

Model Klasifikasi Kenaikan Pangkat Pegawai Negeri Sipil Menggunakan Decision Tree C4.5

Kamaruddin^{1,*}, Suman Wijaya², Mashud³, Sitti Aisa⁴

¹ Teknologi Informasi, Universitas Teknologi Akba Makassar, Makassar, Indonesia

² Teknik Informatika, Universitas Teknologi Akba Makassar, Makassar, Indonesia

³ Bisnis Digital, Universitas Teknologi Akba Makassar, Makassar, Indonesia

⁴ Teknik Informatika, Universitas Dipa Makassar, Makassar, Indonesia

Email: ^{1*}kamaruddin@unitama.ac.id, ²suman22@mhs.unitama.ac.id, ³mashud@akba.ac.id, ⁴sittiaila28@undipa.ac.id

Email Penulis Korespondensi: Kamaruddin@unitama.ac.id

Abstrak—Proses kenaikan pangkat Pegawai Negeri Sipil (PNS) merupakan bagian penting dalam manajemen sumber daya manusia aparatur yang bertujuan untuk meningkatkan kinerja dan profesionalisme. Namun, dalam praktiknya proses tersebut masih menghadapi permasalahan berupa subjektivitas penilaian serta rendahnya efisiensi dalam pengolahan data. Penelitian ini bertujuan untuk membangun model klasifikasi kelayakan kenaikan pangkat reguler PNS menggunakan algoritma Decision Tree C4.5 berbasis data historis kepegawaian. Data yang digunakan merupakan data sekunder sebanyak 6.193 record ASN yang diperoleh dari Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Kabupaten Pangkajene dan Kepulauan. Atribut yang digunakan meliputi masa kerja golongan, nilai kinerja, dan jenis kepegawaian. Tahapan penelitian meliputi preprocessing data, seleksi atribut, pembentukan model, serta evaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model yang dihasilkan memiliki tingkat akurasi sebesar 95,4%, dengan nilai *precision* 87,6%, *recall* 83,7%, dan *F1-score* 85,6%. Model juga mampu menghasilkan aturan keputusan yang mudah diinterpretasikan. Dengan demikian, pendekatan data mining menggunakan algoritma C4.5 terbukti efektif dalam meningkatkan objektivitas, transparansi, dan efisiensi dalam proses pengambilan keputusan kenaikan pangkat PNS..

Kata Kunci: Data Mining; Decision Tree C4.5; Klasifikasi; Kenaikan Pangkat; PNS

Abstract—The promotion process of Civil Servants (PNS) is a critical component of human resource management aimed at improving performance and professionalism. However, in practice, the process often faces challenges such as subjective assessments and inefficiencies in data processing. This study aims to develop a classification model for determining the eligibility of regular promotions for civil servants using the Decision Tree C4.5 algorithm based on historical personnel data. The dataset consists of 6,193 records obtained from the Regional Personnel and Human Resource Development Agency of Pangkajene and Kepulauan Regency. The attributes used include years of service, performance scores, and employment type. The research stages involve data preprocessing, feature selection, model construction, and evaluation using confusion matrix, accuracy, precision, recall, and F1-score. The results indicate that the proposed model achieves an accuracy of 95.4%, with precision of 87.6%, recall of 83.7%, and F1-score of 85.6%. Additionally, the model generates interpretable decision rules that support decision-making processes. Therefore, the application of data mining using the C4.5 algorithm is proven effective in enhancing objectivity, transparency, and efficiency in civil servant promotion decision-making.

Keywords: Data Mining; Decision Tree C4.5; Classification; Promotion; Civil Servants

1. PENDAHULUAN

Pegawai Negeri Sipil (PNS) merupakan komponen utama dalam penyelenggaraan pemerintahan yang profesional, akuntabel, dan berorientasi pada pelayanan publik. Dalam konteks reformasi birokrasi, pengelolaan sumber daya manusia aparatur negara dituntut untuk semakin adaptif, transparan, dan berbasis kinerja. Salah satu aspek penting dalam manajemen kepegawaian adalah sistem kenaikan pangkat, yang berfungsi sebagai bentuk penghargaan atas kinerja, kompetensi, dan pengabdian pegawai, sekaligus sebagai instrumen strategis dalam pengembangan karier aparatur [1]. Kenaikan pangkat tidak hanya berdampak pada peningkatan kesejahteraan individu, tetapi juga berkontribusi terhadap peningkatan motivasi kerja dan kualitas pelayanan publik secara keseluruhan.

Namun demikian, dalam praktiknya proses kenaikan pangkat PNS masih menghadapi berbagai permasalahan. Salah satu isu utama adalah adanya subjektivitas dalam penilaian kelayakan pegawai, yang sering kali dipengaruhi oleh faktor non-teknis seperti hubungan personal, senioritas, maupun pertimbangan administratif lainnya. Selain itu, proses pengolahan data yang masih dilakukan secara manual menyebabkan rendahnya efisiensi, tingginya risiko kesalahan administrasi, serta lamanya waktu yang dibutuhkan dalam pengambilan keputusan. Kompleksitas ini semakin meningkat seiring dengan bertambahnya jumlah data kepegawaian yang harus dianalisis.

Seiring dengan perkembangan teknologi informasi, pemanfaatan teknik analisis data menjadi solusi potensial untuk meningkatkan kualitas pengambilan keputusan. Salah satu pendekatan yang banyak digunakan adalah data mining, yaitu proses eksplorasi dan analisis data dalam jumlah besar untuk menemukan pola dan pengetahuan baru yang bermanfaat [2], [3]. Data mining memungkinkan organisasi untuk mengubah data historis menjadi informasi yang dapat digunakan sebagai dasar pengambilan keputusan yang lebih objektif, cepat, dan akurat.

Dalam konteks pengelolaan kepegawaian, data mining dapat digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi kenaikan pangkat serta memprediksi kelayakan pegawai berdasarkan atribut tertentu. Salah satu teknik yang umum digunakan adalah klasifikasi, yaitu proses pengelompokan data ke dalam kategori tertentu berdasarkan pola yang ditemukan dalam data [3], [4]. Teknik ini sangat relevan untuk menentukan apakah seorang PNS termasuk dalam kategori layak atau tidak layak untuk mendapatkan kenaikan pangkat.



Algoritma Decision Tree, khususnya C4.5, merupakan salah satu metode klasifikasi yang populer karena kemampuannya dalam menghasilkan model yang mudah dipahami dan diinterpretasikan. Algoritma ini bekerja dengan membangun struktur pohon keputusan berdasarkan nilai entropy dan information gain untuk menentukan atribut terbaik dalam proses klasifikasi [4], [5]. Keunggulan utama dari C4.5 adalah kemampuannya dalam menangani data numerik dan kategorikal serta menghasilkan aturan keputusan yang transparan.

Beberapa penelitian sebelumnya telah mengkaji penerapan data mining dalam pengambilan keputusan kepegawaian maupun bidang klasifikasi lainnya. Algoritma C4.5 mampu digunakan untuk klasifikasi penerima beasiswa dengan tingkat akurasi yang baik [6]. Selain itu, Hozeng et al [7] juga membuktikan bahwa metode *decision tree* efektif dalam prediksi risiko kredit. Namun demikian, penelitian-penelitian tersebut masih terbatas pada domain tertentu dan belum secara spesifik mengkaji penerapan algoritma C4.5 dalam proses kenaikan pangkat reguler PNS di lingkungan pemerintah daerah. Oleh karena itu, masih terdapat kesenjangan penelitian dalam penerapan model klasifikasi berbasis data mining untuk mendukung pengambilan keputusan kenaikan pangkat PNS secara objektif dan terstruktur.

Penelitian ini memberikan kontribusi dalam pengembangan model klasifikasi berbasis algoritma *Decision Tree* C4.5 untuk menentukan kelayakan kenaikan pangkat PNS secara objektif dan berbasis data. Selain itu, penelitian ini juga menghasilkan aturan keputusan (*decision rules*) yang dapat diimplementasikan dalam sistem pendukung keputusan guna meningkatkan transparansi, efisiensi, dan akurasi dalam proses manajemen kepegawaian di lingkungan pemerintah daerah.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini merupakan penelitian terapan (*applied research*) dengan pendekatan kuantitatif yang bertujuan untuk membangun model klasifikasi dalam menentukan kelayakan kenaikan pangkat Pegawai Negeri Sipil (PNS). Pendekatan yang digunakan bersifat analitis-prediktif, dengan memanfaatkan teknik data mining untuk mengekstraksi pola dari data historis kepegawaian sehingga dapat menghasilkan informasi yang mendukung pengambilan keputusan secara objektif dan berbasis data [8], [9], [10], [11].

Pemanfaatan data mining dalam penelitian ini didasarkan pada kemampuannya dalam mengolah data dalam jumlah besar menjadi pengetahuan yang bernilai, serta telah banyak digunakan dalam berbagai studi untuk keperluan klasifikasi dan prediksi dalam sistem pendukung keputusan [8], [9], [11].

Metode utama yang digunakan adalah algoritma Decision Tree C4.5, yang dipilih karena kemampuannya dalam menghasilkan model klasifikasi yang mudah dipahami dan diinterpretasikan melalui struktur pohon keputusan (*decision tree*) [4], [12]. Model ini memungkinkan proses pengambilan keputusan yang transparan, sistematis, dan berbasis aturan (*rule-based system*), sehingga sangat sesuai untuk diterapkan dalam konteks pengambilan keputusan di lingkungan pemerintahan.

2.2 Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Pangkajene dan Kepulauan. Dataset terdiri dari 6.193 record ASN dengan berbagai atribut kepegawaian. Pemilihan atribut dilakukan berdasarkan relevansi terhadap proses kenaikan pangkat sesuai dengan regulasi manajemen kepegawaian yang berlaku.

Dari keseluruhan atribut yang tersedia, penelitian ini menggunakan atribut utama yang relevan dengan proses kenaikan pangkat reguler, Atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

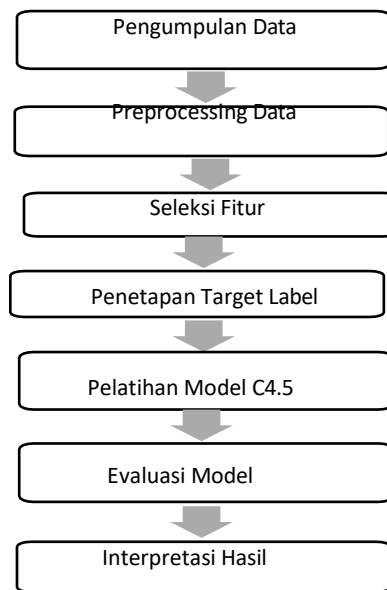
Tabel 1. Atribut Data Penelitian

No	Atribut	Tipe Data	Keterangan
1	Masa Kerja (MKG)	Numerik	Lama kerja dalam tahun
2	Nilai Kinerja	Numerik	Skor evaluasi kinerja
3	Jenis Kepegawaian	Kategorik	Pelaksana, Struktural, Fungsional
4	Status Kelayakan	Kategorik	Layak / Tidak Layak / Evaluasi

Berdasarkan Tabel 1, atribut yang digunakan terdiri dari variabel numerik dan kategorikal yang relevan dengan proses kenaikan pangkat.

2.3 Diagram Alur Penelitian

Diagram alur penelitian menggambarkan tahapan proses data mining mulai dari pengumpulan data, preprocessing, seleksi fitur, pembentukan model, evaluasi, hingga interpretasi hasil. Pendekatan ini sejalan dengan metodologi Knowledge Discovery in Database atau KDD, yaitu proses sistematis untuk mengekstraksi pola, informasi, dan pengetahuan yang bermanfaat dari sekumpulan data [9], [13]. Diagram alur penelitian dapat dilihat pada Gambar 1 berikut :



Gambar 1. Diagram Alur Penelitian

Gambar 1 menunjukkan alur tahapan penelitian yang dilakukan secara sistematis dalam mengimplementasikan algoritma Decision Tree C4.5 untuk klasifikasi kelayakan kenaikan pangkat Pegawai Negeri Sipil (PNS). Proses penelitian diawali dengan tahap pengumpulan data, di mana data kepegawaian diperoleh dari Badan Kepegawaian dan Pengembangan Sumber Daya Manusia (BKPSDM) Kabupaten Pangkajene dan Kepulauan dalam bentuk dataset historis.

Tahap berikutnya adalah preprocessing data, yang bertujuan untuk meningkatkan kualitas data sebelum dilakukan proses analisis. Pada tahap ini dilakukan pembersihan data dari nilai yang hilang (missing values), transformasi data kategorikal menjadi numerik melalui proses encoding, serta seleksi atribut yang relevan dengan kebutuhan penelitian.

Selanjutnya dilakukan penetapan label target, yaitu menentukan kategori kelayakan kenaikan pangkat yang terdiri dari layak, tidak layak, dan evaluasi manual berdasarkan kriteria yang telah ditetapkan. Tahap ini menjadi dasar dalam proses klasifikasi yang akan dilakukan oleh model.

Tahapan berikutnya adalah pelatihan model (*training*) menggunakan algoritma Decision Tree C4.5. Pada tahap ini, model dibangun dengan menentukan atribut terbaik berdasarkan nilai entropy dan information gain untuk membentuk struktur pohon keputusan secara optimal. Setelah model terbentuk, dilakukan evaluasi model menggunakan metode *confusion matrix* untuk mengukur kinerja klasifikasi. Parameter yang digunakan meliputi *accuracy*, *precision*, *recall*, dan *F1-score* guna mengetahui tingkat ketepatan dan keseimbangan performa model.

Tahap akhir adalah interpretasi model, di mana hasil klasifikasi divisualisasikan dalam bentuk pohon keputusan serta aturan keputusan (*decision rules*) berbasis *IF-THEN*. Tahap ini penting untuk memastikan bahwa model yang dihasilkan tidak hanya akurat, tetapi juga mudah dipahami dan dapat digunakan sebagai dasar dalam sistem pendukung keputusan.

Secara keseluruhan, diagram alur penelitian ini menggambarkan proses terstruktur dalam mengolah data menjadi informasi yang bernilai melalui pendekatan data mining, sehingga mampu mendukung pengambilan keputusan yang lebih objektif, transparan, dan efisien.

2.4 Tahapan Penelitian

Tahapan penelitian dilakukan secara sistematis untuk memastikan bahwa proses klasifikasi kelayakan kenaikan pangkat PNS dapat berjalan secara terstruktur, objektif, dan dapat dievaluasi. Adapun tahapan penelitian adalah sebagai berikut.

1) Pengumpulan Data

Data diperoleh dari database kepegawaian dalam format CSV. Data tersebut berisi informasi historis ASN yang relevan dengan proses kenaikan pangkat.

2) Preprocessing Data

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses klasifikasi. Tahapan ini meliputi penghapusan data kosong (missing value), transformasi data kategorikal menjadi numerik melalui proses encoding, serta seleksi atribut yang relevan. Preprocessing merupakan langkah penting dalam data mining karena kualitas data sangat menentukan akurasi dan keandalan model klasifikasi yang dihasilkan [2], [9], [10], [14].

3) Penetapan Label Target

Label klasifikasi ditentukan berdasarkan kriteria kelayakan kenaikan pangkat, yaitu:

- Layak, apabila masa kerja golongan atau MKG ≥ 4 dan nilai kinerja berada pada kategori baik;
- Tidak Layak, apabila pegawai tidak memenuhi kriteria kelayakan; dan
- Evaluasi Manual, apabila pegawai berada pada jabatan fungsional atau membutuhkan pertimbangan lebih lanjut.

4) Pelatihan Model

Model dilatih menggunakan algoritma Decision Tree C4.5. Algoritma ini membangun struktur pohon keputusan berdasarkan data training dengan memilih atribut terbaik melalui konsep entropy dan information gain [4].

5) Evaluasi Model

Model diuji menggunakan beberapa metrik evaluasi, yaitu

- a) *confusion matrix*,
- b) *accuracy*,
- c) *precision*,
- d) *recall*, dan
- e) *F1-score*.

Metrik tersebut digunakan untuk menilai tingkat ketepatan, sensitivitas, serta keseimbangan performa model dalam melakukan klasifikasi.

6) Interpretasi Model

Hasil model divisualisasikan dalam bentuk pohon keputusan dan aturan keputusan berbasis *IF-THEN*. Interpretasi ini bertujuan agar hasil klasifikasi tidak hanya akurat, tetapi juga mudah dipahami dan dapat digunakan sebagai dasar pengambilan keputusan.

2.5 Algoritma Decision Tree C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang digunakan untuk membentuk pohon keputusan berdasarkan pemilihan atribut terbaik dalam proses klasifikasi. Pemilihan atribut dilakukan dengan memanfaatkan konsep entropy dan information gain, yaitu ukuran yang digunakan untuk menentukan atribut yang paling informatif dalam memisahkan data ke dalam kelas tertentu [4], [12]. Dalam konteks klasifikasi, algoritma ini bekerja dengan cara mengidentifikasi pola dari data latih, kemudian menyusun aturan keputusan yang dapat digunakan untuk mengelompokkan data baru ke dalam kelas yang telah ditentukan.

Dalam algoritma C4.5, atribut dengan nilai gain tertinggi akan dipilih sebagai akar pohon (*root node*) karena dianggap memiliki kemampuan paling besar dalam membedakan kelas data. Setelah akar pohon terbentuk, proses pembentukan cabang dilakukan secara rekursif terhadap setiap subset data hingga data dapat terklasifikasi secara optimal. Setiap simpul dalam pohon keputusan merepresentasikan atribut, setiap cabang menunjukkan kondisi atau nilai atribut, sedangkan daun pohon menunjukkan hasil klasifikasi. Mekanisme ini membuat proses klasifikasi menjadi lebih sistematis dan dapat ditelusuri secara logis.

Algoritma Decision Tree C4.5 banyak digunakan dalam berbagai studi klasifikasi karena mampu menghasilkan model yang relatif akurat, sederhana, dan mudah diinterpretasikan. Struktur pohon keputusan yang dihasilkan memungkinkan pengguna memahami hubungan antara atribut dan hasil klasifikasi secara transparan. Metode *decision tree* memiliki keunggulan dalam interpretabilitas, kemudahan implementasi, serta kemampuan menangani data dengan atribut numerik maupun kategorikal [5]. Keunggulan tersebut menjadikan C4.5 relevan digunakan dalam penelitian yang membutuhkan model klasifikasi yang tidak hanya akurat, tetapi juga dapat dijelaskan kepada pengguna.

Dalam penelitian ini, algoritma C4.5 digunakan untuk mengklasifikasikan kelayakan kenaikan pangkat Pegawai Negeri Sipil berdasarkan atribut kepegawaian yang telah ditentukan. Atribut seperti masa kerja golongan, nilai kinerja, dan jenis kepegawaian dianalisis untuk menghasilkan pola keputusan yang dapat membedakan kategori layak, tidak layak, dan evaluasi manual. Dengan karakteristiknya yang berbasis aturan, algoritma C4.5 dapat membantu menghasilkan keputusan yang lebih objektif karena proses klasifikasi didasarkan pada pola data, bukan semata-mata pada penilaian subjektif. Hal ini penting dalam konteks manajemen kepegawaian yang menuntut transparansi, akuntabilitas, dan konsistensi keputusan.

Selain itu, berbagai studi menunjukkan bahwa algoritma C4.5 efektif diterapkan pada proses klasifikasi di berbagai bidang. Penerapan algoritma C4.5 untuk penentuan penerima beasiswa mahasiswa dan menunjukkan bahwa metode ini mampu menghasilkan proses klasifikasi yang dapat digunakan sebagai dasar pengambilan keputusan [6]. Sementara itu, Hozeng et al. [7] menggunakan metode decision tree untuk memprediksi risiko kredit dan menunjukkan bahwa pendekatan ini dapat membantu proses penilaian berbasis data. Temuan tersebut memperkuat relevansi penggunaan algoritma C4.5 dalam penelitian ini untuk mengklasifikasikan kelayakan kenaikan pangkat Pegawai Negeri Sipil secara objektif, sistematis, dan mudah diinterpretasikan.

2.6 Pembagian Data

Dataset dalam penelitian ini dibagi menjadi dua bagian utama, yaitu data training dan data testing. Pembagian data dilakukan untuk memastikan bahwa model klasifikasi tidak hanya mampu mengenali pola pada data yang digunakan dalam proses pelatihan, tetapi juga mampu mengklasifikasikan data baru yang belum pernah diproses sebelumnya. Dalam konteks penelitian ini, pembagian data menjadi tahap penting karena model yang dibangun diharapkan memiliki kemampuan generalisasi yang baik dalam menentukan kelayakan kenaikan pangkat Pegawai Negeri Sipil.

Data training digunakan untuk membangun dan melatih model klasifikasi. Pada tahap ini, algoritma Decision Tree C4.5 mempelajari pola hubungan antara atribut masukan, seperti masa kerja golongan, nilai kinerja, dan jenis kepegawaian, terhadap label target yang telah ditentukan. Melalui data training, model membentuk struktur pohon keputusan berdasarkan pola yang ditemukan dalam data. Dengan demikian, kualitas dan representativitas data training sangat menentukan kemampuan model dalam menghasilkan aturan klasifikasi yang tepat.



Data testing digunakan untuk menguji dan mengevaluasi performa model terhadap data yang tidak digunakan pada proses pelatihan. Pengujian ini bertujuan untuk mengetahui sejauh mana model mampu melakukan klasifikasi terhadap data baru secara akurat. Dengan adanya data testing, performa model dapat dinilai secara lebih objektif karena hasil evaluasi tidak hanya didasarkan pada data yang telah dipelajari oleh model, tetapi juga pada data yang berfungsi sebagai simulasi kasus baru.

Adapun proporsi pembagian data yang digunakan dalam penelitian ini adalah 80% untuk data training dan 20% untuk data testing. Data training sebesar 80% digunakan agar model memiliki cukup data untuk mengenali pola dan hubungan antaratribut dalam dataset. Sementara itu, data testing sebesar 20% digunakan untuk mengukur kemampuan model dalam melakukan klasifikasi terhadap data baru. Skema ini dipilih karena mampu memberikan keseimbangan antara kebutuhan pelatihan model dan kebutuhan evaluasi kinerja model.

Pembagian data dilakukan secara proporsional agar distribusi data tetap representatif terhadap keseluruhan dataset. Hal ini penting untuk menghindari model yang terlalu menyesuaikan diri terhadap data pelatihan atau overfitting. Overfitting merupakan kondisi ketika model memiliki performa yang sangat baik pada data training, tetapi menurun ketika diterapkan pada data baru. Oleh karena itu, pemisahan data training dan testing menjadi langkah penting untuk menilai kemampuan generalisasi model secara lebih objektif [13].

Lebih lanjut, penggunaan skema pembagian data 80:20 merupakan pendekatan yang umum digunakan dalam penelitian data mining karena dapat memberikan porsi data yang cukup untuk proses pelatihan sekaligus menyediakan data uji yang memadai untuk evaluasi. Dengan pembagian tersebut, model yang dihasilkan diharapkan tidak hanya mampu membangun pola klasifikasi berdasarkan data historis, tetapi juga dapat digunakan secara efektif dalam mendukung pengambilan keputusan berbasis data. Dalam penelitian ini, skema pembagian data tersebut menjadi dasar untuk mengukur akurasi, ketepatan prediksi, serta keandalan model Decision Tree C4.5 dalam mengklasifikasikan kelayakan kenaikan pangkat PNS.

2.7 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui sejauh mana algoritma Decision Tree C4.5 mampu mengklasifikasikan kelayakan kenaikan pangkat Pegawai Negeri Sipil secara akurat dan konsisten. Dalam penelitian ini, evaluasi dilakukan menggunakan confusion matrix karena metode ini mampu memberikan gambaran rinci mengenai perbandingan antara kelas aktual dan kelas hasil prediksi model. Melalui confusion matrix, peneliti dapat mengetahui jumlah data yang diklasifikasikan dengan benar maupun data yang mengalami kesalahan klasifikasi.

Confusion matrix merupakan metode evaluasi yang banyak digunakan dalam penelitian klasifikasi karena dapat menggambarkan performa model pada setiap kategori kelas. Dalam konteks penelitian ini, kategori klasifikasi meliputi pegawai yang dinyatakan layak, tidak layak, dan evaluasi manual. Hasil prediksi model kemudian dibandingkan dengan label aktual untuk mengetahui tingkat ketepatan klasifikasi. Dengan demikian, confusion matrix tidak hanya menunjukkan performa model secara umum, tetapi juga membantu mengidentifikasi kategori mana yang paling sering mengalami kesalahan prediksi.

Parameter evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengukur proporsi jumlah prediksi yang benar terhadap keseluruhan data uji. Precision mengukur tingkat ketepatan model dalam memprediksi suatu kelas tertentu, terutama dalam menghindari kesalahan prediksi positif. Recall menggambarkan kemampuan model dalam mengidentifikasi seluruh data yang benar-benar termasuk dalam kelas tertentu. Sementara itu, F1-score merupakan nilai harmonik antara precision dan recall yang digunakan untuk memberikan ukuran performa yang lebih seimbang, khususnya ketika distribusi data tidak sepenuhnya merata [3], [13].

Secara teknis, penggunaan beberapa metrik evaluasi diperlukan karena nilai accuracy saja belum cukup untuk menggambarkan kualitas model secara menyeluruh. Model dapat saja memiliki akurasi tinggi, tetapi belum tentu memiliki kemampuan yang baik dalam mengidentifikasi kelas tertentu apabila data tidak seimbang. Oleh karena itu, precision, recall, dan F1-score digunakan sebagai ukuran tambahan untuk menilai ketepatan dan sensitivitas model. Dengan kombinasi metrik tersebut, evaluasi model menjadi lebih komprehensif dan mampu memberikan gambaran yang lebih objektif mengenai kinerja klasifikasi.

Hasil evaluasi digunakan untuk menentukan apakah model yang dibangun mampu mengklasifikasikan kelayakan kenaikan pangkat secara tepat berdasarkan data kepegawaian yang tersedia. Selain itu, analisis terhadap confusion matrix juga memungkinkan identifikasi kesalahan klasifikasi atau misclassification yang terjadi pada model. Informasi ini penting karena kesalahan klasifikasi dalam konteks kenaikan pangkat dapat berdampak pada kualitas pengambilan keputusan. Oleh karena itu, hasil evaluasi dapat dijadikan dasar untuk memperbaiki model, meninjau kembali atribut yang digunakan, atau menambah data pendukung pada penelitian selanjutnya.

Lebih lanjut, metode klasifikasi dalam data mining telah banyak digunakan dalam pengembangan sistem pendukung keputusan karena mampu membantu proses evaluasi dan pemilihan alternatif secara lebih objektif, terstruktur, dan berbasis data. Sabir et al. [15] menunjukkan bahwa metode berbasis pengambilan keputusan dapat digunakan untuk mendukung proses promosi jabatan karyawan secara lebih sistematis. Sementara itu, Saputra et al. [16] menunjukkan bahwa sistem pendukung keputusan juga dapat digunakan dalam penilaian kinerja karyawan dengan pendekatan komputasional. Dengan demikian, penggunaan metrik evaluasi yang komprehensif dalam penelitian ini tidak hanya bertujuan untuk mengukur performa model, tetapi juga untuk memastikan bahwa model Decision Tree C4.5 memiliki tingkat keandalan yang memadai dalam mendukung pengambilan keputusan di lingkungan pemerintahan.

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan dan Persiapan Data

Tahap awal dalam penelitian ini adalah pengolahan dan persiapan data yang bertujuan untuk memastikan bahwa dataset memiliki kualitas yang memadai sebelum digunakan dalam proses klasifikasi. Dataset yang digunakan merupakan data kepegawaian sebanyak 6.193 record ASN yang diperoleh dari Badan Kepegawaian dan Pengembangan Sumber Daya Manusia atau BKPSDM Kabupaten Pangkajene dan Kepulauan. Data tersebut memuat berbagai atribut kepegawaian, namun tidak seluruh atribut digunakan dalam proses pemodelan. Penelitian ini hanya memilih atribut yang relevan dengan proses kenaikan pangkat reguler agar model yang dibangun tetap fokus pada tujuan klasifikasi kelayakan kenaikan pangkat.

Pengolahan data dilakukan karena data mentah yang diperoleh dari sumber administrasi kepegawaian umumnya masih memerlukan penyesuaian sebelum dapat diproses oleh algoritma klasifikasi. Data mentah dapat mengandung nilai kosong, format yang tidak seragam, atribut yang kurang relevan, atau kategori data yang belum sesuai dengan kebutuhan pemodelan. Apabila kondisi tersebut tidak ditangani, maka hasil klasifikasi berpotensi menjadi kurang akurat dan sulit diinterpretasikan. Oleh karena itu, tahap persiapan data menjadi bagian penting untuk memastikan bahwa data yang digunakan benar-benar siap dianalisis.

Tahapan preprocessing yang dilakukan meliputi pembersihan data (*data cleaning*), transformasi data, dan seleksi fitur. Pembersihan data dilakukan dengan menghapus atau menangani nilai kosong (*missing values*) serta memperbaiki inkonsistensi format data. Langkah ini bertujuan untuk mengurangi gangguan pada proses pembentukan model dan memastikan bahwa setiap record memiliki informasi yang sesuai dengan kebutuhan klasifikasi. Selain itu, data juga diperiksa untuk memastikan tidak terdapat kesalahan format yang dapat menghambat proses analisis.

Setelah proses pembersihan data dilakukan, tahap berikutnya adalah transformasi data. Transformasi diperlukan karena algoritma klasifikasi membutuhkan data dalam format yang dapat diproses secara komputasional. Dalam penelitian ini, atribut kategorikal seperti jenis kepegawaian diubah menjadi format numerik menggunakan teknik label encoding. Proses ini memungkinkan kategori data, seperti pelaksana, struktural, atau fungsional, direpresentasikan dalam bentuk angka sehingga dapat dikenali oleh algoritma Decision Tree C4.5. Dengan transformasi tersebut, data kategorikal tetap dapat digunakan tanpa menghilangkan makna klasifikasinya.

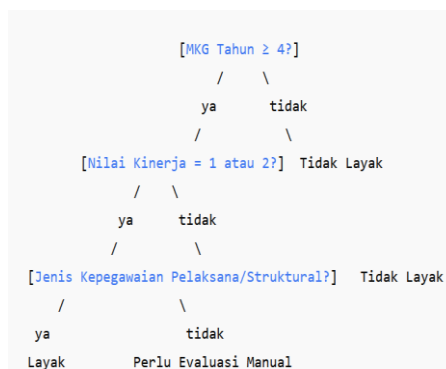
Seleksi fitur dilakukan untuk memilih atribut yang paling relevan terhadap kelayakan kenaikan pangkat. Berdasarkan analisis awal dan kesesuaian dengan regulasi kepegawaian, atribut yang digunakan dalam model meliputi Masa Kerja Golongan (MKG), Nilai Kinerja, dan Jenis Kepegawaian. Masa Kerja Golongan digunakan karena berkaitan dengan lamanya pegawai berada dalam golongan tertentu. Nilai Kinerja digunakan untuk merepresentasikan capaian kerja pegawai. Sementara itu, Jenis Kepegawaian digunakan karena karakteristik jabatan dapat memengaruhi proses penilaian dan kebutuhan evaluasi.

Pemilihan atribut tersebut bertujuan untuk menyederhanakan model tanpa mengurangi kemampuan klasifikasi. Model yang terlalu banyak menggunakan atribut berisiko menjadi kompleks dan sulit diinterpretasikan, sedangkan model yang menggunakan atribut relevan dapat menghasilkan aturan keputusan yang lebih jelas. Dalam konteks penelitian ini, penyederhanaan atribut juga penting karena model yang dihasilkan diharapkan dapat digunakan sebagai dasar sistem pendukung keputusan. Dengan demikian, hasil klasifikasi tidak hanya menekankan akurasi, tetapi juga kemudahan pemahaman oleh pihak pengelola kepegawaian.

Secara keseluruhan, tahap pengolahan dan persiapan data berperan penting dalam membangun model klasifikasi yang objektif, efisien, dan mudah diinterpretasikan. Data yang telah melalui proses pembersihan, transformasi, dan seleksi fitur menjadi dasar bagi pembentukan model Decision Tree C4.5. Dengan kualitas data yang lebih baik, model diharapkan mampu mengidentifikasi pola kelayakan kenaikan pangkat secara lebih akurat serta menghasilkan aturan keputusan yang dapat digunakan dalam proses pengambilan keputusan kepegawaian.

3.2 Pohon Keputusan Manual Berdasarkan Logika

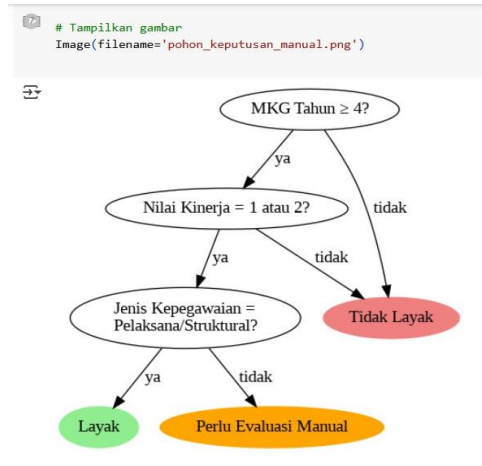
Untuk memperjelas logika klasifikasi, berikut adalah struktur pohon keputusan manual yang digunakan:



Gambar 2. Pohon Keputusan Manual

Gambar 2 menunjukkan struktur pohon keputusan (decision tree) yang dihasilkan oleh algoritma C4.5 berdasarkan data kepegawaian yang digunakan dalam penelitian ini. Pada pohon tersebut, atribut yang memiliki nilai information gain tertinggi dipilih sebagai akar (root node), yang dalam penelitian ini didominasi oleh atribut masa kerja (MKG). Hal ini menunjukkan bahwa masa kerja merupakan faktor utama dalam menentukan kelayakan kenaikan pangkat Pegawai Negeri Sipil.

Selanjutnya, pohon keputusan membentuk cabang-cabang berdasarkan atribut lain seperti nilai kinerja dan jenis kepegawaian. Setiap cabang merepresentasikan kondisi tertentu yang mengarah pada keputusan klasifikasi, yaitu Layak, Tidak Layak, atau Evaluasi. Proses pembentukan pohon dilakukan secara rekursif hingga seluruh data terklasifikasi secara optimal. Struktur pohon keputusan ini memberikan gambaran yang jelas mengenai pola hubungan antar atribut serta memungkinkan interpretasi model secara transparan melalui aturan keputusan (decision rules) yang dihasilkan.



Gambar 3. Tampilan pohon keputusan manual melalui

Gambar 3 menunjukkan hasil evaluasi model klasifikasi menggunakan confusion matrix, yang digunakan untuk mengukur kinerja algoritma C4.5 dalam mengklasifikasikan data kepegawaian. Matriks ini menggambarkan perbandingan antara nilai aktual dan hasil prediksi model dalam setiap kategori kelas, yaitu Layak, Tidak Layak, dan Evaluasi.

Berdasarkan hasil yang ditampilkan, sebagian besar data berhasil diklasifikasikan dengan benar, yang ditunjukkan oleh nilai diagonal pada matriks. Hal ini menunjukkan bahwa model memiliki tingkat akurasi yang baik dalam memprediksi kelayakan kenaikan pangkat. Selain itu, nilai precision dan recall yang diperoleh juga menunjukkan bahwa model mampu mengidentifikasi data positif dengan cukup baik serta meminimalkan kesalahan klasifikasi.

Dengan demikian, hasil evaluasi ini menunjukkan bahwa algoritma C4.5 memiliki performa yang efektif dan dapat diandalkan untuk digunakan sebagai dasar dalam sistem pendukung keputusan kenaikan pangkat Pegawai Negeri Sipil.

3.3 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk mengukur tingkat performa klasifikasi yang dihasilkan oleh algoritma Decision Tree C4.5. Pendekatan ini digunakan untuk menilai sejauh mana model mampu mengklasifikasikan data ke dalam kategori yang tepat berdasarkan label aktual. Berdasarkan hasil pengujian, model menunjukkan tingkat performa yang baik dalam mengklasifikasikan data kelayakan kenaikan pangkat PNS. Secara umum, performa model dapat dijelaskan sebagai berikut:

- Accuracy menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model. Nilai akurasi yang tinggi mengindikasikan bahwa model memiliki tingkat ketepatan yang baik secara keseluruhan.
- Precision menunjukkan kemampuan model dalam memprediksi kategori layak secara tepat. Nilai precision yang cukup tinggi menunjukkan bahwa kesalahan dalam memprediksi pegawai yang tidak layak sebagai layak relatif rendah.
- Recall menunjukkan kemampuan model dalam mendeteksi seluruh data yang benar-benar layak. Nilai recall yang baik mengindikasikan bahwa model mampu mengidentifikasi sebagian besar pegawai yang memang memenuhi syarat kenaikan pangkat.
- F1-Score merupakan ukuran yang menggabungkan precision dan recall, sehingga memberikan gambaran keseimbangan performa model dalam klasifikasi.

Tabel 2. Hasil Evaluasi Model

Metrik	Nilai
Accuracy	95.4%
Precision	87.6%
Recall	83.7%
F1-Score	85.6%

Berdasarkan Tabel 2, model Decision Tree C4.5 menunjukkan performa yang sangat baik dengan tingkat akurasi sebesar 95,4%, yang berarti sebagian besar data berhasil diklasifikasikan dengan benar. Nilai precision sebesar 87,6% menunjukkan bahwa model cukup akurat dalam memprediksi pegawai yang layak, dengan tingkat kesalahan klasifikasi positif yang relatif rendah. Sementara itu, nilai recall sebesar 83,7% mengindikasikan bahwa model mampu mengidentifikasi sebagian besar pegawai yang benar-benar layak untuk naik pangkat, meskipun masih terdapat beberapa data yang tidak terdeteksi secara optimal. Nilai F1-score sebesar 85,6% menunjukkan bahwa model memiliki keseimbangan yang baik antara precision dan recall, sehingga dapat dianggap cukup stabil dalam melakukan klasifikasi.

Tingginya nilai akurasi pada model ini dipengaruhi oleh kesesuaian antara aturan pelabelan (labeling rules) dengan pola data yang digunakan dalam proses pelatihan. Hal ini menunjukkan bahwa atribut seperti masa kerja dan nilai kinerja memiliki kontribusi signifikan dalam membentuk pola klasifikasi. Namun demikian, kondisi ini juga perlu diperhatikan karena berpotensi menimbulkan bias model (overfitting), terutama jika data yang digunakan terlalu mengikuti aturan yang telah ditentukan sebelumnya. Oleh karena itu, diperlukan pengujian lebih lanjut dengan dataset yang lebih beragam serta penambahan atribut lain agar model memiliki kemampuan generalisasi yang lebih baik.

Selain itu, hasil penelitian ini sejalan dengan beberapa penelitian sebelumnya yang menunjukkan bahwa algoritma Decision Tree, khususnya C4.5, mampu menghasilkan model klasifikasi dengan tingkat akurasi yang tinggi serta mudah diinterpretasikan. Keunggulan ini menjadikan C4.5 sebagai metode yang sesuai untuk diterapkan dalam sistem pendukung keputusan, terutama dalam konteks pengambilan keputusan administratif di lingkungan pemerintahan.

3.4 Implikasi Hasil Penelitian

Hasil penelitian ini memberikan kontribusi praktis yang signifikan dalam pengembangan sistem pendukung keputusan di lingkungan pemerintahan, khususnya dalam proses penentuan kenaikan pangkat Pegawai Negeri Sipil (PNS). Model klasifikasi yang dihasilkan melalui penerapan algoritma Decision Tree C4.5 terbukti mampu meningkatkan objektivitas dalam proses penilaian, karena keputusan yang dihasilkan didasarkan pada pola data historis yang terstruktur, bukan semata-mata pada pertimbangan subjektif.

Selain itu, penerapan model ini juga berpotensi mengurangi tingkat subjektivitas dalam pengambilan keputusan kepegawaian, sehingga dapat meningkatkan transparansi dan akuntabilitas dalam tata kelola sumber daya manusia aparatur. Proses evaluasi data kepegawaian yang sebelumnya dilakukan secara manual dan memerlukan waktu yang relatif lama, dapat dipercepat melalui pemanfaatan model klasifikasi yang mampu mengolah data secara otomatis dan efisien.

Lebih lanjut, model yang dihasilkan mampu menghasilkan aturan keputusan (decision rules) yang jelas dan mudah dipahami oleh pengguna. Hal ini memberikan nilai tambah dalam implementasi sistem pendukung keputusan, karena hasil analisis tidak hanya bersifat prediktif, tetapi juga dapat dijadikan sebagai dasar pertimbangan yang rasional dan dapat dipertanggungjawabkan. Dengan demikian, penelitian ini berkontribusi dalam mendorong penerapan pengambilan keputusan berbasis data (data-driven decision making) di lingkungan pemerintahan.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma Decision Tree C4.5 untuk mengklasifikasikan kelayakan kenaikan pangkat Pegawai Negeri Sipil (PNS) berdasarkan data historis kepegawaian. Model yang dihasilkan menunjukkan kinerja yang sangat baik dengan tingkat akurasi sebesar 95,4%, serta didukung oleh nilai precision, recall, dan F1-score yang menunjukkan keseimbangan performa klasifikasi. Selain itu, model mampu menghasilkan aturan keputusan (decision rules) yang mudah diinterpretasikan, sehingga dapat digunakan sebagai dasar dalam sistem pendukung keputusan. Hasil penelitian ini membuktikan bahwa pendekatan data mining efektif dalam meningkatkan objektivitas, transparansi, dan efisiensi proses pengambilan keputusan kenaikan pangkat PNS di lingkungan pemerintah daerah. Oleh karena itu, model yang dikembangkan berpotensi untuk diimplementasikan secara lebih luas sebagai bagian dari sistem digitalisasi manajemen kepegawaian berbasis data.

REFERENCES

- [1] R. Indonesia, "Peraturan Pemerintah Republik Indonesia Nomor 11 Tahun 2017 tentang Manajemen Pegawai Negeri Sipil," Republik Indonesia, Jakarta, 2017.
- [2] J. Han, M. Kamber, and J. Pei, *Data mining: Concepts and techniques*, 3rd ed. in The Morgan Kaufmann Series in Data Management Systems. Burlington, MA: Morgan Kaufmann, 2012.
- [3] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data mining: Practical machine learning tools and techniques*, 4th ed. in The Morgan Kaufmann Series in Data Management Systems. Cambridge, MA: Morgan Kaufmann, 2016.
- [4] J. R. Quinlan, *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufmann, 1993.
- [5] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261-283, 2013, doi: 10.1007/s10462-011-9272-4.
- [6] E. Hasmin and S. Aisa, "Penerapan algoritma C4.5 untuk penentuan penerima beasiswa mahasiswa," *CogITo Smart J.*, vol. 5, no. 2, pp. 308-320, 2019, doi: 10.31154/cogito.v5i2.219.308-320.
- [7] S. Hozeng, S. Aisa, and S. Dipanegara, "Aplikasi data mining dengan menggunakan metode decision tree untuk prediksi penentuan resiko kredit," *SISITI Semin. Ilm. Sist. Inf. dan Teknol. Inf.*, vol. 5, no. 2, pp. 1-10, 2016, doi: 10.36774/sisiti.v5i2.159.
- [8] F. Gorunescu, *Data mining: Concepts, models and techniques*. Berlin, Germany: Springer, 2011. doi: 10.1007/978-3-642-19721-



- 5.
- [9] D. T. Larose and C. D. Larose, *Discovering knowledge in data: An introduction to data mining*, 2nd ed. Hoboken, NJ: Wiley, 2014.
- [10] B. Santosa, *Data mining: Teknik pemanfaatan data untuk keperluan bisnis*, 1st ed. Yogyakarta: Graha Ilmu, 2007.
- [11] D. P. Kusumaningtyas and N. Legowo, "Data mining techniques to predict high leadership promotion of civil servants in Indonesia based on talent pool assessments," *J. Syst. Manag. Sci.*, vol. 13, no. 6, pp. 336–346, 2023, doi: 10.33168/JSMS.2023.0620.
- [12] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man. Cybern.*, vol. 21, no. 3, pp. 660–674, 1991, doi: 10.1109/21.97458.
- [13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Mag.*, vol. 17, no. 3, pp. 37–54, 1996, doi: 10.1609/aimag.v17i3.1230.
- [14] S. Wahyuddin *et al.*, *Data warehouse dan data mining*. Medan: Yayasan Kita Menulis, 2023.
- [15] F. M. Sabir, M. Mashud, A. Halid, A. Asrul, and R. Rumallang, "Penerapan metode multi criteria decision making sebagai sistem penunjang keputusan promosi jabatan karyawan di PT. Fastfood Indonesia Tbk," *J. Minfo Polgan*, vol. 13, no. 1, pp. 150–163, 2024, doi: 10.33395/jmp.v13i1.13500.
- [16] H. Saputra, A. Y. Muniar, and M. Mashud, "Sistem pendukung keputusan penilaian kinerja karyawan menggunakan fuzzy logic Tsukamoto," *Neptunus J. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 3, pp. 278–288, 2025, doi: 10.61132/neptunus.v3i3.1028.