

KLASIFIKASI *MACHINE LEARNING* UNTUK ANEMIA MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DAN *RANDOM FOREST*

Nur Wahida Zam¹⁾, Irwan²⁾, Muh. Irwan³⁾

^{1,2,3}*Program Studi Matematika, Fakultas Sains dan Teknologi UIN Alauddin Makassar, Jl.
Sultan Alauddin No.63, Gowa, Sulawesi Selatan*

¹zamnurwahida@gmail.com, ²irwan.msi@uin-alauddin.ac.id, ³muhirwan@uin-alauddin.ac.id

Abstract

This research involves the classification of anemia using the support vector machine (SVM) and random forest (RF) methods. SVM has the ability to separate two classes of data by finding the optimal hyperplane. On the other hand, Random Forest can handle incomplete or missing data, reduce overfitting, minimize errors, and efficiently handle large training datasets. The classification of anemia in this study uses a 70% training data and 30% testing data split. The accuracy rate of classification using the SVM method is 94%. In comparison, the classification using the RF method achieves an accuracy rate of 98%.

Keywords: Anemia, Classification, Support Vector Machine (SVM), Random Forest (RF)

Abstrak

Penelitian ini berisi tentang klasifikasi menggunakan metode *support vector machine* (SVM) dan *random forest* (RF) pada penyakit anemia. SVM memiliki kemampuan untuk memisahkan dua kelas pada data dengan menemukan *hyperplane* yang optimal. Dan *Random Forest* dapat mengatasi data yang tidak lengkap atau *missing value*, mampu mengurangi *overfitting*, dan meminimalisir *error* yang dihasilkan, serta dapat mengatasi data *training* yang sangat besar. Klasifikasi penyakit anemia pada penelitian ini menggunakan pembagian data *training* sebesar 70% dan *testing* sebesar 30%. Tingkat akurasi hasil klasifikasi menggunakan metode SVM sebesar 94%. Untuk klasifikasi menggunakan metode RF memperoleh tingkat akurasi sebesar 98%.

Kata Kunci : Anemia, Klasifikasi, Support Vector Machine (SVM), Random Forest (RF)

Cara Menulis Sitasi: Zam, N., W., Irwan, & Irwan, M. (2025). Klasifikasi Machine Learning untuk Anemia Menggunakan Metode Support Vector Machine dan Random Forest. *Jurnal Edukasi dan Sains Matematika (JES-MAT)*, 11 (1), 62 - 76.

PENDAHULUAN

Anemia menjadi salah satu jenis kelainan darah yang banyak ditemui sebagai isu kesehatan masyarakat, terlebih pada negara berkembang. Prevalensi anemia di dunia menurut WHO lebih besar dari 40% atau sekitar 40% hingga 88%. Penyebab utama terjadinya peningkatan kasus anemia di dunia disebabkan oleh rendahnya tingkat konsumsi zat besi, sehingga menyebabkan rendahnya produksi *hemoglobin* dalam sel darah merah. Hal ini dikarenakan kurangnya asupan makanan bergizi serta pola makan yang buruk.

Perkembangan teknologi saat ini, sangat membantu dalam proses penyimpanan data yang terus bertambah terus menerus. Hal ini berdampak pada ukuran data menjadi semakin besar termasuk data kesehatan. Data dengan jumlah besar seringkali membutuhkan kemampuan untuk mengekstraksi dan memilih informasi penting yang terdapat dalam data. Oleh karena itu, digunakan salah satu teknik *data mining* yang dikembangkan beberapa tahun terakhir yaitu metode klasifikasi. Untuk mengolah data yang ada, *machine learning* memungkinkan komputer dalam mempelajari pola dan membangun model yang dapat digunakan dalam mengklasifikasikan data.

Terdapat suatu metode klasifikasi yang dapat diterapkan yaitu SVM dan RF. Pada Penelitian Achmad Rizal et al (2019) yang mengklasifikasikan wajah dengan SVM menghasilkan 90% deteksi wajah benar dan 10% salah. Penelitian yang dilakukan oleh Kasim Ahmad Anita & Sudarsono Muhammad (2019) mendapatkan akurasi sebesar 98% pada klasifikasi tingkat ekonomi pada penduduk pemerintah di

kecamatan Simpang Raya, Sulawesi Tengah menggunakan algoritma SVM. Utami Putri & Redi Susanto, (2020) dalam mengklasifikasikan jenis kayu dengan SVM menunjukkan akurasi sebesar 91,3% dan error sebesar 8,7%. Penelitian yang dilakukan Son & Kim (2021) mengenai klasifikasi emosi berbasis EGG untuk memverifikasi klip film emosional Korea dengan SVM menghasilkan akurasi sebesar 94,27%.

Penelitian yang dilakukan Harahap et al (2021) yang mengkalifikasi hasil diagnosa penyakit jantung menggunakan algoritma *Random Forest* memperoleh akurasi sebesar 85,3%. Penelitian Xu & Yin (2021) dalam pengaplikasian algoritma *Random Forest* pada pendidikan jasmani memperoleh akurasi sebesar 88,55%, presisi sebesar 88,21%, dan *recall* sebesar 95,8%. Khasanah et al (2021) dalam mengklasifikasikan kanker kulit dengan algoritma RF menghasilkan nilai akurasi sebesar 85,08%.

Penelitian di atas menjelaskan bahwa SVM memiliki kemampuan untuk memisahkan dua kelas pada data atau menemukan *hyperplane* yang optimal pada data. Dan *Random Forest* berkemampuan dalam mengatasi data yang tidak lengkap atau *missing value*, mampu mengurangi *overfitting*, dan meminimalisir *error* yang dihasilkan, serta dapat mengatasi data *training* yang sangat besar. Berdasarkan uraian tersebut, peneliti akan melakukan klasifikasi menggunakan metode SVM dan RF pada data kesehatan yaitu data anemia.

LANDASAN TEORI

Anemia

Menurut WHO anemia adalah keadaan di mana jumlah sel darah merah atau kadar *hemoglobin* berada di bawah tingkat normal. Anemia tahap 1 (ringan) ditandai dengan terjadinya penurunan *ferritine* dalam tubuh, lalu akan terjadi penurunan *serum iron* terjadi pada anemia tahap 2 (sedang). Selanjutnya anemia tahap 3 (berat) disusul dengan terjadinya penurunan konsentrasi *hemoglobin*, *hematocrit*, MCV (*Mean Corpuscular Volume*), MCH (*Mean Corpuscular Hemoglobin*), MCHC (*Mean Corpuscular Hemoglobin Concentration*) dalam tubuh.

Normalisasi Data

Normalisasi data diperlukan untuk mengurangi tingkat *error* pada *dataset* dengan cara mengolah data aktual menjadi nilai yang memiliki rantang interval [0,1]. Teknik normalisasi pada *dataset* ini menggunakan cara *min-max scaling* dengan persamaan sebagai berikut.

$$X' = \frac{(X - \min_X)}{(\max_X - \min_X)} \quad (1)$$

Keterangan:

X = Data yang akan dinormalisasikan

X' = Data setelah dinormalisasikan

$\max_X$ = Nilai maksimum *dataset*

$\min_X$ = Nilai minimum *dataset*

Machine Learning

Machine learning adalah mesin pembelajar yang terus menerus akan belajar dan meningkatkan kemampuan berdasarkan data lampau menggunakan algoritma khusus. Menurut Arthur Samuel *Machine Learning* adalah suatu bidang studi atau bidang keilmuan yang memampukan suatu komputer untuk mempelajari tanpa perlu

deprogram secara eksplisit. *Machine learning* bertujuan untuk menemukan pola dan hubungan data tersembunyi, mengekstrak informasi dan memberikan pengetahuan tentang data untuk memungkinkan keputusan berdasarkan informasi yang diperoleh.

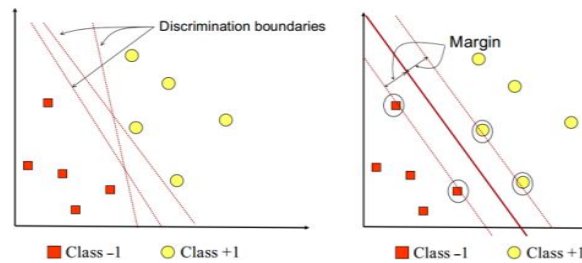
Klasifikasi

Klasifikasi adalah membedakan dua kelompok dengan sangat jelas. Klasifikasi adalah teknik mengelompokkan data yang berarti mengumpulkan suatu objek ataupun entitas yang sama pada data dalam satu kelas dan memisahkan objek ataupun entitas yang tidak sama pada kelas yang berbeda. Klasifikasi akan membangun model berdasarkan data pelatihan yang tersedia, lalu menggunakan model tersebut untuk melakukan klasifikasi pada data baru. (Utomo & Mesran, 2020). Tugas utama klasifikasi adalah menetapkan sebuah label kelas diantara kemungkinan kategori untuk sampel yang diwakili oleh sekumpulan fitur atribut dan diselesaikan oleh model klasifikasi.

Support Vector Machine (SVM)

Support vector machine merupakan bagian dari metode *supervised learning* yang digunakan untuk mengklasifikasikan data. SVM pertama kali dikenalkan oleh Vapnik tahun 1992 bersama Bernhard Boser dan Isabelle Guyon. Tujuan utama SVM adalah untuk memaksimalkan *margin* sehingga dapat mengklasifikasikan pola yang diberikan dengan benar, yaitu semakin besar ukuran *margin*, akan semakin tepat dalam mengklasifikasikan pola. *Margin* adalah jarak antara *hyperplane* optimal dengan titik-titik terdekat dari setiap kelas. *Hyperplane* optimal adalah *hyperplane* yang dipilih dari himpunan *hyperplane* untuk mengklasifikasikan pola yang

memaksimalkan *margin*. *Hyperplane* yaitu jarak *hyperplane* ke titik terdekat dari setiap pola yaitu *support vector*.



Gambar 1 : Mencari *Hyperplane* Terbaik

Pada **Gambar 1** terlihat bahwa ada dua kelas yaitu kelas positif dan negatif dan ditandai dengan lingkaran berwarna kuning dan kotak berwarna merah. Pada Gambar tersebut, terdapat beberapa alternatif garis pemisah yang disebut *discrimination boundaries* pada gambar sebelah kiri. Pola yang lebih dekat dengan garis *discrimination boundaries* yaitu gambar sebelah kanan yang ditandai dengan lingkaran berwarna hitam yang disebut *support vector*. Gambar sebelah kanan garis tegas yang berwarna merah menampilkan *hyperplane* terbaik berada antara kelas positif dan kelas negatif.

Margin ditentukan dengan memaksimalkan nilai jarak *hyperplane* dengan *support vector* seperti pada rumus :

$$\frac{1}{\|w\|} \quad (2)$$

Untuk mencari *margin* terbesar:

$$\frac{2}{\|w\|} \quad (3)$$

Untuk menentukan *hyperplane* kedua kelas :

$$\frac{1}{2} \|w\| \quad (4)$$

Dengan syarat :

$$y_i(x_i \cdot w + b - 1) \geq 0, \quad i = 1, \dots, n \quad (5)$$

Solusi untuk mengoptimalkan *hyperplane* dapat diselesaikan dengan *Quadratic Program* dengan meminimalkan *hyperplane* kedua kelas menggunakan *Lagrange Multiplier*. Pada kasus klasifikasi menggunakan SVM ini mungkin saja *hyperplane* tidak dapat memisahkan kelas dengan sempurna atau terjadi *misclassification*, sehingga dilakukan penambahan variabel *slack* yang kemudian diminimumkan. Variabel *slack* dikenal juga dengan *soft margin hyperplane* yang merupakan ukuran sebuah kesalahan pada klasifikasi. Persamaan yang digunakan dinamakan *Primal Lagrange*.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (w^T x_i + b) - 1) - C \sum_{i=1}^n \mu_i \xi_i \quad (6)$$

Keterangan :

L = *Primal Lagrange*

α_i = Nilai *support vector* ke- i

w^T = Transpose w

n = Banyak *Support Vector*

C = Koefisien penentuan besaran akibat kesalahan klasifikasi

μ_i = *Lagrange Multiplier* yang dikalikan dengan *non-negative* ξ_i

ξ_i = *Soft Margin Hyperplane*

Nilai optimasi ini dihitung dengan dengan memaksimalkan L terhadap α dan meminimalkan L terhadap w dan b . Selanjutnya, untuk memprediksi kelas pada data, dapat digunakan persamaan model berikut :

$$D(X) = \sum_{i=1}^n \alpha_i y_i k(x_i, x) + b \quad (7)$$

Keterangan :

$D(X)$ = Fungsi keputusan klasifikasi

i = banyak data *support vector*

Fungsi kernel

Untuk menyelesaikan masalah non-linear pada SVM digunakan fungsi kernel. Octaviani et al (2014) menyatakan bahwa Menurut Prasetyo (2012) macam-macam *fungsi kernel* diantaranya:

1. Kernel Gaussian *Radial Basic Function* (RBF)

$$K(x_i, x) = \exp(-\gamma ||x_i - x||^2), \gamma > 0 \quad (8)$$

2. Kernel *Polynomial*

$$K(x_i, x) = (\gamma x_i^T x + r)^d, \gamma > 0 \quad (9)$$

Keterangan :

K = Fungsi Kernel

γ = Koefisien

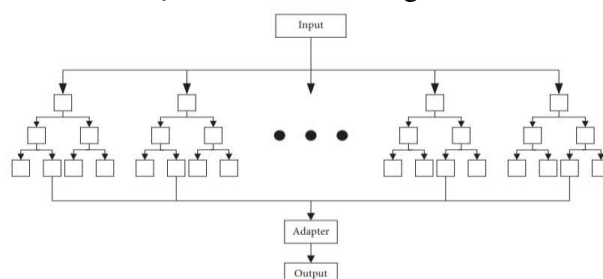
Menurut Hsu dan Lin (2016) menyatakan bahwa, *fungsi kernel radial basic function*

(RBF) adalah *fungsi kernel* yang direkomendasikan. Hal ini karena RBF mampu melihat hubungan antara kelas dengan variabel yang non-linear serta dapat menggunakan parameter yang berbeda.

Random Forest

Random Forest merupakan kumpulan model prediktif berbasis pohon keputusan, di mana setiap pohon menggunakan vektor acak yang dihasilkan secara independen dan terdistribusi secara seragam di seluruh pohon. *Random forest* berdasarkan pohon keputusan berbasis ensemble adalah metode pembelajaran ensemble yang banyak digunakan. *Random Forest* adalah algoritma klasifikasi yang diusulkan oleh Leo Breiman yang berisi beberapa *decision tree*. Proses klasifikasinya dilakukan dengan memilih hasil berdasarkan suara terbanyak dari keputusan yang diperoleh dari pohon keputusan yang dikembalikan.

Algoritma *Random Forest* menggabungkan beberapa *decision tree* secara bersamaan. Setiap dataset dipilih secara acak sebagai input. **Gambar 2** menunjukkan bentuk spesifik dari algoritma *Random Forest*, di mana kombinasi tersebut memilih sebagian besar hasil klasifikasi sebagai hasil akhir dalam klasifikasi.



Gambar 2 : Algoritma *Random Forest*

Proses *Classification and Regression Tree* (CART) sama dengan pembentukan *decision tree* dalam algoritma

random forest, akan tetapi *random forest* tidak terdapat pemangkasan. Pada pemilihan fitur untuk setiap pohon yang digunakan pada *random forest* digunakan rumus :

$$m = \sqrt{M} \quad (10)$$

Dengan M merupakan jumlah keseluruhan fitur yang ada dan m adalah jumlah fitur yang akan digunakan untuk membangun setiap pohon. Perhitungan nilai *Information Gain* yang digunakan sebagai *split point* atau titik pemisah antar simpul yang dihasilkan, dapat dihitung menggunakan persamaan berikut.

$$\text{GAIN}(S, A) = \text{Entropy}(S)$$

$$- \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (11)$$

dengan

$$\text{Entropy}(S) = \sum_i^c -p_i \log_2 p_i \quad (12)$$

Keterangan :

$\text{GAIN}(S, A)$: nilai *information gain* dari fitur

$\text{Entropy}(S)$: nilai sampel entropi

A : fitur yang digunakan
 v : nilai yang mungkin untuk A
 $\text{values}(A)$: himpunan nilai yang mungkin dari A
 S_v : jumlah nilai sampel dari v
 $|S|$: jumlah seluruh sampel data
 p_i : jumlah sampel pada kelas ke- i
 c : jumlah nilai dalam kelas klasifikasi
 $\text{Entropy}(S_v)$: sampel entropi dengan nilai dari v

METODE PENELITIAN

Dataset

Pada penelitian ini digunakan data sebanyak 15.000 baris dengan 8 kolom, dengan 7 kolom merupakan variabel x atau atribut dan 1 kolom merupakan variabel y atau target. Berikut adalah atribut yang digunakan pada penelitian ini :

Tabel 1. Atribut Penelitian

Atribut	Definisi Operasional
<i>Hemoglobin</i> (X1)	Konsentrasi <i>hemoglobin</i> normal untuk pria adalah 13 <i>gr/dL</i> dan wanita sebesar 12 <i>gr/dL</i>
<i>Hematocrit</i> (X2)	Konsentrasi <i>hematocrit</i> normal berada sekitar 38 – 42%
MCH (<i>Mean Corpuscular Hemoglobin</i>) (X3)	Konsentrasi MCH normal sekitar 26 – 33 <i>pg</i>
MCHC (<i>Mean Corpuscular Hemoglobin Concentration</i>) (X4)	Konsentrasi MCHC normal berada sekitar 33 – 37 <i>gr/dL</i>
MCV (<i>Mean Corpuscular Volume</i>) (X5)	Konsentrasi MCV normal sebesar 77 – 98 <i>fL</i>
SD (<i>Serum Iron</i>) (X6)	Konsentrasi <i>serum iron</i> yang normal dalam tubuh sekitar 65 – 165 $\mu\text{g/dL}$
<i>Ferritin</i> (X7)	Konsentrasi <i>ferritin</i> normal untuk pria adalah 15 – 200 $\mu\text{g/L}$ dan wanita sebesar 15 – 150 $\mu\text{g/L}$
Status Anemia (Y1)	Anemia tahap 1 (ringan) ditandai dengan terjadinya penurunan <i>ferritin</i> dalam tubuh, lalu akan terjadi penurunan <i>serum iron</i> terjadi pada anemia tahap 2 (sedang). Selanjutnya anemia tahap 3 (berat) disusul dengan terjadinya penurunan konsentrasi

	<i>Hemoglobin, hematocrit, MCV, MCH, MCHC dalam tubuh (Amalia & Tjiptaningrum, 2016).</i> Keterangan : 1 = Anemia 0 = Tidak Anemia
--	---

Prosedur

Prosedur pada penelitian ini :

- Pengumpulan data pada web :
<https://www.kaggle.com/datasets/se-rhathoca/anemia-disease>.a web :
- Membagi *dataset* menjadi dua bagian, yaitu :
 - Data *training*
 - Data *testing*
- Melakukan klasifikasi data menggunakan metode *Support Vector Machine* (SVM)
 - Inisialisasi parameter
 - Melakukan klasifikasi dengan metode SVM menggunakan data training
 - Menentukan banyaknya nilai support vector
 - Menghitung nilai bias
 - Evaluasi model yang dihasilkan menggunakan data testing
 - Evaluasi Performa menggunakan Confusion Matrix
- Melakukan klasifikasi data menggunakan metode Random Forest
 - Inisialisasi parameter
 - Memilih subset sebanyak m dari subset total yaitu M , dengan syarat $m \leq M$.
 - Menentukan nilai Information Gain terbaik
 - Melatih model RF menggunakan data training
 - Evaluasi model yang dihasilkan menggunakan data testing
 - Evaluasi Performa menggunakan Confusion Matrix dengan nilai *accuracy*, *precision*, *recall*, dan *F1-Score*.
- Melakukan uji statistik untuk mengetahui perbedaan akurasi SVM dan RF
- Penarikan kesimpulan

HASIL PENELITIAN DAN PEMBAHASAN

Statistik Deskriptif

Statistik deskriptif dari variabel-variabel yang digunakan disajikan pada **Tabel 2** di bawah ini.

Tabel 2. Statistik Deskriptif

Variabel	Min	Standar Deviasi	Median
HGB	2,99	2,097	12,82
HCT	7,21	5,602	39,135
MCV	49,93	7,564	83,4
MCH	12,55	3,386	27,965
MCHC	18,07	2,058	32,9
Serum Iron	5	39,853	61,635
Ferritte	0,5	407,38	40,2

Variabel	Mean	Max
HGB	12,78	22,45
HCT	38,958	67,4
MCV	83,425	125,3
MCH	27,375	139
MCHC	32,9	168
Serum Iron	67,635	461,64
Ferritte	129,762	27332

Berdasarkan **Tabel 2**, terlihat bahwa nilai minimum terendah adalah 0,5 pada variabel *ferritte* dan dilai maksimum paling tinggi adalah 27332 juga pada variabel *ferritte* .

Normalisasi Data

Normalisasi digunakan data untuk menyeragamkan skala (rentang nilai) yang dimiliki setiap variabel. Berikut adalah data hasil Normalisasi.

Tabel 3. Data Hasil Normalisasi

No	HGB	HCT	MCH	MCHC
1	0,498	0,504	0,134	0,105
2	1,504	0,493	0,134	0,111
...	...	...	...	...
15211	0,55	0,565	0,119	0,102
15212	0,555	0,546	0,129	0,108
No	MCV	SD	Ferritte	
1	0,494	0,208	0,007	
2	0,465	0,245	0,002	
...	...	...	...	
15211	0,441	0,212	0,0007	
15212	0,451	0,162	0,0004	

Membagi Data *Training* dan Data *Testing*

Pada penelitian ini menggunakan data sebanyak 70% untuk *training* dan 30% untuk *testing* dari total 15.212 baris data.

Melakukan Klasifikasi Menggunakan Metode *Support Vectro Machine* (SVM)

Pada penelitian ini, pertama kali dilakukan klasifikasi menggunakan metode SVM.

Inisialisasi Parameter

Berikut adalah parameter-parameter yang akan dilatih untuk melakukan parameter terbaik dalam pengklasifikasian menggunakan metode SVM.

Tabel 4: Inisialisasi Paramter SVM

Gamma	Cost
0,1	0,1
1	1
10	10
100	100
1000	1000

Hasil Klasifikasi

Dalam proses *training* dilakukan pencarian kombinasi parameter terbaik yang dapat digunakan menggunakan metode SVM. Berikut adalah 10 kombinasi data

parameter terbaik menggunakan metode *grid search* dengan nilai akurasi masing-masing.

T

abel 5: 10 Kombinasi Parameter Terbaik

Rank	Parameter			Akurasi
	C	γ	Kenel	
1	1000	10	RBF	0,932
2	10	100	RBF	0,931
3	100	100	RBF	0,931
4	1	100	RBF	0,929
5	100	10	RBF	0,928
6	1000	100	RBF	0,926
7	10	10	RBF	0,925
8	1000	1	RBF	0,923
9	0,1	100	RBF	0,922
10	1	10	RBF	0,921

Pada **Tabel 5** dapat dilihat bahwa nilai akurasi tertinggi terdapat pada rank 1 baris pertama yaitu sebesar 0,9322877, menunjukkan bahwa kombinasi berisi parameter terbaik yang akan digunakan.

Evaluasi Performa

Evaluasi performa hasil klasifikasi dapat dilihat dari beberapa aspek, berikut adalah *confusion matrix* dari hasil klasifikasi menggunakan data *testing*.

Tabel 6: *Confusion Matrix*

Kelas Aktual	Kelas Prediksi		Total
	Negatif	Positif	
Negatif	3205	99	3304
Positif	170	1090	1260
Total	3375	1189	4564

Berdasarkan *confusion matrix* pada **Tabel 6** dapat dilihat bahwa terdapat 1.090 pengamatan yang dikategorikan *true positive* yang artinya terkena anemia dan 3.205 pengamatan yang dikategorikan *false negative* atau tidak terkena anemia. Selanjutnya menghitung *accuracy*, *precision*, *recall* hasil klasifikasi menggunakan metode SVM, sebagai berikut.

$$accuracy = \frac{1090}{1090 + 3205 + 99 + 170} = 0,94$$

$$precision = \frac{1090}{1090 + 99} = 0,91674$$

$$recall = \frac{1090}{1090 + 170} = 0,865$$

$$\begin{aligned} F1 - Score &= 2 \times \frac{precision \times recall}{precision + recall} \\ &= 2 \times \frac{0,91674 \times 0,865}{0,91674 + 0,865} \\ &= 0,8901187 \end{aligned}$$

Berdasarkan hasil perhitungan diperoleh *accuracy* sebesar 94% menunjukkan bahwa

model benar dalam 94% dari seluruh prediksi yang dibuat. *Precision* sebesar 91,674% berarti dari semua pasien yang diprediksi menderita anemia, 91,6% benar-benar memiliki penyakit tersebut. *Recall* sebesar 86,5% berarti dari semua pasien yang benar-benar memiliki anemia, model berhasil mendeteksi 86,5% dari mereka. *F-1 score* sebesar 89,01187% artinya model cukup seimbang dalam mengenali pasien anemia (*recall*) dan menghindari kesalahan diagnosis (*precision*).

Melakukan Klasifikasi Menggunakan Random Forest (RF)

Kemudian dilakukan klasifikasi menggunakan metode RF.

Inisialisasi Parameter

Berikut adalah parameter-parameter yang akan dilatih untuk melakukan parameter terbaik dalam pengklasifikasian menggunakan metode RF.

Tabel 7 : Inisialisasi Parameter RF

n estimator	max depth	min samples split	min samples leaf
100	10	8	2
200	50	10	5
500	60	12	10
1000	100	15	15

Hasil Klasifikasi

Berikut adalah 10 kombinasi data parameter terbaik menggunakan metode

grid search dengan nilai akurasi masing-masing.

Tabel 8: 10 Kombinasi Parameter Terbaik

Rank	n estimator	max depth	min sample split
1	200	50	8
2	200	100	8

3	200	60	8
4	1000	50	8
5	1000	60	8
6	1000	100	8
7	1000	50	15
8	1000	60	15
9	1000	100	15
10	500	50	15

Rank	min sample leaf	Akurasi
1	2	0,95398178
2	2	0,95398178
3	2	0,95398178
4	2	0,95388789
5	2	0,95388789
6	2	0,95388789
7	5	0,9538878
8	5	0,9538878
9	5	0,9538878
10	5	0,95379395

Kombinasi nilai parameter yang digunakan adalah kombinasi yang berada pada rank 1 yang mana telah dilakukan 5-fold cross validation dan menunjukkan bahwa rank 1 adalah parameter terbaik dengan akurasi sebesar 0,95398178. Hal ini berarti bahwa

parameter pada baris pertamalah yang akan digunakan.

Evaluasi Performa

Evaluasi performa hasil klasifikasi dapat dilihat dari beberapa aspek, berikut adalah *confusion matrix* hasil klasifikasi menggunakan metode RF.

Tabel 9 : Confusion Matrix SVM

Kelas Aktual	Kelas Prediksi		Total
	Negatif	Positif	
Negatif	3291	13	3304
Positif	71	1188	1260
Total	3363	1201	4564

Berdasarkan *confusion matrix* pada **Tabel 9** dapat dilihat bahwa terdapat 1.188 pengamatan yang dikategorikan *true positive* yang artinya terkena anemia dan 3.291 pengamatan yang dikategorikan *false*

negative atau tidak terkena anemia. Selanjutnya menghitung *accuracy*, *precision*, *recall*, *F1-Score* hasil klasifikasi menggunakan metode RF, sebagai berikut.

$$accuracy = \frac{1188 + 3291}{1188 + 3291 + 13 + 72} = 0,98$$

$$precision = \frac{1188}{1188 + 13} = 0,989$$

$$recall = \frac{1188}{1188 + 72} = 0,94$$

$$\begin{aligned} F1 - Score &= 2 \times \frac{precision \times recall}{precision + recall} \\ &= 2 \times \frac{0,989 \times 0,94}{0,989 + 0,94} \\ &= 0,965 \end{aligned}$$

Berdasarkan hasil perhitungan diperoleh *accuracy* sebesar 98% berarti model dapat mengklasifikasikan 98% dari seluruh pasien dengan benar, baik pasien anemia maupun yang tidak. *Precision* sebesar 98,9% artinya dari semua pasien yang diprediksi positif (menderita anemia), 98,9% benar-benar memiliki anemia, sementara 1,1% lainnya sebenarnya tidak (*False Positive*). *Recall* sebesar 94% berarti dari semua pasien yang benar-benar memiliki anemia, 94% berhasil dideteksi sebagai anemia oleh model, sementara 6% lainnya tidak terdeteksi (*False Negative*). *F1 score* sebesar 96,5%, *F1-score* yang tinggi menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara mendeteksi anemia (*recall*) dan menghindari kesalahan diagnosis (*precision*).

KESIMPULAN

Berdasarkan hasil analisis kinerja kedua model, metode *random forest* menunjukkan performa yang lebih baik dibandingkan metode *support vector machine* dalam mendeteksi anemia. Hal ini dibuktikan pada tingkat *accuracy*, *precision*, *recall*, dan *F1-score* yang dimiliki *random forest* lebih tinggi dibanding *support vector machine*. *Random forest* mampu mengklasifikasikan hampir seluruh pasien dengan benar, sangat jarang memberikan

prediksi *false positive*, cukup baik dalam menangkap kasus anemia yang sebenarnya, serta memiliki keseimbangan yang sangat baik antara *precision* dan *recall*.

Oleh karena itu, RF lebih direkomendasikan untuk digunakan dalam klasifikasi anemia karena memiliki keseimbangan yang lebih baik antara mengidentifikasi pasien anemia dan menghindari kesalahan diagnosis.

DAFTAR PUSTAKA

- Abu-Baker, N. N., Eyadat, A. M., & Khamaiseh, A. M. (2021). The impact of nutrition education on knowledge, attitude, and practice regarding iron deficiency anemia among female adolescent students in Jordan. *Heliyon*, 7(2).
- Achmad Rizal, R., Sanjaya Girsang, I., & Apriyadi Prasetyo, S. (2019). Klasifikasi Wajah Menggunakan Support Vector Machine (SVM). *Riset Dan E-Jurnal Manajemen Informatika Komputer*, 3(2).
- Aldi, M. W. P., Jondri, J., & Aditsania, A. (2018). Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin. *eProceedings of Engineering*, 5(2).
- Amalia, A., & Tjiptaningrum, A. (2016). Diagnosis dan Tatalaksana Anemia Defisiensi Besi. *MAJORITY*, 166-169.
- Bjerre, E., Fienen, M. N., Schneider, R., Koch, J., & Højberg, A. L. (2022). Assessing spatial transferability of a random forest metamodel for predicting drainage fraction. *Journal of Hydrology*, 612, 128177.
- Breiman, L. (2001). Random Forests. *Machine learning*, 45, 5-32.
- Burges, C. J. (1998). A tutorial on support vector machine for pattern recognition. *Data mining and knowledge discovery*, 2(2), 955-974.

- Erdiansyah, U., Irmansyah Lubis, A., & Erwansyah, K. (2022). Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil. *Jurnal Media Informatika Budidarma*, 6(1), 208. <https://doi.org/10.30865/mib.v6i1.3373>
- Firasari, E., Khasanah, N., Khultsum, U., Kholifah, D. N., Komarudin, R., & Widyastuty, W. (2020). Comparison of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithm for the Classification of the Poor in Recipients of Social Assistance. *Journal of Physics: Conference Series*, 1641(1).
- Fitriany, J., Saputri, A. I., Ilmu, S., & Anak, K. (2018a). Anemia Defisiensi Besi. In *Jurnal Averrous*, 4(2).
- Harahap, A. H., Muttaqin, I., Malik, B. A., Irfan, M., Imam, N., Thariq, M., Bilhaq, S., Nur, A. A., Lutfia, S., & Agustini, D. (2021). Klasifikasi Diagnosa Penyakit Jantung menggunakan Algoritma Random Forest. *Gunung Djati Conference Series*, 3.
- Hsu, C., Chang, C., & Lin, L. (2016). A Practical Guide to Support Vector Classification. Departement of Computer Science.
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.
- Kasim Ahmad Anita, & Sudarsono Muhammad. (2019). Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi Penduduk Penerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah. In *Seminar Nasional Aptikom (SEMNASTIK) 2019*, 568–573.
- Khaidir, M. (2007). Anemia defisiensi besi. *Jurnal Kesehatan Masyarakat Andalas*, 2(1), 140-145.
- Khan, Z. A., Khan, T., Bhardwaj, A., Aziz, S. J., & Sharma, S. (2018). Underweight as a risk factor for nutritional anaemia—A cross-sectional study among undergraduate students of a medical college of haryana. *Indian journal of community health*, 30(1), 63-69.
- Khasanah, N., Komarudin, R., Afni, N., Maulana, Y. I., & Salim, A. (2021). Skin Cancer Classification Using Random Forest Algorithm. *SISFOTENIKA*, 11(2), 137.
- Lestari, D. T., Khomsan, A., Anwar, F., & Damayanti, D. S. (2022). Protein Intake and Menstruation with Anemia Status in Young Women Based on Economic Status in Cianjur District. *AI GIZZAI: PUBLIC HEALTH NUTRITION JOURNAL*, 75-84.
- Lin, S. W., Ying, K. C., Chen, S. C., & Lee, Z. J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert systems with applications*, 35(4), 1817-1824.
- Liu, Y., Xie, Z., & Li, M. (2023). AUC Optimization from Multiple Unlabeled Datasets.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9, 381-386.
- Meilanie, A. D. (2019). Different of Hematocrit Value Microhematocrit Methods and Automatic Methods In Dengue Hemorrhagic Patients With Hemoconcentration. *Journal of Vocational Health Studies*, 67-71.
- Michael, A., Palelleng, S., Damayanti, I. D., & Rusman, J. (2023). Kombinasi Pretrained Model dan Random Forest Pada Klasifikasi Bakso Mengandung Boraks dan Non-Boraks Berbasis Citra. *Teknika*, 12(1), 27-32.
- Muhammad, A., & Sianipar, O. (2005). Penentuan Defisiensi Besi Anemia Penyakit Kronis Menggunakan Peran Indeks sTfR-F. *Indonesian Jurnal of Clinical and Medical Laboratory*, 9-15.
- Muslim, M. A. (2020). Support vector machine (svm) optimization using grid search and unigram to improve e-commerce review accuracy. *Journal of Soft Computing Exploration*, 1(1), 8-15.

- Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1.
- Octaviani, P. A., Wilandari, Y., & Ispriyanti, D. (2014). Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. *Jurnal Gaussian*, 3(4), 811-820.
- Pradhan, A. (2012). Support Vector Machine-A Survey. *International Journal of Emerging Technology and Advanced Engineering*, 81-85.
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3), e1301.
- Putra, A. F., & Rahman, Y. A. (2022). Pendekatan Diagnosis Anemia pada Pasien dengan Penyakit Ginjal Kronik. *Majority*, 60-64.
- Ramadhan, M. M., Sitanggang, I. S., Nasution, F. R., & Ghifari, A. (2017). Parameter tuning in random forest based on grid search method for gender classification based on voice frequency. *DEStech transactions on computer science and engineering*, 10(2017).
- Ratnawati, L., & Sulistyningrum, D. R. (2020). Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel. *Jurnal Sains dan Seni ITS*, 8(2), A71-A77.
- Religia, Y., Nugroho, A., & Hadikristanto, W. (2021). Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 187-192.
- Ritonga, A. S., & Purwaningsih, E. S. (2018). Penerapan Metode Support Vector Machine (Svm) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding). *Jurnal Ilmiah Edutic*, 17-25.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75-82.
- Rusman, J., Haryati, B. Z., & Michael, A. (2023). Optimisasi Hiperparameter Tuning pada Metode Support Vector Machine untuk Klasifikasi Tingkat Kematangan Buah Kopi. *J-Icon: Jurnal Komputer dan Informatika*, 11(2), 195-202.
- Santosa, B., Rosidi, A., Anggraini, H., Latrobdiba, Z. M., Damayanti, F. N., & Nugroho, H. S. (2022). Mask Protection Againts Lead Exposure and Its Correlations with Erythropoiesis in Automotive Body Painters at Ligu District, Semarang, Indonesia. *Jurnal of Blood Medicine*, 4(2), 113-119.
- Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest. *Infotekmesin*, 14(1), 155-162.
- Silitonga, S. S., & Ginting, M. (2022). Hubungan tingkat pengetahuan terhadap kejadian anemia pada wanita usia subur (WUS) di Puskesmas Putri Ayu Kota Jambi Tahun 2021. *Scientia Journal*, 11(1), 354-360.
- Simeone, O. (2018). A Very Brief Introduction to Machine Learning with Applications to Communication Systems. *IEEE Transactions on Cognitive Communications and Networking*, 4(4), 648-664.
- Son, G., & Kim, Y. (2021). EEG-Based Emotion Classification for Verifying the Korean Emotional Movie Clips with Support Vector Machine (SVM). *Complexity*, 2021.
- Sumathi, B. (2020). Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction. *International*

- Journal of Advanced Computer Science and Applications*, 11(9).
- Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 14(4), 1502-1509
- Utami Putri, N., & Redi Susanto, E. (2020). Klasifikasi Jenis Kayu Menggunakan Support Vector Machine Berdasarkan Ciri Tekstur Local Binary Pattern. *CYBERNETICS*, 4(02), 93–100.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437-444.
- Virro, H., Kmoch, A., Vainu, M., & Uuemaa, E. (2022). Random forest-based modeling of stream nutrients at national level in a data-scarce region. *Science of the Total Environment*, 840, 156613.
- Wang, X., Gong, G., Li, N., & Qiu, S. (2019). Detection analysis of epileptic EEG using a novel random forest model combined with grid search optimization. *Frontiers in human neuroscience*, 13, 52.
- Widiawati, R., Hafy, Z., & Liana, P. (2021). Hubungan Kadar Komponen Besi Darah Pendorong Terhadap Kualitas Packed Red Cells (PRC) si UDD PMI Provinsi Sumatera Selatan. *Jurnal Surya Medika (JSM)*, 66-71.
- Worachartcheewan, A., Shoombuatong, W., Pidetcha, P., Nopnithipat, W., Prachayasittikul, V., & Nantasenamat, C. (2015). Predicting metabolic syndrome using the Random Forest method. *Scientific World Journal*, 2015.
- Xie, Z., Liu, Y., He, H.-Y., Li, M., & Zhou, Z.-H. (2023). Weakly Supervised AUC Optimization: A Unified Partial AUC Approach.
- Xu, Q., & Yin, J. (2021). Application of Random Forest Algorithm in Physical Education. *Scientific Programming*, 2021.
- Yang, J., Han, S., & Chen, Y. (2023). Prediction of Traffic Accident Severity Based on Random Forest. *Journal of Advanced Transportation*, 2023. 1-8.
- Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192.
- Zhu, N., Zhu, C., Zhou, L., Zhu, Y., & Zhang, X. (2022). Optimization of the random forest hyperparameters for power industrial control systems intrusion detection using an improved grid search algorithm. *Applied Sciences*, 12(20), 10456.