

## Penerapan SMOTE dan XGBoost untuk Klasifikasi Penyakit Ginjal Kronis pada Data yang Tidak Seimbang

Muhammad Ghufron Syifa <sup>1\*</sup>, Nadia Anissa Maori <sup>2</sup>, Adi Sucipto <sup>3</sup>

<sup>1\*,2,3</sup> Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Nabdlatul Ulama Jepara, Kota Jepara, Provinsi Jawa Tengah, Indonesia.

### article info

#### Article history:

Received 9 September 2025

Received in revised form

20 September 2025

Accepted 1 November 2025

Available online January 2026.

#### Keywords:

Chronic Kidney Disease;  
XGBoost; SMOTE.

#### Kata Kunci:

Penyakit Ginjal Kronis;  
XGBoost; SMOTE.

### abstract

Chronic kidney disease (CKD) is a global health issue that requires accurate diagnosis to prevent errors and unwanted side effects. This study aims to develop a reliable classification model using the XGBoost algorithm and to explore the effectiveness of the SMOTE method in addressing data imbalance. The dataset is sourced from the UCI Machine Learning Repository, consisting of 400 patient records with 25 attributes. The results indicate that the developed model performs well, with evaluation metrics (Accuracy, Precision, Recall, F1-Score, and AUC-ROC) nearing 1.0. The research findings reveal that the model without SMOTE is slightly superior, achieving an accuracy of 0.9874 compared to 0.9811 for the model with SMOTE. Analysis shows that the data imbalance is not significant, and XGBoost is resilient to unbalanced data. This study also identifies key factors influencing CKD diagnosis, such as hemoglobin and albumin, which can assist medical professionals in making more accurate diagnoses.

### abstract

Penyakit ginjal kronis (CKD) adalah isu kesehatan global yang memerlukan diagnosis yang tepat untuk mencegah kesalahan dan efek samping. Penelitian ini bertujuan mengembangkan model klasifikasi andal menggunakan algoritma XGBoost dan mengeksplorasi efektivitas metode SMOTE dalam mengatasi ketidakseimbangan data. Dataset berasal dari UCI Machine Learning Repository, terdiri dari 400 data pasien dengan 25 atribut. Hasilnya menunjukkan model yang dikembangkan memiliki performa baik dengan nilai evaluasi (Akurasi, Presisi, Recall, F1-Score, dan AUC-ROC) mendekati 1.0. Melalui hasil penelitian diketahui bahwa model tanpa SMOTE sedikit lebih unggul, dengan akurasi 0.9874 dibandingkan 0.9811 untuk model dengan SMOTE. Analisis menunjukkan bahwa ketidakseimbangan data cukup signifikan, dan XGBoost tahan terhadap data tidak seimbang. Penelitian ini juga mengidentifikasi faktor-faktor utama yang memengaruhi diagnosis CKD, seperti hemoglobin dan albumin, yang dapat membantu profesional medis dalam melakukan diagnosis yang lebih tepat.

\*Corresponding Author. Email: [ghufronsyifa453@gmail.com](mailto:ghufronsyifa453@gmail.com) <sup>1\*</sup>.

## 1. Pendahuluan

Penyakit Ginjal Kronis (PGK) atau *Chronic Kidney Disease* (CKD) merupakan persoalan kesehatan global yang terus meningkat dari tahun ke tahun. Gangguan ini ditandai oleh penurunan fungsi ginjal secara bertahap dan bersifat permanen. Deteksi lebih awal memiliki peran penting dalam mencegah komplikasi berat serta menekan beban ekonomi pada sektor kesehatan. Perkembangan *machine learning* memberi peluang besar bagi dunia medis dalam menghasilkan diagnosis yang lebih cepat dan presisi. Namun, tantangan muncul ketika data pasien tidak seimbang jumlah penderita PGK jauh lebih sedikit dibandingkan pasien sehat. Ketimpangan semacam itu dapat menyebabkan model klasifikasi cenderung bias terhadap kelas dengan jumlah data lebih banyak, sehingga kemampuan mengenali kasus PGK menjadi lemah. Berbagai penelitian telah berupaya mengatasi masalah tersebut melalui teknik penyeimbangan data. Irfannandhy (2024), Sidiq (2025), dan Faska (2025) membuktikan bahwa metode SMOTE (*Synthetic Minority Over-sampling Technique*) mampu memperbaiki distribusi kelas dan meningkatkan performa model pada prediksi penyakit metabolik seperti diabetes dan NAFLD. Teknik ini menambah data sintesis pada kelas minoritas melalui interpolasi antar titik data yang berdekatan, sehingga model memperoleh representasi yang lebih proporsional.

Sementara itu, algoritma *Extreme Gradient Boosting* (XGBoost) dikenal efektif dalam mengolah data kompleks serta menangani nilai yang hilang dengan efisien (Abdurrahman, 2022; Hakim & Zyen, 2025). Beberapa penelitian lanjutan (Kurniawan, 2025; Maulana, 2025; Surono, 2025) menunjukkan bahwa penerapan XGBoost dengan penyesuaian parameter mampu menghasilkan tingkat akurasi tinggi pada berbagai bidang kesehatan. Sebagian besar studi terdahulu masih berfokus pada penyakit lain seperti diabetes mellitus. Penelitian tentang penerapan XGBoost dan SMOTE untuk diagnosis PGK masih jarang dilakukan (Abdurrahman, 2022; Irfannandhy, 2024; Hakim & Zyen, 2025; Sidiq, 2025; Faska, 2025; Kurniawan, 2025; Maulana, 2025; Surono, 2025). Bhirawa (2025) pernah mengimplementasikan SMOTE pada kasus PGK, tetapi belum menguji secara langsung efektivitas XGBoost dibandingkan algoritma lain.

Penelitian ini difokuskan untuk menilai sejauh mana penerapan SMOTE dapat meningkatkan kinerja XGBoost pada data diagnosis PGK yang tidak seimbang. Hasil yang diperoleh diharapkan dapat menjadi dasar empiris bagi pengembangan sistem pendukung keputusan klinis yang lebih akurat dan efisien dalam membantu proses diagnosis penyakit ginjal kronis.

## 2. Metodologi Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental terapan yang bertujuan mengukur efektivitas metode SMOTE dalam menangani ketidakseimbangan data serta menilai kinerja algoritma XGBoost dalam klasifikasi penyakit ginjal kronis. Data yang digunakan bersifat sekunder dan diperoleh dari sumber terbuka (*open dataset*), dengan rancangan eksperimen berbasis *data mining* dan *machine learning*.

### Waktu dan Tempat Penelitian

Kegiatan penelitian berlangsung pada Mei hingga Juli 2025. Seluruh proses pengolahan data, pelatihan model, serta evaluasi dilakukan secara daring menggunakan perangkat pribadi dengan lingkungan pemrograman Python di Google Colab dan Jupyter Notebook.

### Populasi dan Sampel Penelitian

Populasi penelitian mencakup seluruh data pasien dalam *dataset* penyakit ginjal kronis. Sampel terdiri atas 400 data pasien yang diambil dari *Chronic Kidney Disease Dataset* di *UCI Machine Learning Repository*. Data tersebut memperlihatkan ketidakseimbangan antara kelompok penderita PGK dan non-PGK. Seluruh data digunakan sebagai sampel melalui teknik total sampling, sehingga setiap catatan pasien berperan dalam tahap pelatihan dan pengujian model.

### Teknik Pengumpulan Data

Data diperoleh dari *dataset* publik berjudul "*Chronic Kidney Disease*" yang tersedia di Kaggle. Data ini mencakup 400 pasien dengan dua label target, yaitu *ckd* (penderita PGK) dan *notckd* (non-PGK). Terdapat 26 atribut, terdiri atas 25 variabel prediktor dan 1 variabel target dengan tipe data numerik dan kategorikal.

## Teknik Analisis Data

Analisis dilakukan melalui empat tahap utama: pra-pemrosesan data, penerapan SMOTE, pembangunan model XGBoost, dan evaluasi hasil. Setiap tahap dirancang untuk memastikan data yang digunakan bersih, seimbang, dan menghasilkan model klasifikasi yang andal.

### 1) Pra-pemrosesan Data

Langkah awal dilakukan dengan mengidentifikasi nilai hilang dan melakukan imputasi. Nilai kosong pada data numerik diganti dengan rata-rata (*mean*), sedangkan pada data kategorikal diganti dengan modus. Selanjutnya, dilakukan deteksi dan pembersihan *outlier* untuk menjaga konsistensi distribusi data. Setelah proses tersebut, seluruh fitur dinormalisasi menggunakan *Min-Max Scaling* agar setiap variabel memiliki skala yang sebanding. Fitur kategorikal diubah ke bentuk numerik melalui *encoding*, kemudian data dibagi menjadi 60% untuk pelatihan dan 40% untuk pengujian. Distribusi kelas dianalisis untuk memastikan tingkat ketidakseimbangan sebelum dilakukan penyeimbangan dengan SMOTE.

### 2) SMOTE (*Synthetic Minority Over-sampling Technique*)

SMOTE diterapkan untuk memperbaiki ketidakseimbangan data dengan menambah sampel sintesis pada kelas minoritas. Prosesnya meliputi identifikasi kelas minoritas, penentuan tetangga terdekat menggunakan algoritma *k-nearest neighbors*, pembentukan sampel baru melalui interpolasi, serta penggabungan hasilnya dengan data asli.

### 3) XGBoost (*Extreme Gradient Boosting*)

Model klasifikasi dibangun menggunakan XGBoost, yang dikenal efektif dalam menangani data berskala besar dan kompleks. Tahapan pengembangannya meliputi inisialisasi parameter, pelatihan model pada data latih, serta penerapan regularisasi untuk mencegah *overfitting*. Model kemudian diuji pada data uji guna menilai kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

### 4) Evaluasi Model dan Validasi

Kinerja model diukur menggunakan lima metrik utama: akurasi, presisi, recall, F1-score, dan AUC-ROC. Selain itu, dilakukan validasi silang (*k-fold cross-validation*) untuk memastikan hasil evaluasi stabil dan representatif terhadap seluruh data.

Nilai rata-rata dari setiap lipatan digunakan sebagai ukuran performa akhir model.

## 3. Hasil dan Pembahasan

### Hasil

#### Gambaran Umum Data

Penelitian menggunakan *dataset* yang bersumber dari UCI *Machine Learning Repository*, terdiri atas 400 data pasien dengan 25 atribut. Dari jumlah tersebut, 24 atribut berperan sebagai variabel prediktor yang menggambarkan kondisi klinis pasien, sedangkan satu atribut menjadi variabel target yang mengklasifikasikan pasien ke dalam dua kategori, yaitu penyakit ginjal kronis (*ckd*) dan tidak mengalami penyakit ginjal kronis (*notckd*). Distribusi kelas menunjukkan ketimpangan yang cukup jelas: 250 data termasuk kelas *ckd* dan 150 data termasuk kelas *notckd*. Ketidakseimbangan tersebut menjadi dasar penggunaan metode SMOTE untuk meningkatkan proporsi kelas minoritas sebelum proses pelatihan model dilakukan.

#### Implementasi Pra-Pemrosesan Data

Tahapan pra-pemrosesan dilakukan untuk memastikan kualitas data yang akan digunakan dalam pelatihan model. Proses ini meliputi penanganan nilai hilang, identifikasi *outlier*, normalisasi, dan konversi data kategorikal ke bentuk numerik. Langkah-langkah ini diperlukan agar model *machine learning* dapat mempelajari pola data secara lebih optimal.

#### Penanganan Nilai Hilang dan *Outlier*

Beberapa atribut dalam *dataset* mengandung nilai kosong yang dapat memengaruhi hasil analisis. Nilai hilang pada variabel numerik diganti menggunakan rata-rata (*mean*), sedangkan nilai kosong pada variabel kategorikal diganti menggunakan modus. Setelah imputasi, seluruh atribut memiliki nilai lengkap. Selain itu, dilakukan identifikasi *outlier* dengan membandingkan nilai ekstrem terhadap referensi klinis dari literatur medis. Data yang menyimpang secara ekstrem dihapus agar tidak memengaruhi kestabilan model. Setelah proses pembersihan, diperoleh 396 data yang valid dan siap digunakan pada tahap selanjutnya.

### Normalisasi Data

Seluruh fitur numerik dinormalisasi menggunakan *Min-Max Scaling* sehingga setiap variabel berada dalam rentang 0 hingga 1. Proses ini memastikan semua fitur memiliki pengaruh yang seimbang terhadap model, mencegah dominasi variabel dengan skala yang lebih besar, serta mempercepat konvergensi algoritma saat pelatihan dilakukan.

### Pembagian Dataset

Data dibagi menjadi dua bagian: 60% untuk pelatihan (237 data) dan 40% untuk pengujian (159 data). Pembagian ini dirancang agar model dapat dilatih secara optimal sekaligus memberikan evaluasi yang objektif terhadap kemampuan generalisasinya pada data baru.

### Hasil Penyeimbangan Data dengan SMOTE

Metode SMOTE (*Synthetic Minority Over-sampling Technique*) digunakan untuk memperbaiki ketimpangan distribusi kelas dengan menambahkan sampel sintesis pada kelas minoritas. Proses ini dilakukan melalui interpolasi antar data pada kelas *notckd*, menghasilkan proporsi kelas yang setara dengan kelas *ckd*. Setelah penerapan SMOTE, dataset pelatihan menjadi seimbang sehingga model dapat mempelajari pola dari kedua kelas dengan proporsi yang adil. Hal ini diharapkan meningkatkan sensitivitas model terhadap kasus penyakit ginjal kronis yang sebelumnya kurang terwakili.

### Hasil Pelatihan dan Pengujian Model XGBoost

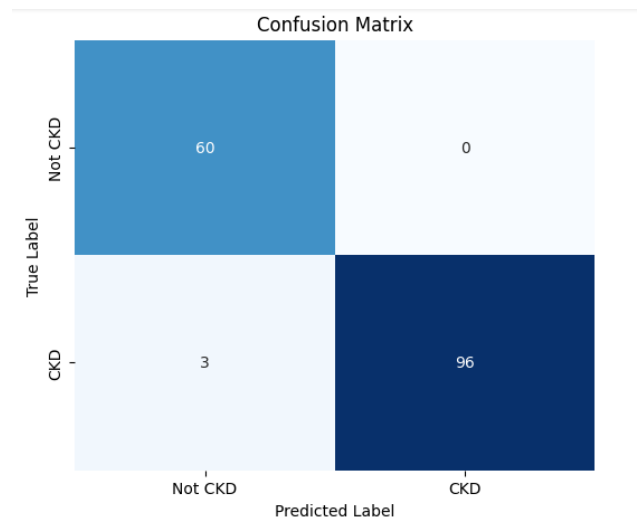
Model *XGBoost* dibangun menggunakan data pelatihan yang telah melalui proses pra-pemrosesan dan penyeimbangan. Prosedur pelatihan mencakup penyesuaian parameter dasar, proses pembelajaran bertahap melalui *boosting*, serta penerapan regularisasi untuk mencegah *overfitting*. Hasil evaluasi menunjukkan bahwa model memiliki performa yang sangat baik berdasarkan lima metrik utama: akurasi, presisi, *recall*, F1-score, dan AUC-ROC. Nilai yang diperoleh berada di kisaran tinggi dan konsisten pada data pengujian, menandakan kemampuan model dalam mengklasifikasikan pasien dengan dan tanpa penyakit ginjal kronis secara efektif.

Tabel 1. Hasil Metrik Evaluasi

| No | Metrik    | Nilai  |
|----|-----------|--------|
| 1  | Accuracy  | 0.9811 |
| 2  | Precision | 0.9820 |
| 3  | Recall    | 0.9811 |
| 4  | F1-Score  | 0.9812 |
| 5  | AUC-ROC   | 0.9997 |

### Confusion Matrix

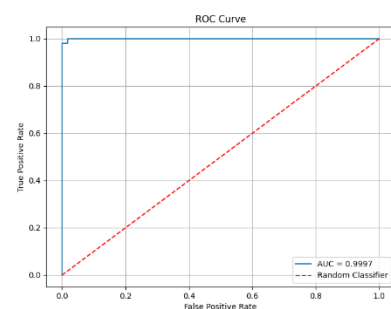
*Confusion matrix* memberikan rincian hasil prediksi model, membantu memahami jenis kesalahan yang dibuat oleh model. Dari 159 data uji, model berhasil mengidentifikasi 96 True Positives dan 3 False Positives.



Gambar 1. *Confusion Matrix*

### Kurva ROC

Kurva ROC menunjukkan hubungan antara *True Positive Rate* dan *False Positive Rate*. Nilai AUC-ROC sebesar 0.9997 menunjukkan kemampuan model yang sangat baik dalam membedakan antara pasien CKD dan notCKD.



Gambar 2. Kurva ROC

### Analisis Hasil Penelitian

Analisis hasil menunjukkan efektivitas metode SMOTE dalam meningkatkan kemampuan model XGBoost dalam mengklasifikasikan penyakit ginjal kronis. Meskipun model dengan SMOTE menunjukkan performa yang sangat baik, model tanpa SMOTE memiliki akurasi yang sedikit lebih tinggi (0.9874) dibandingkan dengan model yang menggunakan SMOTE (0.9811).

Tabel 2. Perbandingan Hasil Metrik Evaluasi dengan dan tanpa SMOTE

| No | Metrik    | Nilai (dengan SMOTE) | Nilai (tanpa SMOTE) |
|----|-----------|----------------------|---------------------|
| 1  | Accuracy  | 0.9811               | 0.9874              |
| 2  | Precision | 0.9820               | 0.9874              |
| 3  | Recall    | 0.9811               | 0.9874              |
| 4  | F1-Score  | 0.9812               | 0.9874              |
| 5  | AUC-ROC   | 0.9997               | 0.9998              |

Perbedaan hasil menunjukkan bahwa pada *dataset* dengan tingkat ketidakseimbangan yang ringan, penerapan SMOTE tidak memberikan peningkatan berarti terhadap performa model. Temuan ini sejalan dengan penelitian sebelumnya yang menegaskan bahwa SMOTE paling efektif diterapkan pada data dengan ketimpangan kelas yang ekstrem. Dengan demikian, dapat disimpulkan bahwa *XGBoost* memiliki kemampuan adaptif yang baik dalam menghadapi distribusi data yang tidak seimbang secara moderat. Model ini mampu mempertahankan tingkat akurasi tinggi tanpa bantuan teknik penyeimbangan tambahan, sementara penggunaan SMOTE pada kondisi data yang relatif stabil justru tidak memberikan keuntungan signifikan.

### Pembahasan

Hasil penelitian menunjukkan bahwa algoritma XGBoost mampu mengklasifikasikan penyakit ginjal kronis dengan tingkat ketepatan yang sangat tinggi. Nilai akurasi, presisi, *recall*, F1-score, dan AUC-ROC yang hampir sempurna menandakan kemampuan model dalam mengenali pola data secara konsisten. Kinerja tersebut sejalan dengan hasil penelitian Abdurrahman (2022) serta Hakim dan Zyen (2025) yang menemukan bahwa *XGBoost* efektif menangani data medis beragam karena pendekatannya yang berlapis dalam memperbaiki kesalahan prediksi. Penerapan metode SMOTE dalam penelitian ini

memang menghasilkan model yang stabil, tetapi peningkatan performa tidak terlalu besar. Model tanpa SMOTE justru mencapai akurasi lebih tinggi, yaitu 0,9874 dibandingkan 0,9811 pada model yang menggunakan SMOTE. Hasil tersebut menunjukkan bahwa ketika perbandingan jumlah kelas masih relatif seimbang, XGBoost dapat bekerja dengan baik tanpa perlu data sintesis tambahan. Kesimpulan serupa ditemukan oleh Haryawan (2023), yang menyatakan bahwa efektivitas SMOTE akan menurun ketika perbedaan jumlah kelas tidak ekstrem. Variabel hemoglobin dan albumin muncul sebagai faktor dominan dalam diagnosis penyakit ginjal kronis. Temuan tersebut mendukung penelitian Gliselda (2021) dan Yuniarti (2021), yang menegaskan bahwa kadar hemoglobin rendah serta gangguan albumin mencerminkan penurunan fungsi ginjal. Hal itu memperlihatkan bahwa model *XGBoost* tidak hanya unggul secara statistik, tetapi juga selaras dengan pola klinis yang diakui dalam literatur kedokteran.

Meskipun SMOTE tidak memberikan lonjakan performa yang signifikan, metode ini tetap membantu meningkatkan sensitivitas model terhadap kasus pasien dengan jumlah data yang lebih sedikit. Sidiq (2025) dan Faska (2025) juga menemukan bahwa SMOTE mampu memperbaiki kemampuan model dalam mengenali kelas minoritas, terutama pada penyakit kronis yang jarang muncul di *dataset*. Perannya lebih terasa dalam memperluas ruang pembelajaran bagi algoritma, meskipun tidak selalu berdampak langsung pada peningkatan akurasi keseluruhan. Penelitian ini memperkuat pemahaman bahwa XGBoost merupakan algoritma yang tangguh, efisien, dan relatif tahan terhadap distribusi data yang tidak seimbang. Hasil serupa ditunjukkan oleh Soelistijadi *et al.* (2024), yang membuktikan keandalan model berbasis ensemble seperti *XGBoost* dan *LightGBM* dalam analisis data medis. Nilai AUC-ROC yang mencapai 0,9997 menunjukkan kemampuan model untuk membedakan pasien dengan penyakit ginjal kronis dari yang sehat dengan presisi tinggi. Temuan tersebut membuka peluang bagi pengembangan sistem pendukung keputusan berbasis pembelajaran mesin di bidang kesehatan. Model seperti *XGBoost* dapat diadaptasi dalam aplikasi klinis untuk membantu dokter menilai risiko penyakit ginjal lebih cepat berdasarkan data laboratorium pasien.

Hasil serupa pernah dilaporkan oleh Maulana dan Ernawati (2025), yang menunjukkan potensi *ensemble learning* dalam membantu diagnosis otomatis dengan tingkat keakuratan tinggi. Secara keseluruhan, hasil penelitian menegaskan bahwa *XGBoost* tetap menjadi pilihan efektif untuk klasifikasi penyakit ginjal kronis, bahkan tanpa penyeimbangan data tambahan. Metode SMOTE memiliki nilai praktis pada kondisi ketidakseimbangan yang lebih berat, tetapi pada data yang proporsional, algoritma *XGBoost* sudah mampu mencapai stabilitas prediksi yang baik tanpa intervensi tambahan.

#### 4. Kesimpulan dan Saran

Penelitian menunjukkan bahwa penggunaan metode SMOTE tidak memberikan peningkatan yang signifikan terhadap performa algoritma *XGBoost* dalam klasifikasi penyakit ginjal kronis. Model tanpa SMOTE mencatat akurasi sebesar 0,9874, sedikit lebih tinggi dibandingkan model dengan SMOTE yang mencapai 0,9811. Hasil tersebut menunjukkan bahwa penambahan data sintesis tidak selalu diperlukan, terutama ketika perbandingan jumlah data antar kelas masih seimbang. Algoritma *XGBoost* terbukti mampu mengenali pola pada kelas minoritas dengan baik tanpa proses oversampling. Hal ini menegaskan bahwa *XGBoost* memiliki ketahanan yang kuat terhadap distribusi data yang tidak merata dan dapat menghasilkan prediksi yang akurat pada kasus medis. Sebagai tindak lanjut, penelitian berikutnya dapat menambahkan variasi tingkat ketidakseimbangan data atau menggabungkan SMOTE dengan metode lain seperti *Adaptive Synthetic Sampling* (ADASYN) atau *SMOTE-ENN* untuk melihat perbandingan hasil yang lebih luas. Selain itu, pengujian menggunakan dataset klinis yang lebih besar dan beragam dapat memberikan gambaran yang lebih realistis terhadap penerapan model dalam sistem diagnosis berbasis pembelajaran mesin.

#### 5. Ucapan Terima Kasih

Kami ingin mengucapkan terima kasih kepada UNISNU Jepara dan para dosen atas kerjasama dan dukungan yang sangat berharga selama penelitian ini.

#### 6. Daftar Pustaka

- Abdurrahman, G., Oktavianto, H., & Sintawati, M. (2022). Optimasi algoritma *XGBoost* classifier menggunakan hyperparameter gridsearch dan random search pada Klasifikasi penyakit diabetes. *Informal Informatics J*, 7(3), 193.
- Akbar, I., Supriadi, F., & Junaedi, D. I. (2025). Pemanfaatan Machine Learning di Bidang Kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 1744-1749.
- Anggraini, D. (2022). Aspek klinis dan pemeriksaan laboratorium penyakit ginjal kronik. *An-Nadaa: Jurnal Kesehatan Masyarakat (e-Journal)*, 9(2), 236-239.
- Cahyaningtyas, C., Nataliani, Y., & Wideasari, I. R. (2021). Analisis sentimen pada rating aplikasi Shopee menggunakan metode Decision Tree berbasis SMOTE. *Aiti*, 18(2), 173-184. <https://doi.org/10.24246/aiti.v18i2.173-184>.
- Ernawati, S., & Maulana, I. (2025). Meningkatkan Klasifikasi Penyakit Diabetes Menggunakan Metode Ensemble Softvoting Dengan SMOTE-ENN dan Optimasi Bayesian. *Evolusi: Jurnal Sains dan Manajemen*, 13(1), 71-86. <https://doi.org/10.31294/evolusi.v13i1.8267>.
- Faska, R. M., Gusti, S. K., Budianita, E., & Syafria, F. (2025). PENGARUH TEKNIK PENYEIMBANGAN DATA PADA KLASIFIKASI PENYAKIT NAFLD DENGAN ALGORITMA SVM. *Jurnal Informatika Teknologi dan Sains (Jinteks)*, 7(2), 858-867. <https://doi.org/10.58602/jima-ilkom.v4i1.41>.
- Gliselda, V. K. (2021). Diagnosis dan Manajemen Penyakit Ginjal Kronis (PGK). *Jurnal Medika Hutama*, 2(04 Juli), 1135-1141.
- Haryawan, C., & Ardhana, Y. M. K. (2023). Analisa Perbandingan Teknik Oversampling SMOTE pada Imbalanced Data. *Jurnal Informatika dan Rekayasa Elektronik*, 6(1), 73-78.

- Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 714-723. <https://doi.org/10.29408/edumatic.v8i2.27990>.
- Juwita, D. A., Rachmaini, F., Abdillah, R., & Meliani. (2022). *Drugs related problems (DRPs) pada pasien penyakit ginjal kronik (PGK) di RSUP Dr. M. Djamil. Jurnal Sains Farmasi & Klinis*, 9(Sup), 184. <https://doi.org/10.25077/jsfk.9.sup.184-189.2022>.
- Kurniawan, H. (2025). Evaluasi performa Random Forest, XGBoost, dan LightGBM dalam diagnosis dini diabetes mellitus. *Jurnal JUPITER*, 17(2), 835–844.
- Ningrum, R. Y. S., Astuti, D., & Puspasari, F. D. (2022). Literature review pengaruh relaksasi dzikir terhadap stres pada pasien chronic kidney disease (CKD) yang menjalani hemodialisa. *Journal of Nursing and Health (JNH)*, 7(2), 169–175.
- Ramadhan, K. G., Wibowo, G. W. N., & Maori, N. A. (2025). Komparasi deteksi penyakit ginjal kronis menggunakan algoritma Support Vector Machine dan Random Forest. *Jurnal Informatika dan Rekayasa Elektronik*, 8(1), 13–21.
- Siboro, O., Banjarnahor, Y. P., Gultom, A., Siagian, N. A., & Silitonga, P. D. (2024, Juni). Penanganan data ketidakseimbangan dalam pendekatan SMOTE guna meningkatkan akurasi algoritma K-NN. *Seminar Nasional Inovasi Sains Teknologi Informasi Komputer*, 473–478.
- Sidiq, S., Alfian, A., & Maburur, N. S. (2025). Pengembangan Model Prediksi Risiko Diabetes Menggunakan Pendekatan AdaBoost dan Teknik Oversampling SMOTE. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, 4(1), 13-23. <https://doi.org/10.58602/jima-ilkom.v4i1.41>.
- Soelistijadi, R., Wismarini, T. D., Eniyati, S., & Sunardi, S. (2024). Pemodelan prediktif menggunakan metode ensemble learning XGBoost dalam peningkatan akurasi klasifikasi penyakit ginjal. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 5(4), 1866–1875.
- Surono, M., Fadli, M., Purwanti, D. S., & Susanto, E. R. (2025). Hybrid XGBoost-SVM Model untuk Sistem Pendukung Keputusan dalam Prediksi Penyakit Diabetes. *INSOLOGI: Jurnal Sains dan Teknologi*, 4(3), 443-454. <https://doi.org/10.55123/insologi.v4i3.5410>.
- Utami, M. P. S., & Dwi Susanti, B. A. (2022). *Awareness questionnaire versi bahasa Indonesia untuk pasien chronic kidney disease (CKD) dengan hemodialisa: Pengembangan dan validitas. Jurnal Ilmiah Kesehatan Sandi Husada*, 11, 175–181. <https://doi.org/10.35816/jiskh.v11i1.720>.
- Wijayanti, N. P. Y. T., Kencana, E. N., & Sumarjaya, I. W. (2021). SMOTE: potensi dan kekurangannya pada survei. *E-Jurnal Matematika*, 10(4), 235.
- Yulianti, S. E. H., Soesanto, O., & Sukmawaty, Y. (2022). Penerapan Metode Extreme Gradient Boosting (XGBOOST) pada Klasifikasi Nasabah Kartu Kredit. *Journal of Mathematics: Theory and Applications*, 21-26. <https://doi.org/10.31605/jomta.v4i1.1792>.
- Yuniarti, W. (2021). Anemia in Chronic Kidney Disease Patients. *Journal Health & Science: Gorontalo Journal Health and Science Community*, 5(2), 341-347.
- Zhao, Y., Wong, Z. S. Y., & Tsui, K. L. (2018). A framework of rebalancing imbalanced healthcare data for rare events' classification: a case of look-alike sound-alike mix-up incident detection. *Journal of healthcare engineering*, 2018(1), 6275435.