

PERBANDINGAN KINERJA ALGORITMA NAÏVE BAYES DAN DECISION TREE UNTUK PREDIKSI PENYAKIT BATU GINJAL

Hafsah Mukaromah

Fakultas Teknologi dan Informatika, Universitas Aisyah Pringsewu, Jl. A Yani No. 1A Tambak Rejo, Pringsewu, Indonesia

Info Artikel

Riwayat Artikel:

Received April 28, 2025

Revised May 08, 2025

Accepted May 27, 2025

Corresponding Author:

Hafsah Mukaromah

Email:

hafsahmukaromah@aisyahuniversity.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract – Abstract in English

Kidney stones consist of calcium crystals and are mineral deposits that can form in the urinary tract. The prevalence of kidney stones is estimated to be 21.11%, with the Standard Age Prevalence in men and women being 24.3% and 18.7%, respectively. This study involved participants with an average age of 52.15 years, and a higher prevalence of kidney stones was observed in women aged 40-50 years and individuals with a moderate socioeconomic status. Logistic regression results indicate that the likelihood of kidney stones is higher in individuals with diabetes, hypertension, fatty liver, and overweight. The Basic Health Research (Riskesdas) data for the year 2013 shows the prevalence of Kidney Failure and Kidney Stones in Indonesia. Therefore, a profound understanding of the factors causing kidney stone formation is crucial. This research was conducted to compare the performance of the naive Bayes and decision tree algorithm models. Based on the experimental results, the decision tree algorithm showed a higher accuracy rate of 72.50% and an AUC value of 0.740%, while naive Bayes had an accuracy rate of 68.57% and an AUC value of 0.697. These results support the conclusion that the decision tree has better predictive performance than naive Bayes.

Keywords: kidney stones, comparison, prediction, naive bayes, decision tree

Abstrak – Abstrak

Batu ginjal terdiri dari kristal-kristal kalsium, merupakan deposit mineral yang dapat terbentuk di saluran kemih. Prevalensi batu ginjal diperkirakan mencapai 21,11%, dengan Angka Prevalensi Standar Usia pada pria dan wanita sebesar 24,3% dan 18,7%. Studi ini melibatkan peserta dengan usia rata-rata 52,15 tahun, dan prevalensi batu ginjal lebih tinggi terlihat pada wanita berusia 40-50 tahun dan individu dengan status sosial ekonomi tingkat sedang. Hasil regresi logistik menunjukkan bahwa peluang terjadinya batu ginjal lebih tinggi pada penderita diabetes, hipertensi, perlemakan hati, dan kelebihan berat badan. Data Riset Kesehatan Dasar (Riskesdas) tahun 2013 menunjukkan prevalensi Gagal Ginjal dan Batu Ginjal di Indonesia. Oleh karena itu, pemahaman mendalam tentang faktor-faktor penyebab pembentukan batu ginjal sangat penting. Penelitian ini dilakukan untuk mengetahui komparasi kinerja model algoritma naive bayes dan decision tree. Berdasarkan hasil eksperimen yang telah dilakukan, algoritma decision tree menunjukkan tingkat akurasi yang lebih baik yaitu sebesar 72,50% dan nilai AUC 0,740%, sementara naive Bayes memiliki tingkat akurasi 68,57% dan nilai AUC 0,697. Hasil ini mendukung kesimpulan bahwa Decision Tree memiliki kinerja prediksi yang lebih baik daripada Naive Bayes.

Kata Kunci: batu ginjal, perbandingan, prediksi, naive bayes, decision tree

I. PENDAHULUAN

Batu ginjal adalah batu yang terdapat di saluran kemih, batu yang sering dijumpai tersusun dari kristal-kristal kalsium. Batu ginjal atau kalkulus adalah bentuk deposit mineral, paling umum oksalat Ca^{2+} dan fosfat Ca^{2+} , namun asam urat dan kristal juga pembentuk batu dalam saluran kemih, batu ini umumnya ditemukan pada pelvis dan kaliks ginjal. [1]

Prevalensi batu ginjal diperkirakan 21,11%. Selain itu, Angka Prevalensi Standar Usia pada pria dan wanita dihitung masing-masing sebesar 24,3% dan 18,7%. Usia rata-rata peserta adalah 52,15 tahun. Prevalensi batu ginjal yang lebih tinggi terlihat pada wanita berusia 40-50 tahun (40,47%, $p = 0,0001$) dan status sosial ekonomi tingkat sedang (31,47%, $p = 0,03$), pria dengan berat badan berlebih (44,69%, $p < 0,0001$) dan mereka yang berada pada tingkat status sosial ekonomi sangat tinggi (35,75%, $p = 0,001$).

Hasil regresi logistik ganda menunjukkan peluang terjadinya batu ginjal 1,17 kali lebih tinggi pada penderita diabetes, 1,43 kali lebih tinggi pada penderita hipertensi, 2,21 kali lebih tinggi pada penderita perlemakan hati, dan 1,35 kali lebih tinggi pada penderita kelebihan berat badan. Tingkat status sosial ekonomi, jenis kelamin laki-laki, dan usia merupakan faktor lain yang berhubungan dengan batu ginjal. Data Riset Kesehatan Dasar (Riskesdas) tahun 2013, menunjukkan bahwa prevalensi penduduk Indonesia yang menderita Gagal Ginjal sebesar 0,2% atau 2 per 1000 penduduk dan prevalensi Batu Ginjal sebesar 0,6% atau 6 per 1000 penduduk atau 1.499.400 penduduk Indonesia menderita Batu Ginjal (Riskesdas, 2013). Oleh sebab itu, diperlukan pengetahuan yang mendalam mengenai faktor-

faktor pemicu pembentukan batu di ginjal. Dengan mengidentifikasi faktor penyebab terjadinya batu ginjal, maka tindakan pencegahan dapat dilakukan dengan mudah.

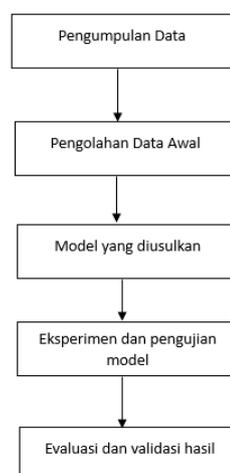
Dalam upaya untuk memperbaiki dan meningkatkan proses diagnosis penyakit batu ginjal, pemilihan algoritma prediksi yang efektif dan akurat menjadi sangat penting. Algoritma prediksi seperti Naïve Bayes dan Decision Tree telah menjadi fokus penelitian dalam domain kesehatan, termasuk dalam konteks prediksi penyakit. 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.[2] sedangkan algoritma decision tree adalah algoritma klasifikasi yang merupakan proses pembelajaran suatu fungsi tujuan yang memetakan setiap kelompok atribut dari kelas yang didefinisikan sebelumnya.[3] Penelitian sebelumnya yang dilakukan oleh [4] dengan judul “Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4.5 Dengan Membandingkan Hasil Uji Akurasi” menghasilkan tingkat akurasi sebesar 95,71 % menggunakan data latih sebanyak 70%. Penelitian selanjutnya yang berjudul “Perbandingan Metode Klasifikasi Dalam Memprediksi Penyakit Ginjal Kronis” yang dilakukan oleh [5] bertujuan untuk mengklasifikasi seseorang apakah pasien tergolong ginjal kronis atau tidak dengan menggunakan algoritma *decision tree* C4.5 dan *Naïve bayes* yang kemudian dievaluasi menggunakan *confussion matrix*.

Hasil penelitian menunjukkan algoritma decision tree C4.5 mendapatkan hasil akurasi sebesar 93,00 %, *precision* 84,27%, *recall* 100% dan AUC sebesar 94,40%. Sedangkan algoritma naïve bayes menghasilkan tingkat akurasi sebesar 93,50%, *precision* 85,23%, *recall* 100% dan AUC sebesar 94,80%. dari hasil uji coba perbandingan kedua algoritma tersebut menunjukkan algoritma naïve bayes memiliki nilai akurasi yang lebih tinggi daripada algoritma decision tree C4.5. Selanjutnya penelitian dengan judul “Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik” dilakukan oleh [6] menggunakan dataset sebanyak 400 record dari UCI Machine Learning Repository. Melalui tahapan preprocessing data yang sudah siap uji sebanyak 280 data berupa 196 data *training* dan 84 data *testing*. Hasil pengujian diperoleh algoritma Naïve Bayes Classifier (NBC) memiliki akurasi yang tinggi dengan presentase 96.43%. Penelitian lainnya yang dilakukan oleh [7] dengan judul “Komparasi Metode Klasifikasi Data Mining untuk Prediksi Penyakit Ginjal” bertujuan untuk memprediksi penyakit ginjal kronis dengan menggunakan algoritma klasifikasi seperti Support Vector Machine dan Algoritma Decision Tree C4.5. klasifikasi menggunakan algoritma C4.5 menunjukkan performa yang lebih unggul daripada SVM.

Perbandingan hasil pengujian menunjukkan akurasi sebesar 95.00% dengan menggunakan algoritma SVM, sementara dengan algoritma C4.5 akurasi mencapai 97.25%. Oleh karena itu, dapat disimpulkan bahwa pengujian dengan menggunakan Algoritma C4.5 menghasilkan tingkat akurasi yang lebih tinggi dibandingkan dengan SVM. Dari hasil penelitian, terlihat bahwa Dengan demikian, perbandingan kinerja antara algoritma Naïve Bayes dan Decision Tree untuk prediksi penyakit batu ginjal menjadi relevan untuk dieksplorasi. Diharapkan penelitian ini dapat memberikan pemahaman yang lebih mendalam mengenai keunggulan dan kelemahan masing-masing algoritma dalam konteks prediksi penyakit batu ginjal. Hasil penelitian ini dapat menjadi dasar untuk pengembangan metode diagnostik yang lebih efisien dan akurat dalam menangani kasus penyakit batu ginjal.

II. METODE

Penelitian ini menggunakan metode penelitian eksperimen dengan tahapan penelitian sebagai berikut:



Gambar 1. Tahapan Penelitian

Berdasarkan gambar diatas, berikut alur penelitian yang dilakukan:

2.1 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. [2] Dalam penelitian ini penulis menggunakan dataset publik untuk memprediksi keberadaan batu ginjal berdasarkan analisis urin yang diunduh dari laman <https://www.kaggle.com/datasets/harshghadiya/kidneystone>. Dataset ini berisi 89 data yang terdiri dari 6 atribut, diantaranya: gravity, ph, osmo, cond, urea, calc.

1.2 Pengolahan Data Awal

Pada tahap ini dilakukan pengolahan data untuk mendapatkan data dengan kualitas terbaik, beberapa tahapan yang dilakukan yaitu:

1. Cleaning Data/Pembersihan Data
Mengidentifikasi dan menangani nilai yang hilang, outlier, atau data yang tidak lengkap (*missing value*).
2. Data Integration/Integrasi Data
Menggabungkan data dari berbagai sumber yang mungkin memiliki format atau skema yang berbeda dan membuat satu set data yang bersatu.
3. Data Selection/Seleksi Data
Memilih subset data yang relevan untuk dianalisis.
4. Transformasi Data
Mengubah data ke dalam format yang lebih sesuai atau informatif untuk dianalisis, sehingga dapat dilakukan persiapan dalam pembuatan model.

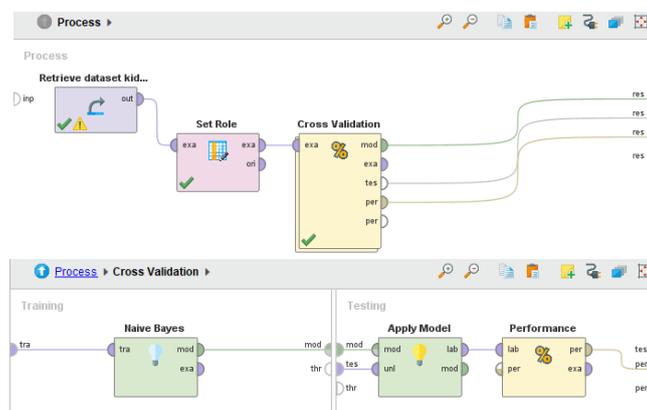
1.3 Model Yang Diusulkan

Pada tahap ini dilakukan analisis data lalu diterapkan model yang sesuai dengan jenis data yang digunakan. Data dibagi menjadi data testing dan data training. Algoritma yang digunakan dalam penelitian ini yaitu decision tree dan naïve bayes. Dalam penelitian ini akan dilakukan komparasi algoritma decision tree dan naïve bayes dengan menggunakan cross validation untuk pengujian model. Selanjutnya hasil pengujian di evaluasi dengan confusion matrix untuk mendapatkan hasil algoritma terbaik dalam memprediksi penyakit batu ginjal.

III. HASIL DAN PEMBAHASAN

a. Eksperimen dan Pengujian Algoritma Naïve Bayes

Pengolahan data menggunakan algoritma naïve bayes dan tools rapidminer dengan 10 fold validation dapat dilihat pada gambar berikut:

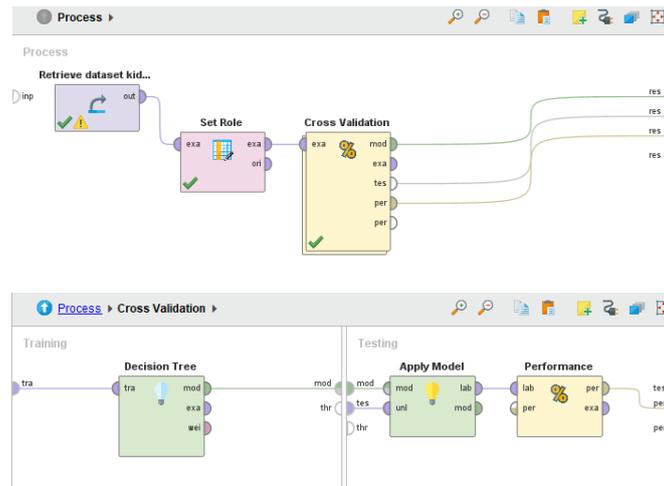


Gambar 2. Pengujian Algoritma Naïve Bayes dengan Rapidminer

Pada pengujian menggunakan algoritma naïve bayes, terdapat beberapa operator yang digunakan yaitu operator set role untuk **Menetapkan Atribut Target (Label)**, operator operator apply model untuk **mengaplikasikan model algoritma naïve bayes ke dalam data training** dan operator performance digunakan untuk mengevaluasi kinerja model algoritma naïve bayes.

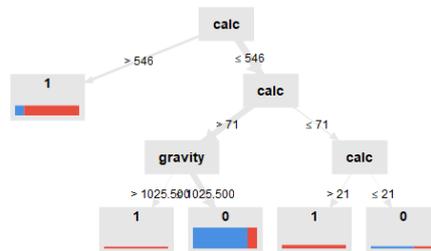
3.1 Eksperimen dan Pengujian Algoritma Decision Tree

Pengolahan data menggunakan algoritma decision tree dan tools rapidminer dengan 10 fold validation dapat dilihat pada gambar berikut:



Gambar 3. Pengujian Algoritma Decision Tree dengan Rapidminer

Berdasarkan pengujian menggunakan algoritma decision tree maka diperoleh pohon Keputusan sebagai berikut:



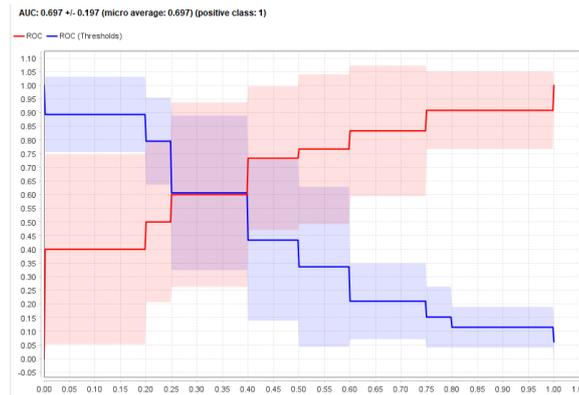
Gambar 4. Pohon Keputusan

3.2 Evaluasi dan Validasi Hasil

Setelah melalui tahapan pengolahan data dengan menggunakan tools rapidminer maka diperoleh hasil akurasi dari kedua model tersebut. Hasil akurasi naïve bayes sebesar 68,57%, precision 61,50% dan recall 57,50%. Sedangkan hasil akurasi algoritma decision tree sebesar 72,50%, precision 66,83% dan recall 70,00%. Untuk pengujian akurasi algoritma naïve bayes menggunakan confusion matrix yang ditunjukkan pada gambar 5 berikut.

		true 0	true 1	class precision
pred. 0		34	14	70.83%
pred. 1		11	20	64.52%
class recall		75.56%	58.82%	

Gambar 5. Nilai Akurasi Algoritma Naïve Bayes



Gambar 6. Kurva AUC Algoritma Naïve Bayes

Dari gambar 5 menunjukkan bahwa dengan algoritma naïve bayes, berdasarkan 79 data pasien diperoleh 34 pasien yang diprediksi batu ginjal hasilnya positive batu ginjal. Lalu, 14 orang yang diprediksi batu ginjal ternyata tidak menderita batu ginjal. Sebanyak 20 orang yang diprediksi tidak batu ginjal hasilnya benar tidak menderita batu ginjal, namun ada 11 orang yang diprediksi tidak batu ginjal ternyata positive batu ginjal. Maka, didapatkan nilai akurasi algoritma naïve bayes adalah 68,57 %.

Gambar 6 menunjukkan kurva AUC dari model algoritma naïve bayes. Garis biru pada kurva ROC umumnya menunjukkan performa model atau kelas pertama (kelas positif). Semakin dekat kurva ke sudut kiri atas, semakin baik kinerja model dalam membedakan kelas positif dari kelas negatif. Sedangkan garis merah, menunjukkan performa model atau kelas kedua (kelas negatif). Pada umumnya, semakin dekat kurva ke sudut kiri atas, semakin baik model dalam meminimalkan jumlah false positive (mengklasifikasikan kelas negatif sebagai kelas positif). Pada gambar diatas garis biru bergerak mendekati sudut kiri atas yang berarti kinerja model algoritma naïve bayes semakin baik dalam membedakan kelas positif dari kelas negative. Namun garis merah bergerak menjauhi sudut kiri atas dan bergerak ke sudut kanan atas yang berarti semakin buruk kinerja model dalam mengklasifikasikan kelas negatif sebagai kelas positif.

Pengujian dengan algoritma Decision Tree menghasilkan nilai akurasi yang ditunjukkan pada tabel confusion matrix berikut.

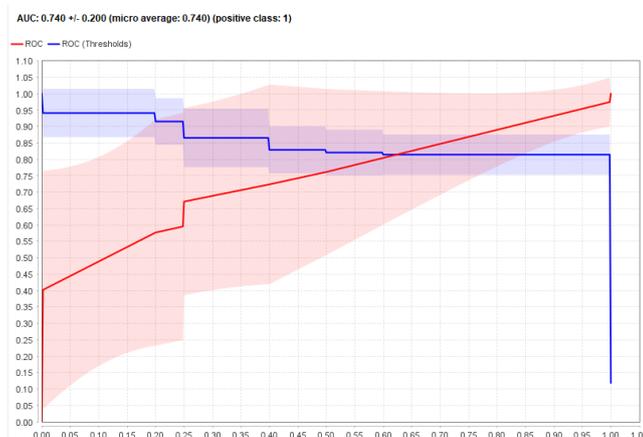
nceVector (Performance) x Tree (Decision Tree) x

Table View Plot View

accuracy: 72.50% +/- 19.36% (micro average: 72.15%)

	true 0	true 1	class precision
pred. 0	33	10	76.74%
pred. 1	12	24	66.67%
class recall	73.33%	70.59%	

Gambar 7. Nilai Akurasi Algoritma Decision Tree



Gambar 8. Kurva AUC Algoritma Decision Tree

Pada gambar 7 menunjukkan bahwa dengan algoritma Decision Tree, sebanyak 33 pasien yang diprediksi menderita batu ginjal hasilnya benar positive batu ginjal. 11 pasien yang diprediksi batu ginjal tetapi tidak menderita batu ginjal. Sementara 12 pasien yang diprediksi tidak batu ginjal ternyata positive batu ginjal. Sedangkan 24 pasien yang diprediksi negative batu ginjal hasilnya benar tidak menderita batu ginjal/negative batu ginjal. Nilai akurasi yang diperoleh menggunakan algoritma decision tree sebesar 72,50 %.

Gambar 8 menunjukkan nilai AUC untuk model decision tree sebesar 0,740 %.

IV. SIMPULAN

Berdasarkan hasil eksperimen pada pasien batu ginjal menunjukkan bahwa kinerja prediksi decision tree sebagai algoritma klasifikasi lebih baik daripada algoritma naïve bayes. Dimana tingkat akurasi decision tree sebesar 72,50 % dan nilai AUC 0,740 %. Sedangkan, pendekatan naïve Bayes menunjukkan tingkat akurasi sebesar 68,57% dan nilai AUC 0,697, yang lebih rendah dibandingkan dengan decision tree. Untuk meningkatkan kualitas penelitian ini, perlu dilakukan pengembangan melalui penerapan metode optimasi guna mencapai tingkat akurasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] U. Hasanah, "Mengenal Penyakit Batu Ginjal," *J. Kel. Sehat Sejah.*, vol. 14, no. 28, pp. 76–85, 2016, [Online]. Available: <https://jurnal.unimed.ac.id/2012/index.php/jkss/article/view/4698/4129>
- [2] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naïve Bayes Classifier Berbasis Particle Swarm Optimization," *J. Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019, doi: 10.36787/jti.v13i1.97.
- [3] J. Bonardo Junior, R. Rohmat Saedudin, and V. P. Widharta, "Perbandingan Akurasi Algoritma Decision Tree Dan Algoritma Support Vector Machine Pada Penyakit Diabetes," *e-Proceeding Eng.*, vol. 8, no. 5, pp. 9749–9756, 2021.
- [4] Y. Widiastwi and I. Ernawati, "Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4 . 5 Dengan Membandingkan Hasil Uji Akurasi," *J. IKRA-ITH Inform.*, vol. 5, no. 2, p. 128, 2021.
- [5] P. G. Kronis, "1 , 2 , 3".
- [6] Q. A'yuniyah *et al.*, "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 72, 2022, doi: 10.30865/json.v4i1.4781.
- [7] S. Handayani, "No Title," *Komparasi Metod. Klasifikasi Data Min. untuk Prediksi Penyakit Ginjal*, vol. 6, no. 1, pp. 1–7, 2017.
- [8] K. Pemikiran, "Penerapan Algoritma K-Means," vol. 7, no. 2, pp. 2–9, 2011.
- [9] A. Rohman and M. Rochcham, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Kelulusan Mahasiswa," *Neo Tek.*, vol. 5, no. 1, pp. 34–40, 2019, doi: 10.37760/neoteknika.v5i1.1379.
- [10] Qadrini L, Sepperwali A, and Aina A, "Decision Tree dan Adaboostpada Klasifikasi Penerima Program Bantuan Sosial," *Decis. Tree Dan Adab. Pada Klasifikasi Penerima Progr. Bantu. Sos.*, vol. 2, no. 7, pp. 1959–1966, 2021.
- [11] A. Nur Rais and Warjiyono, "Optimasi Akurasi Klasifikasi Pada Prediksi Smokte Detection dengan Menggunakan Algoritma Adaboost," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, pp. 343–348, 2022, doi: 10.30865/json.v4i2.5154.
- [12] H. Bugis, "Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 6, no. 1, pp. 8–14, 2022, doi: 10.47970/siskom-kb.v6i1.317.
- [13] T. D. Kaniya and D. Uyun, "Jurnal Ilmiah Kesehatan Sandi Husada Ct-Scan Non Kontras Pada Pasien Batu Saluran Kemih Pendahuluan," vol. 11, no. 1, pp. 284–291, 2020, doi: 10.35816/jiskh.v10i2.272.