikasi Lama Menghafal Al-Quran Pada Santri,” *J. Teknol. Inf.*, vol. 14, no. 2, pp. 92–103, 2018.
 - [12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Third Edit. USA: Morgan Kaufmann, 2012.
 - [13] Suyanto, *Data mining : untuk klasifikasi dan klasterisasi data*, 1st ed. Bandung: Infoinformatika, 2017.
 - [14] Kusri and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.