

KLASIFIKASI DIAGNOSIS PENYAKIT GINJAL KRONIS DENGAN MENERAPKAN KONSEP ALGORITMA NAÏVE BAYES

Siti Noor Chotimah¹⁾, Ade Ricky Rozzaqi²⁾

¹⁾Universitas Nasional Karangturi Semarang

Email: siti.chotimah@unkartur.ac.id

¹⁾Universitas PGRI Semarang

Email : zaqi@upgris.ac.id

Abstrak

Tingginya angka kematian yang disebabkan oleh Penyakit Ginjal Kronis (PGK) membuat makin banyaknya penelitian mengenai teknologi di bidang kesehatan untuk membantu mengklasifikasinya. Penelitian ini memiliki tujuan untuk melakukan klasifikasi penyakit ginjal kronis dengan menggunakan algoritma klasifikasi Naïve Bayes dan mengetahui faktor yang mempengaruhi hasil akurasi. Tahapan pertama *dataset* mentah diolah dalam tahap *pre-processing* untuk menghilangkan *missing value* dan transformasi data. Sementara pada tahapan kedua, *dataset* yang telah melalui tahap *pre-processing* diklasifikasikan dengan dua perlakuan. Perlakuan pertama menggunakan *cross validation* $k=5$ dan perlakuan kedua menggunakan $k=10$. Hasil yang diperoleh menunjukkan bahwa perlakuan jumlah *folds* pada *cross validation* $k=5$ memiliki hasil akurasi yang lebih tinggi yaitu 80%. Sedangkan penggunaan $k=10$ memiliki hasil akurasi sebesar 76%. Penggunaan jumlah $k=5$ ini dianggap dapat mengurangi terjadinya *overfitting* dari *dataset* yang digunakan. Besaran hasil akurasi ini dapat dipengaruhi dari perlakuan yang diberikan pada *dataset* yang digunakan pada saat penelitian dilakukan.

Kata kunci: Penyakit Ginjal Kronis, Klasifikasi Penyakit, Algoritma Klasifikasi, Algoritma Naïve Bayes

Abstract

The high mortality rate caused by chronic kidney disease (CKD) is doing more research on technology in the health sector to help classify it. This study aims to classify chronic kidney disease using Bayes' naïve classification algorithm and discover the factors that affect the accuracy of results. The first step of the raw dataset is processed in the preprocessing phase to eliminate missing values and transform the data. While in the second stage, datasets that have gone through the preprocessing stage are classified with two treatments. The first processing uses cross-validation $k=5$ and the second uses $k=10$. The results showed that treating the number of folds in cross-validation $k=5$ had a higher accuracy result of 80%. While the use of $k=10$ has an accuracy result of 76%. Using the number of $k=5$ is considered to reduce the occurrence of an overfit of the dataset used. The number of precision results can be influenced by the treatment given to the dataset used at the time the search was conducted.

Keywords: Chronic Kidney Disease, Disease Classification, Classification Algorithm, Naïve Bayes Algorithm

1. PENDAHULUAN

Terjadinya peningkatan jumlah pasien penderita Penyakit Ginjal Kronis (PGK) setiap tahunnya menjadikan penyakit ini termasuk ke dalam salah satu penyakit permasalahan kesehatan publik dengan cakupan global. Berdasarkan data yang dihimpun pada tahun 2018, dari total 34 provinsi yang ada di Indonesia prevalensi rata-rata dari penyakit ini sebesar 0,38%, yang dapat diartikan bahwa ada 4 dari total 1000 orang yang terdiagnosa mengalami penyakit ginjal kronis (Hidayangsih dkk., 2023). Sementara data dari *World Health Organization* (WHO) yang dihimpun tahun 2000 hingga tahun 2012 menunjukkan bahwa penyakit ginjal kronis termasuk 10 besar penyakit yang dapat menyebabkan kematian di Indonesia (WHO dkk., 2015). Kondisi yang dialami oleh pasien penderita penyakit ginjal kronis ini memiliki tingkatan keparahan yang beragam. Penanganan penyakit yang kurang serta masih kurangnya kesadaran diri dalam masyarakat tentang penyakit ginjal kronis ini dapat mengakibatkan rusaknya fungsi ginjal yang lebih parah di kemudian hari hingga kasus terparah dapat menyebabkan kematian. Terdapat faktor-faktor yang dapat termasuk dalam diagnosis untuk penyakit ginjal kronis.

Perkembangan teknologi di bidang kesehatan saat ini semakin berkembang pesat. Mulai dari pengklasifikasian penyakit berdasarkan diagnosa dari suatu penyakit, hingga ke suatu sistem pendukung keputusan yang dapat membantu memberi keputusan pendukung untuk melakukan prediksi suatu penyakit. Perkembangan teknologi tersebut dapat memanfaatkan berbagai algoritma *machine learning*. Meski tidak menjadi keputusan atau jawaban akhir untuk mendiagnosa penyakit, teknologi di bidang kesehatan ini dapat membantu menjadi alat bantu pemberian keputusan.

Penelitian terdahulu terkait dengan pengklasifikasian penyakit ginjal kronis telah banyak dilakukan. Salah satu contohnya adalah penelitian mengenai penerapan beberapa algoritma klasifikasi untuk memprediksi penyakit ginjal kronis berdasarkan 24 fitur diagnosa dalam *dataset* yang digunakan dari *UCI Machine Learning Repository* (Almustafa, 2021). Penelitian lain dengan *dataset* serupa pernah digunakan untuk melakukan komparasi beberapa algoritma dengan menambahkan perlakuan dalam *dataset* menggunakan *Principal Component Analysis* (Islam dkk., 2023). Penelitian lain terkait penyakit ginjal kronis pernah dilakukan dengan menerapkan konsep

hybrid dalam proses seleksi fitur untuk mendapatkan fitur terbaik untuk diagnosa PGK (Chotimah dkk., 2021). Oleh karena itu, fokus dari penelitian ini adalah melakukan pengklasifikasian penyakit ginjal kronis berdasarkan dari diagnosa penyakit dengan memanfaatkan salah satu algoritma *machine learning* yaitu Algoritma Naïve Bayes.

2. METODE

2.1. Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah salah satu metode untuk melakukan proses klasifikasi dengan menerapkan konsep metode probabilitas dan statistik (Sanubari dkk., 2020; Sidiq dkk., 2020). Probabilitas yang dihitung menunjukkan jumlah frekuensi serta kombinasi dari *dataset* yang digunakan untuk penelitian (Sidiq dkk., 2020). Algoritma ini dianggap dapat bekerja dengan sangat baik dibandingkan dengan metode klasifikasi lainnya, dan serta dapat melakukan prediksi peluang di masa yang akan datang dengan berdasarkan pada pengalaman di masa lampau (Sanubari dkk., 2020). Algoritma ini mengandalkan adanya probabilitas atau peluang dari *record* yang dihasilkan oleh *dataset* yang digunakan (Sidiq dkk., 2020).

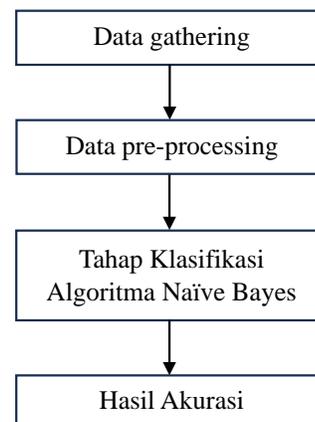
Algoritma Naïve Bayes termasuk algoritma yang dinilai simple dan masing-masing fitur atau atribut bersifat independen (Permana dkk., 2023). Perhitungan dalam Algoritma Naïve Bayes dapat menggunakan persamaan (Chen dkk., 2019):

$$P(c|\mathbf{x}) = \frac{P(c)P(\mathbf{x}|c)}{P(\mathbf{x})} \quad (1)$$

dimana $P(c)$ adalah probabilitas *priori* dari kelas data yang digunakan, $P(\mathbf{x}|c)$ adalah probabilitas kondisional (bersyarat) dari sampel tes \mathbf{x} yang dimiliki kelas c , sementara $P(\mathbf{x})$ adalah faktor normalisasi. Semua variabel tersebut diasumsikan sebagai variabel independen.

2.2. Alur Penelitian

Alur yang digunakan dalam melakukan penelitian ini dapat terlihat pada Gambar 1.



Gambar 1 Alur Penelitian Klasifikasi PGK

Tahap awal adalah mengumpulkan *dataset* mentah untuk selanjutnya masuk dalam *pre-processing*. Tahapan ini akan menghilangkan *missing value* dengan menggunakan data *average* pada masing-masing fitur yang digunakan. *Dataset* tersebut diolah dengan Algoritma Naïve Bayes dalam aplikasi WEKA.

3. HASIL DAN PEMBAHASAN

3.1. *Dataset* Penelitian

Dataset yang digunakan dalam penelitian ini sejumlah 50 data yang terdiri dari 18 fitur diagnosa penyakit dan 1 kelas (Chotimah dkk., 2021).

Tabel 1. *Dataset* PGK

No	Fitur	Jenis	Value
1.	<i>Age</i>	Numerik	Tahun
2.	<i>Blood Pressure</i>	Numerik	mm/Hg
3.	<i>Anemia</i>	Nominal	Ya, tidak
4.	<i>Nausea</i>	Nominal	Ya, tidak
5.	<i>Blown</i>	Nominal	Ya, tidak
6.	<i>VAS Scale</i>	Numerik	0, 2, 4, 6, 8, 10
7.	<i>Edema</i>	Nominal	Ya, tidak
8.	<i>Dehydration</i>	Nominal	Ya, tidak
9.	<i>Pale Skin</i>	Nominal	Ya, tidak
10.	<i>Weight</i>	Numerik	kg
11.	<i>Pulse</i>	Nominal	Regular, irregular
12.	<i>Motion</i>	Nominal	Normal, lemah, terganggu
13.	<i>Appetite</i>	Nominal	Normal, abnormal
14.	<i>Level of Consciousness</i>	Nominal	Compos mentis,

No	Fitur	Jenis	Value
			apatis, delirium, somnolent, sopor, coma
15.	<i>Temperature</i>	Numerik	°C
16.	<i>Gender</i>	Nominal	Laki-laki, perempuan
17.	<i>Another Disease</i>	Nominal	Ya, tidak
18.	<i>Falls (within 3 months)</i>	Nominal	Ya, tidak

Dataset PGK tersebut terbagi ke dalam 2 kategori kelas yaitu kelas CKD untuk pasien dengan penyakit ginjal kronis, dan CKD-A untuk pasien dengan penyakit ginjal akut. Setelah tahapan *pre-processing* data, Algoritma Naïve Bayes diterapkan untuk mengetahui hasil akurasi dari klasifikasi penyakit tersebut.

3.2. Implementasi Metode

Penelitian ini memiliki dua urutan perlakuan (tahapan) yang harus dilakukan. Tahapan pertama merupakan tahapan persiapan data. Data mentah (*raw data*) diolah dengan menggunakan tahapan *data cleaning* dan *data transformation*. Proses *data cleaning* dalam penelitian digunakan untuk menghilangkan *missing value* dengan cara mengisi data yang hilang tersebut dengan hasil rata-rata data masing-masing fitur dalam *dataset*. Sementara

proses *data transformation* dalam penelitian akan mengubah data yang berbentuk ordinal (ada, tidak ada, ya, tidak, regular, irregular, laki-laki, perempuan, normal, tidak normal) dalam bentuk biner 0 dan 1. 0 untuk positif dan 1 untuk negatif. Pada jenis kelamin seperti tertera pada Tabel 2.

Tabel 2. Konversi data jenis kelamin
Jenis Kelamin

Kategori	Numerik
Laki-laki	0
Perempuan	1

Untuk level kesadaran pasien sesuai keterangan *value* dari *dataset* pada Tabel 1 dengan nomor fitur 14 memiliki nilai rentang secara berturut-turut dari 0 sampai 5.

Tahapan kedua merupakan tahapan pengklasifikasian dengan menggunakan Algoritma Naïve Bayes. Data yang telah siap diolah masuk ke dalam sistem sebagai *input* untuk menghasilkan hasil akurasi terbaik. Tahapan kedua ini adalah tahapan yang memberikan hasil akhir berupa akurasi klasifikasi dengan menggunakan Algoritma Naïve Bayes berdasarkan perlakuan *pre-processing data* yang dilakukan sebelumnya.

3.3. Hasil dan Pembahasan

Dari *dataset* yang digunakan dalam penelitian ini berjumlah 50 data yang memiliki 18 fitur dan 1 kelas. Dalam penerapan Algoritma Naïve Bayes, terdapat dua buah perlakuan yang akan dibandingkan. Perlakuan ini menggunakan dua buah *k-fold cross validation* yang berbeda dan umum digunakan dalam penelitian, yaitu $k=5$ dan $k=10$ (Wang dkk., 2023). Perbandingan komparasi hasil akurasi yang diperoleh dari penelitian dapat terlihat pada Tabel 3.

Tabel 3. Perbandingan Hasil Akurasi

Jumlah <i>k-folds</i>	Hasil Akurasi
$k = 5$	80%
$k = 10$	76%

Dari hasil penelitian seperti yang nampak pada Tabel 3, hasil akurasi yang diperoleh dengan memberikan perlakuan *k-fold cross validation* sebesar 5 *folds* memiliki besar akurasi yang lebih baik dibandingkan dengan menggunakan 10 *folds*. Dalam penggunaan jumlah $k = 5$, *dataset* yang digunakan dalam penelitian dibagi ke dalam 5 bagian yang sama untuk *data training* dan *data testing*. Sementara $k=10$ memiliki perlakuan membagi

dataset yang digunakan ke dalam 10 bagian yang sama untuk *data training* dan *data testing*.

Hasil perhitungan dengan menggunakan parameter evaluasi untuk data yang diolah dengan menggunakan Algoritma Naïve Bayes yang mana memiliki jumlah *k-folds* sebanyak 5 *folds* dapat terlihat pada Tabel 4.

Tabel 4. Parameter Evaluasi k=5

	Prediksi CKD	Prediksi CKD-A	Recall	Precision
Benar CKD	30	5		
Benar CKD- A	5	10	0,857	0,857

Sementara pada Tabel 5 merupakan hasil perhitungan untuk parameter evaluasi data yang diolah dengan menggunakan Algoritma Naïve Bayes yang mana memiliki jumlah *k-folds* sebanyak 10 *folds*.

Tabel 5. Parameter Evaluasi k=10

	Prediksi CKD	Prediksi CKD-A	Recall	Precision
Benar CKD	29	6		
Benar CKD- A	6	9	0,829	0,829

Dalam penelitian, hasil akurasi yang dihasilkan oleh sistem dapat dipengaruhi oleh perlakuan yang diberikan terhadap data dan juga jenis data yang digunakan. Hasil *k-folds* sebesar 5 *folds* memiliki hasil yang lebih besar jika dibandingkan dengan jumlah *k-folds* sebesar 10 *folds*. Penggunaan *k-fold cross validation* sebanyak 5 *folds* dianggap lebih baik dalam melakukan penelitian karena dapat melakukan generalisasi *dataset* yang digunakan dalam penelitian dan dapat mencegah terjadinya *overfitting* (Sejuti & Islam, 2023).

4. KESIMPULAN

Berdasarkan dari penelitian yang dilakukan, dapat ditarik kesimpulan bahwa hasil akhir klasifikasi penyakit ginjal kronis dengan menggunakan Algoritma Naïve Bayes dapat dipengaruhi oleh beberapa faktor. Faktor pertama mengenai jenis data yang digunakan. Dalam penelitian ini, *dataset* yang digunakan memiliki banyak sekali fitur diagnosa penyakit yang bernilai biner (0 dan 1). Faktor kedua adalah mengenai perlakuan saat melakukan proses klasifikasi dengan menggunakan Algoritma Naïve Bayes. Dengan penggunaan *k=5 folds*, hasil akurasi yang diperoleh sebesar 80%. Hasil tersebut lebih baik jika dibandingkan

dengan perlakuan penggunaan $k=10$ folds yang memiliki hasil akurasi sebesar 76%. Dengan hasil demikian, Algoritma Naïve Bayes mampu diterapkan untuk melakukan klasifikasi terhadap penyakit ginjal kronis.

5. DAFTAR PUSTAKA

Almustafa, K. M. (2021). Prediction of chronic kidney disease using different classification algorithms. *Informatics in Medicine Unlocked*, 24, 100631. <https://doi.org/10.1016/j.imu.2021.100631>

Chen, G., Liu, Y., & Ge, Z. (2019). K-means Bayes Algorithm For Imbalanced Fault Classification and Big Data Application. *Journal of Process Control*, 81, 54–64. <https://doi.org/10.1016/j.jprocont.2019.06.011>

Chotimah, S. N., Warsito, B., & Surarso, B. (2021). Chronic Kidney Disease Diagnosis System using Sequential Backward Feature Selection and Artificial Neural Network. *E3S Web of Conferences*, 317, 1–8. <https://doi.org/10.1051/e3sconf/202131705030>

Hidayangsih, P. S., Tjandrarini, D. H., Sukoco, N. E. W., Sitorus, N., Dharmayanti, I., & Ahmadi, F. (2023). Chronic Kidney Disease in Indonesia:

Evidence From a National Health Survey. *Osong Public Health and Research Perspectives*, 14(1), 23–30. <https://doi.org/10.24171/j.phrp.2022.0290>

Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics*, 14(September 2022), 100189. <https://doi.org/10.1016/j.jpi.2023.100189>

Permana, A. A., S, W., Santoso, L. W., Wibowo, G. W. N., Wardhani, A. K., Wahidin, A. J., Rahmaddeni, R., Yuliasuti, G. E., Wijayanti, R. R., Elisawati, E., & Abdurrasyid, A. (2023). *Machine Learning*. Global Eksekutif Teknologi. https://books.google.co.id/books?id=1%5C_muEAAAQBAJ

Sanubari, T., Prianto, C., & Riza, N. (2020). *Odol (One Desa One Product Unggulan Online) Penerapan Metode Naive Bayes Pada Pengembangan Aplikasi E-Commerce Menggunakan Codeigniter*. Kreatif. https://books.google.co.id/books?id=s4j%5C_DwAAQBAJ

Sejuti, Z. A., & Islam, M. S. (2023). A Hybrid

- CNN–KNN Approach For Identification of COVID-19 With 5-Fold Cross Validation. *Sensors International*, 4(November 2022), 100229.
<https://doi.org/10.1016/j.sintl.2023.100229>
- Sidiq, Y. N. S., Fathonah, R. N. S., & Riza, N. (2020). *Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur Dengan Algoritma Naive Bayes Pada Sistem JURAGAN Berbasis Komunitas*. CV. Kreatif Industri Nusantara.
<https://books.google.co.id/books?id=BIv9DwAAQBAJ>
- Wang, Y., Khodadadzadeh, M., & Zurita-Milla, R. (2023). Spatial+: A New Cross-Validation Method to Evaluate Geospatial Machine Learning Models. *International Journal of Applied Earth Observation and Geoinformation*, 121(February).
<https://doi.org/10.1016/j.jag.2023.103364>
- WHO, Romualdez Jr., A. G., dela Rosa, J. F. E., Flavier, J. D. a., Quimbo, S. L. a., Hartigan-Go, K. Y., Lagrada, L. P., David, L. C., Kongsap, A., Paphassarang, C., Phoxay, C., Vonglokham, M., Phommavong, C., Pholsena, S., (WHO), W. H. O., Surgery, E., Hedges, J. P., Mock, C. N., Cherian, M. N., ... Who. (2015). Indonesia: WHO statistical profile. *Health Systems in Transition*, 34(9), 80.
[https://doi.org/10.1016/S0140-6736\(15\)60160-X](https://doi.org/10.1016/S0140-6736(15)60160-X)