

PERBANDINGAN ALGORITMA C4.5 DAN SVM DALAM KLASIFIKASI PENYAKIT ANEMIA

Dina Elly Yanti¹, Lizah Framesti², Anita Desiani^{*3}

^{1,2,3}Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sriwijaya

¹dinaeliyanti77@gmail.com, ²lizahframesti@gmail.com, ³anita_desiani@unsri.ac.id

Abstrak

Anemia merupakan keadaan saat hemoglobin pada tubuh tidak dapat berfungsi dengan baik. Anemia berdampak buruk bagi kesehatan, salah satunya pada daya tahan tubuh. Untuk mencegah terjadinya penyakit anemia dapat dilakukan deteksi dini dengan memanfaatkan pendekatan matematika menggunakan data mining. Data mining memiliki metode-metode klasifikasi yang bisa digunakan untuk deteksi dini penyakit anemia. Metode yang bisa digunakan untuk klasifikasi diantaranya yaitu algoritma *Support Vector Machines* (SVM) dan algoritma C4.5. Penelitian ini menerapkan algoritma SVM dan algoritma C4.5 untuk klasifikasi deteksi dini penyakit anemia. Tujuan penelitian ini adalah mendapatkan metode paling tepat antara algoritma SVM dan algoritma C4.5 dalam klasifikasi penyakit anemia. Penelitian ini menerapkan teknik pengujian *percentage split* dan *k-fold cross validation*. Pada *percentage split* dipilih split sebesar 80% sebagai data latih dan 20% sebagai data uji. Pada *k-fold cross validation* dipilih nilai k sebesar 10. Hasil penerapan kedua metode menunjukkan bahwa *k-fold cross validation* bekerja lebih baik dibandingkan *percentage split* dengan persentase nilai akurasi, presisi, dan *recall* lebih tinggi dari masing-masing algoritma. Untuk kinerja kedua algoritma, C4.5 dalam penerapannya bekerja lebih baik dengan nilai akurasi, presisi, dan *recall* secara berturut-turut, yaitu 99.29%, 98.7%, dan 99.69% dibandingkan algoritma SVM. Dari hasil yang diperoleh dapat disimpulkan bahwa algoritma C4.5 dengan teknik pengujian *k-fold cross validation* menghasilkan nilai performa yang paling baik untuk klasifikasi penyakit anemia dibanding algoritma dan teknik pengujian lainnya.

Kata kunci : Data Mining, Algoritma C4.5, SVM, Penyakit Anemia.

1. Pendahuluan

Di Indonesia angka kejadian anemia tergolong relatif tinggi (Priyanto, 2018). Berdasarkan data Riskesdas tahun 2018, prevalensi penyakit anemia mencapai 32% pada remaja. Hal ini berarti terdapat 3-4 dari sepuluh remaja di Indonesia mengalami penyakit anemia. Anemia adalah penyakit yang umum terjadi dikalangan masyarakat. Anemia merupakan keadaan saat eritrosit atau massa hemoglobin (Hb) pada tubuh tidak bisa melakukan fungsinya sebagai penyedia oksigen untuk jaringan tubuh (Prasetya & Wihandani, 2019). Anemia berdampak buruk pada kesehatan reproduksi, daya tahan tubuh, perkembangan motorik, mental, menyebabkan tinggi badan yang tidak maksimal (Adriani & Wirjatmadi, 2016). Berdasarkan dampak yang diakibatkan oleh anemia sangat penting dilakukan deteksi dini untuk mencegah penyakit anemia sejak dini (Hastuty *et al.*, 2021). Menurut dr. Fadhl Rizal Makarim, apabila tidak dilakukan pencegahan, anemia akan menjadi masalah kesehatan jangka panjang dan berakibat fatal hingga kematian.

Dalam melakukan deteksi dini terhadap penyakit anemia dapat menggunakan pendekatan matematika yaitu data mining (Nurmasani & Pristyanto, 2021).

Data mining yaitu sebuah proses dengan teknik statistik, *mechine learning*, matematika serta kecerdasan buatan dalam mengidentifikasi informasi berguna dari berbagai database besar (Febriani & Sulistiani, 2021). Data mining terdiri dari beberapa metode, salah satunya klasifikasi yang dapat dimanfaatkan untuk melakukan deteksi dini (Desiani *et al.*, 2022).

Klasifikasi bekerja melalui pengenalan pola atau model dari sebuah kelas. Klasifikasi bertujuan agar pola tersebut bisa dipakai dalam melakukan prediksi ataupun klasifikasi pada seseorang yang didasarkan pada analisis data latih (Nurmasani & Pristyanto, 2021). Metode klasifikasi yang banyak digunakan dalam deteksi dini yaitu algoritma *Support Vector Machine* (SVM) dan algoritma C4.5 (Rahayu & Purnama, 2022).

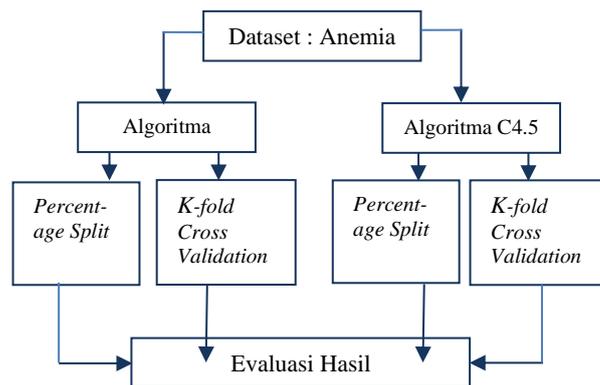
Algoritma SVM memiliki kelebihan yaitu dapat bekerja dengan baik pada data berdimensi tinggi dengan memanfaatkan teknik kernel (Drajana, 2017). Penelitian sebelumnya yang menggunakan algoritma SVM dilakukan oleh Nur Ghaniaviyanto Ramadhan dan Azka Khoirunnisa (Ramadhan & Khoirunnisa, 2021) yang menerapkan metode SVM untuk deteksi dan klasifikasi penyakit malaria. Penelitian tersebut menggunakan dengan teknik pengujian *k-fold cross*

validation dan memperoleh nilai akurasi sebesar 92.3%. Tetapi algoritma SVM memiliki kekurangan yaitu tidak cocok dipakai untuk jumlah sampel yang berskala besar (El Morr *et al.*, 2022). Selain itu algoritma SVM tidak mudah dalam mengakomodasi data hilang sehingga berdampak pada akurasi dari klasifikasi yang dihasilkan (Stewart *et al.*, 2018). Metode lain yang bisa digunakan untuk klasifikasi penyakit anemia yaitu algoritma C4.5.

Berbeda dengan algoritma SVM, Algoritma C4.5 memiliki kelebihan yaitu dapat menangani data hilang dan dapat menangani atribut yang kontinu dan diskrit. Algoritma C4.5 mengolah data dari pohon keputusan menjadi aturan sederhana yang mudah dipahami. Selain itu dalam penerapannya algoritma C4.5 mudah untuk mengetahui atribut paling berpengaruh (Edsel Barito *et al.*, 2022). Penelitian sebelumnya yang menerapkan algoritma C4.5 dilakukan oleh Anita dkk, (Yahdin *et al.*, 2019) dalam penelitian untuk mengetahui pola masa studi mahasiswa jurusan Matematika Universitas Sriwijaya. Penelitian tersebut menggunakan teknik pengujian *k-fold cross validation* dan *percentage split* dengan akurasi hasil yang diperoleh secara berturut-turut sebesar 84.29% dan 79.92%. Kekurangan dari algoritma C4.5 yaitu data mining hanya bisa digunakan dan disimpan pada saat bersamaan. Selain itu, karena algoritma C4.5 menghasilkan pohon keputusan yang kompleks, maka semakin besar data yang digunakan maka memori yang digunakan semakin besar. Sehingga algoritma C4.5 sering mengalami *overfitting* (Sihaan *et al.*, 2020). Pada penelitian ini akan dilakukan perbandingan metode algoritma SVM dan algoritma C4.5 dalam klasifikasi penyakit anemia. Penelitian ini dilakukan dengan teknik pengujian *percentage split* dan *k-fold cross validation*. Penelitian bertujuan untuk mengetahui algoritma mana yang lebih akurat dan mempunyai performa terbaik berdasarkan pengujian akurasi, presisi dan *recall*. Hasil klasifikasi pada penelitian ini menghasilkan dua kelas yaitu anemia dan tidak anemia.

2. Metode

Metodologi penelitian sangat diperlukan dalam penelitian dengan tujuan untuk mendapatkan hasil yang sesuai dengan alur yang telah ditetapkan. Adapun tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

2.1 Dataset Anemia

Data yang digunakan pada penelitian ini yaitu dataset yang didapatkan dari situs Kaggle (<https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset>) dengan menggunakan format csv. Data tersebut merupakan data terkait Penyakit Anemia yang terdiri dari 1.421 data serta memiliki 5 atribut dan satu label yaitu Result dengan 0 berarti tidak anemia dan 1 menyatakan anemia. Sebanyak 801 data tidak mengalami anemia dan 620 data mengalami anemia. Adapun atribut-atribut yang digunakan berikut pada Tabel 1.

Tabel 1. Informasi atribut

Atribut	Keterangan	Tipe Data	Range
Gender	Kategori Gender	Kategori	0=laki-laki, 1=perempuan
Hemoglobin	Parameter dari Hemoglobin	Numerik	6.6 – 16.9
MCH	Parameter dari Mean Corpuscular Hemoglobin (MCH)	Numerik	16 – 30
MCHC	Parameter dari Mean Corpuscular Hemoglobin Concentration (MCHC)	Numerik	27.8 – 32.5
MCV	Parameter dari Mean Corpuscular Volume (MCV)	Numerik	69.4 – 101.6
Result	Kategori klasifikasi Anemia atau tidak Anemia	Kategori	0=tidak anemia, 1=anemia

2.2 Metode Training

1. Percentage Split

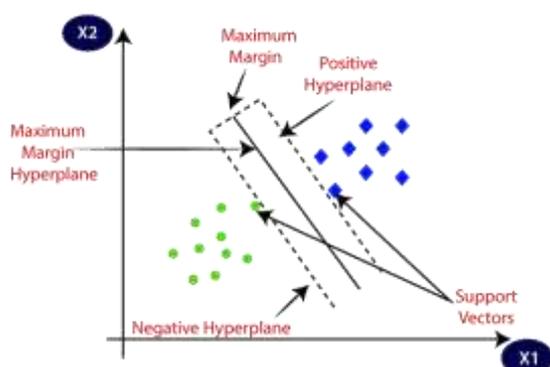
Klasifikasi dengan *percentage split* akan dapat dilakukan dengan pengujian menggunakan k% data set sebagai data latih. Dalam *percentage split* akan dilakukan pembagian terhadap dataset sebagai data latih dan data uji (Primajaya *et al.*, 2020). Adapun *percentage split* yang digunakan pada penelitian ini yaitu 80% dari data menjadi data latih dan 20% menjadi data uji.

2. K-fold Cross Validation

K-fold cross validation yaitu teknik validasi untuk menilai hasil analisis statistika dalam menggeneralisasi kumpulan data (Inan & Uzer, 2021). Pada metode *k-fold cross validation* akan digunakan k=10. Data dibagi 10 bagian yang dilakukan latihan pada suatu bagian dan uji pada data lainnya dan dilakukan secara bergantian.

2.3 Algoritma Support Vector Machine (SVM)

Support Vector Mechine (SVM) merupakan algoritma yang digunakan untuk melakukan pengklasifikasian terhadap suatu objek dengan atribut yang tersedia. Metode ini melakukan pencarian terhadap *hyperplane* yang optimal guna membagi antara dua kelas pada klasifikasi dengan memaksimalkan margin antara dua kelas tersebut (Mahendro & Abimanto, 2022). Berikut merupakan gambar *hyperlane* yang memisahkan dua kelas dapat dilihat pada Gambar 2.



Gambar 2. *Hyperline* SVM yang memisahkan kedua kelas bernilai 0 dan 1

Berdasarkan Gambar 2. Kelas 0 dilambangkan dengan warna hijau sedangkan kelas 1 dilambangkan dengan warna merah. Klasifikasi dengan algoritma SVM dapat dilakukan perhitungan berdasarkan beberapa tahapan berikut (Naufal *et al.*, 2020):

1. Terdapat $\vec{x}_i \in \vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ yang mana x_i adalah data dari n atribut dan dua kelas, $y_i \in +1, -1$.

2. Anggap data linear serta kelas bisa dibagi oleh *hyperplane* yang ditunjukkan persamaan (1).

$$\vec{w} \cdot \vec{x} + b = 0 \tag{1}$$

Berdasarkan persamaan (1) didapatkan persamaan (2) dan persamaan (3) berikut :

$$\vec{w} \cdot \vec{x} + b \geq 1, \text{ untuk kelas } +1 \tag{2}$$

$$\vec{w} \cdot \vec{x} + b \geq -1, \text{ untuk kelas } -1 \tag{3}$$

Keterangan :

\vec{w} = Bobot

b = Bias

\vec{x} = Nilai input

3. Mencari *hyperplane* pemisah dengan memaksimalkan jarak kedua kelas dengan mencari titik minimum, ditunjukkan persamaan (4).

$$\min_w \frac{1}{2} (\|\vec{w}\|)^2 \tag{4}$$

Berdasarkan kendala persamaan (5)

$$y_i (\vec{w} \cdot \vec{x}_i + b) - 1 \geq 0 \tag{5}$$

Keterangan :

\vec{w} = Bobot

b = Bias

\vec{x}_i = Nilai input ke-i

y_i = kelas label ke-i

4. Dalam mengatasi data nonlinear dapat menggunakan karnel yang dapat melakukan transformasi *input space* ke *feature space*. Penelitian ini akan menggunakan *kernel polynomial* seperti persamaan (6).

$$K(\vec{x}, \vec{y}) = (\vec{x} \cdot \vec{y} + c)^d \tag{6}$$

Keterangan :

\vec{x} dan \vec{y} = Vektor dari ruang fitur

d = Derajat

c = Parameter bebas

2.4 Algoritma C4.5

Algoritma C4.5 ialah salah satu metode klasifikasi yang merepresentasi hasil dalam struktur pohon (*tree*) yang mana tiap node mempresentasikan nilai atribut, cabangnya mempresentasikan nilai dari atribut serta daun mempresentasikan kelas (Ginting *et al.*, 2020). Gina Sonia & Indriyani (2023), berpendapat bahwa pohon keputusan algoritma C4.5 dapat dibuat melalui beberapa tahapan diantaranya yaitu:

1. Mengekstrak kelas dari konsep yang diterapkan. Gunakan *entropy* pada ketidakstabilan S, berdasarkan rumus pada persamaan (7) :

$$Entropy(S) = \sum_{i=1}^n (-P_i) \cdot \log_2 P_i ; P_i \neq 0 \tag{7}$$

Keterangan rumus :

S = Himpunan kasus

n = Total partisi S

P = Proporsi S_i oleh S

- Penerapan metode *gain*.
 Dalam penerapan metode *gain* dapat digunakan dengan rumus pada persamaan (8) :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|s_i|}{|s|} Entropy(S_i) \quad (8)$$

Keterangan rumus :

S = Himpunan dari Kasus

A = Fitur

n = Total partisi atribut A

s_i = Total kasus partisi ke- i

s = Total kasus dalam S

- Ulangi dari langkah ke-2 sampai seluruh baris terpisahkan.
- Progres pemisahan pohon keputusan mengalami pemberhentian saat :
 - Baris ke- n memiliki kelas yang sama.
 - Baris tidak memiliki atribut yang dipisahkan.
 - Semua baris dicabang tidak ada yang tidak berisi.

2.5 Evaluasi Hasil

Pengujian akurasi dapat dilakukan dengan *confusion matrix*. *confusion matrix* merupakan matriks yang memiliki fungsi untuk menampilkan visualisasi kinerja dari algoritma, serta digunakan untuk menghitung kinerja perfromansi dari suatu model algoritma dalam suatu prediksi actual dari informasi. Bentuk dari *confusion matrix* ditunjukkan oleh Tabel 2 (Markoulidakis *et al.*, 2021).

Tabel 2. *Confusion Matrix*

Kelas		Nilai Aktual	
		Negative	Positive
Nilai Prediksi	Negative	True Negative (TN)	False Positive (FP)
	Positive	Flase Negative (FN)	True Positive (TP)

Berdasarkan Tabel 2. *True Positive* (TP) yaitu total data positif yang diklasifikasikan positif. *False Negative* (FN) yaitu total data negatif yang diklasifikasikan positif. *False Positive* (FP) yaitu total data positif yang diklasifikasikan negatif. *True Negative* (TN) yaitu total data negatif yang diklasifikasikan negatif.

Ukuran evaluasi kinerja yang digunakan dalam klasifikasi berdasarkan nilai-nilai dalam *confusion matrix*, yaitu akurasi, presisi, dan *recall*. Menurut (Rosandy, 2016) Akurasi merupakan angka yang digunakan untuk memprediksi data yang benar dan salah. Perhitungan nilai akurasi dapat dilihat pada persamaan (9)

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

Presisi merupakan tingkat ketepatan informasi yang diperlukan pengguna berdasarkan jawaban yang diberikan sistem. Perhitungan nilai presisi dapat dilihat pada persamaan (10)

$$Presisi = \frac{TP}{TP+FP} \quad (10)$$

Recall yaitu suatu tingkat keberhasilan sistem untuk menemukan kembali sebuah informasi. Perhitungan nilai *recall* dapat dilihat pada persamaan (11)

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

3. Hasil dan Pembahasan

3.1 Algoritma SVM

Pada penelitian ini kernel yang digunakan yaitu kernel linier dari model algoritma SVM. Untuk mendapatkan nilai akurasi, presisi, dan *recall* terhadap klasifikasi dataset anemia dilakukan perhitungan terhadap *confusion matrix* dengan metode training *percentage split* dan *K-fold cross validation*. Hasil perhitungan *confusion matrix* pada algoritma SVM dapat dilihat pada Tabel 3.

Tabel 3. *Confusion Matrix* algoritma SVM

SVM							
Percentage Split				K-fold Cross Validation			
Kelas		Nilai Aktual		Kelas		Nilai Aktual	
		0	1			0	1
Nilai Prediksi	0	157	14	Nilai Prediksi	0	773	32
	1	3	110		1	8	608

Dari Tabel 3. dapat dilihat *Confusion matrix* pada *percentage split* ada 110 data diprediksi secara benar sebagai kelas anemia. 14 data yang seharusnya masuk kelas anemia tetapi dikenali sebagai kelas tidak anemia. Lalu terdapat 157 data berhasil diklasifikasi secara benar sebagai kelas tidak anemia. Terdapat juga 3 data diprediksi dalam kelas anemia yang seharusnya masuk dalam klasifikasi kelas tidak anemia. *Confusion matrix* pada *k-fold cross validation* ada 608 data diprediksi secara benar sebagai kelas anemia. 32 data yang seharusnya masuk kelas anemia tetapi dikenali sebagai kelas tidak anemia. Lalu terdapat 773 data berhasil diklasifikasi secara benar sebagai kelas tidak anemia. Terdapat juga 8 data diprediksi dalam kelas anemia yang seharusnya masuk dalam klasifikasi kelas tidak anemia.

Selanjutnya dilakukan perhitungan untuk mendapatkan nilai akurasi, presisi dan *recall* dengan menggunakan persamaan (9), (10), (11). Maka didapatkan nilai akurasi, presisi dan *recall* untuk

setiap teknik pengujian yang ditunjukkan oleh Tabel 4.

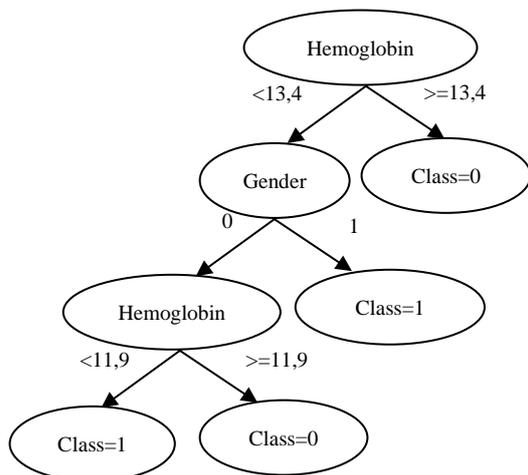
Tabel 4. Perbandingan kedua teknik pengujian pada algoritma SVM

Teknik Pengujian	Akurasi	Presisi	Recall
Percentage Split	94.01%	94.58%	93.41%
K-fold Cross Validation	97.8%	95.74%	98.7%

Berdasarkan Tabel 4. Didapatkan hasil dari kedua teknik pengujian pada algoritma SVM dalam klasifikasi anemia. Teknik pengujian *k-fold cross validation* menghasilkan nilai akurasi, presisi, dan *recall* yang lebih tinggi dibandingkan dengan teknik pengujian *percentage split*. Nilai akurasi, presisi, dan *recall* secara terurut yaitu 97.8%, 95.74% dan 98.7%.

3.2 Algoritma C4.5

Pada perhitungan menggunakan algoritma C4.5 kita terlebih dahulu mencari nilai *gain* tertinggi yang dimiliki oleh data. Dari hasil running program, didapatkan pohon keputusan dengan *gain* tertinggi yaitu hemoglobin. Dimana jika hemoglobin didalam tubuh tinggi, maka resiko terjadinya penyakit anemia lebih rendah dan sebaliknya, jika hemoglobin dalam tubuh rendah maka resiko penyakit anemia lebih tinggi. Berikut pohon keputusan yang didapatkan pada Gambar 3.



Gambar 3. Pohon keputusan *percentage split* dan *k-fold cross validation* algoritma C4.5

Berdasarkan Gambar 3. didapatkan pohon keputusan algoritma C4.5 untuk klasifikasi penyakit anemia. Sehingga diperoleh aturan linguistik dalam menentukan penyakit anemia adalah sebagai berikut :

- IF hemoglobin ≥ 13.4 THEN Result=0, artinya jika hemoglobin lebih besar atau sama dengan 13.4 maka diperoleh hasil tidak.

- IF hemoglobin < 13.4 AND Gender=0 AND Hemoglobin < 11.9 THEN Result=1, artinya jika hemoglobin lebih kecil dari 13.4 dan gender laki-laki dan hemoglobin lebih kecil dari 11.9 maka diperoleh hasil iya.
- IF hemoglobin < 13.4 AND Gender=0 AND Hemoglobin ≥ 11.9 THEN Result=0, artinya jika hemoglobin lebih kecil dari 13.4 dan gender laki-laki dan hemoglobin lebih besar sama dengan 11.9 maka diperoleh hasil tidak.
- IF hemoglobin < 13.4 AND Gender=1 THEN Result=1, artinya jika hemoglobin lebih kecil dari 13.4 dan gender perempuan maka diperoleh hasil iya.

Selanjutnya untuk mendapatkan nilai akurasi, presisi dan *recall* dilakukan perhitungan terhadap klasifikasi dataset anemia Berdasarkan metode training *percentage split* dan *k-fold cross validation*. Hasil perhitungan *confusion matrix* algoritma C4.5 dapat dilihat pada tabel 5.

Tabel 5. Confusion Matrix algoritma C4.5

C4.5							
Percentage Split				K-fold Cross Validation			
Kelas	Nilai Aktual			Kelas	Nilai Aktual		
	0	1			0	1	
Nilai Prediksi	0	160	6	Nilai Prediksi	0	795	8
	1	0	118		1	2	616

Dari Tabel 5. dapat dilihat *Confusion matrix* pada *percentage split* ada 118 data diprediksi secara benar sebagai kelas anemia. 14 data yang seharusnya masuk kelas anemia tetapi dikenali sebagai kelas tidak anemia. Lalu terdapat 160 data berhasil diklasifikasi secara benar sebagai kelas tidak anemia. Terdapat juga 0 data dikenali sebagai kelas anemia yang seharusnya masuk dalam klasifikasi kelas tidak anemia. *Confusion matrix* pada *k-fold cross validation* ada 616 data diprediksi secara benar sebagai kelas anemia. 8 data yang seharusnya masuk kelas anemia tetapi dikenali sebagai kelas tidak anemia. Lalu terdapat 795 data berhasil diklasifikasi secara benar sebagai kelas tidak anemia. Terdapat juga 2 data dikenali sebagai kelas anemia yang seharusnya masuk dalam klasifikasi kelas tidak anemia.

Selanjutnya untuk mendapatkan nilai akurasi, presisi dan *recall* dilakukan perhitungan dengan menggunakan persamaan (9), (10), (11). Maka didapatkan nilai akurasi, presisi dan *recall* untuk setiap teknik pengujian yang ditunjukkan oleh Tabel 6.

Tabel 6. Perbandingan kedua teknik pengujian pada algoritma C4.5

Teknik Pengujian	Akurasi	Presisi	Recall
<i>Percentage Split</i>	97.89%	95.16%	100%
<i>K-fold Cross Validation</i>	99.29%	98.7%	99.67%

Berdasarkan Tabel 6. Didapatkan hasil dari kedua teknik pengujian pada algoritma C4.5 dalam klasifikasi anemia. Teknik pengujian *k-fold cross validation* menghasilkan nilai akurasi, presisi, dan *recall* yang lebih tinggi dibandingkan dengan teknik pengujian *percentage split*. Nilai akurasi, presisi, dan *recall* secara terurut yaitu 99.29%, 98.7% dan 98.67%.

3.3 Perbandingan Hasil Kedua Metode

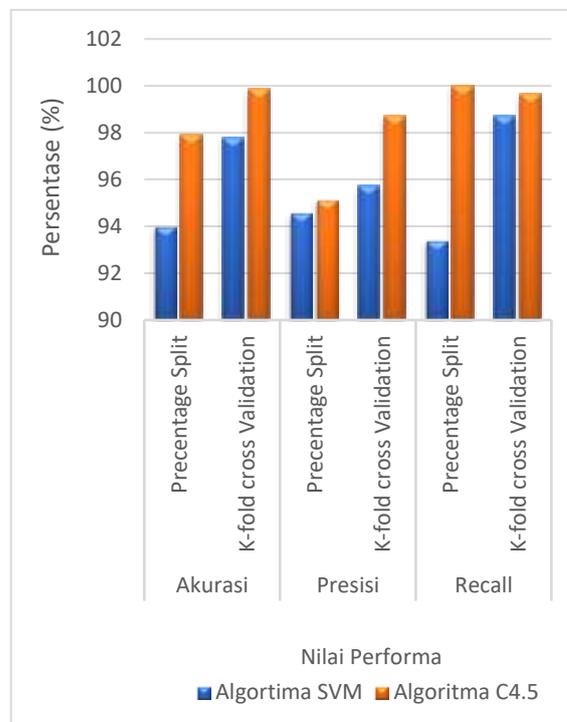
Setelah dilakukan pemodelan menggunakan algoritma SVM dan algoritma C4.5 dengan menggunakan teknik pengujian *percentage split* dan *k-fold cross validation* didapatkan perbandingan hasil dari nilai akurasi, presisi dan *recall* pada kedua algoritma yang ditunjukkan oleh Tabel 7.

Tabel 7. Perbandingan kedua algoritma dengan *Percentage Split* dan *K-fold Cross Validation*

Teknik Pengujian	<i>Percentage Split</i>		<i>K-fold Cross Validation</i>	
	SVM	C4.5	SVM	C4.5
Akurasi	94.01%	97.89%	97.8%	99.29%
Presisi	94.58%	95.16%	95.74 %	98.7%
Recall	93.41%	100%	98.7%	99.69%

Dari Tabel 7 dapat dilihat pada teknik pengujian *percentage split* algoritma C4.5 menghasilkan nilai akurasi, presisi, dan *recall* lebih tinggi dibanding algoritma SVM dengan nilai masing-masing berturut 97.89%, 95.16% dan 100%. Sedangkan algoritma SVM hanya menghasilkan nilai akurasi, presisi, dan *recall* masing-masing berturut 94.01%, 94.58% dan 93.41%. Pada teknik pengujian *k-fold cross validation* algoritma C4.5 menghasilkan nilai akurasi, presisi, dan *recall* lebih tinggi dibanding algoritma SVM dengan nilai masing-masing berturut 99.29%, 98.7% dan 99.69%. Sedangkan algoritma SVM hanya menghasilkan nilai akurasi, presisi, dan *recall* masing-masing berturut 97.8%, 95.74% dan 98.7%. Berdasarkan hasil kedua teknik pengujian metode algoritma C4.5 memiliki tingkat akurasi, presisi dan *recall* yang lebih tinggi dibandingkan algoritma SVM. Teknik pengujian *k-fold cross validation*

menghasilkan nilai yang lebih tinggi dari teknik pengujian *percentage split* untuk kedua metode klasifikasi. Untuk mempermudah dalam membaca hasil perbandingan kedua metode dapat dilihat pada Gambar 4.



Gambar 4. Diagram perbandingan algoritma SVM dan algoritma C4.5

Berdasarkan Gambar 4. Diketahui bahwa algoritma C4.5 memiliki persentase untuk setiap nilai akurasi, presisi, dan *recall* yang lebih tinggi dibandingkan dengan algoritma SVM. Dapat dilihat juga bahwa teknik pengujian *k-fold cross validation* memiliki persentase lebih tinggi dibandingkan *percentage split* untuk masing-masing algoritma.

4. Kesimpulan

Berdasarkan penelitian ini dapat disimpulkan bahwa klasifikasi penyakit anemia dengan menggunakan algoritma SVM dan algoritma C4.5 tergolong baik. Hal ini terlihat dari akurasi, presisi, dan *recall* kedua metode. Dalam metode training *percentage split* algoritma SVM menghasilkan akurasi sebesar 94.01% lebih kecil dibanding algoritma C4.5 yang menghasilkan akurasi sebesar 97.89%. Sedangkan, Dalam metode training *K-fold Cross validation* algoritma SVM menghasilkan akurasi sebesar 97.8% lebih kecil dibanding algoritma C4.5 yang menghasilkan akurasi sebesar 99.29%. Berdasarkan kedua metode training yang dilakukan, dapat disimpulkan bahwa metode *k-fold cross validation* menghasilkan nilai akurasi lebih besar dan algoritma C4.5 menunjukkan nilai performa yang lebih baik dibanding algoritma SVM dalam klasifikasi penyakit anemia.

Daftar Pustaka:

- Desiani, A., Lestari, A. A., Al-Ariq, M., Amran, A., & Andriani, Y. (2022). Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification. *Pattimura International Journal of Mathematics (PIJMath)*, 1(1), 33–42. <https://doi.org/10.30598/pijmathvol1iss1pp33-42>
- Drajana, I. C. R. (2017). Metode Support Vector Machine Dan Forward Selection Prediksi Pembayaran Pembelian Bahan Baku Kopra. *ILKOM Jurnal Ilmiah*, 9(2), 116–123. <https://doi.org/10.33096/ilkom.v9i2.134.116-123>
- Edsel Barito, E., Tji Beng, J., & Arisandi, D. (2022). Penerapan Algoritma C4.5 Untuk Klasifikasi Mahasiswa Penerima Bantuan Sosial Covid-19. *Ilmu Komputer Dan Sistem Informasi*, 10(1). <https://doi.org/https://doi.org/10.24912/jiksi.v10i1.17819>
- El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). Machine Learning for Practical Decision. 385-411. *International Series in Operations Research & Management Science*.
- Febriani, S., & Sulistiani, H. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(4), 89–95. <https://doi.org/10.33365/jtsi.v2i4.1373>
- Gina Sonia, & Indriyani, A. (2023). Analisis Penerapan Data Mining Dengan Metode Algoritma C4.5 Untuk Pendataan Karyawan Tetap Di Koni Sumatera Utara. *Journal of Computer Science and Informatics Engineering (CoSIE)*, 02(1), 48–55. <https://doi.org/10.55537/cosie.v2i1.523>
- Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44. <https://doi.org/10.35585/inspir.v10i1.2535>
- Hastuty, Y. D., Khodijah, D., & Hasibuan, Y. (2021). Edukasi Dan Deteksi Dini Anemia Remaja Putri Di Pancur Batu Kabupaten Deli Serdang. *GEMAKES Jurnal Pengabdian Kepada Masyarakat*, 1(2), 70–82. <https://doi.org/10.36082/gemakes.v1i2.384>
- Inan, O., & Uzer, M. S. (2021). A Method of Classification Performance Improvement Via a Strategy of Clustering-Based Data Elimination Integrated with k-Fold Cross-Validation. *Arabian Journal for Science and Engineering*, 46(2), 1199–1212. <https://doi.org/10.1007/s13369-020-04972-y>
- Mahendro, I., & Abimanto, D. (2022). Analisa Kepuasan Mahasiswa Terhadap E-Learning Menggunakan Algoritma Support Vector Machine. *Jurnal Sains dan Teknologi Maritim (JSTM)*, 23(1), 97–108. <http://dx.doi.org/10.33556/jstm.v23i1.333>
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem. *Technologies*, 9(4), 412–419. <https://doi.org/10.1145/3453892.3461323>
- Naufal, S. A., Adiwijaya, A., & Astuti, W. (2020). Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 162. <https://doi.org/10.30865/jurikom.v7i1.2014>
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Jurnal Pseudocode*, 8(1), 21–26. <https://doi.org/10.33369/pseudocode.8.1.21-26>
- Prasetya, K. A. H., & Wihandani, D. M. (2019). Hubungan Antara Anemia Dengan Prestasi Belajar Pada Siswi Kelas Xi Di Sman I Abiansemal Badung. *E-Jurnal Medika Udayana*, 8(1), 46. <https://doi.org/10.24922/eum.v8i1.45757>
- Primajaya, A., Sari, B. N., & Khususari, A. (2020). Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 6(2), 188. <https://doi.org/10.26418/jp.v6i2.37834>
- Priyanto, L. D. (2018). Hubungan Umur, Tingkat Pendidikan, dan Aktivitas Fisik Antriwati Husada Dengan Anemia. *Jurnal Berkala Epidemiologi*, 6(2), 139. <https://doi.org/10.20473/jbe.v6i22018.139-146>
- Rahayu, S., & Purnama, J. J. (2022). Klasifikasi Konsumsi Energi Industri Baja Menggunakan Teknik Data Mining. *Jurnal Teknoinfo*, 16(2), 395. <https://doi.org/10.33365/jti.v16i2.1984>
- Ramadhan, N. G., & Khoirunnisa, A. (2021). Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 5(4), 1580. <https://doi.org/10.30865/mib.v5i4.3347>
- Siahaan, S. W., Sianipar, K. D. R., R.H Zer, P. P. P. A. N. . F. I., & Hartama, D. (2020). Penerapan Algoritma C4.5 Dalam Meningkatkan Kemampuan Bahasa Inggris Pada Mahasiswa. *Petir*, 13(2), 229–239. <https://doi.org/10.33322/petir.v13i2.1029>
- Stewart, T. G., Zeng, D., & Wu, M. C. (2018).

Constructing support vector machines with missing data. *WIREs Computational Statistics*, 10(4), 1–27. <https://doi.org/10.1002/wics.1430>

Yahdin, S., Desiani, A., Amran, A., Rodiah, D., & Solehan. (2019). Pattern Recognition for Study Period of Student in Mathematics Department with C4.5 Algorithm Data Mining Technique at the Faculty of Mathematics and Natural Science Universitas Sriwijaya. *Journal of Physics: Conference Series*, 1282(1), 1–6. <https://doi.org/10.1088/1742-6596/1282/1/012014>