

PENERAPAN SMOTE-NCL UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT JANTUNG KORONER

Mariana Dewi¹, Trinado Hamonangan Saragih², Rudy Herteno³, Radityo Adi Nugroho⁴,
Dodon Turianto Nugrahadi⁵

^{1,2} Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat
¹marianadwspn@gmail.com, ²triando.saragih@ulm.ac.id, ³rudy.herteno@ulm.ac.id, ⁴radityo.adi@ulm.ac.id,
⁵dodonturianto@ulm.ac.id

Corresponding author: triando.saragih@ulm.ac.id

Abstrak

Penyakit jantung koroner (PJK) terjadi akibat penyumbatan atau penyempitan pada pembuluh darah jantung karena adanya endapan lemak dan kolesterol sehingga mengakibatkan suplai darah ke jantung menjadi terganggu. PJK masih merupakan masalah kesehatan yang penting dan berdampak secara sosioekonomi karena biaya obat-obatan yang cukup mahal dan lamanya waktu perawatan serta pengobatannya. Upaya pencegahan melalui deteksi dini dan pengendaliannya sangat penting untuk dilakukan. Salah satu cara untuk mendeteksi penyakit jantung koroner dengan memanfaatkan teknologi komputasi, yaitu melakukan klasifikasi menggunakan algoritma tertentu. Penelitian ini dilakukan untuk mengklasifikasikan penyakit jantung koroner menggunakan algoritma *Support Vector Machine* (SVM). Pada dataset yang digunakan terdapat masalah ketidakseimbangan data, yaitu jumlah data negatif PJK lebih banyak daripada data positif PJK. Sehingga untuk mengatasi permasalahan tersebut diterapkan metode penanganan ketidakseimbangan data menggunakan SMOTE dan SMOTE-NCL. Penelitian ini dilakukan dengan membandingkan kinerja dari klasifikasi SVM tanpa dilakukan penyeimbangan data, klasifikasi SVM dengan penyeimbangan data SMOTE, dan klasifikasi SVM dengan penyeimbangan data SMOTE-NCL. Hasil yang didapatkan dari penelitian ini adalah pada klasifikasi SVM tanpa dilakukan penyeimbangan data menghasilkan nilai akurasi sebesar 76,60%. Klasifikasi SVM dengan penyeimbangan data SMOTE menghasilkan nilai akurasi sebesar 80,85%. Dan klasifikasi SVM dengan penyeimbangan data SMOTE-NCL menghasilkan kinerja terbaik jika dibandingkan dengan model klasifikasi lain dengan nilai akurasi sebesar 85,10%.

Kata kunci : SVM, SMOTE, SMOTE-NCL

1. Pendahuluan

Penyakit jantung koroner (PJK) adalah penyakit yang disebabkan oleh adanya plak yang menumpuk di dalam pembuluh darah jantung, yang menyebabkan terhambatnya pasokan darah untuk mensuplai nutrisi dan oksigen ke otot jantung. Data WHO tahun 2019 menyebutkan ada 17,9 juta orang meninggal karena penyakit kardiovaskular dan salah satunya disebabkan oleh penyakit jantung koroner. Menurut data Riskesdas tahun 2018 menunjukkan bahwa sebesar 1,5% atau 15 dari 1.000 penduduk Indonesia menderita penyakit jantung koroner.

Jika dilihat dari penyebab kematian tertinggi di Indonesia, menurut Survei Sample Registration System tahun 2014 menunjukkan 12,9% kematian disebabkan oleh penyakit jantung koroner. Berdasarkan pemaparan tersebut pencegahan melalui deteksi dini dan pengendalian sangat penting untuk dilakukan. Data mining merupakan bidang yang dapat menjadi solusi sebagai alat deteksi dini penyakit, salah satu caranya adalah dengan

melakukan klasifikasi menggunakan algoritma tertentu (Nurmasani & Pristyanto, 2021).

Support Vector Machine (SVM) merupakan salah satu metode dalam *supervised learning* yang bisa digunakan untuk klasifikasi dan regresi. Prinsip kerja dari metode ini adalah mencari ruang pemisah yang paling optimal dari suatu dataset dalam kelas yang berbeda (Chazar & Widhiaputra, 2020). Penelitian terdahulu menyatakan bahwa algoritma klasifikasi SVM lebih unggul dengan menghasilkan nilai akurasi yang paling optimal dibandingkan algoritma *Naive Bayes* pada klasifikasi penyakit malaria (Ramadhan & Khoirunnisa, 2021)

Dalam data medis sering ditemukan permasalahan data tidak seimbang, yaitu apabila salah satu kelasnya mendominasi dibandingkan dengan kelas lainnya. Dataset yang tidak seimbang dapat menyebabkan kesalahan dalam klasifikasi, dimana data kelas minoritas sering diklasifikasikan sebagai kelas mayoritas (Sulistiyono et al., 2021). Sehingga, untuk mengatasi masalah tersebut diperlukan penyeimbangan data sebelum dilakukan klasifikasi. Salah satu pendekatan untuk menangani

data yang tidak seimbang adalah dengan pendekatan level data yaitu *oversampling*, *undersampling* dan gabungan keduanya (hybrid) (Hairani et al., 2020).

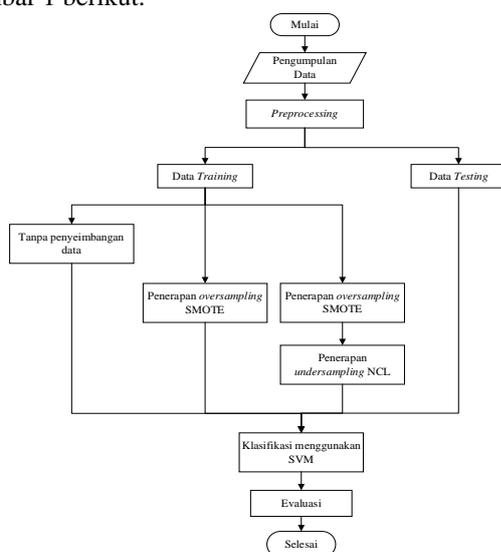
Synthetic Minority Oversampling Technique (SMOTE) merupakan salah satu metode *oversampling* untuk mensintesis dataset minoritas hingga jumlah dataset pada kelas mayoritas seimbang (Gumelar et al., 2021). Penerapan *oversampling* SMOTE mampu mengatasi ketidakseimbangan kelas pada dataset penyakit Hepatitis C serta meningkatkan nilai akurasi dan *recall* pada hasil klasifikasi (Syukron et al., 2020).

Penelitian sebelumnya menyebutkan bahwa *oversampling* SMOTE memiliki kelemahan, yaitu secara acak memilih *instance* kelas minoritas untuk di *oversampling* sehingga rentan menghasilkan data *noise*. Untuk mengatasi data *noise* yang dihasilkan oleh SMOTE dapat menggunakan metode *undersampling* (Hairani et al., 2020). Salah satu metode *undersampling* yang bisa digunakan untuk menangani ketidakseimbangan kelas adalah *Neighborhood Cleaning Rule* (NCL), metode tersebut melakukan penyeimbangan dengan cara mereduksi data berdasarkan pembersihan. Penerapan metode penyeimbangan NCL mampu meningkatkan nilai akurasi pada klasifikasi dataset pendidikan menggunakan algoritma *Decision Tree* C4.5 (Agustianto & Destarianto, 2019).

Berdasarkan uraian diatas, pada penelitian ini mengusulkan penanganan ketidakseimbangan data menggunakan SMOTE dan SMOTE-NCL pada dataset penyakit jantung koroner dan kemudian diklasifikasi menggunakan algoritma *Support Vector Machine*.

2. Metode Penelitian

Adapun alur penelitian ini dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur penelitian

2.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini adalah data *Coronary Heart Disease* yang didapat dari situs Kaggle. Memiliki jumlah data sebanyak 462 data dengan 9 fitur dan 2 kelas. Dataset ini memiliki masalah ketidakseimbangan data yaitu 160 data kelas positif PJK dan 302 data kelas negatif PJK.

2.2 Preprocessing

Preprocessing merupakan sebuah tahap awal yang harus dilakukan pada data mining. Tujuannya adalah untuk mempersiapkan data mentah sebelum dilakukan ke proses selanjutnya. *Preprocessing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem (Nasution et al., 2019). Pada tahap ini data akan diproses terlebih dahulu agar format data sesuai dengan yang diperlukan oleh algoritma klasifikasi yang digunakan. Pada dataset yang digunakan terdapat satu fitur yang memiliki tipe data kategori, yaitu fitur “famhist” yang terdiri dari absent dan present yang akan diubah menjadi data dalam bentuk numerik menggunakan metode label encoding.

2.3 Pembagian Data

Dataset yang digunakan dalam penelitian ini akan dibagi menjadi data training dan data testing dengan jumlah perbandingan 90% sebagai data training dan 10% sebagai data testing.

2.4 Penyeimbangan Data

Data tidak seimbang disebut juga *imbalanced*. Dalam data mining, *imbalanced* diartikan sebagai rasio jumlah data kelas mayoritas dengan kelas minoritas tidak sama. Kelas mayoritas memiliki jumlah data lebih banyak dibandingkan dengan data pada kelas minoritas (Saputra et al., 2021). Cara yang umum untuk mengatasi permasalahan data tidak seimbang adalah melakukan *resampling* dengan mengubah jumlah data pada tiap kelas sehingga mencapai jumlah data yang seimbang pada seluruh kelas (Indrawati et al., 2021).

Pada tahap ini data training akan dilakukan penyeimbangan data menggunakan metode *oversampling* SMOTE dan kombinasi SMOTE-NCL. Pertama, melakukan penyeimbangan menggunakan metode SMOTE, yaitu dengan cara menambahkan data replika pada kelas minoritas sehingga jumlah data menjadi seimbang dengan data kelas mayoritas. *Synthetic Minority Over-sampling Technique* (SMOTE) merupakan teknik *oversampling* dalam menyeimbangkan jumlah distribusi dataset pada kelas minoritas dengan cara mensintesis dataset minoritas hingga jumlahnya sama dengan data kelas

mayoritas dengan memanfaatkan ketetangaan terdekat (Gumelar et al., 2021).

Kedua, melakukan penyeimbangan menggunakan kombinasi SMOTE-NCL, sebelumnya akan dilakukan penyeimbangan menggunakan metode SMOTE dan didapatkan *dataset* baru, kemudian untuk meningkatkan kualitas data akan dilakukan *undersampling* data menggunakan metode NCL. *Neighborhood Cleaning Rule* (NCL) merupakan salah satu teknik *undersampling* yang dapat digunakan untuk menangani ketidakseimbangan data dengan cara mereduksi data berdasarkan pembersihan (Agustianto & Destarianto, 2019).

2.5 Klasifikasi

Klasifikasi adalah sebuah proses yang melibatkan penentuan item data ke dalam kategori atau kelas yang sudah didefinisikan terlebih dahulu, dengan tujuan untuk memprediksi kategori atau kelas yang tepat untuk seluruh data yang ada dalam dataset (Sutoyo & Fadlurrahman, 2020). *Support vector machine* (SVM) merupakan salah satu metode klasifikasi dengan prinsip mencari *hyperplane* yang memiliki margin terbesar (Wilantikasari et al., 2019). *Hyperplane* adalah suatu garis yang memisahkan data antar kelas, sedangkan margin merupakan jarak antara *hyperplane* dengan data terdekat yang berada pada masing-masing kelas (Irmada & Astritama, 2021). Pada penelitian ini akan dilakukan tiga skenario klasifikasi, yang pertama adalah klasifikasi SVM tanpa dilakukan penyeimbangan data. Kedua, klasifikasi SVM dengan penerapan penyeimbangan data SMOTE. Dan ketiga, klasifikasi SVM dengan penerapan penyeimbangan data kombinasi SMOTE-NCL.

2.6 Evaluasi

Tahap evaluasi merupakan tahapan untuk mengukur kinerja dari klasifikasi SVM tanpa dilakukan penyeimbangan data, klasifikasi SVM dengan penerapan penyeimbangan data SMOTE dan klasifikasi SVM dengan penerapan penyeimbangan data kombinasi SMOTE-NCL. Evaluasi kinerja klasifikasi dilakukan menggunakan perhitungan nilai akurasi, presisi, recall dan *f1-score* menggunakan confusion matrix. Pada Tabel 1 berikut ditampilkan confusion matrix untuk kelas biner, yaitu dataset dengan dua jenis kelas.

Tabel 1. Confusion matrix kelas biner

Class	Predictive Positive	Predictive Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

Akurasi adalah ketetapan sistem dalam melakukan proses klasifikasi dengan benar menggunakan persamaan berikut.

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

Presisi adalah rasio jumlah dokumen yang relevan dengan total jumlah dokumen yang ditemukan pada sistem klasifikasi menggunakan persamaan berikut.

$$\text{Presisi} = \frac{TP}{TP+FP} \tag{4}$$

Recall adalah rasio jumlah dokumen yang ditemukan kembali oleh sistem klasifikasi dengan total jumlah dokumen yang relevan menggunakan persamaan berikut.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

F1-score menggabungkan *recall* dan presisi sehingga menghasilkan metrik yang efektif untuk mencari kembali informasi dalam himpunan yang mengandung ketidakseimbangan. Untuk menghitung nilai *f1-score* dengan menggunakan persamaan berikut. (Kasanah et al., 2019).

$$F1 - score = \frac{2x(\text{presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \tag{6}$$

3. Hasil dan Pembahasan

3.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah data Coronary Heart Disease yang diperoleh dari situs Kaggle. Memiliki total data sebanyak 462 data dan 9 fitur. Dataset ini memiliki permasalahan ketidakseimbangan data. Pada kelas 0 (negatif PJK) terdiri dari 302 data dan kelas 1 (positif PJK) sebanyak 160 data. Adapun deskripsi *dataset* dapat dilihat pada Tabel 2 berikut.

Tabel 2. Deskripsi dataset

No	Fitur	Deskripsi
1	sbp (systolic blood pressure)	Tekanan darah sistolik
2	Tobacco	Akumulasi tembakau dalam tubuh
3	ldl (low density lipoprotein)	Kolesterol jahat
4	Adiposity	Persentase lemak tubuh
5	famhist (family history)	Riwayat keluarga penyakit jantung
6	typea (type-A behavior)	Karakteristik seseorang yang kompetitif, tidak sabar dan marah
7	Obesity	Indeks massa tubuh
8	Alcohol	Konsumsi alkohol
9	Age	Usia
10	Chd	Variabel kelas (0 dan 1)

3.2 Preprocessing

Sebelum pembagian data training dan data testing, pada dataset dilakukan preprocessing terlebih dahulu. Dalam tahapan ini fitur yang mengandung data kategori akan diubah menjadi dalam bentuk numerik menggunakan metode label encoding. Pada dataset terdapat satu fitur yang memiliki data kategori yaitu fitur "famhist", terdiri dari kategori absent atau tidak memiliki riwayat keluarga mengalami PJK dan present atau memiliki riwayat keluarga mengalami PJK. Present dilabelkan menjadi 1 dan absent menjadi 0.

3.3 Pembagian Data

Pada tahap ini, data dibagi menjadi dua bagian yaitu data training dan data testing. Pembagian data dilakukan menggunakan teknik stratified untuk memastikan data didistribusikan dengan benar pada setiap kelasnya berdasarkan proporsi data training dan data testing. Proporsi data yang digunakan adalah 90% data training dan 10% data testing.

3.4 Penyeimbangan Data

a. Penyeimbangan Data menggunakan SMOTE

Pada data training, akan dilakukan oversampling data menggunakan SMOTE untuk membuat data replika diantara data yang dipilih secara acak dengan data terdekatnya. Penggunaan metode SMOTE dilakukan untuk memperbanyak data pada kelas 1 (positif PJK) agar jumlahnya setara dengan kelas 0 (negatif PJK). Perbandingan data sebelum dan setelah dilakukan oversampling SMOTE pada Tabel 3 berikut.

Tabel 3. Perbandingan data sebelum dan setelah penerapan SMOTE

Kelas	Data	
	Sebelum oversampling	Sesudah oversampling
Kelas 0	271	271
Kelas 1	144	271
Jumlah	415	542

b. Penyeimbangan Data menggunakan SMOTE- NCL

Pada data training akan dilakukan oversampling-undersampling menggunakan SMOTE-NCL. Tahapan pertama adalah melakukan oversampling SMOTE dengan langkah pertama melakukan perhitungan jarak terdekat data pada setiap kelas minoritas menggunakan euclidean distance, kemudian membuat data replika dan didapatkan dataset baru. Selanjutnya, dataset hasil SMOTE akan dilakukan undersampling NCL untuk meningkatkan kualitas data, dengan cara mereduksi data berbasis

pembersihan. Langkah pertama pada undersampling NCL adalah memilih data observasi, kemudian menghitung jarak terdekat data dari seluruh data menggunakan euclidean distance. Setelah ditentukan tetangga terdekatnya, maka data observasi yang dipilih sebelumnya akan dicek label kelasnya apakah sama dengan label kelas data tetangga terdekatnya, jika label kelas data observasi berbeda dengan label tetangganya, maka data observasi dianggap sebagai noise, dan akan direduksi. Perbandingan data sebelum dan setelah dilakukan SMOTE-NCL dapat dilihat pada Tabel 4 berikut.

Tabel 4. Perbandingan data sebelum dan setelah penerapan SMOTE-NCL

Kelas	Data	
	Sebelum oversampling-undersampling	Sesudah oversampling-undersampling
Kelas 0	271	271
Kelas 1	144	223
Jumlah	415	486

3.5 Klasifikasi

Klasifikasi yang digunakan pada penelitian ini menggunakan metode klasifikasi Support Vector Machine dengan trik kernel Polynomial. Klasifikasi pada penelitian ini terbagi menjadi tiga skenario yaitu klasifikasi SVM tanpa dilakukan penyeimbangan data, klasifikasi SVM dengan penyeimbangan data menggunakan SMOTE, dan klasifikasi SVM dengan penyeimbangan data menggunakan SMOTE-NCL.

a. Klasifikasi SVM tanpa Penyeimbangan Data

Untuk mendapatkan hasil kinerja klasifikasi terbaik, maka dilakukan pengujian parameter. Parameter uji pada klasifikasi SVM tanpa penyeimbangan data menggunakan parameter C dan degree dengan rentang nilai parameter uji yang dapat dilihat pada Tabel 5 berikut.

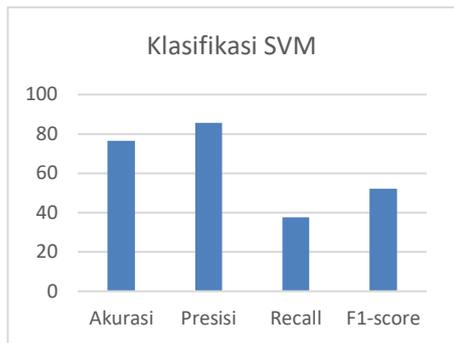
Tabel 5. Nilai parameter uji klasifikasi SVM

Parameter	Nilai
C	0,1, 0,5, 1, 1,5
Degree	1, 2, 3, 4, 5

Setelah dilakukan pengujian nilai parameter C dan degree diatas, maka didapatkan hasil performa klasifikasi SVM tanpa penyeimbangan data yang dapat dilihat pada Tabel 6 dan Gambar 2 berikut.

Tabel 6. Hasil performa klasifikasi SVM

Akurasi	Presisi	Recall	F1-score
76,60	85,71	37,5	52,17



Gambar 2. Grafik performa klasifikasi SVM

b. Klasifikasi SVM dengan Penyeimbangan Data SMOTE

Pengujian parameter yang dilakukan pada klasifikasi SVM dengan penyeimbangan data menggunakan SMOTE dapat dilihat pada Tabel 7.

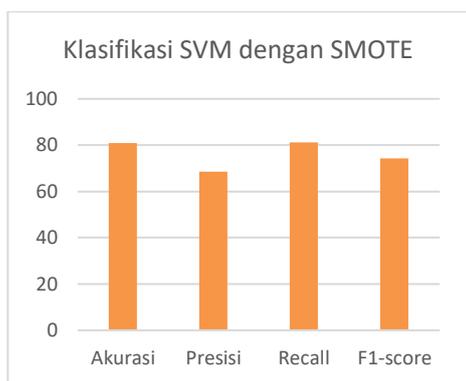
Tabel 7. Nilai parameter uji klasifikasi SVM dengan SMOTE

Parameter	Nilai
C	0,1, 0,5, 1, 1,5
Degree	1, 2, 3, 4, 5
K	3, 5, 7, 9, 11

Setelah dilakukan pengujian nilai parameter C, degree dan k didapatkan performa klasifikasi SVM dengan SMOTE yang dapat dilihat pada Tabel 8 dan Gambar 3 berikut.

Tabel 8. Hasil performa klasifikasi SVM dengan SMOTE

Akurasi	Presisi	Recall	F1-score
80,85	68,42	81,25	74,29



Gambar 3. Grafik performa klasifikasi SVM dengan SMOTE

c. Klasifikasi SVM dengan Penyeimbangan Data SMOTE-NCL

Pengujian parameter yang dilakukan pada klasifikasi SVM dengan penyeimbangan data SMOTE-NCL dapat dilihat pada Tabel 9 berikut.

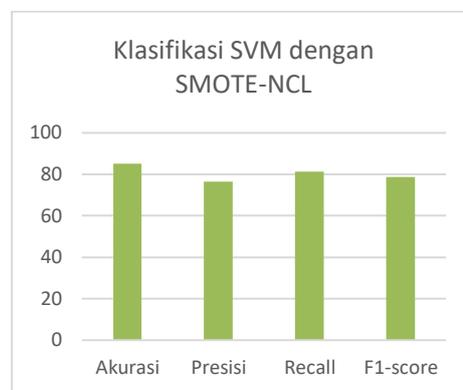
Tabel 9. Nilai uji parameter klasifikasi SVM dengan SMOTE-NCL

Parameter	Nilai
C	0,1, 0,5, 1, 1,5
degree	1, 2, 3, 4, 5
K	3, 5, 7, 9, 11
n_neighbors	3, 5, 7

Setelah dilakukan pengujian nilai parameter C, degree, k dan n_neighbors didapatkan performa klasifikasi SVM dengan penyeimbangan data menggunakan SMOTE-NCL yang dapat dilihat pada Tabel 10 dan Gambar 4 berikut.

Tabel 10. Hasil performa klasifikasi SVM dengan SMOTE-NCL

Akurasi	Presisi	Recall	F1-score
85,10	76,47	81,25	78,78



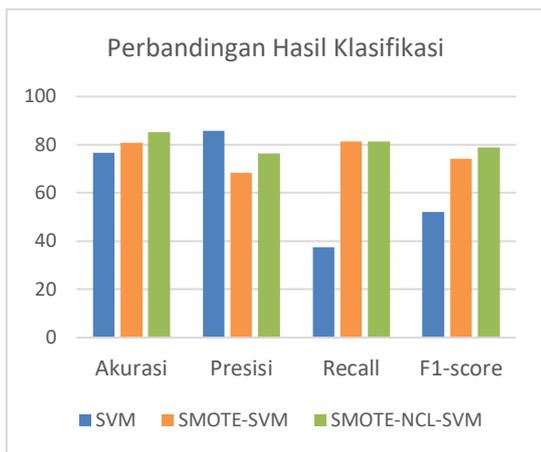
Gambar 4. Grafik performa klasifikasi SVM dengan SMOTE-NCL

3.6 Pembahasan

Terdapat 3 skenario yang akan dilakukan pada penelitian ini, pertama adalah klasifikasi SVM tanpa dilakukan penyeimbangan data, kedua adalah klasifikasi SVM dengan penerapan SMOTE dan ketiga adalah klasifikasi SVM dengan penerapan SMOTE-NCL. Kemudian, untuk mengetahui kinerja klasifikasi maka dievaluasi menggunakan confusion matrix. Perbandingan hasil kinerja dari ketiga skenario klasifikasi yang telah dilakukan dapat dilihat pada Tabel 11 dan Gambar 5 berikut.

Tabel 11. Perbandingan hasil klasifikasi

	Akurasi	Presisi	Recall	F1-score
SVM	76,60	85,71	37,5	52,17
SMOTE-SVM	80,85	68,42	81,25	74,29
SMOTE-NCL-SVM	85,10	76,47	81,25	78,79



Gambar 5. Grafik perbandingan hasil klasifikasi

Tabel 11 menunjukkan bahwa nilai akurasi, recall dan f1-score mengalami peningkatan pada klasifikasi SVM dengan SMOTE dan klasifikasi SVM dengan SMOTE-NCL jika dibandingkan dengan klasifikasi SVM tanpa penyeimbangan data. Nilai presisi yang tinggi pada klasifikasi SVM tanpa penyeimbangan data disebabkan karena data belum seimbang yang mengakibatkan model cenderung fokus pada kelas mayoritas (negatif PJK) sehingga menganggap kelas minoritas (positif PJK) sebagai kelas mayoritas. Dimana dari hasil confusion matrix pada pada klasifikasi SVM tanpa dilakukan penyeimbangan data menghasilkan data FN (false negative) yang lebih tinggi yaitu 10 data dan FP (false positive) yang sedikit yaitu 1 data.

Penerapan penyeimbangan SMOTE dan SMOTE-NCL memiliki pengaruh yang baik pada kinerja klasifikasi. Berdasarkan confusion matrix pada klasifikasi SVM menggunakan SMOTE dan klasifikasi SVM menggunakan SMOTE-NCL menunjukkan bahwa model klasifikasi mampu melakukan prediksi dengan jumlah ketepatan prediksi lebih banyak dan kesalahan prediksi lebih sedikit dibandingkan dengan klasifikasi pada data yang tidak diseimbangkan. Penyeimbangan data menggunakan SMOTE-NCL menghasilkan kinerja klasifikasi yang lebih baik dibandingkan dengan SMOTE karena data yang dihasilkan SMOTE rentan terhadap data noise. Penerapan NCL setelah SMOTE dapat membersihkan data noise yang dihasilkan SMOTE. Dengan demikian, NCL mampu membantu memperbaiki kualitas dataset dan meningkatkan kemampuan model klasifikasi untuk mengenali dan memprediksi data kelas minoritas (positif PJK) dengan lebih baik.

Pada klasifikasi SVM dapat memprediksi data dengan benar untuk kelas positif PJK sebanyak 6 data dan kelas negatif PJK sebanyak 30 data. Sedangkan klasifikasi SVM dengan SMOTE dapat memprediksi data dengan benar untuk kelas positif PJK sebanyak 13 data dan kelas negatif PJK sebanyak 25 data, dimana jumlah kesalahan dalam memprediksi data berkurang daripada saat klasifikasi SVM. Dan klasifikasi SVM dengan SMOTE-NCL dapat

memprediksi data dengan benar untuk kelas positif PJK sebanyak 13 data dan kelas negatif PJK sebanyak 27 data, dimana jumlah kesalahan dalam memprediksi data semakin berkurang daripada saat klasifikasi SVM maupun klasifikasi SVM dengan SMOTE.

4. Kesimpulan dan Saran

Kesimpulan dari penelitian ini adalah penerapan penyeimbangan data memiliki pengaruh yang baik terhadap kinerja klasifikasi. Klasifikasi SVM dengan penerapan penyeimbangan data menggunakan SMOTE-NCL menghasilkan kinerja terbaik dibanding klasifikasi SVM dengan penyeimbangan SMOTE dan tanpa dilakukan penyeimbangan data dengan nilai akurasi terbaik sebesar 85,10%.

Adapun saran untuk penelitian selanjutnya adalah dapat menggunakan metode penyeimbangan data atau melakukan kombinasi metode yang berbeda, serta menggunakan algoritma klasifikasi lain sehingga dapat diketahui perbandingan hasil kerjanya

Daftar Pustaka:

Agustianto, K., & Destarianto, P. (2019). Imbalance Data Handling using Neighborhood Cleaning Rule (NCL) Sampling Method for Precision Student Modeling. *Proceedings - 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering, ICOMITEE 2019, 1*, 86–89. <https://doi.org/10.1109/ICOMITEE.2019.8921159>

Chazar, C., & Widhiaputra, B. E. (2020). Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine. *INFORMASI (Jurnal Informatika Dan Sistem Informasi), Volume 12*, 67–80.

Gumelar, G., Norlaila, Ain, Q., Marsuciati, R., Bambang, S. A., Sunyoto, A., & Mustafa, M. S. (2021). *Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance*. 250–255.

Hairani, Saputro, K. E., & Fadli, S. (2020). K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi Dan Sistem Komputer*, 8(2), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>

Indrawati, A., Ilmu, L., Indonesia, P., & Diabetes, P. I. (2021). Penerapan Teknik Kombinasi Oversampling Dan Undersampling Untuk Mengatasi Permasalahan Imbalanced Dataset. *JIKO (Jurnal Informatika Dan Komputer)*, 4(1), 38–43. <https://doi.org/10.33387/jiko>

- Irmanda, H. N., & Astritama, R. (2021). Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM). *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)(Rekayasa Sistem Dan Teknologi Informasi)*, 4, 915–922.
- Kasanah, A. N., Muladi, & Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *JURNAL RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3, 196–201.
- Nasution, D. A., Khatimah, H. H., & Chamidah, N. (2019). *Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-Nn*. 4(1), 78–82.
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking untuk Klasifikasi Penyakit Jantung pada Dataset Imbalanced Class. *Jurnal Pseudocode*, VIII, 21–26.
- Ramadhan, N. G., & Khoirunnisa, A. (2021). Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 5(4), 1580. <https://doi.org/10.30865/mib.v5i4.3347>
- Saputra, P. Y., Malang, P. N., Abdullah, M. Z., Malang, P. N., Kirana, A. P., & Malang, P. N. (2021). Improvisasi Teknik Oversampling MWMOTE Untuk Penanganan Data Tidak Improvisasi Teknik Oversampling MWMOTE Untuk Penanganan Data Tidak Seimbang. *Jurnal Media Informatika Budidarma*, 5(April). <https://doi.org/10.30865/mib.v5i2.2811>
- Sulistiyono, M., Pristiyanto, Y., Adi, S., & Gumelar, G. (2021). Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi. *SISTEMASI: Jurnal Sistem Informasi*, Volume 10, 445–459.
- Sutoyo, E., & Fadlurrahman, M. A. (2020). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(3), 379. <https://doi.org/10.26418/jp.v6i3.42896>
- Syukron, M., Santoso, R., & Widiharah, T. (2020). Perbandingan Metode SMOYE Random Forest dan Smote XGboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C pada Imbalance Class Data. *Jurnal Gaussian*, 9, 227–236.
- Wilantikasari, Y., Cholissodin, I., & Santoso, E. (2019). *Klasifikasi Penyakit Kulit Kucing menggunakan Metode Support Vector Machine*. 3(5), 4500–4507.

