Jurnal Teknologi Informasi dan Multimedia

p-ISSN : 2715-2529

e-ISSN : 2684-9151

https://journal.sekawan-org.id/index.php/jtim

Optimization of Support Vector Machine Using SMOTE and Grid Search for Kidney Health Data Classification

Muhammad Maulana [1]*, Zulkipli [2], Tanwir [1], Dading Oktaviadi Resmiranta [1], Naufal Hanif [3], Raisul Azhar [1]

1. Program Studi Ilmu Komputer, Universitas Bumigora, Indonesia.
2. Program Studi Pendidikan Teknologi Informasi, Universitas Bumigora, Indonesia.
3. Program Studi Teknologi Informasi, Universitas Bumigora, Indonesia.

* Korespondensi: muhammadmaulana@universitasbumigora.ac.id

Sitasi: M. Maulana, Z. Zulkipli, T. Tanwir, D. O. Resmiranta, N. Hanif, and R. Azhar, "Optimization of Support Vector Machine Using SMOTE and Grid Search for Kidney Health Data Classification," *Jurnal Teknologi Informasi Dan Multimedia*, vol. 8, no. 2, pp. 431–443, 2026, https://doi.org/10.35746/jtim.v8i2.993

Diterima: 19-04-2026
Direvisi: 20-05-2026
Disetujui:30-05-2026

Abstract: Kidney disease is a highly prevalent health problem that can seriously impact the quality of life of those affected. To improve diagnostic accuracy, machine learning methods are widely used to classify patient data. Class imbalance (imbalanced data) is one of the problems that often occurs in the classification process and can affect the performance of machine learning models, especially in detecting minority classes. This study aims to improve the performance of the Support Vector Machine (SVM) algorithm by applying the SMOTE (Synthetic Minority Over-sampling Technique) and Grid Search methods in the data classification process. SMOTE is used to balance the class distribution by adding synthetic data to the minority class, while Grid Search is used to obtain optimal model parameters. The results show that the SVM model without handling data imbalance produces relatively low performance with an accuracy value of 51%, precision 17%, recall 33%, and F1-score 23%. After applying the SMOTE method, the model performance increases significantly to 81% accuracy, 81% precision, 80% recall, and 81% F1-score. Furthermore, the application of Grid Search to the SVM + SMOTE model provides the best results with an accuracy of 84%, precision 82%, recall 81%, and F1-score 81% with an AUC value of 0,92. The findings of this study indicate that the combination of SMOTE and Grid Search is effective in improving the performance of the SVM algorithm in data classification. The novelty of this study demonstrates that data imbalance management and hyperparameter optimization play a crucial role in producing more accurate and optimal classification models.

Keywords: Support Vector Machine, SMOTE, Grid Search, imbalanced data, machine learning

Abstrak: Penyakit ginjal merupakan salah satu masalah kesehatan yang memiliki tingkat prevalensi cukup tinggi dan dapat berdampak serius terhadap kualitas hidup penderitanya. Dalam upaya peningkatan akurasi diagnosis, metode machine learning banyak digunakan untuk melakukan klasifikasi data pasien. Ketidakseimbangan kelas (*imbalanced data*) merupakan salah satu permasalahan yang sering terjadi pada proses klasifikasi dan dapat memengaruhi performa model machine learning, khususnya dalam mendeteksi kelas minoritas. Penelitian ini bertujuan untuk meningkatkan kinerja algoritma Support Vector Machine (SVM) melalui penerapan metode SMOTE (Synthetic Minority Over-sampling Technique) dan Grid Search pada proses klasifikasi data. SMOTE digunakan untuk menyeimbangkan distribusi kelas dengan menambahkan data sintetis pada kelas minoritas, sedangkan Grid Search digunakan untuk memperoleh parameter model yang optimal. Hasil penelitian menunjukkan bahwa model SVM tanpa penanganan imbalance data menghasilkan performa yang relatif rendah dengan nilai accuracy 51%, precision 17%, recall 33%, dan F1-score 23%. Setelah diterapkan metode SMOTE, performa model meningkat secara signifikan

menjadi accuracy 81%**,** precision 81%**,** recall 80%**,** dan F1-score 81%**.** Selanjutnya, penerapan Grid Search pada model SVM+SMOTE memberikan hasil terbaik dengan nilai accuracy 84%, precision 82%, recall 81%, dan F1-score 81% dengan nilai AUC 0,92. Temuan dari penelitian adalah kombinasi SMOTE dan Grid Search efektif dalam meningkatkan kinerja algoritma SVM pada proses klasifikasi data. Kebaruan dalam Penelitian ini menunjukkan bahwa penanganan ketidakseimbangan data serta optimasi hyperparameter memiliki peran penting dalam menghasilkan model klasifikasi yang lebih akurat dan optimal.

Kata kunci: Support Vector Machine, SMOTE, Grid Search, imbalanced data, machine learning

1. Pendahuluan

Penyakit ginjal merupakan salah satu masalah kesehatan yang memiliki tingkat prevalensi cukup tinggi dan dapat berdampak serius terhadap kualitas hidup penderitanya[1],[2]. Jika tidak terdeteksi dan ditangani sejak dini, gangguan kesehatan ginjal berpotensi berkembang menjadi penyakit ginjal kronis yang membutuhkan perawatan jangka panjang serta biaya medis yang besar[3]. Oleh karena itu, deteksi dini kondisi kesehatan ginjal menjadi hal yang sangat penting untuk mendukung pengambilan keputusan medis yang lebih cepat dan akurat.

Perkembangan teknologi informasi, khususnya di bidang machine learning, telah membuka peluang besar dalam pengolahan dan analisis data kesehatan[4],[5]. Pemanfaatan teknik klasifikasi berbasis machine learning mampu membantu tenaga medis dalam mengidentifikasi pola-pola tersembunyi pada dataset kesehatan ginjal, sehingga proses diagnosis dapat dilakukan secara lebih objektif dan efisien[6],[7],[8]. Salah satu algoritma yang sering digunakan dalam tugas klasifikasi adalah SVM, karena kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan performa yang baik pada berbagai permasalahan klasifikasi[9],[10].

Namun, penerapan SVM pada data kesehatan ginjal sering menghadapi beberapa tantangan. Salah satu permasalahan utama adalah ketidakseimbangan kelas (imbalanced data), di mana jumlah data pasien sehat dan pasien dengan gangguan ginjal tidak seimbang[10][11]. Kondisi ini dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas, sehingga menurunkan kemampuan model dalam mendeteksi kelas minoritas yang justru lebih penting dalam konteks medis[12]. Oleh karena itu, untuk menyelesaikan masalah ketidakseimbangan data tersebut, diperlukan pendekatan khusus.

SMOTE merupakan salah satu metode yang efektif untuk mengatasi ketidakseimbangan data dengan cara menghasilkan data sintetis pada kelas minoritas[13]. Dengan penerapan SMOTE, distribusi data antar kelas menjadi lebih seimbang sehingga model klasifikasi, termasuk SVM, dapat belajar secara lebih optimal[14]. Selain itu, performa SVM juga sangat dipengaruhi oleh pemilihan parameter, seperti nilai C, kernel, dan gamma. Pemilihan parameter yang tidak tepat dapat menyebabkan model mengalami overfitting atau underfitting.

Untuk mengoptimalkan kinerja SVM, diperlukan teknik pencarian parameter yang sistematis dan terstruktur. Grid Search merupakan metode optimasi parameter yang bekerja dengan mengevaluasi seluruh kombinasi parameter yang telah ditentukan untuk memperoleh konfigurasi terbaik berdasarkan metrik evaluasi tertentu. Dengan mengombinasikan SMOTE untuk penanganan ketidakseimbangan data dan Grid Search untuk optimasi parameter, diharapkan kinerja SVM dalam mengklasifikasikan data kesehatan ginjal dapat meningkat secara signifikan.

Beberapa penelitian terdahulu telah dilakukan terkait penerapan metode machine learning dalam mendiagnosa penyakit ginjal. Penelitian oleh [15] menggunakan metode optimasi Binary Coati (BCOA), GLCM, GLRLM, GLSZM, GLDM, NGTDM, dan orde pertama, untuk pemilihan fitur pada CKD, kemudian di evaluasi menggunakan metode klasifikasi Random Forest (RF), SVM, Decision Tree (DT), K-nearest Neighbor (KNN), XGboost (XGB) dan Naïve Bayes (NB). Hasil penelitian menunjukan BCOA-V berkinerja lebih baik dalam hal akurasi, presisi, recall, spesifisitas, skor F1, dan kurva AUC masingmasing sebesar 99%, 100%, 97%, 100%, 98%, dan 98%. Penelitian lain oleh[8] menerapkan algoritma machine learning linier diskriminan analisis (LDA), Naïve Bayes,C4.5, dan Random Forest menggunaka dataset yang berisi 1.659 instance dan 52 fitur, yang mencakup data demografis, gaya hidup, dan klinis. Hasil penelitian menunjukan C4.5 mencapai akurasi tertinggi sebesar 92,5%, diikuti oleh Random Forest (92,2%), dengan Naïve Bayes tetap berada di 92,1%, LDA tetap yang paling unggul mencapai akurasi 92,8%.

Penelitin oleh [16] menerapkan algoritma Machine Learning DT, Neural Network (NN), RF, XGB, Gaussian Naive Bayes (GNB), dan pengklasifikasi CatBoost (CB). Hasil penelitian menunjukkan Hasilnya menunjukkan RF, XGB, dan CB mendapatkan akurasi lebih tinggi sebesar 95% yang lebih cocok dalam memprediksi CKD daripada DT, NN, dan GNB. Selanjutnya penelitian oleh [7] menerapkan metode machine learning Regresi Logistic, Gradien Boosting dan RF. Hasil penelitian menunjukkan bahwa penggabungan dari ketiga algoritma machine learning menghasilkan RMSE 0,2111 dan MSE 0,0446. Penelitian lain oleh [17] menggunakan metode DT, NB dan SVM. Hasil penelitian menunjukkan bahwa SVM memiliki tingkat akurasi paling tinggi yaitu 98% dari pada DT dan NB.

Meskipun berbagai penelitian telah menunjukkan efektivitas metode *machine learning* dalam mendiagnosis penyakit *Chronic Kidney Disease* (CKD), selain itu penelitian sebelumnya banyak melakukakan klasifikasi biner yang hanya membandingkan kategori CKD dan Not CKD. Penelitian ini yang secara khusus menggabungkan metode SMOTE dan *Grid Search* dengan algoritma SVM dalam klasifikasi penyakit CKD berdasarkan gejala klinis masih relatif terbatas. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada penerapan teknik SMOTE sebagai tahap *preprocessing* data untuk meningkatkan kualitas serta menyeimbangkan distribusi kelas dalam dataset. Selain itu penelitian ini melakukan klasifikasi multikelas dengan tiga tingkat keparahan (normal, sedang, tinggi),selain itu klasifikasi multikelas memiliki tingkat kesulitan yang jauh lebih tinggi karena batas keputusan antar kelas yang berdekatan antara sedang dan tinggi bersifat *fuzzy* dan tumpang tindih secara klinis sehingga penurunan akurasi merupakan konesekuensi ilmiah yang wajar dari peningkatan kompleksitas tugas klasifikasi . Selanjutnya dataset yang digunakan dalam penelitian ini memiliki distribusi asimetris (379 normal, 728 sedang, 393 tinggi) yang tidak seimbang,Selain itu, metode *Grid Search* digunakan untuk melakukan *hyperparameter tuning* guna memperoleh kombinasi parameter terbaik sehingga dapat meningkatkan performa model. Penerapan kedua teknik tersebut diharapkan mampu memaksimalkan keunggulan algoritma SVM dalam menangani data berdimensi tinggi serta mengurangi risiko *overfitting.*

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi efektivitas penerapan SMOTE dan *Grid Search* dalam meningkatkan kinerja model klasifikasi penyakit CKD. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan kualitas data pelatihan serta mengoptimalkan performa algoritma klasifikasi, sehingga mampu menghasilkan sistem diagnosis yang lebih akurat dan andal untuk membantu praktisi kesehatan dalam mendeteksi penyakit CKD secara lebih dini.

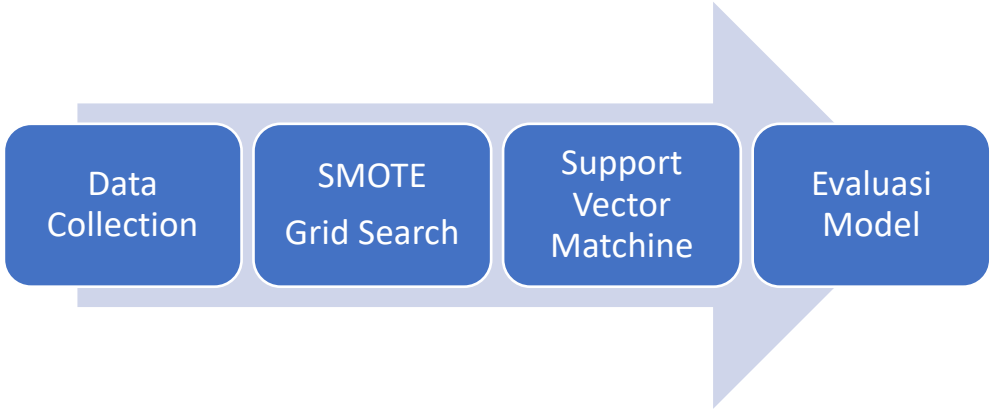
Table 1 Perbandingan penelitian ini dan karya terkait

Penulis	Optimasi	Imbalance Data	Metode Klasifikasi	Hasil
Fizhan Kausar et al.[15]	Binary Coati	-	RF,SVM,DT, KNN,XGboost dan NB	RF=99% SVM=100% DT=97% KNN=100%, XGBoost=98%, NB=98%.
Abrahem et al.[8]	-	-	LDA,NB,C4.5, dan RF	LDA=92,8% C4.5=92,5% RF=92,2% NB=92,1%
Yamini et al.[16]	-	-	DT,NN,RF,XGBoost,GNB, Cat-Boost.	RF,XGBoost, CB=95%
Batini et al. [7]	-	-	Regresi Logistic, Gradien Boost-ing dan RF	RMSE= 0,2111 MSE= 0,0446
Syarif et al.[17]	-	-	DT, NB, dan SVM	SVM=98%
Our/research	Grid Search	SMOTE	SVM	Original = 51% After Optimasi = 84%

2. Bahan dan Metode

Riset ini menempuh serangkaian prosedur sistematis guna menciptakan model klasifikasi penyakit ginjal kronis (CKD) yang memiliki presisi tinggi. Metodologi yang diterapkan mengintegrasikan algoritma SVM sebagai pengklasifikasi utama, yang dioptimalkan dengan teknik SMOTE dan *Grid Search*. Alur penelitian diawali dengan akuisisi data dari dataset CKD, yang kemudian dilanjutkan ke fase pra-pemrosesan. Pada tahap ini, data dibersihkan, nilai-nilai yang hilang diatasi, dan dilakukan transformasi data untuk memastikan kesiapan input dalam pemodelan. Guna menanggulangi kendala ketidakseimbangan kelas yang sering ditemukan pada data medis, teknik SMOTE diimplementasikan untuk menyeimbangkan distribusi sampel sebelum masuk ke tahap klasifikasi.

Tahapan berikutnya dalam penelitian ini adalah proses pemodelan klasifikasi dengan menggunakan algoritma SVM. Pada tahap ini dilakukan optimasi parameter melalui metode Grid Search untuk mendapatkan kombinasi parameter yang optimal sehingga dapat meningkatkan kinerja model. Selanjutnya, model dievaluasi menggunakan beberapa metrik pengujian, yaitu accuracy, precision, recall, dan F1-score, untuk mengukur kemampuan model dalam melakukan klasifikasi data. Secara keseluruhan, alur tahapan penelitian yang dilakukan disajikan pada Gambar 1.



Gambar 1 Metodologi Penelitian

2.1. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan menggunakan dataset penyakit ginjal yang diperoleh dari platform Kaggle terdiri dari 1500 sampel data. Setiap sampel memiliki 36 atribut yang mencerminkan kondisi klinis pasien, parameter laboratorium, serta karakteristik farmakokinetik dan toksikologi yang berkaitan dengan gangguan fungsi ginjal. Atribut yang digunakan meliputi *patient age*, *gender*, *blood pressure systolic*, *blood pressure diastolic*, *blood urea*, *serum creatinine*, *albumin*, *blood glucose random*, *diabetes*, *hypertension*, *drug name*, *drug dosage (mg)*, *exposure days*, *nephrotoxic label*, *molecular weight*, *logP*, *hydrogen bond donors*, *hydrogen bond acceptors*, *rotatable bonds*, *topological polar surface area (TPSA)*, *shape index 3D*, *inertia x*, *inertia y*, *inertia z*, *charge distribution*, *clearance rate*, *half-life (hour)*, *bioavailability (%)*, *volume of distribution*, *kidney cell viability (%)*, *mitochondrial damage*, *oxidative stress*, *protein binding (%)*, *serum creatinine change (%)*, *toxicity score composite*, dan *PK toxic interaction score*. Atribut-atribut tersebut dipilih karena secara klinis dan farmakologis memiliki keterkaitan dengan fungsi ginjal serta potensi terjadinya gangguan ginjal kronis. Dataset ini terdiri dari tiga kelas, yaitu kategori normal (kelas 0), risiko sedang (kelas 1), dan risiko tinggi (kelas 2). Gambaran umum dataset dapat dilihat pada Tabel 2. Data yang telah diperoleh selanjutnya digunakan sebagai input utama pada tahap pra-pemrosesan (*preprocessing*) dan pengembangan model *machine learning* dalam penelitian ini.

Table 2 Data Penyakit Ginjal

No	patient_age	gender	bp_systolic	…..	pk_toxic_interaction_score	ckd_risk_label
0	69	1	136.05.00	…..	00.29	2
1	32	1	125.00.00	…..	00.25	0
2	89	0	124.02.00	…..	00.35	1
3	78	1	98.07.00	…..	0,049305556	0
4	38	1	147.07.00	…..	00.38	2
5	41	0	128.04.00	…..	0,059027778	0
…..	…..	…..	…..	…..	…..	…..
…..	…..	…..	…..	…..	…..	…..
1491	43	1	141.06.00	…..	00.25	1
1492	20	0	115.09.00	…..	0,054166667	0
1493	58	0	147.01.00	…..	0,059722222	1
1494	77	1	163.00.00	…..	0,043055556	2
1495	31	0	151.04.00	…..	00.51	2
1496	29	0	115.04.00	…..	00.28	2
1497	29	0	137.02.00	…..	00.39	2
1498	30	0	104.01.00	…..	00.01	1
1499	42	1	141.04.00	…..	0,063888889	0

2.2. Preprocessing Data

Sebelum menerapkan teknik SMOTE, tahap awal yang dilakukan dalam penelitian ini adalah pemeriksaan kualitas data, meliputi identifikasi data duplikat dan penanganan data kosong (missing values). Proses ini bertujuan untuk memastikan bahwa dataset telah

melalui tahapan pembersihan sehingga berada dalam kondisi yang siap untuk digunakan dalam proses pemodelan. Setelah dilakukan penanganan missing value dan data duplikat, dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Data latih digunakan untuk proses pelatihan model, sedangkan data uji digunakan untuk mengukur performa model. Setelah proses pembersihan data selesai, SMOTE diterapkan sebagai salah satu teknik penyeimbangan data pada tahap preprocessing. Teknik ini digunakan untuk mengatasi permasalahan ketidakseimbangan kelas dengan menghasilkan data sintetis pada kelas minoritas. Kondisi ketidakseimbangan kelas berpotensi menyebabkan model lebih cenderung memprediksi kelas mayoritas, yang berdampak pada menurunnya akurasi prediksi pada kelas minoritas.

Tahap berikutnya adalah hyperparameter tuning menggunakan metode Grid Search untuk menentukan kombinasi parameter terbaik pada algoritma klasifikasi. Metode ini mengevaluasi berbagai kombinasi parameter dan memilih yang menghasilkan performa optimal. Dengan demikian, model diharapkan menjadi lebih stabil dan akurat. Penerapan SMOTE dan Grid Search pada tahap awal juga bertujuan untuk meningkatkan kualitas data serta performa model klasifikasi.

2.3. Modelling

SVM merupakan salah satu algoritma *machine learning* yang digunakan untuk tugas klasifikasi dan regresi dengan cara menentukan *hyperplane* optimal yang mampu memisahkan data ke dalam kelas yang berbeda secara maksimal. SVM bekerja dengan memaksimalkan margin antara dua kelas sehingga menghasilkan batas keputusan yang paling optimal. Algoritma ini efektif dalam menangani data berdimensi tinggi serta dapat menggunakan fungsi *kernel* untuk memetakan data non-linear ke ruang berdimensi lebih tinggi.

Dalam penelitian ini, performa algoritma SVM ditingkatkan melalui proses *hyperparameter tuning* menggunakan metode *Grid Search*. Metode ini digunakan untuk mencari kombinasi parameter terbaik dari beberapa parameter utama pada SVM, seperti parameter C**,** gamma, dan jenis kernel yang digunakan. *Grid Search* bekerja dengan mengevaluasi berbagai kombinasi nilai parameter yang telah ditentukan sebelumnya untuk menemukan konfigurasi yang menghasilkan performa model terbaik. Dengan penerapan *Grid Search*, diharapkan model SVM yang dihasilkan memiliki tingkat akurasi yang lebih tinggi serta mampu mengurangi risiko *overfitting* dalam proses klasifikasi.

2.4. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai performa model klasifikasi yang telah dibangun. Pada penelitian ini, evaluasi model dilakukan menggunakan *confusion matrix* sebagai dasar perhitungan berbagai metrik kinerja. *Confusion matrix* memberikan informasi mengenai jumlah prediksi benar dan salah yang dihasilkan model untuk masing-masing kelas, sehingga dapat digunakan untuk menganalisis kemampuan model secara lebih rinci. Rumus perhitungan confusion matrix evaluasi tersebut disajikan pada Tabel 3.

Berdasarkan table 3 *confusion matrix*, Evaluasi performa model dilakukan dengan menggunakan metrik accuracy, precision, recall, dan F1-score. Accuracy merepresentasikan rasio prediksi yang benar terhadap seluruh data uji. Recall mengukur kemampuan model dalam mengidentifikasi seluruh instance yang relevan dalam suatu kelas, sedangkan precision menunjukkan tingkat keakuratan prediksi positif yang dihasilkan model. F1-score, sebagai rata-rata harmonis dari precision dan recall, digunakan untuk mengevaluasi keseimbangan antara kedua metrik tersebut. Formula perhitungan masing-masing metrik ditunjukkan pada Persamaan (1) hingga Persamaan (4).

Tabel 3 Confusion Matrik

Aktual	**Prediksi**	
	Judul 2	**Judul 3**
Positif	TP	FN
Negatif	FP	TN

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

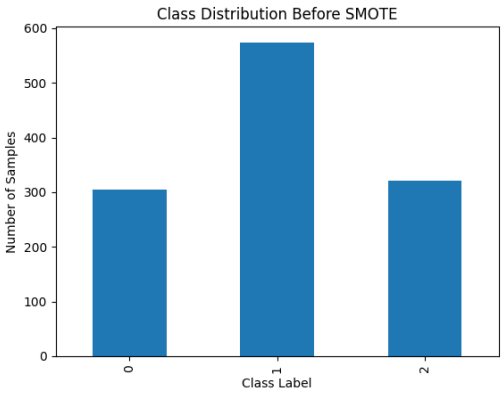
$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1\ Score = 2 \times \frac{Precision\ \times Recall}{Precision\ \times\ Recall} \quad (4)$$

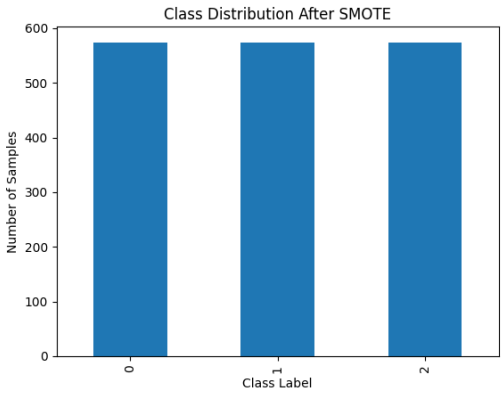
3. Hasil dan Pembahasan

Bagian ini menguraikan hasil analisis dan temuan yang diperoleh dari penelitian. Dataset yang digunakan adalah data CKD yang bersumber dari Kaggle, terdiri dari 1500 sampel dengan 36 atribut independen dan satu atribut target sebagai label kelas. Penelitian ini gunakan dataset Chronic Kidney Disease (CKD) dataset tersebut terdiri dari tiga kelas, yaitu normal sebanyak 379 data**,** sedang sebanyak 728 data**,** dan tinggi sebanyak 393 data. Berdasarkan distribusi kelas tersebut, terlihat bahwa data memiliki ketidakseimbangan jumlah pada masing-masing kategori, di mana kelas sedang mendominasi jumlah data dibandingkan kelas lainnya. Dataset dapat diakses pada link berikut https://drive.google.com/file/d/1X4PbSHrFLBKS9CAxwLahavIL_oXK5pqO/view?usp=sharing. Label tersebut merepresentasikan hasil pemeriksaan laboratorium yang transformasi ke dalam tiga kategori, yaitu 0 (normal), 1 (sedang), dan 2 (tinggi).

Tahap awal penelitian adalah preprocessing data yang bertujuan untuk meningkatkan kualitas, konsistensi, serta kesiapan dataset sebelum digunakan dalam proses pemodelan. Proses ini meliputi pemeriksaan data duplikat, pengecekan missing values, serta transformasi data ke dalam format yang sesuai untuk proses klasifikasi. Dalam penelitian ini memiliki karakteristik data awal yang baik, sehingga tidak ditemukan data yang missing value, selanjutnya transformasi data yang dilakukan untuk mengubah atribut yang bertipe string ke dalam bentuk numerik menggunakan teknik label encoding. Selanjutnya, diterapkan teknik SMOTE untuk mengatasi permasalahan ketidakseimbangan kelas (class imbalance). Ketidakseimbangan data dapat menyebabkan model cenderung bias terhadap kelas mayoritas sehingga menurunkan performa prediksi pada kelas minoritas[13]. Dengan penerapan SMOTE, distribusi data pada setiap kelas menjadi lebih seimbang sehingga model dapat belajar secara lebih optimal. Distribusi data pada dataset Kelas 0 memiliki jumlah data sebanyak 305 sampel, kelas 1 sebanyak 574 sampel, dan kelas 2 sebanyak 321 sampel. Setelah proses SMOTE dilakukan, jumlah data pada masing-masing kelas meningkat menjadi 574 sampel untuk kelas 0, kelas 1, dan kelas 2 Perbandingan distribusi kelas sebelum dan sesudah penerapan SMOTE seperti yang ditunjukkan pada Gambar 2 dan 3.



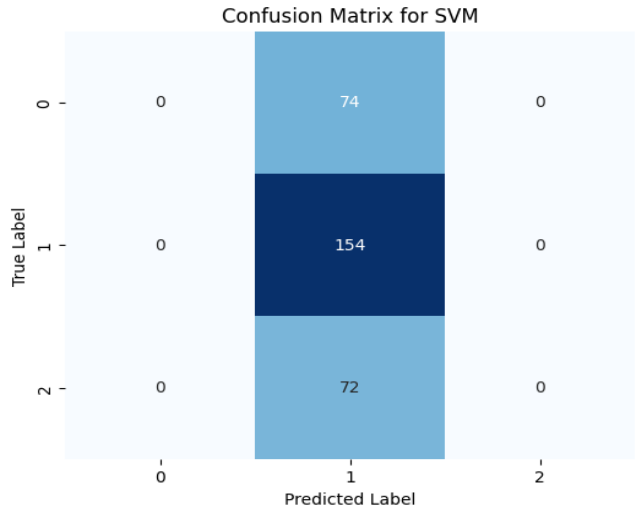
Gambar 2 Sebelum Smote



Gambar 3 Setelah Smote

Setelah tahap preprocessing data selesai dilakukan, langkah selanjutnya adalah mengimplementasikan metode klasifikasi menggunakan SVM. Implementasi metode ini dilakukan dalam dua skenario, yaitu tanpa penerapan SMOTE dan Grid Search, serta dengan penerapan SMOTE dan Grid Search. Penerapan SMOTE bertujuan untuk mengatasi permasalahan ketidakseimbangan kelas pada dataset, sehingga distribusi data antar kelas menjadi lebih seimbang. Sementara itu, Grid Search digunakan untuk melakukan optimasi parameter pada model SVM guna memperoleh kombinasi parameter terbaik yang dapat meningkatkan performa klasifikasi. Optimasi parameter model dilakukan menggunakan Grid Search dengan 5-fold Cross Validation (CV = 5) untuk memperoleh parameter terbaik. Hasil optimasi menggunakan Grid Search, diperoleh kombinasi parameter terbaik yaitu C = 100**,** gamma = 0,001**,** dan kernel = rbf. Nilai parameter tersebut menunjukkan bahwa model menggunakan kernel Radial Basis Function (RBF) dengan nilai C yang tinggi untuk meningkatkan kemampuan model dalam memisahkan kelas, serta nilai gamma yang kecil untuk menghasilkan batas keputusan yang lebih optimal sehingga dapat meningkatkan kinerja model klasifikasi.

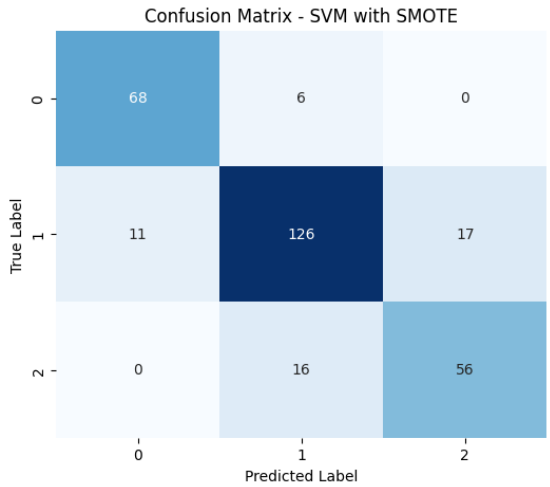
Tahapan selanjutnya adalah membandingkan performa kedua skenario metode klasifikasi berdasarkan metrik accuracy, precision, recall, dan F1-score. Metrik ini digunakan untuk menilai ketepatan model serta konsistensinya dalam mengidentifikasi setiap kelas. Hasil perbandingan kedua skenario disajikan pada Gambar 4 sampai Gambar 6.



Gambar 4 Hasil SVM Original

Gambar 4 menunjukkan confusion matrix hasil klasifikasi menggunakan metode Support Vector Machine tanpa penerapan teknik SMOTE dan Grid Search. Confusion matrix digunakan untuk menggambarkan kinerja model dalam mengklasifikasikan data berdasarkan perbandingan antara label aktual (true label) dan label prediksi (predicted label). Berdasarkan confusion matrix yang diperoleh, terlihat bahwa seluruh data dari setiap kelas diprediksi oleh model sebagai kelas 1. Pada kelas 0, sebanyak 74 data yang sebenarnya termasuk kelas 0 diprediksi sebagai kelas 1. Demikian pula pada kelas 1, sebanyak 154 data diprediksi sebagai kelas 1. Selanjutnya pada kelas 2, sebanyak 72 data yang sebenarnya termasuk kelas 2 juga diprediksi sebagai kelas 1.

Hasil tersebut menunjukkan bahwa model Support Vector Machine tanpa penerapan teknik penyeimbangan data mengalami bias terhadap kelas mayoritas, sehingga model cenderung memprediksi seluruh data ke dalam satu kelas saja. Kondisi ini umumnya disebabkan oleh ketidakseimbangan distribusi kelas (class imbalance) pada dataset yang digunakan. Dengan demikian, hasil confusion matrix ini menunjukkan bahwa model SVM tanpa penerapan teknik penyeimbangan data seperti SMOTE memiliki performa klasifikasi yang kurang optimal, terutama dalam membedakan antar kelas. Oleh karena itu, diperlukan penerapan metode penyeimbangan data dan optimasi parameter untuk meningkatkan kemampuan model dalam melakukan klasifikasi secara lebih akurat.

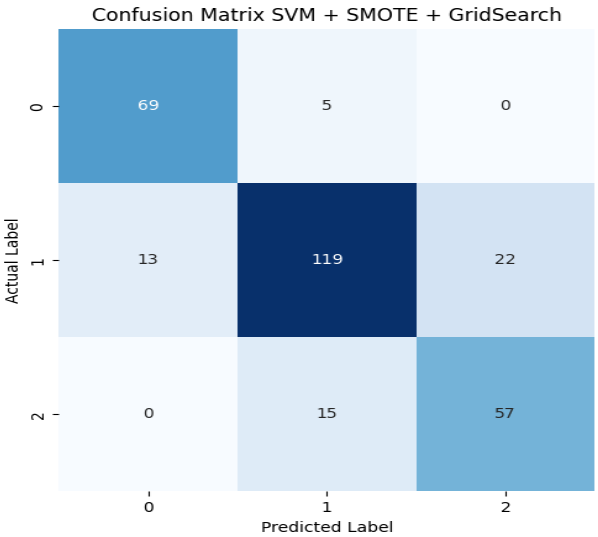


Gambar 5 Model SVM dengan SMOTE

Selanjutnya Gambar 5 menjelaskkan confusion matrix hasil klasifikasi menggunakan metode SVM dengan penerapan teknik SMOTE. Berdasarkan confusion matrix yang diperoleh, pada kelas 0 terdapat 68 data yang berhasil diklasifikasikan dengan benar sebagai kelas 0. Selain itu terdapat 6 data yang sebenarnya termasuk kelas 0 namun diprediksi sebagai kelas 1, dan tidak terdapat data yang diprediksi sebagai kelas 2.

Pada kelas 1, model berhasil mengklasifikasikan 126 data dengan benar sebagai kelas 1. Namun terdapat 11 data yang sebenarnya termasuk kelas 1 tetapi diprediksi sebagai kelas 0, serta 17 data yang diprediksi sebagai kelas 2. Selanjutnya pada kelas 2, model berhasil mengklasifikasikan 56 data dengan benar sebagai kelas 2. Selain itu terdapat 16 data yang sebenarnya termasuk kelas 2 tetapi diprediksi sebagai kelas 1, dan tidak terdapat data yang diprediksi sebagai kelas 0.

Secara umum, nilai pada diagonal utama confusion matrix yaitu 68, 126, dan 56 menunjukkan jumlah prediksi yang benar untuk masing-masing kelas. Hasil ini menunjukkan bahwa model Support Vector Machine dengan penerapan SMOTE mampu melakukan klasifikasi dengan cukup baik pada dataset yang digunakan, meskipun masih terdapat beberapa kesalahan klasifikasi terutama pada kelas 1 dan kelas 2.



Gambar 6 Hasil Model SVM, SMOTE dan Grid Search

Gambar 6 tersebut menunjukkan confusion matrix hasil klasifikasi menggunakan metode SVM dengan penerapan SMOTE dan Grid Search. Berdasarkan confusion matrix yang diperoleh, dapat diketahui bahwa pada kelas 0 terdapat sebanyak 69 data yang berhasil diklasifikasikan dengan benar sebagai kelas 0. Selain itu terdapat 5 data yang sebenarnya termasuk kelas 0 namun diprediksi sebagai kelas 1, dan tidak terdapat data yang diprediksi sebagai kelas 2.

Pada kelas 1, model berhasil mengklasifikasikan 119 data dengan benar sebagai kelas 1. Namun terdapat 13 data yang sebenarnya termasuk kelas 1 tetapi diprediksi sebagai kelas 0, serta 22 data yang diprediksi sebagai kelas 2. Selanjutnya pada kelas 2, model berhasil mengklasifikasikan 57 data dengan benar sebagai kelas 2. Selain itu terdapat 15 data yang sebenarnya termasuk kelas 2 tetapi diprediksi sebagai kelas 1, dan tidak terdapat data yang diprediksi sebagai kelas 0.

Secara umum, nilai pada diagonal utama confusion matrix menunjukkan jumlah prediksi yang benar, yaitu 69, 119, dan 57 untuk masing-masing kelas. Hal ini menunjukkan bahwa model Support Vector Machine dengan penerapan SMOTE dan Grid Search mampu melakukan klasifikasi dengan cukup baik pada dataset yang digunakan.

Berdasarkan table 4 hasil pengujian, model SVM tanpa penerapan SMOTE memperoleh nilai *accuracy* sebesar 51%, *precision* sebesar 17%, *recall* sebesar 33%, dan *F1-score* sebesar 23%. Nilai tersebut menunjukkan bahwa model memiliki performa klasifikasi yang relatif rendah. Hal ini disebabkan oleh adanya ketidakseimbangan distribusi kelas pada dataset, sehingga model cenderung bias terhadap kelas mayoritas dan kurang mampu mengenali kelas lainnya. Setelah diterapkan teknik SMOTE, performa model mengalami peningkatan yang signifikan. Model SVM + SMOTE memperoleh nilai *accuracy* sebesar 81%, *precision* sebesar 81%, *recall* sebesar 80%, dan *F1-score* sebesar 81%. Peningkatan ini menunjukkan bahwa teknik SMOTE berhasil menyeimbangkan distribusi data antar kelas sehingga model dapat mempelajari pola data dengan lebih baik.

Selanjutnya, ketika dilakukan optimasi parameter menggunakan Grid Search, performa model kembali mengalami peningkatan. Model SVM, SMOTE dan Grid Search memperoleh nilai *accuracy* sebesar 84%, *precision* sebesar 82%, *recall* sebesar 81%, dan *F1-score* sebesar 81%. Hasil ini menunjukkan bahwa optimasi parameter mampu meningkatkan kemampuan model dalam melakukan klasifikasi secara lebih optimal.

Secara keseluruhan, hasil penelitian menunjukkan bahwa kombinasi antara teknik penyeimbangan data menggunakan SMOTE dan optimasi parameter menggunakan Grid Search mampu meningkatkan kinerja model Support Vector Machine dalam melakukan klasifikasi data secara lebih akurat dan konsisten. Hasil penelitian ini sejalan dengan penelitian yang dilakukan oleh [18], [2], [19] dan [20] yang menunjukkan bahwa penerapan metode SMOTE dan Grid Search terbukti mampu meningkatkan kinerja model klasifikasi. Teknik SMOTE berperan dalam mengatasi permasalahan ketidakseimbangan kelas (*class imbalance*) dengan menambahkan sampel sintetis pada kelas minoritas, sehingga distribusi data menjadi lebih seimbang. Sementara itu, Grid Search digunakan untuk melakukan optimasi parameter pada algoritma Support Vector Machine, sehingga model dapat menemukan kombinasi parameter yang paling optimal.

Penerapan kedua metode tersebut terbukti mampu meningkatkan nilai accuracy, memperbaiki keseimbangan antara precision dan recall, serta menghasilkan nilai F1-score yang lebih tinggi. Dengan demikian, model yang dihasilkan menjadi lebih stabil dan andal dalam melakukan proses klasifikasi data dibandingkan dengan model yang tidak menggunakan teknik penyeimbangan data maupun optimasi parameter.

Table 4. Hasil Pengujian

Metode	Accuracy	Precission	Recall	F1-Score
SVM	51%	17%	33%	23%
SVM+SMOTE	81%	81%	80%	81%
SVM+SMOTE+Grid Search	84%	82%	81%	81%

4. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa, penerapan teknik SMOTE dan Grid Search mampu meningkatkan kinerja algoritma SVM pada proses klasifikasi data. Penelitian ini dilakukan untuk mengatasi permasalahan ketidakseimbangan kelas (imbalanced data) yang dapat memengaruhi kemampuan model dalam melakukan klasifikasi secara optimal.

Hasil pengujian menunjukkan bahwa model SVM tanpa penanganan imbalance data menghasilkan performa yang relatif rendah, dengan nilai accuracy 51%, precision 17%, recall 33%, dan F1-score 23%. Setelah diterapkan metode SMOTE, performa model meningkat secara signifikan menjadi accuracy 81%, precision 81%, recall 80%, dan F1-score 81%. Selanjutnya, optimasi parameter menggunakan Grid Search pada model SVM+SMOTE memberikan hasil terbaik, yaitu accuracy 84%, precision 82%, recall 81%, dan F1-score 81% dengan nilai AUC 0,92.

Dengan demikian, penelitian ini membuktikan bahwa penggunaan SMOTE efektif dalam menangani data yang tidak seimbang, sedangkan Grid Search berperan dalam mengoptimalkan parameter model agar menghasilkan performa klasifikasi yang lebih baik. Saran pengembangan penelitian selanjutnya adalah menggunakan dataset dengan jumlah data yang lebih besar dan lebih beragam agar model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik. Selain itu dapat membandingkan metode SMOTE dengan teknik penyeimbangan data lainnya, seperti ADASYN, Borderline-SMOTE, Random Oversampling, atau metode hybrid lainnya untuk mengetahui pendekatan yang paling efektif dalam menangani ketidakseimbangan data.

Ucapan Terima Kasih: Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung pelaksanaan penelitian ini, institusi universitas bumigora yang telah memberikan arahan dan fasilitas

Referensi

[1] S. Pal, "A deep analysis of chronic kidney disease for early detection using machine learning classifiers," *Int. J. Med. Eng. Inform.*, vol. 17, no. 3, pp. 279–291, 2025, https://doi.org/10.1504/IJMEI.2025.145849.

[2] D. A. Angggoro and S. S. Mukti, "Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 6, pp. 198–207, Dec. 2021, https://doi.org/10.22266/ijies2021.1231.19.

[3] D. A. Ajalkar, J. Y. Deshmukh, M. V. Shelke, S. V. Wankhade, and S. K. Patil, "Detection of chronic kidney disease based on ensemble approach with optimal feature selection using machine learning," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 14, no. 5, p. 4017, Oct. 2025, https://doi.org/10.11591/ijai.v14.i5.pp4017-4031.

[4] M. Chitra, A. K. Parveen, M. Elavarasi, J. Sangeetha, and R. Vaittilingame, "Chronic kidney disease prediction model using machine learning approach," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 12, no. 2, p. 162, Aug. 2023, https://doi.org/10.11591/ijict.v12i2.pp162-170.

[5] A. A. Sutariya and D. B. Rathod, "Early Phase, Multi Diseases Detection, Using AI & Intelligent Hybrid Supervised Machine Learning Classifier Model," *International Journal of Electronics and Communication Engineering*, vol. 11, no. 11, pp. 45–53, Nov. 2024, https://doi.org/10.14445/23488549/IJECE-V11I11P105.

[6] G. U. Nneji *et al.*, "FFS-IML: fusion-based statistical feature selection for machine learning-driven interpretability of chronic kidney disease," *International Journal of Machine Learning and Cybernetics*, vol. 16, no. 9, pp. 6215–6248, Sep. 2025, https://doi.org/10.1007/s13042-025-02621-0.

[7] B. Dhanwanth, B. Vivek, M. Abirami, S. M. Waseem, and C. Manikantaa, "Forecasting Chronic Kidney Disease Using Ensemble Machine Learning Technique," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 5s, pp. 336–344, Jun. 2023, https://doi.org/10.17762/ijritcc.v11i5s.7035.

[8] A. P. Anqui, "Classifying chronic kidney disease using selected machine learning techniques," *International Journal of Advanced And Applied Sciences*, vol. 12, no. 2, pp. 72–79, Feb. 2025, https://doi.org/10.21833/ijaas.2025.02.008.

[9] D. A. Otchere, T. O. Arbi Ganat, R. Gholami, and S. Ridha, "Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models," *J. Pet. Sci. Eng.*, vol. 200, p. 108182, May 2021, https://doi.org/10.1016/j.petrol.2020.108182.

[10] Q. Shi and H. Zhang, "Fault Diagnosis of an Autonomous Vehicle With an Improved SVM Algorithm Subject to Unbalanced Datasets," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 6248–6256, Jul. 2021, https://doi.org/10.1109/TIE.2020.2994868.

[11] C. C. Colmenares-Mejía *et al.*, "Multivariable prediction model of complications derived from diabetes mellitus using machine learning on scarce highly unbalanced data," *Int. J. Diabetes Dev. Ctries.*, vol. 44, no. 3, pp. 528–538, Sep. 2024, https://doi.org/10.1007/s13410-023-01264-7.

[12] C.-A. Tsai and Y.-J. Chang, "Efficient Selection of Gaussian Kernel SVM Parameters for Imbalanced Data," *Genes (Basel).*, vol. 14, no. 3, p. 583, Feb. 2023, https://doi.org/10.3390/genes14030583.

[13] J.-B. Wang, C.-A. Zou, and G.-H. Fu, "AWSMOTE: An SVM-Based Adaptive Weighted SMOTE for Class-Imbalance Learning," *Sci. Program.*, vol. 2021, pp. 1–18, May 2021, https://doi.org/10.1155/2021/9947621.

[14] W. Cai, M. Cai, Q. Li, and Q. Liu, "Three-way imbalanced learning based on fuzzy twin SVM," *Appl. Soft Comput.*, vol. 150, p. 111066, Jan. 2024, https://doi.org/10.1016/j.asoc.2023.111066.

[15] F. Kausar and B. Ramamurthy, "Machine Learning Based Optimal Feature Selection for Pediatric Ultrasound Kidney Images Using Binary Coati Optimization," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 6, pp. 1300–1313, Dec. 2024, https://doi.org/10.22266/ijies2024.1231.94.

[16] B. Yamini, T. Saraswathi, P. Radhakrishnan, M. Nalini, M. Shanmuganathan and Siva Subramanian. R, "Machine learning algorithms for predicting of chronic kidney disease and its significance in healthcare," *International Journal of Advanced Technology and Engineering Exploration*, vol. 11, no. 112, Mar. 2024, https://doi.org/10.19101/IJATEE.2023.10101788.

[17] A. Syarif, O. D. Riana, D. A. Shofiana, and A. Junaidi, "A Comprehensive Comparative Study of Machine Learning Methods for Chronic Kidney Disease Classification: Decision Tree, Support Vector Machine, and Naive Bayes," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023, https://doi.org/10.14569/IJACSA.2023.0141063.

[18] B. K. Swain, S. Mohapatra, M. Mishra, and R. Sharma, "A unified approach for Parkinson's disease recognition: imbalance mitigation and grid search optimized boosting with LightGBM," *Med. Biol. Eng. Comput.*, vol. 62, no. 11, pp. 3471–3491, Nov. 2024, https://doi.org/10.1007/s11517-024-03139-3.

[19] S. EL Ferouali, Z. Elamrani Abou Elassad, S. Qassimi, and A. Abdali, "From Baseline to Best Practice: An Advanced Feature Selection, Feature Resampling and Grid Search Techniques to Improve Injury Severity Prediction," *Applied Artificial Intelligence*, vol. 39, no. 1, Dec. 2025, https://doi.org/10.1080/08839514.2025.2452675.

[20] D. Santhadevi and B. Janet, "SDB-RGSO: Swarm-Based Data Balancing and Randomized Grid Search Optimization for IoT NetFlow Malware Detection with Ensemble Machine Learning Model," 2023, pp. 615–631. https://doi.org/10.1007/978-981-99-6550-2_46.