

PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION

Toni Arifin¹, Daniel Ariesta²

^{1,2}Universitas BSI Bandung

¹toni.tfn@bsi.ac.id

²arsdaniel.id@gmail.com

Abstrak - Penyakit ginjal kronis (PGK) merupakan masalah kesehatan masyarakat global dengan prevalensi dan insiden gagal ginjal yang meningkat, prognosis yang buruk dan biaya yang tinggi. nilai prevalensi di seluruh Indonesia untuk penyakit gagal ginjal memiliki nilai rata - rata berkisar kurang lebih 0.2 persen. Langkah pertama dalam pengelolaan penyakit ginjal adalah penetapan diagnosis yang tepat. Maka dibutuhkan sebuah metode untuk memprediksi penyakit ginjal kronis. *Naïve Bayes* memiliki beberapa kelebihan, yaitu cepat dalam perhitungan, algoritma yang sederhana dan berakurasi tinggi. *Naïve Bayes Classifier* lebih tepat diterapkan pada data yang besar dan dapat menangani data yang tidak lengkap (missing value) serta kuat terhadap atribut yang tidak relevan dan noise pada data. Untuk meningkatkan akurasi maka digunakan *Particle Swarm Optimization* untuk pembobotan atribut. Dari hasil penelitian *Naïve Bayes Classification* berbasis *Particle Swarm Optimization* memiliki akurasi confusion matrix sebesar 98,75% dan AUC sebesar 99%. sedangkan *Naïve Bayes* memiliki akurasi confusion matrix 97.00% dan AUC sebesar 99.8%.

Kata kunci : Prediksi penyakit ginjal, Naive bayes, *Particle Swarm Optimization*.

Abstract - *Chronic kidney disease (CKD) is a global public health problem with an increased prevalence and incidence of kidney failure, poor prognosis and high costs. prevalence values throughout Indonesia for kidney failure have an average value of around 0.2 percent. The first step in managing kidney disease is determining the right diagnosis. Then we need a method to predict chronic kidney disease. Naïve Bayes has several advantages, which are fast in calculations, simple and high accuracy algorithms. The Naïve Bayes Classifier is more appropriately applied to large data and can handle incomplete data (missing value) and is strong against irrelevant attributes and noise in the data. To improve accuracy, Particle Swarm Optimization is used for weighting attributes. From the results of the Naïve Bayes Classification research based on Particle Swarm Optimization, the accuracy of confusion matrix is 98.75% and AUC 99% . while Naive Bayes has an accuracy of 97.00% confusion matrix and AUC 99.8%.*

Keywords : Prediction of kidney disease, Naive bayes, *Particle Swarm Optimization*.

1. PENDAHULUAN

Ginjal adalah salah satu organ yang sangat penting yang berfungsi untuk menjaga komposisi darah dengan mencegah menumpuknya limbah atau kotoran dan mengatur keseimbangan cairan di dalam tubuh, menjaga tingkatan elektrolit seperti sodium, potasium dan fosfat tetap stabil, serta memproduksi hormon dan enzim yang membantu tubuh dalam mengendalikan tekanan darah, membuat sel darah merah dan menjaga tulang tetap kuat dan sehat (Kemenkes, 2017).

Penyakit ginjal kronis (PGK) adalah masalah kesehatan masyarakat dunia dengan prevalensi dan insiden gagal ginjal yang sering selalu meningkat setiap tahun. Prevalensi PGK selalu meningkat seiring meningkatnya jumlah penduduk usia lanjut dan penyakit diabetes melitus serta tekanan darah tinggi. Sekitar 1 dari 10 orang populasi dunia mengalami PGK pada stadium tertentu (Kemenkes, 2017). Nilai prevalensi di seluruh Indonesia untuk penyakit gagal ginjal memiliki nilai rata - rata berkisar kurang lebih 0.2 persen (Foundation, 2017). PGK adalah penyebab kematian peringkat ke-27 di dunia pada tahun 1990 dan naik menjadi urutan ke-18 pada tahun 2010. Sedangkan di Indonesia, perawatan

penyakit ginjal merupakan ranking kedua dengan pembiayaan terbesar dari BPJS kesehatan setelah penyakit jantung (Kemenkes, 2017). Dengan banyaknya jumlah penderita penyakit ginjal kronis dan perlunya diagnosis yang tepat maka perlu dilakukan penelitian untuk mencari metode yang memberikan nilai akurasi tertinggi.

Algoritma klasifikasi *Naïve Bayes* cocok untuk diterapkan pada data dengan jumlah yang banyak dan mampu menangani data yang kosong (*missing value*) dan dapat menangani atribut yang tidak sama dan gangguan yang ada pada data. Tetapi, metode klasifikasi dengan *Naïve Bayes* memiliki kekurangan dimana probabilitasnya tidak akan bisa menghitung seberapa besar akurasi sebuah klasifikasi. Algoritma klasifikasi *Naïve Bayes* memiliki kekurangan pada pemilihan atribut yang ada di dalam data sehingga dapat mempengaruhi hasil akhir yang berupa nilai akurasi (Amalia, 2018). *Naïve Bayes Classifier* perlu dioptimalkan dengan cara memberikan nilai bobot pada setiap atribut dengan menggunakan algoritma *particle swarm optimization* (PSO) yang bertujuan untuk meningkatkan akurasi dan membantu ahli dalam identifikasi penyakit ginjal kronis.

2. KAJIAN PUSTAKA

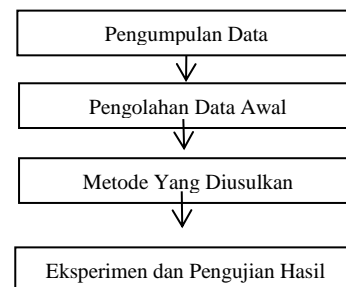
Penelitian sebelumnya (Widiastuti, Stefanus, & Catur, 2010) melakukan penelitian dengan menggunakan algoritma *Naïve Bayes Classifier* berbasis *Particle Swarm Optimization* (PSO) untuk deteksi penyakit jantung. Pada percobaan awal menghasilkan nilai akurasi untuk metode *Naïve Bayes Classifier* sebesar 82.14% dengan nilai area under cover (AUC) 0.686. Pada eksperimen kedua dengan menggunakan algoritma *Naïve Bayes Classifier* berbasis *Particle Swarm Optimization* (PSO) hasil akurasi meningkat menjadi 92.86% dengan nilai AUC 0.839. Penelitian selanjutnya (Arifin, 2017) melakukan optimasi *Naïve Bayes Classifier* berbasis *Particle Swarm Optimization* (PSO). dari hasil penelitiannya didapatkan hasil akurasi 95.70% untuk klasifikasi dengan dua kelas yaitu kelas normal dan abnormal. Sedangkan untuk Klasifikasi kelas 1 sampai dengan kelas 7 didapatkan akurasi 62.67%. Dapat disimpulkan bahwa algoritma *Naïve Bayes Classifier* berbasis *Particle Swarm Optimization* (PSO) dapat meningkatkan akurasi penelitian dan cocok digunakan untuk dataset yang mempunyai dua kelas. Dibawah ini beberapa penelitian terdahulu yang terkait dengan penelitian ini

- Penelitian terdahulu melakukan penelitian tentang penyakit jantung dengan menggunakan metode *Naive Bayes* berbasis *Particle Swarm Optimization* hasil dari penelitian ini didapatkan hasil akurasi sebesar 82.14% untuk *Naive Bayes* dan 92.86 % untuk *Naive Bayes* berbasis PSO (Widiastuti et al., 2010).
- Penelitian terdahulu melakukan penelitian penyakit ginjal kronis menggunakan metode *Back Propagation Neural Network*, *Radial Basis Function* dan *Random Forest*. Pada penelitian itu *Back Propagation Neural Network* menghasilkan akurasi sebesar 80.4, *Radial Basis Function* sebesar 78.6%, dan *Random Forest* sebesar 85.3% (Alasaf et al., 2018).
- Penelitian terdahulu melakukan penelitian tentang Sel Tunggal Pap Smear dengan mengoptimalkan metode *Naive Bayes Classifier* (NBC) dan *Particle Swarm Optimization* (PSO). Penelitian tersebut didapatkan hasil akurasi untuk Kelas 1-7 = 62.67% dan Dua kelas = 95.70% (Hidayatulloh, Herliana, & Arifin, 2016).
- Penelitian terdahulu melakukan penelitian untuk Komparasi Metode Klasifikasi Data Mining untuk Prediksi Penyakit Ginjal. Metode yang di komparasi adalah C.45 dan *Support Vector Machine*. Hasil dari penelitian itu adalah C.45 sebesar 97.25% sedangkan SVM menghasilkan akurasi 95% (Handayani, 2017).
- Penelitian terdahulu melakukan penelitian dengan penggabungan metode *Particle Swarm Optimization* dan Teknik Bagging. Hasil yang didapatkan pada penelitian ini adalah 95.05% untuk sel 2 kelas dan 64.24% untuk sel 7 kelas (Arifin, 2017).
- Penelitian terdahulu melakukan penelitian tentang data iris dengan menggunakan metode *Naive Bayes* dan *Particle Swarm Optimization* untuk pembobotan atribut. Pada penelitian ini dihasilkan nilai fitness sebesar 97.3 (Muhamad et al., 2017).
- Penelitian terdahulu melakukan penelitian Penyakit Gagal Ginjal menggunakan Gabungan *Genetic Algorithm* dan *Fuzzy K-Nearest Neighbor*. Pada penelitian ini dihasilkan akurasi sebesar 98% (Ubaidillah, Ratnawati, & Dewi, 2018).
- Penelitian terdahulu melakukan penelitian tentang penyakit ginjal kronis dengan membandingkan Metode Data Mining Svm Dan Nn. Pada penelitian ini SVM memiliki akurasi tertinggi dengan akurasi 95.16%, sedangkan NN menghasilkan akurasi sebesar 93.37% (Amalia, 2018).
- Penelitian terdahulu melakukan penelitian tentang penyakit ginjal kronis dengan menggunakan Menggunakan Metode *Extreme Learning Machine* (ELM). Pada penelitian ini ELM menghasilkan akurasi sebesar 96.7% (Fadilla, Adikara, & Perdana, 2018).

Berdasarkan permasalahan diatas akan dilakukan penelitian penyakit gagal ginjal kronis dengan menggunakan metode *Naive Bayes Classification* berbasis *Particle Swarm Optimization*.

3. METODE PENELITIAN

Dalam penelitian ini dilakukan beberapa tahapan penelitian, adapun tahapan penelitian yang dilakukan ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Penelitian

A. Pengumpulan data

Pada tahap ini ditentukan data yang akan diproses. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan

semua data kedalam dataset, termasuk variabel yang diperlukan dalam proses. Dalam penelitian ini, digunakan dataset yang didapatkan dari website The UCI *Machine Learning Repository*. Dataset dalam penelitian ini adalah 400 orang pasien yang terdiri dari 24 atribut dan 1 label. Hal ini ditunjukkan pada Tabel 1.

Tabel 1 Penjelasan Atribut

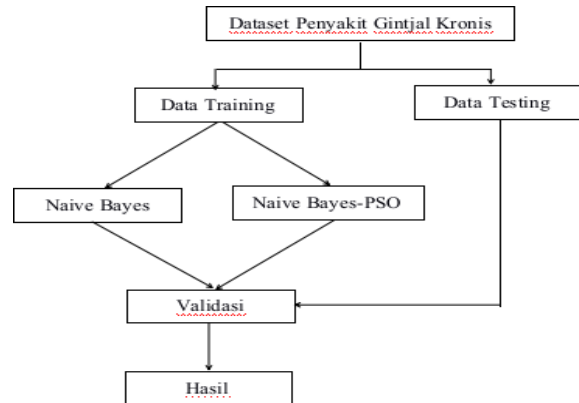
No	Singkatan	Data Input
1	Age	Age
2	Bp	Blood Pressure
3	Sg	Specific Gravity
4	Al	Albumin
5	Su	Sugar
6	Rbc	Red Blood Cells
7	Pc	Pus Cell
8	Pcc	Pus Cell Clumps
9	Ba	Bacteria
10	Bgr	Blood Glucose Random
11	Bu	Blood Urea
12	Sc	Serum Creatinine
13	Sod	Sodium
14	Pot	Potassium
15	Hemo	Hemoglobin
16	Pcv	Packed Cell Volume
17	Wc	White Blood Cell Count
18	Rc	Red Blood Cell Count
19	Htn	Hypertension
20	Dm	Diabetes Mellitus
21	Cad	Coronary Artery Disease
22	Appet	Appetite
23	Pe	Pedal Edema
24	Ane	Anemia

B. Pengolahan Data Awal

Pada tahapan pengolahan data awal, dilakukan tiga tahapan yaitu: (1) Validasi data, (2) Integrasi data dan (3) Pengubahan data, ketiga tahap diatas diimplementasikan untuk memperoleh hasil *dataset* yang mewakili dan diharapkan dapat digunakan untuk mendapatkan informasi yang berharga dari *dataset* yang ada.

C. Metode Yang Diusulkan

Tahap ini data akan dianalisis dan dikelompokkan berdasarkan dari variabel yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu diterapkan model yang sesuai dengan jenis data yang akan digunakan. Jenis data pada penelitian ini dibagi menjadi *training* data dan *testing* data juga diperlukan untuk pembuatan model.



Gambar 2 Kerangka Penelitian

Pada kerangka penelitian Gambar 2 dataset penyakit ginjal kronis akan dibagi menjadi dua bagian yaitu data *training* dan data *testing*. data set akan dibagi menjadi 10 bagian dengan menggunakan 10-Fold x validation. Kemudian data *training* akan diuji dengan metode *Naive Bayes Classifier* dan *Naive Bayes Classifier* berbasis *Particle Swarm Optimization*.

D. Eksperimen dan Hasil Pengujian

Setelah mendapatkan metode yang akan digunakan maka dalam tahap ini model yang diusulkan akan diuji untuk melihat hasil berupa *rule* yang akan digunakan dalam proses pengambilan keputusan. Pengujian akurasi menggunakan *tools* RapidMiner. Kemudian dari hasil pengujian akan didapatkan *Confusion Matrix* dan AUC.

E. Algoritma

• Naive Bayes

Naive Bayes Classifier adalah algoritma *Bayesian Learning* yang paling cepat dan sederhana. Hal ini berasal dari teorema Bayes, yang menghasilkan *classifier* berdasarkan peluang. Metode Bayes merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi (Amalia, 2018). Metode *Naive Bayes* ditemukan oleh Thomas Bayes di abad ke-18. Dalam teorema Bayes, probabilitas atau peluang bersyarat dinyatakan sebagai (Hidayatulloh et al., 2016).

$$p(h|x) = \frac{p(x|h).p(h)}{P(X)} \quad (1)$$

Dimana X adalah bukti, H adalah hipotesis, $P(H|X)$ adalah probabilitas bahwa hipotesis H benar untuk bukti X atau dengan kata lain $P(H|X)$ merupakan probabilitas posterior H dengan syarat X, $P(H|X)$ adalah Probabilitas bahwa X benar untuk hipotesis H atau probabilitas X dengan syarat H, $P(H)$ adalah probabilitas prior bukti X.

- *Particle Swarm Optimization*

Particle Swarm Optimization atau sering disebut juga sebagai PSO, merupakan sebuah metode optimasi yang terinspirasi dari gerak dan perilaku hewan seperti ikan dan burung dalam berperilaku seperti mencari mangsa, yang pertama kali dikemukakan oleh James Kennedy dan Russell C. Eberhart pada tahun 1995. PSO terdiri dari partikel-partikel yang mencari posisi terbaik, yang merupakan posisi terbaik untuk masalah optimasi. Kelebihan metode optimasi *Particle Swarm Optimization* mempunyai konsep sederhana, mudah diimplementasikan, dan efisien dalam perhitungan jika dibandingkan dengan algoritma matematika dan teknik optimisasi heuristik lain (Arifin, 2017).

$$v_{i,m} = w.v_{i,m} + c_1 * R * (pbest_{i,m} - x_{i,m}) + c_2 * R * (gbest_m - x_{i,m})$$

$$x_{i,d} = x_{i,m} + v_{i,mu} \quad (2)$$

4. HASIL DAN ANALISIS

Dari dataset yang telah diolah kemudian dilakukan perhitungan dengan menggunakan RapidMiner untuk mengetahui hasil akurasi metode yang digunakan pada penelitian ini.

Dataset yang telah diolah kemudian dibagi menjadi data *training* dan data *testing*. yaitu 300 data untuk data *training* dan 100 untuk data *testing*. Kemudian dilakukan perhitungan dengan metode *Naive Bayes Classification* dan didapatkan hasil confusion matrix seperti ditunjukkan pada Tabel 2.

Tabel 2 *Confusion Matrix Naive Bayes*

	true ckd	true notckd	Class precision
Pred,ckd	239	1	99.58%
Pred,notckd	11	149	93.12%
Class recall	95.60%	99.33%	

Untuk pembobotan atribut digunakan algoritma *Particle Swarm Optimization*. Maka untuk mengetahui akurasi dari *Naive Bayes* berbasis *Particle Swarm Optimization* dilakukan perhitungan dengan dengan Rapidminer dan didapatkan hasil Confusion Matrix yang ditunjukkan pada Tabel 3.

Tabel 3 *Confusion Matrix Naive Bayes - PSO*

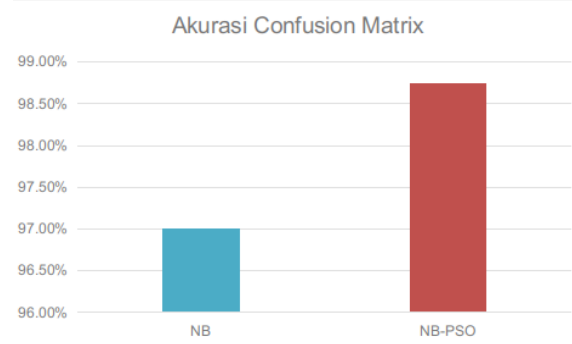
	true ckd	true notckd	Class precision
Pred,ckd	239	1	99.58%
Pred,notckd	11	149	93.12%
Class recall	95.60%	99.33%	

Table 4 adalah tabel perbandingan akurasi dari *Naive Bayes* dan *Naive Bayes* Berbasis PSO:

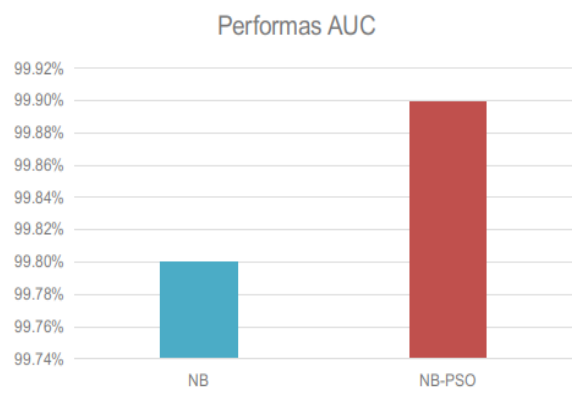
Tabel 4 Komparasi Hasil NB dan NB-PSO

Perbandingan	NB	NB-PSO
Akurasi <i>Confusion Matrix</i>	97.00%	98.75%
Performas AUC	99.80%	99.90%

Untuk mempermudah dalam memahami perbedaan akurasi dari hasil perhitungan Confusion Matrix dan Performa AUC metode *Naive Bayes* dan *Naive Bayes-PSO* maka perlu dibuatkan grafik. Berikut adalah grafik perbandingan Counfusion Matrix antara *Naive Bayes* dan *Naive Bayes-PSO*:



Gambar 3 Grafik Perbedaan Akurasi antara NB dengan NB-PSO



Gambar 4 Grafik Perbedaan Performa AUC antara NB dengan NB-PSO

5. KESIMPULAN DAN SARAN

Pada penelitian ini dilakukan permodelan menggunakan *Naive Bayes Classification* dan *Naive Bayes Classification* berbasis *Particle Swarm Optimization* dengan menggunakan dataset penyakit ginjal kronis yang diambil dari UCI Repository.

Hasil perhitungan dengan *Naive Bayes Classification* yang diterapkan pada dataset Penyakit ginjal kronis menghasilkan nilai akurasi confusion matrix sebesar 97.00% dan AUC sebesar 99.8%. setelah ditambahkan *Particle Swarm Optimization* untuk pembobotan atribut pada dataset penyakit ginjal kronis hasil akurasi dari *Naive Bayes Classification* meningkat menjadi 98,75% dan AUC sebesar 99%. sehingga dapat disimpulkan bahwa

Particle Swarm Optimization dapat meningkatkan akurasi dari algoritma *Naive Bayes* pada dataset penyakit ginjal kronis.

Metode *Naive Bayes Classification* berbasis *Particle Swarm Optimization* dengan hasil akurasi 99.8%. dapat digunakan oleh ahli Patologi dalam membuat program untuk memprediksi penyakit ginjal kronis.

Berdasarkan hasil dari penelitian yang telah dilakukan, maka peneliti mengajukan saran untuk Melakukan eksperimen dengan menggunakan metode lainnya seperti *Decision Tree*, *Super Vector Machine* (SVM). atau algoritma optimasi seperti *genetic algorithm*, *ant colony optimization*.

6. DAFTAR PUSTAKA

- Alassaf, R. A., Alsulaim, K. A., Alroomi, N. Y., Alsharif, N. S., Aljubeir, M. F., Olatunji, O., ... Alturayef, N. S. (2018). Preemptive Diagnosis of Chronic Kidney Disease Using Machine Learning Techniques. *2018 International Conference on Innovations in Information Technology (IIT)*, 99–104. <https://doi.org/10.1109/INNOVATIONS.2018.8606040>
- Amalia, H. (2018). Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis. *Jurnal Pilar Nusa Mandiri*, 14(1), 1–6. Retrieved from <http://ejournal.nusamandiri.ac.id/ejurnal/index.php/pilar/article/view/788/pdf>
- Arifin, T. (2017). Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear. *Jurnal Informatika*, 4(2), 155–162.
- Fadilla, I., Adikara, P. P., & Perdana, R. S. (2018). Klasifikasi Penyakit Chronic Kidney Disease (CKD) Dengan Menggunakan Metode Extreme Learning Machine (ELM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10), 3397–3405.
- Foundation, N. K. (2017). About Chronic Kidney Disease. Retrieved from www.kidney.org
- Handayani, S. (2017). Komparasi Metode Klasifikasi Data Mining untuk Prediksi Penyakit Ginjal. *Jurnal Sistem Informasi STMIK Antar Bangsa*, 6, 34–41.
- Hidayatulloh, T., Herliana, A., & Arifin, T. (2016). Klasifikasi Sel Tunggal Pap Smear Berdasarkan Analisis Fitur Berbasis *Naive Bayes Classifier* Dan *Particle Swarm Optimization*. *Swbumi*, IV(September), 186–193.
- Kemenkes, P. D. dan I. (2017). *No Title*.
- Muhamad, H., Prasajo, C. A., Sugianto, N. A., Surtiningsih, L., Cholissodin, I., Ilmu, F., ... Optimization, P. S. (2017). Optimasi *Naive Bayes Classifier* dengan Menggunakan *Particle Swarm Optimization* pada Data Iris. *Teknologi Informasi Dan Pendidikan*, 4(3), 180–184.
- Ubaidillah, M., Ratnawati, D. E., & Dewi, C. (2018). Deteksi Dini Penyakit Gagal Ginjal menggunakan Gabungan Genetic Algorithm dan Fuzzy K-Nearest Neighbor (GAFKNN), 2(2), 577–583.
- Widiastuti, nur aeni, Stefanus, S., & Catur, S. (2010). Optical imaging of the native brain. *Nature Methods*, 7(1), 34–34. <https://doi.org/10.1038/nmeth.f.284>