



PREDIKSI PENGANGKATAN KARYAWAN DENGAN METODE KLASIFIKASI ALGORITMA C5.0 (STUDI KASUS CV. T-PICO JAYA MANDIRI)

Suprpto¹, Fahrul Fahrezi J.T², Edora³

Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa
¹suprpto@pelitabangsa.ac.id, ²fahrulfahrezi21@gmail.com, ³edora@pelitabangsa.ac.id

Abstrak

Dalam keputusan pengangkatan karyawan tetap yang bisa dilakukan pada perusahaan ditentukan dari segi absensi dan kedisiplinan yang baik. Penentuan pengangkatan karyawan tetap di CV. T-PICO JAYA MANDIRI masih terlihat kurang tepat dan membutuhkan waktu yang cukup lama. Maka dilakukan sebuah penelitian yang bertujuan untuk mengetahui informasi prediksi pengangkatan karyawan tetap dengan dilihat dari kriteria-kriteria yang ditetapkan oleh perusahaan menggunakan klasifikasi algoritma C5.0 dengan metode decision tree. Data yang digunakan pada penelitian ini adalah data karyawan yang dimiliki CV. T-PICO JAYA MANDIRI sebanyak 403 data, proses pengujian metode menggunakan RapidMiner9.5. Berdasarkan hasil pengujian terhadap penelitian dalam memprediksi pengangkatan karyawan pada CV. T-PICO JAYA MANDIRI, diperoleh hasil dari algoritma C5.0 atau decision tree yaitu accuracy sebesar 90,00%. Maka dapat disimpulkan bahwa teknik algoritma C5.0 dengan metode decision tree dinilai berhasil untuk memprediksi pengangkatan karyawan pada CV. T-PICO JAYA MANDIRI.

Kata kunci: Rekrutmen Karyawan, Klasifikasi, Algoritma C5.0.

Abstract

In the decision to appoint permanent employees that can be made at the company is determined in terms of attendance and good discipline. Determination of the appointment of permanent employees at CV. T-PICO JAYA MANDIRI still looks imprecise and takes a long time. Then a study was conducted that aims to determine the prediction information on permanent employee appointments by looking at the criteria set by the company using the C5.0 classification algorithm with the decision tree method. The data used in this study are employee data owned by CV. T-PICO JAYA MANDIRI as many as 403 data, the process of testing the method using Rapid Miner9.5. Based on the results of testing on research in predicting employee appointments at CV. T-PICO JAYA MANDIRI, obtained the result from the C5.0 algorithm or decision tree, which is an accuracy of 90.00%. So it can be concluded that the C5.0 algorithm technique with the decision tree method is considered successful in predicting employee appointments at CV. T-PICO JAYA MANDIRI.

Keywords: Employee Recruitment, Classification, C5.0 Algorithm.

1. Pendahuluan

Karyawan dan perusahaan merupakan dua hal yang sangat bergantung dan saling membutuhkan satu sama lain (Liana, 2020). Karyawan membutuhkan perusahaan sebagai tempat dimana ia dapat bekerja serta memperoleh imbalan yang dapat digunakan untuk memenuhi kebutuhan-kebutuhan hidupnya (Desi Kristanti, 2019). Perusahaan membutuhkan karyawan sebagai sumber daya manusia yang menggerakkan roda perusahaan sehingga perusahaan dapat melakukan produksi dan menghasilkan produk dan kinerja yang memiliki dan pengalaman kerja Karyawan (Nugraha, 2022). Jika karyawan kontrak tersebut telah memenuhi nilai kriteria-kriteria yang sudah ditentukan maka karyawan kontrak tersebut dapat ditentukan sebagai karyawan tetap pada CV. T-Pico Jaya Mandiri.

Permasalahan yang ada dalam Cv. T-Pico Jaya Mandiri dalam penentuan karyawan kontrak menjadi karyawan tetap selama ini dilakukan secara bertahap dengan penyeleksian berkas, seperti tes lisan, tes fisik,

atau tertulis, wawancara dan lain sebagainya. Cara ini cukup membutuhkan waktu yang lama, sehingga kemungkinan akan sering terjadinya kesalahan dalam hasil akhir dari penentuan karyawan kontrak menjadi karyawan tetap sering tidak memenuhi kriteria yang dibutuhkan perusahaan dan menghambat kinerja perusahaan. Karena tugas ini melibatkan banyak manager untuk memutuskan karyawan yang tepat bagi perusahaan, dengan keputusan tersebut sangatlah tidak mudah, karena perlu pertimbangan dari berbagai indikator-indikator seperti kinerja materi, kinerja lapangan, absensi, kehadiran, loyalitas, dan kedisiplinan dengan demikian penulis ingin menerapkan sebuah teknik data mining untuk membantu dalam proses pengangkatan karyawan tetap.

Solusi untuk mengatasi permasalahan di atas peneliti menggunakan metode klasifikasi algoritma C5.0. Hal ini berdasarkan penelitian terdahulu yang dilakukan oleh Yogi Yusuf W (2007) dengan judul “perbandingan performansi algoritma decision tree C5.0, CHART, dan CHAID: kasus prediksi status resiko kredit di Bank X” menyatakan bahwa algoritma C5.0 menghasilkan tingkat keakuratan yang lebih tinggi dalam hal prediksi. Memprediksi permintaan pengambilan keputusan untuk pengangkatan karyawan tetap yaitu dengan metode klasifikasi algoritma C5.0 adalah didasarkan pada konsepnya dimana dari beberapa faktor-faktor yang menjadi atribut penunjang dalam memprediksi pengangkatan karyawan tetap. Konsepnya sederhana dan mudah dipahami, komputasinya efisien, dan memiliki kemampuan untuk mengukur kinerja relatif dari alternatif-alternatif keputusan dalam bentuk matematis yang sederhana.

2. Landasan Pemikiran

Data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data warehouse, atau media penyimpanan informasi yang lain (Rerung, 2018).

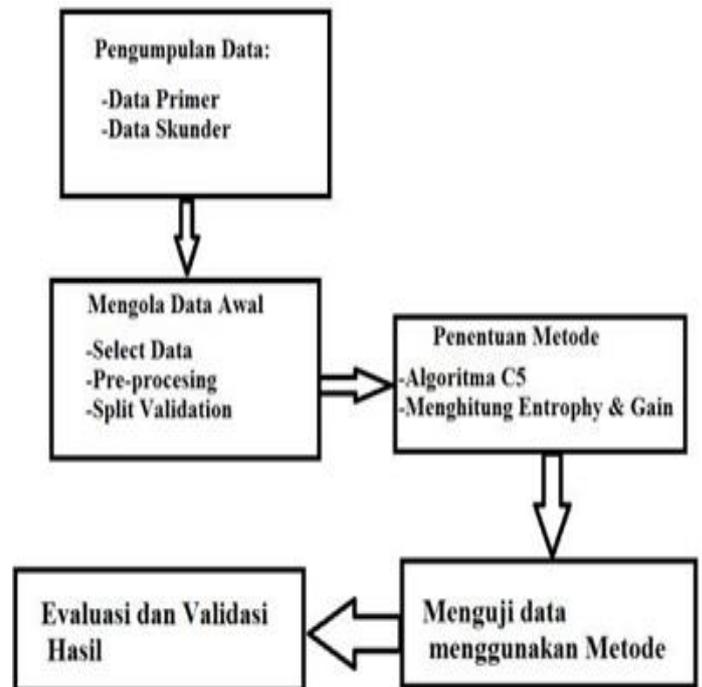
Mengaplikasikan data mining tentunya memiliki beberapa tahapan proses di dalamnya. Tahapan-tahapan tersebut bersifat saling aktif yang dapat di artikan bahwa pengguna atau user mendapatkan peran secara langsung atau dengan perantara pengetahuan dasar yang di dapatkan (Gintara, 2021).

Serangkaian proses tahapan sebagai berikut: data *selection*, *pre-processing* (*cleaning*), *transformation*, data mining, *interpretation* (*evaluation*) (Wibowo & Mahartika, 2019). Berikut ini Teknik data mining *association*, *clustering*, *description*, *prediction*, *estimation*, dan *classification* (Takdirillah, R, 2020).

Penelitian ini menggunakan metode klasifikasi dengan algoritma C5.0. Algoritma C5.0 Merupakan algoritma klasifikasi data mining yang mana pemilihan atribut yang akan diproses menggunakan information gain sebagai kriterianya (Sari N, 2021).

3. Metode Penelitian

Pada penelitian ini akan dilakukan pemodelan menggunakan Metode Algoritma C5.0 untuk mengolah data Karyawan. Metode Algoritma C5.0 dipilih karena satu kelebihanannya adalah dapat menangani data numerik dan diskret (Amalda, R. N, dkk, 2022) Dengan menghitung konsep entropy untuk membentuk pohon keputusan maka dibuat alur atau tahapan dalam penelitian ini sebagai berikut.



Gambar 1 Tahapan Penelitian

Tahap pertama dalam pengolahan data awal yaitu *Select Data*. Pada tahap ini akan dilakukan pemilihan variabel data yang akan dianalisis, karena sering ditemukan bahwa tidak semua data yang dibutuhkan dengan mempertimbangkan tujuan penulisan, sehingga diperoleh beberapa variabel yang akan digunakan untuk menjadi masukan variabel input. 403 data record akan diambil 7 variabel, yaitu: absen, lama kerja, kreatifitas, loyalitas, skill, Kinerja Lapangan, dan Disiplin hasil seleksi akan digunakan dalam proses data mining.

Tahap kedua yaitu *Pre-processing*. Pada proses *pre-processing* akan dilakukan pembersihan data untuk membuang data yang missing value yaitu data yang tidak konsisten dan juga memperbaiki data yang rusak. Proses pembersihan data dilakukan secara manual untuk memastikan bahwa data yang telah dipilih layak untuk dilakukan proses pemodelan.

Split Validation merupakan tahapan ketiga dalam pengolahan data awal. *Split Validation* merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian data training dan sebagian data testing. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik sampling random untuk data training dan data testing Contoh

perhitungan untuk pengambilan data testing adalah sebagai berikut :

Jumlah data keseluruhan (N) = 403

Jumlah data testing = 100

Data Training= 303

Dari hasil diatas diperoleh data testing sebanyak 100 data Karyawan, maka sisanya dijadikan data training sebanyak 403 – 100 = 303 data Karyawan.

Pada penelitian ini akan dilakukan pemodelan menggunakan Metode Algoritma C5.0 untuk mengolah data Karyawan. Metode Algoritma C5.0 dipilih karena satu kelebihanannya adalah dapat menangani data numerik dan diskret. Dengan menghitung konsep entropy untuk membentuk pohon keputusan. Data kemudian dihitung menggunakan algoritma sesuai dengan metodenya kemudian dicari hasil akurasi.

Ada beberapa tahap dalam membentuk pohon keputusan dengan Algoritma C5.0 antara lain:

1. Menyiapkan data training, dimana data terbesar akan diklasifikasikan.
2. Menentukan akar dari pohon, akar akan diperoleh dari atribut yang terpilih dengan cara menghitung gain dari masing-masing atribut. Nilai gain tertinggi akan dijadikan akar pertama dalam pohon keputusan. Sebelum menghitung nilai gain, hitung dulu entropinya.
3. Kemudian menghitung nilai gain.
4. Untuk langkah 2 hingga semua record terpartisi.
5. Proses partisi akan berhenti saat:
 - a. Semua record pada simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam record yang akan dipartisi lagi.
 - c. Tidak ada record dicabang yang kosong. (Aswan Supriyadi Sunge., 2018)

Pengujian Data Metode dilakukan untuk mengetahui hasil perhitungan yang dianalisa dan untuk melihat apakah fungsi bekerja dengan baik atau tidak. Setelah data dihitung secara manual, kemudian data diuji menggunakan tools RapidMiner untuk memastikan bahwa apakah hasil perhitungan manual dengan hasil yang diperoleh RapidMiner sesuai atau tidak.

Evaluasi Validasi Hasil dilakukan dengan cara mengamati dan menganalisa hasil dari algoritma yang digunakan untuk memastikan bahwa hasil pengujian itu benar atau tidak sesuai dengan pembahasan. Sedangkan, validasi dilakukan dengan mengukur hasil prediksi untuk mengetahui tingkat akurasi, presisi dan recall.

1. Akurasi adalah presentase dari catatan yang diklasifikasikan dengan benar oleh pengujian dataset.
2. Presisi adalah presentase data yang diklasifikasikan sebagai model yang baik yang sebenarnya juga baik.

3. Recall adalah hasil pengukuran tingkat pengenalan positif.

4. Pembahasan

4.1 Select Data

Tahap pertama yang dilakukan dalam penelitian ini adalah select data. Data yang sudah tersedia selanjutnya pemilihan terhadap parameter yang akan di analisis. Parameter yang diambil adalah atribut dari data karyawan yang telah didapatkan sebelumnya dari sumber yang terpercaya yaitu CV. T-Pico Jaya Mandiri, yang akan digunakan untuk menjadi masukan atau variable input.

Tabel 1 Select Data

| No | Atribut | Indikator | Detail Pengguna |
|----|------------------|-----------|-----------------|
| 1 | Nama | OK | ID |
| 2 | Absen | OK | Nilai Model |
| 3 | Lama Kerja | OK | Nilai Model |
| 4 | Kreatifitas | OK | Nilai Model |
| 5 | Loyalitas | OK | Nilai Model |
| 6 | Skill | OK | Nilai Model |
| 7 | Kinerja Lapangan | OK | Nilai Model |
| 8 | Disiplin | OK | Nilai Model |

4.2 Transformation Data

Setelah data dipilih maka akan dilakukan tahapan untuk melakukan transformasi terhadap atribut. Selanjutnya atribut data hasil proses cleaning data akan di transformasi ke dalam bentuk ordinal agar memudahkan proses pemodelan.

4.3 Split Validation

Split Validation merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian data training dan sebagian data testing. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik sampling random untuk data training (80%) dan data testing (20%). Untuk pengambilan data testing adalah sebagai berikut.

Jumlah data keseluruhan (N) = 403

Jumlah data testing = 20% x 403 = 100

Jumlah sample (n) = 100

Dari hasil diatas diperoleh data testing sebanyak 100 data karyawan, maka sisanya dijadikan data training sebanyak 403 – 100 = 400 data karyawan

4.4 Perhitungan Nilai Entropy dan Gain

1. Nilai Entropy

Jumlah kasus(all total) = 100

Jumlah hasil YA = 44 Jumlah

hasil TIDAK = 56

Maka nilai entropy total adalah

$$=((-Jumlah\ YA/all\ total)*IMLOG2(Jumlah\ YA/all\ total))+(-Jumlah\ TIDAK/all\ total)$$

$$\begin{aligned} & \text{total}) * \text{IMLOG2}(\text{JumlahTIDAK/all total}) \\ & = ((-44/100) * \text{IMLOG2}(44/100)) + (- \\ & 56/100) * \text{IMLOG2}(56/100) = 0,989588 \end{aligned}$$

2. Nilai Gain

Lama Kerja

Jumlah total data = 100

Jumlah Lama = 53 Jumlah

Sebentar = 47

Entropy Lama = 0,9687

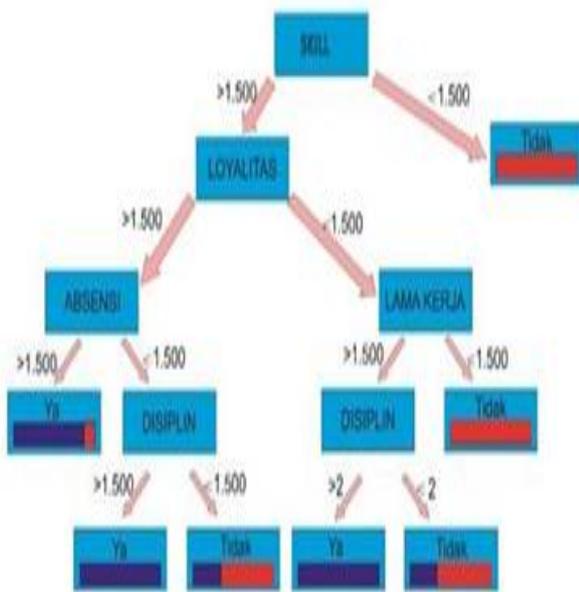
Entropy Sebentar = 0,8196

$$\begin{aligned} & = (0,9896) - ((53/100) * 0,9687) - ((47/100) * 0,8196) \\ & = 0,090977 \end{aligned}$$

Tabel 2 Perhitungan Data Testing

| | | Jumlah (s) | Ya (s) | Tidak (s) | Entropy | Gain |
|---------------------|----------|------------|--------|-----------|----------|----------|
| Nilai Entropy total | | 100 | 44 | 56 | 0,989588 | |
| Lama Kerja | Lama | 53 | 32 | 21 | 0,9687 | 0,090977 |
| | Sebentar | 47 | 12 | 35 | 0,8196 | |
| Absen | Baik | 47 | 26 | 21 | 0,99182 | 0,084879 |
| | Cukup | 28 | 14 | 14 | 1 | |
| | Kurang | 25 | 4 | 21 | 0,63431 | |
| Skill | Baik | 27 | 21 | 6 | 0,764205 | 0,164143 |
| | Cukup | 51 | 20 | 31 | 0,966177 | |
| | Kurang | 22 | 3 | 19 | 0,574636 | |

4.5 Decision Tree (Pohon Keputusan)

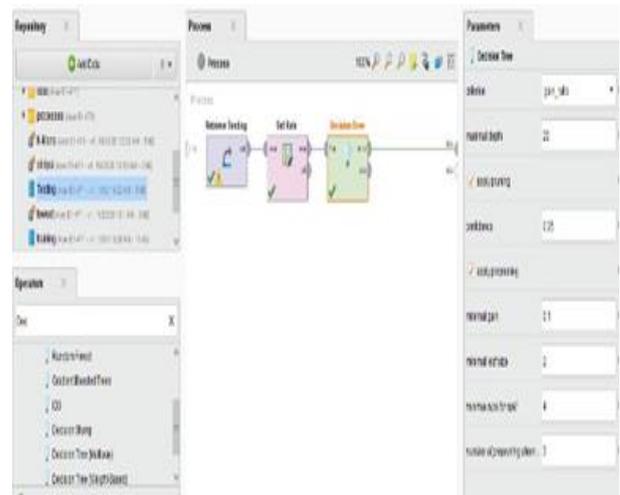


Gambar 1 Pohon Keputusan

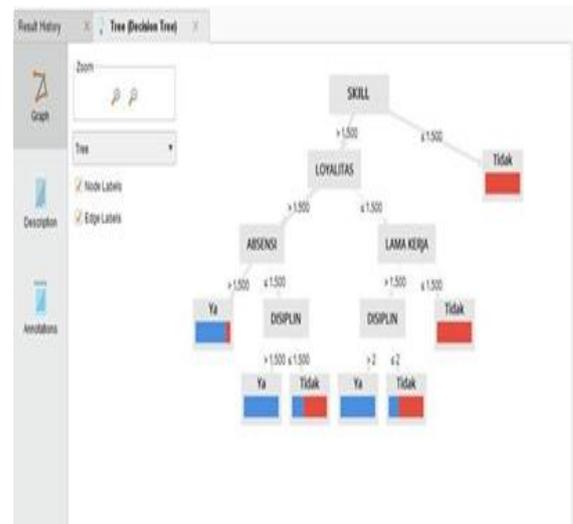
$$\begin{aligned} b_1 &= \frac{n \sum(x_1 y) - (\sum x_1)(\sum y)}{n(\sum x_2)^2 - (\sum x)^2} \\ b_1 &= \frac{100(2419) - (5402)(47)}{241900 - 253894} \\ b_1 &= \frac{30071400 - 29181604}{-11994} \\ b_1 &= \frac{889796}{-11994} \\ b_1 &= -0.01348 \end{aligned}$$

4.6 Proses Pengujian Data

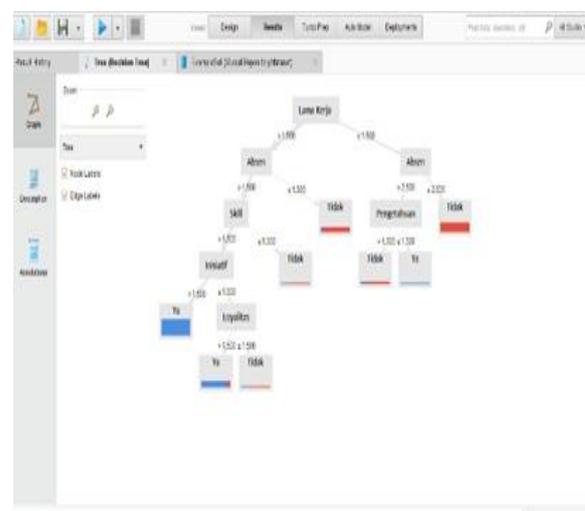
Melakukan select attributes yaitu untuk mengetahui hasil prediksi dari rapidminer, hasil perhitungan manual dan hasil uji di rapidminer.



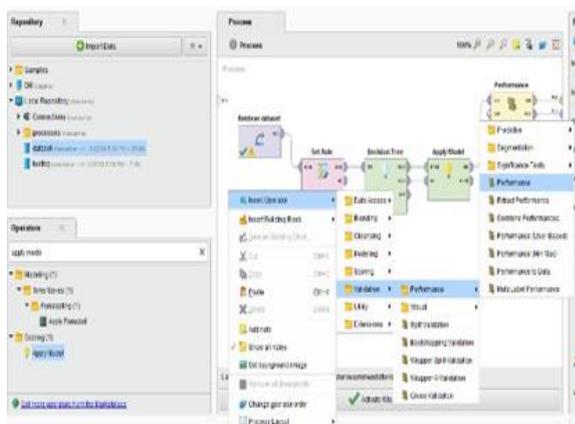
Gambar 2 Proses Pengujian



Gambar 3 Hasil Pohon Keputusan Data Testing



Gambar 4 Hasil Pohon Keputusan Data Training



Gambar 5 Setting Performance

| Index | Nominal value | Absolute count | Fraction |
|-------|---------------|----------------|----------|
| 1 | Tidak | 56 | 0.560 |
| 2 | Ya | 44 | 0.440 |

Gambar 8 Nominal Value Pengangkatan Karyawan

5. Penutup

Dari penelitian diatas yang berjudul Prediksi Pengangkatan Karyawan Dengan Metode Klasifikasi Algoritma C5.0, dapat disimpulkan bahwa dengan menggunakan klasifikasi algoritma C5.0 maka pengangkatan karyawan dapat di prediksi dengan hal tersebut dapat dijadikan kontribusi terhadap proses pengambilan keputusan bagi pihak perusahaan. Penerapan *decision tree* dapat mempercepat dan tingkat keakuratan dalam pengambilan keputusan kepada calon maupun karyawan yang sedang bekerja. Evaluasi hasil klasifikasi dalam prediksi pengangkatan karyawan menggunakan data training dan data testing dengan *Confusion Matrix* dimana data testing dengan tingkat accuracy sebesar 91.00% dan data training dengan tingkat accuracy 96.75%. Jika melihat dengan kurva ROC menggunakan data training dengan akurasi Excellent Clasification sebesar 0.989.

Daftar Pustaka

[1] Liana, W. (2020). Pengaruh Motivasi terhadap Produktivitas Karyawan PT Telkom Indonesia, Tbk Cabang Palembang. *Jurnal Nasional Manajemen Pemasaran & SDM*, 1(01), 65-72.

[2] Desi Kristanti, S. E., & Pangastuti, R. L. (2019). Kiat-kiat merangsang kinerja karyawan bagian produksi. *Media Sahabat Cendekia*.

[3] Nugraha, I. S., & Bidarti, A. (2022). Pengembangan Sumber Daya Manusia Dan Prospek Unit Pengolahan Dan Pemasaran Bokar (Uppb) Menuju Era Baru. *Jurnal Ilmu Manajemen*, 19(1), 23-29.

[4] Yogi Yusuf, W. (2007). Perbandingan Performansi Algoritma Decision Tree C5. 0, CART, dan CHAID: Kasus Prediksi Status Resiko Kredit di Bank X. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.

[5] R. R. Rerung, “Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk,” *J. Teknol. Rekayasa*, vol. 3, no. 1, p. 89, 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.

[6] Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y.,

| | True Ya | True Tidak | class precision |
|--------------|---------|------------|-----------------|
| pred Ya | 40 | 0 | 80.00% |
| pred Tidak | 4 | 50 | 92.99% |
| class recall | 91.81% | 88.28% | |

Gambar 6 Hasil Accuracy Data Training

| | True Ya | True Tidak | class precision |
|--------------|---------|------------|-----------------|
| pred Ya | 23 | 7 | 87.00% |
| pred Tidak | 4 | 18 | 81.00% |
| class recall | 87.00% | 87.78% | |

Gambar 7 Hasil Accuracy Data testing

- & Negara, E. S. (2021). Data mining dan penerapan algoritma. Yayasan Kita Menulis.
- [7] Wibowo, A., & Mahartika, I. R. (2019). Data Mining Klasterisasi dengan Algoritme K-Means untuk Pengelompokan Provinsi Berdasarkan Konsumsi Bahan Bakar Minyak Nasional. Prosiding SISFOTEK, 3(1), 87-91.
- [8] Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. Edumatic: Jurnal Pendidikan Informatika, 4(1), 37-46.
- [9] Sari, N. (2021). Penerapan Algoritma C5. 0 Dalam Pencarian Pola Hasil Audit Laporan Keuangan Perusahaan Publik (Studi Kasus: PT. Bursa Efek Indonesia). Pelita Informatika: Informasi dan Informatika, 9(4), 250-256.
- [10] Amalda, R. N., Millah, N., & Fitria, I. (2022). Implementasi Algoritma C5. 0 Dalam Menganalisa Kelayakan Penerima Keringanan Ukt Mahasiswa Itk. Teorema: Teori dan Riset Matematika, 7(1), 101-116.