

## Optimasi Random Forest Untuk Diagnosis Penyakit Ginjal Kronik Dengan Menggunakan Particle Swarm Optimization

Sheva NaufalRifqi

Prodi S1 Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati Raya, Pd. Labu, Kec. Cilandak, Kota Depok, Daerah Khusus Ibukota Jakarta 12450

shevanaufalrifqi28@gmail.com

**Abstrak.** Penyakit Ginjal Kronik (PGK) merupakan salah satu penyakit yang penderitanya terus meningkat dalam skala global. Penyakit ginjal kronik ini menyebabkan kemampuan cairan elektrolit pada tubuh tidak dapat mempertahankan metabolisme dalam tubuh dengan baik. Penyebab penyakit ini terus meningkat dikarenakan sifatnya yang sangat progresif dan irreversible. Untuk mengatasi hal ini, diperlukannya metode cepat dan akurat dalam mendiagnosa penyakit ginjal kronik ini, agar penanganan terhadap penderitanya bisa cepat ditangani. Salah satu metode yang tepat dalam memprediksi diagnosa ginjal kronik ini adalah dengan membangun model klasifikasi dengan menggunakan berbagai macam algoritma, salah satunya dengan menggunakan random forest. Tetapi dalam penerapannya diperlukan metode lain untuk mengoptimasi algoritma tersebut agar menjadi lebih akurat. Untuk mengatasi hal tersebut digunakan algoritma Particle Swarn Optimization untuk dilakukan seleksi fitur terhadap data yang memiliki fitur yang banyak. Hasil evaluasi dalam pengujian performa dengan menggunakan Particle Swarn Optimization dalam pengklasifikasian CKD dan Non CKD menghasilkan akurasi sebesar 99.167%.

**Kata Kunci:** ginjal kronik, *Random Forest*, *Particle Swarm Optimization*

### 1 Pendahuluan

Penyakit Ginjal Kronik (PGK) adalah salah satu dari masalah kesehatan berskala global yang hingga saat ini masih terus meningkat kasusnya. Hal tersebut terjadi seiringnya dengan angka mortalitas PGK yang mengalami peningkatan selama 10 tahun terakhir yang menjadikannya sebagai penyebab kematian tertinggi di dunia bersamaan dengan penyakit diabetes [8]. Tercatat lebih dari 2 juta penduduk mendapatkan perawatan akibat dari penyakit ginjal kronis ini, dan hanya 10% dari 2 juta penduduk tersebut yang benar mendapatkan perawatan dengan baik. Bahkan di Amerika Serikat, 87,3% individu mengalami dialisis peritoneal dan 2.5% dari jumlah tersebut menerima transplantasi ginjal. Untuk itu dibutuhkan suatu metode yang cepat dalam mendiagnosa penyakit ginjal kronis agar penyakit cepat ditangani [1].

Algoritma random forest cocok untuk diterapkan pada data dengan jumlah yang besar, termasuk dalam masalah diagnosis penyakit yang dalam penelitian ini adalah penyakit ginjal kronis. Penelitian terkait yang telah dilakukan dengan menggabungkan random forest-SVM menghasilkan akurasi sebesar 83.4% [10]. Particle Swarn Optimization merupakan suatu algoritma yang berbasis teknik populasi stochastic yang terinspirasi dari perilaku kawanan burung atau kawanan ikan yang berguna untuk masalah optimasi. Kawanan burung ini seperti partikel, dan setiap partikel memanfaatkan memori individu dan pengetahuan yang diperoleh dengan kawanan secara keseluruhan untuk menemukan solusi terbaik (Raghuwanshi, 2018).

Terdapat banyak metode lain yang telah dilakukan dalam studi kasus penyakit ginjal kronis ini. Misalnya pada penelitian yang telah dilakukan oleh Warid Yunus dengan menggunakan algoritma k-NN yang berbasis Particle Swarn Optimization didapatkan akurasi hingga 97.25%, lebih unggul dibandingkan dengan menggunakan metode k-NN saja, yang hanya mendapatkan akurasi sebesar 78.25% [13]. Pada tahun 2020 juga dilakukan penelitian yang sama dengan kasus yang sama dengan metode hybrid yang menggabungkan bootstrap dengan teknik imputasi k-NN yang menggunakan metode C4-5 oleh Ahmad Ilham. Bootstrap yang digunakan oleh Ahmad untuk mengatasi data yang hilang (missing) dan prediksinya sendiri menggunakan algoritma C4-5. Hasil akurasi yang didapatkan sebesar 97.89% dibandingkan tanpa Bootstrap dan imputasi yang hanya sebesar 87.25% [6]. Terdapat

juga penelitian lain yang menerapkan PSO pada algoritma naïve bayes yang dilakukan oleh Toni Arifin dan Daniel Ariesta. Hasil penelitian yang dilakukan memperlihatkan bahwa dengan dilakukannya pembobotan pada atribut menggunakan PSO didapatkan akurasi sebesar 98.75%, hasil meningkat sebesar 1.75% [4].

Oleh sebab itu diperlukannya pengkajian lebih lanjut untuk mencari metode yang lebih akurat yang memberikan akurasi diagnosis lebih tinggi. Dalam hal ini, penelitian akan menggunakan PSO dan Random Forest karena diperlukannya metode hybrid yang lain dalam optimasi algoritma untuk diagnosis penyakit ginjal kronis ini [10].

## 2

## Tinjauan

## Pustaka

### 2.1 Penyakit Ginjal Kronis

Penyakit ginjal merupakan kelainan yang mempengaruhi fungsi ginjal. Penyakit ginjal kronik (PGK) adalah suatu penyakit yang banyak dialami oleh masyarakat dunia dengan prevalensi serta insiden gagal ginjal yang semakin tinggi, biaya pengobatan tinggi dan prognosis yang buruk. Prevalensi PGK semakin tinggi seiring bertambahnya jumlah penduduk usia lanjut yang disertai dengan hipertensi dan bahkan hingga penyakit diabetes melitus, kurang lebih 1 dari 10 populasi dunia mengalami PGK di stadium tertentu [7]. PGK dapat menyebabkan diabetes, tekanan darah tinggi, serta gaya hidup yang tidak sehat sehingga menyebabkan meningkatnya jumlah pasien dengan PGK [5]. Pada awalnya, PGK ini tidak menampakkan gejala serta tanda – tanda komplikasi, namun nyatanya, penyakit ini bisa tetap berjalan tanpa sadar hingga menjadi gagal ginjal. Penyakit ginjal dapat ditanggulangi dan dicegah serta memungkinkan untuk mendapatkan pengobatan terapi yang lebih efektif apabila bisa diketahui lebih awal [7].

Dalam mendiagnosis PGK di tahap awal, dapat dilakukan dengan teknik pembelajaran mesin (machine learning). Maka dari itu, krusial bagi semua orang untuk tetap memperhatikan kesehatan ginjal. Diperlukan suatu sistem yang menunjang keputusan dalam mendiagnosis terhadap pengambilan keputusan penyakit ginjal kronik [2].

### 2.2 Praproses Data

Praproses data merupakan salah satu metode penting dalam machine learning. Hal tersebut disebabkan oleh dataset yang awalnya bersifat mentah, kurang diolah, dan kurang berkualitas karena data – data tersebut masih banyak yang kurang, tidak lengkap, tidak konsisten, dan banyak noise-nya [12]. Maksud dari tidak lengkap ini adalah, data tersebut memiliki data yang masih hilang ataupun kosong, atribut atau fitur yang tidak lengkap bahkan tidak sesuai. Selain itu, arti dari data tersebut memiliki banyak noise adalah, ada tersebut masih memuat data – data yang outlier (data jauh berbeda dengan data yang lain). Sedangkan, data tidak konsisten dimaknai bahwa dalam satu kolom data tersebut terdapat data yang tidak sama dalam pengkodean atau penamaan. Dalam proses praproses data, sangat dipengaruhi oleh tekniknya karena akan memengaruhi kinerja proses selanjutnya. Dalam proses ini, data akan dapat dikurangi kesalahan – kesalahan yang ada sehingga dapat menjadi data yang memiliki hasil akurasi dan analisis yang lebih baik. Adapun beberapa cara yang biasa dilakukan ke dalam praproses data yaitu pembersihan data, transformasi data, reduksi data, dan integrasi data.

### 2.3 Algoritma Random Forest

Algoritma Random Forest (RF) merupakan suatu metode yang dikembangkan dari Classification dan Regression Tree (CART). Pengembangan yang dilakukan dengan menerapkan random feature selection dan teknik bagging (bootstrap aggregating). *Random forest* merupakan classifier yang terdiri atas kumpulan pohon klasifikasi  $\{h(x, S^b), b = 1, \dots, B\}$  dengan  $\{S^b\}$  yang merupakan vektor acak yang tidak saling berkaitan dan terdistribusi secara identik pada setiap pohon yang memberikan *vote* untuk kelas paling tinggi pada input  $x$  [11].

Diberikan *ensemble classifier*  $h_1(x), h_2(x), \dots, h_b$  dengan data *training* yang diambil secara random berdasarkan distribusi vektor acak  $X, Y$ , didefinisikan pada Persamaan 1.

$$mg(X, Y) = av_b I(h_b(X) = Y) - \max_{j \neq Y} av_b I(h_b(X) = j) \dots \dots \dots (1)$$

dengan  $I(\cdot)$  merupakan fungsi indikator dan  $av_b$  merupakan hasil rata – rata dari fungsi indikator, dengan  $h_b(X) = Y$  adalah hasil dari prediksi  $Y$  dan  $h_b(X) = j$  adalah hasil dari prediksi  $j$ . Margin yang ditentukan digunakan sebagai ukuran seberapa besar nilai rata – rata pada vote  $X, Y$  untuk kelas yang tepat agar dapat melebihi rata – rata vote kelas lainnya. Karena, semakin besar jarak margin maka semakin akurat nilainya. Setelah itu, *generalization error* diberikan Persamaan 2.

$$PE^* = P_{X,Y}(mg(X,Y) < 0) \dots\dots\dots (2)$$

dengan  $PE^*$  merupakan *generalization error* dan  $P_{X,Y}$  sebagai indikasi untuk probabilitas yang melebihi ruang  $X, Y$ .

Sedangkan, dalam RF,  $h_b(X) = h(X, Y^b)$ . Untuk pohon dengan jumlah yang banyak, terdapat *Tree Structurer* dan *Strong Law of Large Numbers*. Semakin tingginya jumlah pohon, maka hampir dalam semua barisan  $S^1, \dots$  akan menyebabkan nilai  $PE^*$  menjadi *konvergen* ke Persamaan 3.

$$P_{X,Y}(P_S(h(X, S^b) = Y) - \max_{j \neq Y} P_S(h(X, S^b) = j) < 0) \dots\dots\dots (3)$$

## 2.4 Algoritma Particle Swarm Optimization

Pada tahun 1955, terdapat penelitian yang dilakukan oleh James Kennedy dan Russel C. Eberhart dengan melakukan observasi terhadap perilaku hewan, terutama terhadap gerak ikan dan burung dalam mencari mangsanya, hasil observasi tersebut menghasilkan sebuah algoritma optimasi yang disebut dengan *Particle Swarm Optimization* (PSO). PSO terdiri atas partikel yang berusaha menemukan posisi terbaik dalam memecahkan masalah optimasi [3]. Penerapan *Particle Swarm Optimization* dalam penelitian dengan menggunakan *library niapy*, ditawarkan solusi untuk merepresentasikan sebuah subset fitur dengan rumus:

$$x = [x_1, x_2, \dots, x_d]; x_i \in [0, 1] \dots\dots\dots (4)$$

$x$  merupakan jumlah fitur pada set yang akan diolah untuk pencarian skor terbaik partikel pada fitur. Dalam pencarian tersebut, dibutuhkan *threshold* sebagai batas/persyaratan bahwa fitur dengan nilai skor partikel  $> 0.5$  akan terpilih sebagai fitur yang terseleksi, apabila  $< 0.5$  maka dicari kembali posisi  $x_i$  fitur lainnya seperti pada Persamaan 5.

$$x_i = \begin{cases} 1, & \text{if } x_i > 0.5 \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots (5)$$

Lalu dioptimalkan dengan fungsi untuk mencari skor dari jumlah fitur yang telah dipilih nantinya, maka dari itu dilakukan minimalisasikan fungsi dengan Persamaan 6.

$$f(x) = \alpha \times (1 - P) + (1 - \alpha) \times \frac{N_{selected}}{N_{features}} \dots\dots\dots (6)$$

$\alpha$  sebagai parameter yang memutuskan pembagian antara kinerja klasifikasi P (sebagai akurasi model klasifikasi) dan jumlah fitur yang dipilih antara jumlah semua fitur.

## 2.5 Confusion Matrix

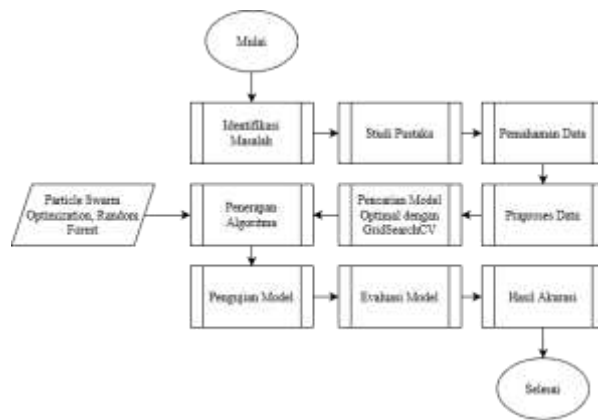
Pengukuran kinerja sistem klasifikasi sangat diperlukan sebagai gambaran atau tolak ukur sudah seberapa baik sistem proses klasifikasi data [11]. Pada penelitian ini, model random forest yang dihasilkan akan diukur lagi performanya dengan cara kuantitatif melalui pengukuran akurasi dari pemodelan yang telah dihasilkan. Confusion matrix pada penelitian ini diperoleh nilai akurasi melalui rumus dengan Persamaan 7.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (7)$$

## 3 Metodologi

Untuk dapat mencapai tujuan penelitian, peneliti akan menguraikan tahapan dan metode pada Gambar 1 yang akan dipakai dalam penyusunan penelitian.

Dalam penyusunan serta penulisan pada pengkajian jurnal ini dibutuhkan informasi dan data yang mendukung kebenaran proses penelitian yang dilakukan. Maka penjelasan tahapan dan metode yang digunakan sesuai penelitian agar penulisan jurnal ini dapat tersusun rapi dan baik. Dimana alur penelitian dijelaskan dalam Gambar 1.



Gambar 1. Alur Penelitian

**Identifikasi Masalah.** Tahap ini dilakukan untuk pencarian permasalahan sebagai dasar pokok pembahasan penelitian yaitu bagaimana penerapan algoritma *particle swarm optimization* sebagai seleksi fitur pada pemodelan *random forest*

**Studi Pustaka.** Penelitian ini disusun dengan melakukan studi pustaka dengan menghimpun data dan informasi yang berhubungan dengan prediksi dari penyakit ginjal kronis, algoritma *Particle Swarn Optimization*, algoritma *Random Forest* serta alat pendukung yang digunakan dalam proses pengolahan data yang diteliti seperti *GitHub*, *UCI Machine Learning Repository*, dan perangkat lunak untuk membantu pemodelan data. Adapun sumber pustaka dalam penelitian ini dapat dilihat pada bagian daftar pustaka.

**Pemahaman Data.** Data yang dipakai pada penelitian ini adalah data yang didapat dari *UCI Machine Learning Repository* dengan nama data *Chronic Kidney Disease* dataset. Dataset yang ada memiliki 25 atribut.

**Praproses Data.** Setelah data dipahami, data yang ada dilakukan pembersihan data untuk menghilangkan noise data (missing) dan normalisasi data agar data yang diinginkan menjadi data yang lebih optimal untuk diproses pada tahap selanjutnya.

**Pencarian Model Terbaik Menggunakan GridSearchCV.** metode *GridSearchCV* ini akan mencari kombinasi yang memungkinkan terhadap nilai parameter yang ada, dan nantinya akan dievaluasi untuk ditemukannya kombinasi parameter terbaik. Pada penelitian yang dilakukan penulis, parameter yang akan dicari model kombinasi terbaiknya diantaranya, *n\_estimators*, *criterion*, *min\_samples\_split*, dan *max\_features*.

**Penerapan Algoritma.** Setelah dilakukannya praproses data, data akan diseleksi dengan menerapkan algoritma optimasi yang penulis pilih, yaitu Algoritma *particle swarn optimization*. *Particle Swarn Optimization (PSO)* adalah algoritma yang sanggup mengoptimalkan data dan menyeleksi atribut yang memiliki banyak kelas. Data yang akan diseleksi nantinya terbagi menjadi dua, yaitu data training dan data testing yang selanjutnya akan diuji dengan menggunakan *random forest*.

**Pengujian Model.** Pengujian dilakukan setelah selesainya tahap seleksi fitur yang dilakukan dengan membaginya menjadi dua data, yaitu data latih dan data uji dengan menggunakan *Random Forest Classifier* yang masing – masing dibagi menjadi 70% data latih dan 30% data uji.

**Evaluasi Model.** Pada tahap ini dilakukan evaluasi akurasi dengan menggunakan *confusion matrix* agar dapat mengetahui nilai akurasi. Perhitungan akurasi dengan *confusion matrix*.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots \dots \dots (7)$$

dengan,

1. TP : Positif Benar, adalah jumlah data dengan kasus saat Non

CKD dideteksi Non CKD

2. TN : Negatif Benar, adalah jumlah data dengan kasus saat CKD dideteksi CKD
3. FP : Positif Salah, adalah jumlah data dengan kasus saat Non CKD dideteksi CKD
4. FN : Negatif Salah, adalah jumlah data dengan kasus saat CKD dideteksi Non CKD

#### 4 Hasil dan Pembahasan

Data chronic kidney disease yang didapatkan dari repository UCI Machine Learning punya beberapa data yang hilang (*missing values*), data hilang tersebut perlu proses agar memudahkan penelitian di tahapan selanjutnya. Data yang hilang ini kosong, dalam artian memang tidak ada nilainya, data yang hilang tersebut diisi menjadi nilai median dan dilakukan mapping untuk data yang bernilai 0 dan 1. Tabel 1 merupakan tabel data setelah yang telah dilakukan praproses dan tidak ada lagi data yang hilang.

**Tabel 1.** Atribut Data Setelah Pengolahan

Kolom	Jumlah Non-Null	Dtype
Age	400 non-null	float64
Bp	400 non-null	float64
Sg	400 non-null	float64
Al	400 non-null	float64
Su	400 non-null	float64
Rbc	400 non-null	float64
pc	400 non-null	float64
pcc	400 non-null	float64
Ba	400 non-null	float64
Bgr	400 non-null	float64
Bu	400 non-null	float64
Sc	400 non-null	float64
Sod	400 non-null	float64
Pot	400 non-null	float64
Hemo	400 non-null	float64
Pcv	400 non-null	float64
Wc	400 non-null	float64
Rc	400 non-null	float64
Htn	400 non-null	float64
Dm	400 non-null	float64
Cad	400 non-null	float64
Appet	400 non-null	float64
Pe	400 non-null	float64
Ane	400 non-null	float64
Classification	400 non-null	int64

Data praproses yang telah diolah akan dilakukan pemodelan dengan menggunakan *Random Forest* untuk diimplementasikan dengan membagi data *training* dan data *testing* sebanyak 70% dan 30% untuk masing – masing data berjumlah 280 untuk data *training* dan 120 untuk *data testing*. Dari pembagian data tersebut, akan dilakukan pencarian model random forest yang paling optimal dengan menguji parameter *random forest* menggunakan *grid search*. Parameter yang akan diuji yaitu, *n\_estimators*, *criterion*, *min\_samples\_split*, dan *max\_features*. Masing – masing parameter tersebut memiliki value yang akan diuji, dimana *n\_estimators* akan menguji jumlah pohon sebesar 10, 20, 30, 40, 50, 60, 70, 80, 90, 100. Lalu, *criterion* yang memiliki nilai *gini*, dan *entropy*. *Min\_samples\_split* yang memiliki nilai 1, 2, 3, 4, 5 dan *max\_features* yang berisi *auto*, *sqrt*, dan *log2*. Hasil dari pengujian yang dilakukan akan menemukan parameter terbaik untuk digunakan dan model terbaik setelah dilakukan *gridsearch*.

**Tabel 2.** Hasil Pengujian Parameter

criterion	max_features	min_sample_split	n_estimators
gini	auto	2	50

Dari pengujian yang telah dilakukan, pada Tabel 4.3 ditemukan parameter terbaik dengan *criterion = gini*, *max\_features = auto*, *min\_sample\_split = 2*, dan *n\_estimators = 50*.

Setelah ditemukannya model *random forest* yang memiliki parameter optimal, model *random forest* tersebut akan dievaluasi menggunakan *confusion matrix*. Hasil prediksi yang dilakukan pada program akan menghasilkan sebuah *classification report* yang berisi informasi mengenai *precision*, *recall*, *f1\_score* dan akurasinya.

	precision	recall	f1-score	support
0	1.00	0.96	0.98	45
1	0.97	1.00	0.99	75
accuracy			0.98	120
macro avg	0.99	0.98	0.98	120
weighted avg	0.98	0.98	0.98	120

**Gambar 2.** Hasil *Classification Report*

Pada Gambar 2 menunjukkan bahwa hasil klasifikasi yang dilakukan random forest dengan model terbaiknya, mendapatkan akurasi sebesar 98% dengan precision 97%, recall 100%, dan f1-score sebesar 99%.

Hasil *Classification Report* yang telah dilakukan akan dibandingkan dengan *Random Forest* yang menggunakan *Particle Swarm Optimization* sebagai seleksi fitur. Pada tahap ini, dilakukan pembangunan model data training dan data testing yang dibagi menjadi 70% dan 30% masing – masing. Model tersebut akan dilakukan iterasi sebanyak 100 kali terhadap fitur – fitur yang akan diseleksi. Untuk mencari fitur terbaik dan teroptimal, Dari hasil iterasi tersebut, ditemukan akurasi pada tiap fiturnya seperti pada Tabel 3. Hasil tersebut memperlihatkan bahwa dari 24 fitur yang dilakukan seleksi fitur, ditemukan sebanyak 12 fitur.

**Tabel 3.** Fitur yang Terseleksi

Fitur Terpilih
Age
Sg
Rbc
Ba
Sc
Sod
Pot
Hemo
Htn
Dm
Appet
Pe

Hasil dari pemodelan dengan menggunakan *Random Forest* terhadap fitur yang telah terseleksi oleh *Particle Swarn Optimization*, Tabel 4 menunjukkan bahwa hasil akurasi terhadap fitur yang telah terseleksi lebih baik dibandingkan dengan hasil akurasi yang menggunakan semua fitur ketika diuji.

**Tabel 4.** Perbandingan Akurasi model yang Terseleksi dengan Semua Fitur

Subset Accuracy	All Features Accuracy
0.991667	0.983

Dari hasil pemodelan algoritma yang telah dijalankan, didapatkan akurasi optimasi sebesar 0.867% lebih besar dibandingkan dengan pemodelan *random forest* tanpa penggunaan seleksi fitur menggunakan *Particle Swarm Optimization*.

## 5 Kesimpulan

Kesimpulan yang didapatkan dari penelitian ini yaitu:

1. Particle Swarn Optimization (PSO) berhasil mengoptimasi akurasi pemodelan yang dilakukan dengan metode Random Forest, dengan hasil akurasi 0.867% lebih baik dibandingkan dengan pemodelan yang hanya dilakukan dengan Random Forest saja yang menghasilkan akurasi sebesar 98%.
2. Implementasi seleksi fitur menghasilkan akurasi yang berbeda dibandingkan dengan menggunakan keseluruhan 24 fitur untuk diolah dengan perbedaan akurasi sebesar 0.867%
3. Fitur – fitur yang terseleksi berpengaruh terhadap kesuksesan peningkatan akurasi, fitur tersebut diantaranya: Age, Spesific Gravity, Red Blood Cells, Bacteria, Serum Creatinine, Sodium, Potassium, Hemoglobin, Hypertension, Diabetes Mellitus, Appetite, Pedal Edema.

## 6 Daftar Pustaka

- [1] Aldy Fauzi, Zahrah Maulidia Septimar, & H.A.Y.G Wibisono. (2021). Literature Review : Pengaruh Mengunyah Xylitol Terhadap Ph Saliva Dan Rasa Haus Pada Pasien Penyakit Ginjal Kronis Yang Menjalani Hemodialisa Di Rumah Sakit. *Jurnal Kesehatan*, 10(1), 51–73. <https://doi.org/10.37048/kesehatan.v10i1.336>
- [2] Amalia, H. (2018). Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis. *Maret*, 14(1), 1. [www.bsi.ac.id](http://www.bsi.ac.id)
- [3] Arifin, T. (2017). Implementasi Algoritma PSO Dan Teknik Bagging Untuk Klasifikasi Sel Pap Smear. *Jurnal Informatika*, 4(2), 155–162.
- [4] Arifin, T., & Ariesta, D. (2019). *PREDIKSI PENYAKIT GINJAL KRONIS MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER BERBASIS PARTICLE SWARM OPTIMIZATION*. 13(1), 26–30.
- [5] Chen, Z., Zhang, X., & Zhang, Z. (2016). Clinical risk assessment of patients with chronic kidney disease by using clinical data and multivariate models. *International Urology and Nephrology*, 48(12), 2069–2075. <https://doi.org/10.1007/s11255-016-1346-4>
- [6] Ilham, A. (2020). Hybrid Metode Bootstrap Dan Teknik Imputasi Pada Metode C4-5 Untuk Prediksi Penyakit Ginjal Kronis. *Statistika*, 8(1), 43–51. <https://jurnal.unimus.ac.id/index.php/statistik/article/view/5765>
- [7] Kemenkes RI. (2017). Info datin ginjal. *Situasi Penyakit Ginjal Kronik*, 1–10.
- [8] Neuen, B. L., Chadban, S. J., Demaio, A. R., Johnson, D. W., & Perkovic, V. (2017). Chronic kidney disease and the global NCDs agenda. *BMJ Global Health*, 2(2). <https://doi.org/10.1136/bmjgh-2017-000380>
- [9] Raghuvanshi, K. S. (2018). *A Qualitative Review of Two Evolutionary Algorithms Inspired by Heuristic Population Based Search Methods: GA & PSO BT - Smart Trends in Systems, Security and Sustainability* (X.-S. Yang, A. K. Nagar, & A. Joshi (eds.); pp. 169–175). Springer Singapore.
- [10] Rustam, Z., Sudarsono, E., & Sarwinda, D. (2019). Random-Forest (RF) and Support Vector Machine (SVM) Implementation for Analysis of Gene Expression Data in Chronic Kidney Disease (CKD). *IOP Conference Series: Materials Science and Engineering*, 546(5). <https://doi.org/https://doi.org/10.1088/1757-899X/546/5/052066>
- [11] Saragih, G. (2018). *Prediksi Kebangkrutan Bank dengan Menggunakan Random Forest*. Universitas Indonesia.
- [12] Suvarchla, K., Madhubala, M., Padmaja, B., & Anjaiah, P. (2018). *Lecture On Data Warehouse and Data Mining*.
- [13] Yunus, W. (2018). Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik. *Jurnal Teknik Elektro CosPhi*, 2(2), 51–55.