PENERAPAN ALGORITMA MENGGUNAKAN RAPIDMINER UNTUK KATEGORISASI KOMPOTENSI DASAR CPNS

Sidik Mahesa Putra, Muhammad Rendi, Depi Rusda
Program Studi Sistem Informasi S1, Fakultas Ilmu Komputer
Universitas Darwan Ali, Jalan Batu Berlian No. 10 Kota Sampit, Indonesia
sidikmp88@gmail.com

ABSTRAK

Penelitian ini mengkaji penggunaan data mining dalam seleksi Calon Pegawai Negeri Sipil (CPNS) diIndonesia, khususnya Universitas Pattimura pada tahun 2021. Pemerintah dihadapkan pada tuntutan untuk merekrut PNS berkualitas dan kompeten. Dalam konteks ini, penelitian membandingkan dua algoritma, yaitu K-Nearest Neighbor dan C4.5, untuk mengklasifikasikan kompetensi dasar CPNS. Untuk empat kelas klasifikasi, hasil penelitian menunjukkan bahwa metode Decision Tree dengan algoritma C4.5 mempunyai tingkat akurasi sebesar 75%, tingkat kesalahan klasifikasi sebesar 5%, recall sebesar 97,14 persen, kappa sebesar 0,947, dan presisi 93,94 persen. Temuan ini memberikan gambaran tentang efektivitas algoritma C4.5 dalam mengurutkan konsekuensi Pilihan Kemampuan Penting (SKD) anggota CPNS Perguruan Tinggi Pattimura. Penelitian ini menekankan pentingnya penggunaan teknik data mining dalam meningkatkan efisiensi dan objektivitas dalam proses seleksi PNS. Hasil ini mungkin memiliki implikasi positif dalam memastikan bahwa PNS yang direkrut memenuhi standar kompetensi yang diinginkan oleh pemerintah.

Kata kunci: *Pegawai Negeri, RapidMiner, CPNS, Kompetensi*

1. PENDAHULUAN

Pejabat berwenang Republik Indonesia berwenang mengangkat warga negara Indonesia menjadi pegawai dan menugaskannya dalam jabatan yang sesuai dengan keahlian dan kualifikasinya, serta menerima gaji sesuai dengan undang-undang dan peraturan yang berlakua dalah Pejabat Pegawai Negeri Sipil (PNS) di Indonesia[1]. Pengalaman sebagai seorang pegawai negeri merupakan hal menarik bagi berbagai kelompok di Indonesia, tidak lain adalah lulusan baru dan mantan siswa sekolah menengah pertama dan menengah atas di seluruh negeri. Oleh sebab itu, setiap tahun dalam seleksi CPNS, ribuan orang dapat bersaing untuk satu posisi di lembaga tertentu. Keuntungan menjadi pegawai negeri meliputi pendapatan yang tetap, manfaat pensiun, perlindungan dari pemecatan pegawai tanpa alasan yang jelas, dan jalan karir yang jelas. Selain itu, sebagai pegawai negeri, terdapat banyak keuntungan lainnya[2]. Saat ini, keterbatasan jumlah PNS yang berkualifikasi adalah masalah yang dihadapi pemerintah. Itu telah menjadi suatu hal penting dalam tata kelola Indonesia. Reformasi rekrutmen CPNS adalah salah satu reformasi penting yang harus segera dilakukan. Hal ini mengingat sebagian besar pegawai negeri sipil di Indonesia sedang menjalani proses pendaftaran CPNS. Pendaftaran adalah salah satu bagian penting dari teknik dewan bantuan umum. Pemilihan calon pegawai negeri sipil yang mempunyai integritas tinggi dan kemampuan sumber daya manusia yang unggul merupakan tujuan rekrutmen CPNS. Hal ini dicapai melalui proses rekrutmen yang bertanggung jawab dan terbuka.

Bersamaan dengan perkembangan teknologi, metode data mining menggunakan teknik klasifikasi data juga telah berkembang. Berbagai penelitian telah dilakukan dalam konteks ini. Sebagai contoh, sebuah penelitian tentang algoritma klasifikasi untuk validasi uang kertas mencapai akurasi yang sangat tinggi, dengan hasil akurasi sebesar 98.5% menggunakan metode Decision Tree C4.5, sedangkan Neural Network mencapai 95%, dan Naive Bayes mencapai 85%[3]. Selain itu, penelitian lain menggunakan teknik data mining dengan SVM (Support Vector Machine) untuk memprediksi kompetensi dasar CPNS dan mencapai akurasi sebesar 43.33%[4]. Pada eksplorasi kali ini, pelaksanaan penanganan gambar dan pemesanan menggunakan K-Closest Neighbor yang diunggulkan adalah dengan membangun aplikasi online untuk memisahkan daging sapi dan daging babi, dan berhasil mengklasifikasikan perbedaan antara daging sapi dan daging babi dengan akurasi 88.75% berdasarkan ekstraksi fitur tekstur[5].

2. TINJAUAN PUSTAKA

2.1. Data Mining

Proses memperoleh pengetahuan dari data fundamental disebut sebagai data mining. Ini adalah suatu proses di mana informasi berharga dan pengetahuan terkait ditemukan dan diidentifikasi dalam basis data besar dengan memanfaatkan teknik-teknik seperti statistik, matematika, kecerdasan buatan, dan pembelajaran mesin[6]. Data Mining juga dapat dipahami sebagai rangkaian siklus yang bertujuan untuk memisahkan nilai tambah sebagai informasi yang sebelumnya tidak dapat dibedakan secara fisik dari indeks informasi.

Pertumbuhan pesat dalam bidang data mining didorong oleh berbagai faktor yang meliputi:

1. Pertumbuhan Volume Data dalam Batch: Kuantitas data yang dihasilkan secara terus-menerus semakin meningkat, baik dalam skala perusahaan maupun global.
2. Penyimpanan Data dalam Gudang Data: Perusahaan semakin cenderung menyimpan data mereka dalam gudang data yang dapat diandalkan.
3. Meningkatnya Akses Data melalui Web dan Intranet: Akses mudah ke data melalui web dan intranet memungkinkan organisasi untuk mengintegrasikan data dari berbagai sumber dan menggunakannya untuk analisis yang lebih baik.
4. Tekanan Persaingan Bisnis dalam Ekonomi Global: Persaingan bisnis yang ketat dalam ekonomi global mendorong organisasi untuk mencari cara-cara baru untuk mendapatkan keunggulan kompetitif.
5. Kemajuan Teknologi Perangkat Lunak: Kemajuan dalam perangkat lunak data mining membuat alat-alat ini lebih mudah diakses dan digunakan oleh berbagai kalangan, tidak hanya oleh para ahli.
6. Pertumbuhan Kapasitas Komputasi dan Media Penyimpanan: Perkembangan komputasi dan kapasitas media penyimpanan telah memungkinkan analisis data yang lebih kompleks dan menyeluruh.

Informasi Data Mining dapat dipecah menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi: Tugas ini fokus pada usaha untuk menggambarkan pola dan tren yang terdapat dalam data. Peneliti dan analis data mencoba untuk memahami karakteristik data yang ada.
2. Estimasi: Estimasi mirip dengan klasifikasi, namun berfokus pada variabel target numerik daripada kategorikal.
3. Prediksi: Serupa dengan klasifikasi dan estimasi, namun dalam prediksi, nilai hasilnya terletak di masa depan. Tujuannya adalah untuk memprediksi nilai atau peristiwa yang akan terjadi.
4. Klasifikasi: Dalam klasifikasi, terdapat variabel target berupa kategori.
5. Pengklusteran: Pengklusteran melibatkan pengelompokan catatan atau objek yang memiliki kesamaan.
6. Asosiasi: Dalam konteks data mining, asosiasi adalah tentang menemukan atribut yang sering muncul bersama-sama dalam data.

2.2. Klasifikasi

Pengklasifikasian adalah suatu teknik yang digunakan untuk memeriksa perilaku dan atribut kelompok yang sudah ditentukan sebelumnya. Dengan menggunakan teknik ini, data yang telah dikelompokkan dapat diolah, dan berdasarkan hasil analisisnya, dapat dibuat aturan-aturan untuk mengklasifikasikan data baru ke dalam kelompok yang sesuai. Teknik ini sangat berguna dalam mengorganisir dan mengelompokkan data serta membuat proses klasifikasi data yang lebih efisien[7]. Pengklasifikasian sering juga merujuk pada proses pencarian model atau fitur yang dapat menggambarkan serta membedakan kelas atau konsep data. Model ini berasal dari analisis serangkaian data pelatihan (yaitu objek data yang diidentifikasi dengan label kelas) dan digunakan untuk memprediksi label kelas objek yang tidak diketahui.

Pada dasarnya, pengklasifikasian adalah menentukan label kelas dari rekaman data baru berdasarkan data yang telah diberi label kelas sebelumnya [8]. Untuk membuat klasifikasi dari data, langkah-langkah berikut dilakukan:

1. Kelas adalah variabel dependen atau hasil dari pengklasifikasian.
2. Variabel prediktor adalah variabel independen.
3. Pola atau ciri dari data yang akan diklasifikasikan digunakan untuk menentukan.
4. Kumpulan Data Pelatihan: Data pelatihan terdiri dari contoh-contoh data yang sudah memiliki label kelas yang diketahui.
5. Kumpulan DataUji: adalah informasi modern yang akan dijalankan melalui tampilan klasifikasi setelah penanganan persiapan selesai. Tujuannya adalah untuk mengklasifikasikan data ini dan menguji sejauh mana model dapat melakukan klasifikasi dengan benar.
6. Akurasi model yang dihasilkan dievaluasi untuk menentukan seberapa baik klasifikasi dilakukan.

2.3. Decision Tree (Pohon Keputusan)

Pohon Keputusan (decision tree) adalah diagram berstruktur seperti pohon yang mewakili setiap simpul sebagai atribut yang diuji, setiap cabang sebagai hasil dari pengujian tersebut, dan setiap daun sebagai kelas atau distribusi kelas tertentu[9]. klasifikasi menggunakan pohon keputusan adalah teknik klasifikasi yang sederhana namun banyak digunakan dalam berbagai bidang. Bagian ini akan menjelaskan cara kerja pohon keputusan dan metode pembuatannya. Untuk mengklasifikasikan objek, sering kali sejumlah pertanyaan dilakukan sebelum menentukan grup.

Pohon keputusan memiliki banyak variasi dalam modelnya dengan berbagai tingkat kemampuan dan persyaratan. Namun, beberapa karakteristik umum yang cocok untuk mengimplementasikan pohon keputusan adalah:

1. Data diwakili sebagai pasangan atribut dan nilai. Sebagai contoh, atribut data bisa berupa suhu dengan nilai rendah. Biasanya, untuk satu atribut, terdapat hanya beberapa tipe

nilai yang mungkin. Misalnya, Sifat warna produk alami mungkin memiliki beberapa nilai yang dapat dibayangkan seperti hijau, kuning, dan kemerahan.

2. Data Keluaran Diskrit: Pohon keputusan biasanya digunakan untuk kasus di mana data keluaran atau label adalah diskrit atau memiliki beberapa kelas. Contohnya adalah klasifikasi "ya" atau "tidak," "sakit" atau "tidak sakit," atau "diterima" atau "ditolak." Meskipun pohon keputusan sering digunakan untuk masalah biner, mereka juga dapat menangani lebih dari dua kelas.
3. Terdapat nilai yang hilang dalam data. Misalnya, dalam beberapa kasus, kita mungkin tidak tahu nilai suatu atribut. Dalam situasi ini, pohon keputusan masih dapat memberikan solusi yang baik.

2.4. Rapid Miner

RapidMiner adalah sebuah aplikasi data mining yang bersifat open source atau sumber terbuka. Lisensi yang digunakan untuk RapidMiner adalah AGPL (GNU Affero General Public License) versi 3. Penelitian dan pengembangan alat ini dimulai oleh beberapa individu, termasuk Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer dari departemen kecerdasan buatan Universitas Dortmund pada tahun 001, dan kemudian diambil alih oleh SourceForge. Pada tahun 010- 011, dalam survei KDnuggets, sebuah media data mining, Pencapaian RapidMiner menduduki peringkat pertama sebagai alat data mining untuk proyek nyata adalah prestasi yang signifikan. Ini menunjukkan bahwa RapidMiner telah terbukti sangat efektif dan relevan dalam dunia praktis analisis data dan data mining.

Dalam penerapannya, Rapid Miner memberikan metode penambangan informasi dan pembelajaran mesin yang menggabungkan ETL (Extricate, Change, Stack), preprocessing informasi, visualisasi, pemodelan dan penilaian. Persiapan penambangan informasi di Rapid Miner digambarkan dalam format XML yang dapat diatur melalui antarmuka grafis (GUI). Hal ini membuat penggunaan dan pengawasan alur kerja penambangan informasi menjadi lebih mudah. Instrumen Rapid Miner disusun dalam dialek pemrograman Java dan berkoordinasi dengan usaha dan wawasan penambangan informasi Weka.

Beberapa solusi yang ditawarkan oleh Rapid Miner adalah sebagai berikut:

- Integrasi data, Analitis ETL, Data Analisis, dan pelaporan dalam satu suite tunggal.
- Powerfull tetapi memiliki antarmuka pengguna grafis yang intuitif untuk desain anakisis proses.
- Repositori untuk prose, data dan penanganan meta data.
- Hanya solusi dengan transformasi meta data: lupakan trail and arror dan memeriksa hasil yang telah diinspeksi selama desain.
- Hanya solusi yang mendukung on-the-fly kesalahan dan dapat melakukan perbaikan dengan cepat. Lengkap dan fleksibel: ratusan loading data, transformasi data, pemodelan data dan metode visualisasi data.

2.5. Algoritma C4.5

Merupakan salah satu perhitungan yang digunakan dalam susunan pohon pilihan, yang mengubah kebenaran dari sejumlah besar data menjadi struktur pohon pilihan yang menyesuaikan dengan aturan. Tujuan utama dari pembentukan pohon keputusan menggunakan algoritma C4.5 adalah untuk memudahkan pemecahan masalah dengan mengorganisir data dan aturan ke dalam struktur yang dapat dengan mudah diinterpretasikan dan digunakan untuk pengambilan keputusan. Terdapat beberapa tahapan umum dalam menggunakan algoritma C4.5. Algoritma C4.5 merangkum prosesnya dalam tiga langkah: mengubah tabel data menjadi pohon keputusan, mengonversi pohon menjadi aturan, dan menyederhanakan aturan[4 Secara garis besar, langkah-langkah membuat pohon keputusannya pilihan dengan perhitungan C4.5 adalah:

1. Cari tahu berapa banyak informasi berdasarkan kualitas hasil individu dengan keadaan tertentu dalam berapa banyak informasi.
2. Pilih sorotan sebagai hub.
3. Buatlah cabang untuk setiap bagian hub.
4. Periksa apakah nilai entropi bagian hub adalah 0. Jika ya, tentukan daun yang bersangkutan.
5. Jika nilai entropi hub lebih besar dari 0, ulangi siklus dari awal hingga semua anggota hub bernilai 0.
6. Sebagai node, pilih atribut dengan gain terbesar.

Rumusnya menghitung gain atribut ialah sebagai berikut:

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

Informasi:
S : entropi dari himpunan data
A : gain atribut yang akan dihitung
n : Jumlah segmen-A
|Si| : Jumlah kasus dalam paket ke-i
|S| : Jumlah kasus di-S

Perhitungan nilai entropi dapat dilihat dalam persamaan berikut.

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

Informasi:
S : Nilai entropi dari himpunan data
A : Fitur
n : Jumlah Individu S
pi : Sebatas Si sampai S

3. METODE PENELITIAN

3.1. Tipe Penelitian

Jenis penelitian yang diterapkan dalam studi ini adalah penelitian kuantitatif. Pendekatan ini fokus pada pengumpulan dan analisis data dalam bentuk tabel, angka, serta diagram untuk menghasilkan temuan penelitian. Selain itu, penelitian ini juga menggunakan pendekatan studi literatur, yang melibatkan pencarian dan pengumpulan referensi serta ide dari literatur yang ada sebagai panduan untuk menyelesaikan penelitian. Dengan demikian, kedua pendekatan ini digunakan secara bersamaan untuk memberikan pemahaman komprehensif dan mendukung hasil penelitian.

3.2. Sumber dan Bahan Penelitian

Sumber dan bahan penelitian untuk studi ini mencakup data hasil seleksi kompetensi dasar (SKD) pada seleksi CPNS Universitas Pattimura tahun 2021. Terdapat 350 peserta yang menjadi subjek penelitian. Untuk mengolah data, penelitian ini menggunakan aplikasi Rapid Miner.

3.3. Variabel Penelitian

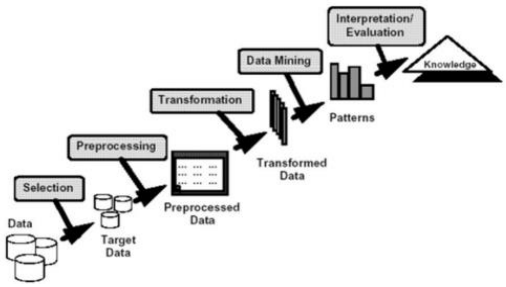
Variabel penelitian yang ada dalam penelitian ini terdiri dari 8 atribut yang diambil dari hasil tes SKD. Berikut merupakan daftar atribut data yang ada pada Tabel 1:

Tabel 1 Atribut Data

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

3.4. Tahapan Penelitian

Tahapan penelitian dilakukan dengan pendekatan Pengungkapan Informasi dalam Basis Data (KDD). KDD dapat menjadi strategi untuk menemukan dan membedakan desain informasi yang dapat dimanfaatkan sebagai informasi. KDD melibatkan lima tahap utama, yaitu Pemilihan (Selection), Pra-pemrosesan (Preprocessing), ransformasi (Transformation), Data mining, dan Evaluasi[10]. Untuk melihat struktur KDD yang dijelaskan dalam berikut:



Gambar 1 Knowledge Discovery in Database

Informasi:

1. Tahap utama adalah memilih informasi yang relevan dan sesuai dengan sasaran yang diselidiki.
2. Pra-pemrosesan adalah pada tahap ini, data yang telah terpilih akan dipersiapkan untuk analisis lebih lanjut.
3. Transformasi Data yang sudah dipersiapkan akan diubah menjadi format yang sesuai untuk analisis data mining.
4. Data mining adalah inti dari proses KDD. Data mining melibatkan penggunaan teknik-teknik khusus untuk menggali pola dan informasi berharga dari data yang telah dipersiapkan.
5. Evaluasi adalah hasil dari analisis data mining dievaluasi untuk memastikan keakuratan dan relevansinya terhadap tujuan penelitian.

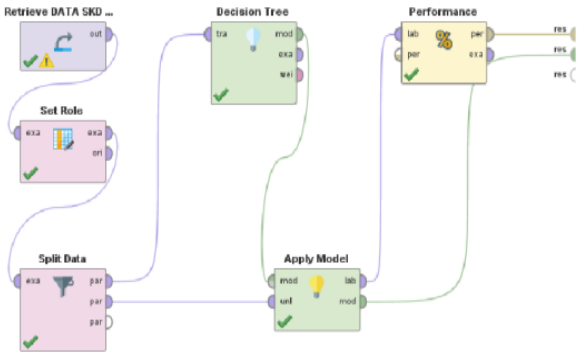
4. HASIL PEMBAHASAN

4.1. Pengujian Data

Informasi yang dapat jika digunakan untuk mengklasifikasikan ujian Pilihan Kompetensi Dasar (SKD) CPNS di Perguruan Tinggi Pattimura Tahun 2021 terdiri dari 350 informasi dengan 8 kualitas. Persiapan klasifikasi dilakukan dengan menggunakan program Rapid Miner Studio adaptasi 9. Dengan data ini dan menggunakan RapidMiner Studio, penelitian ini bertujuan untuk melakukan analisis dan klasifikasi data untuk mendapatkan pemahaman yang lebih baik tentang hasil ujian SKD CPNS.

4.1.1. Model Algoritma Decision Tree

Model yang dibuat dengan menggunakan perangkat lunak Rapid Miner melibatkan beberapa operator seperti Set Role, Split Data, Decision Tree, Apply Model, dan Performance. Gambar 3 menunjukkan operator Retrieve Data yang digunakan dalam proses data mining.



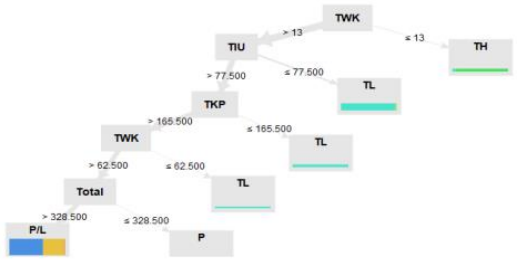
Gambar 2 Model Algoritma Decision Tree

Operator pengatur peran digunakan untuk mengubah peran dari satu atau lebih atribut yang berperan sebagai 'deskripsi', yaitu atribut yang mengubah peran dari target peran. Operator pemisahan data digunakan untuk membagi data menjadi partisi dengan rasio 0,7 dan ,3. Dalam penelitian ini, Operator Pohon Keputusan digunakan

untuk membuat model Pohon Keputusan. Selanjutnya, Operator Penerapan Model digunakan untuk menerapkan model Pohon Keputusan tersebut ke dalam kumpulan data. Evaluasi kinerja model dilakukan dengan menggunakan Operator Kinerja, yang mencakup metrik-metrik seperti Akurasi, Galat Klasifikasi, Kappa, Ingat, dan Presisi. Melalui evaluasi ini, penelitian dapat menilai sejauh mana model Pohon Keputusan mampu mengklasifikasikan hasil ujian SKD CPNS dengan tingkat akurasi dan efisiensi yang sesuai.

4.1.2. Model Tree

Model Pohon dalam konteks ini merupakan representasi struktural hasil dari pemrosesan data SKD menggunakan algoritma Pohon Keputusan dalam RapidMiner. Dalam struktur pohon ini, kelas-kelas hasil klasifikasi dibedakan menjadi empat kategori: P/L (memenuhi ambang batas dan memenuhi syarat untuk berpartisipasi dalam Seleksi Kompetensi Bidang/SKB), P (memenuhi ambang batas), TL (tidak memenuhi ambang batas), dan TH (tidak ada). Model Pohon ini memberikan penjelasan tentang bagaimana hasil klasifikasi dipengaruhi oleh variabel-variabel seperti TIU, TWK, TKP, dan total. Dalam representasi visualnya, Model Pohon dimulai dari akar, yang dalam konteks ini adalah variabel TWK, dan memiliki cabang-cabang yang mengarah ke daun-daun pohon, masing-masing mewakili kategori klasifikasi seperti P/L, TL, P, dan TH.. Gambar menunjukkan Model Pohon yang dihasilkan dari proses dalam RapidMiner, sementara Tabel menyajikan hasil klasifikasinya. Model Pohon ini membantu dalam memahami faktor-faktor yang memengaruhi hasil klasifikasi SKD CPNS, sementara tabel memberikan detail hasil klasifikasinya berdasarkan kategori yang telah disebutkan.



Gambar 3 Model Tree

Dalam Tabel, terdapat penjelasan mengenai hasil klasifikasi yang dihasilkan oleh Model Pohon. Hasil klasifikasi ini dinyatakan melalui sejumlah kondisi nilai yang digunakan untuk menentukan keanggotaan dalam kelas-kelas tertentu. Ketika salah satu dari kondisi nilai tersebut terpenuhi, data akan diklasifikasikan ke dalam kelas yang sesuai. Kelas P/L, P, dan TH memiliki satu kondisi nilai tunggal untuk klasifikasinya, yang berarti hanya ada satu syarat yang harus dipenuhi agar data termasuk dalam kelas tersebut. Di sisi lain, Kelas TL memiliki empat kondisi nilai yang berbeda untuk klasifikasinya. Ini berarti ada empat kriteria yang harus dipenuhi agar data diklasifikasikan ke dalam Kelas TL. Dengan demikian, Tabel memberikan detail tentang kondisi-kondisi ini dan cara data diklasifikasikan ke dalam berbagai kelas berdasarkan nilai-nilai yang ada.

Tabel 1 Hasil Klasifikasi

No	Kelas	Nilai
1	P/L	TWK >13&&TIU >77&&TKP >165&&TWK 6&&
	TH	TWK <13
3	P	Total <38
4	TL	TWK >13&&TIU<77
		TWK >13&&TKP >165
		TWK>13&&TIU<77&&TKP>165&&TWK <6

4.1.3. Performa Model Decision Tree

Kinerja model Pohon Keputusan dalam mengklasifikasikan hasil ujian SKD dievaluasi dengan menggunakan matriks kebingungan dan beberapa indikator kinerja, seperti akurasi, galat klasifikasi, kappa, ingat, dan presisi. Ini membantu dalam menilai sejauh mana model tersebut efektif dalam mengidentifikasi dan mengklasifikasikan hasil ujian SKD dengan akurat. Tabel 3 menampilkan kinerja model Pohon Keputusan, sementara Tabel 4 berisi matriks kebingungannya.

Tabel 2 Performa Model Decision Tree

No	Jenis Perfoma	Nilai
1	Accuracy	75%
	Classification error	5%
3	Kappa	0,947
4	Recal	97,14%
5	Precision	93,94%

Tabel 3 Confusion Matrix

True	P/L	P	TL	TH
P/L	34	0	0	0
P	1	1	0	0
TL	0	0	31	0
TH	0	0	0	9

Sebagian dari kinerja dalam Tabel 3 diambil dari Matriks Kebingungan dalam Tabel 4. Kinerja akurasi adalah nilai akurasi dari model yang digunakan. Nilai akurasi yang diperoleh adalah 75,0%, menunjukkan bahwa model ini sangat baik dalam memprediksi hasil. Kinerja model dapat diukur dengan beberapa metrik. Galat klasifikasi merupakan ukuran tingkat kesalahan yang ditemukan dalam model. Dalam konteks ini, jika terdapat sedikit kesalahan dalam model, nilai galat klasifikasi yang diperoleh adalah sebesar 5,0%.. Kinerja Kappa adalah ukuran yang dinormalisasi dari akurasi model. Nilai Kappa yang diperoleh adalah sekitar 0,947, mendekati nilai 1. Nilai Kappa mendekati kesempurnaan, menandakan bahwa model ini memiliki kinerja yang sangat baik dalam mengklasifikasikan data. Selanjutnya, nilai Recall dan Presisi juga merupakan metrik penting. Nilai Recall sebesar 94,84% menunjukkan kemampuan model

dalam mengidentifikasi semua contoh positif secara benar. Sementara itu, nilai Presisi sebesar 95,79% mengukur kemampuan model untuk mengidentifikasi contoh positif dengan benar dari semua yang diidentifikasi sebagai positif. Dengan hasil ini, dapat disimpulkan bahwa ini mempunyai kinerja yang sangat baik dalam mengklasifikasikan ujian SKD yang dihasilkan, dengan tingkat akurasi yang tinggi, nilai Kappa yang mendekati kesempurnaan, serta nilai Recall dan Presisi yang sangat baik.

5. KESIMPULAN

Berdasarkan kejadian dan wacana, pemeriksaan ini melibatkan 350 anggota SKD dengan 8 kualitas dalam klasifikasi CPNS Perguruan Tinggi Pattimura 2021. Hasil penelitian menyimpulkan bahwa hasil SKD peserta terpilih dapat diklasifikasikan menjadi 4 kelas dengan menggunakan model pohon keputusan. Nilai kinerja dari model Pohon Keputusan yang diperoleh adalah Akurasi 75,0%, Galat Klasifikasi = 5,0%, Kappa = 0,947, Recall = 97,14%, Presisi = 93,94%. Hal ini membuktikan bahwa Algoritma Pohon Keputusan cocok digunakan sebagai model klasifikasi dalam penelitian ini. Penelitian ini juga dapat dikembangkan dengan mengimplementasikan algoritma data mining lain seperti Naive Bayes, K-Means, KNN untuk melakukan perbandingan kinerja di antara algoritma-algoritma tersebut. Selanjutnya, penelitian ini memiliki potensi untuk pengembangan lebih lanjut dengan mengaktualisasikan perhitungan penambangan informasi lainnya seperti Naive Bayes, K-Means, dan K-Nearest Neighbors (KNN) untuk melakukan perbandingan kinerja di antara berbagai algoritma tersebut. Hal ini dapat memberikan wawasan tambahan tentang metode mana yang paling sesuai untuk mengklasifikasikan hasil ujian SKD CPNS Universitas Pattimura atau bahkan untuk penelitian lebih lanjut dalam domain ini.

DAFTAR PUSTAKA

[1] D. M. Kusuma, "Kinerja Pegawai Negeri Sipil (PNS) Di Kantor Badan Kepegawaian Daerah Kabupaten Kutai Timur," *eJournal Adm. Negara*, pp. 1388–1400, 2013.

[2] T. Kristiana, "Penerapan Profile Matching Untuk Penilaian Kinerja Pegawai Negeri Sipil (Pns)," *J. Pilar Nusa Mandiri Vol. XI, No.2 Sept. 2015 PENERAPAN*, vol. 11, no. 2, pp. 161–170, 2015.

[3] K. Sani, W. Wahyu Winarno, and S. Fauziati, "Analisis Perbandingan Algoritma Classification Untuk Authentication Uang Kertas (Studi Kasus: Banknote Authentication)," *J. Inform.*, vol. 10, no. 1, pp. 1130–1139, 2016, doi: 10.26555/jifo.v10i1.a3344.

[4] R. R. Fiska, "Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)," *SATIN - Sains dan Teknol. Inf.*, vol. 3, no. 1, pp. 15–23, 2017, doi: 10.33372/stn.v3i1.200.

[5] D. Alverina, A. R. Chrismanto, and R. G. Santosa, "Perbandingan Algoritma C4.5 dan CART dalam Memprediksi Kategori Indeks Prestasi Mahasiswa," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 2, pp. 76–83, 2018, doi: 10.14710/jtsiskom.6.2.2018.76-83.

[6] S. H. Ha and S. H. Joo, "A hybrid data mining method for the medical classification of chest pain," *World Acad. Sci. Eng. Technol.*, vol. 37, pp. 608–613, 2010.

[7] A. Haditsah, "Klasifikasi Masyarakat Miskin menggunakan Metode Naïve Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.

[8] T. Dudkina, I. Meniailov, K. Bazilevych, S. Krivtsov, and A. Tkachenko, "Classification and prediction of diabetes disease using decision tree method," *CEUR Workshop Proc.*, vol. 2824, pp. 163–172, 2021.

[9] E. Turban, J. E. Aronson, and T.-P. Liang, "Decision Support System and Intelligent System," *Penerbit: Andi, Yogyakarta*. p. 936, 2010.

[10] R. D. Syah, "Metode Decision Tree Untuk Klasifikasi Hasil Seleksi Kompetensi Dasar Pada Cpns 2019 Di Arsip Nasional Republik Indonesia," *J. Ilm. Inform. Komput.*, vol. 25, no. 2, pp. 107–114, 2020, doi: 10.35760/ik.2020.v25i2.2750.