

Optimasi Akurasi Algoritma C4.5 Berbasis *Particle Swarm Optimization* dengan Teknik Bagging pada Prediksi Penyakit Ginjal Kronis

Optimization of C4.5 Algorithm Based On Particle Swarm Optimization with Bagging Technique on Prediction of Chronic Kidney Disease

Ita Yulianti¹, Rizal Amegia Saputra², Muhammad Sukrisno Mardiyanto³, Ami Rahmawati⁴

^{1,2,4}Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika

³Fakultas Teknologi Informasi, Institut Teknologi Batam

E-mail: ¹ita.iyi@bsi.ac.id, ²rizal.rga@bsi.ac.id,

³sukrisno@informatika.org, ⁴ami.amv@bsi.ac.id

Abstrak

Penyakit Ginjal Kronis merupakan penyakit dengan tingkat pengeluaran biaya terbesar di dunia. Penyakit tersebut sering kali tidak menunjukkan adanya gejala yang terjadi seperti penyakit pada umumnya. Oleh karena itu, penelitian ini dilakukan dengan tujuan agar dapat mendeteksi dini penyakit tersebut sebelum didiagnosis ke tahap yang lebih serius. Penerapan model algoritma C4.5 individual dan algoritma C4.5 berbasis PSO dengan teknik bagging dilakukan guna mengetahui model mana yang memberikan hasil terbaik dalam mendeteksi penyakit ginjal kronis. Pemilihan kedua model tersebut dipertimbangkan karena algoritma C4.5 merupakan salah satu algoritma terbaik data mining, namun cenderung memiliki kelemahan pada data yang overlap, kelas dan atribut yang banyak. Maka dari itu, teknik optimasi *Particle Swarm Optimization* (PSO) dan teknik bagging juga dipilih sebagai alternatif dalam mengatasi kelemahan yang ada dalam algoritma C4.5. Dari hasil penelitian diperoleh bahwa model algoritma C4.5 berbasis PSO dengan teknik bagging mampu menyeleksi atribut sehingga dapat meningkatkan nilai akurasi yang lebih baik dengan hasil sebesar 99,70% dibandingkan dengan model individual algoritma C4.5 yang menghasilkan nilai akurasi sebesar 91,72% saja.

Kata kunci: Algoritma C4.5, PSO, Bagging, Ginjal, Penyakit Ginjal Kronis

Abstract

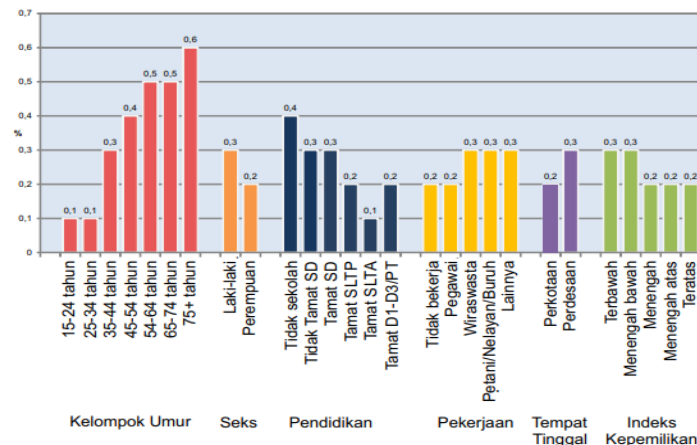
Chronic kidney disease is a disease with the largest rate of expenditure in the world. The disease often does not indicate any symptoms that occur as a disease in general. Therefore, this research was conducted with the aim of being able to detect the disease early before being diagnosed to a more serious stage. The application of individual C4.5 algorithm models and PSO-based C4.5 algorithms with bagging techniques are carried out to determine which models provide the best results in detecting chronic kidney disease. The selection of both models is considered because the C4.5 algorithm is one of the best data mining algorithms, but tends to have weaknesses in overlapping data, classes and many attributes. Therefore, Particle Swarm Optimization (PSO) and bagging techniques are also chosen as alternatives in overcoming the weaknesses in the C4.5 algorithm. From the results of the study it was found that the algorithm model C4.5 based PSO with bagging technique is able to select attributes so so that it could improve the value of accuracy better with a result of 99.70% compared to individual models of C4.5 algorithms that produce an accuracy value of 91.72% only.

Keywords: C4.5 Algorithm, PSO, Bagging, Kidney, Chronic Kidney Disease

1. PENDAHULUAN

Ginjal merupakan salah satu organ yang mempunyai peranan penting dalam mengeluarkan sampah metabolisme dan racun tubuh menjadi bentuk urine [1]. Selain itu, ginjal juga bertugas mengatur cairan tubuh dan keseimbangan elektrolit untuk dapat menjaga stabilitas tubuh[2]. Mengingat akan pentingnya keberadaan ginjal yang sehat didalam tubuh, menyebabkan semua orang harus sadar akan kesehatan ginjal secara menyeluruh. Pasalnya, penyakit ginjal dapat beresiko terhadap nyawa manusia karena biasanya sulit untuk dideteksi [3]. Penderita penyakit tersebut terkadang tidak mengalami gejala spesifik, akan tetapi pada saat kondisi terburuk penderita dapat terancam nyawanya sehingga harus melakukan hemodialisis (cuci darah) secara teratur atau melakukan transplantasi ginjal untuk menggantikan organ ginjalnya yang sudah rusak parah[4].

Penyakit Ginjal Kronis (PGK) atau *Chronic Kidney Diseases* (CKD) merupakan penyakit dengan prevalens dan insidens gagal ginjal yang meningkat dan prognosis yang buruk dengan tingkat pengeluaran biaya terbesar di dunia. Berdasarkan hasil laporan dari *World Health Organization* (WHO) pada tahun 2016 menyatakan bahwa 10 besar kasus di dunia yang menyebabkan kematian di negara berpenghasilan tinggi salah satunya adalah penyakit ginjal. Tidak hanya itu, penyakit ginjal juga dikaitkan sebagai kasus penyakit dengan pengeluaran kesehatan katastropik yang terjadi di negara-negara berpenghasilan rendah dan menengah yang jumlah sekitar 188 juta jiwa. Pada tahun 2013, menurut Kementerian Kesehatan RI sebanyak 2 per 1000 penduduk atau 499.800 penduduk di Indonesia dinyatakan sebagai penyakit Penyakit Gagal Ginjal. Seiring dengan bertambahnya umur seseorang, prevalensi penyakit ginjal kronis (PGK) menjadi kian meningkat. Hal ini dapat dibuktikan berdasarkan dengan hasil penelitian Riskesdas tahun 2013 yang digambarkan pada grafik berikut.



Sumber: [5]

Gambar 1. Prevalensi Penyakit Ginjal Kronis Menurut Karakteristik di Indonesia Tahun 2013

Di Indonesia, perawatan penyakit ginjal menjadi ranking kedua setelah penyakit jantung dengan tingkat pengeluaran biaya terbesar dari BPJS kesehatan. Jika dibandingkan dengan tahun sebelumnya, sebanyak 2,68 triliun rupiah dikeluarkan oleh BPJS pada tahun 2015 untuk pembiayaan pelayanan kesehatan bagi para penderita gagal ginjal, baik untuk keperluan biaya rawat inap ataupun rawat jalan[5]. Oleh Karena itu, untuk dapat mencegah dan menanggulangi penyakit tersebut kedalam tahap yang lebih serius, maka pendeteksian dini penyakit ginjal kronis (PGK) amat sangat diperlukan.

Ada banyak peneliti yang sudah melakukan penelitian untuk memprediksi Penyakit Ginjal Kronis (PGK) atau *Chronic Kidney Diseases* (CKD), seperti penelitian yang dilakukan oleh [6] tentang Komparasi dari K-Nearest Neighbour (KNN) dan Support Vector Mechine (SVM) dimana hasil yang lebih akurat untuk memprediksi penyakit ginjal kronis (PGK) yaitu KNN, penelitian selanjutnya yang dilakukan oleh [7] tentang tentang klasifikasi diagnosis

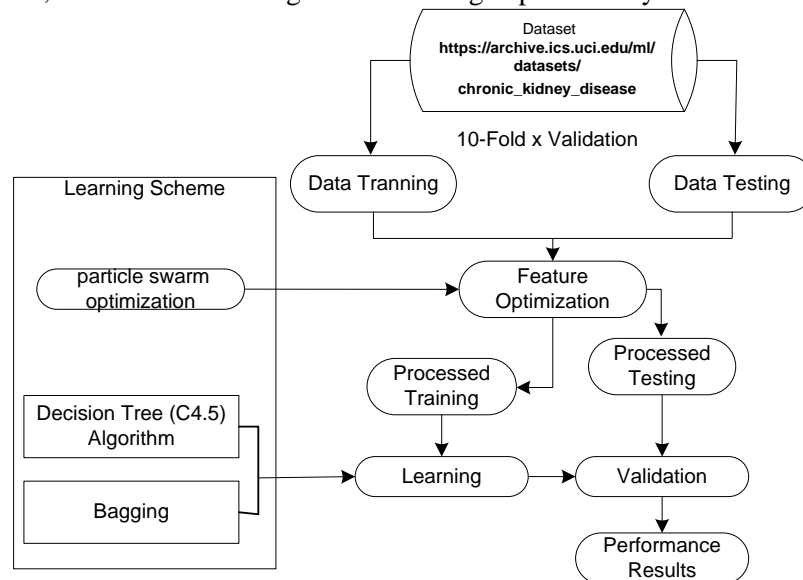
penyakit ginjal kronis (PGK) dengan menggunakan Support Vector Machine (SVM) Neural Networks, dan terakhir, penelitian yang dilakukan oleh [8] tentang rancang bangun aplikasi pendeteksi penyakit ginjal kronis dengan menggunakan algoritma C4.5.

Penelitian penyakit ginjal kronis terbilang cukup banyak, namun belum diketahui akurasi yang paling terbaik. Untuk itu dalam penelitian ini bertujuan menerapkan algoritma C4.5. Algoritma C4.5 memiliki kelebihan dalam mengolah data yang memiliki nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain [9], namun algoritma C4.5 memiliki kelemahan yaitu sering terjadi overlap terutama pada kelas dan kriteria yang digunakan jumlahnya cukup banyak, sehingga membutuhkan waktu dan jumlah memori yang meningkat [10]. Optimasi Particle Swarm Optimization (PSO), salah satu algoritma yang dapat menyeleksi atribut dan mengoptimalkan data yang cukup banyak kelasnya[11], dan teknik bagging akan mengatasi dataset yang tidak seimbang meskipun tidak secara khusus dirancang untuk masalah data yang tidak seimbang[12], untuk itu PSO dan Bagging dapat mengatasi masalah pada algoritma C4.5.

2. METODE PENELITIAN

2.1. Kerangka Penelitian

Pada penelitian ini diperlukan kerangka penelitian untuk menggambarkan penelitian yang akan dilakukan, dibawah ini adalah gambar 2 kerangka penelitian yaitu:



Gambar 2. Model yang Diusulkan

Pada kerangka penelitian diatas menjelaskan tentang dataset yang akan dibagi menjadi 10 bagian dengan menggunakan *10-Fold x validation*, data pada bagian pertama menjadi data testing dan data training, setelah itu data akan di seleksi dengan menerapkan metode seleksi fitur PSO dan akan menghasilkan data training dan testing yang terseleksi, selanjutnya data training akan diuji dengan metode data mining klasifikasi C4.5 dan Teknik bagging untuk menangani masalah dataset yang tidak seimbang kelasnya, lalu di validasi dengan menggunakan *confusion matrix* untuk mengetahui nilai akurasi dan ROC.

2.2 Algoritma C4.5

Algoritma C4.5 yaitu algoritma yang digunakan untuk membentuk sebuah pohon keputusan[13]. Didalam algoritma ini, pohon-pohon keputusan yang dibuat berdasarkan parameter-paramater yang dapat membangun keputusan [14]. Beberapa elemen pembentuk pohon keputusan yaitu *Root*, *Node* dan *Relantioship* [15].

Nilai *gain* tertinggi dari atribut digunakan sebagai penentu atribut yang akan dijadikan akar. Cara menghitung *gain* digunakan persamaan berikut[16]:

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

Keterangan:

S : Himpunan Kasus |S_i| : Jumlah kasus pada partisi ke-i
 A : Atribut |S| : Jumlah kasus dalam S
 n : Jumlah partisi atribut A

Sementara itu, untuk mencari nilai entropi digunakan persamaan beriku[16]t:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Keterangan:

S : Himpunan Kasus
 n : Jumlah partisi S
 p_i : proporsi dari S_i terhadap S

Gain ratio merupakan modifikasi dari *information gain* untuk mengurangi bias atribut yang pada algoritma C4.5[17]. Untuk memilih akar pada nilai *gain ratio* harus mencari nilai *split info* terlebih dahulu. *Gain ratio* dan *split info* dinyatakan dalam persamaan berikut:

$$Split Info (S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (3)$$

Keterangan:

S : Himpunan Kasus
 A : Atribut
 S_i : Jumlah sampel untuk atribut i

$$Gain Ratio (a) = \frac{gain(a)}{split(a)} \quad (4)$$

Keterangan:

a : Atribut
 Gain (a) : Nilai gain pada atribut a
 Split (a) : Nilai split pada atribut a

2.3 Partical Swarm Optimization (PSO)

PSO adalah metode berbasis populasi seperti GA dengan persepsi dasarnya yaitu kooperatif, bukan kompetitif [18]. Ada beberapa cara teknik PSO dalam melakukan pengoptimasian diantaranya penambahan bobot atribut (*attribute weight*) terhadap semua atribut atau variabel yang dipakai, penyeleksian atribut (*attribute selection*) dan *feature selection* [19].

$$V_{i,m} = W * V_{i,w} + C_1 * R * (pbest_{i,m} - X_{i,m}) + C_2 * R * (gbest_m - X_{i,m}) \quad (5)$$

$$X_{id} = X_{i,m} + V_{i,m}$$

Keterangan:

V_{i,m} : kecepatan partikel ke-i pada iterasi ke-i
 w : faktor bobot inersia
 c1, c2 : konstanta akselerasi (learning rate)
 R : bilangan random (0-1)
 X_{i,d} : posisi saat ini dari partikel ke-i pada iterasi ke-i
 Pbest_i : posisi terbaik sebelumnya dari partikel ke-i
 gbest : partikel terbaik diantara semua partikel dalam satu kelompok atau populasi.

2.4. Teknik Bagging

Teknik *Bagging* merupakan teknik yang sukses untuk menangani dataset yang tidak seimbang [12][20]. Teknik *bagging* merupakan salah satu teknik *ensemble* yang digunakan pada klasifikasi untuk memisahkan data training ke dalam beberapa data training baru dengan

random sampling dan membuat model berbasis data training baru[21]. Berikut persamaan teknik bagging untuk klasifikasi[22]:

Input: Data set $D = \{(X_i, Y_i)\}_i^n = 1$ (6)

Algoritma Pembelajaran \mathcal{L}

Jumlah Iterasi T

1) for t=1 to T do

2) $ht = \mathcal{L}(D, Dbs)$

3) end for

4) Output: $H(x) = \max_y \sum_{t=1}^n I(ht(x) = y)$

3. HASIL DAN PEMBAHASAN

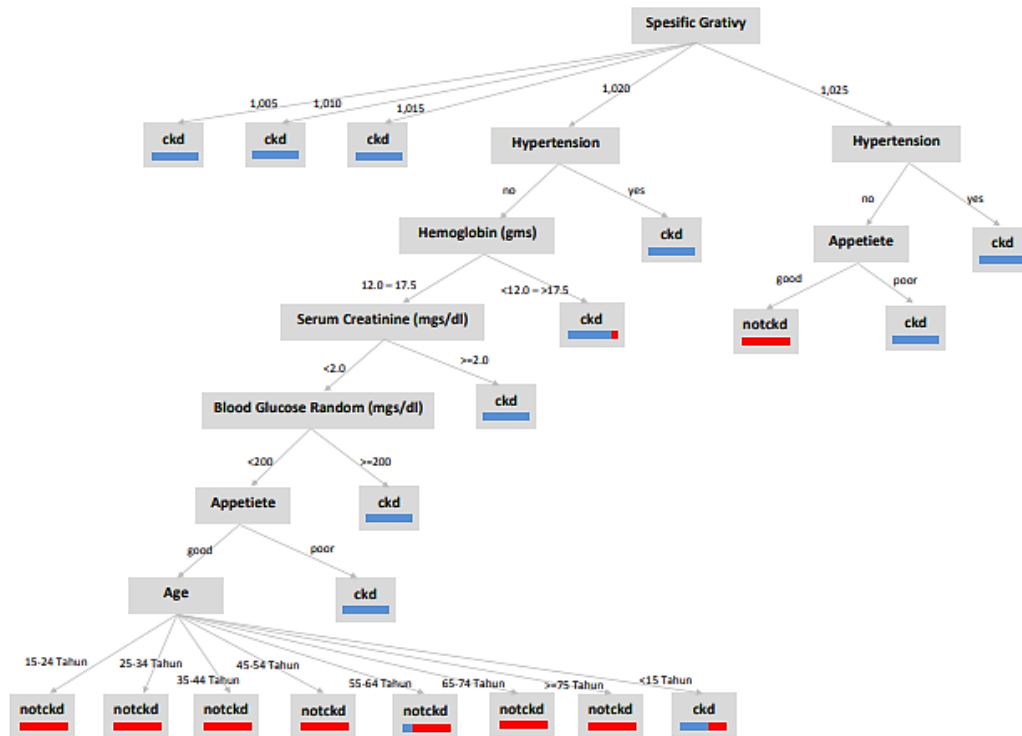
Penelitian ini bertujuan untuk memperoleh hasil prediksi dan akurasi terbaik dengan menganalisis hasil penerapan model individual algoritma C4.5 dengan algoritma C4.5 berbasis bagging yang akan dioptimalkan menggunakan *Particle Swarm Optimization* (PSO) dalam memprediksi penyakit ginjal kronis.

3.1 Dataset

Eksperimen dan pengujian dalam penelitian ini diawali dengan melakukan analisis terhadap dataset yang divalidasi berdasarkan hasil data *preprocessing* yang telah dilakukan menggunakan *tools* RapidMiner dengan tahapan mengisi *value* yang kosong, menemukan outlier dan menghapus data duplikasi.

3.2 Hasil Pemodelan dengan Algoritma C4.5

Langkah pertama yang dilakukan yaitu mencari jumlah *class* yang dinyatakan positif penyakit CKD dan negative NOTCKD. Selanjutnya, data tersebut akan dihitung dengan menggunakan rumus persamaan (1) (2) untuk mendapatkan nilai *entropy* dari masing-masing atribut dan *gain* total supaya dapat terbentuk pohon keputusan berikut:



Gambar 3. Tahapan Validasi Data Menggunakan *Tools* RapidMiner

Dari dataset yang digunakan dalam pemodelan tersebut diperoleh sejumlah 327 tupel dengan rincian data *True Positive (TP)* = 223, *False Negative (FN)* = 10, *False Positive (FP)* = 17 dan *True Negative (TN)* = 77. Berdasarkan rincian data tersebut, maka dapat diperoleh nilai *accuracy*, *sensitivity*, *specifity*, *Positive Predictive Value (PPV)* dan *Negative Predictive Value (NPV)* yang disajikan dalam tabel berikut:

Tabel 1. Nilai *Accuracy*, *Sensitivity*, *Specificity*, *PPV* dan *NPV* C4.5

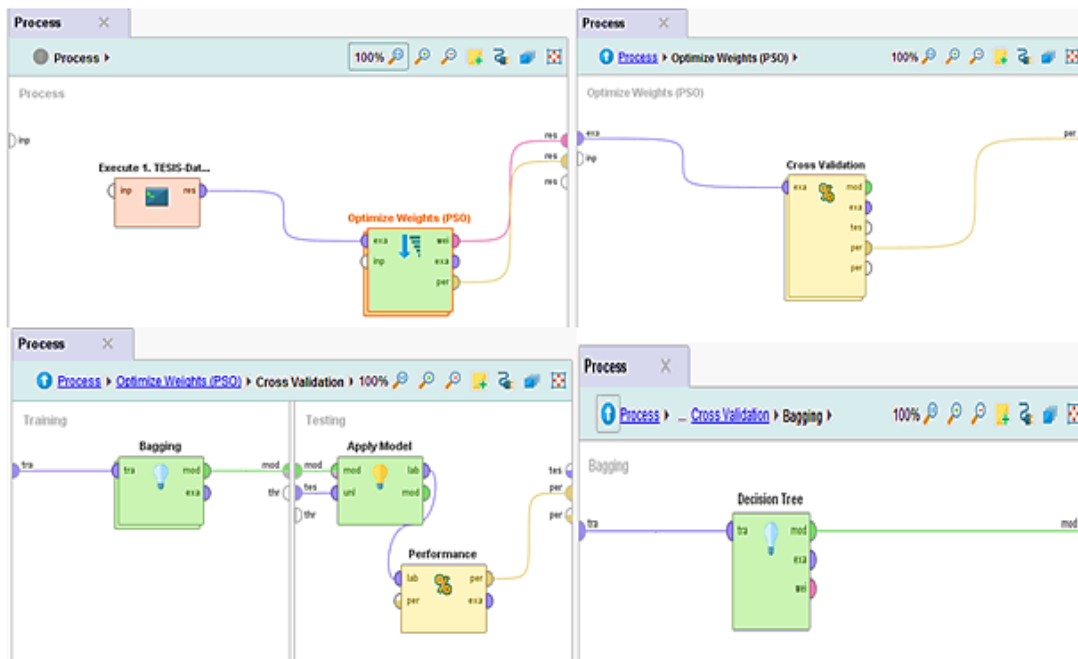
	Nilai
<i>Accuracy</i>	0,917
<i>Sensitivity</i>	0,957
<i>Specificity</i>	0,819
PPV	0,929
NPV	0,885

Selain nilai-nilai diatas, pengujian model ini juga dilakukan dengan melihat grafik ROC yang dinyatakan dalam nilai AUC sebesar 0,931 yang menunjukkan bahwa akurasi pengujian model individual algoritma C4.5 yang termasuk kedalam tingkat *Excellent Classification*.

3.3 Hasil Pemodelan dengan Algoritma C4.5-PSO-Bagging

Indikator yang ada pada seleksi atribut PSO berdasarkan *Optimize Weight (Evolutionary)* yaitu *population*, *max number of generations*, dan *tournament size* yang dapat mempengaruhi hasil akurasi maksimal. Populasi yang digunakan dalam penelitian ini adalah 5 populasi. Dalam penyesuaian indikator PSO, nilai *max number of generations* berisi 30 dan nilai *tournament size* bernilai 0.25. Nilai *c1* dan *c2* masing-masing 0 karena partikel berada dalam putaran pertama. Selain itu, pada proses seleksi ini juga diterapkan teknik bagging yang bertujuan guna menanggulangi data *noise* yang dihasilkan dari proses klasifikasi sehingga akurasi yang diperoleh menjadi optimal.

Desain model C4.5 berbasis PSO dan teknik bagging pada pembobotan atribut dalam RapidMiner dapat dilihat pada gambar berikut:



Gambar 4. Desain Model Validasi Seleksi Atribut C4.5 Menggunakan PSO dan Teknik *Bagging*

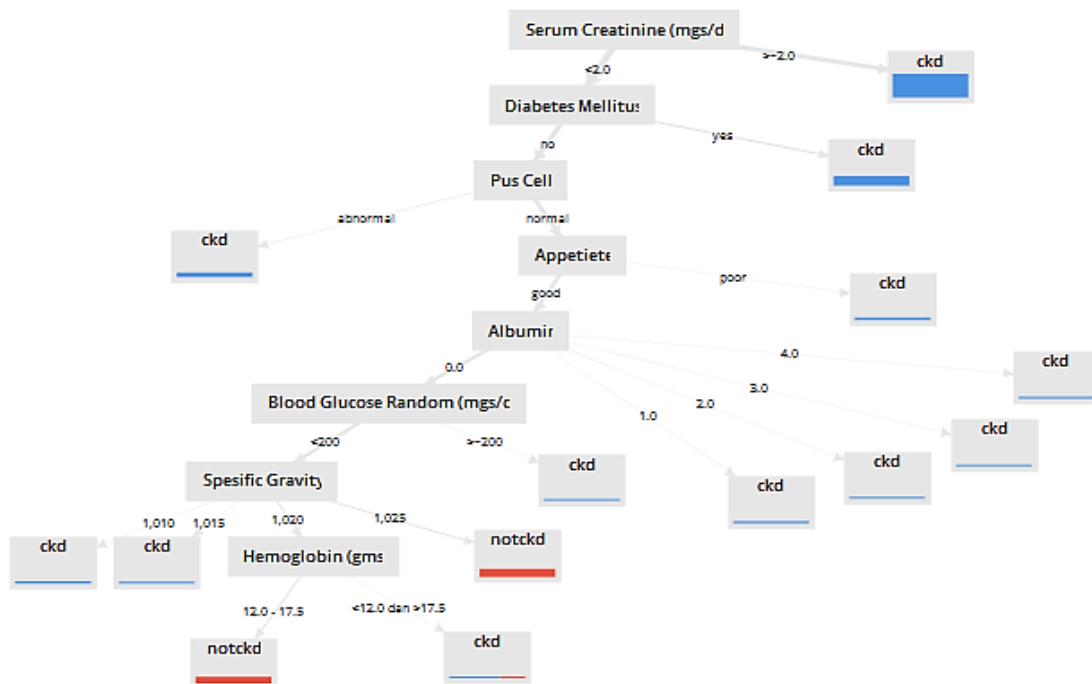
Berdasarkan model proses yang telah berhasil dilakukan, maka diperoleh data seleksi pembobotan atribut sebagai berikut:

Tabel 2. Hasil Seleksi Atribut

Attribute	Weight
Age	0.8459747988213644
Blood Pressure (mmHg)	0.0
Spesific Gravity	1.0
Albumin	1.0
Sugar	0.05568831379927773
Red Blood Cells	1.0
Pus Cell	1.0
Pus Cell Clumps	0.0
Bacteria	1.0
Blood Glucose Random (mgs/dl)	1.0
Blood Urea (mgs/dl)	1.0
Serum Creatinine (mgs/dl)	1.0
Sodium (mEq/L)	0.0
Potassium (mEq/L)	1.0
Hemoglobin (gms)	0.7003268687394637
Packed Cell Volume	0.0
White Blood Cell Count (cells/cumm)	0.0
Red Blood Cell Count (millions/cmm)	1.0
Hypertension	0.0
Diabetes Mellitus	0.4344933068880421
Corony Artery Disease	0.0
Appetiete	1.0
Pedal Edema	0.0
Anemia	0.9653624958912448

Hasil dari tabel diatas, kemudian diterapkan kembali kedalam RapidMiner dengan menggunakan desain model validasi algoritma C4.5. Setiap atribut yang memiliki nilai bobot sama dengan 0 dinyatakan terseleksi sehingga otomatis dihapus dan tidak digunakan pada dataset terbaru.

Berbeda dengan model individual algoritma C4.5, kriteria dari operator *decision tree* pada pemodelan C45-PSO dan Bagging tidak lagi menggunakan *information gain*, melainkan *gain ratio*. Namun, sama halnya dengan pemodelan sebelumnya, pemodelan C4.5 berbasis PSO dan Bagging juga membentuk pohon keputusan yang akan dijadikan sebagai acuan dalam pengujian data baru dan pembuatan aplikasi.



Gambar 5. Pohon Keputusan Klasifikasi Penyakit CKD C4.5 Berbasis PSO dan Bagging

Dari 16 atribut hasil seleksi yang digunakan dalam pembentukan model C4.5 berbasis PSO dan bagging diperoleh sejumlah 327 tupel dengan rincian jumlah *True Positive (TP)* = 239, *False Negative (FN)* = 0, *False Positive (FP)* = 1 dan *True Negative (TN)* = 87. Maka dapat dihitung nilai *accuracy*, *sensivity*, *specifity*, *PPV* dan *NPV* yang disajikan dalam tabel berikut:

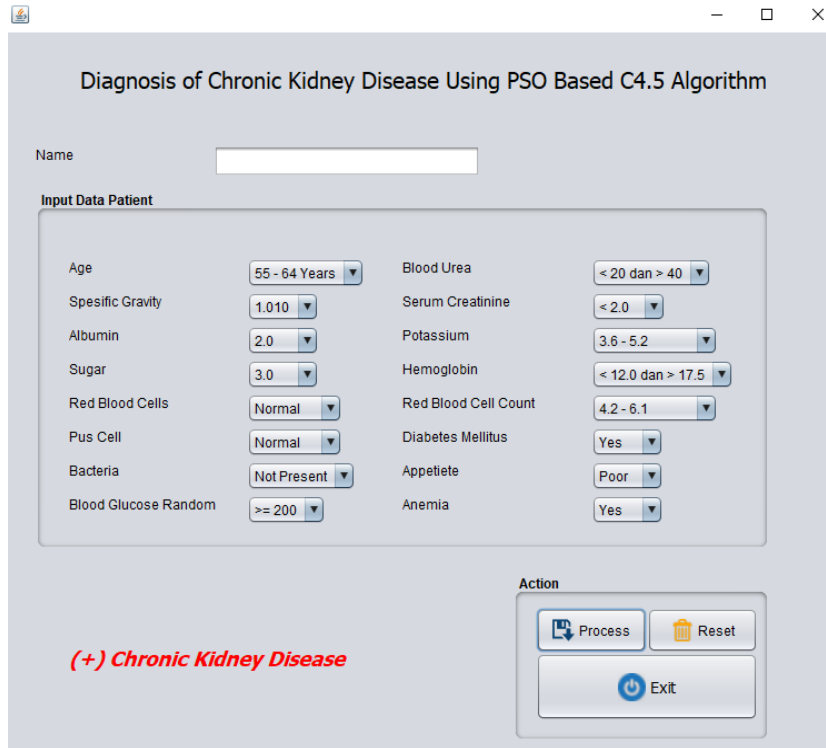
Tabel 3. Nilai *Accuracy*, *Sensitivity*, *Specificity*, *PPV* dan *NPV* C4.5-PSO dan Bagging

	Nilai
Accuracy	0,997
Sensitivity	1
Specificity	0,988
PPV	0,996
NPV	1

Hasil pengujian untuk algoritma C4.5-PSO dan Bagging terhadap grafik ROC dapat diketahui melalui nilai AUC yang dihasilkan sebesar 1,000 yang membuktikan bahwa akurasi pada model optimasi ini memiliki tingkat *Excellent Classification*.

3.3 Pengembangan Aplikasi Model dari Algoritma C4.5 berbasis PSO dan Bagging

Berdasarkan hasil eksperimen yang diperoleh, model dengan akurasi terbaik yang dihasilkan adalah model algoritma C4.5 berbasis PSO dan bagging akan dikembangkan menjadi sebuah aplikasi dengan menggunakan bahasa pemrograman JAVA. Pengujian model aplikasi dilakukan dengan menggunakan salah satu sampel dari data baru pada dataset CKD.



Gambar 6. Pengujian Aplikasi Model Algoritma C4.5 berbasis PSO dan Bagging

3.3 Analisis Evaluasi dan Hasil

Berdasarkan hasil pengujian baik secara *confusion matrix* maupun *ROC curve* terbukti bahwa model algoritma C4.5 berbasis PSO dan bagging memiliki nilai akurasi yang lebih tinggi dibandingkan menggunakan individual algoritma C4.5 saja. Hal ini dapat dilihat perbandingannya dari tingkat akurasi dan nilai AUC yang dihasilkan (Lihat Tabel 4.)

Tabel 4. Perbedaan Hasil Pengujian Algoritma Klasifikasi C4.5 dan C4.5 Berbasis PSO dan Bagging

	Accuracy	AUC
C4.5	91,72%	0,931
C4.5-PSO-Bagging	99,70%	1,000

Selain itu, uji kecermatan GUI berupa matrik galat juga dilakukan untuk menganalisis sejauh mana tingkat prediksi yang dihasilkan dari pengembangan aplikasi yang dapat dilihat pada Tabel berikut:

Tabel 5. Matrik Galat

RapidMiner		Aplikasi		Hasil Diagnosis
a	b	a	B	Keterangan
239	0	239	0	a = CKD

Berdasarkan pengujian tersebut menunjukkan bahwa aplikasi yang dihasilkan sudah sesuai dengan hasil olah RapidMiner. Jika dibandingkan dengan akurasi metode diagnostik dengan penelitian sebelumnya, model dan aplikasi yang dihasilkan pada penelitian ini, memiliki tingkat akurasi yang tertinggi.

4. KESIMPULAN

Dari hasil penelitian yang diperoleh, akurasi yang dihasilkan algoritma C4.5 dalam prediksi penyakit ginjal kronis sebesar 91,72%, dengan AUC bernilai sebesar 0,931 sedangkan untuk akurasi optimasi algoritma C4.5 berbasis PSO dan bagging sebesar 99,70% dengan AUC bernilai 1,000, sehingga didapat selisih peningkatan akurasi sebesar 7,98% dan nilai AUC sebesar 0,069.

Berdasarkan penjelasan tersebut, dapat disimpulkan bahwa penerapan teknik optimasi PSO dan Bagging mampu menyeleksi atribut dan dapat mengatasi kelas yang tidak seimbang pada algoritma C4.5, sehingga menghasilkan tingkat akurasi diagnosis yang lebih baik dibandingkan dengan metode individual algoritma C4.5 saja.

5. SARAN

Berdasarkan proses pengujian dan kesimpulan yang diperoleh, maka terdapat beberapa saran pada penelitian ini diantaranya:

1. Melibatkan sejumlah data yang lebih kompleks dan metode optimasi lain, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi.
2. Melakukan pengembangan perangkat lunak sesuai perkembangan IPTEK supaya dapat digunakan pada gawai terkini.

DAFTAR PUSTAKA

- [1] A. Hamidi and K. Kusnadi, "Sistem Pakar Untuk Mendiagnosa Penyakit Gagal Ginjal," *J. Digit*, vol. 4, no. 2, pp. 149–160, 2014.
- [2] H. Amalia, "Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis," *J. Pilar Nusa Mandiri*, vol. 14, no. 1, pp. 1–6, 2018, doi: <https://doi.org/10.33480/pilar.v14i1.80>.
- [3] F. A. Tarigan, "Sistem Pakar Untuk Mendiagnosa Penyakit Ginjal dengan Metode Backward Chaining," *J. TIMES*, vol. III, no. 2, pp. 25–29, 2014.
- [4] Kementerian Kesehatan RI, *Riset Kesehatan Dasar*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI, 2013.
- [5] Pusat Data dan Informasi Kementerian Kesehatan RI, *Situasi Penyakit Ginjal Kronik di Indonesia*. Jakarta: Pusat Data dan Informasi Kementerian Kesehatan RI, 2017.
- [6] Parul Sinha and Poonam Sinha, "Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM," *Int. J. Eng. Res.*, vol. V4, no. 12, pp. 608–612, 2015, doi: 10.17577/ijertv4is120622.
- [7] B. V. Ravindra, N. Sriraam, and M. Geetha, "Classification of nonchronic and chronic kidney disease using SVM neural networks," *Int. J. Eng. Technol.*, vol. 7, no. 1, pp. 191–194, 2018, doi: 10.14419/ijet.v7i1.3.10669.
- [8] R. Rianto and N. M. S. Iswari, "Rancang Bangun Aplikasi Pendeteksi Penyakit Ginjal Kronis dengan Menggunakan Algoritma C4.5," *J. Ultim.*, vol. 9, no. 1, pp. 10–18, 2017, doi: 10.31937/ti.v9i1.558.
- [9] P. M. Hasugian, "Pengujian algoritma c 4.5 dengan aplikasi weka dalam pembentukan pohon keputusan," *J. Matik Penusa*, vol. 22, no. 1, pp. 7–10, 2018.
- [10] R. S. Rohman, R. A. Saputra, and D. A. Firmansaha, "Komparasi algoritma c4.5 berbasis pso dan ga untuk diagnosa penyakit stroke," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 1, pp. 155–161, 2020.
- [11] R. A. Zuama and I. A. Sobari, "Neural Network Optimization With Particle Swarm Optimization and Bagging Methods on Classification of Single Pap Smear Image Cells," *J. Pilar Nusa Mandiri*, vol. 16, no. 1, pp. 129–134, 2020, doi: 10.33480/pilar.v16i1.1308.
- [12] T. Arifin, "Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap

- Smear,” *J. Inform.*, vol. 4, no. 2, pp. 155–162, 2017, doi: 10.31311/JI.V4I2.2129.
- [13] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [14] D. Nofriansyah, *Konsep Data Mining VS Sistem Pendukung Keputusan*. Yogyakarta: Deepublish, 2014.
- [15] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: CV. Andi Offset, 2009.
- [16] D. T. Larose and C. D. Larose, *Discovering Knowledge In Data: An Introduction To Data Mining-Second Edition (2nd Ed.)*. New Jersey: John Wiley & Sons, Inc, 2014.
- [17] M. Yusa, E. Utami, and E. Luthfi. Taufiq, “Evaluasi Performa Algoritma Klasifikasi Decision Tree Id3 ,” *InfoSys J.*, vol. 4, no. 1, pp. 23–34, 2016.
- [18] E. H. Hasibuan, H. Mawengkang, and S. Efendi, “Optimization Of Feature Weight TheVoting Feature Intervals 5 Algorithm Using Partical Swarm Optimization Algorithm Optimization Of Feature Weight TheVoting Feature Intervals 5 Algorithm Using Partical Swarm Optimization Algorithm.”
- [19] L. A. Utami, “Melalui Komparasi Algoritma Support Vector Machine Dan K-Nearest Neighbor Berbasis Particle Swarm Optimization,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 103–112, 2017.
- [20] I. H. Laradji, M. Alshayeb, and L. Ghouti, “Software defect prediction using ensemble learning on selected features,” *Inf. Softw. Technol.*, vol. 58, pp. 388–402, 2015, doi: 10.1016/j.infsof.2014.07.005.
- [21] R. S. Wahono and N. Suryana, “Combining particle swarm optimization based feature selection and bagging technique for software defect prediction,” *Int. J. Softw. Eng. its Appl.*, vol. 7, no. 5, pp. 153–166, 2013, doi: 10.14257/ijseia.2013.7.5.16.
- [22] X. Y. Liu and Z. H. Zhou, *Ensemble methods for class imbalance learning*. 2013.