

IMPLEMENTASI TEKNIK *RESAMPLING* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA TERHADAP KLASIFIKASI ANEMIA MENGGUNAKAN *SUPPORT VECTOR MACHINE*

Nadia Wulandari^{1*}, Badieah²

Universitas Islam Sultan Agung^{1,2}

nadiawulandari@std.unissula.ac.id^{1*} badieah.assegaf@unissula.ac.id²

Received: 15-01-2025

Revised: 29-01-2025

Approved: 11-02-2025

ABSTRAK

Penelitian ini bertujuan untuk mengatasi permasalahan ketidakseimbangan data dalam klasifikasi anemia dengan menerapkan teknik *Synthetic Minority Over-sampling Technique (SMOTE)* dan algoritma *Support Vector Machine (SVM)*. Anemia merupakan kondisi medis yang terjadi akibat rendahnya kadar hemoglobin dalam darah, sehingga menghambat distribusi oksigen ke seluruh tubuh. Dalam penelitian ini, data yang digunakan berasal dari dataset "*Anemia Type Classification*" yang tersedia di Kaggle, dengan total 1281 sampel, 14 fitur, dan 9 kelas klasifikasi anemia. Metode penelitian melibatkan beberapa tahapan, yaitu *preprocessing data*, penanganan data yang tidak seimbang, klasifikasi menggunakan *SVM* dengan pendekatan *One vs One (OvO)* dan *One vs Rest (OvR)*, serta evaluasi performa model. Teknik *SMOTE* diterapkan untuk meningkatkan jumlah sampel pada kelas minoritas, sehingga mengurangi bias model terhadap kelas mayoritas. Model diuji dalam tiga skenario: tanpa *resampling*, *oversampling* dengan *SMOTE*, dan *undersampling*. Hasil penelitian menunjukkan bahwa penerapan teknik *SMOTE* dengan pendekatan *OvO* pada algoritma *SVM* menghasilkan performa terbaik dengan akurasi sebesar 85%, *precision* 87%, *recall* 90%, dan *F1-score* 87%. Sementara itu, pendekatan *OvR* memiliki performa yang lebih rendah dibandingkan *OvO*. Pengujian tanpa *resampling* menghasilkan akurasi yang cukup baik, namun masih menunjukkan bias terhadap kelas mayoritas. Teknik *undersampling* menghasilkan performa yang lebih rendah dibandingkan metode lainnya. Simpulan, bahwa penerapan *SMOTE* dengan algoritma *SVM* dan pendekatan *OvO* dapat meningkatkan akurasi klasifikasi anemia pada dataset yang tidak seimbang. Temuan ini dapat menjadi dasar bagi pengembangan sistem prediksi anemia yang lebih akurat dan efektif dalam mendukung diagnosis medis.

Kata Kunci : Anemia, Klasifikasi Multiclass, *Support Vector Machine*, *SMOTE*, *One Vs One*, *One Vs Rest*

PENDAHULUAN

Anemia merupakan kondisi di mana jumlah sel darah merah dalam tubuh berada di bawah normal. Sel darah merah mengandung hemoglobin, yang berperan dalam mengangkut oksigen ke seluruh jaringan tubuh (Isnaini et al., 2021). Anemia memiliki beberapa jenis contohnya seperti anemia hipokromik normositik, anemia normokromik normositik, anemia defisiensi besi, trombositopenia, anemia makrositik, dan sebagainya. Oleh karena itu, identifikasi dan klasifikasi anemia berdasarkan data kesehatan pasien sangat penting agar dapat memberikan intervensi medis yang tepat waktu.

Salah satu tantangan dalam mengklasifikasikan penyakit anemia adalah adanya ketidakseimbangan data (*imbalanced data*), di mana jumlah data pada satu kelas jauh lebih kecil dibandingkan kelas lainnya. Ketidakseimbangan ini dapat berdampak pada kinerja algoritma machine learning, karena model cenderung memprediksi kelas mayoritas dan kurang memperhatikan kelas minoritas, meskipun kelas minoritas justru krusial untuk dideteksi dengan akurat. Untuk mengatasi masalah tersebut, dapat menggunakan teknik *resampling*, yaitu *oversampling* dan *undersampling*. Teknik *Synthetic Minority Over-sampling Technique (SMOTE)* adalah salah satu teknik yang dapat diterapkan untuk melakukan *oversampling*. Pada penelitian yang dilakukan oleh (Syukron et al., 2020) memanfaatkan teknik *SMOTE* menunjukkan bahwa penerapan

algoritma SMOTE mampu memperbaiki model dan meningkatkan akurasi dalam memprediksi seluruh kelas. Tingkat akurasi yang lebih tinggi dihasilkan oleh SMOTE Random Forest dan menunjukkan kelas sirosis lebih baik pada performa *recall*. Riset yang dilaksanakan oleh (Dharmendra et al., 2024) membuktikan jika SMOTE lebih optimal dalam memperbaiki keandalan model. Pemanfaatan SMOTE berhasil memperbaiki *nilai akurasi, presisi, recall*, dan *f1-score* pada seluruh model. SMOTE memberi manfaat dalam memperkuat keandalan model dan menandakan kemampuan yang lebih besar dibandingkan dengan teknik resampling lainnya juga dibuktikan oleh (Undersampling, 2024).

Tantangan lain yang muncul dalam proses klasifikasi anemia yaitu data yang memiliki lebih dari dua kelas. Permasalahan klasifikasi ini dikenal sebagai klasifikasi *multiclass* sehingga perlu adanya algoritma untuk proses pengklasifikasian data salah satunya adalah *Support Vector Machine (SVM)*. Klasifikasi data dapat diselesaikan menggunakan algoritma SVM, sehingga diharapkan kombinasi dari teknik SMOTE dan algoritma SVM dapat meningkatkan tingkat akurasi dalam klasifikasi anemia (Gumelar et al., 2021). Salah satu keunggulan algoritma SVM adalah kemampuannya bekerja secara efektif dalam data berukuran tinggi dengan menggunakan kernel. Sebelumnya, sudah ada penelitian yang menerapkan algoritma SVM pada klasifikasi penyakit seperti anemia. Pada penelitian yang dilakukan oleh (Yanti et al., 2022), perhitungan nilai akurasi, presisi, dan recall dalam klasifikasi anemia dilakukan dengan menganalisis confusion matrix, menggunakan metode training untuk membagi data latih dan data uji. Kajian yang menerapkan algoritma SVM juga dilakukan oleh (Sopiatul Ulum et al., 2023), terbukti unggul dalam mengklasifikasikan kelayakan air minum dengan akurasi sebesar 69,764%. Klasifikasi penyakit menggunakan SVM pada kajian yang dilaksanakan oleh (Pradana et al., 2022) menyatakan bahwa algoritma SVM kompeten dalam mengklasifikasikan penyakit kardiovaskular dengan akurasi terbaik sebesar 87%.

Klasifikasi menggunakan SVM *multiclass* dapat dioptimasi dengan menggunakan pendekatan *One vs One (OvO)* dan *One vs Rest (OvR)*. Penelitian mengenai klasifikasi menggunakan algoritma SVM juga dilakukan oleh (Alita et al., 2020). Pada penelitian tersebut SVM OvO memberikan performa terbaik untuk metrik presisi, recall, dan *f1-score*. Di sisi lain, untuk akurasi SVM OvR memperoleh nilai lebih tinggi dengan perbedaan 0,06 dibandingkan OvO. Pada penelitian yang dilakukan oleh (Tantika & Kudus, 2022) membandingkan model SVM *multiclass One Against One* dengan menggunakan kernel linear dan model *One Against Rest* dalam menangani permasalahan klasifikasi *multiclass*. Dalam penelitian tersebut membuktikan bahwa model SVM *One Against One* memperoleh nilai akurasi lebih tinggi dibandingkan dengan *One Against Rest*. Nilai akurasi yang dihasilkan oleh *One Against One* yaitu sebesar 99,53% sedangkan nilai akurasi yang dihasilkan oleh *One Against Rest* yaitu sebesar 87,69%. Penelitian lain yang dilakukan oleh (Rabbani et al., 2023) menunjukkan bahwa metode *One Against One* yang dipadukan dengan kernel polinomial menghasilkan presisi tertinggi sebesar 97%.

Berdasarkan tantangan yang telah disebutkan di atas, penelitian ini akan bereksperimen menyelesaikan persoalan data yang tidak seimbang dengan memanfaatkan SMOTE. Di samping itu, dalam menyelesaikan tantangan pada kasus *multiclass classification*, penelitian ini akan bereksperimen dengan memanfaatkan pendekatan *One vs One (OvO)* dan *One vs Rest (OvR)*. Kemudian hasil akan dievaluasi menggunakan akurasi, *precision, recall*, dan *f1-score*, serta membandingkan seluruh

teknik dengan *baseline* model.

KAJIAN TEORI

Anemia

Anemia merupakan kondisi medis yang umum dijumpai di masyarakat, selain berbagai masalah utama lainnya yang berdampak signifikan pada kesehatan (Sulardi & Witanti, 2020). Anemia terjadi ketika tubuh kekurangan sel darah merah dalam jumlah yang normal, yang dapat dipicu oleh kadar hemoglobin yang rendah. Hemoglobin, yang merupakan protein kaya zat besi, memiliki peran krusial dalam produksi sel darah merah dan mendukung kemampuan darah untuk membawa oksigen ke organ dan jaringan tubuh. Secara klinis, anemia dapat diidentifikasi melalui penurunan kadar hemoglobin dan hematokrit (Marselina, 2022).

Imbalanced Data

Data yang tidak terdistribusi secara seimbang merupakan salah satu tantangan yang sering muncul dalam klasifikasi, termasuk klasifikasi penyakit anemia. Permasalahan ini muncul ketika jumlah sampel pada salah satu kelas tidak sebanyak kelas lain. Ketidakseimbangan tersebut berakibat pada model pengklasifikasian mengarah pada kelas paling banyak dan tidak valid (Aryanti et al., 2023). Ketidakseimbangan data dapat diatasi menggunakan teknik *resampling*, salah satunya adalah teknik SMOTE.

Teknik *Resampling*

Teknik *resampling* adalah teknik yang dapat mengatasi masalah data yang tidak seimbang. Terdapat tiga pendekatan utama, antarlain meningkatkan jumlah sampel pada kelas minoritas, mengurangi jumlah sampel pada kelas mayoritas, atau menggunakan metode yang menggabungkan keduanya. Tujuan dari pendekatan ini adalah untuk menyesuaikan distribusi kelas minoritas agar kelas tersebut tidak dianggap lagi sebagai kelas minoritas (Syaripudin & Walad, 2024). Salah satu teknik untuk meningkatkan jumlah data pada kelas minoritas yaitu SMOTE.

Teknik *Synthetic Minority Over-sampling Technique (SMOTE)*

SMOTE ialah satu diantara strategi yang digunakan untuk memperbaiki distribusi data. Teknik *oversampling* pada kelas minoritas ini dapat membantu meningkatkan performa klasifikasi secara keseluruhan (Nurdian et al., 2022). SMOTE berfungsi menyeimbangkan distribusi data pada kelas mayoritas. SMOTE juga dapat mencegah terjadinya *overfitting* yang disebabkan oleh data ganda pada kelas minoritas, sehingga memungkinkan pemakaian data pelatihan serupa secara lebih efektif (Kurniadi et al., 2022).

Klasifikasi *Multiclass*

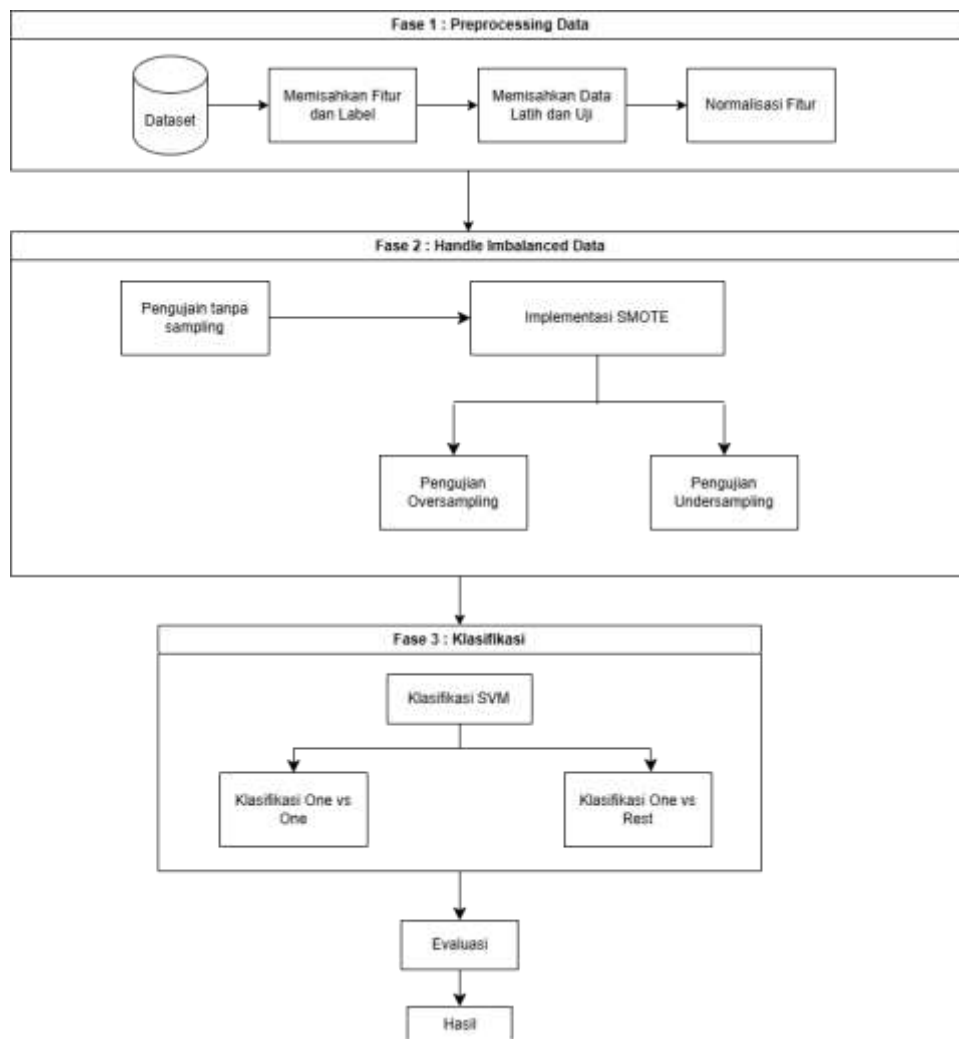
Multiclass Classification (Klasifikasi *Multiclass*) adalah metode klasifikasi yang melibatkan lebih dari dua kelas, di mana setiap sampel hanya dapat diberi label pada salah satu kelas (Lunak & Informatika, 2024). Klasifikasi *multiclass* membagi data ke dalam lebih dari dua kelompok yang saling eksklusif, di mana setiap sampel hanya dapat memiliki satu kelas. Agar algoritma SVM dapat mengklasifikasikan lebih dari dua kelas, digunakan pendekatan *One vs One (OvO)* dan *One vs Rest (OvR)*.

Support Vector Machine (SVM)

Algoritma SVM digunakan untuk membangun sistem prediksi dalam klasifikasi. Konsep inti SVM yakni penggunaan pengklasifikasi linier, di mana klasifikasi dilakukan dengan pemisahan linier. Namun, SVM d untuk menanganirancang untuk mengatasi masalah nonlinier dengan menerapkan konsep kernel ke ruang berdimensi tinggi (Pratama et al., 2020). Keunggulan utama dari algoritma SVM adalah sangat efektif untuk klasifikasi karena tidak terpengaruh oleh banyaknya atribut dan mampu memecahkan masalah dengan skala yang tinggi. SVM juga dapat melakukan pelatihan dengan cepat dan mampu mengatasi tantangan dalam menghadapi data yang ambigu (Septhya et al., 2023).

METODE PENELITIAN

Penelitian ini, menggunakan metode SVM, dengan penerapan teknik SMOTE dalam menyelesaikan kendala data yang tidak seimbang dalam klasifikasi anemia. Untuk klasifikasi SVM *multiclass*, penelitian ini menerapkan pendekatan OvO dan OvR. Gambar 1 di bawah ini menggambarkan proses yang diterapkan dalam penelitian ini.



Gambar 1. Alur Perancangan Sistem

1. Fase 1: *Preprocessing Data*

Pada fase 1 merupakan tahap *preprocessing* data, di mana dataset perlu dilakukan *preprocessing* data agar data menjadi lebih terstruktur untuk melatih model. Berikut penjelasan dari setiap tahapan dalam *preprocessing data* yang dilakukan:

- Pengumpulan Data

Data yang diterapkan pada penelitian ini merupakan data klasifikasi anemia. Data tersebut merupakan data sekunder yang didapatkan dari Kaggle dengan nama Anemia Type Classification.

- Memisahkan Fitur dan Label

Pada tahap ini, data dibagi menjadi dua bagian, yaitu X dan Y, di mana X berisi kolom fitur yang digunakan untuk proses prediksi dan klasifikasi, yaitu kolom WBC (*White Blood Cell*), LYMp (*Lymphocytes Percentage*), NEUTp (*Neutrophils Percentage*), LYMn (*Lymphocytes Number*), NEUTn (*Neutrophils Number*), RBC (*Red Blood Cell*), HGB (*Hemoglobin*), HCT (*Hematocrit*), MCV (*Mean Corpuscular Volume*), MCH (*Mean Corpuscular Hemoglobin*), MCHC (*Mean Corpuscular Hemoglobin Concentration*), PLT (*Platelet*), PDW (*Platelet Distribution Width*), dan PCT (*Plateletcrit*). Sedangkan data Y merupakan kolom target yaitu variabel yang ingin diprediksi yaitu kolom Diagnosis.

- Memisahkan Data Latih dan Uji

Pada tahap ini, data dibagi menjadi dua bagian, yaitu data training dan data testing. Data training diperuntukkan melatih model SVM, sementara itu, data testing diperuntukkan pada evaluasi efektivitas model setelah proses pelatihan. Data yang digunakan untuk data training sebanyak 1024 data dan data testing sebanyak 257 data.

- Normalisasi Fitur

Normalisasi data digunakan untuk memastikan seluruh fitur agar memiliki skala yang sama dalam rentang tertentu dan memastikan bahwa jarak antar titik data tidak dipengaruhi oleh fitur dengan skala yang besar. Normalisasi fitur ini bertujuan untuk menskalakan fitur agar berada dalam rentang nilai tertentu, misalnya antara 0 dan 1.

2. Fase 2: *Handle Imbalanced Data*

Pada fase 2 ini merupakan tahapan menguji data tanpa sampling dan untuk imbalanced data menggunakan teknik SMOTE. Setelah data diseimbangkan, data akan diuji dengan 3 pengujian yaitu pengujian tanpa sampling, pengujian *oversampling*, dan pengujian *undersampling*.

3. Fase 3: Klasifikasi

Pada fase 3 akan dilakukan klasifikasi dengan implementasi algoritma SVM dengan pendekatan OvO dan OvR karena data yang digunakan merupakan *multiclass* sehingga perlu adanya pendekatan agar algoritma SVM dapat diterapkan. Pada proses pengklasifikasian ini, akurasi OvO akan dibandingkan dengan OvR.

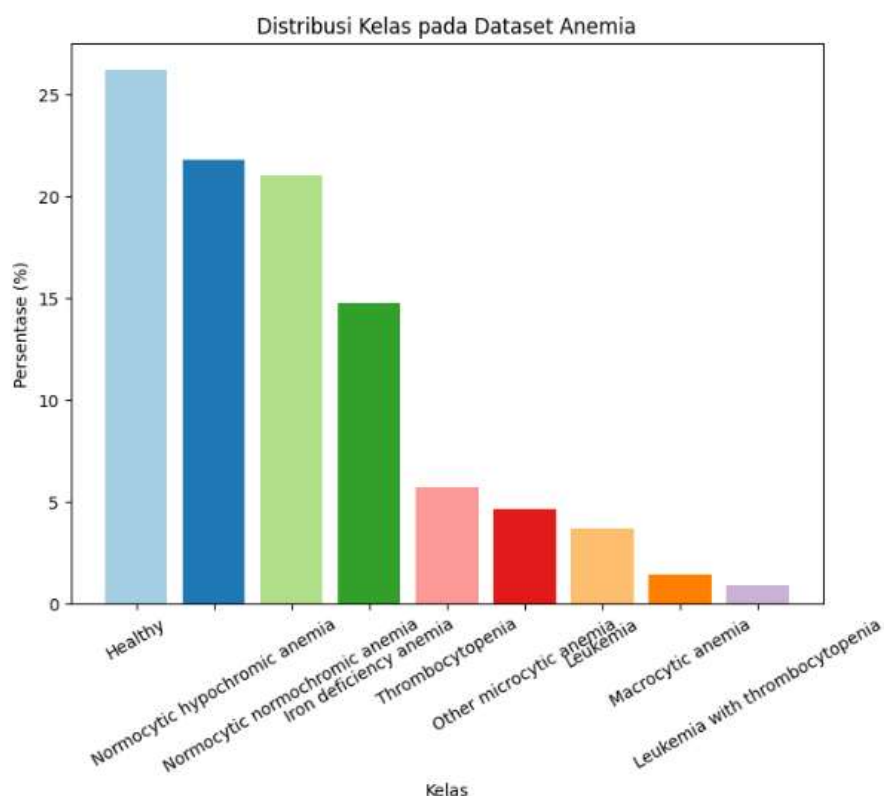
4. Evaluasi

Pada tahapan evaluasi akan dibandingkan hasil performa dari pengujian tanpa *sampling*, pengujian *oversampling*, dan pengujian *undersampling* yang dikombinasikan dengan pendekatan OvO dan OvR untuk mendapatkan hasil terbaik. Selain akurasi, untuk mengevaluasi hasil pengujian juga

- menggunakan *precision*, *recall*, serta *f1-score*.
5. Hasil
Setelah melakukan tahapan evaluasi dari beberapa pengujian, tahap selanjutnya yaitu menentukan hasil. Hasil dari penelitian ini yaitu mengetahui model yang terbaik dari pengujian-pengujian tersebut untuk melakukan klasifikasi anemia multiclass dengan menggunakan teknik SMOTE.

HASIL PENELITIAN DAN PEMBAHASAN

Dalam penelitian ini, algoritma yang diterapkan adalah SVM dengan pendekatan OvO dan OvR. Penelitian ini melibatkan beberapa jenis pengujian, yakni pengujian tanpa *sampling*, *oversampling* dengan SMOTE, dan *undersampling*. Pada kajian ini memanfaatkan data dari kaggle yang bernama Anemia Type Classification. Dataset anemia ini memiliki 1281 data, 14 fitur, dan 9 kelas. Kelas yang terdapat dalam dataset tersebut antarlain, healthy, normocytic hypochromic aemia, normocytic normochromic anemia, iron deficiency anemia, thrombocytopenia, other microcytic anemia, leukimia, macrocytic anemia, dan leukimia with thrombocytopenia. Gambar 2 di bawah ini merupakan Distribusi kelas pada dataset anemia.



Gambar 2. Distribusi Kelas Dataset Anemia

Sebelum digunakan untuk klasifikasi, dataset dilakukan preprocessing data agar lebih terstruktur, seperti pada penjelasan alur penelitian di atas.

Handle Imbalanced data dan Klasifikasi

Pada bagian ini, distribusi data pada tiap kelas diseimbangkan dengan SMOTE. Algoritma yang diterapkan yaitu SVM untuk proses klasifikasi data. Namun, karena data

yang digunakan merupakan data *multiclass*, pendekatan OvO dan OvR juga diterapkan dalam klasifikasi tersebut.

Tabel 1.
Hasil Pengujian

Teknik/Metode	Accuracy	Precision	Recall	F1-score
OvO + SVM (<i>Baseline</i>)	0.82	0.75	0.62	0.64
OvR + SVM (<i>Baseline</i>)	0.73	0.48	0.51	0.48
<i>Oversampling</i> + OvO + SVM	0.85	0.87	0.90	0.87
<i>Oversampling</i> + OvR + SVM	0.72	0.56	0.68	0.58
<i>Undersampling</i> + OvO + SVM	0.59	0.46	0.62	0.49
<i>Undersampling</i> + OvR + SVM	0.42	0.39	0.51	0.38

Pada Tabel 1 merupakan hasil dari seluruh pengujian yang telah dilakukan, diantaranya pengujian tanpa *sampling*, *oversampling*, dan *undersampling*. Penjelasan mengenai tabel di atas adalah sebagai berikut:

1) Pengujian Tanpa Sampling OvO

Pada pengujian ini, model diuji tanpa resampling menggunakan algoritma SVM, sehingga hasilnya menggambarkan performa model pada data asli. Berdasarkan pengujian tersebut, model OvO menghasilkan akurasi mencapai 82%, secara keseluruhan menunjukkan model cukup baik dalam mengklasifikasikan, precision sebesar 75%, model menunjukkan tingkat keakuratan yang tinggi, recall sebesar 62%, dan F1-score sebesar 64%, menunjukkan model memiliki keterbatasan dalam mengklasifikasikan seluruh kelas. Di dalam pengujian ini menunjukkan bahwa terdapat beberapa kelas yang sudah dapat mengklasifikasikan dengan baik, namun ada pula kesalahan klasifikasi sehingga perlu adanya perbaikan.

2) Pengujian Tanpa Sampling OvR

Berdasarkan pengujian tanpa sampling menggunakan pendekatan OvR dan algoritma SVM, model menghasilkan akurasi sebesar 73% yang kurang baik performanya dibandingkan OvO, precision 48% dan recall 51% lebih rendah, dan f1-score 48%. Jika dibandingkan dengan pengujian sebelumnya, yaitu tanpa sampling OvO, model ini memiliki lebih banyak kesalahan klasifikasi, sehingga kurang tepat jika digunakan pada klasifikasi dataset anemia ini.

3) Pengujian Oversampling OvO

Pada pengujian ini, model diuji dengan oversampling menggunakan SMOTE dan algoritma SVM. Data pada kelas minoritas ditambah sehingga data lebih seimbang. Diterapkannya pendekatan OvO di dalam pengujian ini karena data yang digunakan adalah *multiclass*. Berdasarkan hasil pengujiannya, akurasi yang dihasilkan tertinggi 85% , dengan precision 87%, recall mencapai 90%, dan f1-score 87%. Hal ini menunjukkan bahwa model bekerja dengan sangat baik dalam mengklasifikasikan data. Jika dibandingkan dengan pengujian sebelumnya, model ini bekerja terbaik untuk diterapkan pada klasifikasi data anemia.

4) Pengujian Oversampling OvR

Pengujian oversampling menggunakan SMOTE dan algoritma SVM menghasilkan akurasi sebesar 72% yang tidak meningkat dibanding baseline OvR, precision sebesar 56% dan recall 68% yang menunjukkan terdapat perbaikan dibandingkan pada baseline OvR, namun tidak sebaik pada

pengujian oversampling OvO, serta f1-score sebesar 58% yang artinya sedikit lebih baik daripada baseline OvR.

5) Pengujian Undersampling OvO

Pada pengujian undersampling, data pada kelas mayoritas dikurangi agar distribusi kelas lebih seimbang. Pengujian ini menggunakan algoritma SVM dan pendekatan OvO. Berdasarkan hasil pada Tabel 2 menunjukkan bahwa mendapatkan akurasi yang rendah, yakni 59%, precision 46% yang artinya cukup rendah dan menunjukkan tingkat kesalahan yang tinggi dalam mengklasifikasikan kelas, recall 62% yang relatif lebih baik, menunjukkan kemampuan model menangkap sebagian besar data minoritas, serta f1-score 49%, menunjukkan kinerja model secara keseluruhan kurang memadai.

6) Pengujian Undersampling OvR

Pengujian ini menggunakan teknik undersampling, yang menerapkan algoritma SVM dan pendekatan OvR. Pengujian tersebut mendapatkan hasil dengan akurasi terendah 42%, menunjukkan performa yang sangat buruk akibat hilangnya data mayoritas, precision 39%, recall 51%, dan f1-score 38%, menunjukkan metode ini tidak efektif. Sesuai dengan hasil eksperimen yang dijelaskan, menyatakan jika teknik oversampling OvO SVM mengalahkan baseline model yang tidak menerapkan resampling. Hal ini menyatakan bahwasanya tanpa penanganan pada kasus data tidak seimbang, performa klasifikasi menjadi lebih condong pada kelas dominan.

KESIMPULAN

Berdasarkan hasil dari penelitian yang berjudul “Implementasi Teknik Resampling untuk Mengatasi Ketidakseimbangan Data terhadap Klasifikasi Anemia menggunakan *Support Vector Machine*” yang dipadukan dengan pendekatan OvO dan OvR dapat mengatasi permasalahan ketidakseimbangan data pada klasifikasi *multiclass*. Pada penelitian ini, pendekatan OvO berhasil meningkatkan akurasi model dan performa dari klasifikasi *multiclass*. Hasil pengujian dengan *oversampling* menggunakan SMOTE dan pendekatan OvO sebesar 85% dibandingkan dengan pengujian tanpa sampling OvO dan OvR, pengujian *oversampling* OvR, serta pengujian *undersampling* OvO dan OvR yang lebih rendah. Hal ini menunjukkan bahwa teknik SMOTE berhasil mengatasi masalah data yang tidak seimbang. Selain itu, penggunaan pendekatan OvO dapat meningkatkan akurasi dalam mengklasifikasi data *multiclass*.

DAFTAR PUSTAKA

- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter. *Jurnal Tekno Kompak*, 14(2), 86. <https://doi.org/10.33365/jtk.v14i2.792>
- Aryanti, R., Misriati, T., & Hidayat, R. (2023). Klasifikasi Risiko Kesehatan Ibu Hamil Menggunakan Random Oversampling Untuk Mengatasi Ketidakseimbangan Data. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 3(5), 409–416.
- Dharmendra, I. K., Agus, I. M., Putra, W., & Atmojo, Y. P. (2024). *Evaluasi Efektivitas SMOTE dan Random Under Sampling pada Klasifikasi Emosi Tweet*. 9(2), 192–193.
- Gumelar, G., Ain, Q., Marsuciati, R., Agustanti Bambang, S., Sunyoto, A., & Syukri Mustafa, M. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *SISFOTEK: Sistem Informasi Dan Teknologi*, 250–255.

- Isnaini, Y. S., Yuliaprida, R., & Pihahay, P. J. (2021). Hubungan Usia, Paritas Dan Pekerjaan Terhadap Kejadian Anemia Pada Ibu Hamil. *Nursing Arts*, 15(2), 65–74. <https://doi.org/10.36741/jna.v15i2.153>
- Kurniadi, D., Nuraeni, F., & Lestari, S. M. (2022). Implementasi Algoritma Naïve Bayes Menggunakan Feature Forward Selection dan SMOTE Untuk Memprediksi Ketepatan Masa Studi Mahasiswa Sarjana. *Jurnal Sistem Cerdas*, 5(2), 63–82. <https://doi.org/10.37396/jsc.v5i2.215>
- Lunak, R. P., & Informatika, F. (2024). *Perbandingan Kinerja Algoritma Klasifikasi Status Mutu Air*. 2(1), 72–76.
- Marselina, D. (2022). Studi Literatur: Penyebab Terjadinya Anemia pada Remaja Putri. *Prosiding Seminar Nasional Dan CFP Kebidanan Universitas Ngudi Waluyo*, 1(2), 544–556.
- Nurdian, R. A., Mujib Ridwan, & Ahmad Yusuf. (2022). Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru. *Jurnal Teknik Informatika Dan Sistem Informasi*, 8(1), 24–32. <https://doi.org/10.28932/jutisi.v8i1.4004>
- Pradana, M. G., Saputro, P. H., & Wijaya, D. P. (2022). Komparasi Metode Support Vector Machine Dan Naïve Bayes Dalam Klasifikasi Peluang Penyakit Serangan Jantung. *Indonesian Journal of Business Intelligence (IJUBI)*, 5(2), 87. <https://doi.org/10.21927/ijubi.v5i2.2659>
- Pratama, F., Nasir, M., & Sauda, S. (2020). Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Untuk Menentukan Stok Persediaan Barang Pada Koperasi Karyawan Pangan Utama. *Journal of Software Engineering Ampera*, 1(2), 71–81. <https://doi.org/10.51519/journalsea.v1i2.46>
- Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 153–160. <https://doi.org/10.57152/malcom.v3i2.897>
- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19. <https://doi.org/10.57152/malcom.v3i1.591>
- Sopiatul Ulum, Alifa, R. F., Rizkika, P., & Rozikin, C. (2023). Perbandingan Performa Algoritma KNN dan SVM dalam Klasifikasi Kelayakan Air Minum. *Generation Journal*, 7(2), 141–146. <https://doi.org/10.29407/gj.v7i2.20270>
- Sulardi, N., & Witanti, A. (2020). Sistem Pakar Untuk Diagnosis Penyakit Anemia Menggunakan Teorema Bayes. *Jurnal Teknik Informatika (Jutif)*, 1(1), 19–24. <https://doi.org/10.20884/1.jutif.2020.1.1.12>
- Syaripudin, D., & Walad, A. (2024). *Algoritma Random Forest Pada Data Prediksi Kecacatan Perangkat Lunak Resampling Technique To Increase the Accuracy Value of Random Forest Algorithm on Software Defect Prediction*. 4845–4858.
- Syukron, M., Santoso, R., & Widiharih, T. (2020). Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data. *Jurnal Gaussian*, 9(3), 227–236. <https://doi.org/10.14710/j.gauss.v9i3.28915>
- Tantika, R. S., & Kudus, A. (2022). Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid. *Bandung Conference Series: Statistics*, 2(2), 159–166. <https://doi.org/10.29313/bcss.v2i2.3590>

- Undersampling, P. R. (2024). *Effect of Random Under sampling , Oversampling , and SMOTE on the Performance of Cardiovascular Disease Prediction Models terhadap Kinerja Model Prediksi Penyakit Kardiovaskular*. 21(1), 88–102. <https://doi.org/10.20956/j.v21i1.35552>
- Yanti, D. E., Framesti, L., & Desiani, A. (2022). *JIP (Jurnal Informatika Polinema) Perbandingan Algoritma C4.5 Dan Svm Dalam Klasifikasi Penyakit Anemia*. 427–434.