

PENERAPAN ALGORITMA C4.5 UNTUK PENERIMAAN CALON PEGAWAI NEGERI SIPIL KEMENKUMHAM SULAWESI TENGGERA

Abdul Fath Ramadhan¹, Rizal Adi Saputra², L.M. Fid Aksara³

¹Program Studi Teknik Informatika, Universitas Halu Oleo

abdulfathramadhan@gmail.com

Abstrak

Penelitian ini menerapkan algoritma Decision Tree C4.5 dalam seleksi Calon Pegawai Negeri Sipil (CPNS) Kemenkumham Sulawesi Tenggara. Melibatkan Data Mining, metode ini memanfaatkan preprocessing data, pemilihan atribut, dan pengembangan model Decision Tree. Algoritma C4.5, dipilih karena kemudahan dan visualisasi, diaplikasikan pada data pelamar CPNS, termasuk SKD seperti TWK, TIU, dan TKP. Kontribusi utama adalah peningkatan efisiensi dan ketepatan seleksi CPNS, memastikan pegawai sesuai kebutuhan instansi. Visualisasi pohon keputusan dari C4.5 memberikan pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi keputusan model.

Kata Kunci: *algoritma C4.5, Kementerian Hukum dan Hak Asasi Manusia, Penerimaan CPNS, Pohon Keputusan, Sulawesi Tenggara*

Abstract

This research applies the Decision Tree C4.5 algorithm in the selection process of Civil Servant Candidates (CPNS) at the Ministry of Law and Human Rights (Kemenkumham) in Southeast Sulawesi. Involving Data Mining, the method utilizes data preprocessing, attribute selection, and the development of the Decision Tree model. The choice of the C4.5 algorithm, known for its simplicity and visualization capabilities, is applied to CPNS applicant data, including Basic Competency Test (SKD) elements such as TWK, TIU, and TKP. The primary contribution lies in enhancing the efficiency and accuracy of CPNS selection, ensuring that recruited employees align with institutional needs. Visualization of the decision tree from C4.5 provides a better understanding of the factors influencing the model's decisions.

Keywords: *c4.5 algorithm, Ministry of Law and Human Rights, CPNS recruitment, Decision Tree, Southeast Sulawesi.*

1. Pendahuluan

Penerimaan Calon Pegawai Negeri Sipil (CPNS) di Kementerian Hukum dan Hak Asasi Manusia (Kemenkumham) Sulawesi Tenggara memiliki peran strategis dalam pengelolaan sumber daya manusia. [1]Proses penerimaan ini harus memastikan bahwa pegawai yang direkrut memiliki kompetensi dan kualifikasi yang sesuai dengan kebutuhan

instansi. Badan Kepegawaian Daerah memiliki tanggung jawab untuk melaksanakan proses seleksi dan penempatan pegawai sesuai dengan syarat dan prosedur yang telah ditetapkan. Namun, beberapa tantangan masih dihadapi, seperti pemilihan pegawai yang tidak sesuai dengan kemampuan dan bakat mereka.[3] Dalam konteks ini, penelitian ini menggunakan algoritma Decision Tree C4.5 untuk

membantu menentukan hasil penerimaan CPNS dengan lebih akurat. Algoritma ini memanfaatkan atribut seperti hasil Seleksi Kompetensi Dasar (SKD) diantaranya, Tes Wawasan Kebangsaan (TWK), Tes Karakteristik Pribadi (TKP), dan Tes Intelegensi Umum (TIU) dan atribut lulus seleksi yaitu Diterima atau Gagal sebagai keterangan.

Metode penelitian melibatkan proses Data Mining, dengan algoritma C4.5 sebagai fokus utama. Algoritma ini dipilih karena kelebihan yang meliputi kemudahan pemahaman, fleksibilitas, dan visualisasi yang dapat direpresentasikan dalam bentuk pohon keputusan. Proses ini melibatkan langkah-langkah seperti preprocessing data untuk mempersiapkan dataset, pemilihan atribut yang relevan, dan pengembangan model Decision Tree. [6]

2. Metode Penelitian

Data mining

Data mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Data mining merupakan istilah yang sering dikatakan sebagai suatu cara untuk menguraikan serta mencari penemuan berupa pengetahuan didalam suatu database. Data mining adalah proses pemilihan atau “menambang” pengetahuan dari sekumpulan data dalam jumlah yang banyak. [9]

Teknik Klasifikasi

Klasifikasi adalah salah satu bentuk dari teknik atau metode data mining yang termasuk dalam kategori predictive mining yaitu suatu teknik yang dapat digunakan untuk meramalkan atau memprediksi kecenderungan data di masa depan. Proses yang terjadi dalam klasifikasi adalah proses penggolongan data ke dalam variabel target atau variabel tujuan dengan membangun sebuah model penyelesaian dengan memperhatikan atribut yang paling berpengaruh.

Klasifikasi masuk ke dalam supervised induction, dimana pengujian yang

memanfaatkan kumpulan pengujian dari record dan atribut yang terklasifikasi untuk menentukan output dan kelas tambahan. Salah satu contoh algoritmanya adalah decision tree yang terkenal dan mudah dalam implementasinya kedalam bentuk grafik. Komponen-komponen utama dari proses klasifikasi antara lain :

1. Kelas, merupakan variable tidak bebas yang merupakan label dari hasil klasifikasi.
2. Prediktor, merupakan variable bebas suatu model berdasarkan dari karakteristik atribut data yang diklasifikasi.
3. Set data pelatihan, merupakan sekumpulan data lengkap yang berisi kelas dan predictor untuk dilatih agar model dapat mengelompokkan ke dalam kelas yang tepat. Contohnya adalah grup pasien yang telah di-test terhadap serangan jantung, grup 4 pelanggan di suatu supermarket, dan sebagainya.
4. Set data uji, berisi data-data baru yang akan dikelompokkan oleh model guna mengetahui akurasi dari model yang telah dibuat.

Algoritma C4.5

Algoritma pohon keputusan yang terkenal adalah C4.5. Pada akhir tahun 1970 sampai awal tahun 1980 J. Ross Quinlan, seorang peneliti di bidang machine learning, membuat sebuah algoritma Decision Tree yang dikenal dengan ID3 (Iterative Dichotomiser). Quinlan kemudian mengembangkan algoritma ID3 menjadi algoritma C4.5 yang merupakan penyempurnaan algoritma sebelumnya. Algoritma ini memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan). Algoritma C4.5 membuat pohon keputusan dari node atas hingga cabang terakhirnya, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun. Karena algoritma C4.5 digunakan untuk melakukan klasifikasi, jadi hasil dari pengolahan test dataset berupa pengelompokkan data ke dalam kelas kelasnya. [2]

Rumus menghitung entropy pada algoritma C4.5 adalah sebagai berikut :

$$\text{Entropy}(S) = \sum_{i=0}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan :

S : himpunan (dataset) kasus

K : banyaknya partisi S

Pi : probabilitas yang didapat dari Sum(Ya), atau Sum(Tidak) dibagi total kasus.

Setelah mendapatkan entropy dari keseluruhan kasus, lakukan analisis padasetiap atribut dan nilai nilainya dan kemudian hitung entropinya. Kemudian hitung Gain dengan Rumus sebagai berikut :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum n A_i * \text{Entropy}(A) \quad (2)$$

S : himpunan kasus

|Si| : jumlah kasus pada partisi ke-i

A : Atribut

|S| : jumlah kasus dalam S

Langkah algoritma C4.5 untuk membangun pohon keputusan secara umum adalah sebagai berikut :

1. Memilih atribut paling atas sebagai akar.
2. Pilih cabang untuk setiap nilai.
3. Tiap cabang bagi kasus nya.
4. Ulangi proses pada setiap cabang hingga semua kasus yang ada pada cabang mempunyai kelas yang sama.

Decision Tree

Decision Tree (pohon keputusan) adalah salah satu metode dalam data mining yang digunakan untuk klasifikasi dan prediksi. Decision Tree memodelkan keputusan berdasarkan serangkaian aturan keputusan yang diperoleh dari fitur-fitur data. Model ini menggambarkan struktur pohon, di mana setiap node internal mewakili keputusan berdasarkan suatu atribut, setiap cabang mewakili hasil keputusan, dan setiap daun menyatakan kelas atau nilai prediksi [8]

Preprocessing Data

Persiapan atau Preprocessing Data berguna untuk manipulasi dataset sebelum digunakan pada model. Beberapa persiapan data yang dilakukan adalah pembersihan data (Data Cleaning) dengan menghilangkan redundansi data dan memperbaiki data yang tidak konsisten. Pemilihan data dari

banyaknya jenis formasi yang ada pada data seleksi CPNS kemenkumham SULTRA yang akan diujikan menggunakan Data Mining [7]. Transformasi dataset menjadi bentuk yang lebih cocok untuk Data Mining agar lebih mudah untuk mendapatkan pengetahuan dari dataset.

Data Pelamar CPNS

Data yang digunakan berasal dari Instansi Pemerintah Badan Kepegawaian Negara daerah Sulawesi Tenggara, yang mencakup 3848 data pelamar CPNS. Atribut yang dijadikan sebagai acuan adalah Hasil Kompetensi Dasar (SKD) diantaranya, Tes Wawasan Kebangsaan (TWK), Tes Karakteristik Pribadi (TKP), dan Tes Intelegensi Umum (TIU) dan atribut lulus seleksi yaitu Diterima atau Gagal sebagai keterangan.

Tabel 1. Atribut Data

No	Atribut	Type Data
1	No Peserta	16 Digit Number
2	Nama	String
3	Pendidikan	Alphanumeric
4	TWK	Number
5	TIU	Number
6	TKP	Number
7	Total	Number
Keterangan		P/L, P, TL, TH

Pengolahan Data

Data kemudian dilakukan pemilihan atribut dan sebagian dari data dalam atribut yang ada akan dikonversikan untuk memudahkan proses data mining, karena data akan diproses dengan tools bantu data mining.[4]

Tabel 2. Seleksi Atribut

Atribut	Detail Penggunaai	
NIK	×	No
No Register	×	No
No Peserta	√	Nilai Unique
Nama	√	Nilai Unique
Jenis Kelamin	×	No
Tempat Lahir	×	No
Tanggal Lahir	×	No
Kode Pos	×	No
Provinsi	×	No
Kota	×	No

Telepon	×	No
Email	×	No
Asal Instansi	×	No
No Ijasah	×	No
Akreditasi	×	No
Nilai IPK	×	No
Kode Jabatan	×	No
Nama Jabatan	×	No
Kode Pendidikan	×	No
Nama Jabatan	×	No
Kode Pendidikan	×	No
Nama Pendidikan	×	No
TWK	√	Nilai Model
TIU	√	Nilai Model
TKP	√	Nilai Model
Hasil Seleksi	√	Label Target

Pengembangan Model

Model yang dikembangkan menggunakan algoritma C4.5 Decision Tree.[4] Struktur Decision Tree dan nilai akurasi akan didapatkan setelah dataset diproses. Nilai performa diukur dengan confusion matrix.[5]

3. Hasil dan Pembahasan

Data yang digunakan untuk pengklasifikasian hasil tes Seleksi Kompetensi Dasar (SKD) CPNS digunakan berasal dari Instansi Pemerintah Badan Kepegawaian Negara daerah Sulawesi Tenggara sebanyak 3848 data pelamar CPNS dengan 8 atribut. Setelah itu, beberapa library utama seperti pandas, scikit-learn, dan matplotlib diimpor. Library-library ini nantinya akan digunakan dalam proses analisis data serta pembangunan model prediksi. Hasilnya, DataFrame menggambarkan data yang telah dimodifikasi sesuai dengan kebutuhan analisis. [10]

Pada tahap pembangunan model Decision Tree, library scikit-learn digunakan dengan parameter default untuk membuat model yang dapat memprediksi status keterangan para pelamar CPNS berdasarkan fitur tertentu. Proses ini melibatkan pelatihan model menggunakan data latih, sehingga model dapat belajar pola-pola yang ada dalam dataset.

Setelah model dibangun, langkah selanjutnya adalah melakukan prediksi terhadap data uji. Pertama, matrix kecacauan digunakan untuk menunjukkan seberapa baik model dapat membedakan kelas "P/L", "P",

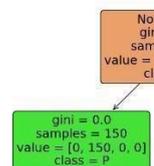
"TL", dan "TH" pada data uji. Selanjutnya, laporan klasifikasi diberikan untuk memberikan informasi lebih lanjut tentang ketepatan, recall, dan skor f1-untuk setiap kelas.

Selain itu, tingkat akurasi juga dihitung sebagai ukuran keseluruhan dari kemampuan model Decision Tree untuk memberikan suatu sebagai ukuran keseluruhan dari kemampuan memberikan prediksi yang dapat diandalkan tentang status Keterangan para pelamar CPNS. Tingkatakurasi menunjukkan persentase dari prediksi yang tepat dibandingkan dengan total jumlah data uji.

Tabel 3. Hasil pemodelan algoritma C4.5

Tingkat akurasi algoritma C4.5				
Kelas	precision	recall	f1-score	support
0	1.00	1.00	1.00	156
1	1.00	1.00	1.00	45
2	1.00	1.00	1.00	136
3	1.00	1.00	1.00	433
accuracy			1.00	770
macro avg	1.00	1.00	1.00	770
weighted avg	1.00	1.00	1.00	770
Tingkat akurasi : 100%				

Visualisasi Pohon Keputusan: Library plot_tree dari scikit-learn digunakan untuk menampilkan pohon keputusan yang telah dibangun dalam bentuk grafik. Pohon ini memberikan pemahaman tentang elemen yang memengaruhi keputusan model.



Gambar 1 Hasil Pohon Keputusan

4. Kesimpulan Dan Saran

Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma Decision Tree C4.5 untuk meningkatkan akurasi dan objektivitas dalam proses penerimaan Calon

Pegawai Negeri Sipil (CPNS) di Kementerian Hukum dan Hak Asasi Manusia Sulawesi Tenggara. Dengan menggunakan atribut seperti Tes Wawasan Kebangsaan, Tes Karakteristik Pribadi, dan Tes Intelegensi Umum, model Decision Tree yang dikembangkan untuk menunjukkan efektivitas dalam menilai kualifikasi pelamar dengan tingkat akurasi 100%.

Hasil menunjukkan bahwa Model ini Decision Tree C4.5 yang dikembangkan berhasil memprediksi hasil tes SKD CPNS dengan tingkat akurasi yang tinggi. Hal ini menunjukkan bahwa algoritma ini dapat menjadi alat yang efektif dalam mendukung keputusan penerimaan CPNS. Atribut yang digunakan dalam model, seperti Tes Wawasan Kebangsaan, Tes Karakteristik Pribadi, dan Tes Intelegensi Umum, membuktikan pentingnya faktor-faktor ini dalam menilai kualifikasi pelamar. Pemilihan atribut yang tepat berkontribusi pada keberhasilan model.

Penggunaan teknik Data Mining, khususnya algoritma Decision Tree, membantu dalam menggali pengetahuan dari dataset CPNS. Proses ini memungkinkan peneliti untuk membuat keputusan yang lebih informasional dan tepat.

Saran

1. Perbaikan Data: Diperlukan perhatian khusus pada perbaikan dan kebersihan data. Proses preprocessing data perlu dipertimbangkan lebih lanjut untuk meminimalkan potensi kesalahan dan memastikan kualitas data yang baik.
2. Pengembangan Model: Terus mengembangkan model dengan memperbarui data pelamar dan melibatkan atribut yang relevan. Penelitian lanjutan dapat mempertimbangkan penambahan atribut atau memperluas cakupan data untuk meningkatkan ketepatan model.
3. Transparansi Model: Memastikan transparansi dalam keputusan yang dihasilkan oleh model. Penjelasan yang jelas tentang faktor-faktor yang memengaruhi keputusan dapat membantu membangun kepercayaan dalam penggunaan model ini. Dengan implementasi saran-saran ini,

diharapkan penelitian ini dapat memberikan kontribusi lebih lanjut dalam meningkatkan efisiensi, objektivitas, dan transparansi dalam proses penerimaan CPNS di Kementerian Hukum dan Hak Asasi Manusia Sulawesi Tenggara..

Daftar Pustaka

- I. S. Sunandar, F. F. Nugraha, and C. Juliane, "Penerapan Data Mining dengan Metode Klasifikasi Menggunakan Algoritma C4.5," vol. 9, no. 4, pp. 2862-2869, 2022, doi: 10.35957/jatisi.v9i4.2399
- F. Ferdian Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," vol. 3, no. 2, pp. 95-103, 2017, doi: 10.35957/jatisi.v3i2.71.
- A. Oktafiqurahman, K. Kusriani, and A. Nasiri, "Implementasi Algoritma C4.5 untuk Tes Kepribadian Penerimaan Karyawan di Dinas Perhubungan Provinsi Sulawesi Tengah," vol. 11, no. 1, pp. 11-16, 2023, doi: 10.30646/tikomsin.v11i1.719.
- H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *Journal of Information Systems Engineering and Business Intelligence*, vol. 7, no. 1, p. 67, Apr. 2021, doi: 10.20473/jisebi.7.1.67-73.
- R. Puspita, S. Putri, I. Waspada, and U. D. Semarang, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 1-7, 2018, doi: 10.23917/khif.v4i1.5975.
- A. M. Husein, and M. Brutu, "Prediksi Penerimaan Calon Karyawan dengan Menggunakan Algoritma C4.5 pada Biro Kesejahteraan Rakyat Provinsi Sumatra Barat," vol. 2, no. 1, pp. 16- 20, 2022, doi: 10.47709/digitech.v2i1.1769.
- M. Idris and J. E. Suseno, "Implementation of C4.5 Algorithm and Forward Chaining Method for Higher Education Performance Analysis", vol. 125, pp. 1-5, 2019, doi: 10.1051/e3sconf/201912521002.
- Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," vol. 2, no. 2, pp. 213-219, 2016, doi : 10.22202/ei.2016.v2i2.1465.

- [9] R. D. Syah, “Metode Decision Tree Untuk Klasifikasi Hasil Seleksi Kompetensi Dasar Pada Cpns 2019 Di Arsip Nasional Republik Indonesia,” *Jurnal Ilmiah Informatika Komputer*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.
- [10] E. Gustiarto, “Penerapan Algoritma C4.5 dalam Memprediksi Calon Siswa Baru,” vol. 3, no.4, pp. 242-248, 2023, doi: 10.30998/jrkt.v3i04.9930.