

Implementasi Algoritma C4.5 Untuk Tes Kepribadian Penerimaan Karyawan Di Dinas Perhubungan Provinsi Sulawesi Tengah

Andi Oktafiqurahman¹⁾; Kusri²⁾; Asro Nasiri³⁾

^{1), 2), 3)} Magister Teknik Informatika, Universitas AMIKOM Yogyakarta

¹⁾andi.oktafiqurahman@students.amikom.ac.id; ²⁾kusri@amikom.ac.id ; ³⁾asro.nasiri@amikom.ac.id

ABSTRACT

The employee who has bad performance in working can influence the other employee to disturb many jobs. The purpose of this research is creating *prediction system which it makes easier for HRD to select prospective employees who have appropriate personality based on the position, by utilizing the C4.5 algorithm known to be the personality of employees of the Central Sulawesi Provincial Transportation Office with a level of accuracy in the implementation of the C.45 algorithm. Testing using the confusion matrix, obtained a system accuracy of 79.167%.*

Keywords: C 4.5, prediction, personality, the relationship of the Central Sulawesi province

I. Pendahuluan

Dinas Perhubungan Provinsi Sulawesi Tengah kantor ini mempunyai banyak karyawan yang berbeda-beda suku, ras, agama dan kepribadian. dan dibagi dalam bidang darat, laut, dan udara. sehingga interaksi sangat perlu dilakukan di pekerjaan ini. Di pekerjaan seperti ini interaksi sangat diperlukan interaksi karna berhubungan langsung dengan Gubernur atau Sekda. Namun terkadang manusia memiliki kepribadian yang tidak memungkinkan bekerja seperti ini karena dia merasa aneh berinteraksi dengan banyak orang sehingga ketika dia bekerja di Dinas Perhubungan, pekerjaan tidak maksimal. Di dunia psikologi, setiap orang memiliki sifat yang berbeda-beda. Ada yang memiliki sikap eksklusif dan individual. Sehingga dapat disimpulkan bahwa setiap orang memiliki perilaku yang berbeda-beda yang terbentuk dan terbangun dari perjalanan hidupnya. Setiap orang memiliki karakteristik kepribadiannya masing-masing. Sebagai contoh, ada yang memiliki kepribadian melankolis, sanguin, koleris, plegmatis, dan lain-lain. dimasa sekarang banyak teknologi yang hampir menyerupai manusia seperti *artificial intelligence* (kecerdasan buatan) banyak mesin dibuat hingga menyerupai kecerdasan manusia bahkan memiliki kelebihan yang tidak bisa dimiliki manusia sendiri. Salah satunya sistem prediksi, yah banyak teknologi yang bisa memprediksi apa yang akan dilakukan kedepannya. Mesin ini memprediksi berdasarkan data yang diperoleh dari beberapa tahun kebelakang. Banyak algoritma yang digunakan untuk membangun mesin prediksi ini. Salah satu algoritma yang digunakan adalah C4.5. Algoritma C4.5 adalah

sekumpulan algoritma *Dececion Tree*. Algoritma ini mempunyai input berupa *training sample* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *Tree* yang telah diuji kebenarannya. Pohon keputusan (*Dececion Tree*) merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan ini mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksikan aturan. Aturan dapat dengan mudah dengan menggunakan bahasa yang alami. Penelitian ini bertujuan untuk bisa mengetahui kepribadian apa yang dimiliki oleh para calon karyawan dengan menggunakan metode C4.5. Dengan cara memprediksi berdasarkan hasil kuisisioner yang diberikan kepada orang tersebut. Berdasarkan hasil yang sudah di olah maka akan didapatkan hasil atau prediksi seorang itu berkepribadian apa dan apakah cocok dengan jabatannya saat ini. Dengan mengetahui kepribadian seseorang maka kita akan lebih mudah untuk berhubungan dengan orang tersebut.

Tujuan dalam penulisan penelitian ini adalah sebagai berikut:

1. Algoritma C4.5 dipakai pada aplikasi sehingga dapat mengetahui kepribadian anggotanya atau kariawannya.
2. Menganalisa hasil prediksi dengan *confusion matrix*.

Dalam penulisan penelitian ini, batasan masalah akan dirumuskan sebagai berikut:

1. Membuat sistem prediksi kepribadian dengan jenis kepribadian Sanguin, Melankolis, Koleris, dan Plegmatis.
2. Kriteria yang digunakan seperti umur, nama, tempat tinggal, asal sekolah dan jabatan yang diinginkan.

3. Menggunakan Algoritma C4.5 untuk membuat sistem prediksi ini.
4. Output yang dihasilkan berupa prediksi kepribadian dari calon karyawan yang akan menjalani tes.

II. Tinjauan Pustaka

2.1 Penelitian Terkait

Penerapan algoritma C4.5 digunakan untuk memprediksi perilaku aparatur sipil negara dalam menempati jabatan fungsional. Untuk pengujian algoritma C4.5 menggunakan confusion matrix memperoleh hasil akurasi sebesar 92,54% dengan rasio 80% data latih dan 20 data uji. Nilai information gain yang diperoleh dari atribut Nama Pendidikan sebagai factor utama dari penentuan dalam menempati jabatan fungsional [1].

Sistem pakar digunakan untuk menganalisis kepribadian siswa menggunakan metode Certainty Factor. Untuk menganalisis kepribadian siswa menggunakan metode certainty factor dengan ditunjukkan beberapa gejala yang setiap gejala memiliki nilai keyakinan yang berbeda, dimana metode tersebut untuk membuktikan ukuran kepastian terhadap suatu aturan ataupun fakta, yang menggambarkan tingkat keyakinan seorang pakar terhadap suatu masalah. Dalam sistem pakar analisis kepribadian diri ini dengan metode precision and recall mendapatkan hasil 70% untuk tingkat precissionnya dan 100% untuk hasil recallnya. Dengan adanya sistem pakar untuk mengidentifikasi perilaku dan kepribadian siswa SMA didapatkan rata-rata waktu dalam konsultasi mengenai karakter siswa secara langsung sebelum menggunakan sistem anantara 30-60 menit, setelah menggunakan sistme rata-rata waktu yang digunakan hanya 10-20 menit sehingga sistem pakar ini dapat membantu memudahkan pendidik menangani masalah yang ada mengenai karakter siswa [2].

Algoritma C4.5 digunakan untuk menentukan calon karyawan yang memenuhi kriteria maka dibutuhkan rekomendasi yang tepat. Adapun alur yang tepat dalam proses pengerjaan penelitian ini dimulai dari pengumpulan data, kemudian setelah terkumpul dilakukan preprocessing untuk proses pengolahan data kemudian dilakukan proses perhitungan baik manual maupun menggunakan sistem. Setelah pengujian dilakukan beberapa kali maka operator yang ditetapkan adalah dengan menggunakan operator split validatin dan menggunakan stratified sampling dengan training data 90%

dan testing data 10% dengan tingkat akurasi sebesar 81,82% dan classification error sebesar 18,18% [3].

2.2 Asal Mula Kepribadian

Sebagian psikolog bersikeras bahwa kepribadian adalah masalah genetik. Sedangkan sebagian yang lain berpendapat bahwa kepribadian lebih dipengaruhi oleh cara kita dibesarkan atau cara orang tua mendidik kita serta pengalaman hidup kita sendiri. Dan ada juga psikolog berpendapat bahwa kepribadian adalah gabungan dari kedua pendapat sebelumnya. ketiga pendapat diatas kita dapat mempelajari bahwa akan terlihat aspek-aspek perilaku yang tampak muncul secara otomatis. Hal ini menyebabkan kita dapat mendeskripsikan tindakan tertentu sebagai tindakan alami manusia atau lumrah terhadap manusia. Yaitu implikasi bahwa beberapa pola perilaku berasal dari fungsi otomatis otak manusia, hal tersebut adalah implikasi bahwa beberapa pola perilaku berasal dari fungsi otomatis otak manusia. Akan tetapi, tidak seperti binatang. Bahkan sekitar tahun 1950-an psikolog bernama Sheldon mengelompokkan orang ke dalam tiga tipe sesuai bentuk tubuh dan penampilan fisik mereka, diantaranya adalah [5].

1. *Endomorphs*. Adalah jenis perawakan yang ditandai dengan lebih menonjolnya jaringan lemak dan kadang-kadang bersifat bawaan atau keturunan.
2. *Ectomorphs*. Adalah jenis perawakan yang ditandai dengan lebih menonjolnya jaringan-jaringan kulit ari pada orang yang tinggi kurus.
3. *Mesomorphs*. Adalah jenis perawakan yang ditandai dengan lebih unggulnya jaringan-jaringan tubuh yang berasal dari otot.

2.3 Jenis-Jenis Kepribadian

Kita bisa mengenal kepribadian apa yang kita miliki lewat berbagai macam ter kepribadian. Baik melalui tes langsung atau menggunakan tes psikolog yang lainnya. Salah satunya adalah tes *personality plus* yang mengategorikan manusia kedalam empat karakter diantaranya Sanguinis, Koleris, Melankolis, Dan Plegmatis [4].

1. Sanguin

Kepribadian ini adalah kepribadian paling populer dari kepribadian-kepribadian yang lainnya. Biasanya orang yang memiliki kepribadian ini dapat berkomunikasi dengan baik dan mengasikkan dan ini merupakan nilai tambah yang membuat orang lain iri hati, tetapi

jika dibawa sampai ke ujung yang ekstrim orang dengan kepribadian ini dapat berbicara secara terus-menerus, memonopoli, menyela, dan menyimpang terlalu jauh dari kebenaran.

2. Koleris

Kepribadian koleris yang handal dengan kepemimpinan yang cepat dan tajam sangat diperlukan dalam setiap tahap kehidupan pada zaman sekarang. Namun jika kita melihat ke titik ekstrim, orang dengan kepribadian koleris dia mempunyai sifat sok berkuasa, mendominasi, dan manipulatif.

3. Melankolis

Pemikiran analitis yang mendalam dari orang melankolis yang sempurna merupakan ciri khas yang jenius yang dimiliki orang dengan kepribadian ini, banyak dihormati oleh orang yang memiliki pikiran yang dangkal, namun jika kita melihat dari sudut terburuknya, orang dengan kepribadian melankolis ini adalah tidak pandai dalam bergaul sehingga membuat orang lain merasa bosan terhadap dirinya, dikarenakan semua hal dikerjakan secara teliti dan teratur, segala sesuatu dikerjakan dengan langkah-langkah yang jelas dan terstruktur. Kalau lama dalam proses pengerjaan bukan karena malas tapi karena hasilnya harus sempurna sesuai keinginannya

4. Plegmatis

Sifat orang yang berkepribadian plegmatis ini adalah damai dan mudah bergaul merupakan perpaduan yang mengagumkan dan menjadikannya orang yang paling disukai dalam kelompok mana saja, namun kalau dibawah sampai ke titik ekstrim, orang plegmatis yang damai tidak pernah peduli melupakan apapun, masa bodoh, dan tidak punya kepastian. Ciri-ciri dari orang yang memiliki kepribadian plegmatis yang mudah dilihat adalah tidak mempunyai sikap yang jelas dalam keputusan, tidak mau secara tegas memihak sesuatu yang belum diketahuinya.

2.4 Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk membuat atau membentuk pohon keputusan (*decision tree*).

2.5 Pohon Keputusan

Menurut [6], Pohon keputusan merupakan metode prediksi dan klasifikasi yang sangat terkenal dan kuat. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi

kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain (Berry & Linoff, 2004) Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya.

2.6 Entropy

Sebuah obyek yang diklasifikasi dalam pohon keputusan harus diuji nilai entropinya. *Entropy* adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari *impurity* dan *homogeneity* dari kumpulan data [7]. *Entropy* digunakan untuk memprediksi rata-rata jumlah bit per symbol yang diperlakukan untuk mentransmisikan data. *Entropy* juga merupakan jumlah bit yang diperkirakan dibutuhkan untuk dapat mengekstrak suatu kelas (+ atau -) dari beberapa data acak pada ruang sampel. Untuk perhitungan entropi dapat dilihat pada persamaan 1 sebagai berikut [6].

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \tag{1}$$

Dengan penjelasan dari persamaan diatas kita bisa lihat pada tabel 1 dibawah ini.

Tabel 1. Penjelasan Dari Persamaan 1

| Symbol | Quantity |
|--------|--------------------------------|
| S | Himpunan Data Uji |
| n | Banyaknya partisi S |
| pi | Rasio dari kelas S1 terhadap S |

2.7 Gain

Setelah menemukan nilai *entropy* untuk suatu kumpulan data, maka kita dapat mengukur efektivitas suatu atribut dalam mengklasifikasikan data. Ukuran efektifitas ini disebut *information gain*. Secara matematis, *information gain* dari suatu atribut A, rumusnya didefinisikan pada persamaan 2 [6].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \tag{2}$$

Untuk penjelasan dari persamaan 2 dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Penjelasan dari persamaan 2

| Symbol | Quantity |
|--------|----------|
|--------|----------|

| | |
|--------------|--|
| S | Entropy dari data training |
| A | Atribut dari data training |
| <i>n</i> | Jumlah kasus pada partisi ke- <i>i</i> |
| S1 | Jumlah kasus pada partisi ke- <i>i</i> |
| S | Jumlah kasus dalam dalam S |
| Entropy (S1) | Entropy untuk sample-sample yang memiliki nilai <i>i</i> |

2.8 Confusion Matrix

Confusion matrix adalah sebuah metode untuk evaluasi yang menggunakan tabel matrix [13], dapat dilihat Tabel 3.

Tabel 3. Confusion Matrix

| Klasifikasi Benar | Diklasifikasikan Sebagai | |
|-------------------|--------------------------|---------------|
| | + | - |
| + | Benar Positif | Salah Negatif |
| - | Salah Positif | Benar Negatif |

Pada Tabel 3 diatas, diperlihatkan metode pengujian menggunakan confusion matrix menghasilkan nilai accuracy. Nilai accuracy adalah presentase jumlah record data yang diklasifikasikan secara benar oleh salah satu algoritma. Pada tabel confusion matrix diatas, Benar Positif adalah jumlah record positif yang diklasifikasikan sebagai positif, dan Salah Positif adalah jumlah record negatif yang diklasifikasikan sebagai positif, Salah Negatif adalah jumlah record positif yang diklasifikasikan negatif, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam confusion matrix, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah akurasi.

III. Metode Penelitian

3.1 Tahapan Penelitian

Pada penelitian ini, penulis melakukan tahapan-tahapan penelitian sebagai berikut:

1. Metode Analisis

Metode Analisis yang digunakan penelitian untuk menganalisa adalah Analisa fungsional dan non fungsional. Analisa fungsional adalah analisa untuk mengetahui apa saja hal yang bisa dilakukan oleh sistem. Sedangkan untuk analisa non fungsional adalah analisa untuk mengetahui spesifikasi yang dibutuhkan oleh sistem.

2. Metode Perancangan

Metode perancangan adalah metode yang digunakan untuk membangun sebuah sistem prediksi ini yaitu menggunakan Flowchart, Data Flow Diagram (DFD) dan Normalisasi Database.

3. Metode Implementasi

Metode ini dilaksanakan dengan melakukan implementasi rancangan sistem ke dalam Bahasa pemrograman PHP dan Database MySql Server.

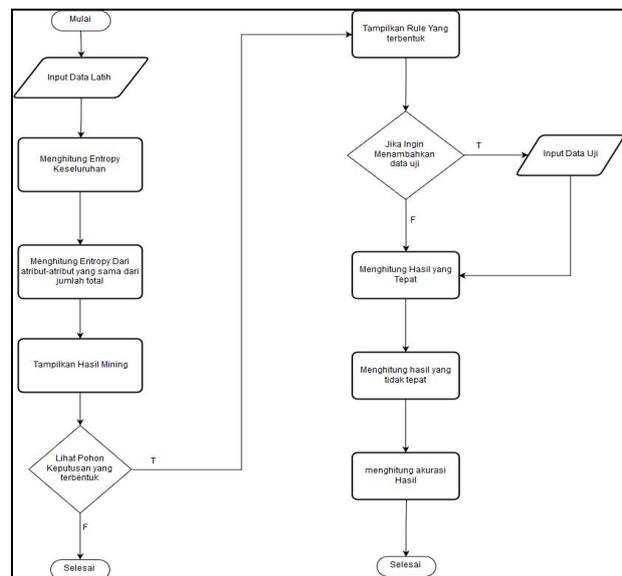
4. Metode Pengujian

Metode ini adalah metode yang paling akhir dari semua metode penelitian yang digunakan pada penelitian ini. Metode adalah suatu proses pengujian sistem prediksi yang telah di buat. Metode yang digunakan adalah metode confusion matrix untuk menghitung nilai akurasi dari hasil pengujian.

IV. HASIL DAN PEMBAHASAN

4.1. Flowchart

Proses kerja dari C4.5 dan proses menguji akurasi hasil klasifikasi. Seperti yang dapat dilihat pada Gambar 1 dibawah ini.



Gambar 1. Flowchart

Adapun penjelasan dari flowchart pada Gambar 1 sebagai berikut:

a. Input data

Keseluruhan data yang telah diinputkan pada databse, dan akan digunakan dalam proses C4.5 sebanyak 60 dataset.

b. Menghitung Entropy keseluruhan

Menghitung entropy keseluruhan, jumlah entropy keseluruhan berdasarkan klasifikasi yang dibentuk dari data training. Jumlah kelas yang terdapat pada data set yaitu 4 kelas. Diantaranya jumlah sanguin, Jumlah koleris, jumlah melankolis, jumlah plegmatis.

Perhitungan entropy keseluruhan menggunakan persamaan (1). Dimana jumlah data sebanyak (S) = 42, jumlah sanguin (P1)=

17, jumlah koleris (P2)= 6, jumlah melankolis (P3) = 5, jumlah plegmatis (P4)= 14.

$$\begin{aligned}
 entropy(Total) = & \left(-\frac{17}{42} * \log_2 \left(\frac{17}{42} \right) \right) \\
 & + \left(-\frac{6}{42} * \log_2 \left(\frac{6}{42} \right) \right) \\
 & + \left(-\frac{5}{42} * \log_2 \left(\frac{5}{42} \right) \right) \\
 & + \left(-\frac{14}{42} * \log_2 \left(\frac{14}{42} \right) \right)
 \end{aligned}$$

$$Entropy(Total) = 1.823$$

c. Menghitung Gain

Menghitung nilai entropy dari atribut-atribut yang sama dari jumlah kepribadian (jumlah sanguin, Jumlah koleris, jumlah melankolis, jumlah plegmatis) Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada.

- d. Setelah selesai system akan menampilkan hasil dari proses mining tersebut.
- e. Kemudian akan ada link yang akan merujuk ke form untuk proses melihat rule pohong keputusan yang terbentuk dari hasil *mining*.
- f. Jika pilihan ya maka akan di alihkan ke halaman rule untuk menguji ke akuratan.
- g. Jika pilihan tidak maka proses C4.5 selesai.
- h. Sistem akan menampilkan rule yang terbentuk dalam proses C4.5 untuk menguji ke akuratan dari system.
- i. Selanjutnya akan menampilkan data uji, akan terdapat pilihan ingin menambahkan data uji, jika TRUE maka akan melakukan proses penginputan data uji, jika FALSE maka akan dilakukan proses selanjutnya
- j. Menghitung jumlah jawaban yang bernilai benar
- k. Menghitung jumlah total jawaban yang bernilai salah, atau datanya kosong
- l. Kemudian menghitung akurasi menggunakan rumus persamaan (4).
- m. Setelah menemukan hasil akurasi maka proses akan selesai

Tabel 4. Pengujian Keakuratan

| No | Hasil Sistem | Data | Kesesuaian |
|----|--------------|------------|--------------|
| 1 | Sanguin | Sanguin | Sesuai |
| 2 | Melankolis | Melankolis | Sesuai |
| 3 | Plegmatis | Koleris | Tidak Sesuai |
| 4 | Plegmatis | Plegmatis | Sesuai |
| 5 | Plegmatis | Koleris | Tidak Sesuai |
| 6 | Sanguin | Sanguin | Sesuai |
| 7 | Melankolis | Plegmatis | Tidak Sesuai |

| No | Hasil Sistem | Data | Kesesuaian |
|----|--------------|------------|--------------|
| 8 | Sanguin | Sanguin | Sesuai |
| 9 | Plegmatis | Plegmatis | Sesuai |
| 10 | Plegmatis | Melankolis | Tidak Sesuai |
| 11 | Melankolis | Melankolis | Sesuai |
| 12 | Melankolis | Plegmatis | Tidak Sesuai |
| 13 | Plegmatis | Plegmatis | Sesuai |
| 14 | Koleris | Koleris | Sesuai |
| 15 | Plegmatis | Plegmatis | Sesuai |
| 16 | Koleris | Koleris | Sesuai |
| 17 | Melankolis | Melankolis | Sesuai |
| 18 | Plegmatis | Plegmatis | Sesuai |
| 19 | Koleris | Koleris | Sesuai |
| 20 | Sanguin | Sanguin | Sesuai |
| 21 | Sanguin | Sanguin | Sesuai |
| 22 | Plegmatis | Plegmatis | Sesuai |
| 23 | Koleris | Koleris | Sesuai |
| 24 | Plegmatis | Plegmatis | Sesuai |

Dari Tabel 4 diperoleh data diklasifikasikan salah sebanyak 5 data yang tidak benar dengan total data sebanyak 24 data. Maka dapat kita hitung sebagai berikut.

$$Accuracy = \frac{19}{24} \times 100\% = 79.167\%$$

Dari hasil pengujian *confussion Matrix*, diperoleh akurasi kecocokan kelas sebenarnya terhadap kelas prediksi sebesar 79.167% dan ketidakcocokan kelas prediksi yang diperoleh sebesar 20.833%.

V. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan, maka kesimpulan yang diambil adalah sebagai berikut:

1. Proses C4.5 dengan total 60 data yang menghasilkan data sebanyak (S) = 42, jumlah sanguin (P1)= 17, jumlah koleris (P2)= 6, jumlah melankolis (P3) = 5, jumlah plegmatis (P4)= 14 dengan nilai akurasi gain sebesar 75%.
2. Proses pengujian dilakukan dengan menggunakan *confusion matrix* berdasarkan data uji kepribadian karyawan. Metode C4.5 berhasil mengklasifikasi 19 data dari 24 data yang diuji sehingga metode C4.5 berhasil memprediksi tingkat keakuratan dengan akurasi sebesar 79.167%.

5.2. Saran

Saran untuk penelitian ini, adalah:

- Menggunakan *k-fold cross validation* dengan angka perbandingan yang berbeda sehingga bisa membandingkan akurasi yang dihasilkan.

- Dapat menggunakan metode data mining yang lain untuk melakukan prediksi kepribadian.

Daftar Pustaka

- [1] M. Isra, "Analisis Dan Prediksi Perilaku Aparatur Sipil Negara Dalam Menempati Jabatan Fungsional Menggunakan Algoritma C4.5", Jurnal Informasi Dan Teknologi, Vol 4.1, 2022.
- [2] A. P. Nurjannah dan F. Asharudin, "Sistem Pakar Untuk Mengidentifikasi Perilaku dan Kepribadian Siswa Menggunakan Metode *Certainty Factor* Pada SMA Sunan Kalijogo", Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM), Vol 3(2), 2022.
- [3] A. M. Husein and M. Brutu, "Prediksi Penerimaan Calon Karyawan Dengan menggunakan algoritma C4.5 Pada Biro Kesejahteraan Rakyat provinsi Sumatera Utara," Digital Transformation Technology, vol. 2, no. 1, pp. 16–20, 2022.
- [4] Sudaryo, "Interpersonal Skill: Kecakapan Antarpersonal", Prenada Media, Jakarta, 2022.
- [5] Irwan "Etika dan Perilaku Kesehatan" Absolute Media, Yogyakarta, 2020.
- [6] Yendrizar. "Monograf Algoritma C4.5 Pada Teknik Klasifikasi Penyusutan Volume Pupuk" CV AZKA PUSTAKA, Sumatra Barat, 2022.
- [7] A. Wanto, "Data Mining: Algoritma dan Implementasi", Yayasan Kita Menulis, Kota Medan, 2020.