

**IMPLEMENTASI TEKNIK *RESAMPLING* UNTUK  
MENGATASI *IMBALANCED DATA* TERHADAP KLASIFIKASI  
ANEMIA MENGGUNAKAN *SUPPORT VECTOR MACHINE*  
(SVM)**

**LAPORAN TUGAS AKHIR**

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh  
Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika  
Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang



**DISUSUN OLEH :**

**NADIA WULANDARI**

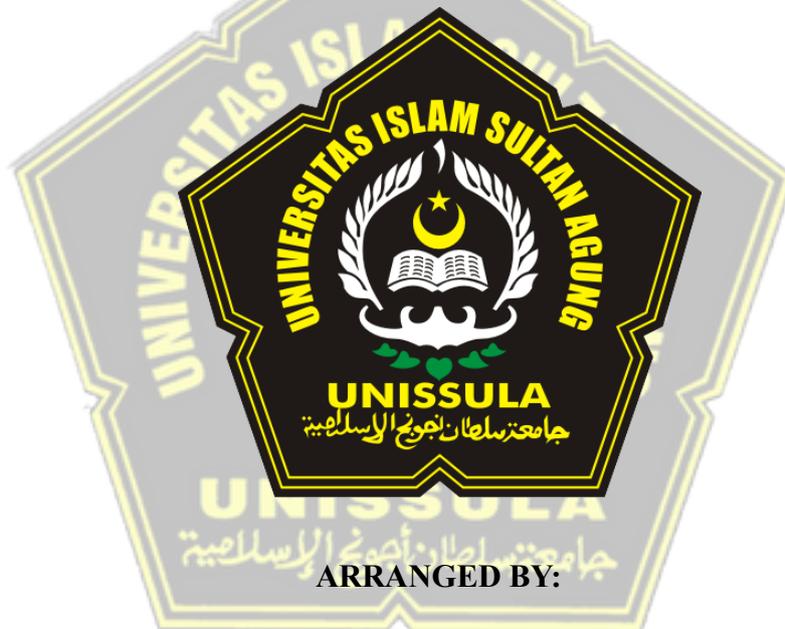
**NIM 32602100095**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG  
JANUARI 2025**

***IMPLEMENTATION OF RESAMPLING TECHNIQUE TO  
OVERCOME IMBALANCED DATA FOR ANEMIA  
CLASSIFICATION USING SUPPORT VECTOR MACHINE  
(SVM)***

***FINAL PROJECT***

*This report is prepared to fulfill one of the requirements to obtain a Bachelor's Degree (S1) in the Informatics Engineering Study Program, Faculty of Industrial Technology, Sultan Agung Islamic University, Semarang*



**ARRANGED BY:**

**NADIA WULANDARI**

**NIM 32602100095**

***MAJORING OF INFORMATICS ENGINEERING  
FACULTY OF INDUSTRIAL TECHNOLOGY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG  
JANUARY 2025***

**LEMBAR PENGESAHAN  
TUGAS AKHIR**

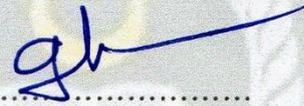
**IMPLEMENTASI TEKNIK *RESAMPLING* UNTUK MENGATASI  
*IMBALANCED DATA* TERHADAP KLASIFIKASI ANEMIA  
MENGUNAKAN *SUPPORT VECTOR MACHINE (SVM)***

**NADIA WULANDARI  
NIM 32602100095**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir  
Program Studi Teknik Informatika  
Universitas Islam Sultan Agung  
Pada tanggal : 17 Februari 2025

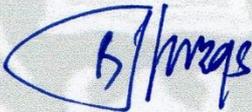
**TIM PENGUJI UJIAN SARJANA :**

**Ghufron, S.T., M.Kom**  
NIDN 0602079005  
(Ketua Penguji)



24-02-2025

**Bagus Satrio WP, S.Kom, M.Cs**  
NIDN 1027118801  
(Anggota Penguji)



29-02-2025

**Badie'ah, S.T., M.Kom**  
NIDN 0619018701  
(Pembimbing)

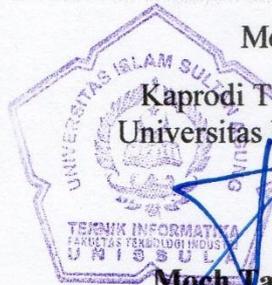


24-02-2025

Semarang, 25 Februari 2025

Mengetahui,

Kaprodi Teknik Informatika  
Universitas Islam Sultan Agung



**Moch Faufik, S.T., M.IT**  
NIDN 0622037502

## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Nadia Wulandari

NIM : 32602100095

Judul Tugas Akhir : Implementasi Teknik *Resampling* untuk Mengatasi *Imbalanced Data* terhadap Klasifikasi Anemia Menggunakan *Support Vector Machine* (SVM)

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 25 Februari 2025

Yang Menyatakan,



Nadia Wulandari

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Nadia Wulandari

NIM : 32602100095

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Semarang

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : Implementasi Teknik *Resampling* untuk Mengatasi *Imbalanced Data* terhadap Klasifikasi Anemia Menggunakan *Support Vector Machine* (SVM)

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 25 Februari 2025



Nadia Wulandari

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada ALLAH SWT, yang telah memberikan rahmat, taufik serta hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi Teknik *Resampling* untuk mengatasi *Imbalanced Data* terhadap Klasifikasi Anemia menggunakan *Support Vector Machine (SVM)*” ini dengan baik. Dengan penuh rasa hormat, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala rahmat dan karunia-Nya.
2. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang mengizinkan penulis menimba ilmu di kampus ini.
3. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T
4. Dosen pembimbing penulis Ibu Badie’ah, ST., M.Kom yang telah memberikan arahan, bimbingan, dan saran yang sangat berarti dalam proses penelitian hingga penulisan laporan ini.
5. Dosen-dosen Program Studi Teknik Informatika, Fakultas Teknologi Industri UNISSULA yang telah membekali penulis dengan ilmu dan pengetahuan selama masa studi.
6. Orang tua penulis, Bapak Kurrohman Wahidin dan Ibu Ana Wahyu Setiawati yang selalu memberikan doa, dukungan, dan motivasi selama proses penyelesaian tugas akhir ini.
7. Teman-teman dan rekan seperjuangan, yang telah memberikan moral, motivasi, serta semangat dalam proses penyelesaian tugas akhir ini.
8. Dan kepada seluruh pihak yang tidak dapat saya sebutkan satu persatu.

Semarang, 19 Januari 2025

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN JUDUL .....</b>	<b>ii</b>
<b>LEMBAR PENGESAHAN TUGAS AKHIR .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vi</b>
<b>DAFTAR TABEL .....</b>	<b>viii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>x</b>
<b>ABSTRAK .....</b>	<b>xi</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	2
1.3 Pembatasan Masalah .....	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	4
<b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....</b>	<b>5</b>
2.1 Tinjauan Pustaka.....	5
2.2 Dasar Teori .....	6
2.2.1 Anemia .....	6
2.2.2 <i>Imbalanced Data</i> .....	7
2.2.3 Teknik SMOTE ( <i>Synthetic Minority Over-Sampling Technique</i> ).....	7
2.2.4 Klasifikasi <i>Multiclass</i> .....	8
2.2.5 Pendekatan <i>One vs One</i> (OvO) .....	8
2.2.6 Pendekatan <i>One vs Rest</i> (OvR) .....	8
2.2.7 <i>Support Vector Machine</i> (SVM).....	8
<b>BAB III METODE PENELITIAN .....</b>	<b>10</b>

3.1	Fase 1: <i>Preprocessing Data</i> .....	11
3.2	Fase 2: <i>Handle Imbalanced Data</i> .....	12
3.3	Fase 3: Klasifikasi .....	13
3.4	Evaluasi .....	13
3.5	Hasil.....	13
<b>BAB IV HASIL DAN ANALISIS PENELITIAN.....</b>		<b>14</b>
4.1	HASIL PENELITIAN .....	14
4.1.1	<i>Preprocessing Data</i> .....	14
4.1.2	<i>Handle Imbalanced Data</i> dan Klasifikasi .....	17
4.2	ANALISA PENELITIAN .....	36
4.2.1	Evaluasi.....	36
4.2.2	Hasil Implementasi Model .....	36
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>41</b>
5.1	Kesimpulan.....	41
5.2	Saran.....	41
<b>DAFTAR PUSTAKA.....</b>		<b>42</b>
<b>HALAMAN LAMPIRAN.....</b>		<b>45</b>



## DAFTAR TABEL

Tabel 4. 1 Sampel Dataset Anemia .....	15
Tabel 4. 2 Sampel Dataset Anemia .....	15
Tabel 4. 3 <i>Classification Report Tanpa Sampling</i> OvO .....	18
Tabel 4. 4 <i>Classification Report Tanpa Sampling</i> OvR .....	20
Tabel 4. 5 <i>Classification Report Oversampling</i> OvO .....	24
Tabel 4. 6 <i>Classification Report Oversampling</i> OvR.....	26
Tabel 4. 7 <i>Classification Report Undersampling</i> OvO .....	30
Tabel 4. 8 <i>Classification Report Undersampling</i> OvR .....	33
Tabel 4. 9 Evaluasi Perbandingan Pengujian .....	36



## DAFTAR GAMBAR

Gambar 2. 1 Arsitektur SVM (Yohannes & Al Rivan, 2022) .....	9
Gambar 3. 1 <i>Workflow</i> Perancangan Sistem .....	10
Gambar 4. 1 Distribusi Kelas pada Dataset Anemia .....	14
Gambar 4. 2 <i>Confusion Matrix</i> Tanpa <i>Sampling</i> OvO .....	19
Gambar 4. 3 <i>Confusion Matrix</i> Tanpa <i>Sampling</i> OvR .....	22
Gambar 4. 4 <i>Confusion Matrix</i> <i>Oversampling</i> OvO .....	25
Gambar 4. 5 <i>Confusion Matrix</i> <i>Oversampling</i> OvR .....	28
Gambar 4. 6 <i>Confusion Matrix</i> <i>Undersampling</i> OvO .....	31
Gambar 4. 7 <i>Confusion Matrix</i> <i>Undersampling</i> OvR .....	34
Gambar 4. 8 Tampilan <i>UI</i> Klasifikasi Anemia .....	37
Gambar 4. 9 Tampilan <i>UI</i> Klasifikasi Anemia .....	37
Gambar 4. 10 Tampilan <i>UI</i> Klasifikasi Anemia .....	38
Gambar 4. 11 Uji Coba Klasifikasi .....	39
Gambar 4. 12 Uji Coba Klasifikasi .....	39
Gambar 4. 13 Uji Coba Klasifikasi .....	40



## DAFTAR LAMPIRAN



## ABSTRAK

Anemia merupakan penyakit yang disebabkan karena rendahnya kadar kemoglobin dalam darah. Anemia sendiri memiliki beberapa jenis sehingga perlu adanya untuk meningkatkan efektivitas proses pengklasifikasiannya. Namun permasalahan ketidakseimbangan data yang sering muncul dalam klasifikasi anemia dapat mengurangi performa klasifikasi. Tantangan lain dalam proses klasifikasi yaitu banyaknya jumlah kelas pada dataset yang digunakan, sehingga perlu adanya algoritma lain yang dapat diterapkan secara bersamaan dengan algoritma SVM. Penelitian ini menerapkan teknik *resampling*, yaitu dengan teknik *oversampling* dan *undersampling*. Teknik yang digunakan untuk melakukan proses *oversampling* yaitu *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk menangani permasalahan data yang tidak seimbang. Algoritma SVM digunakan dalam penelitian ini yang kemudian dipadukan dengan pendekatan *One vs One* (OvO) dan *One vs Rest* (OvR) karena data yang digunakan merupakan data multiclass. Tujuan dari penelitian ini yaitu membandingkan teknik *resampling* dari pendekatan OvO dan OvR dalam mengklasifikasikan anemia *multiclass*. Hasil penelitian yang dilakukan menunjukkan bahwa pengujian model *oversampling* OvO lebih unggul secara keseluruhan dengan akurasi sebesar 85% . Dengan demikian, penelitian ini menyimpulkan bahwa teknik *oversampling* dan pendekatan OvR adalah teknik yang lebih unggul untuk melakukan klasifikasi dengan data yang memiliki beberapa kelas.

Kata kunci: Klasifikasi, SMOTE, *Support Vector Machine*, *Imbalance Data*, *Multiclass Classification*.

## ABSTRACT

*Anemia is a disease caused by low levels of chemoglobin in the blood. Anemia itself has several types, so it is necessary to increase the effectiveness of the classification process. However, data imbalance problems that often arise in anemia classification can reduce classification performance. Another challenge in the classification process is the large number of classes in the dataset used, so it is necessary to have another algorithm that can be applied simultaneously with the SVM algorithm. This research applies resampling techniques, namely oversampling and undersampling techniques. The technique used to carry out the oversampling process is Synthetic Minority Over-Sampling Technique (SMOTE) to handle the problem of unbalanced data. The SVM algorithm used in this research was then combined with the One vs One (OvO) and One vs Rest (OvR) approaches because the data used was multiclass data. The aim of this study is to compare resampling techniques from the OvO and OvR approaches in classifying multiclass anemia. The results of the research conducted show that the OvO oversampling model testing is superior overall with an accuracy of 85%. Thus, this study concludes that the oversampling technique and the OvR approach are superior techniques for classifying data that has several classes.*

*Keywords: Classification, SMOTE, Support Vector Machine, Imbalance Data, Multiclass Classification.*

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

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. 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 (Isnaini dkk., 2021).

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.

Salah satu pendekatan yang telah digunakan untuk mengatasi masalah ketidakseimbangan data ialah teknik *resampling*. Teknik yang dapat diterapkan untuk menyeimbangkan data yaitu teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE merupakan teknik yang digunakan untuk menghasilkan data minoritas sehingga jumlahnya seimbang dengan data mayoritas. Sedangkan pada *undersampling*, data asli yang terdapat pada kelas mayoritas akan didistribusikan atau dikurangi yang bertujuan agar dataset lebih seimbang (Kasanah dkk., 2019).

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). SVM adalah algoritma yang digunakan untuk melakukan klasifikasi terhadap suatu objek dengan atribut yang tersedia. Metode SVM dapat melakukan pencarian terhadap *hyperplane* yang optimal untuk membagi antara dua kelas pada klasifikasi dengan memaksimalkan margin antara dua kelas tersebut. Algoritma SVM mempunyai kelebihan yaitu dapat bekerja dengan baik pada data berdimensi tinggi dengan memanfaatkan kernel.

Namun algoritma SVM memiliki kekurangan yaitu tidak cocok digunakan untuk jumlah sampel yang berskala besar (Yanti dkk., 2022). Klasifikasi menggunakan SVM *multiclass* dapat dioptimasi dengan menggunakan pendekatan *One vs One* (OvO) dan *One vs Rest* (OvR). Pada penelitian ini akan membandingkan penggunaan pendekatan OvO dan OvR karena pada penelitian sebelumnya pada OvO dapat mengklasifikasi dengan jumlah kelas yang lebih banyak sedangkan pada OvR lebih cocok digunakan untuk jumlah kelas yang tidak terlalu banyak (Alita dkk., 2020).

Penelitian ini bertujuan untuk mengimplementasikan teknik SMOTE untuk mengatasi masalah ketidakseimbangan data pada klasifikasi anemia dengan menggunakan algoritma SVM, yang dipadukan dengan pendekatan OvO dan OvR. Diharapkan bahwa penerapan pendekatan ini dapat meningkatkan akurasi model dalam klasifikasi anemia, khususnya dalam mendeteksi kasus anemia pada klasifikasi *multiclass*.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana pengaruh implementasi teknik SMOTE terhadap ketidakseimbangan data pada akurasi anemia?
2. Bagaimana pengaruh implementasi SVM *One vs One* dan *One vs Rest* terhadap akurasi anemia *multiclass*?
3. Berapa akurasi performa model SVM *One vs One* dan *One vs Rest* yang telah diimbangi dengan teknik SMOTE dibandingkan dengan sebelum diimbangi?

### 1.3 Pembatasan Masalah

Untuk menjaga fokus pada penelitian, terdapat beberapa batasan yaitu:

1. Penelitian ini hanya menggunakan data kesehatan dengan klasifikasi anemia dan tidak mencakup jenis penyakit lainnya. Dataset yang digunakan merupakan data sekunder yang didapat dari Kaggle dan tidak melakukan pengumpulan data primer.
2. Teknik penyeimbang data yang digunakan yaitu teknik resampling yaitu *oversampling* dan *undersampling*.
3. *Output* yang dihasilkan hanya klasifikasi anemia *multiclass*, bukan rekomendasi.

### 1.4 Tujuan

Penelitian ini memiliki beberapa tujuan yaitu:

1. Mengimplementasikan model yang dapat mengklasifikasikan anemia *multiclass*.
2. Mengimplementasikan teknik *resampling* yakni *oversampling* dan *undersampling* untuk mengatasi masalah ketidakseimbangan data pada klasifikasi anemia.
3. Membandingkan performa model SVM sebelum dan sesudah penerapan teknik *resampling*.

### 1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode pengolahan data yang tidak seimbang, khususnya dalam konteks klasifikasi medis, dengan mengaplikasikan teknik SMOTE, dan algoritma SVM.
2. Penelitian ini diharapkan dapat membantu praktisi kesehatan dalam mendeteksi anemia secara lebih efektif dan akurat.

3. Implementasi SMOTE dan SVM dapat membuka peluang bagi pengembangan otomatis berbasis kecerdasan buatan dalam klasifikasi penyakit.
4. Penelitian ini dapat membantu pengelola data medis dalam memahami pentingnya penanganan ketidakseimbangan data, sehingga meningkatkan kualitas analisis dan prediksi dari data pasien.

## 1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam pembuatan laporan tugas akhir adalah sebagai berikut:

### **BAB I : PENDAHULUAN**

Pada BAB I menjelaskan tentang latar belakang, pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, dan sistematika penulisan.

### **BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI**

Pada BAB II berisi penelitian-penelitian terdahulu yang dijadikan tinjauan dalam melakukan penelitian, serta teori-teori mengenai topik pembahasan dalam penelitian seperti *imbalanced data* dalam klasifikasi SVM *multiclass*.

### **BAB III : METODE PENELITIAN**

Pada BAB III menjelaskan proses penelitian yang dimulai dari pengumpulan data hingga evaluasi performa klasifikasi.

### **BAB IV : HASIL DAN ANALISIS PENELITIAN**

Pada BAB IV berisi tentang pemaparan hasil penelitian yang dimulai dari hasil akhir sistem, klasifikasi data uji, dan performa klasifikasi sistem.

### **BAB V : KESIMPULAN DAN SARAN**

Pada BAB V berisi rangkuman keseluruhan proses penelitian dari awal hingga akhir.

## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Algoritma SVM ialah teknik yang efektif untuk menangani tantangan klasifikasi data. Oleh karena itu, gabungan antara teknik SMOTE dan algoritma SVM diharapkan dapat mengoptimalkan akurasi dalam mengklasifikasikan anemia (Gumelar dkk., 2021). Kajian sebelumnya yang menggunakan algoritma SVM telah banyak digunakan pada penelitian untuk klasifikasi penyakit seperti anemia. Pada penelitian yang dilakukan oleh (Yanti dkk., 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 dkk., 2023), terbukti unggul dalam mengklasifikasikan kelayakan air minum dengan akurasi sebesar 69,764%. Klasifikasi penyakit menggunakan SVM pada kajian yang dilaksanakan menyatakan bahwa algoritma SVM kompeten dalam mengklasifikasikan penyakit kardiovaskular dengan akurasi terbaik sebesar 87% (Pradana dkk., 2022).

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 dkk., 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 (Alita dkk., 2020).

Salah satu kendala yang kerap terjadi dalam klasifikasi ialah ketidakseimbangan data. Pada penelitian yang dilakukan oleh (Syukron dkk., 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 dkk., 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 dalam proses klasifikasi yaitu terdapat beberapa kelas sehingga perlu adanya pendekatan untuk mengklasifikasikan kasus *multiclass*. 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 menanganani 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 menunjukkan bahwa metode *One Against One* yang dipadukan dengan kernel polinomial menghasilkan presisi tertinggi sebesar 97% (Rabbani dkk., 2023).

## 2.2 Dasar Teori

### 2.2.1 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. Terdapat beberapa jenis anemia, contohnya seperti anemia hipokromik normositik, anemia normokromik

normositik, anemia defisiensi besi, trombositopenia, anemia makrositik, dan sebagainya. Oleh karena itu, penting untuk melakukan identifikasi dan klasifikasi anemia berdasarkan data kesehatan pasien guna memastikan intervensi medis dapat diberikan secara tepat waktu (Marselina, 2022).

### 2.2.2 *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. Ketidakseimbangan data dapat diatasi menggunakan teknik *resampling*, salah satunya adalah teknik SMOTE, yang menghasilkan sampel tambahan untuk kelas minoritas agar lebih seimbang dengan kelas mayoritas. Selain itu, *undersampling* juga dapat diterapkan dengan mengurangi jumlah sampel pada kelas mayoritas, sehingga proporsi data antara kelas mayoritas dan minoritas menjadi lebih seimbang (Aryanti dkk., 2023).

Ketidakseimbangan dalam data dapat berdampak pada hasil prediksi yang dihasilkan oleh suatu algoritma. Untuk mengukur sejauh mana tingkat ketidakseimbangan tersebut, dapat dilakukan perhitungan menggunakan *Imbalanced Ratio* (IR) dengan perbandingan sebagai berikut (Akbar & Hayaty, 2020):

$$\text{Imbalanced Ratio (IR)} = \frac{\text{majority}}{\text{minority}} \dots \dots \dots (1)$$

Keterangan :

- IR = *Imbalanced Ratio*  
*Majority* = Nilai kelas mayoritas  
*Minority* = Nilai kelas minoritas

### 2.2.3 Teknik SMOTE (*Synthetic Minority Over-Sampling Technique*)

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 dkk., 2022). SMOTE berfungsi menyeimbangkan distribusi data pada kelas mayoritas. SMOTE juga dapat mencegah terjadinya *overfitting* yang disebabkan oleh duplikasi data di

kelas minoritas, sehingga memungkinkan penggunaan data pelatihan yang sama secara lebih efektif (Kurniadi dkk., 2022).

#### **2.2.4 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. 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* dan *One vs Rest* (Lunak & Informatika, 2024).

#### **2.2.5 Pendekatan *One vs One* (OvO)**

Masalah dalam klasifikasi *multiclass* adalah mengklasifikasikan data ke lebih dari dua kelas. Salah satu metode yang dapat digunakan untuk memungkinkan algoritma SVM mengklasifikasikan lebih dari dua kelas adalah pendekatan OvO. Prinsip utama dari pendekatan OvO adalah membangun model SVM biner, di mana setiap pasangan kelas akan dilatih secara terpisah. Keuntungan dari pendekatan ini adalah kemudahan implementasinya, terutama ketika menggunakan algoritma yang dirancang khusus untuk klasifikasi biner (Tantika & Kudus, 2022).

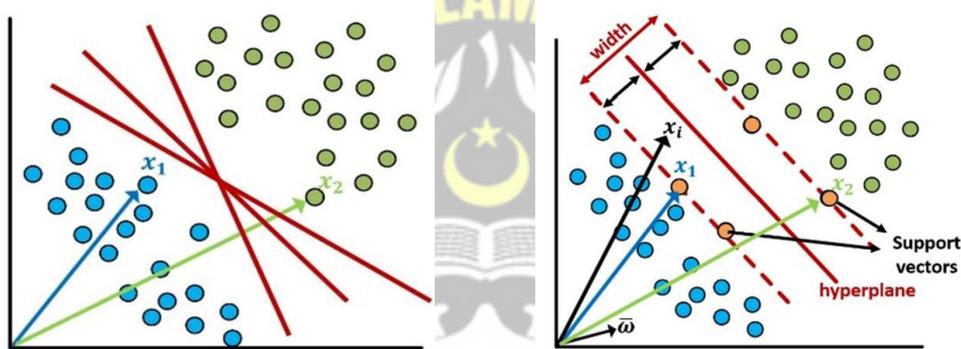
#### **2.2.6 Pendekatan *One vs Rest* (OvR)**

Pendekatan lain yang dapat diterapkan agar algoritma SVM mampu mengklasifikasikan lebih dari dua kelas adalah OvR. Pendekatan OvR biasanya digunakan untuk data dengan jumlah kelas yang relatif sedikit. Dalam metode ini, seluruh data akan dianalisis dalam satu formulasi secara langsung. Prinsip dasar dari OvR adalah melatih data pada setiap kelas secara terpisah dengan membandingkannya melawan kelas lainnya (Tantika & Kudus, 2022).

#### **2.2.7 *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 dirancang untuk menangani masalah nonlinier dengan menerapkan konsep kernel ke ruang berdimensi tinggi (Pratama dkk., 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 dkk., 2023). SVM memiliki dua jenis *hyperplane* utama, yakni *hyperplane* linear dan *hyperplane* non-linear. Jika suatu dataset dipisahkan secara sempurna menggunakan *hyperplane* linear, maka SVM disebut SVM linear. Namun, jika pemisahan secara linear tidak memungkinkan, SVM menerapkan teknik transformasi kernel untuk memetakan data ke dalam dimensi fitur yang lebih tinggi, sehingga memungkinkan pembentukan *hyperplane* linear. Fungsi kernel berperan dalam mentransformasikan data ke ruang berdimensi lebih tinggi guna meningkatkan struktur data dan mempermudah pemisahan (Yohannes & Al Rivan, 2022).



Gambar 2. 1 Arsitektur SVM (Yohannes & Al Rivan, 2022)

Pada Gambar 2.1 merupakan gambar arsitektur SVM. Algoritma SVM melakukan klasifikasi data dengan menentukan *hyperplane* optimal yang dapat memisahkan data dari satu atau lebih kelas. Pendekatan SVM didasarkan pada konsep jarak minimum antara batas keputusan dan salah satu sampel data. Secara umum, rumus SVM linear dapat dinyatakan sebagai berikut (Rabbani dkk., 2023):

$$f(x) = \text{sign}(w \cdot x + b) \dots \dots \dots (2)$$

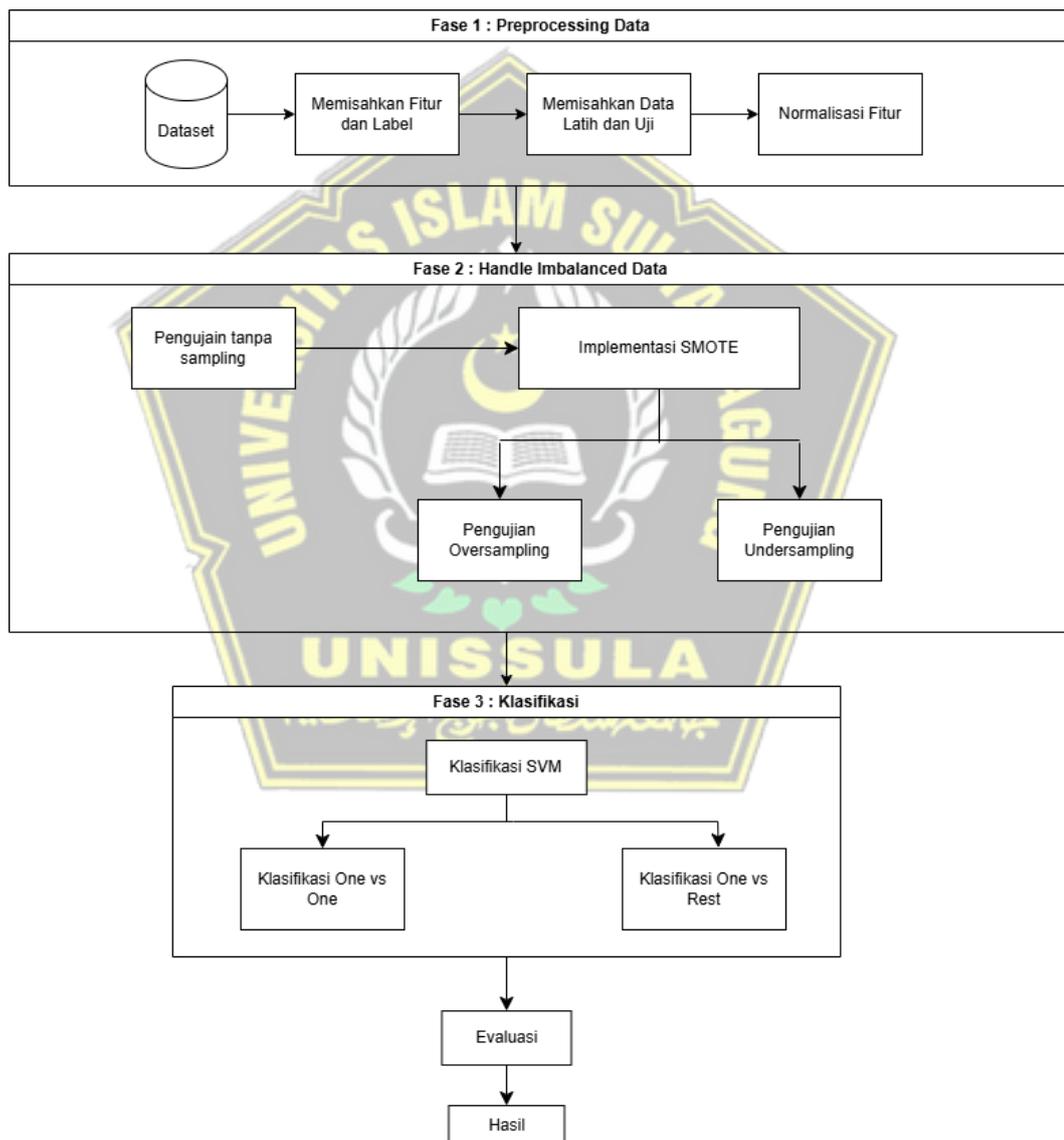
Keterangan :

- $f(x)$  = Fungsi prediksi
- $w$  = Vektor normal *hyperplane*
- $x$  = Vektor fitur input
- $b$  = Bias atau *intercept*

### BAB III

## 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 3.1 menggambarkan proses yang diterapkan dalam penelitian ini.



Gambar 3. 1 *Workflow* Perancangan Sistem

### 3.1 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:

#### 1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah data kesehatan yang berkaitan dengan klasifikasi anemia. Dataset ini akan diperoleh dari Kaggle dan mencakup informasi penting, seperti jumlah hemoglobin, jumlah trombosit, jumlah sel darah putih, jumlah sel darah merah, rata-rata sel darah merah, rata-rata hemoglobin per sel darah merah, rata-rata konsentrasi hemoglobin dalam sel darah merah, pengukuran variabilitas distribusi ukuran trombosit dalam darah, tes prokalsitonin, dan diagnosis anemia berdasarkan parameter CBC. Dataset ini memiliki 1281 data dengan 9 kelas dan 14 fitur, antarlain WBC (*White Blood Cell*), LYMp (*Lymphocyte Percentage*), NEUTp (*Neutrophil Percentage*), LYMn (*Lymphocyte Number*), NEUTn (*Neutrophil Number*), RBC (*Red Blood Cell*), HGB (*Hemoglobin*), HCT (*Hematocrit*), MCV (*Mean Corpuscular*), MCH (*Mean Corpuscular*), MCHC (*Mean Corpuscular Hemoglobin Concentration*), PLT (*Platelet*), PDW (*Platelet Distribution Width*), PCT (*Plateletcrit*), serta Diagnosis.

Aboelnaga, E. (n.d.). *Anemia Types Classification - Anemia type prediction based on CBC data*.

#### 2. Memisahkan Fitur dan Label

Dalam proses memisahkan fitur dan label, data yang digunakan untuk proses klasifikasi akan dibagi menjadi dua bagian. Bagian pertama, yang disebut X terdiri dari sekumpulan variabel independen yang akan digunakan untuk memprediksi hasil klasifikasi. Bagian kedua, yaitu Y berisi variabel yang ingin diprediksi berupa klasifikasi anemia, sehingga model dapat mempelajari pola-pola yang ada untuk menentukan klasifikasi berdasarkan fitur yang diberikan.

### 3. 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.

### 4. Normalisasi Fitur

Normalisasi bertujuan untuk menskalakan fitur agar berada dalam rentang nilai tertentu, misalnya antara 0 dan 1. 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.

## 3.2 Fase 2: *Handle Imbalanced Data*

Pada fase 2 ini merupakan tahapan menguji data tanpa sampling dan untuk *imbalanced data* menggunakan teknik SMOTE. Pada tahapan ini, data minoritas akan ditambah dan data mayoritas akan dikurangi agar lebih seimbang, sehingga proses pengklasifikasian data dapat lebih akurat. Setelah data diseimbangkan, data akan diuji dengan pengujian di bawah ini:

### 1. Pengujian Tanpa *Resampling* Data

Pada pengujian awal akan dilakukan pengujian model tanpa melakukan teknik resampling. Pada tahap ini data asli akan diuji tanpa ada perubahan pada distribusi kelas, sehingga hasilnya akan menggambarkan performa model pada data asli.

### 2. Pengujian *Oversampling*

Pada pengujian yang kedua akan dilakukan teknik *oversampling* dengan menggunakan teknik SMOTE akan diimplementasikan pada data latih untuk mengatasi masalah ketidakseimbangan kelas. SMOTE akan menghasilkan sampel sintetis baru dari kelas minoritas, sehingga distribusi data menjadi lebih seimbang.

### 3. Pengujian *Undersampling*

Pada pengujian yang ketiga dilakukan teknik undersampling untuk menangani permasalahan *imbalanced data*, namun dengan cara mengurangi jumlah data pada kelas mayoritas agar distribusi kelas menjadi lebih seimbang.

### 3.3 Fase 3: Klasifikasi

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

### 3.4 Evaluasi

Pada tahapan evaluasi akan dibandingkan hasil performa dari pengujian tanpa *sampling*, pengujian *oversampling*, dan pengujian *undersampling* yang dikombinasikan dengan pendekatan *One vs One* dan *One vs Rest* untuk mendapatkan hasil terbaik.

### 3.5 Hasil

Setelah mendapatkan hasil terbaik dari perbandingan tersebut, pada tahapan ini model akan diimplementasikan dalam bentuk web yang bertujuan untuk melakukan klasifikasi penyakit anemia. Pada web tersebut terdapat kolom sesuai dengan dataset untuk memasukkan data yang akan diklasifikasikan. Setelah data diinputkan, sistem akan memproses data dan memberikan hasil klasifikasi anemia.

## BAB IV

### HASIL DAN ANALISIS PENELITIAN

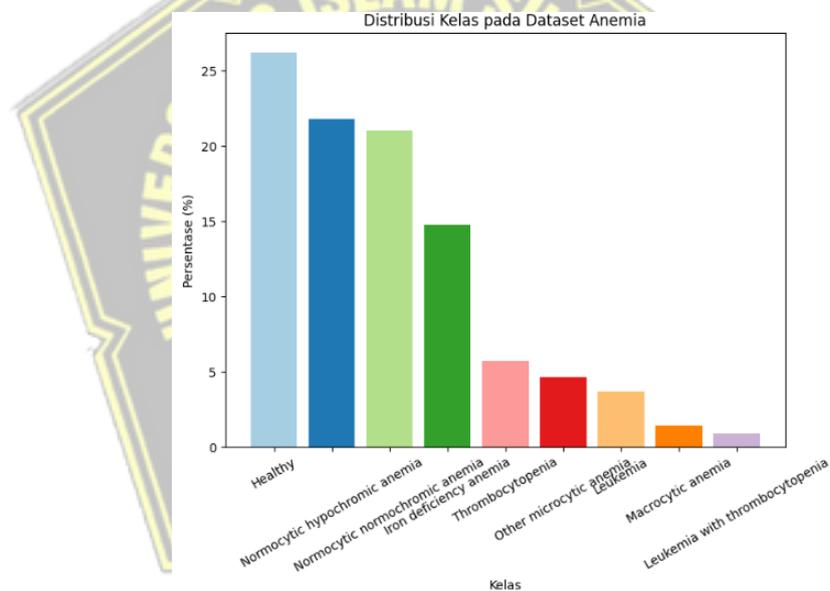
#### 4.1 HASIL PENELITIAN

##### 4.1.1 Preprocessing Data

Fase 1, yang merupakan tahap *preprocessing* data, mencakup beberapa tahapan penelitian, diantaranya:

##### 1. Pengumpulan Data

Pada kajian ini memanfaatkan data dari kaggle yang bernama *Anemia Type Classification*. Dataset anemia ini memiliki 1281 data, 14 fitur, dan 9 kelas. Gambar 4.1 merupakan Distribusi kelas pada dataset anemia.



Gambar 4. 1 Distribusi Kelas pada Dataset Anemia

Fitur yang terdapat di dalam dataset tersebut yaitu WBC (*White Blood Cell*), LYMp (*Lymphocyte Percentage*), NEUTp (*Neutrophil Percentage*), LYMn (*Lymphocyte Number*), NEUTn (*Neutrophil Number*), RBC (*Red Blood Cell*), HGB (*Hemoglobin*), HCT (*Hematocrit*), MCV (*Mean Corpuscular*), MCH (*Mean Corpuscular*), MCHC (*Mean Corpuscular Hemoglobin Concentration*), PLT (*Platelet*), PDW (*Platelet Distribution Width*), PCT (*Plateletcrit*), serta Diagnosis.

Sedangkan kelas yang terdapat pada dataset anemia antaralain healthy dengan jumlah data 336, *normocytic hypochromic anemia* dengan jumlah data 279, *normocytic normochromic anemia* dengan jumlah data 269, *iron deficiency anemia* dengan jumlah data 189, *thrombocytopenia* dengan jumlah data 73, *other microcytic anemia* dengan jumlah data 59, *leukimia* dengan jumlah data 47, *macrocytic anemia* dengan jumlah data 18, dan *leukimia with thrombocytopenia* dengan jumlah data 11.

Adapun sampel data pada dataset *Anemia Types Classification* dapat dilihat pada Tabel berikut.

Tabel 4. 1 Sampel Dataset Anemia

WBC	LYMp	NEUTp	LYMn	NEUTn	RBC	HGB	HCT
10	43.02.00	50.01.00	04.03	5	0,136806	07.03	24.02.00
10	42.04.00	52.03.00	04.02	05.03	0,141667	07.03	25
07.02	30.07.00	60.07.00	02.02	04.04	0,192361	9	30.05.00
6	30.02.00	63.05.00	01.08	03.08	04.22	03.08	32.08.00
04.02	39.01.00	53.07.00	01.06	02.03	0,189583	00.04	316
06.06	27.03.00	65.04.00	01.08	04.03	0,191667	08.08	29.07.00
16.07	19.01	68.02.00	03.02	11.04	05.15	14.02	44.08.00
09.03	27.04.00	64	02.06	05.09	04.39	12	37.09.00
05.02	19.07	72.04.00	1	03.08	0,225694	13.02	41

Tabel 4. 2 Sampel Dataset Anemia

MCV	MCH	MCHC	PLT	PDW	PCT	Diagnosis
87.07.00	26.03.00	30.01.00	189	12.05	00.17	<i>Normocytic hypochromic anemia</i>
88.02.00	25.07.00	20.02	180	12.05	00.16	<i>Normocytic hypochromic anemia</i>
77	22.06	29.05.00	148	14.03	00.14	<i>Iron deficiency anemia</i>

77.09.00	23.02	29.08.00	143	11.03	00.12	<i>Iron deficiency anemia</i>
80.06.00	23.09	29.07.00	236	12.08	00.22	<i>Normocytic hypochromic anemia</i>
75.02.00	22.02	79.06.00	207	11.05	00.18	<i>Other microcytic anemia</i>
87.01.00	27.05.00	31.06.00	151	12.08	00.14	<i>Leukemia</i>
86.04.00	27.03.00	31.06.00	194	15.09	00.19	<i>Normocytic hypochromic anemia</i>
84.07.00	27.02.00	32.01.00	181	10	00.15	<i>Healthy</i>

Di dalam dataset tersebut dilakukan perhitungan *Imbalanced Ratio* (IR) untuk mengetahui data telah seimbang atau tidak agar dapat ditangani dengan teknik *resampling*.

$$IR = \frac{\text{majority}}{\text{minority}}$$

$$IR = \frac{336}{11}$$

$$IR = 30,54$$

Berdasarkan hasil yang diperoleh dalam menghitung IR sebesar 30,54. Hal ini berarti jumlah sampel di kelas mayoritas sekitar 30 kali lebih banyak dibandingkan jumlah sampel di kelas minoritas. Oleh karena itu, perlu adanya teknik untuk menangani ketidakseimbangan data agar model machine learning tidak bias dalam mengklasifikasikan penyakit anemia.

## 2. 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.

### 3. 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.

### 4. 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.

#### 4.1.2 *Handle Imbalanced Data dan Klasifikasi*

Pada tahap ini, dilakukan langkah untuk mengatasi ketidakseimbangan data agar hasil klasifikasi lebih akurat dan tidak condong pada kelas dengan jumlah data yang lebih banyak. Oleh karena itu, teknik SMOTE diterapkan untuk mengatasi permasalahan tersebut.

Pada tahap ini, algoritma SVM diterapkan bersama dua pendekatan, yaitu OvO dan OvR yang hasilnya akan dibandingkan berdasarkan tingkat akurasi.

#### 1. Pengujian Tanpa *Sampling* OvO

Pada pengujian ini, model diuji tanpa *resampling* menggunakan algoritma SVM, sehingga hasilnya menggambarkan performa model pada data asli.

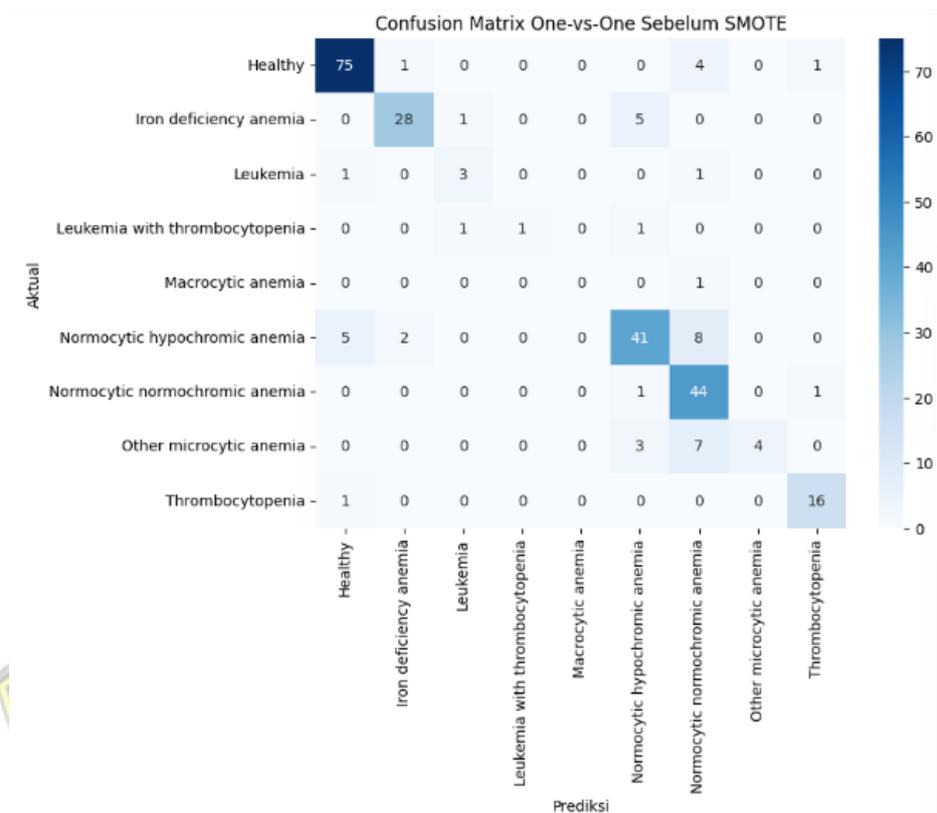
a. *Classification Report*

Tabel 4. 3 *Classification Report* Tanpa Sampling OvO

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.91	0.93	0.92	81
<i>Iron deficiency anemia</i>	0.90	0.82	0.86	34
<i>Leukemia</i>	0.60	0.60	0.60	5
<i>Leukemia with thrombocytopenia</i>	1.00	0.33	0.50	3
<i>Macrocytic anemia</i>	0.00	0.00	0.00	1
<i>Normocytic hypochromic anemia</i>	0.80	0.73	0.77	56
<i>Normocytic normochromic anemia</i>	0.68	0.96	0.79	46
<i>Other microcytic anemia</i>	1.00	0.29	0.44	14
<i>Thrombocytopenia</i>	0.89	0.94	0.91	17
<i>accuracy</i>			0.82	257
<i>macro avg</i>	0.75	0.62	0.64	257
<i>weighted avg</i>	0.84	0.82	0.82	257

Tabel 4.3 merupakan hasil *classification report* dari pengujian tanpa sampling menggunakan OvO. 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.

b. *Confusion Matrix*



Gambar 4. 2 *Confusion Matrix Tanpa Sampling OvO*

Pada gambar 4.2 menunjukkan *confusion matrix* dari pengujian tanpa *resampling* yang menggunakan pendekatan OvO. Hasil dari pengujian tersebut adalah sebagai berikut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 77 sampel, dan kelasahan prediksi sebanyak 1 sampel *leukimia with thrombocytopenia*, 3 sampel *normocytic normochromic anemia*, 1 sampel *other microcytic anemia*, dan 6 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 28 sampel, dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 5 sampel *normocytic hypochromic anemia*.
- 3) Kelas *leukimia* dapat memprediksi secara benar sebanyak 3 sampel, dan kesalahan prediksi sebanyak 1 sampel *healthy* serta 1 sampel *normocytic normochromic anemia*.

- 4) Kelas *leukimia with thrombocytopenia* dapat memprediksi secara benar sebanyak 1 sampel, dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 1 sampel *normocytic hypochromic anemia*.
- 5) Kelas *macrocytic anemia* tidak dapat melakukan prediksi dengan benar dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
- 6) Kelas *normocytic hypochromic anemia* dapat memprediksi secara benar sebanyak 41 sampel, dan kesalahan prediksi sebanyak 5 sampel *healthy*, 2 sampel *iron deficiency anemia* serta 8 sampel *normocytic normochromic anemia*.
- 7) Kelas *normocytic normochromic anemia* dapat memprediksi secara benar sebanyak 44 sampel dan kesalahan prediksi sebanyak 1 sampel *normocytic hypochromic anemia* serta 1 sampel *thrombocytopenia*.
- 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 4 sampel, dan kesalahan prediksi sebanyak 3 sampel *normocytic hypochromic anemia* serta 7 sampel *normocytic normochromic anemia*.
- 9) Kelas *thrombocytopenia* dapat memprediksi secara benar sebanyak 16 sampel, dan kesalahan prediksi sebanyak 1 sampel *healthy*.

## 2. Pengujian Tanpa *Sampling* OvR

Pada pengujian ini, model diuji tanpa *resampling* menggunakan algoritma SVM, sehingga hasilnya menggambarkan performa model pada data asli.

### a. *Classification Report*

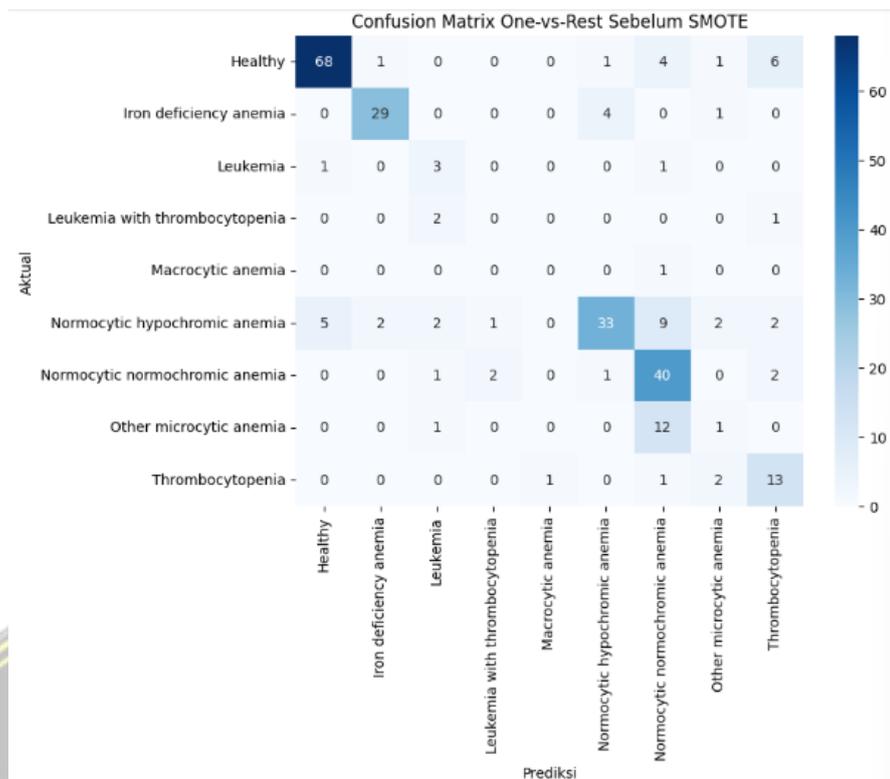
Tabel 4. 4 *Classification Report* Tanpa *Sampling* OvR

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.92	0.84	0.88	81
<i>Iron deficiency anemia</i>	0.91	0.85	0.88	34
<i>Leukemia</i>	0.33	0.60	0.43	5
<i>Leukemia with thrombocytopenia</i>	0.00	0.00	0.00	3

<i>Macrocytic anemia</i>	0.00	0.00	0.00	1
<i>Normocytic hypochromic anemia</i>	0.85	0.59	0.69	56
<i>Normocytic normochromic anemia</i>	0.59	0.87	0.70	46
<i>Other microcytic anemia</i>	0.14	0.07	0.10	14
<i>Thrombocytopenia</i>	0.54	0.76	0.63	17
<i>accuracy</i>			0.73	257
<i>macro avg</i>	0.48	0.51	0.48	257
<i>weighted avg</i>	0.75	0.73	0.73	257

Tabel 4.4 menyajikan hasil *classification report* dari pengujian tanpa resampling menggunakan pendekatan 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.

b. *Confusion Matrix*



Gambar 4.3 *Confusion Matrix* Tanpa *Sampling* OvR

Pada gambar 4.3 menampilkan *confusion matrix* dari pengujian tanpa resampling dengan pendekatan OvR. Adapun hasil dari pengujian tersebut adalah sebagai berikut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 68 sampel dan kesalahan prediksi sebanyak 1 sampel *normocytic hypochromic anemia*, 4 sampel *normocytic normochromic anemia*, 1 sampel *other microcytic anemia* serta 6 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 29 sampel dan kesalahan prediksi sebanyak 4 sampel *normocytic hypochromic anemia*, dan 1 sampel *other microcytic anemia*.
- 3) Kelas *leukimia* dapat memprediksi secara benar sebanyak 3 sampel dan kesalahan prediksi sebanyak 1 sampel *healthy* serta 1 sampel *other microcytic anemia*.

- 4) Kelas *leukimia with thrombocytopenia* dapat memprediksi secara benar sebanyak 2 sampel dan kesalahan prediksi sebanyak 1 sampel *thrombocytopenia*.
- 5) Kelas *macrocytic anemia* tidak dapat memprediksi secara benar dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
- 6) Kelas *normocytic hypochromic anemia* dapat memprediksi secara benar sebanyak 33 sampel dan kesalahan prediksi sebanyak 5 sampel *healthy*, 2 sampel *iron deficiency anemia*, 2 sampel *leukimia*, 1 sampel *leukimia with thrombocytopenia*, 9 sampel *normocytic normochromic anemia*, 2 sampel *other microcytic anemia*, dan 2 sampel *thrombocytopenia*.
- 7) Kelas *normocytic normochromic anemia* dapat memprediksi secara benar sebanyak 40 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia*, 2 sampel *leukimia with thrombocytopenia*, 1 sampel *normocytic hypochromic anemia*, dan 2 sampel *thrombocytopenia*.
- 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 1 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 12 sampel *normocytic normochromic anemia*.
- 9) Kelas *thrombocytopenia* dapat memprediksi secara benar sebanyak 13 sampel dan kesalahan prediksi sebanyak 1 sampel *macrocytic anemia*, 1 sampel *normochromic anemia*, serta 2 sampel *other microcytic anemia*.

### 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*.

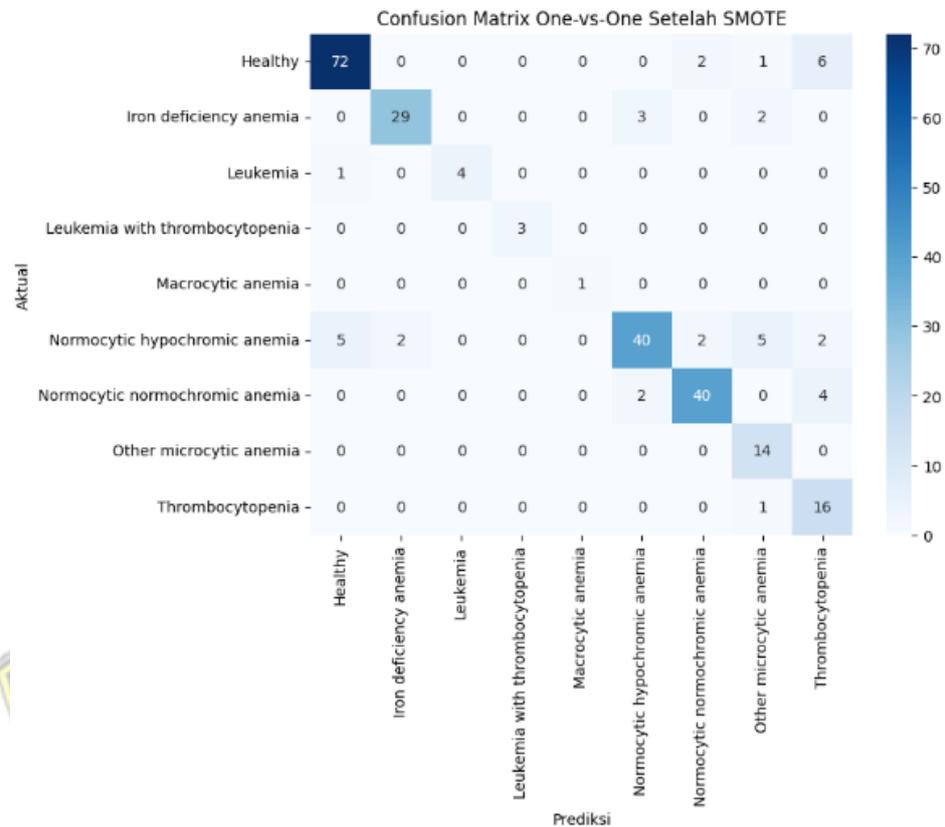
#### a. *Classification Report*

Tabel 4. 5 *Classification Report* *Oversampling* OvO

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.92	0.89	0.91	81
<i>Iron deficiency anemia</i>	0.94	0.85	0.89	34
<i>Leukemia</i>	1.00	0.80	0.89	5
<i>Leukemia with thrombocytopenia</i>	1.00	1.00	1.00	3
<i>Macrocytic anemia</i>	1.00	1.00	1.00	1
<i>Normocytic hypochromic anemia</i>	0.89	0.71	0.79	56
<i>Normocytic normochromic anemia</i>	0.91	0.87	0.89	46
<i>Other microcytic anemia</i>	0.61	1.00	0.76	14
<i>Thrombocytopenia</i>	0.57	0.94	0.71	17
<i>accuracy</i>			0.85	257
<i>macro avg</i>	0.87	0.90	0.87	257
<i>weighted avg</i>	0.88	0.85	0.86	257

Tabel 4.5 menampilkan hasil *classification report* dari pengujian dengan *oversampling* menggunakan pendekatan OvO. 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.

b. *Confusion Matrix*



Gambar 4. 4 *Confusion Matrix Oversampling OvO*

Pada gambar 4.4 menampilkan *confusion matrix* hasil pengujian *oversampling* menggunakan pendekatan OvO. Adapun hasil dari pengujian tersebut adalah sebagai berikut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 72 sampel dan kesalahan prediksi sebanyak 2 sampel *normocytic normochromic anemia*, 1 sampel *other microcytic anemia* serta 6 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 29 sampel dan kesalahan prediksi sebanyak 3 sampel *normocytic hypochromic anemia* serta 2 sampel *other microcytic anemia*.
- 3) Kelas *leukimia* dapat memprediksi sebanyak 4 sampel dan kesalahan prediksi sebanyak 1 sampel *healthy*.

- 4) Kelas *leukimia with trombocytopenia* dapat memprediksi sebanyak 3 sampel.
- 5) Kelas *macrocytic anemia* dapat memprediksi sebanyak 1 sampel.
- 6) Kelas *normocytic hypochromic anemia* dapat memprediksi sebanyak 40 sampel dan kesalahan prediksi sebanyak 5 sampel *healthy*, 2 sampel *iron deficiency anemia*, 2 sampel *normocytic normochromic anemia*, 5 sampel *other microcytic anemia* serta 2 sampel *thrombocytopenia*.
- 7) Kelas *normocytic normochromic anemia* dapat memprediksi sebanyak 40 sampel dan kesalahan prediksi sebanyak 2 sampel *normocytic hypochromic anemia* serta 4 sampel *thrombocytopenia*.
- 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 14 sampel.
- 9) Kelas *thrombocytopenia* dapat memprediksi secara benar sebanyak 16 sampel dan kesalahan prediksi sebanyak 1 sampel *other microcytic anemia*.

#### 4. Pengujian *Oversampling* OvR

Pada pengujian ini, model diuji dengan *oversampling* menggunakan SMOTE dan algoritma SVM. Data pada kelas minoritas ditambah sehingga data lebih seimbang. Diterapkannya pendekatan OvR di dalam pengujian ini karena data yang digunakan adalah *multiclass*.

##### a. *Classification Report*

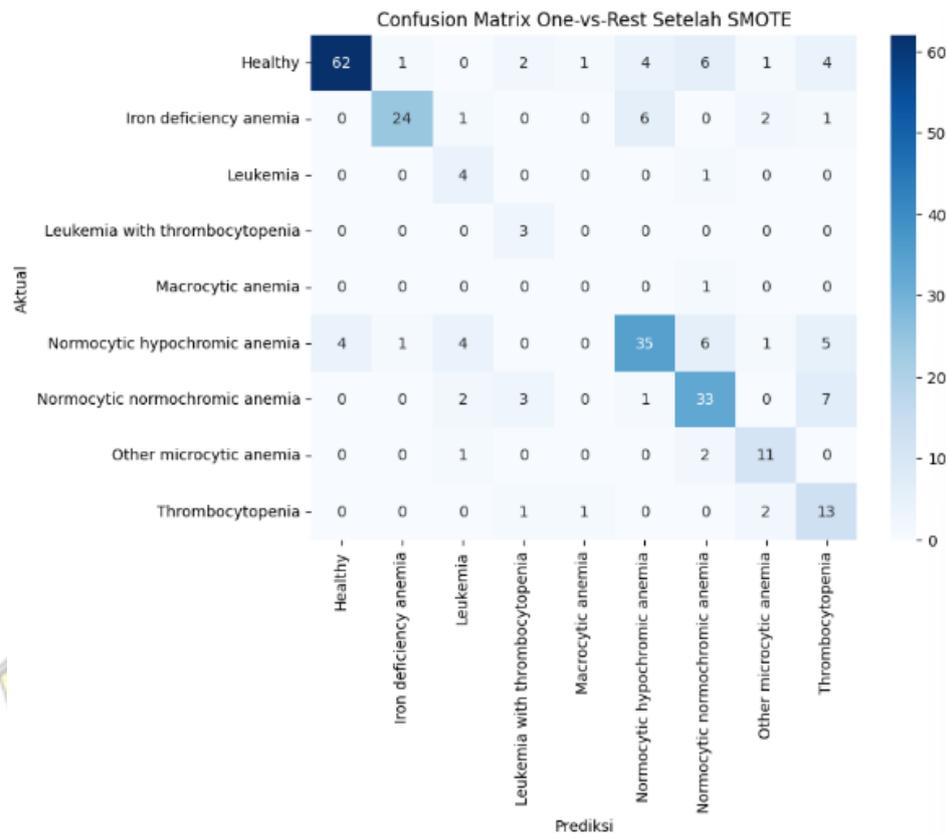
Tabel 4. 6 *Classification Report Oversampling* OvR

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.94	0.77	0.84	81
<i>Iron deficiency anemia</i>	0.92	0.71	0.80	34
<i>Leukemia</i>	0.33	0.80	0.47	5
<i>Leukemia with thrombocytopenia</i>	0.33	1.00	0.50	3
<i>Macrocytic anemia</i>	0.00	0.00	0.00	1

<i>Normocytic hypochromic anemia</i>	0.76	0.62	0.69	56
<i>Normocytic normochromic anemia</i>	0.67	0.72	0.69	46
<i>Other microcytic anemia</i>	0.65	0.79	0.71	14
<i>Thrombocytopenia</i>	0.43	0.76	0.55	17
<i>accuracy</i>			0.72	257
<i>macro avg</i>	0.56	0.68	0.58	257
<i>weighted avg</i>	0.78	0.72	0.74	257

Tabel 4.6 menampilkan hasil *classification report* dari pengujian menggunakan *oversampling* dengan pendekatan 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.

b. *Confusion Matrix*



Gambar 4.5 *Confusion Matrix Oversampling OvR*

Pada Gambar 4.5 menampilkan *confusion matrix* hasil pengujian *oversampling* dengan pendekatan OvR. Adapun hasil dari pengujian tersebut adalah sebagai berikut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 62 sampel dan kesalahan prediksi sebanyak 1 sampel *iron deficiency anemia*, 2 sampel *leukimia with thrombocytopenia*, 1 sampel *macrocytic anemia*, 4 sampel *normocytic hypochromic anemia*, 6 sampel *normocytic normochromic anemia*, 1 sampel *other microcytic anemia* serta 4 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 24 sampel dan kesalahan prediksi sebanyak 1 sampel

- leukimia*, 6 sampel *normocytic hypochromic anemia*, 2 sampel *other microcytic anemia* serta 1 sampel *thrombocytopenia*.
- 3) Kelas *leukimia* dapat memprediksi secara benar sebanyak 4 sampel dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
  - 4) Kelas *leukimia with thrombocytopenia* dapat memprediksi secara benar sebanyak 3 sampel.
  - 5) Kelas *macrocytic anemia* tidak dapat memprediksi secara benar sebanyak dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
  - 6) Kelas *normocytic hypochromic anemia* dapat memprediksi secara benar sebanyak 35 sampel dan kesalahan prediksi sebanyak 4 sampel *healthy*, 1 sampel *iron deficiency anemia*, 4 sampel *leukimia*, 6 sampel *normocytic normochromic anemia*, 1 sampel *other microcytic anemia* serta 5 sampel *thrombocytopenia*.
  - 7) Kelas *normocytic normochromic anemia* dapat memprediksi secara benar sebanyak 33 sampel dan kesalahan prediksi sebanyak 2 sampel *leukimia*, 3 sampel *leukimia with thrombocytopenia*, 1 sampel *normocytic hypochromic anemia* serta 7 sampel *thrombocytopenia*.
  - 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 11 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 2 sampel *normocytic normochromic anemia*.
  - 9) Kelas *thrombocytopenia* dapat memprediksi secara benar sebanyak 13 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia with thrombocytopenia*, 1 sampel *macrocytic anemia* serta 2 sampel *other microcytic anemia*.

##### 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.

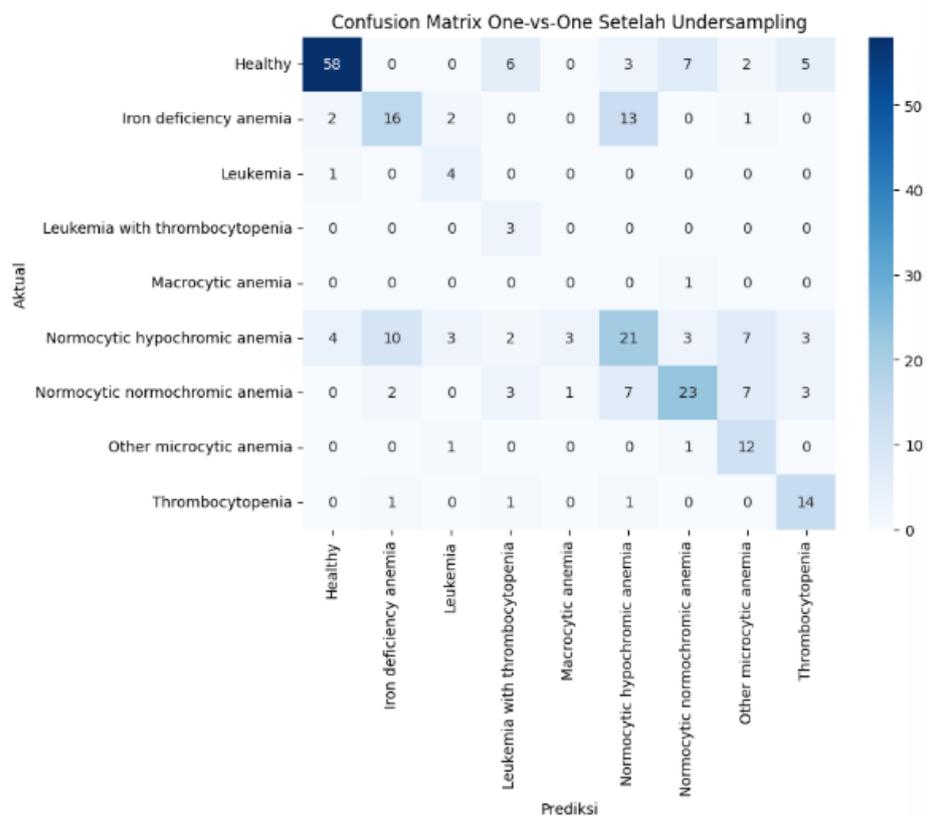
a. *Classification Report*

Tabel 4. 7 *Classification Report Undersampling OvO*

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.89	0.72	0.79	81
<i>Iron deficiency anemia</i>	0.55	0.47	0.51	34
<i>Leukemia</i>	0.40	0.80	0.53	5
<i>Leukemia with thrombocytopenia</i>	0.20	1.00	0.33	3
<i>Macrocytic anemia</i>	0.00	0.00	0.00	1
<i>Normocytic hypochromic anemia</i>	0.47	0.38	0.42	56
<i>Normocytic normochromic anemia</i>	0.66	0.50	0.57	46
<i>Other microcytic anemia</i>	0.41	0.86	0.56	14
<i>Thrombocytopenia</i>	0.56	0.82	0.67	17
<i>accuracy</i>			0.59	257
<i>macro avg</i>	0.46	0.62	0.49	257
<i>weighted avg</i>	0.64	0.59	0.60	257

Tabel 4.7 menunjukkan hasil *classification report* dari pengujian dengan *undersampling* menggunakan pendekatan OvO. Berdasarkan hasil pada tabel tersebut 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.

b. *Confusion Matrix*



Gambar 4. 6 *Confusion Matrix Undersampling OvO*

Pada gambar 4.6 menampilkan *confusion matrix* hasil pengujian *undersampling* dengan pendekatan OvO. Berikut hasil dari pengujian tersebut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 58 sampel dan kesalahan prediksi sebanyak 6 sampel *leukimia with thrombocytopenia*, 3 sampel *normocytic hypochromic anemia*, 7 sampel *normocytic normochromic anemia*, 2 sampel *other microcytic anemia* serta 5 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 16 sampel dan kesalahan prediksi sebanyak 2 sampel *healthy*, 2 sampel *leukimia*, 13 sampel *normocytuc hypochromic anemia* serta 1 sampel *other microcytic anemia*.

- 3) Kelas *leukimia* dapat memprediksi secara benar sebanyak 4 sampel dan kesalahan prediksi sebanyak 1 sampel *healthy*.
- 4) Kelas *leukimia with thrombocytopenia* dapat memprediksi secara benar sebanyak 3 sampel.
- 5) Kelas *macrocytic anemia* tidak dapat memprediksi secara benar dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
- 6) Kelas *normocytic hypochromic anemia* dapat memprediksi secara benar sebanyak 21 sampel dan kesalahan prediksi sebanyak 4 sampel *healthy*, 10 sampel *iron deficiency anemia*, 3 sampel *leukimia*, 2 sampel *leukimia with thrombocytopenia*, 3 sampel *macrocytic anemia*, 3 sampel *normocytic normochromic anemia*, 7 sampel *other microcytic anemia* serta 3 sampel *thrombocytopenia*.
- 7) Kelas *normocytic normochromic anemia* dapat memprediksi secara benar sebanyak 23 sampel dan kesalahan prediksi sebanyak 2 sampel *iron deficiency anemia*, 3 sampel *leukimia with thrombocytopenia*, 1 sampel *macrocytic anemia*, 7 sampel *normocytic hypochromic anemia*, 7 sampel *other microcytic anemia* serta 3 sampel *thrombocytopenia*.
- 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 12 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 1 sampel *normocytic normochromic anemia*.
- 9) Kelas *thrombocytopenia* dapat melakukan prediksi secara benar sebanyak 14 sampel dan kesalahan prediksi sebanyak 1 sampel *iron deficiency anemia*, 1 sampel *leukimia with thrombocytopenia* serta sampel *normocytic hypochromic anemia*.

#### 6. Pengujian *Undersampling* OvR

Pengujian ini menggunakan teknik *undersampling*, yang menerapkan algoritma SVM dan pendekatan OvR.

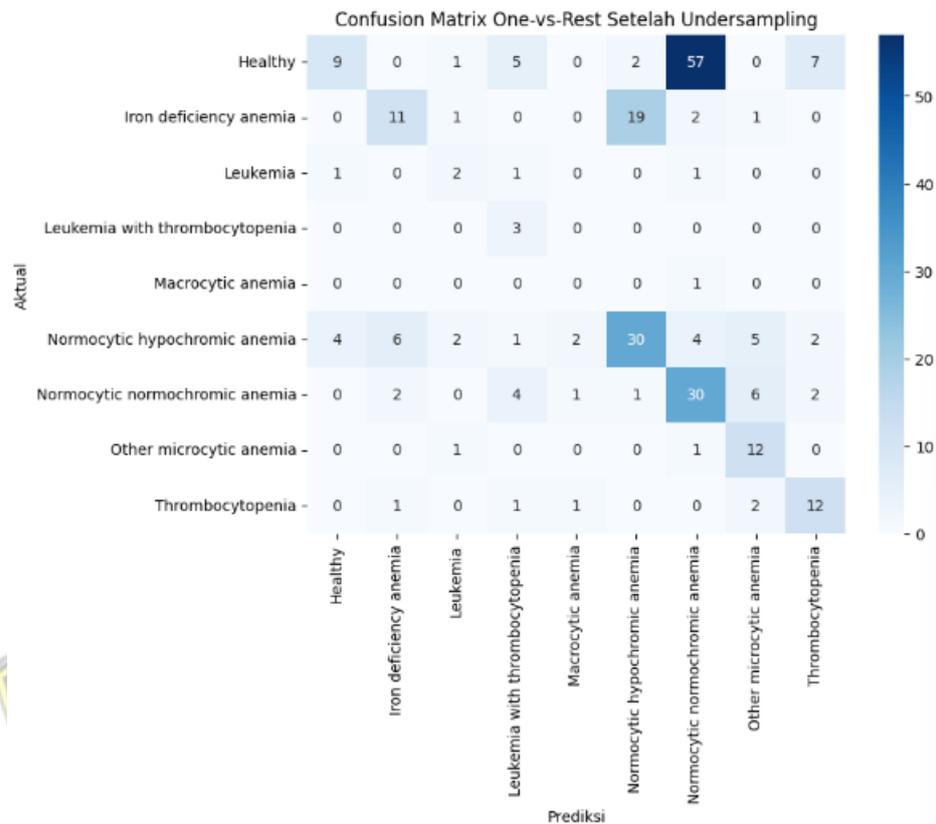
a. *Classification Report*

Tabel 4. 8 *Classification Report Undersampling OvR*

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
<i>Healthy</i>	0.64	0.11	0.19	81
<i>Iron deficiency anemia</i>	0.55	0.32	0.41	34
<i>Leukemia</i>	0.29	0.40	0.33	5
<i>Leukemia with thrombocytopenia</i>	0.20	1.00	0.33	3
<i>Macrocytic anemia</i>	0.00	0.00	0.00	1
<i>Normocytic hypochromic anemia</i>	0.58	0.54	0.56	56
<i>Normocytic normochromic anemia</i>	0.31	0.65	0.42	46
<i>Other microcytic anemia</i>	0.46	0.86	0.60	14
<i>Thrombocytopenia</i>	0.52	0.71	0.60	17
<i>accuracy</i>			0.42	257
<i>macro avg</i>	0.39	0.51	0.38	257
<i>weighted avg</i>	0.52	0.42	0.39	257

Tabel 4.8 menyajikan hasil *classification report* dari pengujian dengan *undersampling* menggunakan 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.

b. *Confusion Matrix*



Gambar 4. 7 *Confusion Matrix Undersampling OvR*

Pada gambar 4.7 menunjukkan *confusion matrix* dari pengujian *undersampling* dengan pendekatan OvR. Berikut adalah hasil dari pengujian tersebut:

- 1) Kelas *healthy* dapat memprediksi secara benar sebanyak 9 sampel dan kesalahan prediksi sebanyak 1 sampel leukimia, 5 sampel *leukimia with thrombocytopenia*, 2 sampel *normocytic hypochromic anemia*, 57 sampel *normocytic normochromic anemia* serta 7 sampel *thrombocytopenia*.
- 2) Kelas *iron deficiency anemia* dapat memprediksi secara benar sebanyak 11 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia*, 19 sampel *normocytic hypochromic anemia*, 2 sampel *normocytic normochromic anemia* serta 1 sampel *other microcytic anemia*.

- 3) Kelas *leukimia* dapat memprediksi secara benar sebanyak 2 sampel dan kesalahan prediksi sebanyak 1 sampel *healthy*, 1 sampel *leukimia with thrombocytopenia*, 1 sampel *leukimia with thrombocytopenia* serta 1 sampel *normocytic normochromic anemia*.
- 4) Kelas *leukimia with thrombocytopenia* dapat melakukan prediksi secara benar sebanyak 3 sampel.
- 5) Kelas *macrocytic anemia* tidak dapat melakukan prediksi secara benar dan kesalahan prediksi sebanyak 1 sampel *normocytic normochromic anemia*.
- 6) Kelas *normocytic hypochromic anemia* dapat melakukan prediksi secara benar sebanyak 30 sampel dan kesalahan prediksi sebanyak 4 sampel *healthy*, 6 sampel *iron deficiency anemia*, 2 sampel *leukimia*, 1 sampel *leukimia with thrombocytopenia*, 2 sampel *microcytic anemia*, 4 sampel *normocytic normochromic anemia*, 5 sampel *other microcytic anemia* serta 2 sampel *thrombocytopenia*.
- 7) Kelas *normocytic normochromic anemia* dapat melakukan prediksi secara benar sebanyak 30 sampel dan kesalahan prediksi sebanyak 2 sampel *iron deficiency anemia*, 4 sampel *leukimia with thrombocytopenia*, 1 sampel *macrocytic*, 6 sampel *other microcytic anemia* serta 2 sampel *thrombocytopenia*.
- 8) Kelas *other microcytic anemia* dapat memprediksi secara benar sebanyak 12 sampel dan kesalahan prediksi sebanyak 1 sampel *leukimia* serta 1 sampel *normocytic normochromic anemia*.
- 9) Kelas *thrombocytopenia* dapat melakukan prediksi secara benar sebanyak 12 sampel dan kesalahan prediksi sebanyak 1 sampel *iron deficiency anemia*, 1 sampel *leukimia with thrombocytopenia*, 1 sampel *macrocytic anemia* serta 2 sampel *other microcytic anemia*.

## 4.2 ANALISA PENELITIAN

### 4.2.1 Evaluasi

Berdasarkan hasil yang telah dijelaskan sebelumnya, perbandingan setiap pengujian dapat dilihat pada Tabel 4.7 berikut:

Tabel 4. 9 Evaluasi Perbandingan Pengujian

Teknik/Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
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
<b><i>Oversampling + OvO + SVM</i></b>	<b>0.85</b>	<b>0.87</b>	<b>0.90</b>	<b>0.87</b>
<i>Oversampling + OvR + SVM</i>	0.72	0.56	0.68	0.58
<i>Undersampling + OvO + SVM</i>	0.59	0.46	0.62	0.49
<i>Undersampling + OvR + SVM</i>	0.42	0.39	0.51	0.38

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.

### 4.2.2 Hasil Implementasi Model

Setelah memperoleh model yang terbaik dari perbandingan pengujian yang telah dilakukan, dengan hasil pengujian *oversampling* menggunakan pendekatan OvO yang menunjukkan akurasi lebih tinggi, model tersebut diimplementasikan ke dalam aplikasi *web* untuk mempermudah penggunaannya dalam proses klasifikasi.

## Anemia Classification One Vs One

**White Blood Cell ( $\times 10^9/L$ ):**

**Lymphocytes (%):**

**Neutrophils (%):**

**Lymphocytes ( $\times 10^9/L$ ):**

**Neutrophils ( $\times 10^9/L$ ):**

Gambar 4. 8 Tampilan UI Klasifikasi Anemia

**Red Blood Cell ( $\times 10^{12}/L$ ):**

**Hemoglobin (g/dL):**

**Hematocrit (%):**

**Mean Corpuscular Volume (fL):**

**Mean Corpuscular Hemoglobin (pg):**

**Mean Corpuscular Hemoglobin Concentration (g/dL):**

Gambar 4. 9 Tampilan UI Klasifikasi Anemia

**Platelet ( $\times 10^9/L$ ):**

**Platelet Distribution Width (fL):**

**Plateletcrit (%):**

**Predict**

Gambar 4. 10 Tampilan UI Klasifikasi Anemia

Pada gambar 4.8 hingga 4.10 menunjukkan tampilan *User Interface* untuk klasifikasi anemia menggunakan pendekatan OvO. Di dalam *web* tersebut, terdapat 14 parameter yang dapat digunakan untuk memasukkan data yang ingin diuji, antarlain *White Blood Cell*, *Lymphocytes*, *Neutrophils*, *Lymphocytes*, *Neutrophils*, *Red Blood Cell*, *Hemoglobin*, *Hematocrit*, *Mean Corpuscular Volume*, *Mean Corpuscular Hemoglobin*, *Mean Corpuscular Hemoglobin Concentration*, *Platelet*, *Platelet Distribution Width*, *Plateletcrit*. *Web* ini juga dilengkapi dengan fitur “*Predict*” untuk menghasilkan hasil pengujian setelah data dimasukkan.

Contoh uji coba untuk pengklasifikasian data dapat dilihat pada gambar 4.11 hingga 4.13. Proses dimulai dengan memasukkan data sesuai dengan dataset yang ada. Setelah data terinput, pengguna dapat menekan tombol “*Predict*” untuk memperoleh hasil klasifikasi.

## Anemia Classification One Vs One

White Blood Cell ( $\times 10^9/L$ ):

5.2

Lymphocytes (%):

19.7

Neutrophils (%):

72.4

Lymphocytes ( $\times 10^9/L$ ):

1.0

Neutrophils ( $\times 10^9/L$ ):

3.8

Gambar 4. 11 Uji Coba Klasifikasi

Red Blood Cell ( $\times 10^{12}/L$ ):

4.85

Hemoglobin (g/dL):

13.2

Hematocrit (%):

41.0

Mean Corpuscular Volume (fL):

84.7

Mean Corpuscular Hemoglobin (pg):

27.2

Mean Corpuscular Hemoglobin Concentration (g/dL):

32.1

Gambar 4. 12 Uji Coba Klasifikasi

Platelet ( $\times 10^9/L$ ):

Platelet Distribution Width (fL):

Plateletcrit (%):

Predict

**The diagnosis is: Healthy**

Gambar 4. 13 Uji Coba Klasifikasi

Pada Gambar 4.11 hingga 4.13 merupakan tampilan dari uji coba klasifikasi anemia dengan menginputkan data WBC 5.2, LYMP 19.7, NEUTP 72.4, LYMN 1.0, NEUTn 3.8, RBC 4.85, HGB 13.2, HCT 41.0, MCV 84.7, MCH 27.2, MCHC 32.1, PLT 181.0, PDW 10.0, PCT 0.15 dengan hasil klasifikasi *Healthy*.



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil dari penelitian yang berjudul “Implementasi Teknik *Resampling* untuk Mengatasi *Imbalanced Data* terhadap Klasifikasi Anemia menggunakan *Support Vector Machine (SVM)*” 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 mengklasifikasikan data *multiclass*.

#### 5.2 Saran

Penelitian ini diharapkan dapat dikembangkan dengan menerapkan teknik lain untuk mengatasi permasalahan ketidakseimbangan data, guna membandingkan efektifitasnya dalam meningkatkan akurasi model. Selain itu, disarankan untuk mengeksplorasi penggunaan algoritma lain yang dapat dipadukan dengan pendekatan *One vs One* dan *One vs Rest*.

## DAFTAR PUSTAKA

- Akbar, K., & Hayaty, M. (2020). Data Balancing untuk Mengatasi Imbalance Dataset pada Prediksi Produksi Padi Balancing Data to Overcome Imbalance Dataset on Rice Production Prediction. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 2(02), 1–14.
- 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. <https://djournals.com/klik>
- 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 Peker 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>
- Kasanah, A. N., Muladi, M., & 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(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>
- 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., Rahmadden, 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>
- 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. <https://www.kaggle.com/datasets/biswaranjanrao/an>
- Yohannes, R., & Al Rivan, M. E. (2022). Klasifikasi Jenis Kanker Kulit Menggunakan CNN-SVM. *Jurnal Algoritme*, 2(2), 133–144. <https://doi.org/10.35957/algoritme.v2i2.2363>