JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i1.8951

KLASIFIKASI CHRONIC KIDNEY DISEASE (CKD) MENGGUNAKAN TOOLS WEKA, RAPIDMINER, DAN ORANGE DATA MINING: ANALISIS PERBANDINGAN KINERJA

Agus Wantoro[1*], Heni Sulistiani[2], Debby Alita[3], Permata[4], Catur Aribowo[5]

[1,5]Universitas Aisyah Pringsewu; Jl. Raya A. Yani No. 1A, Tambahrejo, Gadingrejo, Pringsewu
[2,3,4]Universitas Teknokrat Indonesia; Jl. ZA Pagar Alam No 9-11 Kedaton, Bandar Lampung

Keywords:
Analisis perbandingan;
Klasifikasi;
Ginjal;
Machine Learning;

Corespondent Email:
aguswantoro@aisyahuniversity.ac.id

Abstrak. Chronic Kidney Disease (CKD) merupakan salah satu penyakit tidak menular dengan tingkat prevalensi dan mortalitas yang terus meningkat secara global. Deteksi dini CKD sangat penting untuk mencegah komplikasi dan memperpanjang harapan hidup pasien. Penelitian ini bertujuan untuk membandingkan performa algoritma klasifikasi yang diterapkan pada dua platform data mining populer, yaitu WEKA, RapidMiner, dan Orange dalam menganalisis dataset penyakit ginjal kronis dari UCI Machine Learning (ML) Repository. Lima algoritma klasifikasi digunakan dalam eksperimen, yaitu Naive Bayes, Support Vector Machine (SVM), Random Forest, k-NN, dan Logistic Regression dengan skema validasi silang 10-fold. Kinerja model dievaluasi berdasarkan Confusion Matrix berupa nilai accuracy, precision, dan recall. Hasil menunjukkan bahwa terdapat perbedaan performa antar algoritma pada masing-masing tools. Pada tools WEKA, algoritma Random Forest menunjukkan performa terbaik dengan akurasi 99.81% dan algoritma k-NN menunjukkan performa terburuk. Pada tools RapidMiner, algoritma k-NN justru menampilkan nilai terbaik dengan nilai akurasi 99.5%, sedangkan Niave Bayes menyusul di bawahnya. Pada tools Orange algoritma SVM dan Random Forest memiliki performa terbaik dengan nilai 99.8% dan algoritma terburuk k-NN. Secara umum tools WEKA memiliki kinerja yang lebih baik, disusul Orange, dan RapidMiner. Namun, setiap platform memiliki keunggulan masing-masing. Ketiga tools memiliki potensi yang besar dalam pengembangan sistem pendukung keputusan berbasis ML untuk diagnosis CKD.

Abstract. *Chronic Kidney Disease (CKD) is a non-communicable disease with prevalence and mortality rates that continue to increase globally. Early detection of CKD is very important to prevent complications and prolong patient life expectancy. This research aims to compare the performance of classification algorithms applied to two popular data mining platforms, namely WEKA, RapidMiner, and Orange in analysing chronic kidney disease datasets from the UCI Machine Learning (ML) Repository. Five classification algorithms were used in the experiment, namely Naive Bayes, Support Vector Machine (SVM), Random Forest, k-NN, and Logistic Regression with a 10-fold cross validation scheme. Model performance is evaluated based on the Confusion Matrix in the form of accuracy, precision and recall values. The results show that there are differences in performance between algorithms in each tool. In the WEKA tool, the Random Forest algorithm shows the best performance with an accuracy of 99.81% and the k-NN algorithm shows the worst performance. In the RapidMiner tool, the k-NN algorithm actually*

displays the best value with an accuracy value of 99.5%, while Niave Bayes follows below. In the Orange tools, the SVM and Random Forest algorithms have the best performance with a value of 99.8% and the worst is the k-NN algorithm. In general, WEKA tools have better performance, followed by Orange and RapidMiner. However, each platform has its own advantages. These three tools have great potential in developing ML-based decision support systems for CKD diagnosis.

1. PENDAHULUAN

Chronic Kidney Disease (CKD) merupakan salah satu masalah kesehatan global yang serius, dengan tingkat prevalensi yang terus meningkat setiap tahunnya [1]. Deteksi dini CKD sangat penting untuk mencegah komplikasi lebih lanjut dan meningkatkan kualitas hidup pasien [2].

Deteksi dini CKD sangat penting untuk mencegah komplikasi serius dan meningkatkan kualitas hidup pasien[3]. Namun, proses deteksi CKD seringkali memerlukan serangkaian tes laboratorium yang kompleks dan membutuhkan waktu yang lama [4]. Dalam konteks ini data mining menawarkan solusi yang efisien untuk menganalisis data medis dan membantu dalam proses klasifikasi penyakit, termasuk CKD. Namun, tantangan signifikan yang sering dihadapi dalam penerapan model klasifikasi ini adalah ketidakseimbangan kelas dalam dataset, di mana jumlah data pada kelas mayoritas misalnya "ckd" lebih banyak dibandingkan dengan kelas minoritas "not-ckd"

Ketidakseimbangan kelas dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, sehingga mengurangi kemampuan model dalam mendeteksi kasus pada kelas minoritas [5]. Beberapa teknik telah dikembangkan untuk mengatasi masalah ini, seperti Synthetic Minority Over-sampling Technique (SMOTE) [6][7]. Teknik-teknik ini bertujuan untuk menyeimbangkan distribusi kelas dalam dataset sebelum proses pelatihan algoritma pada data mining [7]

Berbagai tools data mining populer seperti WEKA, RapidMiner, dan Orange telah dikembangkan untuk memfasilitasi proses analisis data [8]. Masing-masing tools memiliki keunggulan dan kekurangan tersendiri dalam hal antarmuka pengguna, algoritma yang tersedia, dan kemampuan dalam menangani data medis. Namun, hingga saat ini, belum banyak penelitian yang secara komprehensif membandingkan kinerja ketiga tools tersebut dalam klasifikasi CKD

Beberapa penelitian sebelumnya telah mengevaluasi kinerja algoritma klasifikasi dalam mendeteksi CKD. Misalnya, penelitian oleh [1] melakukan perbandingan algoritma ML untuk klasifikasi CKD. Hasil perbandingan menujukkan algoritma Decision Tree memiliki akurasi terbaik. Penelitian lain yang melakukan perbandingan algoritma ML pada CKD yaitu [9]. Hasil perbandingan kinerja menunjukkan algoritma Random Fores (RF) memiki akurasi terbaik diikuti Gradient Boosting (GB). Sebagian besar penelitian tersebut fokus pada perbandingan algoritma ML, bukan pada perbandingan tools data mining yang digunakan. Penelitian oleh [10] membandingkan kinerja WEKA dan RapidMiner dalam algoritma klasifikasi, tetapi tidak secara spesifik pada kasus CKD. Sementara itu, penelitian [8] membandingkan WEKA, RapidMiner dan Orange menggunakan Iris, Car Evaluation, and Tic-Tac-Toe end game Dataset namun tidak pada konteks CKD

Dari tinjauan pustaka, terlihat bahwa terdapat kekurangan dalam penelitian yang secara langsung membandingkan kinerja tools data mining WEKA, RapidMiner, dan Orange dalam klasifikasi khususnya CKD. Sebagian besar studi sebelumnya lebih fokus pada perbandingan algoritma atau hanya membandingkan dua tools tanpa mempertimbangkan konteks spesifik CKD. Selain itu, belum ada penelitian yang mengevaluasi aspek seperti kemudahan penggunaan, akurasi klasifikasi, precision, recall, F1-Score, dan waktu pemrosesan secara bersamaan pada ketiga tools tersebut dalam konteks CKD

Penelitian ini menawarkan kontribusi baru dengan melakukan analisis komparatif terhadap tiga tools data mining populer yaitu WEKA, RapidMiner, dan Orange dalam klasifikasi CKD menggunakan dataset UCI.

Keunikan penelitian ini terletak pada evaluasi menyeluruh yang mencakup aspek akurasi klasifikasi, precision, recall dan efisiensi waktu pemrosesan, F1-Score, serta kemudahan penggunaan masing-masing tool. Dengan demikian, penelitian ini memberikan wawasan praktis bagi para profesional medis dan data scientist dalam memilih tool yang paling sesuai untuk deteksi dini CKD

Tujuan dari penelitian ini membandingkan kinerja tools data mining WEKA, RapidMiner, dan Orange dalam klasifikasi CKD berdasarkan akurasi, presisi, recall, dan F1-score. Selain itu, mengevaluasi efisiensi waktu pemrosesan dan kemudahan penggunaan masing-masing tool dalam konteks klasifikasi CKD serta memberikan rekomendasi tool data mining yang paling efektif dan efisien untuk deteksi dini CKD berdasarkan hasil analisis komparatif.***)***

2. TINJAUAN PUSTAKA

2.1. Dataset

Pengumpulan data adalah proses mengumpulkan, mengukur, dan menganalisis informasi dari berbagai sumber [11]. Data yang digunakan berupa Dataset Chronic Kidney Disease (CKD) yang diperoleh dari UCI Machine Learning Repository. Dataset ini berisi 400 sampel pasien, terdiri dari dua class yaitu ("ckd" dan "not-ckd"), 24 atribut seperti Usia, Tekanan darah, Specific Gravity, Albumin, Serum Creatinine, Natrium Sodium, Potassium, Hemoglobin, Volume sel darah merah, Packed Cell Volume, White Blood Cell Count, Red Blood Cell Count, Bilirubin, dan Blood Glucose Random, Blood Urea, Serum Creatinine, Sodium, Potassium, Hemoglobin, Packed cell volume, White blood cell count, Red blood cell count, Hypertension, Diabetes Mellitus, Coronary artery disease, Appetite, Peda edema, dan Anemia[3]. Tabel 1 menampilkan jumlah class dan record dari dataset CKD

Tabel 1. Jumlah class

Ckd	Not-ckd	Total
250	150	400

Tabel 1 menampilkan jumlah *class "ckd"* dan "*not-ckd"* tidak seimbang. Jumlah *class* "ckd" mayoritas dan *class* "not-ckd" minoritas. Hal ini mengakibatkan akurasi klasifikasi algoritma ML tidak optimal karena cenderung pada *class* mayoritas [12]. Selain itu, terdapat fitur-fitur yang digunakan dalam dataset PIDD disajikan dalam Tabel 2

2.2. SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) adalah pengembangan dari metode *oversampling,* dimana cara kerja metode ini adalah dengan membangkitkan sampel baru yang berasal dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang dengan cara sampling ulang sampel kelas minoritas [6], [7]

Table 1. Jumlah class setelah dilakukan balancing class

Method	Ckd	Not-ckd	Total
Non-SMOTE	250	150	400
SMOTE	250	300	550

Tabel 2 menunjukkan bahwa penerapan metode SMOTE meningkatkan jumlah masing-masing class dan record. Selanjutnya kami melakukan pengujian akurasi algoritma ML menggunakan dataset SMOTE

2.3. Cross-validation

Merupakan teknik statistik yang digunakan dalam machine learning dan pemodelan prediktif lainnya untuk menilai kinerja dan kemampuan generalisasi suatu model [6]. Berdasarkan geeksforgeeks.org, pada cross validation, data yang tersedia akan dibagi ke dalam subset, biasanya disebut fold, supaya dapat dilakukan pelatihan dan pengujian model berkali-kali. Teknik ini memberikan estimasi performa model yang lebih akurat pada data yang tidak terlihat [13]. Manfaat penting lainnya dari *cross validation*, yakni membantu data analyst mengatasi masalah *overfitting* atau kondisi saat model terlalu spesifik pada data pelatihan sehingga kurang baik dalam menganalisis data baru [14]. Skema fold cross-validation digunakan untuk mengevaluasi performa model [6]

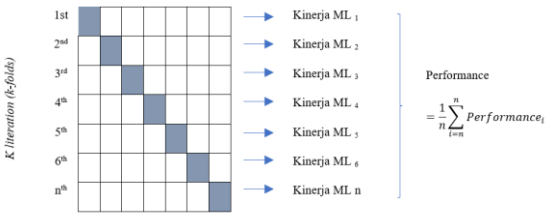


Figure 1. *Prosedur k-fold validation*

2.4. Klasifikasi

Klasifikasi adalah proses pengelompokkan atau penggolongan sesuatu berdasarkan ciri-ciri atau karakteristik yang sama, sehingga dapat diidentifikasi, dibandingkan, dan dipelajari dengan lebih mudah [15][18]. Dalam konteks data, klasifikasi adalah proses memberikan label atau kategori pada data untuk memudahkan analisis dan pemodelan [7]. Tiga algoritma akan diuji yaitu Decision Tree, Naive Bayes, dan SVM. Selanjutnya Proses dilakukan pada dua platform: WEKA, RapidMiner dan Orange

2.5. Evaluasi Pengujian Kinerja

Evaluasi dilakukan untuk mengetahui kinerja algoritma ML. Kami mengevaluasi kinerja algoritma ML menggunakan Confusion Matrix berupa nilai *Accuracy, Precision*, dan *Recall*. Confusion matrix adalah tabel yang digunakan untuk mengevaluasi kinerja suatu model klasifikasi dalam ML [16]. Matriks ini memvisualisasikan perbandingan antara hasil prediksi model dengan nilai aktual, membantu memahami kesalahan dan kelemahan model secara detail [17]. Perbandingan hasil dilakukan untuk melihat keunggulan relatif antar platform dan algoritma. Analisis visual dilakukan dengan grafik perbandingan hasil

Tabel 2. Confusion Matrix

Class	Positif	Negatif
Clas positif	Jumlah true positive (TP)	Jumlah false negative (FN)
Class negatif	Jumlah false positive (FP)	Jumlah true negative (TN)

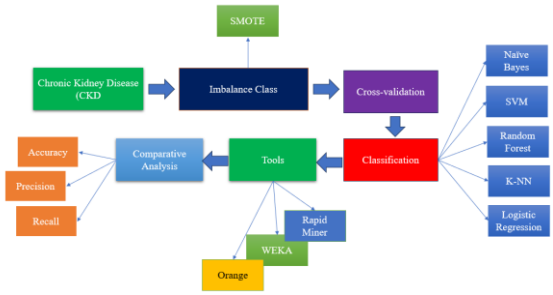
2.6. Tools Data Mining

Kami menggunakan tiga tools data mining open sources yaitu (1) WEKA, (2) RapidMiner, dan (3) Orange. Ketiga tools memiliki algoritma pembelajaran mesin untuk melakukan berbagai tugas data science, seperti eksplorasi, pra-pemrosesan, klasifikasi, regresi, clustering, dan visualisasi data. Versi WEKA yang digunakan 3.9.6, versi RapidMiner Studio 10.1 dan Orange 3.38.1

3. METODE PENELITIAN

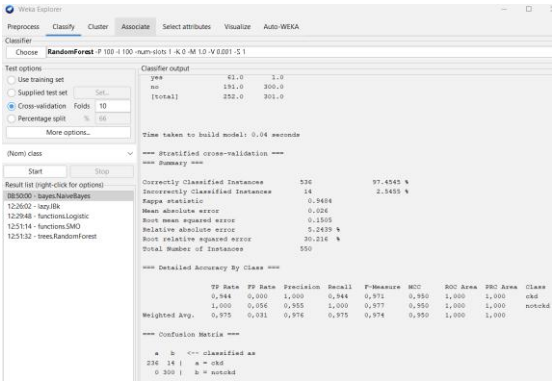
3.1. Kerangka Penelitian

Makalah Tahapan penelitian adalah langkah-langkah yang disusun secara terstruktur dan sistematis untuk mencapai tujuan penelitian. Tahapan ini meliputi berbagai aktivitas, mulai dari pengumpulan data CKD, analisis data, hingga evaluasi penelitian. Penelitian ini menggunakan pendekatan eksperimental kuantitatif dengan beberapa tahapan terstruktur. Tahapan penelitian ditampilkan pada Gambar 1.



Gambar 1. Kerangka Penelitian

Kami melakukan perbandingan kinerja algoritma klasifikasi ML menggunakan tiga tools yaitu WEKA, RapidMiner, dan Orange menggunakan dataset PKG yang telah dilakukan balancing class menggunakan SMOTE.

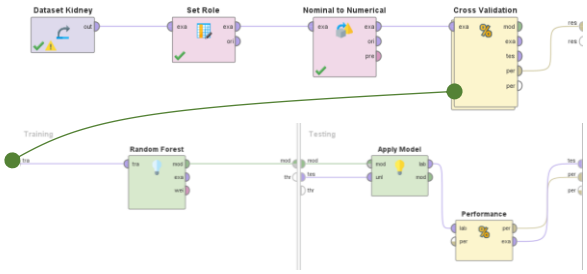


Gambar 2. Pengujian kinerja algoritma ML menggunakan tools WEKA

3.2. Perbandingan Kinerja ML

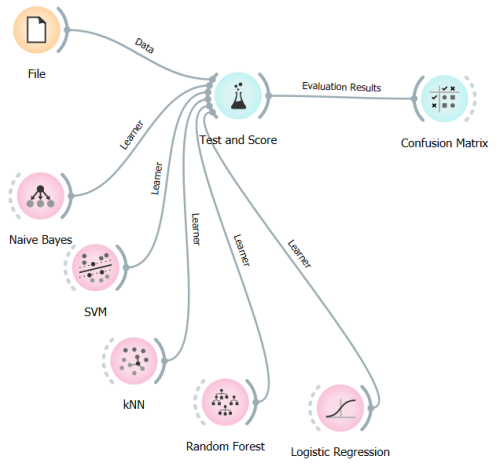
Keunggan tools WEKA yaitu kemudahan penggunaan dan antarmuka grafisnya yang simple. Tools ini cocok untuk pemula maupun ahli di untuk bidang data science. Weka juga mendukung format data yang beragam,

memungkinkan pengguna untuk mengimpor data dari berbagai sumber. Selain WEKA, kami menggunakan tools RapidMiner. Desain model pengujian algoritma ML menggunakan RapidMiner ditampilkan pada Gambar 3.



Gambar 3. Pengujian kinerja ML menggunakan tools RapidMiner

Keunggulan tools RapidMiner yaitu kemudahan akses. RapidMiner memudahkan user dari berbagai latar belakang teknis untuk melakukan analisis data, tanpa perlu skill pemrograman yang mendalam. Tools ini memiliki fleksibilitas yaitu kompatibel dengan berbagai sumber data dan format, memungkinkan integrasi data yang luas. Selain WEKA, kami menggunakan tools RapidMiner. Desain model pengujian algoritma ML menggunakan RapidMiner ditampilkan pada Gambar 4.



Gambar 4. Pengujian kinerja ML menggunakan tools Orange

Tools Orange mempunyai keunggulan dalam hal visualisasi. Orange memiliki tampilan GUI yang menarik dan mudah digunakan oleh para penggali data.

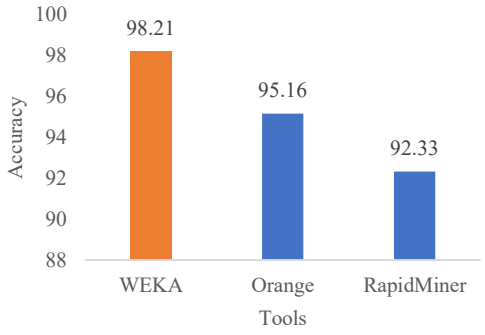
4. HASIL DAN PEMBAHASAN

Kami melakukan perbandingan kinerja ML berupa akurasi menggunakan tiga tools yang ditampilkan dalam Tabel 3.

Tabel 3. Accuracy algoritma ML menggunakan tiga tools

Algoritma	WEKA (%)	RapidMiner (%)	Orange (%)
Naïve Bayes	97.45	70	97.10
Support Vector Machine (SVM)	98.54	98.18	**99.80**
Random Forest	**99.81**	95.64	**99.80**
k-NN	97.09	**99.09**	80.40
Logistic Regression	98.18	98.73	98.70
Rata-rata	**98.21**	**92.33**	**95.16**

Hasil pengujian accuracy menunjukkan bahwa setiap tools menampilkan hasil akurasi algoritma ML yang berbeda. Pengujian menggunakan tools WEKA menghasilkan algoritma Random Forest memiliki nilai akurasi terbaik. Pada tools Rapid Miner algoritma terbaik k-NN, dan pada tools Orange algoritma terbaik yaitu SVM dan k-NN. Kombinasi penggunaan tools dan algoritma ML terbaik yaitu Random Forest dan WEKA. Nilai rata-rata tools berdasarkan akurasi algoritma ditampilkan pada Gambar 5.



Gambar 5. Nilai akurasi rata-rata setiap tools

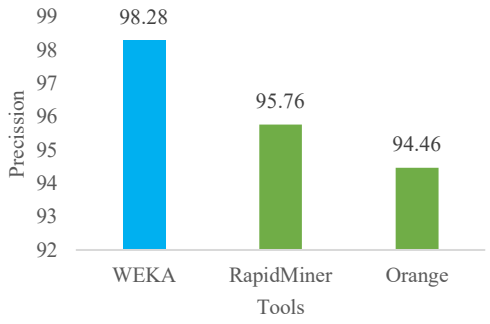
Grafik menampilkan nilai akurasi rata-rata yang berbeda. Tools WEKA memiliki nilai rata-rata akurasi terbaik, disusul dengan tools Orange dan RapidMiner. Selanjutnya kami melakukan

analisis perbandingan nilai precission yang ditampilkan pada Tabel 4.

Tabel 4. Precision algoritma ML menggunakan tiga tools

Algorithm	WEKA (%)	RapidMiner (%)	Orange (%)
Naïve Bayes	97.6	97.2	79.8
Support Vector Machine (SVM)	98.6	**99.8**	98.44
Random Forest	**99.8**	**99.8**	96.29
k-NN	97.2	83.2	**99.14**
Logistic Regression	98.2	**98.8**	98.67

Hasil pengujian precission algoritma ML menunjukkan setiap tools menampilkan hasil yang berbeda. Pengujian menggunakan tools WEKA menghasilkan algoritma terbaik Random Forest. Pada tools Rapid Miner algoritma terbaik SVM, Random Forest dan Logistic Regression, dan pada tools Orange yaitu SVM dan k-NN. Nilai rata-rata tools berdasarkan nilai precission ditampilkan pada Gambar 6.



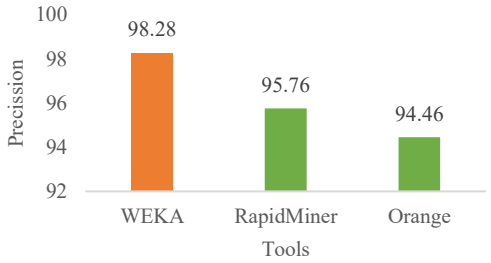
Gambar 6. Nilai rata-rata precission setiap tools

Grafik menampilkan nilai akurasi rata-rata yang berbeda. Tools WEKA memiliki nilai rata-rata akurasi terbaik, disusul dengan tools Orange dan RapidMiner. Selanjutnya kami melakukan analisis perbandingan nilai precission yang ditampilkan pada Tabel 5.

Tabel 5. Recall algoritma ML menggunakan tiga tools

Algoritma	WEKA (%)	RapidMiner (%)	Orange (%)
Naïve Bayes	97.5	72.47	97.1
Support Vector Machine (SVM)	98.5	98	99.8
Random Forest	99.8	95.23	99.8
k-NN	97.1	99.07	80.4
Logistic Regression	98.2	98.82	98.7

Hasil pengujian precission algoritma ML menunjukkan setiap tools menampilkan hasil yang berbeda. Pengujian menggunakan tools WEKA menghasilkan algoritma terbaik Random Forest. Pada tools Rapid Miner algoritma terbaik SVM, Random Forest dan Logistic Regression, dan pada tools Orange yaitu SVM dan k-NN. Nilai rata-rata tools berdasarkan nilai precission ditampilkan pada Gambar 7.



Gambar 7. Nilai rata-rata recall setiap tools

Gambar 7 menampilkan nilai rata-rata yang berbeda. Tools WEKA memiliki nilai rata-rata terbaik, disusul dengan tools RapidMiner, lalu tools Orange. Hasil ini berbeda dengan rata-rata nilai precission

Tabel 6. Perbandingan nilai akurasi, precision, dan recall

Measurement	WEKA (%)	RapidMiner (%)	Orange (%)
Accuracy	98.21	92.33	95.16
Precision	98.28	95.76	94.46
Recall	98.22	92.71	95.16
Average	**98.24**	**93.60**	**94.93**

Berdasarkan rata-rata nilai accuracy, precision, dan recall, menunjukkan tools WEKA memiliki nilai terbaik dari semua sisi disusul Orange dan RapidMiner. Selain itu, WEKA memiliki kemudahan penggunaan dan antarmuka grafisnya yang intuitif, cocok untuk pemula maupun ahli di bidang data science. WEKA juga mendukung format data yang beragam,

memungkinkan pengguna untuk mengimpor data dari berbagai sumber

5. KESIMPULAN

Kami melakukan analisis komparasi terhadap algoritma ML dan tools data mining pada dataset Chronic Kidney Disease (CKD). Algoritma ML yang kami bandingan yaitu Naive Bayes, SVM, k-NN, dan Logistic Regression. Sedangkan untuk tools data mining yang digunakan WEKA, RapidMiner dan Orange.

Berdasarkan hasil analisis terhadap performa algoritma ML pada tools WEKA, RapidMiner, dan Orange, kami menemukan bahwa kinerja setiap algoritma menampilkan hasil yang berbeda antar tools. Pada tools WEKA, algoritma Random Forest menunjukkan performa terbaik dengan akurasi 99.81% dan algoritma k-NN menunjukkan performa terburuk. Pada tools RapidMiner, algoritma k-NN justru menampilkan nilai terbaik dengan nilai akurasi 99.5%, sedangkan Niave Bayes menyusul di bawahnya. Pada tools Orangi algoritma SVM dan Random Forest memiliki performa terbaik dengan nilai 99.8% dan algoritma terburuk k-NN

Secara umum tools WEKA memiliki kinerja yang lebih baik, disusul Orange, dan RapidMiner. Namun, setiap platform memiliki keunggulan masing-masing. WEKA unggul dari segi fleksibilitas eksperimen dan dukungan terhadap tuning parameter yang lebih teknis. RapidMiner lebih ramah pengguna berkat antarmuka grafis yang intuitif dan kemampuan visualisasi proses yang baik, cocok untuk pengguna non-programmer. Sedangkan, Orange mempunyai keunggulan dalam hal visualisasi yang mudah digunakan. Pemilihan platform sebaiknya disesuaikan dengan kebutuhan pengguna. Dengan demikian, ketiga platform memiliki potensi yang besar dalam pengembangan sistem pendukung keputusan berbasis ML untuk diagnosis CKD.

Penelitian lanjutan disarankan untuk menguji kombinasi algoritma ensemble, serta integrasi data klinis real-time untuk mendukung penerapan di lingkungan klinis yang sesungguhnya

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Kesehatan, dan Fakultas Teknologi dan Informatika Kelompok Machine Learning yang telah mendukung terlaksananya hasil penelitian ini. Penulis juga mengapresiasi dukungan dari Universitas Aisyah Pringsewu (UAP) yang telah memfasilitasi pelaksanaan kegiatan ini, serta semua pihak yang telah memberikan kontribusi, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

[1] M. A. Islam, M. Z. H. Majumder, and M. A. Hussein, “Chronic kidney disease prediction based on machine learning algorithms,” *J. Pathol. Inform.*, vol. 14, 2025, doi: 10.1016/j.jpi.2023.100189.

[2] E. K. Bongetti *et al.*, “Classification of chronic kidney disease in older Australian adults by the CKD-EPI 2009 and 2021 equations: secondary analysis of ASPREE study data,” *Med. J. Aust.*, 2024, doi: 10.5694/mja2.52559.

[3] M. S. Arif, A. Mukheimer, and D. Asif, “Enhancing the Early Detection of Chronic Kidney Disease: A Robust Machine Learning Model,” *Big Data Cogn. Comput.*, vol. 7, no. 3, 2023, doi: 10.3390/bdcc7030144.

[4] A. Levin, I. G. Okpechi, F. J. Caskey, C. W. Yang, M. Tonelli, and V. Jha, “Perspectives on early detection of chronic kidney disease: the facts, the questions, and a proposed framework for 2023 and beyond,” *Kidney Int.*, vol. 103, no. 6, pp. 1004–1008, 2023, doi: 10.1016/j.kint.2023.03.009.

[5] C. Karima and W. Anggraeni, “Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalanced Dataset,” *Procedia Comput. Sci.*, vol. 234, pp. 645–653, 2024, doi: https://doi.org/10.1016/j.procs.2024.03.050.

[6] W. Thungrut and N. Wattanapongsakorn, “Diabetes classification with fuzzy genetic algorithm,” *Adv. Intell. Syst. Comput.*, vol. 769, pp. 107–114, 2019, doi: 10.1007/978-3-319-93692-5_11.

[7] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, “Hybrid Prediction Model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, Synthetic Minority Over Sampling Technique (SMOTE), and random forest,” *Appl. Sci.*, vol. 8, no. 8, 2018, doi: 10.3390/app8081325.

[8] A. Abuashour, M. Salem Alzboon, M. Kamel Alqaraleh, and A. Abuashour, “Comparative

Study of Classification Mechanisms of Machine Learning on Multiple Data Mining Tool Kits," *Am. J. Biomed. Sci. Tesearch*, vol. 22, no. 1, pp. 33–43, 2024, doi: 10.34297/AJBSR.2024.22.002913.

[9] M. Rashed-Al-Mahfuz, A. Haque, A. Azad, S. A. Alyami, J. M. W. Quinn, and M. A. Moni, "Clinically Applicable Machine Learning Approaches to Identify Attributes of Chronic Kidney Disease (CKD) for Use in Low-Cost Diagnostic Screening," *IEEE J. Transl. Eng. Heal. Med.*, vol. 9, no. December 2020, pp. 1–11, 2021, doi: 10.1109/JTEHM.2021.3073629.

[10] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.

[11] A. Tejawati, H. S. Pakpahan, and W. Susantini, "Drugs Diagnose Level using Simple Multi-Attribute Rating Technique (SMART)," *Proc. - 2nd East Indones. Conf. Comput. Inf. Technol. Internet Things Ind. EIConCIT 2018*, no. Double L, pp. 357–362, 2018, doi: 10.1109/EIConCIT.2018.8878564.

[12] M. Ambika, G. Raghuraman, and L. SaiRamesh, "Enhanced decision support system to predict and prevent hypertension using computational intelligence techniques," *Soft Comput.*, vol. 24, no. 17, pp. 13293–13304, 2020, doi: 10.1007/s00500-020-04743-9.

[13] H. Sulistiani, A. Syarif, K. Muludi, and Warsito, "Performance evaluation of feature selections on some ML approaches for diagnosing the narcissistic personality disorder," *Bull. Electr. Eng. Informatics*, vol. 13, no. 2, pp. 1383–1391, 2024, doi: 10.11591/eei.v13i2.6717.

[14] X. He *et al.*, "Sample-efficient deep learning for COVID-19 diagnosis based on CT scans," *IEEE Trans. Med. Imaging*, vol. XX, no. Xx, 2020, doi: 10.1101/2020.04.13.20063941.

[15] M. Jasri, A. Wijaya, and R. Sunggara, "Penerapan Data Mining untuk Klasifikasi Penyakit Demam Berdarah Dengue (DBD) Dengan Metode Naïve Bayes (Studi Kasus Puskesmas Taman Krocok)," *SMARTICS J.*, vol. 8, no. 1, pp. 35–42, 2022, [Online]. Available: https://doi.org/10.21067/smartics.v8i1.7375

[16] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalanced data classification," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806–1819, 2017, doi: 10.1109/TKDE.2017.2682249.

[17] I. Düntsch and G. Gediga, "Confusion Matrices and Rough Set Data Analysis," *J. Phys. Conf. Ser.*, vol. 1229, no. 1, 2019, doi: 10.1088/1742-6596/1229/1/012055.

[18] A. Aryanusa, "Evaluasi Kinerja Logistic Regression Dan Multinomial Naïve Bayes Dalam Klasifikasi Sentimen Mobile Banking," J. Inform. dan Tek. Elektro Terap., vol. 13, no. 3S1, pp. 463–474, 2025, [Online]. Available: https://journal.eng.unila.ac.id/index.php/jitet/article/view/7687