JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 3, No. 2, Desember 2024, hlmn. 180-186
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v3i2.5803

Implementasi Prediksi Penyakit Ginjal Kronis dengan Menggunakan Metode Decision Tree

Arifah Amaliyah Qori'ah[1*], Zaehol Fatah[2]

[1] Teknologi Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia
[2] Sistem Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **09-11-2024**
Direvisi : **13-11-2024**
Disetujui : **15-12-2024**

Kata Kunci:

Ginjal Kronis, Decision Tree RapidMiner, Prediksi Penyakit, Data Mining

ABSTRAK

Penyakit ginjal kronis merupakan masalah kesehatan serius yang memerlukan deteksi dini untuk penanganan yang tepat. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit ginjal kronis menggunakan metode Decision Tree dengan bantuan aplikasi RapidMiner. Data yang digunakan berasal dari platform Kaggle dengan 22 parameter kesehatan seperti tekanan darah, kadar gula darah, dan hasil tes laboratorium lainnya. Proses penelitian meliputi empat tahap utama: pengumpulan data, *preprocessing* menggunakan RapidMiner, pembuatan model Decision Tree, dan evaluasi performa. Preprocessing data mencakup pembacaan data dengan Read Excel, pengaturan *Set Role* untuk klasifikasi, dan pembagian *dataset* menggunakan Split Data. Model Decision Tree diimplementasikan dengan parameter *gain ratio*, *maximal depth* 10, serta fitur *pruning* dan *prepruning*. Hasil penelitian menunjukkan tingkat akurasi yang sangat baik yaitu 98.12%, dengan nilai presisi 65.34% dan *recall* 65.88%. Dari *confusion matrix* terlihat bahwa model hanya melakukan 5 kesalahan prediksi dari total *dataset*, yang terdiri dari 2 *False Positive* dan 3 *False Negative*. Model ini diharapkan dapat membantu tenaga medis dalam melakukan screening awal dan diagnosis penyakit ginjal kronis secara lebih efektif dan efisien.

***Keywords*:**

Chronic Kidney, Decision Tree, RapidMiner, Disease Prediction, Data Mining

ABSTRACT

Chronic kidney disease is a serious health concern that requires early detection for proper treatment. This research aims to develop a prediction model for chronic kidney disease using the Decision Tree method with RapidMiner application. The data used comes from the Kaggle platform with 22 health parameters such as blood pressure, blood sugar levels, and other laboratory test results. The research process includes four main stages: data collection, preprocessing using RapidMiner, Decision Tree model development, and performance evaluation. Data preprocessing includes data reading with Read Excel, Set Role configuration for classification, and dataset division using Split Data. The Decision Tree model was implemented with gain ratio parameters, maximal depth of 10, and pruning and prepruning features. The results show excellent accuracy at 98.12%, with precision value of 65.34% and recall of 65.88%. The confusion matrix shows that the model only made 5 prediction errors from the total dataset, consisting of 2 False Positives and 3 False Negatives. This model is expected to assist medical personnel in conducting initial screening and diagnosis of chronic kidney disease more effectively and efficiently.

Penulis Korespondensi:
Arifah Amaliyah Qori'ah,
Program Studi Teknologi Informasi
Universitas Ibrahimy, Situbondo, Indonesia
Email: arifahamaliyah1@gmail.com

This is an open access article under the CC BY-SA license

1. PENDAHULUAN

Penyakit ginjal kronis (*Chronic Kidney Disease*/CKD) saat ini menjadi masalah kesehatan yang sangat serius dan terus meningkat di Indonesia maupun dunia. Data dari Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa sekitar 850 juta orang saat ini hidup dengan berbagai gangguan ginjal. Di Indonesia sendiri, menurut data Kementerian Kesehatan tahun 2023, penderita penyakit ginjal kronis terus bertambah dan menjadi salah satu penyebab kematian tertinggi[1]. Hal yang memprihatinkan adalah banyak orang tidak menyadari mereka mengidap penyakit ini karena gejalanya yang tidak terlihat di awal, dan baru mengetahuinya saat kondisi sudah parah. Untuk mendiagnosis penyakit ginjal kronis, dokter perlu memeriksa banyak hal seperti tekanan darah pasien, kadar gula darah, protein dalam darah (albumin), fungsi ginjal (*kreatinin*), dan berbagai tes laboratorium lainnya. Seiring dengan majunya teknologi, kini kita bisa memanfaatkan teknologi *data mining* dan *machine learning* untuk membantu dokter mendiagnosis penyakit ini lebih cepat dan tepat[2]. Dalam penelitian ini, kami menggunakan data dari Kaggle, sebuah situs yang menyediakan data-data kesehatan dari berbagai rumah sakit dan klinik, khususnya data pasien yang berkaitan dengan penyakit ginjal kronis.

Penelitian kami menggunakan metode Decision Tree yang dijalankan dengan *software* RapidMiner, sebuah program yang mudah digunakan untuk mengolah data. Decision Tree ini bekerja seperti pohon keputusan yang bercabang, mirip dengan cara dokter mengambil keputusan berdasarkan berbagai gejala dan hasil tes pasien. Keunggulan metode ini adalah hasilnya mudah dipahami, tidak hanya oleh ahli teknologi tapi juga oleh dokter dan tenaga medis[3]. RapidMiner membantu kami mengolah data mentah, menguji keakuratan model, dan mengukur seberapa tepat prediksi yang dihasilkan melalui berbagai pengukuran seperti tingkat akurasi, ketepatan, dan kelengkapan hasil prediksi. Penelitian ini bertujuan untuk membuat sistem yang bisa membantu dokter mendeteksi penyakit ginjal kronis lebih awal dan akurat. Kami menggunakan data dari Kaggle dan mengolahnya dengan RapidMiner untuk membuat model prediksi yang handal. Harapannya, hasil penelitian ini bisa membantu dokter dan rumah sakit di Indonesia melakukan pemeriksaan awal pasien yang diduga mengidap penyakit ginjal kronis[4]. Ke depannya, sistem ini bisa dikembangkan menjadi alat bantu yang lebih lengkap untuk mendukung keputusan dokter dalam menangani pasien. Meskipun sudah ada beberapa penelitian yang menggunakan *machine learning* untuk mendeteksi penyakit ginjal kronis, penggunaan Decision Tree dengan RapidMiner yang kami lakukan ini memiliki keunikan tersendiri. Kami fokus pada bagaimana membuat sistem yang tidak hanya akurat, tapi juga mudah dipahami oleh tenaga medis. Penelitian ini juga memperhatikan berbagai aspek penting dalam pengolahan data medis, pemilihan parameter yang tepat untuk Decision Tree, dan pengujian menyeluruh terhadap keakuratan hasil prediksi yang sangat penting dalam dunia kedokteran.

2. METODE PENELITIAN

2.1 Penyakit Ginjal Kronis

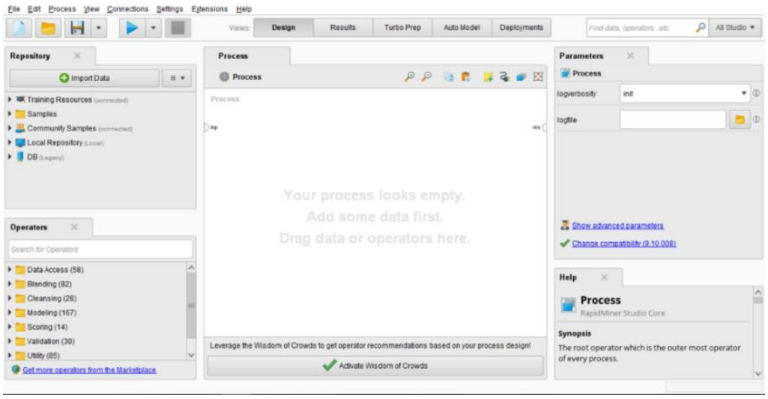
Penyakit ginjal kronis (*Chronic Kidney Disease*/CKD) merupakan kondisi kesehatan serius dimana ginjal mengalami kerusakan dan tidak dapat menyaring limbah serta kelebihan cairan dari darah secara efektif. Kondisi ini berkembang secara perlahan selama berbulan-bulan hingga bertahun-tahun, dan sering kali tidak menunjukkan gejala yang jelas hingga fungsi ginjal mengalami gangguan serius[1]. Diagnosis CKD dilakukan melalui berbagai parameter medis seperti glomerular filtration rate (GFR), kadar *kreatinin*, albumin, tekanan darah, dan berbagai indikator laboratorium lainnya. WHO mengklasifikasikan CKD menjadi 5 stadium berdasarkan tingkat keparahannya, mulai dari kerusakan ginjal ringan dengan fungsi normal hingga gagal ginjal stadium akhir[5].

2.2 *Decission Tree*

Decision Tree merupakan salah satu metode dalam *data mining* yang menggunakan struktur pohon untuk membuat model prediksi. Metode ini bekerja dengan cara memecah *dataset* menjadi *subset* yang lebih kecil berdasarkan kriteria pemisahan tertentu. Setiap *node* dalam pohon mewakili sebuah pertanyaan tentang atribut, sedangkan cabang mewakili jawaban dari pertanyaan tersebut[6]. Decision Tree memiliki keunggulan dalam hal interpretabilitas karena dapat menggambarkan aturan keputusan secara visual dan mudah dipahami. Dalam konteks medis, Decision Tree dapat membantu mengidentifikasi faktor-faktor yang paling berpengaruh dalam diagnosis penyakit dan memberikan alasan yang jelas untuk setiap prediksi yang dihasilkan[7].

2.3 *Rapid Miner*

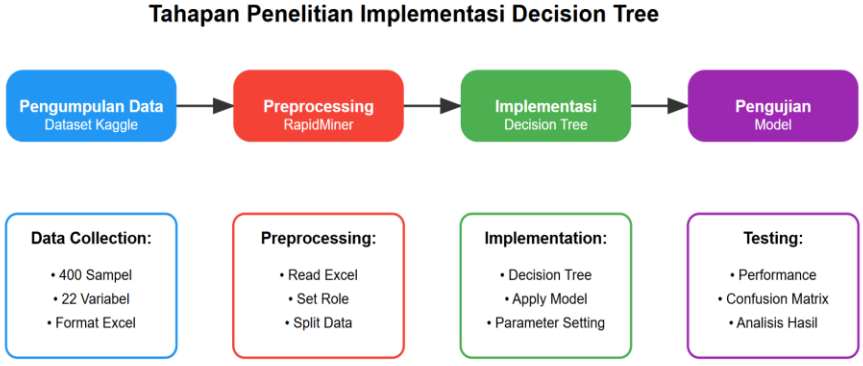
RapidMiner adalah platform *software* yang menyediakan lingkungan terintegrasi untuk *data mining*, *machine learning*, dan analisis prediktif. *Software* ini menawarkan antarmuka visual yang memudahkan pengguna dalam membangun alur kerja analisis data tanpa perlu menulis kode program yang kompleks[8]. RapidMiner mendukung berbagai tahap dalam proses *data mining*, mulai dari preprocessing data, pemodelan, hingga evaluasi dan visualisasi hasil. Dalam konteks klasifikasi medis, RapidMiner menyediakan berbagai operator untuk membangun dan mengevaluasi model Decision Tree, termasuk fitur untuk mengukur performa model melalui metrik seperti akurasi, presisi, *recall*, dan F1-score[9]. Platform ini juga mendukung teknik validasi silang (*cross-validation*) untuk memastikan kehandalan model yang dihasilkan.



Gambar 1. Tampilan Awal *Rapid Miner*

2.5 Tahapan Penelitian

Pengumpulan data merupakan tahap awal penelitian yang melibatkan perolehan *dataset* penyakit ginjal kronis dari platform Kaggle. Data ini mencakup 22 parameter penting untuk diagnosis, yang telah diseleksi dan dibersihkan dari field yang tidak diperlukan. Semua data disusun dalam format Excel untuk memudahkan pengolahan selanjutnya di RapidMiner[10]. Tahap pengolahan data di RapidMiner dimulai dengan proses Read Excel untuk membaca *dataset*, dilanjutkan dengan Set Role untuk menentukan target prediksi, dan Split Data untuk membagi *dataset* menjadi data *training* dan *testing*. Proses ini memastikan data siap untuk digunakan dalam pembuatan model. Pembuatan model Decision Tree menjadi tahap ketiga dengan pengaturan parameter seperti gain ratio sebagai criterion, maximal depth 10, serta mengaktifkan fitur *apply pruning* dan *prepruning* dengan *confidence* 0.1[11]. Model yang dibuat kemudian diterapkan menggunakan *Apply* Model untuk proses pengujian. Tahap terakhir adalah evaluasi performa model menggunakan operator Performance untuk mengukur tingkat keberhasilan prediksi. Hasil evaluasi ini akan mencakup nilai akurasi, presisi, *recall*, dan *confusion matrix* yang akan dibandingkan dengan penelitian-penelitian sejenis untuk menilai efektivitas model yang dibangun[12].

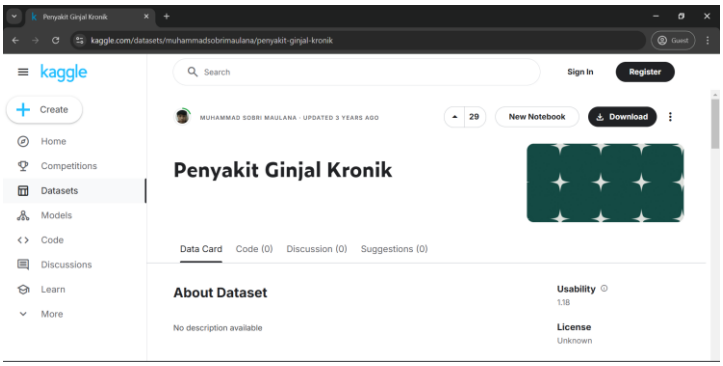


Gambar 2. Tahapan Penelitian

3. HASIL DAN ANALISIS

3.1. Deskripsi Data

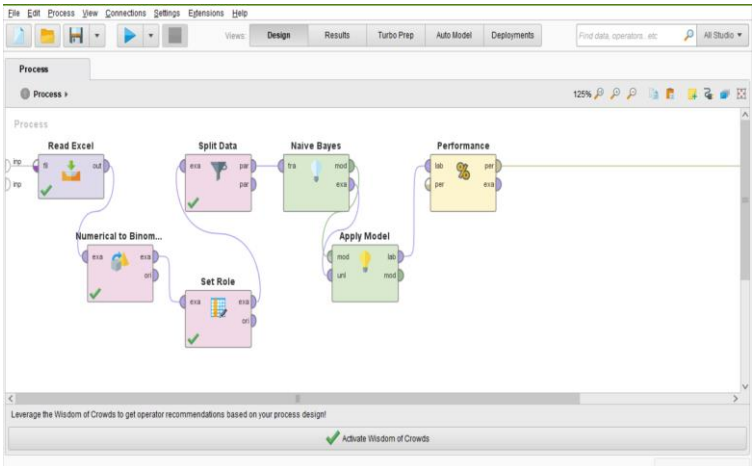
Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle, terdiri dari 400 data pasien yang mencakup berbagai parameter medis terkait penyakit ginjal kronis. *Dataset* ini memiliki total 25 atribut, yang terbagi menjadi 24 atribut input dan 1 atribut target. Dari 24 atribut input tersebut, 13 atribut merupakan data numerik yang mencakup parameter seperti umur, tekanan darah, gravitas, hasil tes laboratorium (*kreatinin*, natrium, kalium, hemoglobin), dan 11 atribut lainnya merupakan data nominal yang menggambarkan kondisi klinis pasien seperti status hipertensi, diabetes, dan kondisi fisik lainnya. Atribut target dalam *dataset* ini adalah klasifikasi yang menunjukkan status pasien, apakah terdiagnosis penyakit ginjal kronis (ckd) atau tidak (notckd). Distribusi kelas dalam *dataset* menunjukkan sebanyak 250 data (62.5%) termasuk dalam kategori ckd dan 150 data (37.5%) termasuk dalam kategori notckd. Meskipun terdapat sedikit ketidakseimbangan kelas, *dataset* masih cukup representatif untuk digunakan dalam pemodelan. Karakteristik data sebelum preprocessing menunjukkan beberapa tantangan yang perlu diatasi. Terdapat beberapa missing value pada beberapa atribut, khususnya pada parameter hasil laboratorium. Beberapa atribut numerik memiliki rentang nilai yang sangat bervariasi, seperti usia pasien yang berkisar antara 2-90 tahun, dan kadar *kreatinin* yang berkisar antara 0.4-15.2 mg/dL. Atribut nominal juga memiliki beberapa variasi format penulisan dan inkonsistensi data yang memerlukan standardisasi dalam tahap preprocessing.



Gambar 3. Halaman Tampilan Kaggle Ginjal Kronik

3.2. Preprocessing Data

Dalam tahap preprocessing data, penelitian ini mengimplementasikan beberapa tahapan menggunakan RapidMiner dengan alur proses yang terdiri dari Read Excel, Set Role, Split Data, Decision Tree, dan *Apply* Model. *Dataset* yang digunakan terdiri dari 22 atribut setelah dilakukan pembuangan field "seldarahmerah yang kedua" dan "cad" untuk menghindari redundansi dan meningkatkan efisiensi model.



Gambar 4. Tahapan proses *Data Mining*

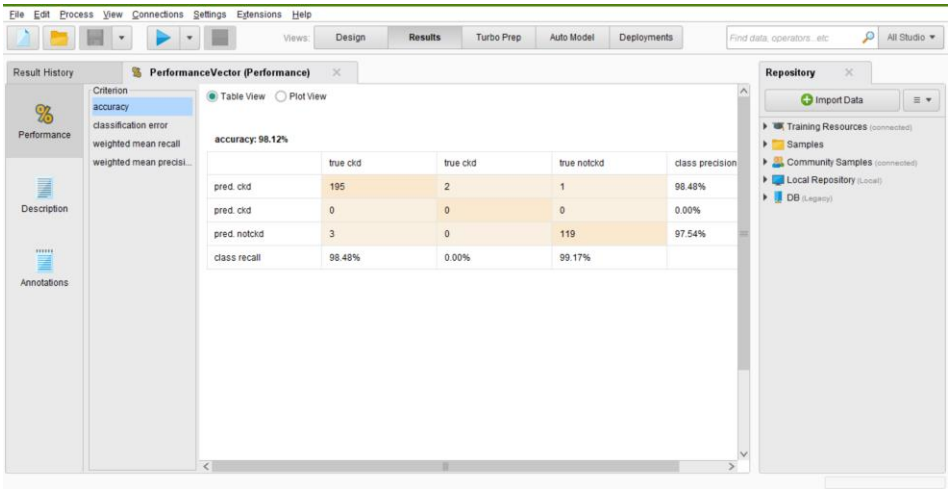
Tahap pertama dimulai dengan proses Read Excel untuk mengimpor *dataset*. Kemudian dilanjutkan dengan Set Role untuk menentukan atribut target yaitu "klasifikasi" sebagai label yang akan diprediksi. Operator Split Data digunakan untuk membagi *dataset* menjadi data *training* dan *testing* dengan rasio yang

telah ditentukan untuk proses validasi model. *Dataset* yang diproses memiliki tiga kategori tipe data yang berbeda. Atribut numerik integer meliputi umur (dalam tahun), tekanan darah (mmHg), albumin (g/dL), sugar, gds (mg/dL), ureum (mg/dL), MCV (fL), dan seldarahputih (cells/mm^3). Atribut numerik real mencakup gravitas, *kreatinin* (mg/dL), natrium (mEq/L), kalium (mEq/L), dan hemoglobin (g/dL). Sedangkan atribut kategorikal terdiri dari seldarahmerah (normal/abnormal), pussel (present/notpresent), puscell (present/notpresent), bakteri (present/notpresent), hipertensi (yes/no), nafsumakan (good/poor), edema (yes/no), dan anemia (yes/no).

3.3 Implementasi *Decission Tree*

Penerapan metode Decision Tree (Pohon Keputusan) dalam penelitian ini menggunakan aplikasi RapidMiner dengan beberapa pengaturan penting untuk mendapatkan hasil prediksi yang optimal. Pengaturan yang digunakan meliputi criterion (kriteria) yang dipilih gain_ratio untuk menentukan cara pemisahan data, kedalaman pohon (maximal depth) dibatasi sampai 10 tingkat agar tidak terlalu kompleks, serta menggunakan fitur pemangkasan pohon (*apply pruning* dan *prepruning*) dengan tingkat kepercayaan 0.1 untuk mencegah model menjadi terlalu rumit dan mengikuti noise data.

Bentuk pohon keputusan yang dihasilkan mirip seperti diagram pohon dengan beberapa tingkatan keputusan. Setiap titik percabangan menunjukkan pertanyaan tentang kondisi pasien (misalnya "Apakah hemoglobin di bawah 12?"), setiap cabang menunjukkan jawaban dari pertanyaan tersebut (misalnya "Ya" atau "Tidak"), dan titik akhir menunjukkan kesimpulan apakah pasien tergolong CKD atau tidak. Cara kerja pohon keputusan ini sangat praktis karena menghasilkan aturan-aturan yang mudah dipahami oleh dokter dan petugas medis.



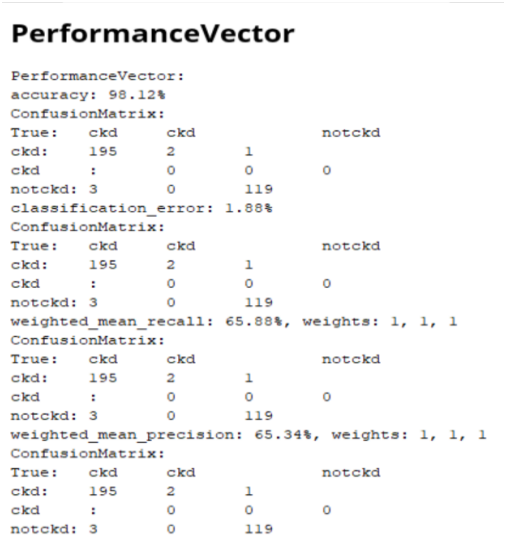
Gambar 5. Hasil Implementasi *Decission Tree*

3.4 Evaluasi Model

Hasil pengujian model Decision Tree untuk memprediksi penyakit ginjal kronis menunjukkan performa yang sangat baik dengan tingkat akurasi mencapai 98.12%. Ini berarti dari setiap 100 prediksi yang dilakukan, sekitar 98 prediksi berhasil memberikan hasil yang tepat. Model juga menunjukkan nilai presisi sebesar 65.34%, yang menggambarkan ketepatan model dalam mengidentifikasi kasus positif CKD, dan nilai *recall* sebesar 65.88% yang menunjukkan kemampuan model dalam mengenali seluruh kasus CKD yang ada. Dari hasil *confusion matrix*, dapat dilihat rincian prediksi sebagai berikut: terdapat 195 kasus True Positive (pasien CKD yang tepat diprediksi sebagai CKD) dan 119 kasus True Negative (pasien non-CKD yang tepat diprediksi sebagai non-CKD). Sementara itu, terdapat beberapa kesalahan prediksi yaitu 2 kasus False Positive (pasien non-CKD yang salah diprediksi sebagai CKD) dan 3 kasus False Negative (pasien CKD yang salah diprediksi sebagai non-CKD). Tingkat kesalahan klasifikasi (classification error) tercatat sebesar 1.88%, yang menunjukkan bahwa model memiliki tingkat kesalahan yang sangat rendah.

Analisis lebih lanjut menunjukkan bahwa model lebih cenderung melakukan kesalahan dalam bentuk False Negative, dimana 3 pasien yang sebenarnya menderita CKD tidak terdeteksi oleh model. Hal ini perlu mendapat perhatian khusus dalam penerapan praktisnya, karena kesalahan tipe ini berisiko mengakibatkan keterlambatan penanganan pasien CKD. Namun secara keseluruhan, dengan tingkat akurasi yang sangat tinggi

dan tingkat kesalahan yang rendah, model ini dapat diandalkan sebagai alat bantu dalam proses diagnosis awal penyakit ginjal kronis.



PerformanceVector

```
PerformanceVector:
accuracy: 98.12%
ConfusionMatrix:
True:   ckd     ckd             notckd
ckd:    195     2       1
ckd     :       0       0       0
notckd: 3       0       119
classification_error: 1.88%
ConfusionMatrix:
True:   ckd     ckd             notckd
ckd:    195     2       1
ckd     :       0       0       0
notckd: 3       0       119
weighted_mean_recall: 65.88%, weights: 1, 1, 1
ConfusionMatrix:
True:   ckd     ckd             notckd
ckd:    195     2       1
ckd     :       0       0       0
notckd: 3       0       119
weighted_mean_precision: 65.34%, weights: 1, 1, 1
ConfusionMatrix:
True:   ckd     ckd             notckd
ckd:    195     2       1
ckd     :       0       0       0
notckd: 3       0       119
```

Gambar 6. Nilai Evaluasi Model

4. KESIMPULAN

Berdasarkan hasil penelitian implementasi metode Decision Tree untuk memprediksi penyakit ginjal kronis menggunakan RapidMiner, dapat disimpulkan bahwa model yang dikembangkan menunjukkan performa yang sangat baik dengan tingkat akurasi mencapai 98.12%. Hasil ini menunjukkan bahwa metode Decision Tree efektif dalam mengklasifikasikan pasien yang berpotensi mengidap penyakit ginjal kronis. Model juga menunjukkan nilai presisi 65.34% dan *recall* 65.88%, yang mengindikasikan kemampuan model dalam mengidentifikasi kasus positif CKD cukup baik. Dari *confusion matrix*, terlihat bahwa model hanya melakukan sedikit kesalahan prediksi, dengan 5 kasus prediksi yang kurang tepat dari total *dataset* yang digunakan.

Untuk pengembangan penelitian selanjutnya, disarankan beberapa poin perbaikan. Pertama, memperbesar jumlah *dataset* untuk meningkatkan keragaman data dan kemampuan generalisasi model. Kedua, menambahkan parameter medis lain yang mungkin berpengaruh dalam diagnosis penyakit ginjal kronis. Ketiga, mengkombinasikan metode Decision Tree dengan algoritma *machine learning* lainnya untuk membandingkan dan potensial meningkatkan akurasi prediksi. Terakhir, mengembangkan antarmuka yang lebih user-friendly agar model dapat diimplementasikan dengan lebih mudah oleh tenaga medis dalam praktik klinis sehari-hari.

REFERENSI

[1] Warite Gire, Eman Sulaiman, and Nirwana, "Hubungan Pengetahuan dan Dukungan Keluarga Dengan Kepatuhan Pasien Gagal Ginjal Kronik Dalam Menjalani Terapi Hemodialisa di RSUD Bahteramas Provinsi Sulawesi Tenggara Tahun 2023," *J. Penelit. Sains dan Kesehat. Avicenna*, vol. 2, no. 3, pp. 60–68, 2023, doi: 10.69677/avicenna.v2i3.62.

[2] F. Mailani, *Edukasi Pencegahan Penyakit Ginjal Kronik (Pgk) Pada Lansia*. 2022.

[3] M. F. R. Assauqi, "DECISION TREE BERDASARKAN GEJALA KLINIS DAN DATA LABORATORIUM," vol. 2, no. 4, pp. 42–53, 2024.

[4] S. K. . M. K. A. M. Siregar and S. K. . M. K. . A. Puspabhuana, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner - Google Buku*. 2021.

[5] Bijak, Hani, and Susianti, *Memahami Interpretasi Penyakit Ginjal Kronis*. 2016.

[6] I. A. Gellysa Urva, Desyanti, *PENERAPAN DATA MINING DI BERBAGAI BIDANG : Konsep, Metode, dan Studi Kasus*. 2023. [Online]. Available:

[7] G. Gunawan, A. Rahmawati, S. Suhada, T. Hidayatulloh, and D. Wintana, "Optimasi Linear Sampling

dan Information Gain pada Algoritma Decision Tree Untuk Diagnosis Penyakit Diabetes," *Multinetics*, vol. 7, no. 2, pp. 124–131, 2022, doi: 10.32722/multinetics.v7i2.3796.

[8] Ismai, *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. 2017.

[9] A. Yogianto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.

[10] A. Wasik *et al.*, "Implementasi *data mining* untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1," vol. 1, no. 2, pp. 469–479, 2024.

[11] A. Jalil, A. Homaidi, and Z. Fatah, "Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2070–2079, 2024, doi: 10.33379/gtech.v8i3.4811.

[12] H. Alrasyid, A. Homaidi, M. Kom, Z. Fatah, and M. Kom, "Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes," vol. 1, no. 1, pp. 447–453, 2024.